Seasonality and Prepackaged Software Price Indexes *

Adam Copeland†

August 2008

Abstract

In this paper, we construct a seasonally-adjusted price index for prepackaged software using detailed and comprehensive scanner data from the NPD Group. We document a large sales surge over the winter-holiday and claim that this seasonality is being driven by consumer heterogeneity. We introduce a novel approach for constructing the software component of the cost-of-living price index which explicitly accounts for this type of consumer heterogeneity. Using this index and the detailed product-level data, we find from 1997 to 2003 constant-quality software prices declined at an average 16.8 percent at an annual rate. To demonstrate the importance of properly accounting for heterogeneity, we compare a Mudgett-Stone price index, a representative-consumer approach to accounting for seasonality, to our index, and find substantial differences in the estimates of constant-quality annual price declines.

*We would like to thank Ana Aizcorbe for many insightful discussions. In addition, comments by Dennis Fixler, Marshall Reinsdorf and other colleagues at the Bureau of Economic Analysis were quite helpful. Finally I would like to thank Andrew Miller for substantive help as a research assistant. The views expressed herein are our own and not necessarily those of the Bureau of Economic Analysis or the US Department of Commerce.

†U.S. Department of Commerce, Bureau of Economic Analysis, Office of the Chief Statistician BE-40, Washington, DC 20230; phone: (202) 606-9948; e-mail: adam.copeland@gmail.com; webpage: http://www.copeland.marginalq.com

1 Introduction

Software plays an important role in the information technology revolution that has swept the US. Yet compared to other information technology products such as computers and semiconductors, relatively little research has been devoted to understanding this sector.[1] This paper aims to help fill this gap by constructing a price index for prepackaged software, an increasingly large segment of software investment in the US.[2] A better understanding of software pricing trends is the first step towards a deeper appreciation of the role of software as a driver of US productivity growth.[3] Further, constructing a price index that properly accounts for the seasonality in software purchases is important for accurately measuring real personal consumption expenditures on prepackaged software.

A number of researchers have already produced price indexes for software; this paper builds upon this small literature in two main ways. First, unlike most previous work, our index is based on detailed, industry-wide scanner data, as opposed to a small subset of products.[4] Hence, our index is representative of price changes throughout the industry, and so generates robust measures of constant-quality price change. Second, we develop and implement an approach to constructing price indexes that accounts for the large amount of seasonality within the software industry. Because previous software price indexes have ignored seasonality, our inclusion of seasonality adjustment is, in itself, a modest improvement on previous empirical work. Perhaps more importantly however, we introduce a new variation to the existing set of empirical methods on accounting for seasonality when constructing a price index. Unlike previous methods, our approach explicitly accounts for consumer heterogeneity, which we claim is the driving force behind software's seasonal fluctuations.[5]

[1]Jorgenson (2001) emphasizes that information gaps remain about understanding software pricing trends.

[2]See Parker and Grimm (2000) for details on the high rate of growth of prepackaged software.

[3]Oliner and Sichel (1994) make the case that software and computer-services labor should be considered along with computer hardware in understanding whether investment in information technology is a main driver of US productivity.

[4]For example, Oliner and Sichel (1994) and McCahill (1997) study price movements of word processors, spreadsheet, and database software applications, Abel, Berndt, and White (2003) examine price movements of Microsoft's personal computer software products, and Gandal (1994) analyzes prices of spreadsheets.

[5]In this paper we consider seasonal adjustment from an index-number approach. Another strategy

The scanner data used in this research came from the NPD Group, and contains monthly data on units-sold and revenue-earned through retail channels nationwide from January 1997 to August 2004.[6] This data set is unusual for its comprehensiveness and detail. The NPD Group claims their data covers 84 percent of all retail sales, including transactions at big-box stores and over the internet. Further, an observation is at the publisher/title/operating-system level, providing us with an intimate look into the software market.

Examining the data, we find the majority of software products experience a significant boost in unit sales over November, December and January, the winter-holiday season. Given the particular correlations of prices and sales over these months, we argue that the arrival of casual, once-a-year shoppers are the main driving force behind the surge in winter-holiday sales. Using this insight, we argue that the appropriate method of accounting for seasonality is to construct two indexes: one index for casual, once-a-year consumers, and another for regular, year-round shoppers. Given the economic behavior behind the seasonality, the Mudgett-Stone index approach to constructing price indexes does not properly account for the seasonality. This is striking because given there is seasonality in the data, Alterman, Diewert, and Feenstra (1999) claim the Mudgett-Stone index should be considered the best measure of annual price change (page 48). The differences between our index and the Mudgett-Stone index can be substantial. For the software market as a whole, our index measures the average constant-quality price decline at 16.8 percent at an annual rate. Using the standard Mudgett-Stone approach, constant-quality annual price change averages 18.6 percent. Naturally, there are larger differences at lower levels of aggregation. For example, for the software categories PC Games, Personal Productivity and System Utilities, the differences in estimates of constant-quality annual price change between our index and the Mudgett-Stone index are 2.5, 3.3, and 8.3 percent, respectively. These differences arise, in part, because our approach uses more of the software data relative to the Mudgett-Stone technique. In particular,

would be to use a statistical approach to seasonal adjustment. This approach, which is used by the Bureau of Labor Statistics, first constructs a price index without regard to seasonality. A statistical algorithm, such as x-12-ARIMA, is then applied to the price index which removes seasonal patterns in the index.

[6] This scanner data is not used by the Bureau of Economics Analysis (BEA) in the construction of the national accounts.

because software typically has a short product life, year-over-year price relatives can be constructed for only 30 percent of products offered in a typical quarter. Quarter-to-quarter price relatives, however, can be calculated for over 80 percent of products offered in a typical quarter. Accordingly, the Mudgett-Stone technique, which uses year-over-year price relatives, only includes information on price changes for a small subset of all products. In contrast, because our technique uses both year-over-year and quarter-to-quarter price relatives, it incorporates price changes for over 80 percent of products offered in a typical period.[7]

Finally, as a point of reference, we compare our index to the Computer Software and Accessories consumer price index published by the Bureau of Labor Statistics (BLS). This BLS price index is used by the Bureau of Economic Analysis to deflate consumer expenditures on software, which is almost entirely prepackaged software. Over the same period, this index falls at an annual average rate of only 7 percent. We show, however, that the difference in measured price change between the BLS and our price indexes is almost entirely related to differences in data, as opposed to the method of index construction.

