Australian Bureau of Statistics

6458.0

Information Paper

The Introduction of Hedonic Price Indexes for Personal Computers

2005

100 years of STATISTICS

www.abs.gov.au

New Issue

Information Paper

The Introduction of Hedonic Price Indexes for Personal Computers

2005

Dennis Trewin
Australian Statistician

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30AM (CANBERRA TIME) TUES 8 FEB 2005

ABS Catalogue No. 6458.0

Produced by the Australian Bureau of Statistics

INQUIRIES

- For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070 or Matthew Berger on Canberra (02) 6252 6170.

CONTENTS

PREFACE

The Australian Bureau of Statistics (ABS) has prepared this information paper as a basis for user consultation on a strategy that is being adopted for enhancing our statistical service in the field of prices statistics.

The ongoing challenge for prices statisticians is to price to constant quality. "A change in the quality of a good or service occurs when there is a change in some of its characteristics or when a replacement product is selected in the sample" (Producer Price Index Manual, paragraph 1.213). When characteristics of goods change, and it is not possible to price "like with like", a quality adjustment must be made to the observed prices. In effect, this process splits the overall price change into a *quality change* component and a *pure price change* component, with the latter entering directly into the price index calculation.

The ABS has long been aware of the need for improvement regarding price indexes for personal computers, and progress in this regard has been discussed in meetings of the ABS Economic Statistics User Group as well as other forums. When faced with measuring prices for products that undergo rapid quality change, international best practice is to develop *hedonic price indexes*, provided suitable source data are available. This approach is being advocated by international agencies such as the Organisation for Economic Cooperation and Development (OECD), the International Labour Organization (ILO) and the International Monetary Fund (IMF). National statistical agencies that have access to suitable data sources have adopted this approach as a standard solution to dealing with rapidly changing products. The ABS has been developing a hedonic index for some years but, prior to obtaining an ongoing reliable data source, has not been able to implement the index until now.

This paper presents the methodology that the ABS proposes to use to adjust computer prices for quality change, and plans for implementation.

Readers wishing to obtain further information on topics covered in this information paper should contact:

Mr Matthew Berger
Director
Producer Price Indexes
Australian Bureau of Statistics
PO Box 10
BELCONNEN ACT 2616

Telephone: 02 6252 6170
Facsimile: 02 6252 7060
email: matthew.berger@abs.gov.au

ABBREVIATIONS

ABS	Australian Bureau of Statistics
APMI	Price Index of Articles Produced by Manufacturing Industries
BEA	US Bureau of Economic Analysis
BLS	US Bureau of Labor Statistics
CPI	Consumer Price Index
Eurostat	Statistical Office of the European Communities
ILO	International Labour Organization
IMF	International Monetary Fund
IPI	Import Price Index
IWGPS	Inter-Secretariat Working Group on Price Statistics
OECD	Organisation for Economic Co-operation and Development
PPI	Producer Price Index
SOP	Stage of Production Producer Price Index
UNECE	United Nations Economic Commission for Europe
USA	United States of America

CHAPTER 1 EXECUTIVE SUMMARY

THE PROBLEM

1.1 Price indexes must measure pure price change, and therefore must make adjustments for changes in the characteristics (and therefore quality) of individual goods. Quality adjustment of the prices of commodities that undergo rapid changes is problematic. This is particularly the case for prices of personal computers. When confronted with this problem of rapid quality change, international best practice is to develop hedonic price indexes. Hedonic price indexes are calculated by forming functions that relate the price of a product to its characteristics.

CURRENT SITUATION

1.2 The ABS is currently using adjusted hedonic price indexes obtained from US statistical agencies. The hedonic indexes from the US are exchange rate adjusted and lagged to increase their suitability for the Australian context. Over the past few years sources of detailed Australian price data for personal computers have become available. These data have enabled the ABS to develop hedonic price indexes for personal computers purchased within Australia.

PROPOSAL

1.3 The ABS proposes to adopt the 'consecutive two period chained time dummy double imputation hedonic price index' for use in price indexes for personal computers. This proposal sees a matched model price index applied for personal computers sold between consecutive periods, combined with a consecutive-period time dummy price index to measure price changes for both discontinued and newly introduced goods.

1.4 The ABS believes that the proposed double imputation hedonic price index presented in this paper is robust and fit for purpose, and represents a substantial improvement over the existing practice based on adjusting hedonic price indexes for the United States by $US/$A exchange rate movements.

1.5 The ABS introduced this new methodology in the Producer Price Indexes (PPI) in September quarter 2003, and plans to include this new methodology in both the Consumer Price Index (CPI) and the national accounts from September quarter 2005. There will be no revisions made to the CPI for earlier periods, but there will be revisions in the national accounts. The effect on gross domestic product (GDP) will, however, be negligible.

1.6 The ABS has prepared this information paper as a basis for user consultation on the planned improvements in the way changes in computer prices are measured.

CHAPTER 2 REQUIREMENTS

REQUIREMENTS OF A PRICE INDEX FOR PERSONAL COMPUTERS

2.1 The ABS requires a price index for personal computers to have the following properties:

1. The price index must measure *pure price change*; in the context of personal computers, this means that the price index must correctly account for rapid changes in the capability of personal computers.

2. The price index must be *representative*; that is, the index should be representative of transactions occurring within the Australian computer market, and must account for both the price changes of continuing models of personal computers, and those price changes associated with new models with new and/or improved characteristics entering the marketplace.

3. Movements in the price index should be *easily explainable* to users. Specifically, the period-to-period movements in the price index must be decomposable to show the impact of the price changes for continuing models of personal computers, and the impact of the introduction of new models (with new and/or improved characteristics) to the marketplace.

4. The price index must be an *improvement over existing methods* that utilise data from United States of America (US) statistical agencies. In particular the price index must avoid the counter intuitive movements that can be observed at times when using US data adjusted by $US/$A exchange rates series.

CHAPTER 3 BACKGROUND

THE NEED FOR PRICE INDEXES FOR PERSONAL COMPUTERS

3.1 In the past few decades, the advancement in information technology has been one of the most dynamic sources of economic change and development throughout the world. Today, the use of personal computers and rapid expansion of the Internet have changed the nature of commerce, the wealth of nations and the way people communicate, work and live.

The importance of computers to Australia's economy

3.2 Australia is experiencing the impact of these technologies throughout its society and economy. Expenditure on computers and peripherals was 0.4% of GDP in 1996. By 2002, this had risen to 1.7%.

3.3 Furthermore, according to the ABS survey on the *Household Use of Information Technology 2001-02* (cat. no. 8146.0), 61% of households in Australia had access to a computer in 2002 compared with 53% in 2000. The proportion of employing businesses using computers also increased substantially, from 76% in 2000 to 84% in 2002 (*Business Use of Information Technology 2001-02* (cat. no. 8129.0)).

PRICE INDEXES FOR PERSONAL COMPUTERS

3.4 “A price index is a measure of the proportionate, or percentage, changes in a set of prices over time” (paragraph 1.1, PPI Manual (ILO et al, 2004)). Price indexes provide information about price movements of items, but provide no information about the price levels of the items. A consumer price index measures changes in the prices of goods and services that households purchase. A producer price index measures changes in the price of goods and services either as they leave the place of production, or as they enter the production process. Producer price indexes thus fall into two categories: input prices (that is, at purchasers' prices) and output prices (that is, at basic or producers' prices).

3.5 Price indexes are frequently used as indicators of short-term inflationary pressure. Although important in this context, a vital use of price indexes is as deflators in the compilation of the national accounts. Deflation is the process of removing price effects from value (i.e. price-volume) data in order to estimate underlying volumes. Price indexes are used in national accounts as deflators of household expenditure (components of the CPI), capital expenditure, inventories and sales (components of the PPI).

3.6 Price indexes for personal computers are used as components of the CPI Recreation group. In the PPIs, prices for personal computers can be found in the Price Index of Articles Produced by Manufacturing Industries (APMI), and the Import Price Index (IPI). Price indexes for personal computers also subsequently contribute to the Stage of Production (SOP) producer price index. Within the national accounts, deflators for personal computers are used in compiling the volume estimates of household final consumption expenditure, capital expenditure and inventories. The resulting measures are provided in the following macro-economic publications:

- *Australian System of National Accounts* (cat. no. 5204.0)
- *Australian National Accounts: National Income, Expenditure and Product* (cat. no. 5206.0)
- *Consumer Price Index, Australia* (cat. no. 6401.0)
- *International Trade Price Indexes, Australia* (cat. no. 6457.0)
- *Producer Price Indexes, Australia* (cat. no. 6427.0)

CHAPTER 4 THE PROBLEM

THE PRICE INDEX

4.1 A price index provides an average of the price changes in a group of products between one period and another. The average price change over time cannot be directly observed and must be estimated by measuring actual prices at different points in time. Price index numbers are compiled from the price observations in each period (i.e. month or quarter); their significance lies in a series of index numbers referencing the comparison of prices between a particular period and a reference base. For an index to provide information on price changes, at least two index numbers for the same series need to be available, and these index numbers should relate to the same "basket of goods".

A PRICE INDEX FOR A SINGLE COMMODITY

4.2 As described earlier, a price index provides a measure of the proportionate, or percentage, changes in a set of prices over time. As the prices of different goods and services do not all change at the same rate, a price index can only reflect their average movement. A price index typically assumes a value of 100 in some reference base period. The values of the index for other periods then show the average proportionate, or percentage, changes in price from the base period.

4.3 Table 1 shows an example of a price index. The index in this example is constructed for a single personal computer priced from a vendor's website over the period January 2004 through August 2004. This price is a purchaser's price - the price paid by a consumer to purchase the computer. The price index shows the relationship of the price of the personal computer in each month as compared with its price in the initial month (or base period) of January 2004.

TABLE 1: EXAMPLE OF A PRICE INDEX FOR A SINGLE MODEL OF PERSONAL COMPUTER

Period	*Price ($)*	*Calculation*	*Price Index*	*% change since Jan 04*	*% change since previous month*
Jan 04	1,649	1,649 / 1,649 x 100 =	100.0	. .	. .
Feb 04	1,630	1,630 / 1,649 x 100 =	98.8	-1.2	–1.2
Mar 04	1,630	1,630 / 1,649 x 100 =	98.8	-1.2	—
Apr 04	1,599	1,599 / 1,649 x 100 =	97.0	-3.0	–1.9
May 04	1,590	1,590 / 1,649 x 100 =	96.4	-3.6	–0.6
Jun 04	1,520	1,520 / 1,649 x 100 =	92.2	-7.8	–4.4
Jul 04	1,430	1,430 / 1,649 x 100 =	86.7	-13.3	–5.9
Aug 04	1,330	1,330 / 1,649 x 100 =	80.7	-19.3	–7.0

— nil or rounded to zero . . null cell

4.4 From this example we see that the price index for the month of August 2004 is 80.7, meaning that the price of the computer decreased by 100.0 - 80.7 = 19.3 % since January 2004.

A PRICE INDEX FOR MANY ITEMS

4.5 Whereas a price index for a single item reflects the price changes for that item alone over a period of time, a price index for many items reflects the average price change for all such items. Table 2 shows price data for ten (10) different personal computers priced from a vendor website in January and February 2004.

TABLE 2: PRICES FOR TEN DIFFERENT PERSONAL COMPUTERS, JANUARY AND FEBRUARY 2004

	Price		*Price relatives*
Model	*Jan 04 ($)*	*Feb 04 ($)*	*Ratio of prices: Feb 04 / Jan 04 x 100*
Computer Model 1	2,130	2,129	100.0
Computer Model 2	1,830	1,830	100.0
Computer Model 3	1,605	1,599	99.6
Computer Model 4	1,180	1,095	92.8
Computer Model 5	1,545	1,395	90.3
Computer Model 6	1,360	1,349	99.2
Computer Model 7	1,680	1,699	101.1
Computer Model 8	1,140	1,140	100.0
Computer Model 9	1,260	1,249	99.1
Computer Model 10	1,595	1,595	100.0

4.6 However, this example immediately draws light to a key issue in constructing price indexes, namely how to determine the average of the price change. One type of measurement might be to determine the average price change by considering the ratio of the arithmetic average of prices.

$$\text{Average Feb 04 price}$$
$$= \frac{2{,}129 + 1{,}830 + 1{,}599 + 1{,}095 + 1{,}395 + 1{,}349 + 1{,}699 + 1{,}140 + 1{,}249 + 1{,}595}{10}$$
$$= 1{,}508.0$$
$$\text{Average Jan 04 price}$$
$$= \frac{2{,}130 + 1{,}830 + 1{,}605 + 1{,}180 + 1{,}545 + 1{,}360 + 1{,}680 + 1{,}140 + 1{,}260 + 1{,}595}{10}$$
$$= 1{,}532.5$$
$$\text{Ratio of average prices (Feb 04/Jan 04)}$$
$$= \frac{1{,}508.0}{1{,}532.5} \times 100$$
$$= 98.4$$

4.7 Another equally viable measure would be to consider the average of the price relatives; that is, the arithmetic average of the changes in price for each of the ten personal computers observed.

$$\text{Average of price relatives}$$
$$= \frac{100.0 + 100.0 + 99.6 + 92.8 + 90.3 + 99.2 + 101.1 + 100.0 + 99.1 + 100.0}{10}$$
$$= \frac{982}{10}$$
$$= 98.2$$

4.8 A third type of measure is the geometric mean. Use of the geometric mean rather than an arithmetic average has several desirable outcomes, particularly when considering averages of price relatives. One property in particular is that when considering a geometric mean, the ratio of the geometric mean of prices and the geometric mean of price relatives are always identical measures.

