

# Seasonality and Prepackaged Software Price Indexes \*

Adam Copeland<sup>†</sup>

January 2010

## Abstract

This paper measures constant-quality price change for prepackaged software in the US using detailed and comprehensive scanner data. Because there is a large sales surge over the winter-holiday, it is important to account for seasonal variation. Using a novel approach to constructing a seasonally-adjusted cost-of-living price index that explicitly accounts for consumer heterogeneity, I find that from 1997 to 2003 constant-quality software prices declined at an average 16.8 percent at an annual rate. As a point of comparison, the Bureau of Labor Statistics reports average annual price declines of only 6.8 percent for prepackaged software.

**Keywords:** seasonal adjustment, software prices, heterogeneity, price indexes

---

\*I would like to thank Dennis Fixler and Marshall Reinsdorf for helpful comments. I also thank Andrew Miller for substantive help as a research assistant. The views expressed herein are my own and not necessarily those of the Federal Reserve Bank of New York or the Federal Reserve System.

<sup>†</sup>Money and Payments Studies, Federal Reserve Bank of New York, 33 Liberty St., New York, NY 10045; e-mail: [adam.copeland@gmail.com](mailto: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.<sup>1</sup> 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.<sup>2</sup> 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.<sup>3</sup> 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.

Using scanner data from the NPD Group to construct a maximum-overlap Fisher price index, I find the average price decline for prepackaged software is 17.2 percent at an annual rate. In the data, however, I find the majority of software products experience a significant boost in unit sales over November, December and January, the winter-holiday season. Given there is seasonality in the data, Alterman, Diewert, and Feenstra (1999) claim the Mudgett-Stone index (i.e. a year-over-year approach) should be considered the best measure of annual price change (page 48). The underlying framework of the Mudgett-Stone approach, however, is a representative-consumer framework.<sup>4</sup> In the prepackaged software market, the winter-holiday variation seems to be driven by consumer heterogeneity. Given the particular correlations of prices and sales over these months, I 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, a better way to account for seasonality in the prepackaged software market is to construct two indexes: one index for casual, once-a-year consumers, and another for regular, year-round shoppers. These two indexes are then averaged using

---

<sup>1</sup>Jorgenson (2001) emphasizes that information gaps remain about understanding software pricing trends.

<sup>2</sup>See Parker and Grimm (2000) for details on the high rate of growth of prepackaged software.

<sup>3</sup>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.

<sup>4</sup>See Diewert (1998), Diewert (1999), and Nesmith (2007).

revenue weights, resulting in what I label the Heterogeneous index. For the software market as a whole, this price index measures the average constant-quality price decline to be 16.8 percent at an annual rate. Using the standard Mudgett-Stone approach, constant-quality annual price change averages 18.6 percent. Hence, properly accounting for the underlying consumer heterogeneity lowers the measured fall in prices over the sample by 1.4 percent at an annual rate, not an inconsequential difference. Naturally, there are larger differences at lower levels of aggregation. For the software categories PC Games, Personal Productivity and System Utilities, the differences in estimates of constant-quality annual price change between the Heterogeneous index and the Mudgett-Stone index are 2.5, 3.3, and 8.3 percent, respectively.

From these results, I draw two main conclusions. First, whether or not you account for seasonal variation, prepackaged software prices are declining at annual rate of more than 16 percent between 1997 and 2004. This is more than two times as large as the 7 percent average annual rate of decline reported by the Bureau of Labor Statistics (BLS) for this same period. The difference in measured price change between the BLS and this paper's price indexes is almost entirely related to differences in data, as opposed to the method of index construction.

Second, while there is little doubt that accounting for seasonal variation is important, the results above demonstrate the importance of understanding the underlying causes in the variation. If seasonal variation is not driven by consumer heterogeneity, then the Mudgett-Stone price index is likely the best available approach. But if consumer heterogeneity is driving the change in price and quantity sold over the year, then the Mudgett-Stone price index can be misleading. I argue that the Heterogeneous index presented here is a more appropriate measure of price change.

The results from the Heterogeneous index, however, should be taken with a note of caution. A crucial step in constructing this index is determining the revenue flows to each type of consumer. Ideally, there would be data describing the different types of consumers in the market. Since the consumer types are not observed, I use an *ad hoc* approach to divide the revenue over the winter-holiday between the two consumer types: the casual, once-a-year shoppers and the regular, year-round shoppers. Reassuringly, the results are robust to alternative approaches to dividing up the revenue.