The construction of cost-of-living price indexes when there is seasonality in the data is an old and well-known problem in economics. In Diewert (1998), Diewert (1999), Erwin Diewert laid down the theoretical foundations for an "economic" approach to constructing price indexes when there is seasonality in the data. In contrast to the atheoretical, statistical method, Diewert formally derived a set of assumptions on the consumer's utility function that are required to justify particular price index formulas. More recently, Nesmith (2007) has extended Diewert's results to a larger class of consumer decision problems. This paper builds upon these efforts by further relying on economic theory to guide the construction of price indexes when seasonality in the data exists. We argue that consumer heterogeneity drives the seasonality in the software market, and adapt existing price index techniques to directly account for the changing mix of consumers over the calendar year.

By taking consumer heterogeneity into account, this paper also builds upon the

[7]Note the Mudgett-Stone index employs a year-over-year approach to account for seasonality. This technique is different from those statistical approaches which compute seasonal factors for each period within the year, where the sum of the seasonal factors are constrained to equal one.

work of Griliches and Cockburn (1994), Fisher and Griliches (1995), and more recently, Aizcorbe, Bridgman, and Nalewaik (2007) and Aizcorbe and Copeland (2006). These works consider the effects of consumer heterogeneity on the construction of price indexes with respect to the introduction of new goods. We differ from these papers because of our focus on seasonality and in how we account for consumer heterogeneity when constructing a price index.

Empirically, our works builds upon a small body of work which, for the most part, studied small subsets of the prepackaged software market. An exception is Prud'homme, Sanga, and Yu (2005), who used samples of transaction level data from the Canadian market. Our data, while not the universe, is more comprehensive than their sample. Further, we have detailed enough information that we can distinguish between different versions of the same software product. In addition to the data differences, the current paper differs from Prud'homme, Sanga, and Yu (2005) because we take the seasonality of the data into account.[8]

The rest of the paper proceeds as follows. In section 2 we describe the data we use and present evidence of the prevalence of a winter-holiday seasonal effect across the software market. We layout our price index in section 3. In section 4 we present our main results. We then compare our price index against the Mudgett-Stone price index. Finally, we compare the results from our index against the relevant consumer price index series published by the Bureau of Labor Statistics. Section 5 concludes.

2 Data

In this section we describe the data on prepackaged software that we use. We then detail how we measure seasonality within the data and show how pervasive the winter-holiday seasonal effect is within the software market. Finally, we discuss how we deal with two data quality issues.

[8]Our results on the average annual decline in constant-quality price differ substantial from Prud'homme, Sanga, and Yu (2005). They report an average annual decline of 7.9 percent relative to our 16.8 percent. Our seasonal adjustments do not effect this gap, leaving data differences as the main explanation.

2.1 Description

We obtained prepackaged software industry data from the NPD Group. Software is prepackaged when it is sold or licensed in standardized form and is delivered in packages or as electronic files downloaded from the Internet. This is opposed to custom and own-account software which require larger degrees of tailoring to the specific application of the user.[9] The data are point-of-sale transaction data (i.e. scanner data) that are sent to the NPD Group from participating outlets. The data we purchased from NPD are retail sales, or transactions from warehouse club stores, internet retailers, office superstores, etc. NPD claims to cover 84 percent of the retail market, and so provides a clear picture of the prepackaged software retail market. The data are monthly observations at the national level, where a record is a product. For each observation, the revenue earned and the number of units sold that month are reported, allowing us to compute the average monthly price of the product. Further, the data include the name of the software publisher, and category and subcategory variables that provide an classification structure for grouping products. The time frame of the data ranges from January 1997 to August 2004 and includes 782,849 observations. Table 1 provides a summary of the data at the category level, showing the number of subcategories and observations within each category as well as the relative size of each category by units sold and revenue generated. PC Games is the largest category by far, accounting for 35 percent of total revenue and and almost half of all sales. Business and Finance, are the next two largest categories and together account for roughly 25 percent of total revenue generated within this market.

The prepackaged software market is turbulent, with many products entering and exiting the market over our 7 year sample. The mean lifespan for an average software product is 22.0 months. This statistic is skewed by a few extremely longed-lived products; the median length of time an average product is sold in the market is 17 months. As shown in table 2, there is a large amount of variation in the length of time a product is sold by category. The median number of months a product is sold ranges from 9 to 35 months, where System Utilities products have the shortest average lifespan and PC

[9]These definitions follow those used to measure prepackaged software in the U.S. national income and product accounts. See Parker and Grimm (2000) for more details.

Category	Subcategories	Observations	Unit sales (millions)	Revenue (millions)	Suppressed (percent)
Business	23	108,216	133	12,940	0.0037
Education	30	150,213	449	9,633	0.0009
Finance	3	13,239	262	11,985	0.0002
Imaging/Graphing	16	76,341	195	8,861	0.0009
Operating System	3	14,068	71	7,032	0.0004
PC Games	13	294,243	1,482	34,505	0.0002
Personal Productivity	33	75,637	183	5,910	0.0017
System Utilities	25	50,892	207	9,754	0.0011
Total	146	782,849	2,983	100,619	0.0007

Table 1: Data Summary

Category	Mean	Median
Business	15.5	10
Education	27.8	26
Finance	23.0	19
Imaging/Graphics	19.3	14
Operating System	16.1	13
PC Games	34.5	35
Personal Productivity	27.7	26
System Utilities	13.7	9
All	22.0	17