Geometric mean of Feb 04 prices

$$=(2,129\times1,830\times1,599\times1,095\times1,395\times1,349\times1,699\times1,140\times1,249\times1,595)^{\frac{1}{10}}$$
$$=1,447.4$$

Geometric mean of Jan 04 prices

$$=(2,130\times1,830\times1,605\times1,180\times1,545\times1,360\times1,680\times1,140\times1,260\times1,595)^{\frac{1}{10}}$$
$$=1,505.2$$

Ratio of geometric means of prices (Feb 04/Jan 04)

$$=\frac{1,477.4}{1,505.2}\times100$$
$$=98.1$$

Geometric mean of price relatives

$$=(100.0\times100.0\times99.6\times92.8\times90.3\times99.2\times101.1\times100.0\times99.1\times100.0)^{\frac{1}{10}}$$
$$=98.1$$

4.9 Further details on the choice of index formula may be found in *Consumer Price Index: Concepts, Sources and Methods 2003* (cat. no. 6461.0), the Consumer Price Index Manual (ILO et al, 2004) and the Producer Price Index Manual (ILO et al, 2004).

THE MATCHED MODEL

4.10 Whilst a price index is used as an indicator of all price changes within the economy, in practice it is not possible to observe price changes for every transaction, and the price indexes are calculated by using a sample of products. In this manner price changes observed in the sample are inferred back to all transactions that occur in the population.

4.11 The traditional method of constructing a price index from a sample of goods is to use the matched model approach. Under this approach, the prices of a sample of goods, in this case, personal computers, are observed in some base period, and are then compared with the prices of goods sold in some later period. Only those items that match between the two periods are included in the price comparison.

4.12 This concept is illustrated in Table 3. Data for fifty-six (56) computer models were collected from a vendor website during both May and June 2004. Each computer model is a unique combination of components such as central processing unit, hard drive, random access memory and so forth. The vendor has assigned a unique descriptor to each model (the "manufacturer number" in Table 3). For convenience and ease of reference, a numeric identifier has also been assigned to each computer model.

TABLE 3: PRICE SAMPLE FROM A VENDOR'S WEBSITE ILLUSTRATING THE MATCHED MODEL APPROACH

		Price ($)		
Numeric	*Manufacturer number*	*May 2004*	*June 2004*	
001	A0116C	726		
002	B0170C	726		
003	B0178C	817		
004	B0173C	817		
005	B0126C	817		
006	K0103C	904		
007	G0147C	909		
008	A0124C	932		
009	K0104C	1,068		
010	K0106C	1,132		*Models of computer*
011	J0178C	1,223		*available only*
012	K0191C	1,235		*in May 2004*
013	J0174C	1,359		
014	J0162C	1,359		
015	B0176C	1,363		
016	K0124C	1,386		
017	A0118C	1,450		
018	B0159C	1,541		
019	K0100C	1,541		
020	B0179C	635	681	
021	B0165C	726	726	
022	B0177C	817	863	
023	B0181C	905	870	
024	B0166C	908	908	
025	A0178C	950	950	
026	K0105C	1,068	1,068	
027	J0170C	1,086	1,068	
028	N0118C	1,090	1,090	
029	A0172C	1,132	1,132	
030	N0112C	1,135	1,135	
031	A0174C	1,156	1,155	
032	J0177C	1,212	1,212	*Models of computer*
033	A0179C	1,250	1,250	*available in both May*
034	J0168C	1,268	1,181	*and June 2004*
035	K0193C	1,268	1,268	
036	J0173C	1,295	1,091	
037	J0175C	1,302	1,302	
038	J0161C	1,313	1,136	
039	K0195C	1,359	1,207	
040	K0194C	1,450	1,298	
041	K0113C	1,450	1,331	
042	K0192C	1,495	1,371	
043	N0119C	1,505	1,591	
044	K0190C	1,541	1,423	
045	B0171C	1,632	1,504	
046	K0107C		590	
047	N0117C		770	
048	N0116C		860	
049	B0174C		1,045	
050	N0115C		1,048	*Models of computer*
051	B0183C		1,135	*available only in*
052	N0122C		1,226	*June 2004*
053	N0121C		1,454	
054	B0172C		1,462	
055	B0175C		1,499	
056	P0103C		1,590	

4.13 As can be seen, nineteen (19) models were available for sale in May only, being discontinued in June; eleven (11) models were introduced in June, and twenty six (26) models were available in both May and June 2004.

4.14 This matched model approach to price indexes utilises only those data which are available in both current and base periods. For a price index measuring the change in the prices of personal computers from May to June 2004, a matched model price index would use only those twenty-six (26) models of computer that were on sale in both periods.

4.15 However, adopting such an approach for products that undergo rapid change, such as personal computers, is problematic. In this example, the sample of "matched models" is less than half the total number of models on sale during the two months. It is unlikely that the matched models alone are *representative* of all transactions that occurred in the marketplace and therefore they are unlikely to accurately record the price changes for all computers.

4.16 Incorporating the new models of personal computer into the price index is not straightforward. Newly introduced models have different characteristics, for example, a larger hard drive, faster processor and so forth. When comparing products from earlier periods, we note that the price may have changed but, more importantly, we know that the product itself has changed. In this context we are no longer comparing "like with like", and the pure price change for a personal computer arises because of both a different list price and because of changes in the computer's characteristics. Price statisticians refer to this issue as the need to *price to constant quality*. A price index must measure pure price change, and therefore must make adjustments for changes in characteristics (or changes in quality) of individual goods. These adjustments are called *quality adjustments*. An example of quality change is provided in Table 4.

TABLE 4: EXAMPLE OF QUALITY CHANGE FOR A SPECIFIC MODEL OF PERSONAL COMPUTER

	Configuration on sale	
Component of computer	*March 2004*	*June 2004*
Processor	Celeron 2.4 Ghz	Celeron 2.4 Ghz
Hard drive	40 Gb	80 Gb
RAM	256 Mb	256 Mb
Price	**$899**	**$849**

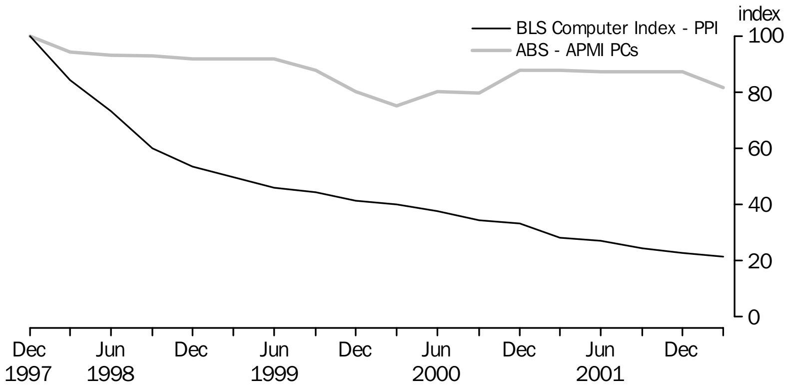
4.17 The list price of the specified model in Table 4 changes from $899 in March 2004 to $849 in June 2004. This is a fall of 5.6%. However, we observe that the configuration being sold in June is different from that being sold in March, with the hard disk drive changing from forty gigabytes (40 Gb) to eighty gigabytes (80 Gb). The larger hard drive is a quality change. We may know from investigations into the market during March that an 80 Gb hard drive adds $50 to the value of the equivalent computer with a 40 Gb hard drive. Thus we may estimate that if the equivalent machine in March had an 80 Gb hard drive rather than 40 Gb, it would have a list price of $899 + $50 = $949. Therefore, a quality adjusted measure of price change would be obtained by comparing the June figure of $849 with this March estimate of $949, or a fall of 10.5%.

4.18 In practice it is difficult, if not impossible, to determine change in prices on a component-by-component basis. In addition to the large number of characteristics that comprise a personal computer, the perceived value of an individual component is often altered by the combination of components. For example, the value of a larger monitor may be different depending upon the type of graphics card available, or the amount of graphics memory, or perhaps a combination of the two, and so forth.

4.19 The quality adjustment problem is applicable to all price indexes, not just those for personal computers. However, for personal computers, technology is driving rapid changes in computer attributes, so that matching computers from one period to the next would require a large number of quality adjustments. Quality adjustments for computers often require subjective estimates of the change in value provided by a new or improved attribute. The application of a large number of subjective estimates has the ability to reduce the accuracy of the index. In particular, failure to account for quality improvements, such as those introduced by larger and faster components, results in a price index that is upwardly biased; that is, it will overstate price increases, with a result that any measure of volumes that uses such a price index as a deflator will be downwardly biased.

4.20 To illustrate the importance of adjusting for quality change, Figure 1 shows the personal computer index for the Price Index of Articles Produced by Manufacturing Industries (APMI) that priced computer models without quality adjustment. Consequently the level of the price index remained static. This is contrasted with the corresponding personal computer index from the United States (US) Bureau of Labor Statistics (BLS) producer price indexes that uses an explicit mechanism for quality adjustment and has shown a substantial decrease from December quarter 1997 to March quarter 2002.

Figure 1 PERSONAL COMPUTER PRODUCER PRICE INDEXES, Base: December 1997 = 100



4.21 In the Australian national accounts the APMI personal computer series was not used as a deflator over this period because of the bias observed here.

THE PROPOSED SOLUTION

4.22 This chapter has highlighted that traditional approaches (such as matched models) to constructing price indexes are not appropriate when dealing with personal computers, since they are neither representative nor do they allow for the change in quality of newly introduced models. When confronted with this problem, international best practice as advocated by international agencies such as the OECD, ILO and the IMF, as well as being adopted by many national statistical organisations, is to develop ***hedonic price indexes***.

CHAPTER 5 THE HEDONIC PRICE INDEX

5.1 Traditional price index methods are well adapted to measuring the price of relatively standardised products, but they encounter problems in terms of data requirements and methods when the characteristics, market shares, and prices of a class of products are changing rapidly. *Hedonic price indexes* provide one means of addressing these empirical and methodological problems. Using the statistical relationship between observed price changes and changes in the characteristics and qualities of the goods, a hedonic price index is developed that measures relative price changes while holding quality and underlying characteristics constant.

5.2 A hedonic price index is any price index that utilises, in some manner, a *hedonic function.* In simple terms a hedonic function identifies the relation between the prices of different varieties of a product, such as differing models of personal computers, and the characteristics within them.

5.3 The principle underlying hedonics is that if purchasers face observable relationships between a product's characteristics and its price, one should be able to use these relationships to disentangle pure price changes from quality changes (Schultze and Mackie 2002). It is important to note that, when applied to personal computers, hedonic techniques identify an empirical relationship between the price and the variation in characteristics from different models of personal computers.

5.4 The assumption that underlies this technique is that a change in a characteristic is of increased value to the consumer and increased cost to the producer[1].

5.5 The hedonic method is particularly well suited for comparing goods that can be thought of as comprising a bundle of underlying attributes, each of which is assumed to have its own intrinsic value. In the case of personal computers, the components inside the "box" itself have several independent, measurable attributes. By comparing prices and features of various computers, a *hedonic regression model* assigns values to each of the particular features that are identified as "price determining" (for example, processor speed, memory, disk capacity etc.).

HEDONIC FUNCTIONS

5.6 A hedonic function is a statistical model that expresses a product's price as a function of its characteristics.

5.7 It is most commonly determined by using regression techniques to estimate the value of specific bundles of individual characteristics that, when packaged together, form products. In application, determining a hedonic function requires regressing prices against observed characteristics that are considered to be price determining. In the simplest case the regression is a "line of best fit", but most hedonic models have more than one explanatory variable. The model resulting from the regression process is the hedonic function.

5.8 The types of statistical models adopted for hedonics frequently express the natural logarithm of the price in terms of a linear combination of its components, or sometimes the logarithms of the components.

---

[1] "The theory of hedonic indexes is built on the proposition that the characteristics are the variables that the buyers of the product want, and that the characteristics of the product also are costly to produce. That is, computer buyers want more speed, other things equal, and faster computers are more costly to produce, with a given production technology." Triplett, J, OECD Hedonics Handbook 2004, p42.

Example of a hedonic function

5.9 Consider that we have a set of data containing information on prices of personal computers, together with information on the amount of RAM in megabytes and the size of the hard drive in gigabytes. For this particular example, it may be found that:

$$\log(price) = \beta_0 + \beta_1 \times RAM + \beta_2 \times Hard\ Drive + \varepsilon$$

5.10 That is, we may find that the log of the price is a function of both RAM and the size of hard drive. For each computer the amount of RAM and the size of the hard drive are known. These data (RAM and hard drive size) are called *explanatory variables*.

