Setting a New Standard in Veterinary Cost Analysis: The Nationwide® | Purdue Veterinary Price Index

A review of methodology, January 2015

In the fall of 2014, Nationwide® entered into an agreement with the Krannert School of Business at Purdue University to provide analysis of the extensive pet insurance claims database of Veterinary Pet Insurance® (VPI), a Nationwide company. The key objective of this collaborative arrangement is to construct various price indices of veterinary services. The initial analysis focuses on creating price indices for the five-year period from January 2009 to December 2013. Below we discuss the nature of the dataset and the methods used to calculate these price indices for veterinary services.

Veterinary services price data

The Nationwide claims database breaks down claims filed by pet insurance policyholders into specific treatments, with the expense of each treatment separately recorded in the claims database. Our analysis starts with this claims database for services purchased by policyholders over the 2009 to 2013 period as it existed as of October 2014. The bulk of the database concerns canines, so our initial focus is on veterinary services for canines. From the complete set of canine treatments in the database, we exclude treatment codes not reflective of specific veterinary services for canines.[1] The resulting dataset contains 1,208 distinct treatments for canines. The dataset classifies 1,128 of these treatments as "medical" treatments; the remaining 80 are classified as "well-care" treatments.

To construct a monthly price index, it is important to have a large number of price observations of specific treatments that are spread over the entire time period being considered. We thus limit our analysis to canine treatments that are commonly observed in the claims data. In particular, we consider only treatments that have at least 5,000 claims over the five-year period from 2009 to 2013 and at least 500 claims each year during this period. This reduces the number of canine medical treatments considered by 89%, from 1,128 to 128, and reduces the number of well-care treatments considered by 53%, from 80 to 30. However, the resulting dataset of "common" treatments for canines still represents more than 80% of the total value recorded across all treatment codes.

It is typical for datasets on prices to have anomalies, specifically cases where prices are well outside the norm. To limit the potential effect of such "outliers," we drop from the dataset such outliers. For our purposes, outliers are defined as claims for a single canine treatment code with a claim value equal to or below one dollar (21,744 cases) or equal to or above $10,000 (119 cases). All 119 cases of "high" outliers involve medical treatments. Note that such outliers are not a significant part of the dataset. In particular, the total claim value of the dataset of common canine treatments that excludes outliners represents more than 99.8% of the total claim value of the dataset that includes the outliers.

Overview of price indices

A price index provides a measure of the average price change for a fixed basket of goods and services. For veterinary services, the calculation of such a fixed-weight price index involves re-pricing

[1] Examples of treatment codes that are excluded include the treatment codes for duplicate claims, taxes and ineligible services such as boarding. These excluded treatment codes represent 17.3 percent of the more than 15 million treatments recorded in the full claims database as being purchased by policyholders during the period 2009-2013.

for each period (month) the same basket of veterinary services and comparing the resulting aggregate expenditures with the total expenditures for the basket of services in a designated base period. Formally, let q_{it} denote the quantity of veterinary services of type *i* purchased by a representative individual in period *t*. Let p_{it} denote the price of service *i* in period *t*. If there are n different services purchased, and we designate the base period as period *o*, then the fixed-weight (Laspeyres) price index for period *t* for this "basket" of *n* services is given by:

$$Index_t = \frac{\sum_{i=1}^{n} p_{it} q_{io}}{\sum_{i=1}^{n} p_{io} q_{io}}$$

Expanding this expression, we have:

$$Index_t = \frac{p_{1t}q_{1o} + p_{2t}q_{2o} + \ldots + p_{nt}q_{no}}{\sum_{i=1}^{n} p_{io} q_{io}}$$

Rearranging, we obtain the following expression for the price index:

$$Index_t = \frac{p_{1o}q_{1o}}{\sum_{i=1}^{n} p_{io}q_{io}}\left(\frac{p_{1t}}{p_{1o}}\right) + \frac{p_{2o}q_{2o}}{\sum_{i=1}^{n} p_{io}q_{io}}\left(\frac{p_{2t}}{p_{2o}}\right) + \ldots.. + \frac{p_{no}q_{no}}{\sum_{i=1}^{n} p_{io}q_{io}}\left(\frac{p_{nt}}{p_{no}}\right)$$