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, this paper uses detailed, industry-wide scanner data, as opposed to a small subset of products.<sup>5</sup> Hence, the price indexes are representative of price changes throughout the industry, and so generates robust measures of constant-quality price change. Second, I 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, the inclusion of seasonality adjustment is, in itself, a modest improvement on previous empirical work. I also introduce a new variation on the existing set of empirical methods used to account for seasonality when constructing a price index. Unlike previous methods, this approach explicitly accounts for consumer heterogeneity, the driving force behind prepackaged software's seasonal fluctuations.

This paper is closest to Prud'homme, Sanga, and Yu (2005), who construct a maximum-overlap Fisher price index for prepackaged software using samples of transaction level data from the Canadian market. This paper's results on the average annual decline in constant-quality price differ substantial from Prud'homme, Sanga, and Yu (2005), however. They report an average annual price decline of 7.9 percent. Using their price index methodology on the NPD scanner data, I find an average annual price decline of 17.2 percent. Hence differences in the data must be driving the result. The NPD scanner data, while not the universe, is more comprehensive than their sample. Further, the data are detailed enough information that different versions of the same software product are observed. Besides the difference in empirical results, the current paper differs from Prud'homme, Sanga, and Yu (2005) because it accounts for the seasonality in the data.

By accounting for consumer heterogeneity, this paper builds upon the 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

---

<sup>5</sup>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.

the introduction of new goods. This paper differs from these works because of its focus on seasonality.

## 2 Data

In this section I describe the data on prepackaged software. I then measure the seasonality within the data and discuss two data quality issues.

### 2.1 Description

The prepackaged software industry data come 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.<sup>6</sup> The data are point-of-sale transaction data (i.e. scanner data) that are sent to the NPD Group from participating outlets. The data 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 me 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 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.

---

<sup>6</sup>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

The prepackaged software market is turbulent, with many products entering and exiting the market over the 7 year sample. The mean lifespan for an average software product is 22.0 months. This statistic is skewed by a few extremely long-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 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, which are all precisely estimated, are graphed in Figure 1. The 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

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)

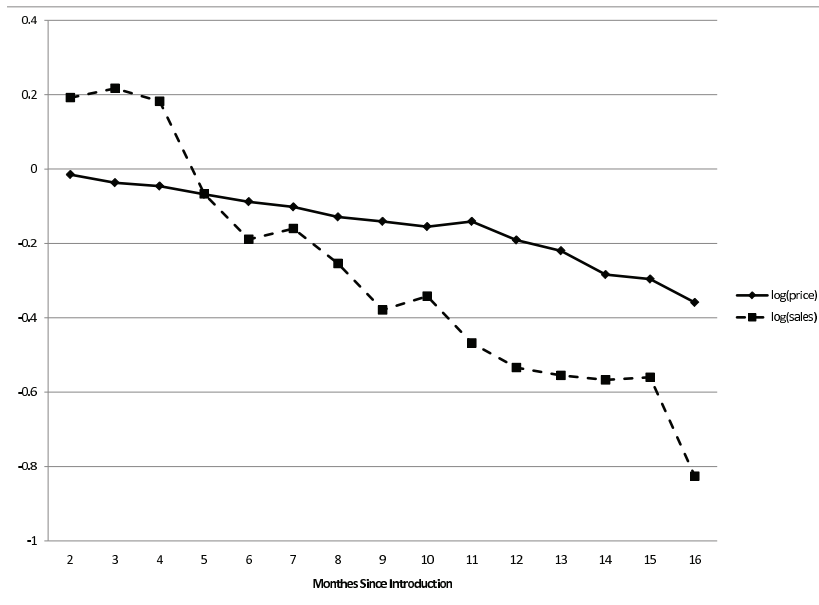


Figure 1: Price and Sales Contours over the Product Cycle

Notes: The figure plots coefficients estimated from regressing the log of price and the log of sales on product cycle dummies, with product-level fixed effects.

unit sales.

