

Trends in the use of centralised pricing data sources for price indexes

Dr. Susan Kluth
Australian Bureau of Statistics

Abstract

The growing availability of administrative datasets such as government transaction databases, centralised business databases and industry datasets provide opportunities for the construction of reliable measures of price change for products for which traditional methods proved unsatisfactory. This paper describes how administrative datasets have been used to improve the reliability and / or coverage of the Australian CPI. Administrative data are used in the measurement of price change for deposit and loan facilities, child care, medical services and property rates. Practical issues of incorporating administrative data into price indexes are discussed as well as future directions.

Keywords: *administrative datasets, deposit and loan facilities, child care, medical services, property rates*

Introduction

The use of administrative datasets for obtaining price information for a consumer price index (CPI) can provide a number of advantages over traditional price collection methods. Point of sale data such as scanner data has the potential to dramatically increase efficiency of price collection and the number of prices collected, while reducing costs. However, these sources do not necessarily provide a better measure of price change. Prices obtained for supermarket items are identical whether collected directly or obtained from scanner data. Other administrative data, such as regulatory data, may be used for price measurement where non-linear pricing structures are used (for example water becomes more expensive as an individual's consumption increases) and traditional price index measurement techniques based on observed prices have proved ineffective or impracticable.

Outlined below are a number of examples where administrative data have been used to obtain improved measures of price change for the Australian CPI. Administrative data are currently used for pricing deposit and loan facilities, child care, medical services and property rates. Existing methods are compared with the alternatives and issues arising from the use of administrative data in this manner are discussed. Other difficult areas to price, such as telecommunications, utilities and insurance are expected to be added in the near future.

Deposit and loan facilities

A major part of bank revenue is derived from the provision of deposit and loan facilities. Banks charge explicit fees for account-keeping services and selected transactions and lend funds at a higher rate of interest than they pay on deposits. The amount paid by a household depends on a number of factors such as: the type of

accounts, transaction patterns, and balances in accounts. For example, a \$6 account keeping fee may be payable on accounts in general but waived on accounts with balances in excess of \$50,000. Furthermore, financial institutions often bundle products together, with the price paid for particular products depending on the bundling arrangements. In these circumstances it is impossible to calculate an average price from observable schedules.

One approach to measure price change would be to develop a number of user profiles and price these according to the charging rules for each financial institution. Hypothetical profiles could be constructed based on a pattern of average use or other characteristics. However, this process is costly, requires a high degree of cooperation from financial institutions and yields only a small number of profiles. The option the ABS has pursued is to make use of sampled actual customer accounts from financial institutions. This provides a much larger volume of data for much less effort than that for generating theoretical profiles. The sampled profiles also reflect actual customer behaviour.

To use the sample of accounts approach, data on balances and interest flows by product and in aggregate are collected each month, to enable calculation of a percentage interest rate margin for each product.

Each month, the samples of accounts are rolled forward by updating the transaction values (and balances) by movements in the CPI (lagged) to derive “real” values. The current percentage margin rate, calculated as described above, is applied to the updated average balances of the sampled accounts to obtain a monetary value for interest margins. These amounts are added to the amounts calculated for fees and charges by applying the charging rules for the relevant financial institution (generally available from the Internet) giving a total price paid for each sampled account.

There are several issues associated with the sample of accounts approach. The high respondent burden and ABS’s high level of reliance on the data make it essential to develop and maintain a good relationship with the service provider. While the accounts are stripped of all identifying information such as names and addresses they still contain real data pertaining to actual customer transactions. The ABS’s excellent data security record gives the providers the necessary assurance that their sensitive data will be handled appropriately. To ensure that the sampled customer accounts continue to reflect contemporary consumer behaviour, consideration needs to be given to how often these samples need to be refreshed. Current intentions are to do this annually.

This approach has required the ABS to enhance its computing system to process and manage the large volumes of data involved in the process. The system was enhanced further to apply the charging rules to the sample of accounts.