Table 2: Prepackaged Software Life (Months)

	log(Price)		log(Sales)	
Parameter	Coefficient	Standard Error	Coefficient	Standard Error
$I_{t=2}$	-0.015	0.00298	0.192	0.00697
$I_{t=3}$	-0.037	0.00299	0.217	0.00670
$I_{t=4}$	-0.046	0.00301	0.182	0.00705
$I_{t=5}$	-0.068	0.00313	-0.067	0.00732
$I_{t=6}$	-0.088	0.00324	-0.189	0.00758
$I_{t=7}$	-0.102	0.00328	-0.160	0.00767
$I_{t=8}$	-0.129	0.00339	-0.254	0.00793
$I_{t=9}$	-0.141	0.00350	-0.379	0.00819
$I_{t=10}$	-0.155	0.00350	-0.342	0.00820
$I_{t=11}$	-0.141	0.00362	-0.468	0.00848
$I_{t=12}$	-0.191	0.00377	-0.534	0.00882
$I_{t=13}$	-0.220	0.00384	-0.555	0.00898
$I_{t=14}$	-0.284	0.00401	-0.567	0.00938
$I_{t=15}$	-0.296	0.00418	-0.560	0.00978
$I_{t>=16}$	-0.359	0.00297	-0.826	0.00695

$I_{t=N}$ is a indicator variable equal to 1 when a product is in month N of its product cycle.

Table 3: Price and Sales Contours over the Product Cycle

Games the longest. The 22 month lifespan of the average prepackaged software product, however, is slightly deceiving. On average, a software product generates 75 percent of its lifetime revenue in the first year of its life. Hence, the tail end of software product's lifespan tends to be unimportant.

Behind this last fact is the declining trend in both price and units sold for prepackaged software over the product cycle. To measure how quickly price and units sales fall over the product cycle, we regressed the log of these variables on product cycle dummy variables, with fixed effects for each software product and using revenue weights. The estimated coefficients for the product cycle dummy variables are listed in table 3, and are precisely estimated. Our results indicate that prices fall over 19 percent over the first year a product is sold while unit sales decrease 50 percent. Hence, prepackaged software is a market where, over the product cycle, prices are rapidly falling alongside plummeting unit sales.

Month	Units Sold	Revenue Generated
Jan	8.98	9.08
Feb	8.85	8.90
Mar	9.42	9.64
Apr	6.66	6.66
May	5.51	5.72
Jun	7.55	7.91
Jul	5.81	5.97
Aug	6.15	6.32
Sep	7.60	7.92
Oct	6.23	6.67
Nov	8.76	8.43
Dec	18.47	16.78

Note: Results computed using data from Jan 1997 to Dec 2003

Table 4: Percent of Units Sold and Revenue Generated by Month

2.2 Seasonality

A priori it is not surprising that some products within the prepackaged software market exhibit strong seasonality over the winter holiday. The rise in U.S. retail sales over the winter holiday is well-known phenomenon.[10] Looking at the raw prepackaged software monthly data, it is not hard to see a significant winter holiday sales surge across many prepackaged software categories. Table 4 reports the percentage of units sold and revenue generated by month for all prepackaged software from January 1997 to December 2003. Clearly, December is a significant month for publishers of software, contributing over 18 percent of units sold annually and almost 17 percent of total revenue for the year. November and January also have a higher-than-average rate of sales, leading us to group November, December, and January together. To better examine the winter-holiday seasonal phenomenon, we define this group of months as the fourth quarter, and then define quarters 1-3 accordingly. We aggregate the data to the quarterly frequency; our sample encompasses 28 quarters, from quarter 1 of 1997 to quarter 4 of 2003.

[10]The U.S. Census Bureau publishes retail sales seasonal factors which show the large surge in sales in December for most kinds of businesses (see http://www.census.gov/svsd/www/adseries.html).

Identifying which products experience a winter-holiday seasonal effect is complicated by prepackaged software's short-lived product cycle. As described above, the median length of time a specific software product is sold is 17 months. Further, the vast majority of the revenue that software generates occurs within the first year, devaluing year-over-year comparisons. Hence, for the majority of cases, we are not able to definitively determine if there is a winter holiday seasonal effect at the product level.

To identify seasonal affects we consider the data at a higher level of aggregation–the subcategory level. For example, we consider whether a group of products such as "Foreign Language" software within the "Education" category experiences a fourth quarter boost in unit sales. We considered several approaches to determining when a subcategory of software experiences a winter-holiday seasonal affect. Our preferred approach, and the one we present here, uses x-12-ARIMA, a seasonal adjustment software packaged used and maintained by the U.S. Census Bureau.[11]

Using x-12-ARIMA, we produce a seasonally-adjusted series of units sold for each subcategory of software. For each subcategory and for each year, we state there is a winter-holiday seasonal effect when fourth quarter units sales in the seasonally-adjusted units sold series are less than fourth quarter unit sales in the non-seasonally-adjusted series.

Using this framework, we find that winter-holiday seasonality is pervasive in the prepackaged software market. Across all subcategories, the median value of the ratio of seasonally-adjusted to non-seasonally-adjusted units sold is 0.83 in the fourth quarter. For quarters 1 through 3, the median values of this ratio are 1.00, 1.21 and 1.11 respectively. Excluding the Business category, over 90 percent of software subcategories experience some degree of winter-holiday seasonality. Even for Business software, roughly 70 percent of software subcategories experience some winter-holiday seasonality.

The magnitude of the winter-holiday seasonality differs substantially across types of software. We denote the difference between the non-seasonally-adjusted and seasonally-adjusted revenue series as the "seasonal" component of revenue. For the fourth quarter, this seasonal component varies between 7 and 36 percent of total revenue (see table 5). Business software has the smallest seasonal component, in-line with our priors that work-

[11]See http://www.census.gov/srd/www/x12a/ for more information.

Category	Seasonal Component (percent)
Business	7
Education	31
Finance	36
Imaging/Graphics	18
Operating System	12
PC Games	34
Personal Productivity	21
System Utilities	15
All	28

Table 5: Seasonal Magnitude of Fourth-Quarter Revenue

related software is not much effected by the winter-holiday season. Education, Finance, and PC Games software have the largest winter-holiday effect. For these categories, more than 30 percent of total revenue in the fourth quarter is attributable to the seasonal component.