## 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.<sup>7</sup> 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. Figure 2 charts 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 me to group November, December, and January together. To better examine the winter-holiday seasonal phenomenon, I define this group of months as the fourth quarter, and then define quarters 1-3 accordingly. With the data aggregated to the quarterly frequency, the sample encompasses 28 quarters, from quarter 1 of 1997 to quarter 4 of 2003.

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, I am not able to definitively determine if there is a winter holiday seasonal effect at the product level.

To identify seasonal effects I consider the data at a higher level of aggregation—the subcategory level.<sup>8</sup> Several approaches were used to determine when a subcategory of software experiences a winter-holiday seasonal affect. The preferred approach, and the one presented here, uses x-12-ARIMA, a seasonal adjustment software packaged used and maintained by the U.S. Census Bureau.<sup>9</sup>

---

<sup>7</sup>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>).

<sup>8</sup>An example of a subcategory is “Foreign Language” software within the “Education” category.

<sup>9</sup>See <http://www.census.gov/srd/www/x12a/> for more information.

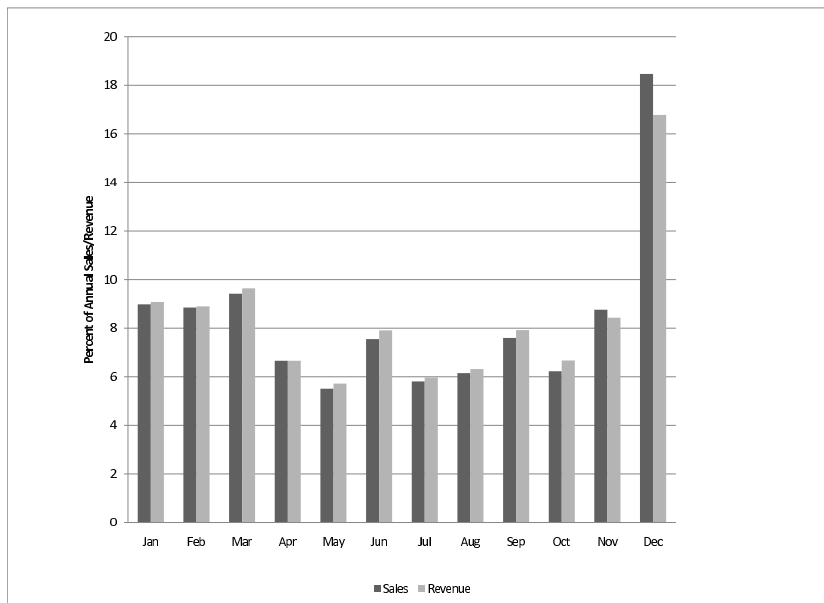


Figure 2: Percent of Units Sold and Revenue Generated by Month

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

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 3: Seasonal Magnitude of Fourth-Quarter Revenue

The x-12-ARIMA algorithm produces a seasonally-adjusted series of units sold for each subcategory of software. For each subcategory and for each year, I 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, I 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. 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 3). Business software has the smallest seasonal component, in-line with priors that work-related software is not much affected by the winter-holiday season. Education, Finance, and PC Games software have the largest winter-holiday effect. For these categories, more

