18th MEETING OF THE VOORBURG GROUP

Tokyo, Japan
October 2003

SESSION 2: PRICE INDEX FOR PRE-PACKAGED SOFTWARE

Fred Barzyk
Statistics Canada

1. Introduction

Until recently, the deflation of pre-packaged software in the Canadian System of National Accounts (CSNA) at Statistics Canada (STC) relied on using the pre-packaged software price indexes produced by the US Bureau of Labor Statistics (BLS), adjusted for the Canada-U.S. exchange rate. As with any proxy series, there is concern that deviations or different behaviours entirely might exist between the proxy series and a series based on actual Canadian pricing.

As a result, in the spring of 2003 development of a Canadian price index was undertaken at STC, resulting in a monthly price index series for pre-packaged software. This new index is produced by combining two sources of data: 1) scanner data from AC Neilsen to provide the weights, and 2) price data obtained from International Data Corporation (IDC) of Canada. It should be noted that because the pricing and expenditure data are for software sold in Canada (domestically produced and imported software) at final purchase or 'street' prices, the index is *not* a producer index for software. Instead, it is to be used for the deflation of investment expenditures of businesses and of governments in the final demand category of GDP.

This paper provides a brief summary of the key research carried out at STC in this area, and describes the development and results of the new price index for pre-packaged software.[1] A conclusion and summary of ongoing work is also provided.

2. Previous Research at STC

In 2002, key research was undertaken for the development of a price index for pre-packaged software, culminating in a research paper by Prud'homme and Yu in 2002 (referred to as PY from this point on). The results of this research were important to the subsequent development of the current software price index used by Statistics Canada.[2]

Their study looked at a number of issues relating to software and to producing a price index for software. PY point out that the growing importance of software in the economy can be demonstrated by the fact that, in terms of its share of investment expenditures in machinery and equipment, pre-packaged software expenditures rose from less than 1% in 1981 to almost 10% in 2001. The growth in software spending has outpaced that of computer hardware, with the ratio of pre-packaged software investment to computer hardware investment rising from 19% in 1986 to 65% in 2001. From a household perspective, software purchases as a percentage of total household expenditure nearly doubled from 0.05% in 1992 to 0.09% in 2000.

---

[1] The author and Statistics Canada would like to acknowledge the substantive effort and contribution of one of its former analysts, Dimitri Sanga, to the development of this index and its documentation.

[2] Pre-packaged software is one of the three sub-categories that make up the overall category for software, used by the System of National Accounts at Statistics Canada. The other two are custom software and own-account software. For more details, see Jackson (2000).

PY conducted a review of literature in order to assess and understand the issues and problems associated with producing a constant-quality price index for software. They found that there was limited research on the use of hedonics for producing constant-quality price indexes for software and, in their opinion, this was largely because hedonic studies on software prices require large amounts of good quality detailed data, software characteristics are difficult to identify, and knowledge and understanding of the software industry by economists is generally lacking.

Some notable studies cited by PY include: Gandal (1994, 1995), Brynjolfosson and Kermerer (1996), Oliner and Sichel (1994), Harhoff and Moch (1997) and Moulton, Parker and Seskin (1999).[3] The highlights of these studies are presented in Figure 1 (*Table 8: Comparison of Price Index Studies in Computer Software)*, where the calculated average annual change (AAC) ranges from -16.0% for Brynjolfosson and Kermerer to -1.5% for Gandal (1995) for databases. With the exception of the outlier results obtained by Gandal (1994) and Brynjolfosson and Kermerer, the AAC is on average -4.5%.[4]

Figure 1 – Reproduction of PY Table 8.

Table 8: Comparison of Price Index Studies in Computer Software

Study	Method	Type	Country	Period	AAC (%)
Gandal (1994)	Hedonic	Spreadsheet	U.S.	1986-1991	-15.0
Gandal (1995)	Hedonic	Spreadsheet	U.S.	1989-1191	-4.4
	Hedonic	Database	U.S.	1989-1991	-1.5
Brynjolfsson and Kemerer (1996)	Hedonic	Spreadsheet	U.S.	1987-1992	-16.0
Harhoff and Moch (1997)	Hedonic	Database	Germany	1986-1994	-7.4
	Matched	Database	Germany	1986-1994	-4.4
Oliner and Sichel (1994)	Matched	Word Processing	U.S.	1985-1993	-2.6
	Matched	Spreadsheet	U.S.	1985-1993	-4.5
	Matched	Database	U.S.	1985-1993	-4.7
BLS	Matched	General	U.S.	1998-2000	-6.6