5.11 The coefficients (or *parameters*) β_0, β_1, and β_2 which relate the explanatory variables to the price are usually unknown. A statistical modelling process such as linear regression allows us to estimate these unknown parameters β_0, β_1, and β_2. Continuing this example, we may say that for a certain set of data on personal computers, the estimated values of the parameters β_0, β_1, and β_2 are as follows:

TABLE 5: EXAMPLE OF PARAMETERS ESTIMATED USING REGRESSION

Parameter	*Estimate determined by regression*
$\hat{\beta}_0$	*6.614*
$\hat{\beta}_1$	*0.00083562*
$\hat{\beta}_2$	*0.00187*

5.12 Using these parameter estimates a hedonic function is then constructed as follows:

$$\begin{aligned}\log(\hat{price}) &= \hat{\beta}_0 + \hat{\beta}_1 \times RAM + \hat{\beta}_2 \times Hard\ Drive \\ &= 6.614 + 0.00083562 \times RAM + 0.00187 \times Hard\ Drive\end{aligned}$$

5.13 Beginning with known values of RAM and hard drive size, the hedonic function can be used to estimate quality adjusted prices. This process is shown in Table 6. In this example a computer with a 120 gigabyte hard drive and 512 megabytes of RAM has an estimated log price 7.326, and hence an estimated price of $1518.68.

TABLE 6: EXAMPLE OF PARAMETERS ESTIMATED USING REGRESSION

		log(Price) = 6.614 + (0.00083562 x RAM) + (0.00187 x Hard Drive)	
Hard Drive (Gb)	*RAM (Mb)*	*Log (Price)*	*Price = exp (log (Price))*
40	*256*	*6.977*	*$1,071.48*
80	*256*	*7.049*	*$1,151.48*
120	*512*	*7.326*	*$1,518.68*

DETERMINING A HEDONIC FUNCTION IN PRACTICE

5.14 Determining a robust hedonic function is a "data hungry" exercise and requires a regular source of detailed information regarding both prices and characteristics of products. Until very recently, such data sources were not available within Australia. However, with vast amounts of marketing and sales now conducted via the Internet, regular, timely and rich data sources are available which provide sufficient information to undertake the modelling necessary to determine hedonic functions.

5.15 Each month a detailed data capture exercise is undertaken by visiting computer vendor websites. Information is collected on prices of personal computers and on a vast array of descriptive characteristic information, such as processor type, hard drive size, RAM and so forth. These data are then formatted for use in statistical modelling software, with the creation of a large number of explanatory variables. These variables include numeric data for hard drive, RAM, graphics memory etc., and include a variety of (0,1) indicator variables. For example, a variable might be constructed to take the value zero (0) if the model of computer does not include a monitor, and take the value one (1) if the model of computer does include a monitor. The numeric data, both prices and explanatory variables, are transformed to a natural logarithm scale. This type of model is frequently called a log-log or double log model. This combination of variables is then used in a regression process utilising the backward elimination model selection technique. The resulting linear model is the hedonic function.

TYPES OF HEDONIC PRICE INDEX

5.16 Given a hedonic function, there are two main types of hedonic price index.

5.17 The first is called the "direct" method, whereby manipulation of the hedonic function directly results in a *price index*. The result of the regression modelling process is itself an index number, which measures the price change from some base period to the current period.

5.18 The second type of hedonic price index is the "indirect" method. This method uses the hedonic function to estimate *prices*, which in turn are included in a more traditional index number formula. An example of the indirect method was included in Table 5.

5.19 The ABS has researched both direct and indirect hedonic price indexes for personal computers. Findings from this research are included in the Appendix.

CHAPTER 6 CURRENT APPROACH

ABS CURRENT PRACTICE

6.1 Chapter 4 illustrated how traditional methods of constructing price indexes encounter problems when applied to rapidly changing products such as personal computers. International best practice has recommended development of hedonic price indexes for some time but until recently the ABS has not been able to obtain sufficient data to either research or construct hedonic price indexes.

6.2 The ABS is instead using adjusted hedonic price indexes obtained from the US Bureau of Economic Analysis (BEA), and from the US Bureau of Labor Statistics (BLS), as an alternative to traditional methods, and as an alternative to developing an Australian hedonic index. The hedonic indexes from the US are exchange rate adjusted and lagged to increase their suitability for the Australian context.

6.3 The rationale behind this strategy is that it was considered that new models or new features for computers have the same impact on prices across the global market. That is, changes in characteristics of personal computers would have the same impact on prices in the US as they would on prices in Australia. The index is lagged as it was considered that new models of personal computer would enter the US market before they would enter the Australian market. An additional assumption underlying this approach is that trade and transport margins move the same way in different countries.

6.4 This rationale is also the basis for the formation of the Jack Triplett[2] "Hedonics group", which for several years has worked to find universal hedonic functions for given products that would be applicable in all countries.

6.5 Currently the publications containing economic accounts that use exchange rate adjusted computer price indexes are:

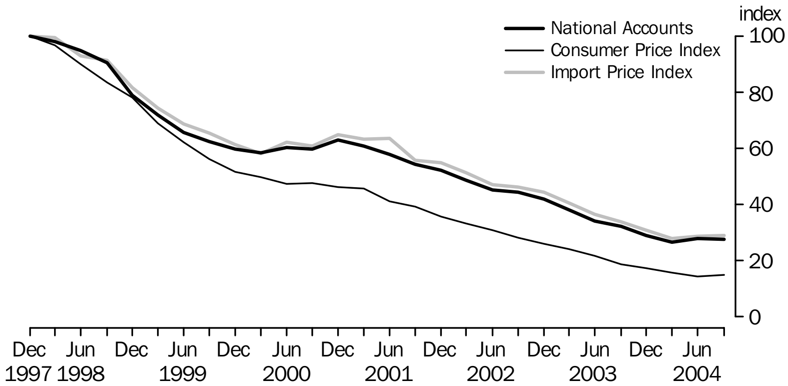
- *Australian System of National Accounts* (cat. no. 5204.0)
- *Australian National Accounts: National Income, Expenditure and Product* (cat. no. 5206.0)
- *Consumer Price Index, Australia* (cat. no. 6401.0)
- *International Trade Price Indexes, Australia* (cat. no. 6457.0)
- *Producer Price Indexes, Australia* (cat. no. 6427.0)

6.6 Figure 2 shows that the adjusted US indexes used by the ABS have not moved in the same manner over the period shown. This difference is due to the ways in which the lagging and exchange rate adjustments are applied in the different indexes.

---

[2] Jack Triplett is a visiting fellow of the Brookings Institution in Washington, DC. He was formerly Chief Economist at the US BEA and has written extensively on problems of economic measurement, particularly on prices and productivity.

Figure 2 COMPARISON OF EXCHANGE RATE ADJUSTED US SERIES, Base: December 1997 = 100



UNIVERSAL HEDONIC FUNCTIONS

6.7 The use of US hedonic price indexes implies that pure price movements observed in the US marketplace would be identically observed in the Australian marketplace, after allowing for fluctuations in the exchange rate. Examining this process in more detail, the implication of this assumption is that not only will list prices of personal computers move at the same rate as in the US, but that the change in characteristics of personal computers would result in exactly the same quality change in the Australian market as in the US market; that is, changes in characteristics have the same impact on pure price change regardless of the market.

6.8 In practice, this assumption is problematic. It is readily observable that list prices of components of personal computers, in particular RAM and processors, vary in different markets. The impact of changes in characteristics on pure price movements differs from market to market, depending not only on the perceived benefits of changes in the products but also on the different factors of supply and demand that apply in local markets. Work done by Jack Triplett's group, amongst others, has suggested that there is no universal hedonic function for personal computers. Examples of this sort of evidence include widely differing hedonic functions for neighbouring countries within Europe. In other words, recent evidence suggests that changes in a given set of characteristics have different impacts on prices in different market places.

6.9 These observations suggest that the conceptual basis for using US price indexes in Australia’s economic statistics is not as strong as had been previously assumed.

USE OF EXCHANGE RATES

6.10 Applying exchange rate movements to US hedonic price indexes assumes that the impacts of any fluctuations in currency are passed on to the Australian purchaser as pure price movements.

6.11 This assumption is also problematic. First, it does not allow for substitution of supply; for example, it assumes that, regardless of movement of the Australian dollar against the US dollar, all purchases of computers will still be undertaken in US dollars, rather than from some other market or in some other currency. Second, the process assumes that the full impact of exchange rate movements will be passed on to purchasers; for example householders in the case for indexes used in the CPI and those used in the national accounts for estimates of household final consumption expenditure. This assumption is flawed, since it neglects the impact of changing margins for distributive trade (that is, wholesale and retail margins) and further neglects the use of financial instruments such as hedging by importers to protect against exchange rate movements.

6.12 This weakness is likely to be exacerbated when there are large fluctuations in currency rates.

BEHAVIOUR OF EXCHANGE RATE ADJUSTED US HEDONIC PRICE INDEXES

6.13 Examination of the US price indexes for personal computers, sourced from the US BEA shows that personal computers have experienced real price falls. This accords with intuition and observation, in that list prices of personal computers have remained static (or even fallen) whilst new or improved features are introduced with every new model.

6.14 When the Australian dollar began to depreciate markedly against the US dollar in the late 1990s, the exchange rate adjusted indexes showed far smaller price falls than the corresponding BEA indexes and, on occasion, actual price rises. These movements are counter-intuitive, in that they accord with neither observation of price movements nor improvements in product characteristics visible in the Australian market.

6.15 Examples of these movements are presented in Figure 3, which shows the original series sourced from the BEA, and the series after applying an exchange rate adjustment. This adjusted series has been used in the Australian national accounts. By way of comparison, Figure 4 details the $US/$A exchange rate for the same period.

Figure 3 COMPARISON OF ADJUSTED AND UNADJUSTED US BEA SERIES, Base: December 1997 = 100

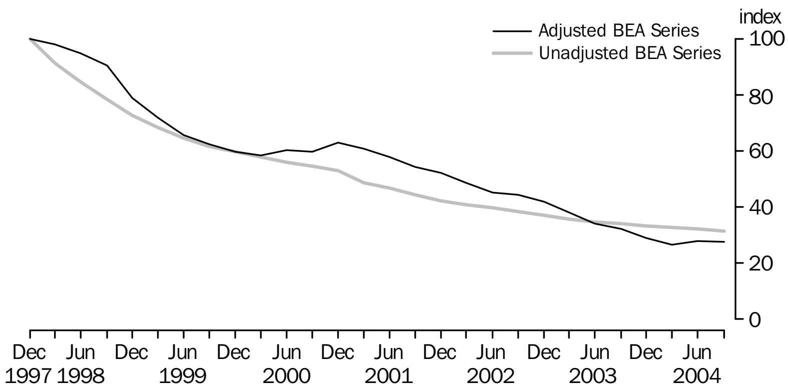
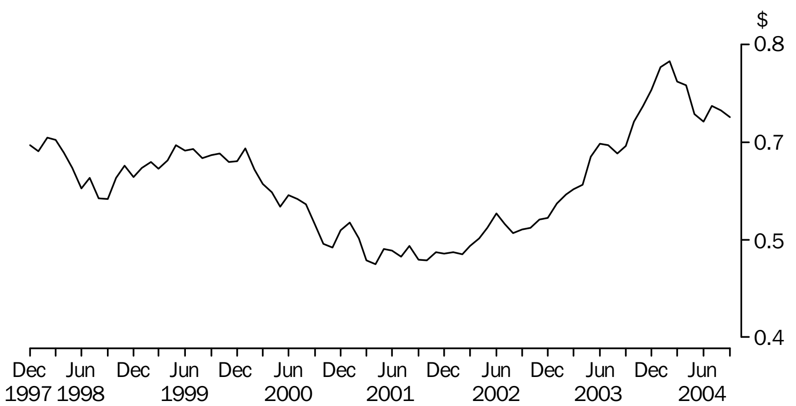


Figure 4 US$/AUSTRALIA$ EXCHANGE RATE, Source: RBA



6.16 In the period from June 2000 to March 2001, the Australian dollar fell relative to the US dollar, and consequently the adjusted price index increased, whereas the unadjusted BEA index fell significantly. By contrast, for the period March 2001 to December 2003 the Australian dollar has appreciated relative to the US dollar resulting in the adjusted index falling at a greater rate than the BEA index.

6.17 The conclusion from these observations is that the exchange rate adjusted hedonic price indexes have been more volatile than their US counterparts. The amount and direction of movement of the adjusted indexes does not accord with expectations, given the large falls in the index in $US terms, and the assumptions underlying their use suggest that alternative measures should be investigated.

CHAPTER 7 THE PROPOSED HEDONIC METHOD FOR PRICE INDEXES FOR PERSONAL COMPUTERS

PROPERTIES OF A HEDONIC PRICE INDEX

7.1 Chapter 2 of this information paper outlines the properties required from a price index for personal computers. Briefly restating, these requirements are:

1. The price index must measure *pure price change*;
2. The price index must be *representative*;
3. Changes in the price index should be *easily explainable* to users; and
4. The price index must be more robust than the current one with the implication being it is based on *an improvement over existing methods*.