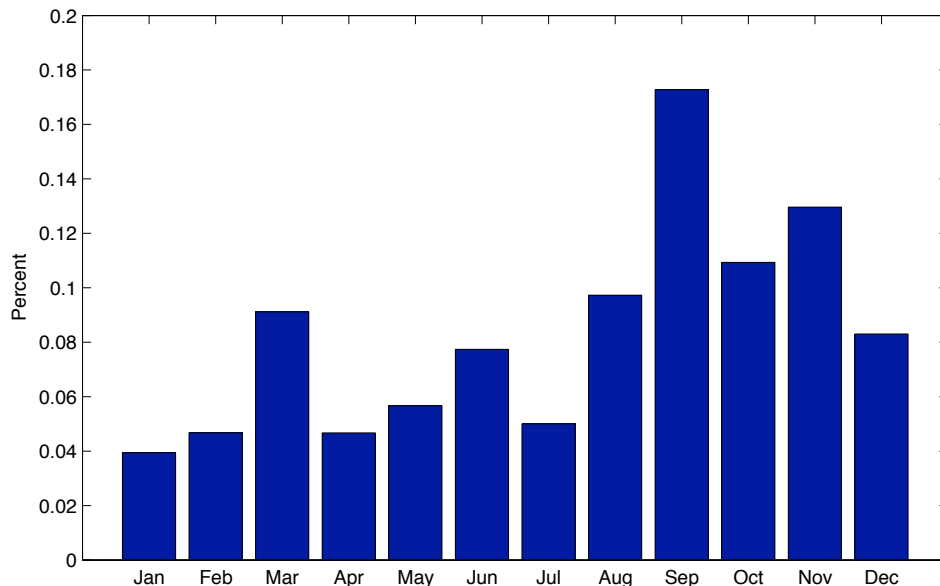


Figure 3: Product Introductions by Month

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, using the lifetime sales weighted average of product introductions by quarter, a disproportionate amount of products are introduced around the beginning of the fourth quarter (see Figure 3). In part because I 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,

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 4: Frequency Distribution of the Price Ratio of Adjoining Quarters’ Prices

something beyond the scope of this paper. Instead, consider the seasonal factors produced by the x-12-ARIMA program to be good, first-cut approximations.

### 2.3 Data Quality

Before discussing how to construct price indexes that adjust for the winter-holiday seasonality, two quality issues in the data are addressed. 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 the analysis. As shown in the last column of Table 1, these observations account for a negligible share of the total units sold.<sup>10</sup>

Second, there are implausible quarterly price changes. As shown in Table 4, the price ratio of adjoining quarters’ prices has some extreme values. Categorizing which quarterly price changes are the result of measurement error can be difficult to discern. I take a conservative approach and drop the observations that are in the top and bottom one percent of the quarterly price ratio distribution. This translates into dropping quarterly price ratios below 0.13 and above 5.07.

## 3 Methods

In this section I describe and compare three different price index approaches: the maximum-overlap Fisher, the Mudgett-Stone and the Heterogeneous price indexes.

---

<sup>10</sup>In addition to the removing the suppressed observations, I 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.

### 3.1 Fisher price index

The maximum overlap Fisher price index is a standard approach to measuring constant quality price change. This is the approach used by Prud'homme, Sanga, and Yu (2005) in their work measuring price change for prepackaged software. The Fisher is an average of a Laspeyres and Paasche price index. I use a moving basket of goods and so the reference period when considering quarter  $t$  is quarter  $t-1$ . Denote  $\mathcal{L}_t^{c,\text{std}}$  as the Laspeyres quarterly price relatives for software in group  $c$  and  $\mathcal{P}_t^{c,\text{std}}$  the Paasche. Letting  $J_{t,s}^c$  be the set of products belonging to the software group  $c$  available in both quarter  $t$  and  $s$ , I compute the Laspeyres and Paasche price relatives using the following standard formulas,

$$\mathcal{L}_t^{c,\text{std}} = \frac{\sum_{j \in J_{t,t-1}^c} P_{jt} Q_{j,t-1}}{\sum_{j \in J_{t,t-1}^c} P_{j,t-1} Q_{j,t-1}}, \quad \mathcal{P}_t^{c,\text{std}} = \frac{\sum_{j \in J_{t,t-1}^c} P_{jt} Q_{jt}}{\sum_{j \in J_{t,t-1}^c} P_{j,t-1} Q_{jt}}, \quad (1)$$

I then compute a quarterly 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 the empirical analysis, I compute a market-level price index.

### 3.2 Mudgett-Stone price index

I construct a Mudgett-Stone annual index following the description in Diewert (1998). Each product is defined 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. The base year is set to be the previous year when constructing price relatives. Denote  $\mathcal{L}_t^{c,\text{MS}}$  as the Laspeyres Mudgett-Stone quarterly price relatives for software in group  $c$  and  $\mathcal{P}_t^{c,\text{MS}}$  the Paasche. The Laspeyres and Paasche price relatives are computed using the following formulas,

$$\mathcal{L}_t^{c,\text{MS}} = \frac{\sum_{j \in J_{t,t-4}^c} P_{jt} Q_{j,t-4}}{\sum_{j \in J_{t,t-4}^c} P_{j,t-4} Q_{j,t-4}}, \quad \mathcal{P}_t^{c,\text{MS}} = \frac{\sum_{j \in J_{t,t-4}^c} P_{jt} Q_{jt}}{\sum_{j \in J_{t,t-4}^c} P_{j,t-4} Q_{jt}}, \quad (2)$$

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

$$\mathcal{L}_a^{c,MS} = \sum_{s=1}^4 w_s^a \mathcal{L}_s^{c,MS}, \quad \mathcal{P}_a^{c,MS} = \sum_{s=1}^4 w_s^a \mathcal{P}_s^{c,MS}. \quad (3)$$

The four quarters summed over in the above equation correspond with a calendar year  $a$  and  $w_s^a$  is the share of annual revenue for calendar year  $a$  earned in quarter  $s$ . Finally, I 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 the empirical analysis, I compute both a market-level and category-level price indexes.