Using x-12-ARIMA, or any statistical approach, to define the seasonal component of revenue at the sub-category level is complicated by the endogenous entry problem. Software publishers regularly release both completely new software programs as well as updated versions of current software. Given the general rise in demand over the winter-holiday, firms may have an incentive to introduce new products over the winter-holiday to take advantage high demand. Indeed, if we take a lifetime sales weighted average of product introductions by quarter, we find a disproportionate amount of products are introduced around the beginning of the fourth quarter (see figure 1). In part because we aggregated the data to the subcategory level, the x-12-ARIMA approach does not take entry into account when it computes seasonal factors. As such we cannot distinguish how much of the winter-holiday sales surge is due to the introduction of new products and how much is a purely seasonal effect. Properly dealing with this endogenous entry problem would require a formal model of both consumer demand and publisher profit-maximization, something beyond the scope of this paper. Instead, we consider the seasonal factors produced by the x-12-ARIMA program to be good, first-cut approxima-

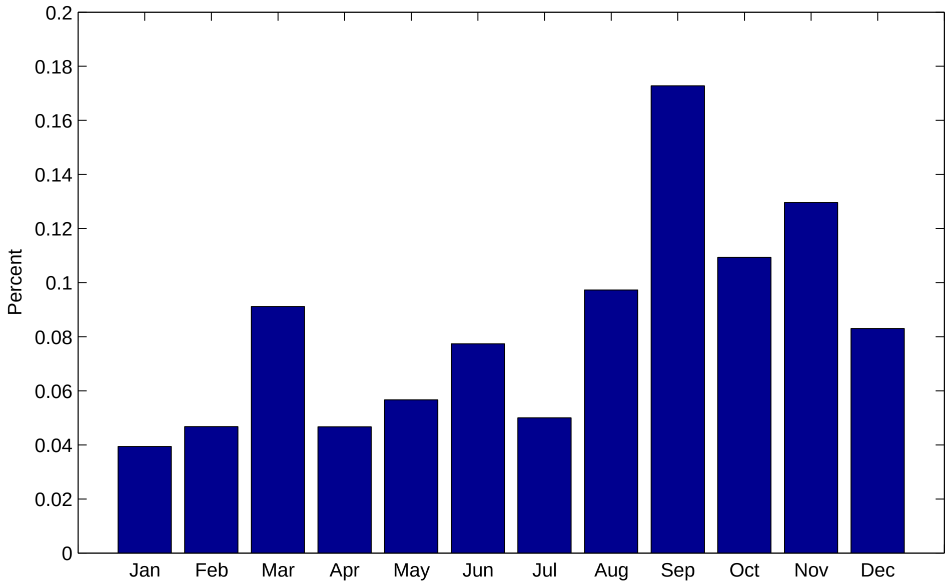

Figure 1: Product Introductions by Month

tions.

2.3 Data Quality

Before discussing how to construct price indexes that adjust for the winter-holiday seasonality, we address two quality issues in the data. First, observations are suppressed by the NPD Group whenever a product's sales for a particular month come from fewer than five retailers. NPD aggregates these suppressed data together into a single observation by subcategory. Because this aggregation mixes products inconsistently over time, these observations are excluded from our analysis. As shown in the last column of table 1, these observations account for a negligible share of the total units sold.[12]

Second, there are implausible quarterly price changes. As shown in table 6, the price ratio of adjoining quarters' prices has some extreme values. Categorizing which quarterly

[12] In addition to the removing the suppressed observations, we also removed four subcategories in which the percentage of suppressed observations accounted for over 60 percent of units sold. These subcategories are Data Center Management, Drivers/Spoolers, Engineering, and Network Resource Sharing, and together they make up an insignificant portion of all units sold.

Quantile	99%	95%	90%	75%	50%	25%	10%	5%	1%
Price Ratio	5.07	1.75	1.27	1.03	0.99	0.82	0.56	0.40	0.13

Table 6: Frequency Distribution of the Price Ratio of Adjoining Quarters' Prices

price changes are the result of measurement error can be difficult to discern. We take a conservative approach and drop the observations that are in the top and bottom 1 percent of the quarterly price ratio distribution. This translates into dropping quarterly price ratios below 0.13 and above 5.07.

3 Method

In this section we outline how we construct a Mudgett-Stone price index, the conventional approach to account for seasonality when measuring constant-quality price change. We then describe our price index.

3.1 Mudgett-Stone Price Index

We construct a Mudgett-Stone annual index following the description in Diewert (1998). We define each product both by its description and the quarter in which it was sold. Hence, a product sold in the fourth quarter of the year is compared with its namesake in the fourth quarter of the base period. Software with the same description but sold in different quarters are considered different products. We do not employ a fixed basket, but rather set the base year be the previous year when constructing price relatives. Denote $\mathcal{L}_t^{c,\mathrm{MS}}$ as the Laspeyres Mudgett-Stone quarterly price relatives for software in group c and $\mathcal{P}_t^{c,\mathrm{MS}}$ the Paasche. Letting $J_{t,s}^c$ be the set of products belonging to the software group c available in both quarter t and s, we compute the Laspeyres and Paasche price relatives using the following formulas,

$$\mathcal{L}_t^{c,\mathrm{MS}} = \frac{\sum_{j\in J_{t,t-4}^c} P_{jt}Q_{j,t-4}}{\sum_{j\in J_{t,t-4}^c} P_{j,t-4}Q_{j,t-4}}, \quad \mathcal{P}_t^{c,\mathrm{MS}} = \frac{\sum_{j\in J_{t,t-4}^c} P_{jt}Q_{jt}}{\sum_{j\in J_{t,t-4}^c} P_{j,t-4}Q_{jt}}, \tag{1}$$

where (P, Q) denote price and quantity. We aggregate to the annual frequency by taking a weighted average,

$$\mathcal{L}_a^{c,\text{MS}} = \sum_{s=1}^{4} w_s \mathcal{L}_s^{c,\text{MS}}, \qquad \mathcal{P}_a^{c,\text{MS}} = \sum_{s=1}^{4} w_s \mathcal{P}_s^{c,\text{MS}}. \tag{2}$$

The four quarters summed over in the above equation correspond with a calendar year a and w_s is the share of annual revenue for calender year a earned in quarter s. Finally, we compute an annual Fisher index by taking the geometric mean of the annual Laspeyres and Paasche price relatives and chaining them together. The above formulas produce a price index for software in group c, which can be defined for any grouping of software. In our empirical analysis, we compute both a market-level and category-level price indexes.