PROPOSED METHODOLOGY

7.2 In May 2003, Statistics Netherlands presented a new hedonic technique to the Inter-Secretariat Working Group on Price Statistics (IWGPS). This technique is known as 'consecutive two period chained time dummy double imputation hedonic price indexes'. This method satisfies all the requirements outlined above. Furthermore, the double imputation price index addresses concerns observed in investigating both direct and indirect hedonic price indexes.

7.3 Given its performance against the requirements outlined above, and its demonstrated ability to mitigate the concerns that arise when considering other methodologies, the ABS proposes to adopt the double imputation method for constructing price indexes for personal computers.

7.4 Details of the formulation of the double imputation index, as well as ABS concerns regarding the direct and indirect methodologies, are provided in the Appendix.

KEY CHARACTERISTICS OF THE DOUBLE IMPUTATION METHOD

7.5 The double imputation method can best be thought of as a traditional matched model index with an explicit adjustment applied because of both the departure of superseded models and the introduction of new models (Chapter 4.11 describes the matched model index). A key deficiency of the basic matched model approach is that it makes no provision for systematically including the effects of price and quality changes in models available in the marketplace, and determines price change by only considering those models which appear in the market in both periods of interest. In other words, any improvement in quality associated with the introduction of a new model will not be measured if only matched models are priced.

7.6 The double imputation price index counters this deficiency by implicitly imputing price movements for both superseded models and newly introduced models. This is where the term "double imputation" arises. The index is then considered *representative* of all transactions, since recently superseded models and new models are included in the determination of price change, in addition to products common to both periods.

7.7 Further, the implicit imputation process at the core of this technique uses a hedonic function to adjust for changes in the characteristics of both the new and superseded models; that is, the prices imputed are adjusted for quality change, and hence the resulting index measures *pure price change.*

7.8 The proposed process utilises price data from Australian vendors of personal computers, and so is not only representative of the Australian marketplace but also avoids issues with both exchange rate fluctuations and arbitrarily lagging prices to take account of shipping times etc. The double imputation index uses a hedonic function based on characteristics of personal computers sold in the Australian marketplace, using prices in Australian dollars, and so furthermore does not rely on the restrictive assumptions underlying a universal hedonic function. As a result the double imputation index is a substantial *improvement over historic methods*.

7.9 Finally, any movements in the double imputation index can be decomposed into the movement due to changes in prices of the matched sample, and the movement due to changes in products in the marketplace. Movements in the index can be explained in terms of changes in list prices of existing products and changes in quality of new products, and so the resulting measures are *easily explainable* to users.

IMPLEMENTATION INTO PPI

7.10 In September quarter 2003 the double imputation index was introduced into PPI in the Price Index of Articles Produced by Manufacturing Industries (APMI), and as a consequence into the Stage of Production Producer Price Indexes (SOP). Further work is required to develop a double imputation index for the Import Price Index (IPI). Until this work is completed, the ABS will continue to use the exchange rate adjusted “Computers and peripheral equipment” from the US BEA as the basis for the personal computer component of the IPI.

PROPOSED IMPLEMENTATION INTO CPI

7.11 To measure price change for personal computers as purchased by households, the CPI utilises a hedonic price index from the US Bureau of Labor Statistics (BLS), "Personal Computers and Peripheral Equipment", which is both lagged and adjusted to account for changes in the $US/$A exchange rate. This process was described in Chapter 6, noting that the assumptions underlying this approach now need reconsidering.

7.12 A comparison of the CPI series for personal computers shows that the proposed double imputation index more plausibly reflects the impact of exchange rate fluctuations. As seen in Figure 4, for the period from June 2000 to March 2001 the Australian dollar fell relative to the US dollar. We observe in Figure 5 that the proposed double imputation price index fell at a greater rate over this period than the exchange rate adjusted BLS price index (as used in the Australian CPI).

7.13 By contrast, for the period March 2002 to December 2003 the Australian dollar appreciated relative to the US dollar. Conversely, the exchange rate adjusted BLS price index used in the CPI fell at a greater rate over this period than the proposed double imputation index.

Figure 5 CPI SERIES FOR PERSONAL COMPUTERS, Base: June 2000 = 100

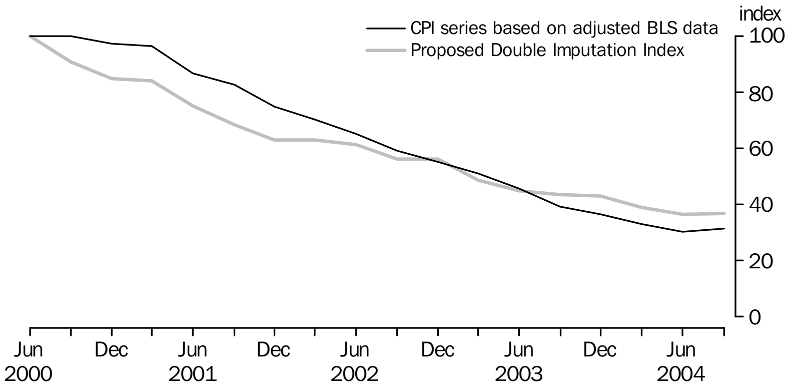
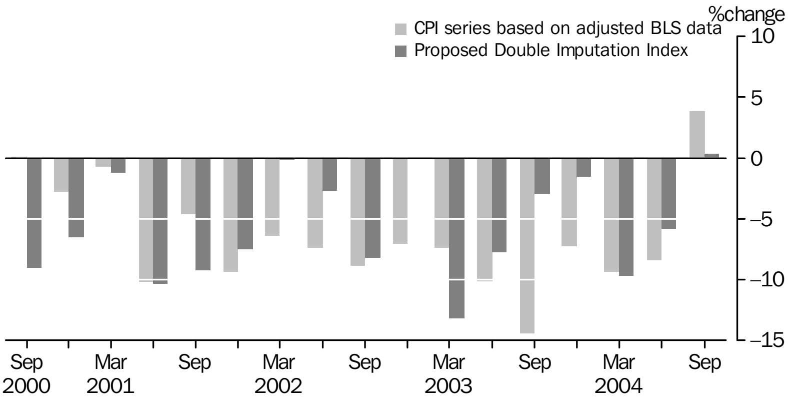


Figure 6 CPI SERIES FOR PERSONAL COMPUTERS, Quarterly % change



7.14 The ABS proposes to adopt the double imputation method for the CPI from September quarter 2005. In line with ABS policy, no revisions will be made to the CPI for earlier periods.

7.15 The introduction of the double imputation index into the CPI will provide a computer index of higher quality, removing the direct effect of the $US/$A exchange rate from the index and providing an index that better reflects the Australian market.

PROPOSED IMPLEMENTATION OF THE DOUBLE IMPUTATION INDEX INTO AUSTRALIAN NATIONAL ACCOUNTS

7.16 In calculating the volume estimates of household final consumption expenditure in the Australian national accounts, the ABS utilises the "computers and peripheral equipment" index from the US BEA as a price deflator. The index is lagged and exchange rate adjusted. This process was discussed in Chapter 6, again noting that the assumptions underlying this approach now need revisiting.

7.17 Figure 4 shows that the Australian dollar fell relative to the US dollar over the period from June 2000 to March 2001. Figure 7 shows that during this period the price index used in the national accounts increased. This increase is counter intuitive to the expected downward movements of the pure prices of personal computers, given that list prices have shown little change over this period, whilst the capabilities of personal computers have improved.

Figure 7 NATIONAL ACCOUNTS SERIES FOR PERSONAL COMPUTERS, Base: June 2000 = 100

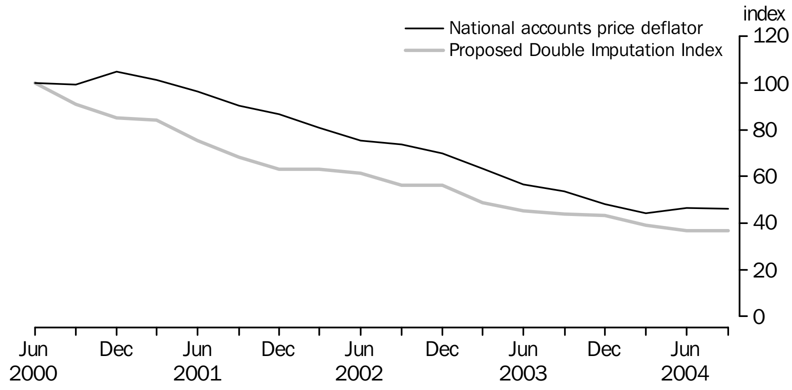
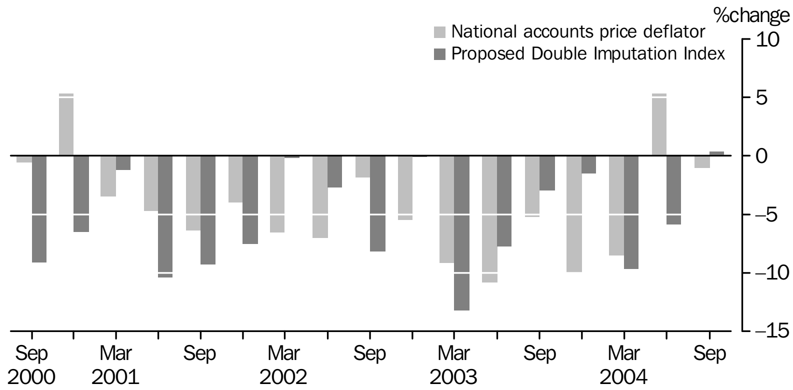


Figure 8 NATIONAL ACCOUNTS SERIES FOR PERSONAL COMPUTERS, Quarterly % change



7.18 By way of contrast, the proposed double imputation index shows decreases over this same period, again more plausibly reflecting price movements of personal computers.

7.19 The ABS proposes to adopt the double imputation method for the Australian national accounts from September quarter 2005.

7.20 The introduction of the double imputation index into the national accounts will provide a higher quality deflator for computers. It will remove the direct effect of the $US/$A exchange rate from the index and will provide an index that better reflects the Australian market.

7.21 Note that, unlike the proposed use for the CPI and PPI, the introduction of the double imputation index into the national accounts will allow an opportunity to revise data to better reflect this improved measurement. The most recent data show that the current level of the double imputation index and the existing price deflator for personal computers used in the national accounts are quite close. The GDP volume estimates are unlikely to be revised significantly, no matter how much the personal computer price index is changed, because the impact on final demand is almost completely offset by imports. However, there will be some revisions to the volume estimates for household final consumption expenditure, gross fixed capital formation and imports as a result of the change.

APPENDIX

OVERVIEW OF THIS APPENDIX

A.1 This appendix contains technical material regarding the research and development of the ABS hedonic price index of personal computers. It is intended for those readers interested in the statistical modelling approach undertaken by the ABS, as well as those readers who are interested in the process used to determine an appropriate methodology for hedonic price indexes within the ABS.

A.2 This appendix is divided into different sections.

A.3 The first section (A.7 – A.38) provides an outline of the symbols and notation used in the remainder of the appendix. Also included in this section is an example of the regression modelling process which is fundamental to hedonic price indexes.

A.4 The second section (A.39 – A.71) deals with investigations into "indirect" hedonic price indexes. As described earlier, indirect methods are those methods which impute prices, which are in turn used in the price index formulation traditionally used by the ABS. This section also highlights ABS concerns with indirect methods.

A.5 The third section (A.72 – A.105) deals with "direct" hedonic price indexes. These indexes use the hedonic function to directly estimate a measure of price change from the base period to the current period. This section also highlights ABS concerns with direct methods.

A.6 The fourth and final section of the Appendix (A.106 – A.131) considers the double imputation price index, and shows how this approach mitigates the concerns raised with both the direct and indirect approaches.

TOOLS AND NOTATION FOR DESCRIBING HEDONIC PRICE INDEXES

A.7 A detailed description of the research that the ABS undertook in developing an appropriate method for hedonic price indexes for personal computers requires discussion of the mathematical concepts underlying the various alternatives. As such it is necessary to outline a common set of nomenclature; that is, to describe in a common way, the terms and concepts encountered and to define symbols and notations. This section of the appendix describes the different populations observed at different points of time, and introduces the terms and formulae we will use as short hand in describing them.

Time

A.8 The first concept that we must describe is the period. In earlier examples we have used specific references such as May to June or January to August. It is necessary in generalising to all possible periods to use a notation that is free from individual months (or quarters or years etc.). When describing time, for any given period we use the letter t. Most discussions regarding price indexes are comparing the current period with some base period, and as a consequence convention compares the current period t with an earlier base period 0.

A sample - prices and characteristics of products

A.9 In the field of price statistics, prices are generally denoted by the letter P or the letter p. When considering any price observation it is important to associate this observation with the period in which it was observed. This is done through use of a superscript t.