While I 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.

### 3.3 The Heterogeneous Price Index

In this subsection, I first lay out the motivation behind the construction of the Heterogeneous price index. I then detail the formulas behind its construction. Lastly, I describe an important issue with regard to practically implementing this price index.

#### 3.3.1 Motivation

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 the software example, this translates into a representative consumer having different tastes in each quarter of the year.

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 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 Figure 2). But

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 5: November to December Price Relatives

if new households show up in the 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.<sup>11</sup>

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, there is at most a slight uptick in price. Using the original monthly data, a maximum-overlap Fisher price index computes an average price increase of only 0.09 percent in December (see Table 5). 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.<sup>12</sup> Given the traditions of gift-giving over the winter-holiday in the U.S., software consumers can be categorized into two types: regular, repeat customers and first-time, casual buyers. While regular customers buy throughout the year, casual buyers crowd 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

<sup>11</sup>“Casual Fans Are Driving Growth of Video Games,” Seth Schiesel, *The New York Times*, September 11, 2007.

<sup>12</sup>Bils (1989) discusses how these results would extend to a version of the model with monopolistic competition.

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.

### 3.3.2 Method

To explicitly account for this type of consumer heterogeneity, I propose constructing separate indexes for each type of consumer. These indexes are then averaged together to produce the Heterogeneous price index. The Implementation subsection that follows details how the two types of consumers can be identified in the data. For now, take as given there are data on units sold for each type of consumer. Both consumers see the same prices in the market, but purchase different amounts. Denote  $\hat{Q}_{it}$  as the unit sales to consumer type  $i = \{1, 2\}$  in period  $t$ , where  $\hat{Q}_t^1 + \hat{Q}_t^2 = Q_t$ . For prepacked software, we are only concerned about the winter-holiday seasonal variation. By assumption then, 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. I 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 construction of  $\mathcal{L}_t^{c,1}$  and  $\mathcal{P}_t^{c,1}$  follow the same formulas detailed in equation 1, where  $\hat{Q}_{js}$  is substituted in place of  $Q_{js}$ . The Laspeyres and Paasche price relatives are then chained together to produce annual price relatives,

$$\mathcal{L}_a^{c,1} = \prod_{s=1}^4 \mathcal{L}_s^{c,1}, \quad \mathcal{P}_a^{c,1} = \prod_{s=1}^4 \mathcal{P}_s^{c,1}. \quad (4)$$

The second index measures the constant-quality price change for the second type of consumer who only shows up in the fourth quarter. I use a year-over-year approach to

measure the constant-quality price change faced by these once-a-year consumers, similar to the formulas outlined in the Mudgett-Stone section above. Because the casual type 2 consumers only show up in the fourth quarter of each year, I construct annual Laspeyres and Paasche indexes for these consumers by looking at the change in prices in the fourth quarter to the previous fourth quarter,

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

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

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

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

where

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

Finally, I take the geometric mean of the annual Laspeyres and Paasche indexes to construct an annual Fisher price index. As evidenced from its construction, this Heterogeneous price index is, roughly speaking, an average of a maximum-overlap Fisher and Mudgett-Stone price index.

### 3.3.3 Implementation

Because the prepackaged software data do not have demographic information, a complication with the Heterogeneous 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. 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, I use an ad hoc approach. I turn back to the seasonally-adjusted series created by the x-12-ARIMA software and 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 I labeled this difference as the “seasonal” component. In essence, I am assuming that the extra bump in units sold in the fourth quarter is attributed to type 2 consumers.

Using this approach to divide total units sold into sales to type 1 and type 2 consumers, I can use the formulas detailed above to construct the Heterogeneous price index. In the appendix there are more details on the construction of the Heterogeneous price index as well as some of the intermediate figures used in the construction of the final index.

