Estimating Regional Price Parities Using New Data on Medical Goods and Services

Authors James P. Choy, Bureau of Economic Analysis

Contact James.Choy@bea.gov

Date July 2021

Abstract I report on a project to improve estimates of Regional Price Parities (RPPs) by making use of a large commercial dataset on health care prices, the Health Marketscan dataset. Using the new data, I obtain estimates of regional price levels for health related goods and services that are stable across years and that vary less across regions than existing estimates obtained using CPI data.

Keywords Regional Price Parities; Health Care Cost Index

JEL Code E31

Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

1. Introduction

The Bureau of Economic Analysis (BEA) regional price parities (RPP) project produces relative price indexes for U.S. states and metropolitan statistical areas. Constructing these indexes requires a large quantity of price data that covers a wide range of goods and services in each of the measured regions. These data would ideally be sufficiently granular to allow price comparisons between identical or nearly identical goods and services across regions, in order to ensure that measured price differences are not due to differences in quality, but that they are also sufficiently comprehensive to be representative of a typical consumer's entire consumption bundle. Finding data with these properties is a challenge.

Currently, the RPPs are estimated using the same data used by the Bureau of Labor Statistics (BLS) to produce the Consumer Price Index (CPI), supplemented with data from the American Consumer Survey (ACS) for rents.[1] These data have the advantage of covering a representative basket of consumer goods and services. However, the data collection process is designed to measure price differences across time, not across space. In particular, while the CPI compares the prices of identical or nearly identical goods at different times, the CPI data collection process is not designed to collect prices of identical or nearly identical goods in different regions. Thus, in order to construct spatial price indexes, it is necessary to correct for differences in the characteristics of the goods and services whose prices are observed in different regions. This correction is performed using hedonic methods. In addition, the CPI does not always have a large number of observations for each combination of area and good or service category; in some cases, the price of a given category of goods in an area must be estimated from fewer than ten observations. The small number of observations reduces the reliability of the estimates for some area-category combinations.

In order to address these problems, the staff of the Regional Price Branch at BEA have begun to explore alternative sources of data for use in generating RPPs. In this note, I present a preliminary report on our efforts to use Health Marketscan data, produced by Truven Health Analytics, to substitute for CPI data on medical goods and services. Relative to CPI data, Health Marketscan data have two advantages. First, the dataset contains a much more granular range of price-determining characteristics, which in some cases allows for comparisons between identical or nearly identical goods or services across areas. Even when comparisons between identical or nearly identical goods or services are not possible, the wider range of price-determining characteristics included in the Health Marketscan data increase the credibility of hedonic corrections for unobserved quality differences. Second, the Health Marketscan dataset contains a much larger sample than the CPI dataset, with hundreds of millions of observations rather than a few thousand observations for the medical categories in the CPI. This larger sample increases the precision and stability of the estimates. A disadvantage of the Health Marketscan data is that the dataset is drawn from a convenience sample and is not nationally representative.

1. The BEA Regional Price Parity statistics are based in part on restricted access Consumer Price Index data from the Bureau of Labor Statistics. The BEA statistics expressed herein are products of BEA and not BLS.

In the following analysis, I describe the Health Marketscan data and produce RPP estimates for U.S. states in 2017 using the new data. I then discuss some of the potential problems with the Health Marketscan data and the extent to which these problems may reduce the value of the estimates.

2. Related work

Two recent papers have attempted to estimate regional price variation in medical goods and services. Dunn, Shapiro, and Liebman (2013) create an index to study variation in the cost of treating a given disease across geographic areas. This index is conceptually distinct from the index in this paper, which is meant to capture variation in the cost of a given medical good or service across geographical areas. Typically, many different goods and services, sometimes spread across multiple medical facility visits, are used to treat a single instance of a disease. Dunn, Shapiro, and Liebman use the same dataset used in this paper. Cooper, Craig, Gaynor, and Van Reenen (2019) develop an index that is conceptually similar to the index in this paper, albeit using different data, to study differences in health prices across Hospital Referral Regions. They use their index to study the effect of market structure on medical pricing across regions. Relative to Cooper et. al., the main contribution of this paper is to use the medical price index as part of the larger RPP index that includes prices for all consumer goods and services.

3. Data

The Health Marketscan Commercial Claims and Encounters dataset includes medical claims data for all employees covered by approximately 350 different private-sector insurers, including large employers, health plans, and government and private organizations. The dataset covers both active employees and early retirees who are not covered by Medicare, as well as dependents of those employees. The sample of payers is a convenience sample and is not statistically representative of the nation as a whole or any other larger population. Among other things, the dataset does not include care paid for through Medicare or Medicaid, or care provided to the uninsured. Since it is not clear how payers are chosen for inclusion in the sample, the criteria for inclusion may change from year to year.

The Health Marketscan data are divided into several sub-datasets, three of which are used to construct RPP estimates. The first dataset contains observations of claims for coverage of inpatient services, where inpatient services are services which require the patient to stay overnight at the medical facility. The second dataset contains observations of claims for outpatient services, which are services that do not require the patient to stay overnight. The third dataset is a dataset of claims for coverage of prescription drugs.

The inpatient and outpatient datasets are further divided into claims for coverage of facility charges and for coverage of physician charges. Facility charges are fees charged by a hospital or other medical facility. They include charges for items like room and board at a hospital, as well as charges for services performed by hospital employees, for example many services performed by nurses. Physician charges are fees for services performed by physicians who are not directly employed by a hospital or other medical facility.

A single observation in each dataset is a charge for a single medical service. For example, a 20-minute examination by a doctor, a blood test, and a knee replacement operation are all distinct services. In a typical visit to a medical facility, a patient receives multiple services, including both facility and physician services. Many medical facilities do not itemize their facility charges in a meaningful way and instead bill a single facility charge for each visit, perhaps depending on the characteristics of the visit such as the number of days spent at the facility and the diagnosis of the patient. For this reason, a visit to a medical facility is the meaningful unit of observation for facility charges. Truven aggregates inpatient data into visits by summing all the charges that are associated with a given visit. These aggregates include separate totals for facility and physician charges, so it is possible to construct a dataset containing only the facility charges associated with each inpatient visit. Truven does not aggregate outpatient data into visits. Thus, I aggregate this data by summing all the facility charges incurred by a given patient on a given day into a single visit.

Table 1 summarizes the five types of data and shows the number of observations for each type of data in 2017, after dropping observations from Puerto Rico and South Carolina, and for observations from the state in which the observation occurs is unknown. Observations from South Carolina are not reported due to privacy restrictions imposed by the Health Marketscan team. Fewer than 0.01 percent of observations are from Puerto Rico for any of the five types of data. However, there are a significant number of observations from the state in which the observation occurs is unknown. The state is unknown for 14.2 percent of inpatient facility charges, 11.5 percent of outpatient facility charges, 12.6 percent of inpatient physician service observations, 15.0 percent of outpatient physician service observations, and 18.1 percent of prescription drug observations. The unit of observation for facility charges is a visit to a medical facility, while the unit of observation for physician charges is a single service. A patient may incur multiple physician charges in a given visit to a medical facility. The unit of observation for prescription drugs is a claim for a specific drug in a specific quantity/package. Table 1 also shows the correspondence between Health Marketscan categories and the categories used to organize CPI data. Inpatient and outpatient facility charges together correspond to the category of hospital services used by the CPI. Inpatient and outpatient physician charges together correspond to the category of physician services used by the CPI. Prescription drugs are a separate category in both the Health Marketscan and the CPI data.

Table 1. Number of Observations by Category, 2017 Data

Dataset	Charge type	Unit of observation	Number of observations	Corresponding BLS-CPI ELI
Inpatient	Facility	Facility visit	1,063,898	Hospital services
	Physician	Individual service	14,262,651	Physician services
Outpatient	Facility	Facility visit	26,281,938	Hospital services
	Physician	Individual service	368,835,206	Physician services
Prescription Drugs	N/A	Individual drug package	172,940,126	Prescription drugs

BLS Bureau of Labor Statistics
CPI Consumer Price Index
ELI Entry-level items

Note. In inpatient data, a “facility visit” consists of all the facility charges incurred by a given patient during a given stay in a medical facility. In outpatient data, a “facility visit” consists of all the facility charges incurred by a given patient on a given day.

Table B.1 shows the number of observations by type in each state in 2017 (table B.1 and all subsequent tables are in Appendix B below). Table B.1.a shows the same data as percentages of the total number of observations by type. These tables highlight the nonrepresentative nature of the sample, as the numbers of observations in the dataset are clearly not proportional to each state’s population. (Texas, and southern states more generally, are heavily overrepresented.) However, it also shows an advantage of the Health Marketscan data, that there are a large number of observations in every state and category.