While we focus on the winter-holiday seasonality, this approach accounts for seasonality in each quarter of the year, and so the resulting index is a useful benchmark to compare against our index, which we label the "Heterogeneous" price index because it explicitly accounts for heterogeneity in consumers.

3.2 The Heterogeneous Price Index

As detailed in Diewert (1998), the Mudgett-Stone approach is based on a representative consumer framework and so accounts for seasonality by assuming that the representative consumer's tastes change from season-to-season. In our software example, this translates into a representative consumer having different tastes in each quarter of the year.

There are two major problems with the Mudgett-Stone annual index. First, because of its year-over-year approach, the revenue-weighted fraction of products we can match in the average quarter is only 32 percent. This low matching percentage is due to the large amount of exit and entry in the software market, described in section 2. In contrast, a quarter-to-quarter matched model index can match 84 percent of products on average. Consequently, using a Mudgett-Stone annual index constrains us to using only a small portion of the data.

Second, the Mudgett-Stone index provides an accurate measure of the change in cost-of-living under the assumption that a representative consumer can provide a good

1997	1998	1999	2000	2001	2002	2003	Average
0.99588	0.99362	1.01027	1.00233	1.00362	1.00641	0.99397	1.00087

Table 7: December Fisher Price Relatives

approximation of consumer behavior. The nature of the winter-holiday seasonality within the prepackaged software market, however, challenges this assumption. The overall surge in units sales in the fourth quarter is too large to be explained by increased shopping intensity from the same pool of households who show up throughout the year (see table 4). But if new households show up in fourth quarter, how are these casual shoppers different from the regular shoppers who buy throughout the year? A New York Times article describes how the video game industry retailers reconfigure stores for the winter holiday to cater these casual, once-a-year shoppers.[13]

There is also empirical evidence that these fourth-quarter casual shoppers are different from regular shoppers. With the surge in fourth-quarter prepackaged software sales, a signal of high demand, we would expect an accompanying rise in price. In the data, however, we see at most a slight uptick in price. Using the original monthly data, a maximum-overlay Fisher price index computes an average price increase of only 0.09 percent in December (see table 7). This dynamic in the data of average price *not* climbing during periods of high demand is a puzzle seen in other retail markets and is an active field of research. Given software is durable and its market is characterized by monopolistic competition, Bils (1989) is most relevant. That paper considers a monopolist selling a good to both first-time and repeat customers and shows that in periods with many new potential customers, the monopolist lowers its markup. This pricing policy generates a time-series for prices that appears to show little response to shifts in demand.[14] Given the traditions of gift-giving over the winter-holiday in the U.S., we believe that the software consumers can be characterized as two types: regular, repeat customers and first-time, casual buyers. While regular customers buy throughout the year, casual buyers crowd

[13] "Casual Fans Are Driving Growth of Video Games," Seth Schiesel, *The New York Times*, September 11, 2007.

[14] Bils (1989) discusses how these results would extend to a version of the model with monopolistic competition.

into the market in the fourth quarter, spurred by the holiday season. According to Bils (1989), this description of consumer demand would explain the puzzling behavior of prices *not* rising over the winter holiday.

This characterization of consumers, however, implies that a representative framework would not provide a good approximation of consumer behavior. This is especially true for those segments of prepackaged software which experience large seasonal effects, such as Education, Finance, and PC Games. Rather, a more accurate way to characterize consumer's behavior would be to separate consumers into 2 types. The first type of consumer would be regular or repeat shoppers who are in the market throughout the year. The second type of consumer only shows up in the fourth-quarter of the year. Importantly, this type of heterogeneity is not nested within the Mudgett-Stone framework because both types of consumers purchase products in the fourth quarter of the year.

Our Heterogeneous price index explicitly accounts for this type of heterogeneity by constructing an index for each type of consumer. We then average the two indexes to get a single, representative price index.

Because our prepackaged software data do not have demographic information, a complication with our index is determining how to split the data between both types of consumers in the fourth quarter of each year. By construction, only the first type of consumer is shopping in quarters 1 through 3. We assume that both types of consumers pay the same price for products in the fourth quarter (i.e. there is one market-clearing price). To split out unit sales of software between consumers, we turn back to the seasonally-adjusted series created by the x-12-ARIMA software. We set type 1 consumer unit sales in the fourth quarter equal to the seasonally-adjusted unit value. The difference between the non-seasonally-adjusted and seasonally-adjusted unit values in the fourth quarter is then defined as type 2 consumer sales. In section 2.2 we labeled this difference as the "seasonal" component. In essence, we are assuming that the extra bump in units sold in the fourth quarter is attributed to type 2 consumers, or the entry of new consumers. We denote $\hat{Q}_{it}$ as the unit sales to consumer type $i = \{1, 2\}$ in period t, where $\hat{Q}_t^1 + \hat{Q}_t^2 = Q_t$. By assumption, when t is not the fourth quarter, $\hat{Q}_t^2 = 0$.