A.10 One of the benefits of using a price index is that it summarises the movement of the prices of many products. We use a subscript i to denote an individual product within an index. Furthermore, prices are not observed in isolation - they are observed along with the many defining characteristics of a product. In the case of personal computers these characteristics are memory (RAM), hard drive size, processor speed, type of processor, graphics card and so forth. We use the letter z to indicate a particular characteristic of a product. When there are multiple characteristics (as is the case with personal computers) we use a second subscript to indicate the k characteristics. We then refer to this set of characteristics for an individual observation. For example, when considering a personal computer, each different model of computer will have a particular combination of memory (RAM), hard drive size, processor speed, type of processor, graphics card and so forth.

A.11 Thus z_{ji} is the jth characteristic for the ith computer.

A.12 In the general case for some period t, we will consider a sample of models of personal computers. From each element of the sample we observe a set of:

- period t prices p_i^t ; and
- the k characteristics of the ith personal computer $\underset{\sim}{z}_i^t = \left(z_{1i}, \ldots, z_{ki}\right)$

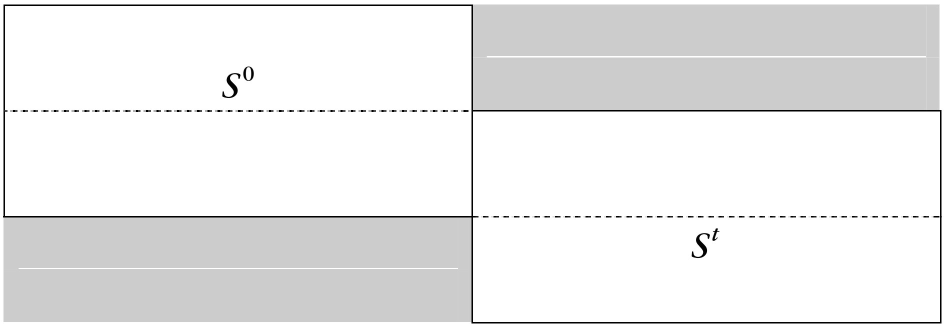
A.13 When we make reference to the entire set of observations (that is, prices of personal computers together with their respective characteristics), we denote it by S:

$$S^t = \left\{p_i^t, \underset{\sim}{z}_i^t = \left(z_{1i}, \ldots, z_{ki}\right), i = 1, \ldots, N\right\}$$

A.14 For each such sample we denote the associated sample size (the total number of sampled personal computers) as N_{S^t}

Comparing samples over time

A.15 Price indexes provide a means of answering the question as to how prices change over time. The price index does this by comparing prices from two different periods. A sample of products from each period is taken. Some products will be available in the earlier period, but not the latter. Other products will be available in the latter period, but not the earlier. There will also be a set of products which are available in both periods. We like to represent this situation in a diagram as follows (0 being the earlier period and t being the later period):



A.16 When considering the samples from the two periods, we note that we can break the samples into four distinct sets. There are two sets of data arising from the observations which are common to both periods, a set of observations with earlier period prices and a set of observations with later period prices.

A.17 We refer to these as overlap sets, or matched sets, which we denote by *M*.

A.18 These two matched sets have the same number of observations, which we denote by $N_{S^0 \cap S^t}$

$$M^0 = \left\{ p_i^0, \underset{\sim}{z}_i' = \left(z_{1i}, \ldots, z_{ki} \right), i \in S^0 \cap S^t \right\}$$

$$M^t = \left\{ p_i^t, \underset{\sim}{z}_i' = \left(z_{1i}, \ldots, z_{ki} \right), i \in S^0 \cap S^t \right\}$$

A.19 The third set of interest will be those models of personal computers from period 0 that do not exist in period *t*. These are product deletions, or superseded models. These are also called deaths, which we denote by *D*. Since they only exist in time 0, there are only price observations from time 0.

A.20 We refer to the set of deaths with prices from time 0 as:

$$D^0 = \left\{ p_i^0, \underset{\sim}{z}_i' = \left(z_{1i}, \ldots, z_{ki} \right), i \in S^0 \not\subset S^t \right\}$$

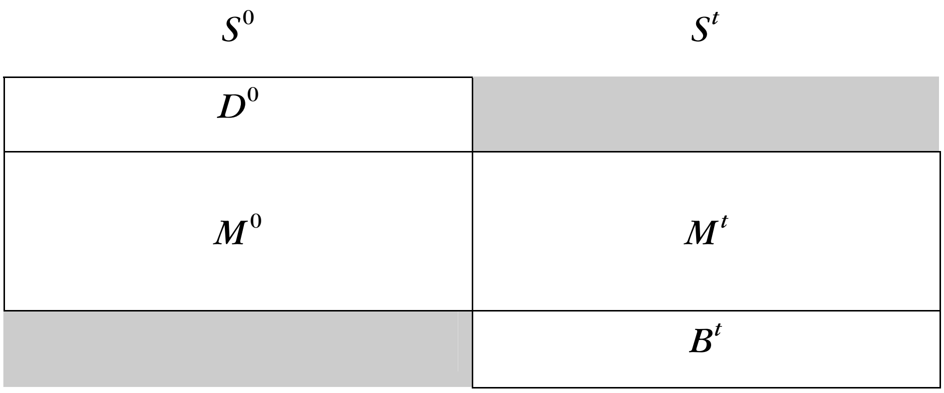
A.21 We denote the associated sample size as $N_{S^0 \not\subset S^t}$

A.22 The fourth set of interest will be those models of personal computers from period *t* that did not exist in period 0. These are product introductions, or new lines or new models. These are also called births, which we denote by *B*. Since they only exist in time *t*, there are only price observations from time *t*.

$$B^t = \left\{ p_i^t, \underset{\sim}{z}_i' = \left(z_{1i}, \ldots, z_{ki} \right), i \in S^t \not\subset S^0 \right\}$$

A.23 We denote the associated sample size as $N_{S^t \not\subset S^0}$

A.24 The relationships between these four sets can be seen from the following diagram:



A.25 Referring back to the samples from each period, we see that the first period sample is the union of the deaths with the matched sample observed with prices from time 0. Similarly, the second period sample is the union of the births with the matched sample observed with prices from time *t*. In set notation we write this as:

$$S^0 = D^0 \cup M^0$$
$$S^t = B^t \cup M^t$$

We can see a similar relation between the different sample sizes.

$$N_{S^0} = N_{S^0 \cap S^t} + N_{S^0 \not\subset S^t}$$
$$N_{S^t} = N_{S^0 \cap S^t} + N_{S^t \not\subset S^0}$$

A.26 The set notation introduced here provides us with tools which help describe the concepts underlying the construction of hedonic price indexes. The other tool required to describe hedonic price indexes is regression modelling, which is detailed below.

Regression modelling

A.27 Regression is a statistical tool for using information observed for one set of phenomena to explain, or to predict, the behaviour in another phenomenon. For example, the amount of RAM and size of hard drive can be used to explain the price of a personal computer through the use of regression. When using regression, the data observed for the sets of phenomena are called variables. If the amount of RAM and size of hard drive are used to explain price, we say that RAM and hard drive size are explanatory variables, or independent variables, and the price is the dependent variable. Regression is used to explain the dependent variable in terms of the explanatory variables.

A.28 Explanatory variables can be quantitative, and take on numeric values; for example, the amount of RAM in megabytes or the number of gigabytes for a hard drive. Explanatory variables can also be indicator variables, and take the value of zero (0) or one (1) signifying the occurrence or not of some condition. An example of an indicator variable would be "Does the computer have a DVD burner (No=0, Yes=1)".

A.29 Regression is used in a variety of ways. It is frequently used as a summary measure of large amounts of data, and is also used to predict values of the dependent variable for known values of the independent variables. For example, if RAM and hard drive size are observed then it is possible through the use of regression to reasonably predict the price of the personal computer.

The linear model

A.30 The basis for regression is that the dependent variable is related to the explanatory variables by some underlying mathematical model. This is called a linear model[1]. For example, we may believe that the logarithm of the price of a personal computer (log(price)) is related to the amount of RAM and hard drive size. This model would be expressed as:

$$\log(price) = \beta_0 + \beta_1 \times RAM + \beta_2 \times Hard\ Drive + \varepsilon$$

A.31 In this formula the term ε represents the error or the variation of any individual observation from the underlying model; using this example, two personal computers with identical RAM and hard drive may be sold at slightly different prices, and they may both vary from that value explained by the model. The error term explains this variation.

A.32 Most importantly the formula above explains the relationship between RAM and hard drive, and the log of the price of the personal computer, in terms of unknown parameters. Whilst RAM and hard drive are observable for any given personal computer, the coefficients β_0, β_1, and β_2 are all unknown. In order to utilise the relationship, we need to estimate the value of β_0, β_1, and β_2.

A.33 If we have a set of data describing personal computers, with values for RAM, hard drive size and price (and hence log(price)), it is possible to estimate the unknown parameters through linear regression. Once these parameter estimates are determined, they can be used to predict log(price).

Estimating values for the unknown parameters using regression

A.34 Taking the example further, linear regression uses sample data to estimate values for β_0, β_1, and β_2. The estimated values are denoted with a caret as $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$.

A.35 Using these values it is possible to predict a value of the dependent variable (in this case log(price)) for any combination of the explanatory variables (RAM and hard drive).

$$\log(\hat{price}) = \hat{\beta}_0 + \hat{\beta}_1 \times RAM + \hat{\beta}_2 \times Hard\ Drive$$

A.36 The parameter estimates are chosen such that when considering the observations in the sample data, the variation between log(price) and the predicted value of log(price) is minimised overall. This technique may be applied using standard statistical software.

Which explanatory variables best explain the dependent variable?

A.37 Part of the regression process involves determining which of all possible explanatory variables best explains the dependent variable. In terms of the regression for prices of personal computers, this involves determining which characteristics of personal computers explain the price of the personal computer. The example to date has used RAM and hard drive to explain variation in computer prices. In practice, many more variables are included in the regression analysis, at least initially. An iterative regression process is then undertaken with the end result that a smaller subset of all possible variables is chosen which best explains the variations in personal computer prices. This model resulting from this variable selection process is called the *parsimonious form*. For hedonic price indexes this form equates to the hedonic function.

[1] The word *linear* in the term *linear model* refers to the linear combination of explanatory variables such as RAM and hard drive, rather than the functions on the left hand side of the expression. To emphasise this we could describe the example in paragraph A.30 as "linear in log(price)".

Regression as a tool

A.38 All hedonic price indexes involve some form of regression, yet regression is simply a tool for summarising large quantities of information. For hedonic price indexes, the information summarised is "how the characteristics of products affect the price of the product". This information is then utilised in estimating prices for goods whose characteristics are changing over time, and when changes of characteristics are accounted for, the prices statistician is then able to *price to constant quality*.

INDIRECT HEDONIC PRICE INDEXES

A.39 An indirect hedonic price index uses the hedonic function to predict *prices*, which in turn are included in more traditional index number formulae. The ABS investigated this method, and undertook research into hedonic imputation. Hedonic imputation is one particular application of an indirect hedonic price index.

A.40 The hedonic imputation approach uses hedonic models to predict prices when product lines discontinued in one period and product lines introduced in a successive period cannot be matched, that is, when the new models of personal computers cannot be compared with the old models. This technique can be separately applied to prices for deaths (or deleted items) and births (or replacement products).

A.41 For example, consider data from two months, June and July. Ideally, we would like to compare all prices in June with all prices in July. However, some personal computers in June have been discontinued, and do not appear in July. The hedonic imputation technique uses those machines that match between June and July to produce a model for July prices. Then the model is used to impute a "July price" for the discontinued lines.

A.42 Continuing the example, we see some personal computers in July have been newly introduced, and did not appear in June. The hedonic imputation technique uses those machines that match between June and July to produce a statistical model for June prices. Then the statistical model is used to impute a "June price" for the newly introduced product lines.

Regression modelling

A.43 The regression models trialled by the ABS were the so-called "double log" models. If we consider quantitative explanators $\left(z_{1i},\ldots,z_{mi}\right)$ and indicator variables $\left(z_{m+1,i},\ldots,z_{ki}\right)$ the models took the form of:

$$\ln\left(p_i^t\right)=\beta_0+\beta_1\ln\left(z_{1i}\right)+\ldots+\beta_m\ln\left(z_{mi}\right)+\beta_{m+1}z_{m+1,i}+\ldots+\beta_k z_{ki}+\varepsilon_i$$

A.44 The methodology researched by the ABS saw a different model determined each period, separately for births and deaths. To clarify, in the case of the hedonic imputation approach as applied by the ABS, the form of the regression is re-determined each period.

A.45 This modelling used the datasets described earlier as

$M^0=\left\{p_i^0,z_i'=\left(z_{1i},\ldots,z_{ki}\right),i\in S^0\cap S^t\right\}$, with sample size $N_{S^0\cap S^t}$

$M^t=\left\{p_i^t,z_i'=\left(z_{1i},\ldots,z_{ki}\right),i\in S^0\cap S^t\right\}$, with sample size $N_{S^0\cap S^t}$

$D^0=\left\{p_i^0,z_i'=\left(z_{1i},\ldots,z_{ki}\right),i\in S^0\not\subset S^t\right\}$, with sample size $N_{S^0\not\subset S^t}$

$B^t=\left\{p_i^t,z_i'=\left(z_{1i},\ldots,z_{ki}\right),i\in S^t\not\subset S^0\right\}$, with sample size $N_{S^t\not\subset S^0}$

Deaths - imputing prices in a later period

A.46 For each of the elements of the death population (elements that last existed at time 0), it is necessary to impute prices for time t.