It is useful to compare the Health Marketscan dataset to the CPI data which is currently used to estimate RPPs. The most important difference between the two datasets is their size—while table 1 shows that the Health Marketscan data contain hundreds of millions of observations in each year, the CPI medical categories only contain a few thousand observations for each time period. (RPPs estimated using the CPI are estimated on 5-year rolling aggregates of yearly surveys.) There are also some differences in the samples of prices collected by each survey. As mentioned above, Health Marketscan data does not include goods and services paid for by Medicare or Medicaid, or goods and services paid for with cash. In contrast, the CPI dataset heavily oversamples goods and services paid for with cash. The CPI also oversamples charges for physician services provided in office settings, while the Health Marketscan physician charge dataset is a more representative sample of charges for physician services provided in a variety of settings including offices, hospitals, and other health care facilities. Finally, as is discussed in more detail below, the two datasets contain different sets of potential control variables.

4. Estimates

I combine the Health Marketscan with CPI data to construct RPP estimates for states for 2017. The first step in constructing RPP estimates is to run hedonic regressions of the form

$$p_{iac} = \beta X_{iac} + \gamma_{ac} + \varepsilon_i$$

Here p_{iac} is the log price of observation i in area a for goods or services category c, X_{iac} is a vector of price determining characteristics, γ_{ac} is a fixed effect for area a and goods or services category c, and ε_i is an error term. The object of interest is γ_{ac}, which is a measure of the price level of goods in category c in area a.

I am interested in price level estimates for three categories of goods as defined by the BLS, namely hospital services, physician services, and prescription drugs. In 2017 these categories represented, respectively, 2.3 percent, 1.7 percent, and 1.4 percent of total consumer expenditure according to weights calculated from CPI data, so that collectively the categories covered by the Health Marketscan data represent 5.4 percent of total consumer expenditure. In the existing RPP estimates, the expenditure class of medical goods and services also includes the categories of dental services, eyeglasses and eyecare, services by other medical professionals, and nonprescription drugs, which collectively represent 1.8 percent of consumer expenditure. Thus, the new data cover 75 percent of spending in the medical expenditure class.

Hospital services correspond to the category of facility charges in the Health Marketscan data, physician services correspond to the category of physician charges in the Health Marketscan data, and prescription drugs correspond (obviously) to the category of prescription drugs in the Health Marketscan data. Since facility charges and physician charges are distributed across two different datasets in the Health Marketscan data, I aggregate these datasets to construct a dataset of facility charges and a dataset of physician charges. That is, the facility charges dataset contains both inpatient and outpatient facility charges, and the physician charges dataset contains both inpatient and outpatient physician charges. These two datasets then correspond respectively to the categories of hospital services and physician services as defined in the CPI.

In the physician services regression, the most important price-determining characteristic is the procedure code. A procedure code is a code, standardized across facilities, that identifies medical services at a very fine level. For example, a 20-minute and a 45-minute examination by a doctor are distinct procedure codes. There are 13,651 distinct procedure codes in the Health Marketscan data. I include a dummy for each procedure code in my regressions, so that the state price level dummies compare prices only within procedure codes across states. This implies that the services compared by the regression are nearly identical. In order to control for any remaining quality variation within procedure codes across states, I also include controls for a number of other price-determining characteristics, including provider specialty and the type of facility in which the service is provided. These controls help to adjust for differences in quality. The control variables and brief descriptions are listed in table B.2. To save space I do not report the coefficients on the variables; however, all control variables enter significantly into all regressions. The third column of table B.2 gives the number of distinct values for each control variable in the data.

In contrast to the Health Marketscan dataset, the CPI dataset does not contain procedure codes. As a result, price estimates using the CPI compare the prices of different procedures, which likely introduces bias into estimates obtained using CPI data. The fact that the Health Marketscan data contain procedure codes is a major advantage of this dataset over the CPI. On the other hand, the CPI contains data on the name of the insurance company that pays for a given procedure, while the Health Marketscan data does not contain this information. The name of the insurance company may be an important price-determining characteristic if some insurance companies are better able to negotiate low prices than others. The lack of data on the name of the insurance company is one disadvantage of the Health Marketscan data relative to the CPI. Other than the procedure codes and the name of the insurance company, the control variables available in the Health Marketscan data and in the CPI data are similar.

In the hospital services regression, there is nothing that corresponds to the procedure code in the physician services regression, since the unit of observation in the hospital services regression is a patient visit and not an individual service. I include controls for additional price determining characteristics, but these controls do not contain as much detailed information as the procedure code. Thus, the hospital services regression is more likely than the physician services regression to suffer from bias due to unobserved differences in quality across states. Table B.3 shows the control variables used in the hospital services regression.

Three variables used in the physician services regression, PAIDNTWK (whether a claim is paid in-network), STDPLAC (type of health facility) and STDPROV (type of provider) are not included in the hospital services regression because these variables are not necessarily constant within a single visit. A patient may be treated by multiple different specialists, some of whom may be in network and while others are not and may also be moved from one type of facility to another during a visit.

One control variable that is contained in the Health Marketscan hospital services data, but not in the CPI hospital services data, is the number of days spent in hospital for a given visit. The price of a hospital visit is strongly correlated with the number of days in hospital, and so the existence of this variable is an advantage of the Health Marketscan data over the CPI. On the other hand, as in the physician services data, the CPI hospital services data contain the name of the insurer that pays for care, while the Health Marketscan data does not. This is an advantage of the CPI data. Otherwise, the available control variables are similar in the two datasets.

In the prescription drug regression, the most important price determining characteristic is the national drug code, a number that identifies the manufacturer, product name, and package size of a drug. There are 38,489 distinct national drug codes in the Health Marketscan data. The inclusion of this variable ensures that the prescription drug regression compares essentially identical goods across states. Table B.4 shows the control variables used in the prescription drug regression.

The national drug code is not contained in the CPI prescription drug dataset, which is a disadvantage of the CPI data relative to the Health Marketscan data. The CPI prescription drug data does not contain any important variables that are not contained in the Health Marketscan data.

Table B.5 shows state coefficients from each of the three regressions, normalized so that the geometric mean of all state coefficients is equal to one.

For physician services, the most expensive states are Alaska, North Dakota, and Wisconsin. The least expensive states are Kentucky, Arizona, and Rhode Island. For hospital services, the most expensive states are Nevada, Texas, and Alaska. The least expensive states are Alabama, Rhode Island, and Michigan. For prescription drugs, the most expensive states are Hawaii, Colorado, and Minnesota. The least expensive states are Alabama, New Hampshire, and Georgia.

It is interesting to compare price levels estimated from hedonic regressions using Health Marketscan data to the analogous results using CPI data. Table B.6 shows price levels estimated using CPI data. The relevant regions for the CPI data are CPI areas, not states, and so it is not possible to compare area estimates directly. However, it is possible to compare the amount of price variation across areas in the two datasets. The estimates using CPI data show much more variation in price levels across areas. For example, in hospital services, using CPI data we estimate that the highest priced area has price levels 3.41 times the national average, and the lowest priced are has price levels 0.39 times the national average, so that there is price difference of nearly nine-fold between the lowest and highest priced areas. In contrast, there is only about a three-fold difference in hospital service prices between the lowest priced state and the highest priced state in the Health Marketscan data. Similar results appear for physician services and prescription drugs. The amount of price variation across areas in the Health Marketscan data is especially low for prescription drugs, which is intuitive since prescription drugs can easily be shipped from one area to another. In contrast, the high amounts of price variation across areas in the CPI data, especially in prescription drugs, may be harder to explain. The lower level of price variation across areas using the Health

Marketscan data may be a reason to prefer the Health Marketscan estimates to the CPI estimates, if the levels of price variation across areas in the CPI data are deemed implausibly large.

I use the Health Marketscan estimates to construct RPPs for 2017, following the methodology in Aten (2007), Aten, Figueroa, and Martin (2011), Aten (2017), and Aten (2019). I omit the details of the RPP estimation procedure, referring readers to the original source. The novel aspect of the present paper is that I replace hedonic regression estimates of prices for hospital services, physician services, and prescription drugs derived from the CPI with hedonic regression estimates of prices for hospital services, physician services, and prescription drugs derived from Health Marketscan data. This replacement process follows the process used by Aten, Figueroa, and Martin (2011) to replace rent data derived from the CPI with rent data derived from the American Community Survey (ACS). There is one difference, namely that in Aten, Figueroa, and Martin (2011), ACS data are used to calculate both expenditure weights and price levels, while in this paper the new Health Marketscan data is used only to estimate price levels, with expenditure weights still derived from CPI data. Note that expenditure weights are calculated separately for each state. Table B.7 compares the RPPs estimated using the Health Marketscan data to the RPPs estimated using only CPI and ACS data. The differences between the estimates with and without Health Marketscan data are small, as can be seen by looking at column 3 of table B.7, with a maximum 2 percent difference between the RPPs estimated with and without Health Marketscan data. This is unsurprising given that medical spending is a relatively small fraction of total spending.