The first index measures the constant-quality price change for regular, type 1, consumers who show up throughout the year. We measure the price change faced by these consumers using a maximum-overlap matched-model approach. Let $\mathcal{L}_t^{c,1}$ be the Laspeyres quarterly price relatives of sub-category c of software products for the type 1 consumer and $\mathcal{P}_t^{c,1}$ the Paasche. The formulas we use are

$$\mathcal{L}_t^{c,1} = \frac{\sum_{j\in J_{t,t-1}^c} P_{jt}\hat{Q}_{j,t-1}^1}{\sum_{j\in J_{t,t-1}^c} P_{j,t-1}\hat{Q}_{j,t-1}^1}, \qquad \mathcal{P}_t^{c,1} = \frac{\sum_{j\in J_{t,t-1}^c} P_{jt}\hat{Q}_{jt}^1}{\sum_{j\in J_{t,t-1}^c} P_{j,t-1}\hat{Q}_{jt}^1}, \tag{3}$$

The Laspeyeres and Paasche price relatives are then chained together to produce annual price relatives,

$$\mathcal{L}_a^{c,1} = \prod_{s=1}^{4} \mathcal{L}_s^{c,1}, \qquad \mathcal{P}_a^{c,1} = \prod_{s=1}^{4} \mathcal{P}_s^{c,1} \tag{4}$$

The second index measures the constant-quality price change for the second type of consumer who only shows up in the fourth quarter. We use a year-over-year approach to measure the constant-quality price change faced by these once-a-year consumers. We construct both Laspeyres and Paasche indexes for these consumers using the following formulas,

$$\mathcal{L}_a^{c,2} = \frac{\sum_{j\in J_{t,t-4}^c} P_{jt}\hat{Q}_{j,t-4}^2}{\sum_{j\in J_{t,t-4}^c} P_{j,t-4}\hat{Q}_{j,t-4}^2}, \qquad \mathcal{P}_a^{c,2} = \frac{\sum_{j\in J_{t,t-4}^c} P_{jt}\hat{Q}_{jt}^2}{\sum_{j\in J_{t,t-4}^c} P_{j,t-4}\hat{Q}_{jt}^2}. \tag{5}$$

While t refers to quarters, the above formulas provide comparisons of prices in the fourth quarter to the previous fourth quarter, because Q_{jt}^2 is equal to zero for quarters 1 through 3 by construction.

Using annual revenue weights, we combine both the Laspeyres and Paasche indexes for each consumer type,

$$\mathcal{L}_a^{c,H} = \mathcal{L}_a^{c,1}(1-\omega_a^L) + \mathcal{L}_a^{c,2}\omega_a^L \tag{6}$$

$$\mathcal{P}_a^{c,H} = \mathcal{P}_a^{c,1}(1-\omega_a^P) + \mathcal{P}_a^{c,2}\omega_a^P \tag{7}$$

where

$$\omega_a^L = \frac{P_{a-1}\hat{Q}_{a-1}^2}{P_{a-1}Q_{a-1}}, \quad \omega_a^P = \frac{P_a\hat{Q}_a^2}{P_aQ_a}. \tag{8}$$

Finally, we take the geometric mean of the annual Laspeyres and Paasche indexes to construct an annual Fisher price index.

This approach accounts for the seasonality in the fourth quarter of the year using the same year-over-year comparison employed by the Mudgett-Stone annual index. But unlike the Mudgett-Stone index, our Heterogeneous index incorporates a majority of the data because the cost-of-living index for type 1 consumers is constructed from a quarter-to-quarter comparison.

There are a number of alternative models of consumer behavior that seek to explain why prices do not climb during periods of high demand. Waren and Barsky (1995) argue that increases in the volume of purchases per household, or “thick markets”, lower search costs and so change consumers' price elasticities. This effect however, does not apply to increases in the number of households in the market, which is surely the driving force behind the large surge in sales over the winter holiday. Further, Chevalier, Kashyap, and Rossi (2003) use retail grocery data to discount the empirical significance of the “thick market” effect in that market.[15] Nevo and Hatzitaskos (2005) advocate that changes in brand-level demand can play a significant role in explaining why high average prices do not rise during periods of peak demand. Such an explanation is consistent with the current paper's modeling of two types of households. The interpretation is that type 2 households hardly ever buy software during the year, except for during the winter holiday. While this leads to an overall increase in fourth-quarter aggregate demand, different products can face dramatically different surges in demand. In particular, if the shift in demand is towards cheaper products, then we would not expect to see large increases in average prices during periods of high demand.

As discussed previously, the seasonal effects measured by x-12-ARIMA procedure do not account for the disproportionate number of introductions that occur at the beginning

[15] Chevalier, Kashyap, and Rossi (2003) argue that loss-leader type models play a large role in explaining why grocery retail prices do not rise during periods of high demand. We believe that loss-leader models are less relevant to a durable goods market like prepackaged software.

of the winter-holiday season. Ignoring entry has an ambiguous effect on our measurement of seasonality. If entry causes more regular, type 1 shoppers to shop in the fourth quarter, then our estimates of fourth quarter sales by type 1 shoppers are too low. On the other hand, firms might time the introduction of software directed at casual, type 2 shoppers to coincide over the fourth quarter. In this case, our estimate of fourth quarter sales by type 1 shoppers is too high. We believe these entry effects are second-order and present the x-12-ARIMA estimates as good approximations of the true seasonal effects.

4 Results

We begin by presenting the details behind the construction of our Hetergenous price index. We then compare this index to the Mudgett-Stone index at both the aggregate and category level. Finally, we compare our results to the published estimates from the US Bureau of Labor Statistics (BLS).

4.1 Heterogeneous Index Details

As detailed in section 3.2, our price index directly reflects the two types of consumers in the software market. In table 8, we display some of the computations behind our price index for the software subcategory Personal Finance, within the Finance category. In the table, the chained Laspeyres is an annual price relative which is the product of the four Laspeyres price relatives from the same year (see equation 4). For example, the chained Laspeyres value in 1998, 0.988 is equal to $0.993 \times 0.986 \times 1.012 \times 0.997$. The chained Paasche is constructed similarly. These annual indexes are comparable to Laspeyres and Paasche price relatives for the type 2 consumer. The last column of table 8 lists the revenue weights needed to aggregate the type 1 and 2 Laspeyres and Paasche indexes. Following equation 6, the combined Laspeyres index value for 1998 is $0.988 \times (1 - 0.16) + 0.965 \times 0.16 = 0.984$. The Paasche indexes are combined in an equivalent way, where the revenue weights are the percentage of annual revenue generated by type 2 consumers in the current year.