The above expression for the price index incorporates two terms. The first is a set of "expenditure weights"

$$w_{io} = \frac{p_{io}q_{io}}{\sum_{i=1}^{n} p_{io}q_{io}}, \quad i = 1, 2, \ldots.., n$$

that denote the expenditures on service *i*, $i = 1, 2, \ldots, n$, as a proportion of the total expenditures on all services in the base period. The second is a set of "price change" terms

$$1 + \pi_{it} = \frac{p_{it}}{p_{io}}, \; i = 1, 2, \ldots, n$$

that equal one plus the rates of change in the price of service *i,* $i = 1, 2, \ldots, n$, between the base period *o* and period *t* (π_{it}). Thus the price index for period *t* can be viewed as simply a weighted average of the price changes for each of the services in the basket between the base period and period *t*, with the weight for each service equal to the service's expenditures in the base period as a proportion of total expenditures on all services during the base period. That is:

$$(1) \quad Index_t = \sum_{i=1}^{n} w_{io}(1 + \pi_{it})$$

Identification of the price index components

The price index formula (1) requires a set of weights across treatments and a set of price changes for each treatment across time. The set of weights is calculated from the annual aggregate claims dataset for common treatments during 2013. In particular, the weights equal the proportion of total claims in the dataset that each of the treatments reflects during 2013. As such, these expenditure weights for the entire sample of treatments sum to one in each period. To calculate weights for various subsets of the treatments (e.g., medical vs. well-care treatments), we rescaled the weights so that within each set of treatments, the expenditure weights sum to one.

With respect to set of price changes, we identify each month the median claim amount across all observations for each treatment. The ratio of this median price for a particular treatment in a particular month to its average median value in the base year represents one plus the rate of price change for that treatment between that month and the base year. The base year price for each treatment is defined as the average of the median prices that existed for that treatment in the base year (2013). Later when we calculate price changes for particular areas, we identify median prices by treatment and period for each area.

Converting price indices to three-month moving averages

To smooth out short-term fluctuations in the price indices computed, we create a three-month moving average of each price index series. Such moving averages are commonly used when examining time series data in economics. The moving average we employ is the simple mean of the current and previous two months values of the price index. As such, it is an equal weight moving average calculation.

Introducing price indices by region

We consider two types of geographic price indices. One considers prices across geographical areas divided by collections of states. In particular, we consider four regions: Northeast, Midwest, South, and West. We adopt the characterization of regions provided by the U.S. Census Bureau. The census bureau's classification (regions, divisions within each region, and states within each division) is as follows:

- **Northeast:** New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont) and Mid-Atlantic (New Jersey, New York, and Pennsylvania)
- **Midwest:** East North Central (Illinois, Indiana, Michigan, Ohio, and Wisconsin) and West North Central (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota)
- **South:** South Atlantic (Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, Washington D.C., and West Virginia), East South Central (Alabama, Kentucky, Mississippi, and Tennessee), and the West South Central (Arkansas, Louisiana, Oklahoma, and Texas)
- **West:** Mountain (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming) and Pacific (Alaska, California, Hawaii, Oregon, and Washington)

Providers of veterinary services are divided into the above four regions based on the state recorded as part of the addresses of the providers. In computing these regional price indices, we use the same weights across the regions; these weights reflect U.S.-wide treatment purchases during 2013 and are identical to the weights used for the aggregate price index. We use an interpolated value for the median price for the few cases when a price observation is not available for a particular treatment in a particular region on a particular date.

Introducing price indices by population density

The second type of geographic price index calculated the groups' geographical areas based on the density of population in the area of which the provider is located. We divide areas into three groups (Urban, Suburban, and Rural). To define urban, suburban, and rural area price indices, we start with the Census Bureau ZIP code tabulation areas (ZCTAs).