RPPs can also be estimated for expenditure categories. Table B.8 compares such medical goods and services category RPP estimates with and without Health Marketscan data. Here the new data make more of a difference, as can be seen by looking at column 3 of table B.8. Medical goods and services appear to be 30 percent cheaper in Michigan using the new data, while medical goods and services in Texas appear to be 15 percent more expensive.

It is interesting to note in table B.8 that the variance of medical RPPs across states is higher using Health Marketscan data than using CPI data, even though tables B.5 and B.6 show the variance of each component of the medical RPP index is higher using the CPI data than using the Health Marketscan data. One possible explanation for this result is that if the prices of the different components of the medical RPP index are correlated, and if there is more measurement error in the estimates using CPI data than in the estimates using Health Marketscan data, then using the Health Marketscan data may increase the measured correlation between the different components of the medical RPP index and hence the variance of measured RPPs across states. See Appendix A below for a more detailed illustration of how this possibility could occur.

4.1 Estimates from previous years

As a robustness check, I run first stage regression estimates for 2014 through 2016 to compare to the estimates from 2017. I report the estimates from 2014 through 2017 for physician services, hospital services, and prescription drugs respectively in tables B.9 through B.11. The last column of each estimates shows the range of estimates over the four years divided by the mean of the estimates over the four years. In most states, the price estimate varies by less than 10 percent over the four years. It is worth emphasizing that the data in each year are collected independently. Nevertheless, the results are quite stable across time, which increases confidence in the validity of the results.

4.2 Is the nonrepresentativeness of the Health Marketscan data a problem?

One difficulty with Health Marketscan is that the Health Marketscan data is not designed to be nationally representative. In this section I discuss the extent to which this is a problem.

Nonrepresentativeness could affect the RPP estimates if some characteristics of goods or services have heterogeneous effects on prices in different states, and if the distribution by characteristic in the sample differs from the distribution by characteristic in the population. For example, suppose that Texas law advantages HMOs over PPOs, so that HMOs in Texas are relatively better at bargaining for low prices than HMOs in other states, and suppose that the Health Marketscan data oversamples HMO members relative to PPO members in Texas. Then the estimated medical price index in Texas is too low relative to the Texas population average price of medical goods and services. This is true even after including a control for health plan type.

In order to give some indication of the size of the bias due to nonrepresentativeness, I perform the following test. In the prescription drug regression, I reweight the observations by health plan type (variable PLANTYP) so that the weighted distribution of observations by health plan type is the same in every state. Table B.12 compares coefficient estimates in the original regression and in the reweighted regression.

Reweighting observations does have an effect on the estimated coefficients. However, this effect is relatively small. The largest change is in California, where the reweighting causes a 6 percent change in the estimated price index. This suggests that nonrepresentativeness of the Health Marketscan data in terms of health plan type does not cause a large bias in the area estimates for prescription drugs. Of course, this is only one dimension of the nonrepresentativeness problem, and it may be useful to perform further tests. Tables B.13 and B.14 report the results of a similar reweighting for hospital services and physician services. The broad patterns in tables B.13 and B.14 are similar to the pattern in table B.11. The change due to reweighting is somewhat larger in the hospital services regression, causing a maximum 11 percent change in the estimated price index in North Dakota, probably because the control variables in the hospital services regression do not explain as much of the variation in hospital prices and so there is greater potential for bias. This suggests that nonrepresentativeness may be more of a problem in hospital services than in the other regressions.

4.3 Other outstanding issues

There are two other outstanding issues with the estimates presented in this paper. First, due to computational limitations, I am unable to calculate marginal effects of state dummies in hedonic regressions. Thus, as discussed above, I use the geometric mean of state dummy coefficients to estimate the normalized price in each state. In this respect, my methodology differs from Aten, Figueroa, and Martin (2011). One way to reduce the computational burden of calculating marginal effects would be to use a smaller number of procedure codes and national drug codes in the physician services and prescription drug regressions.

The other outstanding issue is that I have not attempted to identify or remove outliers. There are two kinds of outliers in the data. First, the data include many observations whose prices are negative, zero, or very small positive numbers. These are essentially accounting fictions which are used to adjust previous claims. Zero and negative values naturally fall out of the dataset when I take logarithms of prices, but small positive prices that do not represent real transactions need to be removed from the data. The data also contain some very large outlier prices. However, it is not feasible to remove these values by hand, as is done for the existing RPP estimates, due to the large size of the dataset. It will be necessary to develop an automated procedure to identify and remove these outliers.

5. Conclusion

This paper is a first step towards using novel "big data" sources to estimate regional differences in prices for medical goods and services. Using a large dataset—the Health Marketscan dataset collected by Truven Health Analytics—I estimate RPPs for U.S. states. Estimates of state price levels for each major component of the medical goods and services category (hospital services, physician services, and prescription drugs) using the new data seem reasonable and are stable across years. The estimated variation in price levels across regions for each category is lower when using Health Marketscan data than when using CPI data. The addition of the new dataset makes relatively little difference to the estimated overall RPPs for all goods, which is not surprising since medical goods and services only account for a small fraction of total expenditure. However, the new dataset more significantly changes the estimates of the medical category RPPs. These results suggest that big data sources can be a useful addition to RPP calculations.

References

Aten, Bettina H. "Estimates of State Level Prices for Consumption Goods and Services: A First Brush," Bureau of Economic Analysis (BEA) working paper, WP2007-6 (Washington, D.C.: BEA November 2007), https://www.bea.gov/research/papers/2007/estimates-state-price-levels-consumption-goods-and-services-first-brush

Aten, Bettina H. 2017. "Regional Price Parities and Real Regional Income for the United States," *Social Indicators Research* 131, no. 1 (March):123–143.

Aten, Bettina H. 2019. "Regional price parities in the United States," in Andrew Haughwout and Benjamin Mandel, eds., *Handbook of US Consumer Economics*, London: Academic Press.

Aten, Bettina H., Eric Figueroa, and Troy Martin. "Notes on Estimating the Multi–Year Regional Price Parities by 16 Expenditure Categories: 2005–2009," BEA working paper (Washington, D.C.: BEA, May 2011) https://www.bea.gov/research/papers/2011/notes-estimating-multi-year-regional-price-parities-16-expenditure-categories

Dunn, Abe, Adam Hale Shapiro, and Eli Liebman. 2013. "Geographic Variation in Commercial Medical-Care Expenditures: A Framework for Decomposing Price and Utilization," *Journal of Health Economics* 32, no. 6 (December): 1153–1165.

Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John van Reenen. 2019. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured," *Quarterly Journal of Economics* 134, no.1 (February): 51–107.

Appendix A. Variance of RPPs Estimated Using CPI and Health Marketscan Data

This appendix provides a simple illustration of how the variance of medical RPPs across regions estimated using Health Marketscan data can be higher than the variance of medical RPPs across regions estimated using CPI data, as shown in table B.8, even though the variance across regions of each component of the medical RPP index is higher in the CPI data than in the Health Marketscan data, as shown in tables B.5 and B.6.

Let P_{ir} be the "true" price level of component i of the medical category (hospital services, physician services, or prescription drugs) in region r, where regions are either states or CPI regions. Suppose that $P_{ir} = 1 + \eta_{ir}$, where η_{ir} is a random variable with mean 0 and variance σ_P^2. It seems reasonable to suppose that the price levels of different components of the medical category are correlated within regions. To model this assumption, suppose that $Cov(\eta_{ir}, \eta_{jr}) = \alpha > 0$ for all regions r and for any components i, j. Let CPI_{ir} be the measured price level for component i in region r using CPI data, and let HM_{ir} be the measured price level for component i in region r using Health Marketscan data. Suppose that there is more measurement error when using CPI data than when using Health Marketscan data. In fact, to make the point most clearly, suppose that the CPI data does not measure the underlying price level at all, so that $CPI_{ir} = 1 + \varepsilon_{ir}$, where ε_{ir} is an error term with mean 0 and variance $\sigma_{CPI}^2 > \sigma_P^2$, and where ε_{ir} and ε_{js} are independently distributed for all components i, j and all regions r, s. On the other hand, suppose that the Health Marketscan data perfectly measures the true underlying price level, so that $HM_{ir} = P_{ir}$. Since $\sigma_{CPI}^2 > \sigma_P^2$, there is more variance in the measured price level of any individual component i across regions when using the CPI data than when using the Health Marketscan data. This observation corresponds to the result from tables B.5 and B.6 that the variance of first stage price level estimates is greater when using the CPI data than when using the Health Marketscan data.