The price indexes can dramatically differ for each type of consumer. To demonstrate, for four categories we construct Fisher price indexes for both types of consumers, based on

	Type 1				Type 2			
							Revenue weight (for Type 2)	
Date	Lasp	Paas	Chained Lasp	Chained Paas	Lasp	Paas	Lasp	Paas
1998 1	0.993	0.995						
1998 2	0.986	0.984						
1998 3	1.012	1.022						
1998 4	0.997	1.002	0.988	1.003	0.965	1.007	0.13	0.16
1999 1	1.000	0.999						
1999 2	0.951	1.011						
1999 3	0.960	0.978						
1999 4	0.952	0.960	0.869	0.948	0.861	0.962	0.16	0.17

Note: Lasp is Laspeyres Index and Paas is Paasche Index

Table 8: Price Indexes for Personal Finance Software

the subcategory-level Laspeyres and Paasche price indexes described above (see table 9). The differences in price declines between types of consumers can be quite large; for System Utilities software, for example, the average difference in price declines between type 1 and 2 consumers is almost 9 percent at an annual rate. Such differences highlight consumer heterogeneity and demonstrate the importance of addressing this feature of the market when constructing price indexes.

4.2 Aggregate Price Indexes

We first display the differences between our Heterogeneous price index and the Mudgett-Stone index at the category level. We then show the difference at the market-level.

Table 10 displays the category-level price indexes constructed using the Mudgett-Stone and our approach. The differences between these two types of indexes are substantial, even at this high level of aggregation. For System Utilities software, the difference between the average price relatives from the two price indexes is a hefty 8.3 percent. The categories of PC Games and Personal Productivity also have substantial differences of 2.5 and 3.3 percent respectively in the average price relatives of the two indexes. Even for categories where the average price relatives are close, significant differences crop up

Year	Operating System		PC Games		Personal Productivity		System Utilities	
	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2
1997	100	100	100	100	100	100	100	100
1998	95.9	91.3	67.6	64.3	81.4	71.8	73.2	67.0
1999	95.4	87.1	49.3	45.3	67.7	56.9	51.0	35.7
2000	90.8	86.6	36.3	32.6	55.7	42.0	50.8	29.7
2001	90.1	86.3	29.1	25.9	49.9	34.3	49.2	27.8
2002	86.3	83.9	22.7	20.2	42.2	27.7	46.1	23.4
2003	84.4	82.3	16.4	14.1	36.2	21.3	44.3	22.3
Avg Price Rel.	0.972	0.968	0.739	0.720	0.844	0.772	0.864	0.762

Note: Type 1 consumers purchase software every quarter. Type 2 consumers only purchase software in the fourth quarter. Avg Price Rel is the harmonized mean of annual price relatives

Table 9: Type-Specific Fisher Price Indexes

at the annual frequency. Business software, for example, has an average price relative of 91.2 under both the Mudgett-Stone and our method. Comparing annual price relatives however, reveals large differences in constant-quality price change. In 2002, for instance, the Mudgett-Stone method results in a price relative of 86.6, while the Heterogeneous index computes a price relative of 94.3. This translates into a 7.7 percent difference in the decline of constant-quality price as measured by the two indexes in 2002.

At the market-level, the Heterogeneous price index exhibits a slower decline relative to the Mudgett-Stone indexes (see table 11). Not surprisingly, the difference between the two indexes is smaller than those seen at the category level. Nevertheless, there is still a substantial 1.8 percent difference between the two indexes average annual price decline. In addition, as seen with the category-level indexes, for certain years the predicted constant-quality price decline is quite different across the two indexes. In 1999 for example, the Mudgett-Stone index measures a price decline of 20.9 percent, while the Heterogeneous price index reports a price decline of only 17.9 percent.

Year	Business		Education		Finance		Imaging/ Graphics	
	MS	Het.	MS	Het.	MS	Het.	MS	Het.
1998	91.6	89.7	75.3	74.6	96.5	94.5	88.2	85.0
1999	91.4	95.9	80.9	89.1	88.5	87.5	83.8	84.2
2000	89.7	91.0	80.3	80.5	91.5	90.7	80.1	80.7
2001	95.9	85.1	85.5	87.0	97.4	92.4	83.8	92.5
2002	86.6	94.3	86.2	86.1	87.7	85.6	86.2	89.7
2003	92.3	92.3	87.2	86.3	85.4	95.7	84.5	85.0
Average	91.2	91.2	82.4	83.6	90.9	90.9	84.4	86.0
Year	Operating System		PC Games		Personal Productivity		System Utilities	
	MS	Het.	MS	Het.	MS	Het.	MS	Het.
1998	94.0	95.9	64.0	67.3	81.0	81.0	71.5	73.0
1999	97.8	99.4	68.2	72.6	77.3	82.9	63.7	69.4
2000	97.0	95.3	70.4	73.5	81.7	81.6	74.8	99.0
2001	101.3	99.3	76.7	80.0	82.2	89.2	88.1	96.9
2002	97.4	95.8	77.1	77.9	84.1	84.2	88.2	93.4
2003	98.1	97.8	71.8	72.0	77.8	85.3	88.4	96.1
Average	97.6	97.2	71.1	73.6	80.6	83.9	77.9	86.2

Note: MS stands for Mudgett-Stone index, Het. stands for our Heterogeneous price index. The average price relative is the harmonized mean of annual price relatives

Table 10: Type-specific Fisher Price Indexes by Category

Year	Price Deflators		Price Indexes	
	Mudgett-Stone	Het.	Mudgett-Stone	Het.
1997	100	100	.	.
1998	78.4	78.2	0.784	0.782
1999	62.0	64.2	0.791	0.821
2000	49.8	52.9	0.802	0.824
2001	42.3	46.0	0.849	0.868
2002	35.4	39.4	0.837	0.856
2003	29.2	33.4	0.825	0.848
Average	.	.	0.814	0.832

Note: Het. stands for our heterogeneous price index. The average price relative is the harmonized mean of annual price relatives

Table 11: Prepackaged Software Fisher Price Indexes

4.3 Comparison to the BLS

Our measure of the decline in annual price change is substantially larger than that reported by the BLS for prepackaged software.[16] From 1998 to 2003 and using a Laspeyres index, the BLS reports an average annual price decline of 6.8 percent. This difference is not explained by our novel seasonal adjustment approach. If we construct a monthly Laspeyres index using all the scanner data without any seasonal adjustment, our measure of annual price decline is more than 17 percent. There are, however, a number of differences between our data sources which might explain this gap. First, our data contains 84 percent of the retail prepackaged software market while the BLS uses a random sample of products. Second, the frequency of the price data differ. Our price data are based on daily transaction data, and so reflect economic activity throughout the month. The BLS index uses price data that is gathered once a month.[17] Third, in constructing the regular Laspeyres index on the scanner data, we use the maximum-overlap method and so update the basket of goods used to construct the index every month. In contrast,

[16]We compare our index to the U.S. city average for Computer Software and Accessories series published by the BLS. The BLS only publishes a non-seasonally adjusted version of this price index.