Child care

The price of child care required for the Australian CPI is the price paid by the consumer less all government subsidies. These subsidies depend on family income, the number of children in care, whether the care is full-time or part-time and whether a child is cared for in a child care centre or a private home. The complexity of the

resulting fee schedule does not support the estimation of aggregate (or the overall average) rate of price change.

The most attractive approach is to obtain profiles of family circumstances. This could be achieved by surveying households. However, this would yield a small sample, be expensive and incur a high respondent burden. As an alternative, the government body responsible for administering the child care subsidies was approached to provide the family profiles. Randomly sampled, confidentialised family profiles that reflect the spread of attributes (such as income, number of children in care and hours in care for each child) throughout the population of families that receive the child care subsidies are obtained. This allows the ABS to apply government subsidy rules to family profiles to calculate the subsidies, and in turn apply these to the prices which are obtained directly from selected child care centres.

To ensure the sample family profiles remain representative of the current population, the incomes in the sample profiles are indexed quarterly using the labour price index. Simply indexing the incomes is not sufficient to ensure long-term representativeness of the sample so each year a new sample of family characteristics is obtained and the sample is refreshed.

As for the deposit and loan sample of accounts, a good relationship must be maintained between the data provider and the ABS to ensure the ongoing supply and quality of the data provided. It was also necessary to build an interface to the current system to process the quantity of data provided.

Medical Services

Prices for medical services in the CPI are collected net of any government subsidies. There are two main subsidies that apply to these services – a flat rate depending on the nature of the consultation, and a safety net, which is an extra percentage subsidy on remaining costs, once an annual cost to the consumer has been exceeded. Historically, prices for medical services were obtained directly from a limited number of medical practitioners. Subsidies were then applied to the prices using average rates provided by the relevant government body. Several problems existed with this approach. First, medical services are represented by a large number of highly specific procedures. It is difficult to directly collect a sufficient volume of prices to accurately reflect true price changes, both from a cost and respondent burden point of view. Second, the data for applying the subsidies were only available in the format appropriate for use by the government body rather than the format required for the CPI. This led to issues such as a lagging and the inability to decompose data to determine the effect of various influences.

To resolve these difficulties a new approach has been adopted. It makes use of aggregate data acquired from the government health insurance scheme. This provides total fees, subsidies, and number of consultations for each of the several thousand types of medical procedures. Data pertaining to a number of representative procedures are extracted from the initial dataset and processed appropriately. The number of prices has increased by several orders of magnitude and the data are sufficiently

detailed to allow in-house calculation of the subsidies, allowing understanding of contributions to price change.

One danger of using aggregate administrative data is that individual prices are not obtained, so direct sources of price change are not available. To minimise this effect, diagnostic tools were developed to attribute price movements to a variety of different sources, for example: changes in prices charged by medical practitioner; changes in subsidies, and changes in the number of concession patients.

Use of aggregate administrative data has distinct limitations. As data are not obtained on a unit record basis it is not possible to obtain a perfectly matched sample, i.e., it is not possible to compare prices between two periods for services provided by the same medical practitioner. This may lead to compositional change in the CPI basket. In general, this is viewed as a small sacrifice to make for the vast improvement in coverage and increase in price reliability. However, the Australian CPI is compiled on a state capital city basis and the population of some of the capital cities is rather small. In these cities, significant problems may be encountered with compositional change. For example, if only a low number of specialists, with different pricing schedules, practise in a field and one took a holiday then the average price will be significantly changed. Likewise, if a provider enters or leaves a small market, false price changes can be observed. These can be managed by appropriate editing procedures. Other potential issues such as quality and timeliness arise from ABS reliance on the government department providing the data.

Property Rates

Property rates are dependent on the land value of the property. Formerly, they were calculated by requesting the current land value for a fixed list of properties from local councils and applying the charging rules to the properties in the list. As land values change differently depending on location within a city, a wide coverage was necessary to accurately determine average price change. This proved burdensome for the providers. The ABS now uses unit record administrative data to facilitate a matched sample approach to calculate property rates in some capital cities. The land values are provided for all dwellings within a city and an appropriate algorithm is applied to calculate the relevant property rates.