Greatly oversimplifying the actual procedure for generating RPPs, suppose that the medical RPP for a given state is just the mean of the measured price levels for each component in the state. That is, let $CPI_r = \frac{1}{n}\sum_i CPI_{ir}$ be the medical RPP for region r calculated using CPI data, and let $HM_r = \frac{1}{n}\sum_i HM_{ir}$ be the medical RPP for region r calculated using Health Marketscan data, where n is the number of components of the index. Then $Var(CPI_r) = \sigma_{CPI}^2/n$, while $Var(HM_r) = \sigma_P^2/n + 2\alpha/n$. Thus, it is possible to have $Var(HM_r) > Var(CPI_r)$ even though $Var(HM_{ir}) < Var(CPI_{ir})$ for all i.

Appendix B. Tables

Table B.1. Number of Observations by Category and State, 2017 Health Marketscan Data

State	Inpatient facility	Inpatient physician	Outpatient facility	Outpatient physician	Prescription drugs
Alabama	19,333	234,420	408,858	7,262,478	4,517,462
Alaska	719	6,807	14,945	235,483	99,680
Arizona	20,575	247,701	271,589	7,998,600	2,851,603
Arkansas	4,735	67,732	92,374	1,443,812	923,894
California	63,160	726,489	1,297,118	27,285,328	10,323,032
Colorado	15,410	182,093	393,318	5,155,426	2,330,393
Connecticut	11,330	126,579	368,312	4,801,151	1,981,754
Delaware	5,335	50,938	126,171	2,069,838	772,356
District of Columbia	683	6,845	12,824	311,616	101,535
Florida	99,998	1,873,247	2,526,212	29,999,257	13,502,473
Georgia	51,637	622,439	1,112,471	22,817,538	12,046,941
Hawaii	99	1,582	2,418	50,104	18,153
Idaho	6,069	55,713	150,478	1,549,026	1,008,232
Illinois	39,723	785,073	1,096,167	13,759,267	5,703,212
Indiana	24,286	256,753	663,917	6,924,349	4,469,730
Iowa	8,817	86,659	243,129	2,599,139	1,413,482
Kansas	10,992	150,016	208,369	3,125,713	1,774,885
Kentucky	19,023	194,951	422,659	6,338,639	3,865,931
Louisiana	29,266	477,462	818,777	10,867,616	5,118,558
Maine	2,317	19,887	92,120	662,267	352,873
Maryland	12,124	129,268	235,833	5,127,117	2,006,798
Massachusetts	22,188	212,753	940,258	6,785,554	3,140,635
Michigan	51,413	575,829	1,815,905	15,000,351	7,661,545
Minnesota	14,054	144,990	270,232	4,289,145	1,760,930
Mississippi	14,553	161,690	296,217	4,347,668	3,059,001
Missouri	29,855	330,455	786,902	7,973,998	4,569,914
Montana	1,375	11,760	35,602	334,759	187,221
Nebraska	4,253	39,872	87,922	1,275,352	681,870
Nevada	7,797	103,585	100,139	2,523,728	1,239,116
New Hampshire	3,689	34,159	142,391	1,084,649	578,400
New Jersey	31,052	473,665	620,889	14,227,354	4,288,847
New Mexico	2,573	28,357	68,202	808,182	421,031
New York	57,777	668,254	1,172,228	24,288,727	7,235,183
North Carolina	31,906	398,233	672,063	12,054,412	6,064,478
North Dakota	798	6,289	16,062	193,385	99,181
Ohio	66,041	703,731	2,020,264	18,858,142	10,602,343
Oklahoma	9,225	115,320	197,857	2,760,850	1,627,163
Oregon	10,989	91,936	305,691	3,875,280	1,824,127
Pennsylvania	40,879	437,448	1,232,710	12,202,310	5,681,686
Rhode Island	2,946	24,479	91,816	931,668	329,254
South Carolina	*	*	*	*	*
South Dakota	1,835	16,482	29,211	443,571	216,620
Tennessee	22,878	263,742	486,854	9,171,126	5,418,983
Texas	110,233	2,254,400	2,331,382	37,418,057	17,141,164
Utah	7,509	63,487	130,104	1,671,290	932,748
Vermont	690	6,514	29,626	187,557	113,589
Virginia	23,625	309,151	512,483	9,329,583	4,253,246
Washington	22,898	218,186	542,015	9,050,442	4,399,300
West Virginia	2,979	34,764	104,441	936,109	666,444
Wisconsin	21,257	221,620	660,391	6,204,715	3,424,352
Wyoming	1,000	8,846	24,022	223,478	138,778
Total	1,063,898	14,262,651	26,281,938	368,835,206	172,940,126
Maximum	110,233	2,254,400	2,526,212	37,418,057	17,141,164
Minimum	99	1,582	2,418	50,104	18,153
Range	110,134	2,252,818	2,523,794	37,367,953	17,123,011

* Not available. Data from South Carolina was not reported due to privacy restrictions.

Note. Totals exclude observations from South Carolina.

Table B.1.a. Percentage of Observations in Each State by Category, 2017 Health Marketscan Data

State	Inpatient facility	Inpatient physician	Outpatient facility	Outpatient physician	Prescription drugs	State population as percentage of national population
Alabama	1.82	1.64	1.56	1.97	2.61	1.50
Alaska	0.07	0.05	0.06	0.06	0.06	0.23
Arizona	1.93	1.74	1.03	2.17	1.65	2.15
Arkansas	0.45	0.47	0.35	0.39	0.53	0.92
California	5.94	5.09	4.94	7.40	5.97	12.14
Colorado	1.45	1.28	1.50	1.40	1.35	1.72
Connecticut	1.06	0.89	1.40	1.30	1.15	1.10
Delaware	0.50	0.36	0.48	0.56	0.45	0.30
District of Columbia	0.06	0.05	0.05	0.08	0.06	0.21
Florida	9.40	13.13	9.61	8.13	7.81	6.44
Georgia	4.85	4.36	4.23	6.19	6.97	3.20
Hawaii	0.01	0.01	0.01	0.01	0.01	0.44
Idaho	0.57	0.39	0.57	0.42	0.58	0.53
Illinois	3.73	5.50	4.17	3.73	3.30	3.93
Indiana	2.28	1.80	2.53	1.88	2.58	2.05
Iowa	0.83	0.61	0.93	0.70	0.82	0.97
Kansas	1.03	1.05	0.79	0.85	1.03	0.89
Kentucky	1.79	1.37	1.61	1.72	2.24	1.37
Louisiana	2.75	3.35	3.12	2.95	2.96	1.44
Maine	0.22	0.14	0.35	0.18	0.20	0.41
Maryland	1.14	0.91	0.90	1.39	1.16	1.86
Massachusetts	2.09	1.49	3.58	1.84	1.82	2.11
Michigan	4.83	4.04	6.91	4.07	4.43	3.06
Minnesota	1.32	1.02	1.03	1.16	1.02	1.71
Mississippi	1.37	1.13	1.13	1.18	1.77	0.92
Missouri	2.81	2.32	2.99	2.16	2.64	1.88
Montana	0.13	0.08	0.14	0.09	0.11	0.32
Nebraska	0.40	0.28	0.33	0.35	0.39	0.59
Nevada	0.73	0.73	0.38	0.68	0.72	0.92
New Hampshire	0.35	0.24	0.54	0.29	0.33	0.41
New Jersey	2.92	3.32	2.36	3.86	2.48	2.76
New Mexico	0.24	0.20	0.26	0.22	0.24	0.64
New York	5.43	4.69	4.46	6.59	4.18	6.09
North Carolina	3.00	2.79	2.56	3.27	3.51	3.15
North Dakota	0.08	0.04	0.06	0.05	0.06	0.23
Ohio	6.21	4.93	7.69	5.11	6.13	3.58
Oklahoma	0.87	0.81	0.75	0.75	0.94	1.21
Oregon	1.03	0.64	1.16	1.05	1.05	1.27
Pennsylvania	3.84	3.07	4.69	3.31	3.29	3.93
Rhode Island	0.28	0.17	0.35	0.25	0.19	0.33
South Carolina	*	*	*	*	*	1.54
South Dakota	0.17	0.12	0.11	0.12	0.13	0.27
Tennessee	2.15	1.85	1.85	2.49	3.13	2.06
Texas	10.36	15.81	8.87	10.14	9.91	8.69
Utah	0.71	0.45	0.50	0.45	0.54	0.95
Vermont	0.06	0.05	0.11	0.05	0.07	0.19
Virginia	2.22	2.17	1.95	2.53	2.46	2.60
Washington	2.15	1.53	2.06	2.45	2.54	2.27
West Virginia	0.28	0.24	0.40	0.25	0.39	0.56
Wisconsin	2.00	1.55	2.51	1.68	1.98	1.78
Wyoming	0.09	0.06	0.09	0.06	0.08	0.18
Total	100%	100%	100%	100%	100%	100%
Maximum	10.36	15.81	9.61	10.14	9.91	12.14
Minimum	0.01	0.01	0.01	0.01	0.01	0.18
Range	10.35	15.80	9.60	10.13	9.90	11.96

* Not available. Data from South Carolina was not reported due to privacy restrictions.