PY chose to forego the use of hedonics in constructing a price index for pre-packaged software for three reasons. First, their data set did not include the details about the characteristics of each product. Second, PY felt they lacked sufficient product knowledge to produce separate hedonic estimates for each of the categories available (34 categories in 1996). Third, the degree of diversification in some categories such as games and education makes it very difficult to find a common set of characteristics to accommodate the hedonic methodology.

Given these deficiencies, PY decided to use the 'maximum overlap' method to construct their price index, first suggested by Turvey (2001). Under this method, the price of a product in the base period is compared with the price of the product with the identical attributes or characteristics in the comparison period. When an existing product disappears or is replaced by a new product with

[3] In a study not cited by PY but pertinent to this topic nonetheless, McCahill (1997) validated the results of earlier hedonic price studies concerning the US market for spreadsheet software and concluded that is was feasible to build a reliable econometric model for the purposes of producing a hedonically-adjusted price index series for word processing software.

[4] In a more recent study Abel, Berndt and White (2003) estimated that prices for Microsoft Products fell at an annual rate of 4.26% between July 1993 and June 2001. The results were derived using a matched model process.

different characteristics, it is deleted from the sample. The new product is included in the sample to be matched in the next period. Matching *all* the products that are commonly available in two adjacent (overlap) periods leads to the maximum overlap method.

The data used by PY consisted of scanner data obtained from AC Nielsen, where samples were collected electronically and manually from twenty-eight major chain stores and independent outlets across Canada. The data are monthly and cover the period January 1996 to June 2000. The software products are categorized according to their type (e.g. games, word processing, etc.). A category typically includes the software title, its sales volume, total generated revenue, and resulting average price (total revenue divided by sales volume for each title). Unfortunately, the data lacks information on who is purchasing the software (i.e. businesses or consumers). In order to produce a business versus consumer split, PY allocated software titles and categories based on what software was presumed to be typically purchased by businesses or by households. In cases where overlap existed they allocated half the expenditures to each purchaser.

Based on this data, PY calculated revenue shares and found the most important business software category to be *Suite* (30.2%), followed by *Operating Systems* (12.2%) and *Networking* (6.3%), with the remaining categories (26 or 51.3%) each having shares under 5%. For the consumer software group, *Games* held the largest share (36.2%), followed by *Operating Systems* (18.0%), *Edutainment* (7.6%), *Education* (7.1%) and *Hardware Utility* (6.1%). A large proportion of categories (9 out of 14 or 64.3%) all had shares of less than 5%.

Using the maximum overlap method, PY calculated Laspeyres, Paasche and Fisher versions of price indexes for the various categories of business and consumer software. The Fisher version for business software results in an average annual decline of 4.4% for the overall index. Within this group, the growth rates vary from -18.7% for *Antivirus* to 18.5% for *Networking*. On the consumer side, the PY index falls at an average annual rate of 7.9%, with the range being -15.5% for *Education* versus 22.9% for *Communication*.

When compared with the BLS monthly price index series for computer software and accessories for the period beginning December 1997 and ending June 2000, PY found that their overall index for business software fell faster (-8.1% versus -5.8%) and in a more erratic fashion. They attribute the divergence in the growth rates to the use of different index formulae and different market conditions, while the lower volatility of the BLS series may be due to the inclusion of the *'other computer accessories'* category having a stability effect on that series.

Apart from the actual index series results themselves, perhaps the most notable contribution of this study is the exposé of the data problems encountered with their scanner data sample. For example, PY encountered labelling problems between versions such that it was not possible to determine which version of the software was in the sample, leading to a substantial effort to subjectively match products and product versions. Additional data anomalies included the discovery of negative prices in the data, extreme increases or decreases in price relatives at the micro-level, and the existence of fractional quantities resulting in exceptionally high prices. All of these problems resulted in the need to clean and trim the data by excluding these types of outliers.