Unit record data makes it possible to match individual dwellings in current and previous periods to allow direct price comparisons. It also reveals properties for which only concessional rates are charged. The unit record approach is extremely effective due to the relative simplicity of the price calculation, and the less sensitive nature of the information. An issue is the different data formats used by different cities or within cities. The data processing system must be able to accommodate these differing formats. It is also important to keep abreast of the various levies charged by individual councils.

Future directions

The ABS is currently exploring options for using the sample of accounts approach to price telecommunications, utilities and general insurance. Pricing of telecommunication products generates similar problems to those associated with

deposit and loan facilities. Each provider offers a large number of different telephone plans with prices differing depending on factors including the time, duration and distance of the call. Compared to the charging rules implemented by financial institutions, charging rules for telephone services are an order of magnitude more complex. Further complications are encountered due to the rapidly changing nature of mobile telephone plans.

Most utilities in Australia are subject to non-linear pricing. For example water and electricity are subject to step function pricing where consumption exceeding a certain usage threshold is charged at a higher rate. Several such steps may exist depending on the utility. In addition, electricity is priced according to peak and off-peak usage as well as at concessional and non-concessional rates. In the future this area is expected to become increasingly complex to measure price change accurately.

The technology exists to enable pricing on an hour-of-the-day basis although this is not yet operational on a wide scale. Price calculation is further complicated by the concept of “bundling”. Many providers offer a number of services, such as electricity, gas, telephone and pay television. Customers consuming more than one service from a single provider may receive multi-service discounts. This is extremely difficult to price reliably in a conventional manner. Work is currently underway to price electricity from a sample of accounts. Once this is operational other utilities will be included as well, including bundling arrangements.

The field of general insurance is rapidly increasing in complexity, as risk calculation becomes increasingly granular in nature. It is envisaged that a point will be reached at which an individual risk factor will be calculated for individual properties. Due to the highly complicated and sensitive nature of the risk algorithms used by insurance companies as well as their rapidly changing natures, it is not proposed that the ABS performs the calculation of the premiums. The methodology currently proposed is to obtain samples of policies from the insurance companies, adjust these each period to reflect current period values of various factors, such as the value of the house, and return these to insurance companies for pricing within their own computer system. Regular random checks will need to be conducted to ensure integrity of the data provided.

As the ABS expands its use of administrative data for the calculation of the CPI, initial investigations into point of sale data such as scanner data will also be continued. While this field has not yet received a great deal of attention, due to other more pressing priorities, its potential for reducing costs and increasing coverage makes it highly attractive.

Summary

The use of administrative data in the compilation of the Australian CPI allows price measures to be constructed in areas, such as service industries, where it is difficult or impossible to price using traditional pricing methods. For successful exploitation of administrative data emphasis must be placed on fostering relationships with providers. Their trust and cooperation is essential. However, it will not be achieved without appropriate security surrounding the data. In this respect, the ABS's excellent track record of preserving the secrecy of information has proven invaluable.

The large volumes of data and variety of data formats demand appropriate technological infrastructure. Consideration needs to be given to how often samples need to be refreshed and the providers' costs in supplying these data to the ABS. Appropriate documentation and the training of personnel to deal with the analysis and validation of such data and the understanding of business processes are also critical. In particular, personnel need to minimise compositional change so that the service is priced to constant quality.

Used appropriately, administrative data are a useful tool and, given the increasing level of complexities being observed in pricing structures, it will be essential for producing a high quality CPI in the future.

References

Australian Bureau of Statistics, *Experimental Price Indexes for Financial Services 1998 to 2003*, cat. no. 6554.0

Australian Bureau of Statistics, *Consumer Price Index September 2007*, cat. no. 6401.0 (Appendix: *Child care services in the CPI*)