Note. This table is generated by regressing the log price of a given good or service on dummies for each state and the control variables described in tables B.2, B.3, or B.4. The coefficients on the state dummies are then normalized so that the geometric mean of the dummies is one. Each entry in the table is a normalized coefficient.

Table B.2. Physician Services Regression Control Variables

Control variable	Brief description	Number of values
PROC1	Procedure code	13651
AGE	Age of patient	66
SEX	Sex of patient	2
DRG	Diagnosis related group—a variable that classifies the patient's diagnosis. This variable is only available for inpatient observations; for outpatient observations the DRG is coded as 0.	753
MDC	Major disease category—a classification of the patient's diagnosis at a higher level of aggregation. This variable is available for both inpatient and outpatient observations.	26
PLANTYP	Type of insurance that covers the patient, e.g. HMO or PPO	8
EECLASS	Employee classification of the beneficiary—salaried, hourly, or other, and union or non-union	9
EESTATU	Employment status of the beneficiary—active full time or part-time, retiree, etc.	9
EMPREL	Relationship of the patient to the primary beneficiary	4
HLTHPLAN	Indicator for whether the data is supplied by a large employer or a health plan	2
INDSTRY	Industry classification of the employer responsible for payment of claim	10
STDPROV	Provider type—a code for the specialty of the medical provider, e.g. dermatologist	133
STDPLAC	Setting where service occurred —a code for the type of medical facility, e.g. birthing center or inpatient hospital	44
PAIDNTWK	An indicator of whether a claim was paid as in-network or not.	2

Table B.3. Hospital Services Regression Control Variables

Control variable	Brief description	Number of values
DAYS	Number of days spent in hospital. This variable is set to 0 for all outpatient observations.	229
AGE	Age of patient	66
SEX	Sex of patient	2
DRG	Diagnosis related group—a variable that classifies the patient's diagnosis. This variable is only available for inpatient observations; for outpatient observations the DRG is coded as 0.	753
MDC	Major disease category— classification of the patient's diagnosis at a higher level of aggregation. This variable is available for both inpatient and outpatient observations.	26
PLANTYP	Type of insurance that covers the patient, e.g. HMO or PPO	8
EECLASS	Employee classification of the beneficiary—salaried, hourly, or other, and union or non-union	9
EESTATU	Employment status of the beneficiary—active full time or part-time, retiree, etc.	9
EMPREL	Relationship of the patient to the primary beneficiary	4
HLTHPLAN	Indicator for whether the data is supplied by a large employer or a health plan	2
INDSTRY	Industry classification of the employer responsible for payment of claim	10

Table B.4. Prescription Drugs Regression Control Variables

Control variable	Brief description	Number of values
NDCNUM	National drug code	38489
AGE	Age of patient	66
SEX	Sex of patient	2
PLANTYP	Type of insurance that covers the patient, e.g. HMO or PPO	8
EECLASS	Employee classification of the beneficiary—salaried, hourly, or other, and union or non-union	9
EESTATU	Employment status of the beneficiary—active full time or part-time, retiree, etc.	9
EMPREL	Relationship of the patient to the primary beneficiary	4
HLTHPLAN	Indicator for whether the data is supplied by a large employer or a health plan	2
INDSTRY	Industry classification of the employer responsible for payment of claim	10
PAIDNTWK	An indicator of whether a claim was paid as in-network or not.	2

Table B.5. First Stage Regression Estimates, 2017 Health Marketscan Data

State	Physician services	Hospital services	Prescription drugs
Alabama	0.83	0.65	0.90
Alaska	2.13	1.52	0.99
Arizona	0.81	1.44	1.00
Arkansas	0.92	0.73	1.02
California	1.03	1.40	1.04
Colorado	0.95	1.07	1.14
Connecticut	1.02	0.84	1.03
Delaware	0.84	1.01	0.99
District of Columbia	0.89	1.14	1.04
Florida	0.84	1.34	0.96
Georgia	0.94	1.21	0.82
Hawaii	1.12	0.86	1.21
Idaho	1.10	0.86	1.02
Illinois	0.97	1.05	1.03
Indiana	0.86	1.29	0.98
Iowa	1.19	0.83	1.01
Kansas	0.96	1.08	0.97
Kentucky	0.82	1.02	0.97
Louisiana	0.84	0.76	1.01
Maine	0.98	0.90	1.04
Maryland	0.86	1.08	1.04
Massachusetts	1.09	0.67	0.90
Michigan	0.88	0.51	0.92
Minnesota	1.35	1.05	1.11
Mississippi	1.00	0.77	0.93
Missouri	0.84	0.90	1.01
Montana	1.23	0.85	1.02
Nebraska	1.26	0.99	1.00
Nevada	0.82	1.59	1.08
New Hampshire	1.06	0.88	0.90
New Jersey	0.95	1.33	0.98
New Mexico	0.92	0.95	1.00
New York	1.00	0.96	0.97
North Carolina	0.96	1.10	1.01
North Dakota	1.47	0.91	1.00
Ohio	0.88	1.00	0.96
Oklahoma	0.90	1.10	1.02
Oregon	1.21	0.93	1.07
Pennsylvania	0.89	0.83	1.07
Rhode Island	0.81	0.59	0.99
South Carolina	*	*	*
South Dakota	1.20	1.17	0.97
Tennessee	0.90	1.02	1.00
Texas	0.88	1.54	0.95
Utah	0.93	0.90	0.97
Vermont	1.06	1.06	0.94
Virginia	0.90	1.12	0.99
Washington	1.10	1.02	1.09
West Virginia	0.99	1.14	1.03
Wisconsin	1.40	1.13	1.04
Wyoming	1.27	1.08	0.99
Geometric Mean	1.00	1.00	1.00
Maximum	2.13	1.59	1.21
Minimum	0.81	0.51	0.82
Range	1.32	1.08	0.39
Standard Deviation	0.23	0.24	0.06

* Not available. Data from South Carolina not reported due to privacy restrictions.

Note. This table is generated by regressing the log price of a given good or service on dummies for each state and the control variables described in tables B.2, B.3, or B.4. The coefficients on the state dummies are then normalized so that the geometric mean of the dummies is one. Each entry in the table is a normalized coefficient.

Table B.6. First Stage Regression Estimates, 2017 CPI Data

CPI Area	Physician services	Hospital services	Prescription drugs
A102: Philadelphia-Wilmington-Atlantic City	0.92	1.24	1.37
A103: Boston-Brockton-Nashua	1.32	*	1.97
A104: Pittsburgh	0.60	0.42	1.00
A109: New York City	1.06	1.99	1.15
A110: New York-Connecticut Suburbs	1.64	1.28	0.84
A111: New Jersey-Pennsylvania Suburbs	1.10	1.57	0.99
A207: Chicago-Gary-Kenosha	1.62	1.10	0.72
A208: Detroit-Ann Arbor-Flint	0.75	1.65	1.06
A209: St. Louis	*	0.82	0.58
A210: Cleveland-Akron	0.53	0.62	0.73
A211: Minneapolis-St. Paul	3.06	0.62	0.68
A212: Milwaukee-Racine	1.45	0.69	1.42
A213: Cincinnati-Hamilton	1.23	3.41	0.68
A214: Kansas City	1.22	*	0.55
A312: District of Columbia	2.13	0.88	1.67
A313: Baltimore	1.10	0.42	1.33
A316: Dallas-Fort Worth	0.53	0.90	0.91
A318: Houston-Galveston-Brazoria	1.00	2.91	0.97
A319: Atlanta	0.76	1.33	0.95
A320: Miami-Fort Lauderdale	0.59	1.84	2.89
A321: Tampa-St. Petersburg-Clearwater	0.50	2.81	*
A419: Los Angeles-Long Beach	0.61	0.46	0.91
A420: Los Angeles Suburbs	1.32	1.89	0.32
A422: San Francisco-Oakland-San Jose	1.09	2.07	1.25
A423: Seattle-Tacoma-Bremerton	1.24	1.07	1.16
A425: Portland-Salem	0.93	0.39	1.09
A426: Honolulu	0.87	*	1.28
A427: Anchorage	0.91	0.60	0.56
A429: Phoenix-Mesa	0.79	*	1.09
A433: Denver-Boulder-Greeley	0.55	0.89	0.92
D200: Midwest nonmetropolitan urban	1.35	0.82	1.14
D300: South nonmetropolitan urban	1.17	*	1.08
D400: West nonmetropolitan urban	1.05	0.84	1.43
X100: Northeast small metropolitan	1.07	0.74	1.07
X200: Midwest small metropolitan	0.88	0.83	0.79
X300: South small metropolitan	0.79	0.70	1.25
X499: West small metropolitan	1.43	0.84	1.00
Geometric Mean	1.00	1.00	1.00
Maximum	3.06	3.41	2.89
Minimum	0.50	0.39	0.32
Range	2.56	3.02	2.57
Standard deviation	0.49	0.74	0.45

* Coefficient not estimated for these areas or the cell is suppressed because the number of quotes does not meet minimum threshold.