3. Current Index Methodology

In developing and implementing a price index for pre-packaged software for Prices Division at STC, several conclusions from the PY research were adapted. It was decided to use the maximum overlap method, for the same reasons PY cite in their research, the main one being a lack of sufficient product characteristics to attempt the use of hedonics.

However, it should be noted that one of the weaknesses of the overlap method is the potential for bias to occur in the index. The premise behind the overlap method, as it applies to quality adjustment, is that the average price ratio of the matched products is the same as that of the products not matched. When new products come into the market, the prices of the existing old versions are bid down, reflecting the difference in quality between the old and new versions. An additional assumption is that the new versions are introduced *excluding* any pure price change. If this does occur, and Triplett (1997) makes the case that it often does, then the pure price changes cannot be captured by the matched samples. Also, PY mention the point (made by others as well) that when the market is slow to adjust to a new equilibrium, a matched model price index will be biased. The direction of the bias depends on whether the new products are priced upward or downward. In their study of quality adjustment methods for the price of computers, Barzyk and MacDonald (2001) compared results obtained using the overlap method to those using several versions of hedonic indexes and found upward bias in the case of the overlap method. In their view, the market for computers does *not* behave to the extent that the entire quality difference between old and new versions of computers is captured in the overlap method, certainly when compared to hedonic-based series.

A second adaptation is that scanner data would be used, but *only* for the calculation of pertinent weights. Scanner data contains information on volumes and unit values, making it highly suitable for constructing purchase weights at the category level. In terms of pricing, the major weakness associated with scanner data is the difficulty in tracking items, as found by PY. While the scanner data does include a variable describing the software title, it is often not detailed enough to determine which version of the software is sampled with any degree of certainty. This can have significant consequences for the construction of a price index because the matching process and results would be suspect.

Instead, price data obtained from the International Data Corporation of Canada (IDC) is used to produce the price relatives for the index. The IDC data has the advantage of having a unique identifier for each particular observation or software product (and version), which makes tracking or matching products very easy. Monthly price data is available on 2,000 products or more with corresponding information on software category, purchasing sector (i.e. consumer or business), manufacturing number, list price and street price. The list price represents the price set by the vendor (e.g. Mcafee, Corel, Microsoft or Adobe) while the street price is an estimate obtained by taking an average of the reseller cost by item.

Unfortunately, there is no accompanying shipment data from which to construct weights, but by combining the scanner data and the IDC data it is possible to benefit from the advantages of each source. As a result, a set of weights derived from the AC Neilsen scanner data for the various software categories form the basis for the index, while the price relatives, calculated using the maximum overlapping matched model method, are obtained from the IDC pricing data.

In order to integrate the two data sources, eight software categories were established and weights were derived for the business and government sector. These are presented in Table 1. In producing the index, monthly price relatives are calculated for each product matched and then the geometric mean is obtained for each category. The category weights, which are based on 2000 purchase values (scanner data), are then used to aggregate the series to produce a total price index for business and government.

Table 1: Software Categories and Weights

Category	Weight
Business	38.6%
Communication	8.2%
Developer	5.3%
Graphics	8.3%
Internet	5.5%
Networking	8.7%
Operating systems	15.2%
Utilities	10.3%
Total	100.0%

4. Results

Using this methodology, a monthly price index for business software was constructed for the period January 2000 to December 2002. The resulting index values are presented graphically in Figure 2. In addition, a quarterly index series was produced using the monthly indexes (see Figure 3). These series were compared to the corresponding BLS price index for pre-packaged software, and also compared to the quarterly exchange rate-adjusted version of the BLS index produced and used by the CSNA. The reason for comparing these series is to see how they move in relation to one another. The assumption behind using the BLS series as a proxy until now has been that the two markets for software behave similarly from a price perspective. Now this can be investigated.