A.47 A regression is run on the dataset M^t to determine a parsimonious form, and to estimate the parameters $\beta_0,\ldots,\beta_k$.

A.48 To clarify, to predict a period t price for an observation last observed at time 0, a regression modelling exercise is undertaken using the matched data at time t, and using the period t prices.

A.49 These parameter estimates so determined are denoted by $\hat{\beta}_0^t,\ldots,\hat{\beta}_k^t$.

A.50 Predictions of period t prices for those observations that last existed at time 0 are constructed as follows:

$$\hat{p}_i^t = \exp\left(\hat{\beta}_0^t + \hat{\beta}_1^t \ln\left(z_{1i}\right) + \ldots + \hat{\beta}_m^t \ln\left(z_{mi}\right) + \hat{\beta}_{m+1}^t z_{m+1,i} + \ldots + \hat{\beta}_k^t z_{ki}\right),\ \ \forall \boldsymbol{i} \in \boldsymbol{D}^0$$

Births - imputing prices in an earlier period

A.51 For each of the elements of the birth sample (new elements at time t), it is necessary to impute prices for time 0.

A.52 A regression is run on the dataset M^0 to determine a parsimonious form, and to estimate the parameters $\beta_0,\ldots,\beta_k$.

A.53 To clarify, to predict a period 0 price for an observation first observed at time t, a regression modelling exercise is undertaken using the matched data at time 0, and using the time 0 prices.

A.54 These parameter estimates determined are denoted by $\hat{\beta}_0^0,\ldots,\hat{\beta}_k^0$.

A.55 Predictions of period 0 prices for those observations first observed at time t are constructed as follows:

$$\hat{p}_i^0 = \exp\left(\hat{\beta}_0^0 + \hat{\beta}_1^0 \ln\left(z_{1i}\right) + \ldots + \hat{\beta}_m^0 \ln\left(z_{mi}\right) + \hat{\beta}_{m+1}^0 z_{m+1,i} + \ldots + \hat{\beta}_k^0 z_{ki}\right),\ \ \forall \boldsymbol{i} \in \boldsymbol{B}^t$$

Constructing a price index using imputed data

A.56 In utilising the imputed data to construct a price index, we begin by constructing price ratios as follows:

A.57 For elements in the common or matched sample, we consider

$$r_i = \frac{p_i^t}{p_i^0},\ \forall \boldsymbol{i} \in S^0 \cap S^t$$

For all elements of the birth sample, we consider

$$r_i = \frac{p_i^t}{\hat{p}_i^0},\ \forall \boldsymbol{i} \in S^t \not\subset S^0$$

For all elements of the death sample, we consider

$$r_i = \frac{\hat{p}_i^t}{p_i^0},\ \forall \boldsymbol{i} \in S^0 \not\subset S^t$$

A.58 We then consider the geometric mean of the price ratios. The hedonic imputation price index is the geometric mean of the price relatives for all observations from the joint sample is:

$$I_{HI}^{t} = \left(\prod_{\forall i} r_i^{\frac{1}{N}} \right) \times 100$$

$$= \left(\left\{ \prod_{i \in S^0 \cap S^t} \left(\frac{p_i^t}{p_i^0} \right)^{\frac{1}{N}} \right\} \times \left\{ \prod_{i \in S^t \not\subset S^0} \left(\frac{p_i^t}{\hat{p}_i^0} \right)^{\frac{1}{N}} \right\} \times \left\{ \prod_{i \in S^0 \not\subset S^t} \left(\frac{\hat{p}_i^t}{p_i^0} \right)^{\frac{1}{N}} \right\} \right) \times 100$$

where N is the sum of the numbers of observations in the common matched sample, the birth sample and the death sample.

A.59 Note that this two period index can be decomposed into a matched model component, and a component that arises because of the need to impute for the new and deleted elements, that is, it can be decomposed into the product of a matched model index and an imputed index.

Indirect Approach - Advantages and Disadvantages

A.60 The advantage of this approach is that the traditional approach of comparing "like with like" is used for those models available in both months and the hedonic function is used only in the case of new or old goods. As a result the imputed prices can be used as part of the regular cycle of price index construction. Furthermore, it is possible to decompose any index movement into a component due to the quality changes observed in the new and/or discontinued products and a component due to price changes in the matched products.

A.61 There are however two disadvantages to this approach:

1. Out of sample predictions (extrapolation); and

2. Poor predictive power in some circumstances

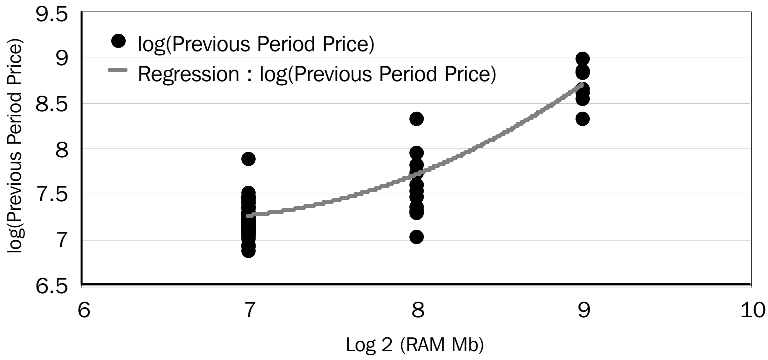
Out of Sample Prediction

A.62 An out of sample prediction occurs when explanatory variables observed in the sample births or sample deaths domains are outside the range of those observed in the matched domain.

A.63 Predictive imputation, such as that being applied, is always sensitive to extrapolation. Predictions occurring on a log scale being brought back to a linear scale exacerbate the problem. These problems are best illustrated with an example.

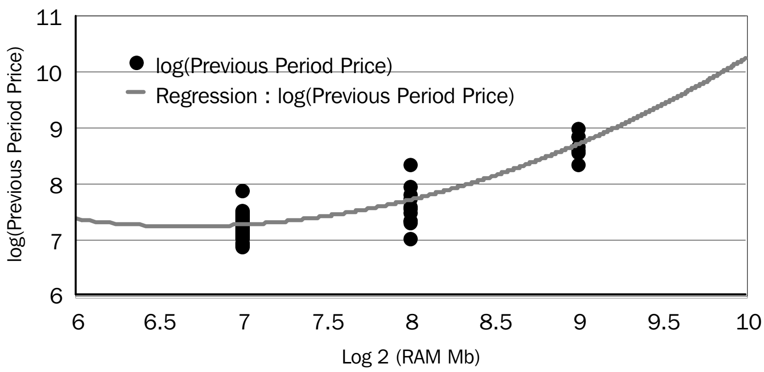
A.64 An example for the month of September 2002 is considered - a regression model that uses RAM only as the explanator to predict price. Note that in practice the models used are more complex, and the choice of one variable here is for illustrative purposes. In this example the single variable plus intercept model fits reasonably well (the percentage of variability in the data explained by the model, the R-squared, takes a value of 0.799 for estimating September 2002 prices, and an R-squared of 0.844 for estimating August 2002 prices). The spread of data and the results of fitting a regression curve, quadratic in $\log_2$(RAM), are illustrated in Figure 9.

Figure 9 REGRESSION MODEL, IMPUTING FOR BIRTHS, SEPTEMBER 2002
Log scale both axes



A.65 The practical application of this model sees the predicted coefficients applied. Now consider the application of this model to impute the previous period price for units that are new models in September 2002 (that is, imputing an August 2002 price for a September 2002 birth). The modelling process applied by the ABS uses only those personal computers that appear in both August 2002 and September 2002. New models introduced in September 2002, which cannot be used to predict prices for August 2002, may have quite different characteristics from those personal computers common to both periods. Some of these models have exceptionally large or small RAM sizes. The curve used to predict these values is illustrated in Figure 10.

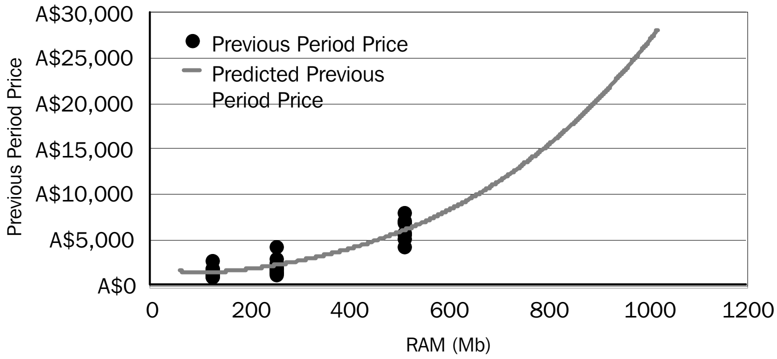
Figure 10 EXTRAPOLATION, IMPUTING FOR BIRTHS, SEPTEMBER 2002
Log scale both axes



A.66 As can be seen, RAM sizes of 64Mb (with a log in base 2 of 6) and 1 Gb = 1024 Mb (with a log in base 2 of 10) are well outside the range of values observed in the matched sample at September 2002. This is particularly a concern for the 1 Gb machines, whose extrapolated value is substantially higher than the observed prices.

A.67 This problem is exacerbated when converting from the log scale back to a linear scale. This is illustrated in Figure 11.

Figure 11 EXTRAPOLATION, IMPUTING FOR BIRTHS, SEPTEMBER 2002
Linear scale both axes



A.68 The model predicts a price for any new unit with 1 Gb of RAM to be over $28,000, when the maximum price observed for any machine in the dataset with 1 Gb of RAM was $7,800. This error is due to the model extrapolating a long way outside the range of the observed data. This is clearly a deficiency with this approach when the characteristics of the product change very rapidly, which was the case with the size of hard drives installed in PCs in the early 2000s.

Poor predictive power

A.69 Empirical studies of the trial of hedonic imputation as used by the ABS have shown that the prediction (imputation) process is sensitive to the level of matching that occurs from period to period. More clearly, the proportion of records from the joint sample that requires imputation frequently exceeds 50%, and in many cases exceeds 60%, 70% and even up to as high as 80%.

A.70 The high levels of imputation mean a low number of matches. Whilst a regression model based on the matches explains the variation in the matched data with some high levels of success, it is not always sufficiently capturing the variation in prices for those machines that require imputation. This is caused by small sample sizes as well as a limited number of combinations of RAM, hard drive and so forth.

A.71 More clearly, the matched data can fail to cover the different combinations that are observed in the births or deaths datasets even when all values of the explanatory variables for the imputed data lie within the ranges of the matched data. Including interaction variables in the modelling process provides some gains in this regard, but empirical studies have indicated that variation in predicted prices can be excessive when a small number of observations are used to impute a large number of prices.

DIRECT HEDONIC PRICE INDEXES

A.72 A direct hedonic price index uses the hedonic function to directly determine a *price* index. The result of the regression modelling process is itself an index number, which measures the price change from the current period over some base period. The ABS investigated this method, and undertook research into the *multi-period pooled time dummy price index* which is one particular application of a direct hedonic price index, and is frequently discussed in statistical literature.

The Time Dummy

A.73 The pooled time dummy index methodology pools data from successive periods, and then uses this large pooled dataset to directly determine an index number which measures price change from the base period to the current period.

A.74 In each period t, we consider a sample of models of personal computers. From each element of the sample we observe a set of period t prices together with the k characteristics of the model. We denote this by

$$S^t = \left\{ p_i^t, \underset{\sim}{z}_i' = \left(z_{1i}, \ldots, z_{ki} \right) \right\}$$

In addition to the k characteristic variables $z_{1i}, \ldots, z_{ki}$, we can consider an additional (0,1) indicator variable d_{ti} such that

$$d_{ti} = \begin{cases} 1, & \text{if } i \in S^t \\ 0, & \text{otherwise} \end{cases}$$

This variable indicates from which period the observation was observed. This type of indicator variable is called a "time dummy" variable.

If we consider many periods, say for $t=0,...,t$, for each sample S^τ we can include a dummy variable for each of the periods.