This table is generated by regressing the log price of a given good or service on dummies for each CPI area and control variables. Each entry in this table is a normalized coefficient. Note that not all CPI areas are estimated in each regression. Cell suppressed if number of quotes does not meet minimum threshold. The CPI data contain different price-determining characteristics than the Health Marketscan data, and so the control variables in the regressions reported in table 7 are not the same as the control variables used in the Health Marketscan regressions. CPI areas are 30 large metropolitan areas (area codes beginning with A), non-metropolitan urban areas in each of four regions (area codes beginning with D), and small metropolitan areas in each of four regions (area codes beginning with X).

Table B.7. RPP Estimates With and Without Health Marketscan Data, 2017

State	RPPs using CPI and ACS data only	RPPs using Health Marketscan data	Ratio of RPP from Marketscan data to RPP from CPI data
Alabama	0.87	0.86	0.99
Alaska	1.04	1.05	1.01
Arizona	0.96	0.96	1.00
Arkansas	0.86	0.86	1.00
California	1.15	1.15	1.00
Colorado	1.03	1.03	1.00
Connecticut	1.08	1.07	0.99
Delaware	1.00	1.00	1.00
District of Columbia	1.17	1.16	0.99
Florida	1.00	1.01	1.01
Georgia	0.92	0.93	1.01
Hawaii	1.18	1.18	1.00
Idaho	0.93	0.92	0.99
Illinois	0.98	0.98	1.00
Indiana	0.90	0.90	1.00
Iowa	0.90	0.89	0.99
Kansas	0.90	0.90	1.00
Kentucky	0.88	0.88	1.00
Louisiana	0.90	0.90	1.00
Maine	0.98	0.98	1.00
Maryland	1.09	1.09	1.00
Massachusetts	1.08	1.07	0.99
Michigan	0.93	0.91	0.98
Minnesota	0.97	0.98	1.01
Mississippi	0.86	0.85	0.99
Missouri	0.89	0.89	1.00
Montana	0.95	0.94	0.99
Nebraska	0.90	0.89	0.99
Nevada	0.98	0.98	1.00
New Hampshire	1.06	1.06	1.00
New Jersey	1.13	1.13	1.00
New Mexico	0.93	0.93	1.00
New York	1.16	1.15	0.99
North Carolina	0.91	0.92	1.01
North Dakota	0.90	0.89	0.99
Ohio	0.89	0.89	1.00
Oklahoma	0.89	0.90	1.01
Oregon	0.99	0.99	1.00
Pennsylvania	0.98	0.98	1.00
Rhode Island	0.99	0.97	0.98
South Carolina	*	*	*
South Dakota	0.88	0.88	1.00
Tennessee	0.90	0.91	1.01
Texas	0.97	0.98	1.01
Utah	0.97	0.97	1.00
Vermont	1.03	1.03	1.00
Virginia	1.02	1.02	1.00
Washington	1.06	1.06	1.00
West Virginia	0.87	0.87	1.00
Wisconsin	0.92	0.92	1.00
Wyoming	0.95	0.95	1.00
Maximum	1.18	1.18	1.01
Minimum	0.86	0.85	0.98
Range	0.33	0.33	0.03
Standard deviation	0.09	0.09	N/A

* Not available. Data from South Carolina was not reported due to privacy restrictions.

N/A Not applicable.

Note. This table shows state level RPPs estimated using BLS and ACS data only and using BLS and ACS rent data combined with Health Marketscan data for physician services, hospital services, and prescription drugs.

Table B.8. Medical RPP Estimates With and Without Health Marketscan Data, 2017

State	RPPs using CPI and ACS data only	RPPs using Health Marketscan data	Ratio of RPP from Marketscan data to RPP from CPI data
Alabama	0.91	0.81	0.89
Alaska	1.12	1.24	1.11
Arizona	1.04	1.02	0.98
Arkansas	0.92	0.87	0.95
California	0.99	1.09	1.10
Colorado	0.99	1.01	1.02
Connecticut	1.04	0.93	0.89
Delaware	1.01	0.96	0.95
District of Columbia	1.16	1.07	0.92
Florida	0.92	1.05	1.14
Georgia	0.97	1.01	1.04
Hawaii	0.95	0.92	0.97
Idaho	1.08	0.95	0.88
Illinois	1.10	1.06	0.96
Indiana	1.06	1.02	0.96
Iowa	1.06	0.90	0.85
Kansas	1.08	1.00	0.93
Kentucky	0.92	0.95	1.03
Louisiana	0.90	0.88	0.98
Maine	0.97	0.93	0.96
Maryland	1.03	1.01	0.98
Massachusetts	0.94	0.87	0.93
Michigan	1.07	0.75	0.70
Minnesota	1.02	1.03	1.01
Mississippi	0.93	0.88	0.95
Missouri	0.99	0.92	0.93
Montana	1.12	0.96	0.86
Nebraska	1.06	0.97	0.92
Nevada	0.97	1.05	1.08
New Hampshire	0.95	0.93	0.98
New Jersey	1.08	1.09	1.01
New Mexico	1.01	0.94	0.93
New York	1.05	0.98	0.93
North Carolina	0.91	1.02	1.12
North Dakota	1.06	0.94	0.89
Ohio	0.97	0.93	0.96
Oklahoma	0.91	1.01	1.11
Oregon	1.00	1.00	1.00
Pennsylvania	0.94	0.89	0.95
Rhode Island	0.97	0.77	0.79
South Carolina	*	*	*
South Dakota	1.05	1.00	0.95
Tennessee	0.91	0.98	1.08
Texas	0.96	1.10	1.15
Utah	0.99	0.93	0.94
Vermont	0.97	0.98	1.01
Virginia	1.00	1.03	1.03
Washington	1.13	1.04	0.92
West Virginia	0.95	1.02	1.07
Wisconsin	1.05	1.04	0.99
Wyoming	1.11	1.02	0.92
Maximum	1.16	1.24	1.15
Minimum	0.90	0.75	0.70
Range	0.25	0.50	0.44
Standard deviation	0.07	0.09	N/A

* Not available. Data from South Carolina was not reported due to privacy restrictions.

N/A Not applicable.

Note. This table shows state level category RPPs for medical goods and services only, estimated using BLS and ACS data only, and using BLS and ACS rent data combined with Health Marketscan data for physician services, hospital services, and prescription drugs.

Table B.9. First Stage Regression Estimates for Physician Services, 2014–2017, Health Marketscan Data