ZCTAs attempt to coincide with ZIP code areas as defined by the U.S. Postal Service. However, because they combine census blocks that have their own boundaries, ZCTAs sometimes do not perfectly coincide with ZIP code areas. In particular, some ZCTAs contain addresses from more than one ZIP code, such that a ZIP code may not have a corresponding ZCTA. When there are multiple ZIP codes in a ZCTA, the ZCTA uses the most frequently occurring ZIP code. If a ZIP code was never the most frequently occurring ZIP code in a ZCTA, then it will not appear in the ZCTA data. For each ZCTA, we determine the square miles of land that it represents as well as the total population level in the ZCTA according to the 2010 U. S. Census. Dividing the population by the land area provides us with a measure of the density of the population for the ZCTA.

We then match the ZCTA data to the ZIP code recorded in the claims dataset as part of the addresses of the providers of veterinary services.[2] For a small fraction of treatment observations (.3%), a provider's ZIP code does not match any ZCTA. In 99% of such cases, we were able to approximate the population density surrounding the provider by using the population density of the county of the provider.

We identify three categories of settings for providers in terms of the population density of the area in which they are located. Urban settings are defined as areas with a population density equal to or above 5,000 people per square mile. Suburban settings are areas with a population density between 1,000 and 5,000 people per square mile. Rural settings are areas with a population density less than or equal to 1,000 people per square mile. This breakdown resulted in 25% of treatments being provided by veterinarians in rural areas, 51% being provided by veterinarians in suburban areas, and 24% being provided by veterinarians in urban areas.

In computing area price indices by population density, we use the same weights across the three types of areas; these weights reflect U.S.-wide treatment purchases during 2013, and are identical to the weights used for the aggregate (country-wide) price index. We use an interpolated value for the price in the few cases when a price observation is not available for a particular population density group on a particular date.

Substitution bias issues

The above "fixed-weight" price index has intuitive appeal. However, such calculations can introduce a "substitution" bias over time when compared to a "cost of living" price index constructed to determine the changes in income necessary to keep individuals at the same level of satisfaction across time. The reason for this bias is that the base year basket of goods and services may not be the optimal basket of goods and services for consumers over time if there exists changes in relative prices within the basket.