Figure 2 - Pre-packaged Computer Software Price Indexes
- Monthly -

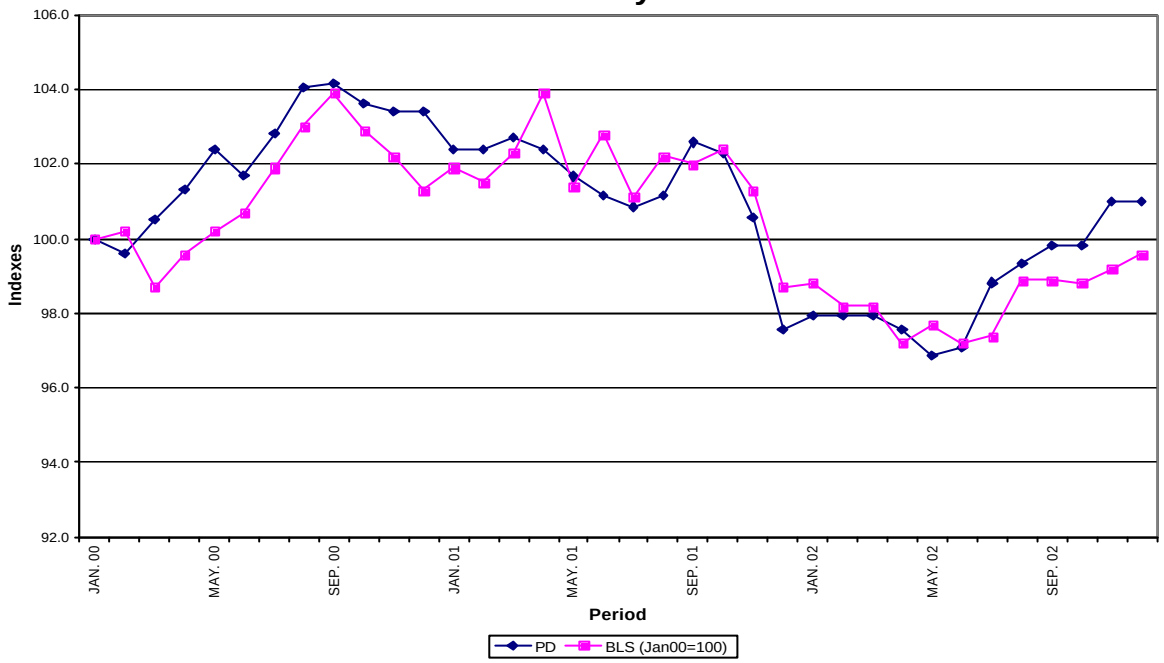
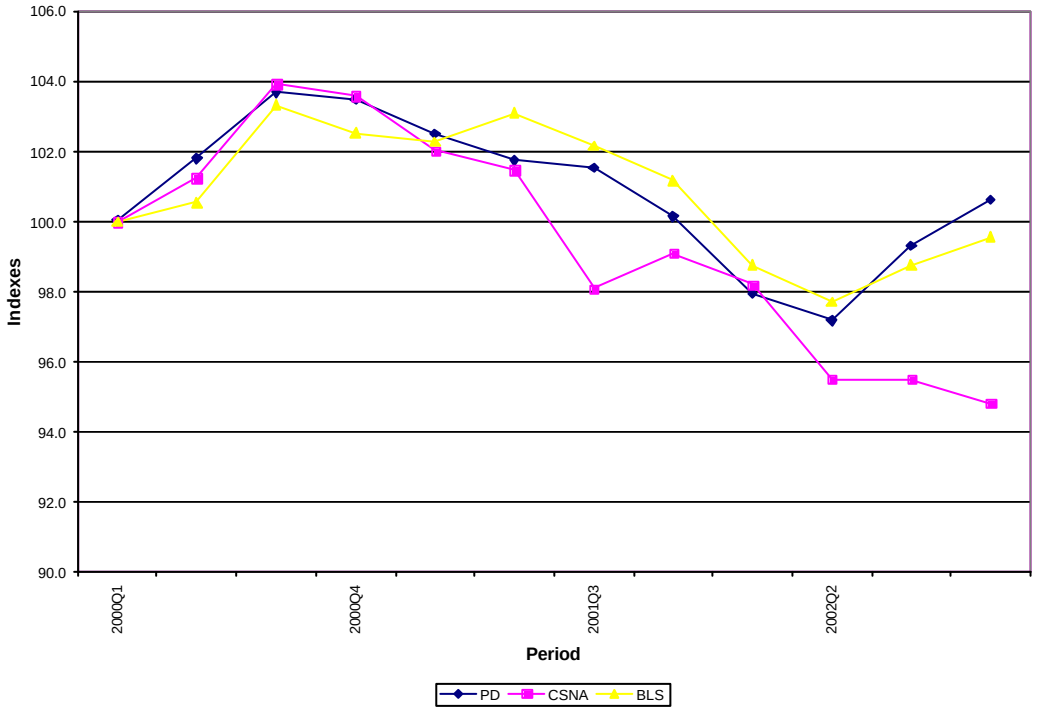