In most cases, there will not be data available which will indicate sales by type of consumer. Hence, some sort of assumption will need to be made to divide total units sold among the types of consumers. My approach of assigning the seasonal component of sales to the casual type 2 consumers has at least two advantages. First, this division of sales accords well with the underlying premise behind the winter holiday seasonality. The data indicate that the burst of sales over the fourth quarter is driven by once-a-year shoppers, and so assigning the extra bump in sales for the quarter to the type 2 consumers seems reasonable. While this division of sales is ad hoc, I argue it provides a good first-cut approximation to the true division of sales between regular and once-a-year shoppers. Second, the approach of relying on x-12-ARIMA is transparent and easy to replicate. Reassuring, the paper’s results are robust to different approaches to computing the seasonal and non-seasonal components of unit sales.

## 4 Results

I present the results from the three price indexes at the aggregate level. To explore the different seasonal-adjustment methods, the Mudgett-Stone and Heterogeneous price indexes are compared.

Year	Price Deflators			Price Indexes		
	Fisher	Mudgett-Stone	Het.	Fisher	Mudgett-Stone	Het.
1997	100	100	100	.	.	.
1998	77.7	78.4	78.2	0.777	0.784	0.782
1999	64.5	62.0	64.2	0.829	0.791	0.821
2000	52.1	49.8	52.9	0.808	0.802	0.824
2001	45.8	42.3	46.0	0.879	0.849	0.868
2002	42.2	35.4	39.4	0.924	0.837	0.856
2003	32.6	29.2	33.4	0.772	0.825	0.848
Average	.	.	.	0.828	0.814	0.832

Table 6: Prepackaged Software Fisher Price Indexes

Notes: Het. stands for the Heterogeneous price index and Fisher stands for a maximum-overlap Fisher price index. The average price relative is the harmonized mean of annual price relatives.

## 4.1 Aggregate Results

As shown in Table 6, all three price indexes measure rapid declines in prepackaged software prices from 1997 to 2003. The maximum-overlap Fisher price index records that constant-quality price declined 67.4 percent from 1997 to 2003, where the average annual price decline was 17.2 percent. The Mudgett-Stone price index measures an even faster decline in prices, with an average annual price decline of 18.6 percent. The Heterogeneous price index records the slowest price decline, with an average annual price decline of 16.8 percent over the sample.

Further differences are revealed by looking at individual years. The Fisher and Heterogeneous price indexes track one another closely from 1997 to 2001. But in 2002 and 2003, these two indexes report fairly significant different price declines, resulting in a divergence in their respective price deflators. The Mudgett-Stone price index systematically records larger price declines compared to the other two indexes, which is worrisome from a measurement perspective.

As a group, these results point to steady and rapid declines in prepackaged software prices. Hence the first main result is that constant-quality prepackaged software prices from 1997 to 2003 declined at an average annual rate of roughly 17 percent. This is

a significant result in itself, since the BLS reports much more timid decline in price.<sup>13</sup> From 1998 to 2003 and using a Laspeyres index, the BLS reports an average annual price decline of 6.8 percent. A monthly Laspeyres index constructed using the scanner data without any seasonal adjustment delivers an average annual price decline of more than 17 percent at an annual rate.

The differences between this paper's result and the BLS reported figures most likely come from the data. There are a number of differences between the data sources which might explain the gap in measured price declines. First, the NPD scanner 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. The NPD 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.<sup>14</sup>

Differences in the base period is another possible explanation behind the measure of constant-quality price decline. I use the maximum-overlap method and so update the basket of goods used to construct the index every month. In contrast, 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 this paper's price indexes. 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. By construction, however, the Mudgett-Stone's year-over-year approach considers an older set of products and yet measures average annual price declines of 19 percent. This suggests that differences in updating the basket of goods does not explain much of the difference between this paper's measure of price declines and those published by the BLS.

---

<sup>13</sup>The BLS index is the U.S. city average for Computer Software and Accessories series. The BLS only publishes a non-seasonally adjusted version of this price index.

<sup>14</sup>See Feenstra and Shapiro (2003) for a collection of articles concerning the promise and challenges of using scanner data to produce economic statistics.