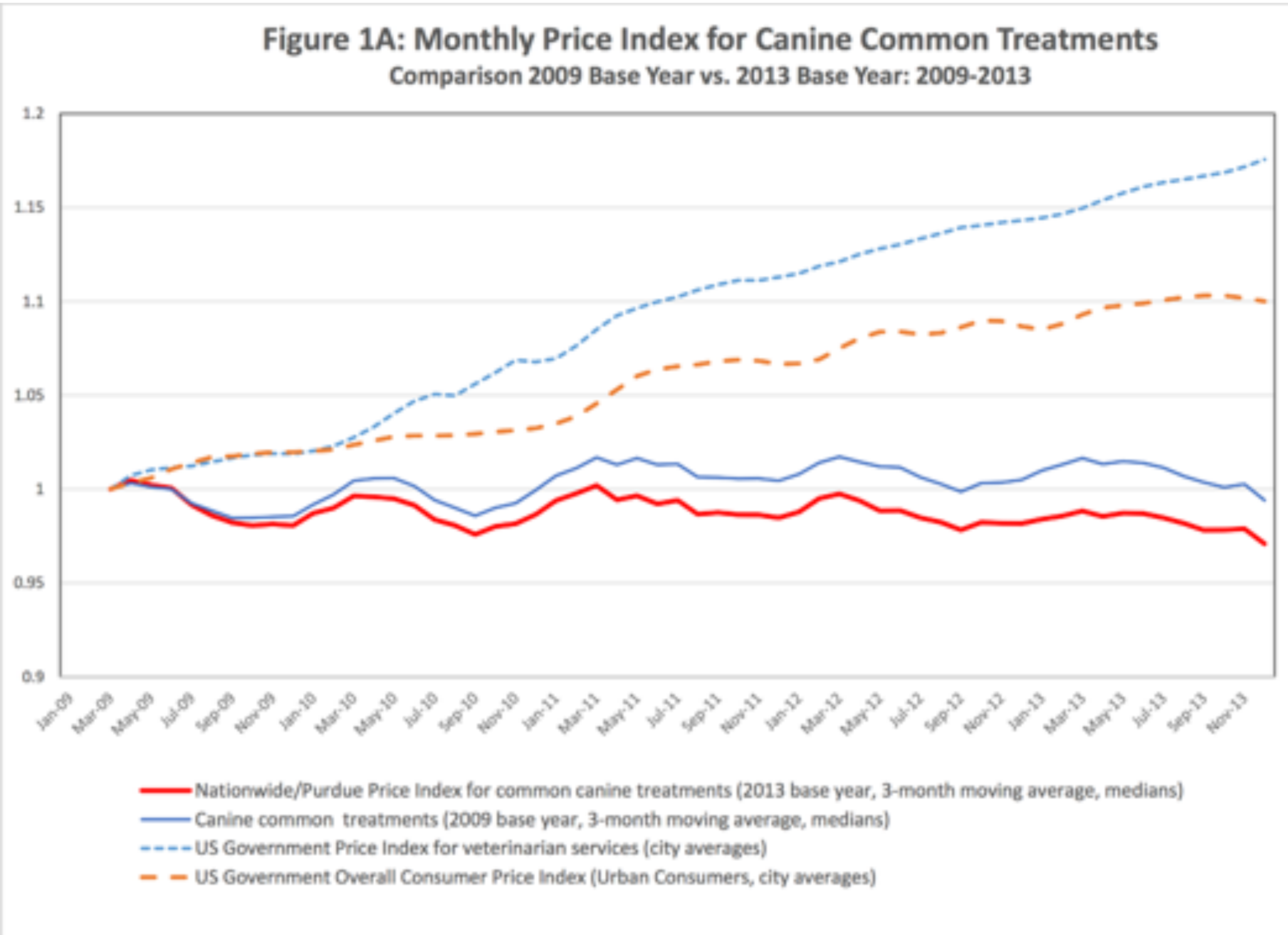
To see why this may be the case, let us first formerly define a "cost of living index" (COLI) for period *t*. This COLI index is the ratio of the expenditures required in period *t* to the expenditures in the base period such that the individual achieves the same level of satisfaction ("utility") in period *t* as in the base period. If prices of the various goods and services do not change at the same rate between the two periods, individuals will find that the basket of goods purchased in the base period can be improved upon by substituting away from goods and services whose relative prices have increased towards goods and services whose relative prices have decreased. The result is that the simple (Laspeyres) price index with fixed quantity weights has a tendency to overstate the "true" cost of living index (COLI), as it fails to account for individuals mitigating the effect of relative price changes on their satisfaction by altering what they purchase.

One way to address this substitution bias issue is to combine an index computed using weights from a base-period (a Laspeyres price index) and an index computed using current period weights (a Paasche price index). However, we have chosen to retain the intuitive appeal of the fixed-weight price index. Given our specification designates the base year to be 2013, the substitution bias suggests that price changes before 2013 will tend to be understated using our 2013base year price index when compared to a "true" cost of living index.

Figure 1A illustrates, as expected, that our price index with the base year of 2013 does indicate a lower rate of inflation over the 2009-2013 period than one that used 2009 the base year. However, the differences between the two price indices in terms of the estimated annual price changes are relatively

[2] There is a small fraction of treatment observations (.3%) that are missing the provider's ZIP code information. For these cases we use the ZIP code information of the policyholder to approximate the location of the provider.

small—fractions of percentages. In particular, the annual rate of price changes over the 2009-2013 period using 2009 as the base year is, on average, 6% higher each year than the annual rate of price changes using 2013 as the base year.



Authors

For Nationwide (VPI): Carol McConnell, DVM, MBA, Chief Veterinary Officer, and Kerry O’Hara, Ph.D., Director of Research, Data and Strategy.

For the Krannert School of Management, Purdue University: John M. Barron, Ph.D., and Kevin Mumford, Ph.D.

NationwideVetChannel.com