Figure 3 - Pre-packaged Computer Software Price Indexes
- Quarterly -



Graphically the series appear to move together and the estimated average monthly rate of change for both series is -0.1%. The quarterly series have very similar growth rates as well, with -0.34% for the Prices Division (PD) series and -0.29% for the BLS. Correlations between the series were also estimated to provide a statistical comparative measure and the results are presented in Table 2. For the monthly data, the correlation value between the BLS and the PD series is 0.876141 and for the quarterly data it is 0.903616. The correlation between the quarterly CSNA and PD series is weaker (0.795055) due to the exchange rate-adjustment applied to the BLS series. From these results it is evident that the PD and the BLS indexes move similarly from month to month and that they display the same trend downward throughout this period.

Table 2: Correlations Results

Monthly Index			
	PD	*BLS*	
PD	1		
BLS	0.876141	1	
Quarterly Index			
	PD	*CSNA*	*BLS*
PD	1		
CSNA	0.795055	1	
BLS	0.903616	0.812523	1

5. Conclusion and Ongoing Work

Previous research at STC has proven extremely useful in the development of a price index for pre-packaged software. This new index behaves in a very similar fashion to its US counterpart, while exchange rate-adjusting the BLS series imparts an unnecessary weakening in this relationship. The PD index would seem to provide some promising results and should be considered as a viable alternative the current series produced and used by the CSNA.

Ongoing work for this series includes readying them for official publication and developing a revision strategy for the weighting portion of the index. A hedonic-based alternative to the maximum overlap method might be considered should data become available as part of the ongoing research effort into producing constant-quality price indexes at STC.

References

Abel, J.R., Berndt, E.R. and White, A.G. (2003) "Price Indexes For Microsoft's Personal Computer Software Products" draft manuscript, (available from <jabel@analysisgroup.com> or <eberndt@mit.edu> or <awhite@analysisgroup.com>).

Barzyk, F. and M. MacDonald (2000). The Treatment of Quality Change for Computer Price Indexes – A Review of Current and Proposed Practices. Unpublished working paper. Prices Division, Statistics Canada.

Brynjolfsson, E. and C. F. Kemerer (1996) 'Network Externalities in Microcomputer Software: An Econometric Analysis of the Spreadsheet Market,' *Management Science*, Vol. 42, No. 12, 1627-1647.

Gandal, N. (1994) "*Hedonic Price Indexes for Spreadsheets and an Empirical Test for Network Externalities*", RAND Journal of Economics, Vol. 25, nº 1, 160-170.

Gandal, N. (1995) "*Competing Compatibility Standards and Network Externalities in the PC Software Market*", Review of Economics and Statistics, Vol. LXXVII, nº 4, 599-608.

Harhoff, D. and D. (1997) 'Price Indexes for PC Database Software and the Value of Code Compatibility,' *Research Policy*, Vol. 26, 4-5, 509-20.

Jackson, C. (2000) "*The Capitalization of Software in the National Accounts*", Statistics Canada Reports.

McCahill, R J. (1997) "A Hedonic Study of Prepackaged Software", Master's Thesis, Virginia Polytechnic Institute and State University.

Moulton, B., R. P. Parker and E. P. Seskin (1999) "*A Preview of the 1999 Comprehensive Revision of the National Income and product Accounts*", *Survey of Current Business*", U.S. Departyment of Commerce, Vol. 79, nº 8, 7-20.

Oliner, S. and D.E. Sichel (1994) "*Computers and Output Growth Revisited: How Big is the Puzzle?*", Brookings Papers on Economic Activity, 2, 273-317.

Prud'homme, M. and K. Yu (2000) "*A Price Index for Computer Software Using Scanner Data*", Discussion Paper, Prices Division, Statistics Canada.

Triplett, J. (1997) 'Measuring Consumption: The Post-1973 Slowdown and the Research Issues,' *Review*, May/June 1997, Federal Reserve Bank of St. Louis, 9-42.

Turvey, R. (2001) 'Consumer Price Index Methodology,' http://www.turvey.demon.co.uk/index.htm.