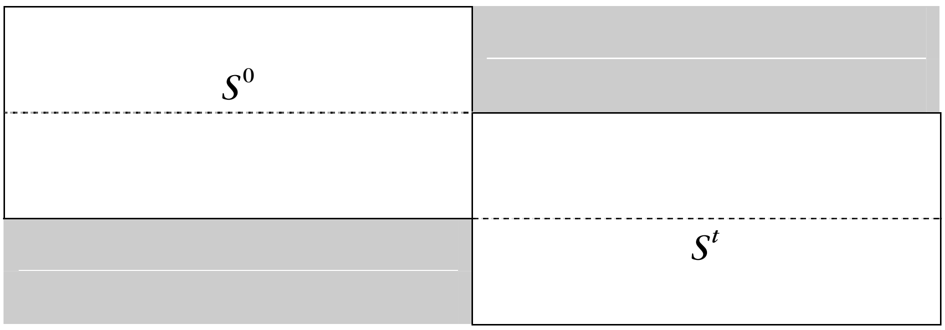
For each period we then observe:

$$S^\tau = \left\{ p_i^\tau, \underset{\sim}{z}_i' = \left(z_{1i}, \ldots, z_{ki} \right), d_{1i}, d_{2i}, \ldots, d_{ti} \right\}$$

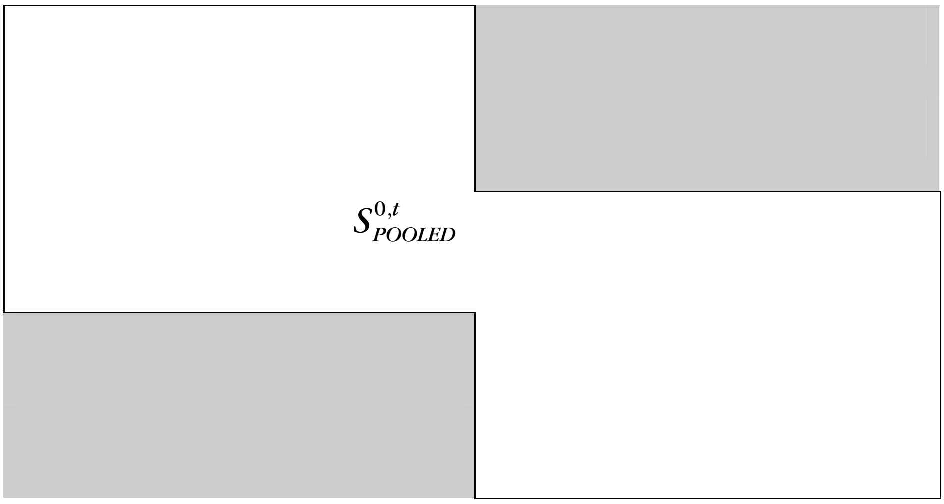
A.75 By convention we do not define a time dummy variable for the "base period", since this may be determined from the set of other time dummy variables.

Pooling data over multiple periods

A.76 If we consider two periods, 0 and t, we have previously described the disjoint samples:



A.77 We can take the union (or join) of these two sets to produce a pooled sample. We can denote this pooled sample by $S^{0,t}_{POOLED}$, which we illustrate by:



A.78 Following this approach, we can successively add sample from more periods.

$$S^{0,t}_{POOLED} = S^0 \cup S^t$$

So,

$$\begin{aligned}
S^{0,1}_{POOLED} &= S^0 \cup S^1 \\
S^{0,1,2}_{POOLED} &= S^0 \cup S^1 \cup S^2 \\
&= S^{0,1}_{POOLED} \cup S^2 \\
S^{0,\ldots,t}_{POOLED} &= \bigcup\nolimits_{\tau=0}^{t} S^{\tau} \\
&= S^{0,\ldots,t-1}_{POOLED} \cup S^t
\end{aligned}$$

Note that the observations in this pooled dataset are

$$S^{0,\ldots,t}_{POOLED} = \left\{\tau : \tau = 0,\ldots,t : p_i^{\tau}, z_i' = \left(z_{1i},\ldots,z_{ki}\right), d_{1i}, d_{2i},\ldots,d_{ti}, i = 1,\ldots,N\right\}$$

A.79 There is an important distinction between the price data in the pooled datasets used for the direct approach and those datasets used for the indirect approach. For each of the datasets used for the indirect approach, the price data come from a single period - all price observations are from the same point in time. For the direct approach, however, the act of pooling data means that the price observations are from many different periods.

Regression modelling for the Time Dummy approach

A.80 Price (or a function of price) can be expressed as a function of descriptive variables. These can either be quantitative variables (such as megabytes of RAM, gigabytes of hard drive space etc.) or indicator variables (such as 1="PC has a DVD burner", 0="Otherwise"). Included amongst the class of indicator variables are the time dummy variables described previously.

A.81 When expressing price as a function of explanatory variables, the function used is called the hedonic function. For the multi-period pooled time dummy approach as researched by the ABS, the functional form is fixed from period to period. Once the model has been determined, the form including which variables are included and which variables are excluded, is held constant from period to period. The model is re-estimated at regular intervals (for example, every eighteen months).

A.82 In the case of the multi-period pooled approach as applied by the ABS, the form of the regression is fixed in advance - that is, the same parameters are re-estimated each period, regardless of their actual contribution to the model. More clearly, for each period no attempt is made to reduce the model to a more parsimonious form. This is in contrast to the indirect approach that saw new models determined each period.

A.83 The models trialled by the ABS were the so-called "double log" models. If we consider quantitative explanators $\left(z_{1i},\ldots,z_{mi}\right)$ and indicator variables $\left(z_{m+1,i},\ldots,z_{ki}\right)$ the models took the form of:

$$\ln\left(p_i^\tau\right)=\beta_0+\beta_1\ln\left(z_{1i}\right)+\ldots+\beta_m\ln\left(z_{mi}\right)+\beta_{m+1}z_{m+1,i}+\ldots+\beta_k z_{ki}+\delta_1 d_{1i}+\ldots+\delta_t d_{ti}+\varepsilon_i$$

A.84 To fix the form of the model, a regression was run on historic data (in practice, an early portion of dataset $S_{POOLED}^{0,\ldots,t}$) to determine a parsimonious form. This process determines any appropriate transformations to be included in the model, and most importantly which explanatory variables are to be included in the final model. Note that, unlike the indirect approach, the method applied by the ABS here was to fix the form of model rather than to update each successive period.

A.85 After the form of the model is determined, the fixed form regression is re-run each period to redetermine parameter estimates by running a regression on the complete dataset, $S_{POOLED}^{0,\ldots,t}$

A.86 In this process we note that for observations from period 0, all time dummy variables take the value zero. That is, $d_{\tau i}=0$, for all $\tau=1,\ldots,t$.

A.87 This means that an estimate of the prices from period 0 is given by:

$$\ln\left(\hat{p}_i^0\right)=\hat{\beta}_0+\hat{\beta}_1\ln\left(z_{1i}\right)+\ldots+\hat{\beta}_m\ln\left(z_{mi}\right)+\hat{\beta}_{m+1}z_{m+1,i}+\ldots+\hat{\beta}_k z_{ki}$$

A.88 Similarly, if we take the parameter estimates determined from the complete dataset, we also note that for observations from period *t* alone, all time variables take the value 0, except that time dummy variables indicating data from period *t*. That is, $d_{\tau i}=0$, for all *t*=1,...,*t*-1, and further $d_{ti}=1$

A.89 This means that an estimate of the prices from period *t* is given by:

$$\ln\left(\hat{p}_i^t\right)=\hat{\beta}_0+\hat{\beta}_1\ln\left(z_{1i}\right)+\ldots+\hat{\beta}_m\ln\left(z_{mi}\right)+\hat{\beta}_{m+1}z_{m+1,i}+\ldots+\hat{\beta}_k z_{ki}+\hat{\delta}_t$$

Taking the difference of these two equations, we get

$$\begin{aligned}\ln\left(\hat{p}_i^t\right)-\ln\left(\hat{p}_i^0\right)&=\hat{\beta}_0+\hat{\beta}_1\ln\left(z_{1i}\right)+\ldots+\hat{\beta}_m\ln\left(z_{mi}\right)+\hat{\beta}_{m+1}z_{m+1,i}+\ldots+\hat{\beta}_k z_{ki}+\hat{\delta}_t\\&\quad-\left[\hat{\beta}_0+\hat{\beta}_1\ln\left(z_{1i}\right)+\ldots+\hat{\beta}_m\ln\left(z_{mi}\right)+\hat{\beta}_{m+1}z_{m+1,i}+\ldots+\hat{\beta}_k z_{ki}\right]\\&=\hat{\delta}_t\end{aligned}$$

A.90 Conducting transforms, we see that

$$\ln\left(\hat{p}_i^t\right) - \ln\left(\hat{p}_i^0\right) = \hat{\delta}_t$$

$$\ln\left(\frac{\hat{p}_i^t}{\hat{p}_i^0}\right) = \hat{\delta}_t$$

$$\frac{\hat{p}_i^t}{\hat{p}_i^0} = \exp\left(\hat{\delta}_t\right)$$

A.91 Note that this quantity does not depend upon any particular observation; it holds true for all values of i. The implication of this phenomenon is that the price change predicted for all observations is the same.

A.92 This property of the hedonic multi-period pooled time dummy price index can best be illustrated with an example. Consider the situation where the variation in price is explained by log(RAM) alone. That is:

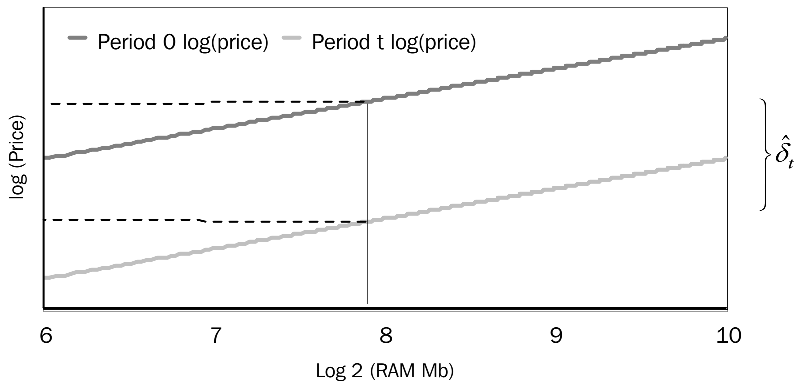
$$\ln\left(\hat{p}_i^t\right) = \hat{\beta}_0 + \hat{\beta}_1 \ln\left(RAM_i\right) + \hat{\delta}_t$$

$$\ln\left(\hat{p}_i^0\right) = \hat{\beta}_0 + \hat{\beta}_1 \ln\left(RAM_i\right)$$

$$\ln\left(\hat{p}_i^t\right) - \ln\left(\hat{p}_i^0\right) = \hat{\delta}_t$$

A.93 These relationships are shown graphically in Figure 12.

Figure 12 MULTI-PERIOD POOLED TIME DUMMY HEDONIC INDEX



A.94 As can be seen, the fitted or predicted lines for $\ln\left(\hat{p}_i^0\right)$ and $\ln\left(\hat{p}_i^t\right)$ differ by the constant value $\hat{\delta}_t$.

A.95 A hedonic multi-period pooled time dummy price index comparing the price at period t with the price at period 0 is given by:

$$I_{PTD}^t = \exp\left(\hat{\delta}_t\right) \times 100$$

Direct approach - advantages and disadvantages

A.96 The advantage of this approach is that the predictive power of the model is quite high since the model uses data from multiple periods (that is, the model error is lower).

A.97 However a number of concerns exist for the multi-period pooled time dummy approach:

1. New estimates for earlier periods;

2. Insensitivity to major changes in underlying characteristics e.g. the introduction of a new feature such as a DVD drive or a large change in an existing feature such as moving to much larger hard drives;

3. Inability to decompose the index movement into that part due to directly observed price changes and that part due to the hedonic estimates; and

4. Lack of transparency to users.

New estimates for earlier periods

A.98 The multi-period pooled time dummy approach is used to produce a price index each period, measuring the price of the current period as compared with the base period. As a side effect of this process, measures are also produced for earlier periods. For example, using all periods pooled up to and including period t, a measure is produced comparing period t to period 0, and as a side effect measures are also produced for period t against each of the periods t-1, t-2, ..., 2, and 1.

A.99 The ABS considers these "ancillary estimates" to be of little or no value because they have no bearing on the current period estimate.

A.100 Some observers suggest that the existence of these data imply the need to revise earlier period index scores. The ABS does not agree with this observation. Consider the following: if a base period is June 2000, it is possible to determine an estimate for December 2000 by pooling observations from June 2000 up to and including December 2000. An estimate for June 2003 is determined by pooling observations from June 2000 up to and including June 2003. Different estimates are also produced for December 2000 at this time. However, the ABS does not consider that the observations over the period January 2001 through June 2003 are informative with regards to price movements over the period June 2000 through December 2000 and so no revisions should be put through for these periods.

A.101 The debate around the importance or otherwise of these data casts uncertainty over the application of this technique.

Insensitivity to major changes

A.102 The ABS research into this method involved the use of a fixed form of the hedonic function. This means that the form of the model for the multi-period pooled fixed time dummy approach is fixed for the entire pooled dataset. The ABS believes that this approach does not pick up turning points: that is, it is insensitive to the changes in technology in the marketplace. In addition to failing to account for new characteristics (for example, DVD burners would not be included in a model determined in 2000 but are standard features today), fixing the model does not allow account to be taken of changes in the relative importance of different characteristics.

A.103 An example of this insensitivity to turning points in the market was seen in the early 2000s when hard drive sizes reached a static situation. A hedonic form determined at this time may not include hard drives (i.e., in the initial modelling process they may have been flagged as being non-significant in terms of explaining price variation and subsequently dropped from the model). Subsequent advances in technology have meant that hard drives have rapidly increased in size, with a result being they have a substantially different impact as a determinant of PC price change. Fixing a model in this manner would mean that a key predictor of prices at a later time would not be included in an initial model.