## 4.2 Differences in Seasonal Adjustment

Given the three price indexes, the natural question is which index provides the most accurate measure of constant-quality price change. I argue the Heterogeneous price index provides the best measure of constant-quality price change for two main reasons. First, from the data, it is clear there is substantial seasonal variation in prepackaged software unit sales over the winter holidays (see Section 2.2). Hence, some adjustment for seasonal variation is necessary. Second, the source of the winter holiday sales spike is likely due to consumer heterogeneity (see Section 3). The Heterogeneous price index directly accounts for consumer heterogeneity, unlike the Mudgett-Stone price index.

While at the aggregate level there do not seem to be substantial differences between the three price indexes, large differences do appear at the disaggregated level. Because there is already a large body of empirical work which examines differences between seasonally-adjusted and non-seasonally-adjusted price indexes, here I focus on the differences between the two seasonally-adjusted price indexes: the Mudgett-Stone and Heterogeneous price indexes.<sup>15</sup> Large differences between these two indexes appear at the category level, as shown in Table 7. 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 at the annual frequency. Business software, for example, has an average price relative of 91.2 under both the Mudgett-Stone and Heterogeneous methods. 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.

As laid out in Section 3, there are fair number of differences in the construction of the Mudgett-Stone and Heterogeneous price indexes which contribute to the discrepancies in

---

<sup>15</sup>For a recent and thorough overview of accounting for seasonality in price indexes, see Diewert, Balk, Fixler, Fox, and Nakamura (2009)

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

Table 7: Price Indexes by Category

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

the measured constant-quality price change. Recall that the Heterogeneous price index is an average of a quarter-to-quarter and a year-over-year index. The year-over-year index measures the price change for type 2 consumers, and tends to track the Mudgett-Stone index, which is also a year-over-year index. Consequently, differences between the Heterogeneous and Mudgett-Stone index are mainly driven by the fact that the Heterogeneous price index is also a function of the quarter-to-quarter index for type 1 consumers. For intuition on why a quarter-to-quarter price index may yield a different measure of constant quality price change compared to a year-over-year price index, refer to Alterman, Diewert, and Feenstra (1999) and Diewert, Balk, Fixler, Fox, and Nakamura (2009).

These differences between the Mudgett-Stone and Heterogeneous price indexes highlight the importance of choosing the right method to account for seasonal variation. Using the Mudgett-Stone price index can present a different, and I argue inaccurate, picture of constant-quality price decline for various categories of prepackaged software.

## 5 Conclusion

In this paper, I examine the prepackaged software market using a detailed and comprehensive data set. Using three different price indexes, I find that constant-quality prices decline at an average annual rate of more than 16 percent in the sample. This is a substantially greater fall in price than reported by the BLS. Second, I consider the seasonal variation in the prepackaged software market. I find that most software products experience a winter-holiday sales surge, which I claim is driven by consumer heterogeneity. Specifically, the data suggests that casual, once-a-year shoppers enter the software market over the winter holiday season. Significantly, the standard, Mudgett-Stone price index does not properly account for this type of heterogeneity. Hence, I 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. This Heterogeneous price index measures an average annual price decline of 16.8 percent for prepackaged software. Properly accounting for this heterogeneity is important, because the Mudgett-Stone and Heterogeneous price index

produce different 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 Heterogeneous index constructed 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. I ignore these new goods with the 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.

		Type 1				Type 2			
Date		Lasp	Paas	Chained	Chained	Lasp	Paas	Revenue weight (for Type 2)	
				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

Table 8: Price Indexes for Personal Finance Software

Notes: Lasp is Laspeyres Index and Paas is Paasche Index

## A Appendix

Table 8 displays some of the computations behind the Heterogeneous 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 I construct Fisher price indexes for both types of consumers, based on 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

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

Table 9: Type-Specific Fisher Price Indexes

Notes: 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.

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.

## 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 Journal of Economics*, 104(4), 699–718.
- 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.
- DIEWERT, W. E., B. M. BALK, D. FIXLER, K. J. FOX, AND A. O. NAKAMURA (2009): *Price and Productivity Measurement: Volume 2 – Seasonality*. Trafford Press.
- 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.
- JORGENSEN, D. W. (2001): “Information Technology and the U.S. Economy,” *American Economic Review*, 91, 1–32.
- MCCAHERILL, 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.
- 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.