State	2014	2015	2016	2017	Range/Mean
Alabama	0.9	0.88	0.83	0.83	0.08
Alaska	2.23	2.15	2.18	2.13	0.05
Arizona	0.83	0.82	0.83	0.81	0.02
Arkansas	0.96	0.97	0.95	0.92	0.05
California	1	1.05	1.08	1.03	0.08
Colorado	0.93	0.93	0.95	0.95	0.02
Connecticut	1.01	1.04	1.01	1.02	0.03
Delaware	0.86	0.84	0.84	0.84	0.02
District of Columbia	0.95	0.94	0.94	0.89	0.06
Florida	0.86	0.85	0.83	0.84	0.04
Georgia	0.93	0.95	0.95	0.94	0.02
Hawaii	1.05	1.13	1.23	1.12	0.16
Idaho	1.06	1.05	1.07	1.1	0.05
Illinois	1.01	0.99	0.96	0.97	0.05
Indiana	0.82	0.82	0.83	0.86	0.05
Iowa	1.13	1.15	1.14	1.19	0.05
Kansas	1.01	0.98	0.96	0.96	0.05
Kentucky	0.77	0.78	0.78	0.82	0.06
Louisiana	0.88	0.86	0.85	0.84	0.05
Maine	0.98	0.92	0.94	0.98	0.06
Maryland	0.86	0.86	0.87	0.86	0.01
Massachusetts	1.06	1.05	1.05	1.09	0.04
Michigan	0.88	0.85	0.86	0.88	0.03
Minnesota	1.33	1.36	1.39	1.35	0.04
Mississippi	1.04	1.03	1.04	1	0.04
Missouri	0.84	0.83	0.82	0.84	0.02
Montana	1.27	1.26	1.27	1.23	0.03
Nebraska	1.25	1.26	1.26	1.26	0.01
Nevada	0.84	0.81	0.81	0.82	0.04
New Hampshire	0.99	0.99	1.02	1.06	0.07
New Jersey	0.97	0.97	0.96	0.95	0.02
New Mexico	0.93	0.94	0.93	0.92	0.02
New York	1	1.02	1.03	1	0.03
North Carolina	0.99	0.98	0.96	0.96	0.03
North Dakota	1.4	1.42	1.39	1.47	0.06
Ohio	0.82	0.84	0.85	0.88	0.07
Oklahoma	0.95	0.93	0.93	0.9	0.05
Oregon	1.18	1.2	1.2	1.21	0.03
Pennsylvania	0.9	0.87	0.86	0.89	0.05
Rhode Island	0.76	0.78	0.79	0.81	0.06
South Carolina	*	*	*	*	*
South Dakota	1.23	1.26	1.22	1.2	0.05
Tennessee	0.91	0.9	0.92	0.9	0.02
Texas	0.88	0.89	0.88	0.88	0.01
Utah	0.94	0.93	0.93	0.93	0.01
Vermont	0.96	1	1.07	1.06	0.11
Virginia	0.91	0.91	0.9	0.9	0.01
Washington	1.09	1.12	1.11	1.1	0.03
West Virginia	0.98	0.97	0.97	0.99	0.02
Wisconsin	1.43	1.41	1.41	1.4	0.02
Wyoming	1.33	1.31	1.29	1.27	0.05
Geometric mean	1	1	1	1	N/A
Maximum	2.23	2.15	2.18	2.13	0.16
Minimum	0.76	0.78	0.78	0.81	0.00
Range	1.46	1.37	1.4	1.32	0.16

* Not available.

N/A Not applicable.

Note. Data from South Carolina was not reported due to privacy restrictions.

Table B.10. First Stage Regression Estimates for Hospital Services, 2014–2017, Health Marketscan Data

State	2014	2015	2016	2017	Range/Mean
Alabama	0.62	0.65	0.64	0.65	0.05
Alaska	1.5	1.48	1.53	1.52	0.03
Arizona	1.41	1.43	1.45	1.44	0.03
Arkansas	0.76	0.79	0.75	0.73	0.08
California	1.31	1.31	1.37	1.4	0.07
Colorado	1.07	1.01	1.07	1.07	0.06
Connecticut	0.86	0.87	0.84	0.84	0.04
Delaware	1.06	1.04	1.03	1.01	0.05
District of Columbia	1.09	1.23	1.36	1.34	0.22
Florida	1.22	1.22	1.22	1.21	0.01
Georgia	0.76	0.86	0.84	0.86	0.12
Hawaii	0.84	0.84	0.88	0.86	0.05
Idaho	1.09	1.05	1.02	1.05	0.07
Illinois	1.23	1.22	1.28	1.29	0.06
Indiana	0.9	0.9	0.87	0.83	0.08
Iowa	1.06	1.05	1.04	1.08	0.04
Kansas	1.11	1.04	1.1	1.02	0.08
Kentucky	0.79	0.77	0.74	0.76	0.07
Louisiana	0.88	0.89	0.87	0.9	0.03
Maine	1.05	1.02	1.03	1.08	0.06
Maryland	0.65	0.66	0.66	0.67	0.03
Massachusetts	0.49	0.49	0.51	0.51	0.04
Michigan	1.16	1.13	1.08	1.05	0.10
Minnesota	0.76	0.75	0.76	0.77	0.03
Mississippi	0.88	0.88	0.89	0.9	0.02
Missouri	0.83	0.83	0.84	0.85	0.02
Montana	1.07	1.05	1.04	0.99	0.08
Nebraska	1.5	1.59	1.62	1.59	0.08
Nevada	0.92	0.91	0.89	0.88	0.04
New Hampshire	1.35	1.31	1.32	1.33	0.03
New Jersey	1.16	1.08	1.06	0.95	0.20
New Mexico	0.95	0.84	0.88	0.96	0.13
New York	1.07	1.29	1.07	1.1	0.19
North Carolina	0.82	0.86	0.93	0.91	0.13
North Dakota	1.05	0.96	0.99	1	0.09
Ohio	1.08	1.05	1.07	1.1	0.05
Oklahoma	0.93	0.91	0.89	0.93	0.04
Oregon	0.8	0.86	0.83	0.83	0.07
Pennsylvania	0.61	0.7	0.63	0.59	0.17
Rhode Island	1.24	1.23	1.24	1.26	0.02
South Carolina	*	*	*	*	*
South Dakota	0.98	0.91	1	1.02	0.11
Tennessee	1.47	1.47	1.52	1.54	0.05
Texas	0.94	0.95	0.93	0.9	0.05
Utah	1.01	1.07	1.06	1.06	0.06
Vermont	1.12	1.11	1.09	1.12	0.03
Virginia	0.98	0.96	1	1.02	0.06
Washington	1.17	1.14	1.11	1.14	0.05
West Virginia	1.13	1.12	1.13	1.14	0.02
Wisconsin	1.14	1.12	1.11	1.13	0.03
Wyoming	1.11	1.05	1.11	1.08	0.06
Geometric mean	1	1	1	1	N/A
Maximum	1.5	1.59	1.62	1.59	0.22
Minimum	0.49	0.49	0.51	0.51	0.01
Range	1	1.09	1.12	1.08	0.21

* Not available.

N/A Not applicable.

Note. Data from South Carolina was not reported due to privacy restrictions.

Table B.11. First Stage Regression Estimates for Prescription Drugs, 2014–2017, Health Marketscan Data

State	2014	2015	2016	2017	Range/Mean
Alabama	0.96	0.94	0.88	0.9	0.09
Alaska	1.03	0.95	0.98	0.99	0.08
Arizona	0.96	0.97	1	1	0.04
Arkansas	0.98	0.99	1.01	1.02	0.04
California	1.01	1.03	1.06	1.04	0.05
Colorado	1.05	1.08	1.09	1.14	0.08
Connecticut	1.01	1.02	1.02	1.03	0.02
Delaware	1.01	1.02	1	0.99	0.03
District of Columbia	0.93	0.93	0.93	0.96	0.03
Florida	0.96	0.91	0.85	0.82	0.16
Georgia	1.05	1.17	1.18	1.21	0.14
Hawaii	1.01	0.96	1.03	1.02	0.07
Idaho	1.02	1.07	1.06	1.03	0.05
Illinois	0.96	0.91	0.95	0.98	0.07
Indiana	1.03	1.03	1.04	1.01	0.03
Iowa	1.03	0.99	0.98	0.97	0.06
Kansas	0.97	0.93	0.94	0.97	0.04
Kentucky	1.06	1.02	1.01	1.01	0.05
Louisiana	1.23	1.23	0.98	1.04	0.22
Maine	1.01	1.06	1.06	1.04	0.05
Maryland	1	0.91	0.91	0.9	0.11
Massachusetts	0.97	0.95	0.94	0.92	0.05
Michigan	1.05	1.1	1.11	1.11	0.05
Minnesota	0.99	0.95	0.96	0.93	0.06
Mississippi	1	1	1.01	1.01	0.01
Missouri	0.99	0.98	1.01	1.02	0.04
Montana	1.02	1.04	1.02	1	0.04
Nebraska	1.01	1.06	1.03	1.08	0.07
Nevada	1.06	1.01	0.92	0.9	0.16
New Hampshire	1.01	1.01	0.99	0.98	0.03
New Jersey	0.97	0.96	1	1	0.04
New Mexico	1.02	1	0.99	0.97	0.05
New York	1	1.01	1.01	1.01	0.01
North Carolina	1	1.01	1.03	1	0.03
North Dakota	0.94	0.95	0.95	0.96	0.02
Ohio	1.02	0.97	1.04	1.02	0.07
Oklahoma	0.92	0.92	1	1.07	0.15
Oregon	0.9	1.07	1.04	1.07	0.17
Pennsylvania	0.99	1.06	1.02	0.99	0.07
Rhode Island	0.93	0.94	0.97	0.95	0.04
South Carolina	*	*	*	*	*
South Dakota	1.01	0.99	0.96	1	0.05
Tennessee	0.99	0.97	0.97	0.95	0.04
Texas	0.96	0.95	0.96	0.97	0.02
Utah	1.04	1.05	0.94	0.94	0.11
Vermont	1.01	1	1.02	0.99	0.03
Virginia	0.9	1.01	1.1	1.09	0.20
Washington	0.99	1.04	1.04	1.04	0.05
West Virginia	1.05	1.02	1.05	1.03	0.03
Wisconsin	1.08	1.1	1.05	1.04	0.06
Wyoming	0.98	0.86	0.97	0.99	0.14
Geometric mean	1	1	1	1	N/A
Maximum	1.23	1.23	1.18	1.21	0.22
Minimum	0.9	0.86	0.85	0.82	0.01
Range	0.34	0.37	0.33	0.39	0.21

* Not available.