Inability to decompose

A.104 The price index produced by the multi-period pooled time dummy approach is a single measure. It is not possible to decompose this measure to show the price movements that come from actual observations as opposed to the component that is imputed or modelled by the process. Therefore, for a given change in an index, it is not possible to ascertain what is causing or driving change, making output editing, data confrontation and validation difficult.

Transparency to users

A.105 The ABS is concerned about the transparency of this approach as it is difficult to explain to users, especially when compared with approaches that use more traditional price index methods such as the indirect approach. The direct approach becomes more cumbersome when considering that the technique can fail to detect turning points and cannot be decomposed into real and imputed movements.

THE DOUBLE IMPUTATION PRICE INDEX

A.106 The double imputation price index was first proposed by de Haan, J, (2003), in a report to the Inter-secretariat Working Group on Price Statistics (IWGPS). The double imputation index is a combination of the direct and indirect approaches: it uses a direct approach to determine a price change over two consecutive periods, and in a manner very similar to indirect approaches this imputed price change is then used to impute price movements for both new and superseded models. These price movements are combined with matched model price movements to produce a price index for two consecutive periods. This price index is then chained to provide a measure of price change from a base period to the current period.

The two-period pooled time dummy index

A.107 The approach begins with the construction of a pooled time dummy index, but restricts pooling to two consecutive periods. That is, the pool would consist of data only from period *t*-1 and period *t*. Thus the dataset used for regression modelling is:

$$S_{POOLED}^{t-1,t}$$

A.108 This differs from the multi-period pooled approach trialled by the ABS in two key ways: first, pooling only occurs for two consecutive periods. Second, the resulting price index is not the final measure of change from period *t*-1 to period *t*, but is instead used in later stages of the calculation.

A.109 In trialling this methodology the ABS again used "double log" models. If we consider quantitative explanators $\left(z_{1i},\ldots,z_{mi}\right)$ and indicator variables $\left(z_{m+1,i},\ldots,z_{ki}\right)$ the models took the form of:

$$\ln\left(p_i^{\tau}\right)=\beta_0+\beta_1\ln\left(z_{1i}\right)+\ldots+\beta_m\ln\left(z_{mi}\right)+\beta_{m+1}z_{m+1,i}+\ldots+\beta_k z_{ki}+\delta_t d_{ti}+\varepsilon_i,\ \tau=t-1,t$$

A.110 Here we are only considering two periods worth of data, so only one dummy variable is required. We use the variable:

$$d_{ti} = \begin{cases} 1, & i \in S^t \\ 0, & i \in S^{t-1} \end{cases}$$

A.111 For this process, a regression is run each period to determine a parsimonious form, and to estimate the parameters $\beta_0, \ldots, \beta_k, \delta_t$.

A.112 This differs from the approach trialled for the multi-period pooled time dummy in that both the form and the parameters are re-determined each period.

A.113 Following the logic from earlier, this means that a prediction of the prices from period *t*-1 is given by:

$$\ln\left(\hat{p}_i^{t-1}\right) = \hat{\beta}_0 + \hat{\beta}_1 \ln\left(z_{1i}\right) + \ldots + \hat{\beta}_m \ln\left(z_{mi}\right) + \hat{\beta}_{m+1} z_{m+1,i} + \ldots + \hat{\beta}_k z_{ki}$$

and prices from period *t* are estimated by:

$$\ln\left(\hat{p}_i^{t}\right) = \hat{\beta}_0 + \hat{\beta}_1 \ln\left(z_{1i}\right) + \ldots + \hat{\beta}_m \ln\left(z_{mi}\right) + \hat{\beta}_{m+1} z_{m+1,i} + \ldots + \hat{\beta}_k z_{ki} + \hat{\delta}_t$$

A.114 Taking the difference of these two equations, we get

$$\begin{aligned} \ln\left(\hat{p}_i^{t}\right) - \ln\left(\hat{p}_i^{t-1}\right) &= \hat{\beta}_0 + \hat{\beta}_1 \ln\left(z_{1i}\right) + \ldots + \hat{\beta}_m \ln\left(z_{mi}\right) + \hat{\beta}_{m+1} z_{m+1,i} + \ldots + \hat{\beta}_k z_{ki} + \hat{\delta}_t \\ &\quad -\left[\hat{\beta}_0 + \hat{\beta}_1 \ln\left(z_{1i}\right) + \ldots + \hat{\beta}_m \ln\left(z_{mi}\right) + \hat{\beta}_{m+1} z_{m+1,i} + \ldots + \hat{\beta}_k z_{ki}\right] \\ &= \hat{\delta}_t \end{aligned}$$

A.115 Note that this quantity does not depend upon any particular observation; it holds true for all values of *i*.

A.116 Conducting transforms, we see that

$$\begin{aligned} \ln\left(\hat{p}_i^{t}\right) - \ln\left(\hat{p}_i^{t-1}\right) &= \hat{\delta}_t \\ \ln\left(\frac{\hat{p}_i^{t}}{\hat{p}_i^{t-1}}\right) &= \hat{\delta}_t \\ \frac{\hat{p}_i^{t}}{\hat{p}_i^{t-1}} &= \exp\left(\hat{\delta}_t\right) \end{aligned}$$

A.117 A hedonic two-period pooled time dummy price index comparing the price at period *t* with the price at period *t*-1 is given by:

$$I_{2TD}^{t-1,t} = \exp\left(\hat{\delta}_t\right) \times 100$$

A.118 As with all price indexes, we interpret the index $I_{2TD}^{t-1,t}$ as giving the price change from period *t*-1 to period *t*.

The matched model

A.119 The matched model price index from the matched sample is:

$$I_{MM}^{t-1,t} = \prod_{i \in S^{t-1} \cap S^t} \left(\frac{p_i^t}{p_i^{t-1}} \right)^{\frac{1}{N_{S^{t-1} \cap S^t}}} \times 100$$

This index is a geometric mean.

Constructing the double imputation index

A.120 The double imputation index is constructed from the two indexes

- $I_{MM}^{t-1,t}$ for observations matched between the two periods; and
- $I_{2TD}^{t-1,t} = \exp\left(\hat{\delta}_t\right)$ for new items or discontinued lines.

A.121 The double imputation price index measuring price change of personal computers from period *t*-1 to period *t* is then a weighted geometric mean of the two component indexes:

$$\begin{aligned} I_{DI}^{t-1,t} &= \left[I_{MM}^{t-1,t} \right]^{f_M} \left[\exp\left(\hat{\delta}_t\right) \times 100 \right]^{1-f_M} \\ &= \left[I_{MM}^{t-1,t} \right]^{f_M} \left[I_{2TD}^{t-1,t} \right]^{1-f_M} \end{aligned}$$

A.122 Here the value f_M is the weighted fraction of matched observations from period *t*-1 to period *t*.

A.123 To construct a price index to some earlier price reference (base) period, a chained series is formed.

$$I_{DI,Chain}^{t} = I_{DI}^{t-1,t} \times I_{DI,Chain}^{t-1}$$

The double imputation index and concerns regarding other methodologies

A.124 In researching appropriate methods for hedonic price indexes, several concerns were raised regarding both the indirect and direct methods for hedonic price indexes. The double imputation price index has been assessed against these concerns, with outcomes detailed below.

Extrapolation is no longer a concern

A.125 The double imputation price index does not suffer from the extrapolation concerns which arise when constructing an indirect hedonic price index. Whereas the modelling process for the indirect hedonic price index estimates coefficients which are used in turn to predict a price measure, the modelling process for the double imputation price index directly predicts a price movement (the value of the two period pooled double imputation price index), which itself is a summary of the price movements of all observations from the two periods. This predicted price movement is then used for those observations that were not matched between the two periods.

Predictive power substantially improved

A.126 The hedonic imputation method was judged deficient in terms of predictive power, which in part arises because of small sample sizes observed when matching between two periods. The double imputation price index avoids this issue. The hedonic imputation price index used the matched sample between two periods to perform regression, which means the sample size can be no bigger than the smaller of the two sample sizes from either period. The double imputation price index uses the pool of the data from both periods, so that the sample size is equal to the sum of the number of observations from each period.

Revision of earlier data not an issue

A.127 A consequence of the multi-period pooled technique used to construct a direct hedonic price index is that additional data are available each period which may be interpreted by some observers as revising earlier period index scores. This is an outcome of pooling data and adding to the pool for each new period. This issue does not arise for the double imputation hedonic price index because pooling is only undertaken for two consecutive periods, and resulting two period movements are then chained.

Sensitivity to changes in characteristics

A.128 The ABS implementation of the multi-period pooled time dummy price index saw the hedonic function held constant over time. The ABS has concerns over this approach in that it does not pick up turning points: that is, it is insensitive to the changes in technology in the marketplace. In addition to failing to account for new characteristics, fixing the model does not allow account to be taken of changes in the relative importance of different characteristics. This issue does not arise for the double imputation price index, since the model and its functional form are re-determined for each period.

Decomposition and transparency to users

A.129 The double imputation hedonic price index can be written as

$$\begin{aligned} I_{DI}^{t-1,t} &= \left[I_{MM}^{t-1,t} \right]^{f_M} \left[\exp\left(\hat{\delta}_t\right) \times 100 \right]^{1-f_M} \\ &= \left[I_{MM}^{t-1,t} \right]^{f_M} \left[I_{2TD}^{t-1,t} \right]^{1-f_M} \end{aligned}$$

A.130 Seeing the index in this form, a sensible interpretation of the double imputation price index is that it:

1. uses the matched model for those observations common to each period;

2. uses a direct index to impute price movements for those observations not common to both periods.

A.131 In this manner, a period-to-period movement is easily decomposed into a portion which can be explained by movements from continuing models of personal computer, and a portion which may be explained by the difference between new and superseded models of personal computers.

BIBLIOGRAPHY

Bureau of Economic Analysis (BEA), 'Indexes for Private Fixed Investment by Type - table 5.3.4.'

Available online:

http://www.bea.gov/bea/dn/nipaweb/SelectTable.asp?Selected=N

de Haan, J 2003, 'Time Dummy Approaches to Hedonic Price Measurement', Paper presented at the Seventh Meeting of the International Working Group on Price indexes (Ottawa Group), Paris 27-29 May, 2003

Diewert, WE 2003, 'Hedonic Regressions: A Consumer Theory Approach', *Scanner Data and Price indexes*, (eds) Robert C. Feenstra and Matthew D. Sharpiro, NBER Studies in Income and Wealth, Vol. 64 (Chicago: University of Chicago Press), pp. 317-48.

International Labour Organization (ILO), International Monetary Fund (IMF), Organisation for Economic Cooperation and Development (OECD), Statistical Office of the European Communities (Eurostat), United Nations, and the World Bank 2004, *Consumer Price Index Manual, Theory and Practice*

Available online:

http://www.ilo.org/public/english/bureau/stat/guides/cpi/index.htm

International Labour Organization (ILO), International Monetary Fund (IMF), Organisation for Economic Cooperation and Development (OECD), United Nations Economic Commission for Europe (UNECE), and the World Bank 2004, *Producer Price Index Manual, Theory and Practice*

Available online:

http://www.imf.org/external/np/sta/tegppi/

Schultze CB, and Mackie C, (eds.) 2002, *At What price? Conceptualising and Measuring Cost-of-Living and Price Indexes*, Committee on National Statistics, National Research Council (Washington DC: National Academy Press).

Triplett, J, 2004, *Handbook on hedonic indexes and quality adjustments in price indexes: special application to information technology products.* OECD STI working paper 2004/9

FOR MORE INFORMATION . . .

INTERNET	**www.abs.gov.au** the ABS web site is the best place to start for access to summary data from our latest publications, information about the ABS, advice about upcoming releases, our catalogue, and Australia Now—a statistical profile.
LIBRARY	A range of ABS publications is available from public and tertiary libraries Australia-wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our web site for a list of libraries.
CPI INFOLINE	For current and historical Consumer Price Index data, call 1902 981 074 (call cost 77c per minute).
DIAL-A-STATISTIC	This service now provides only current Consumer Price Index statistics call 1900 986 400 (call cost 77c per minute).

INFORMATION SERVICE

Data already published that can be provided within five minutes will be free of charge. Our information consultants can also help you to access the full range of ABS information—ABS user pays services can be tailored to your needs, time frame and budget. Publications may be purchased. Specialists are on hand to help you with analytical or methodological advice.

PHONE	1300 135 070
EMAIL	client.services@abs.gov.au
FAX	1300 135 211
POST	Client Services, ABS, GPO Box 796, Sydney NSW 2001

WHY NOT SUBSCRIBE?

ABS subscription services provide regular, convenient and prompt deliveries of selected ABS publications and products as they are released. Email delivery of monthly and quarterly publications is available.

PHONE	1300 366 323
EMAIL	subscriptions@abs.gov.au
FAX	(03) 9615 7848
POST	Subscription Services, ABS, GPO Box 2796Y, Melbourne Vic 3001

2645800001031

RRP $10.00