[17]See Feenstra and Shapiro (2003) for a collection of articles concerning the promise and challenges of using scanner data to produce economic statistics.

the BLS only periodically updates the basket. Hence, new products are introduced into the basket with a delay, which causes the BLS to track an older set of products relative to the basket of goods used to construct the Heterogenous price index. It may be these older software products have slower rates of price declines, contributing to the difference in average annual price decline between the BLS consumer price index and the Laspeyres price index based on the scanner data.

5 Conclusion

In this paper, we examine the software market using a detailed and comprehensive data set. We find that most software products experience a winter-holiday sales surge, which we claim is driven by consumer heterogeneity. Specifically, we argue that casual, once-a-year shoppers enter the software market over the winter holiday season. The standard, Mudgett-Stone price index does not properly account for this type of heterogeneity. We propose a novel approach to constructing a price index that directly accounts for heterogeneity. This approach entails constructing separate price indexes for each type of consumer, and then averaging the price indexes. We then show that properly accounting for this heterogeneity is important, because the Mudgett-Stone and our price index differ on estimates of constant-quality annual price change.

More broadly, this research suggests that real consumer expenditures on software may be understated in the national accounts. This is because the BLS's consumer price index for computer software, which the BEA uses to deflate nominal personal consumption expenditures on software, measures a markedly smaller decline in software prices compared to the Heterogenous index we construct using the NPD Group scanner data. The national accounts then, may not fully reflect software's role as a driver of US productivity. Further research should be done on measuring constant-quality price change for software, to ascertain whether the BLS price index understates the decline of prepackaged software prices and, if so, by how much.

An issue touched upon, but outside the scope of this paper, is the endogenous entry of new products around the beginning of the winter-holiday season. We ignore these new goods with our matched-model approach. But the entry of new goods at the beginning of the fourth quarter is closely related to the seasonality issues discussed in this paper. Untangling these two forces, however, likely requires a formal and sophisticated model of firm and consumer behavior, a promising avenue for future research.

References

ABEL, J., E. BERNDT, AND A. WHITE (2003): "Price Indexes for Microsoft's Personal Computer Software Products," NBER working paper 9966.

AIZCORBE, A., B. BRIDGMAN, AND J. NALEWAIK (2007): "The Implications of Heterogeneous Buyers for Measuring Quality Change," Bureau of Economic Analysis mimeo.

AIZCORBE, A., AND A. COPELAND (2006): "Intermittent Purchases and Welfare-Based Price Deflators for Durable Goods," BEA working paper 2007-02.

ALTERMAN, W., W. E. DIEWERT, AND R. C. FEENSTRA (1999): *International Trade Price Indexes and Seasonal Commodities.* U.S. Department of Labor, Bureau of Labor Statistics, Washington D.C.

BILS, M. (1989): "Pricing in a Customer Market," *Quarterly Jounral of Economics*, 104(4), 699–718.

CHEVALIER, J. A., A. K. KASHYAP, AND P. E. ROSSI (2003): "Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data.," *American Economic Review*, 93(1), 15–37.

DIEWERT, W. E. (1998): "High Inflation, Seasonal Commodities and Annual Index Numbers," *Macroeconomic Dynamics*, 2, 456–471.

——— (1999): "Index Number Approaches to Seasonal Adjustment," *Macroeconomic Dynamics*, 3, 48–68.

FEENSTRA, R., AND M. SHAPIRO (2003): *Scanner Data and Price Indexes.* NBER and University of Chicago Press, 11030 South Langley Avenue, Chicago, IL 60628.

FISHER, F. M., AND Z. GRILICHES (1995): "Aggregate Price Indices, New Goods, and Generics," *Quarterly Journal of Economics*, 110, 229–244.

GANDAL, N. (1994): "Hedonic Price Indexes for Spreadsheets and an Empirical Test of Network Externalities," *RAND journal of Economics*, 24, 160–170.

GRILICHES, Z., AND I. COCKBURN (1994): "Generics and New Goods in Pharmaceutical Price Indexes," *American Economic Review*, 84, 1213–1232.

JORGENSON, D. W. (2001): "Information Technology and the U.S. Economy," *American Economic Review*, 91, 1–32.

MCCAHILL, R. J. (1997): "A Quality-Adjusted Price Index for Prepackaged Software," Master's Thesis.

NESMITH, T. D. (2007): "Rational Seasonality," FEDS working paper 2007-04.

NEVO, A., AND K. HATZITASKOS (2005): "Why Does the Average Price of tuna Fall During Lent?," Northwestern University manuscript.

OLINER, S., AND D. SICHEL (1994): "Computers and Output Growth Revisited: How Big is the Puzzle?," *Brookings Papers on Economic Activity*, 2, 273–317.

PARKER, R., AND B. GRIMM (2000): "Recognition of Business and Government Expenditures for Software as Investment: Methodology and Quantitative Impacts: 1959-98," Bureau of Economic Analysis methodological paper.

PRUD'HOMME, M., D. SANGA, AND K. YU (2005): "A Computer Software Price Index Using Scanner Data," *Canadian Journal of Economics*, 38, 999–1017.

WAREN, E., AND R. BARSKY (1995): "The Timing and Magnitude of Retail Store Markdowns: Evidence from Weekends and Holidays," *Quarterly Journal of Economics*, 110, 321–352.