N/A Not applicable.

Note. Data from South Carolina was not reported due to privacy restrictions.

Table B.12. Coefficients from Original and Reweighted Prescription Drug Regressions, 2017 Health Marketscan Data, Ordered by Ratio Between Original and Reweighted Regression Coefficients

State	Normalized coefficient in original regression	Normalized coefficient in reweighted regression	Ratio of column 2 to column 1
California	1.04	0.98	0.94
Washington	1.09	1.03	0.94
Montana	1.02	0.96	0.95
Colorado	1.14	1.11	0.98
Alaska	0.99	0.97	0.98
Indiana	0.98	0.96	0.98
Oklahoma	1.02	1.00	0.98
Washington, DC	1.04	1.02	0.98
Nevada	1.08	1.07	0.98
Arkansas	1.02	1.00	0.99
Iowa	1.01	1.00	0.99
Illinois	1.03	1.02	0.99
Idaho	1.02	1.02	0.99
South Dakota	0.97	0.96	0.99
Kansas	0.97	0.96	0.99
Oregon	1.07	1.06	0.99
Connecticut	1.03	1.02	0.99
Maine	1.04	1.04	1.00
New Mexico	1.00	1.00	1.00
Kentucky	0.97	0.97	1.00
West Virginia	1.03	1.03	1.00
Florida	0.96	0.96	1.00
Rhode Island	0.99	0.99	1.00
Wisconsin	1.04	1.04	1.00
North Carolina	1.01	1.01	1.00
Arizona	1.00	1.01	1.01
Louisiana	1.01	1.02	1.01
Georgia	0.82	0.82	1.01
Michigan	0.92	0.92	1.01
Maryland	1.04	1.05	1.01
Virginia	0.99	1.00	1.01
Ohio	0.96	0.96	1.01
Texas	0.95	0.96	1.01
Delaware	0.99	1.00	1.01
Wyoming	0.99	1.00	1.01
Hawaii	1.21	1.22	1.01
New York	0.97	0.98	1.01
Utah	0.97	0.99	1.01
North Dakota	1.00	1.01	1.01
South Carolina	*	*	*
Vermont	0.94	0.96	1.01
New Jersey	0.98	1.00	1.01
Pennsylvania	1.07	1.09	1.02
Alabama	0.90	0.92	1.02
Nebraska	1.00	1.01	1.02
Minnesota	1.11	1.13	1.02
New Hampshire	0.90	0.92	1.02
Missouri	1.01	1.03	1.03
Tennessee	1.00	1.03	1.03
Mississippi	0.93	0.96	1.04
Massachusetts	0.90	0.94	1.04
Geometric mean	1.00	1.00	N/A
Maximum	1.21	1.22	1.04
Minimum	0.82	0.82	0.94
Range	0.39	0.39	0.10

* Not available.

N/A Not applicable.

Note. Data from South Carolina was not reported due to privacy restrictions.

Table B.13. Coefficients from Original and Reweighted Physician Services Regressions, 2017 Health Marketscan Data, Ordered by Ratio Between Original and Reweighted Regression Coefficients

State	Normalized coefficient in original regression	Normalized coefficient in reweighted regression	Ratio of column 2 to column 1
Alaska	2.13	2.06	0.96
California	1.03	1.00	0.96
Nebraska	1.23	1.21	0.98
Oregon	0.90	0.89	0.99
Indiana	0.97	0.96	0.99
Arkansas	0.92	0.91	0.99
Colorado	0.95	0.94	0.99
Nevada	1.26	1.25	0.99
Oklahoma	0.88	0.88	0.99
New Jersey	1.06	1.05	0.99
South Dakota	0.91	0.91	0.99
Kansas	1.19	1.18	0.99
Kentucky	0.96	0.96	0.99
Wyoming	1.27	1.26	0.99
Montana	0.84	0.83	1.00
Rhode Island	0.89	0.89	1.00
North Carolina	1.00	0.99	1.00
Hawaii	0.94	0.94	1.00
West Virginia	0.99	0.99	1.00
Illinois	1.10	1.09	1.00
District of Columbia	1.10	1.10	1.00
Vermont	0.93	0.92	1.00
North Dakota	0.96	0.96	1.00
Florida	0.89	0.89	1.00
Maine	0.84	0.84	1.00
Iowa	0.86	0.86	1.00
New York	0.92	0.92	1.00
New Mexico	0.95	0.95	1.00
Idaho	1.12	1.12	1.00
Michigan	1.09	1.09	1.00
Massachusetts	0.86	0.86	1.00
Arizona	0.81	0.82	1.00
Utah	0.88	0.88	1.00
Mississippi	1.35	1.35	1.00
Connecticut	1.02	1.03	1.00
Texas	0.90	0.90	1.00
Alabama	0.83	0.83	1.00
Washington	0.90	0.90	1.01
Louisiana	0.82	0.83	1.01
Wisconsin	1.40	1.41	1.01
Maryland	0.98	0.99	1.01
South Carolina	*	*	*
Delaware	0.84	0.84	1.01
New Hampshire	0.82	0.84	1.01
Tennessee	1.20	1.21	1.01
Georgia	0.84	0.86	1.02
Pennsylvania	1.21	1.23	1.02
Virginia	1.06	1.08	1.02
Ohio	1.47	1.51	1.03
Missouri	1.00	1.03	1.03
Minnesota	0.88	0.91	1.03
Geometric mean	1.00	1.00	N/A
Maximum	2.13	2.06	1.03
Minimum	0.81	0.82	0.96
Range	1.32	1.24	0.07

* Not available.

N/A Not applicable.

Note. Data from South Carolina was not reported due to privacy restrictions.

Table B.14. Coefficients from Original and Reweighted Hospital Services Regressions, 2017 Health Marketscan Data, Ordered by Ratio Between Original and Reweighted Regression Coefficients

State	Normalized coefficient in original regression	Normalized coefficient in reweighted regression	Ratio of column 2 to column 1
Georgia	1.34	1.26	0.94
New Hampshire	1.59	1.50	0.94
Delaware	1.01	0.96	0.95
Virginia	1.06	1.01	0.96
Mississippi	1.05	1.00	0.96
Iowa	1.29	1.24	0.96
Minnesota	0.51	0.49	0.97
Massachusetts	1.08	1.04	0.97
Utah	1.54	1.49	0.97
Montana	0.90	0.88	0.97
Connecticut	0.84	0.82	0.98
North Carolina	0.96	0.94	0.98
South Carolina	*	*	*
Maryland	0.90	0.88	0.98
Kentucky	1.08	1.06	0.98
Florida	1.14	1.12	0.98
Hawaii	1.21	1.19	0.98
South Dakota	1.26	1.24	0.98
New Mexico	1.33	1.31	0.99
Wyoming	1.08	1.07	0.99
California	1.40	1.38	0.99
Indiana	1.05	1.03	0.99
Arizona	1.44	1.42	0.99
West Virginia	1.14	1.13	0.99
New York	0.95	0.94	1.00
Pennsylvania	0.93	0.92	1.00
Alabama	0.65	0.65	1.00
Rhode Island	0.83	0.83	1.00
Arkansas	0.73	0.73	1.00
Michigan	0.67	0.67	1.00
Wisconsin	1.13	1.13	1.00
New Jersey	0.88	0.88	1.00
Washington	1.12	1.13	1.00
Nevada	0.99	0.99	1.01
Maine	0.76	0.76	1.01
Oklahoma	1.00	1.00	1.01
Nebraska	0.85	0.86	1.01
Oregon	1.10	1.11	1.01
Illinois	0.86	0.87	1.02
Alaska	1.52	1.55	1.02
Vermont	0.90	0.93	1.03
Texas	1.02	1.05	1.03
Kansas	0.83	0.85	1.03
District of Columbia	1.02	1.05	1.04
Idaho	0.86	0.89	1.04
Louisiana	1.02	1.06	1.04
Tennessee	1.17	1.23	1.05
Ohio	0.91	0.97	1.07
Missouri	0.77	0.83	1.07
Colorado	1.07	1.16	1.08
North Dakota	1.10	1.22	1.11
Geometric mean	1.00	1.00	N/A
Maximum	1.59	1.55	1.11
Minimum	0.51	0.49	0.94
Range	1.08	1.06	0.18

* Not available.

N/A Not applicable.

Note. Data from South Carolina was not reported due to privacy restrictions.