A New Approach for Quality Adjusting PPI Microprocessors

Steven Sawyer and Alvin So
Bureau of Labor Statistics
2 Massachusetts Ave, NE
Washington DC 20212
202-691-7845
202-691-7893

(Preliminary draft, please do not circulate)

A New Approach for Quality Adjusting PPI Microprocessors

Steven Sawyer   Alvin So

Abstract

June 2017

The Producer Price Index for microprocessors has shown a slower rate of decline since 2009 than in previous years. Between 2000 and 2009 the index fell, on average, 33.66 percent per year. Since 2009 the index has only fallen, on average, 6.28 percent per year. This shift in deflation is a result of a change in pricing behavior by a major producer. The current index is calculated with a matched model methodology. While this methodology might have been effective in the past when microprocessors prices dropped steadily from product introduction to end-of-life, a revised methodology must be explored in the current environment of static prices to accurately account for the changes in quality.

With a data set that includes processors from Q1 2015 to Q4 2016 and benchmark statistics from PassMark along with key characteristics for processors, we estimate seven two-period overlapping quarterly models. To select specifications, we use statistical learning methods of selecting the model with the minimum Bayesian information criterion (BIC) score and the model with the lowest mean squared error (MSE) calculated using 10-fold cross validation. The minimum BIC models yielded an average annualized price decline of 3.68 percent while the models selected with the minimum MSE yielded an average annualized price decline of 3.20 percent. Constructing counterfactual indexes using the models, we estimated that for the period of August 2015 to November 2016, the PPI Semiconductors Primary products fell 4.6 percent with the minimum BIC model and 4.3 percent with the minimum MSE model. The official index fell 3.6 percent.

Introducing a hedonic model for microprocessors would be a novel approach for the PPI. It could potentially serve as a template for hedonic quality adjustment for other industries that see rapid technological change. It would also be the first use of a *time dummy* hedonic model and the first application of statistical learning methods in the PPI.

Introduction

 The PPI is one of the nation's principal Federal economic indicators that measures the average change over time of the selling prices received by domestic producers of goods and services. This family of indexes is made up of approximately 10,000 PPIs for individual products and groups of products that are published each month, one of which is the PPI for semiconductor and related device manufacturing[1]. One subcomponent of semiconductor and related device manufacturing is the index for microprocessors[2].

To ensure the PPI measures only a "pure" price change based on market factors, it must exclude any price change or portion of a price change that is due to a change in the characteristics of a product. A change in the characteristics of a product is also called a quality change. The challenge that the PPI faces is how to split out this "pure" price change from the quality change. For a product, such as a piece of luggage, the quality change from vinyl to leather trim can be easily measured and accounted for[3] by removing the value of the switch to leather from the price change. This is known as quality adjustment. However, quality adjusting for microprocessor characteristics is more difficult because of their technological complexity. This challenge in properly accounting for the quality change has led to debate about whether the PPI for microprocessors is biased because it is not properly accounting for quality change.

This debate has informed our efforts to develop new quality adjustment methods for microprocessors. Some of the quality adjustment methods proposed by other researchers have been helpful in guiding the general direction of our approach to quality adjustment which we then refine and develop to suit the needs of the PPI microprocessors index. These methods involve using a type of statistical model for quality adjustment that has never been deployed in a PPI index before. In addition, we use "statistical learning" techniques in the development of our quality adjustment method which is also a first for the PPI.

In this article, we will (i) discuss the technical reasons for the debate over the possible magnitude of the bias of the PPI microprocessors index; (ii) reexamine some of the results from a study claiming

---

[1] http://www.bls.gov/ppi/ppifaq.htm
[2] The PPI for Microprocessors (including microcontrollers) has not been published since March 2015, since it does not meet publication standards for the PPI.
[3] http://www.bls.gov/ppi/ppifaq.htm

substantial bias in the microprocessors index; and (iii) present a quality adjustment method that both uses new methods and is developed with new methods.

Background

The PPI for microprocessors is a matched model index. Theoretically, a matched model index tracks price changes from period to period for the same set of products. However, producers typically stop producing products after a certain amount of time and introduce new products with different characteristics; this creates a problem of price comparison between different products. Prior to 2009, when new microprocessors were introduced, the prices of existing microprocessors would decline. In this case, a matched model index still captures price changes due to factors such as technological improvements. The market is effectively valuing the technological change.

The index that includes microprocessors has shown a pronounced reduction in its rate of decline, primarily due to a change in the pricing behavior of a heavily weighted producer in the index. After 2009, when new products were introduced, the prices of existing products would usually remain unchanged. Between 2000 and 2009, the annual rate of deflation of the PPI for microprocessors was 33.66 percent per year. Between 2009 and 2014, it was 6.28 percent per year. Some have interpreted slower rate of decline as a slowdown of technological innovation in the semiconductor industry and as a signal for the end of Moore's Law.

Moore's Law states that the number of transistors on a chip doubles every two years, which is significant because it has driven "feature-size reduction (*scaling*) that leads to better performance and cost reduction".[4] This combination of decreasing cost per transistor[5] and increasing performance was responsible for the large declines in the microprocessors index before 2009. However, at very small feature sizes, which companies began to manufacture in the 2005-2007 timeframe, power usage for a given area of a processor begins to increase.[6] This increase in power usage per area makes it difficult to continue increasing performance at the same rate as before. While fabrication technology can address

---

[4] Peter Van Zant, *Microchip Fabrication: A Practical Guide to Semiconductor Processing*, Sixth Edition, (New York: McGraw Hill, 2014), p. 394.
[5] Kenneth Flamm, "Has Moore's Law Been Repealed? An Economist's Perspective", Computing in Science and Engineering, March/April 2017, p. 33. Flamm's Table 1 shows for Intel, the cost per transistor has continued to decline even as wafer processing costs have increased.
[6] For a more in depth description of this phenomenon see Rambus, *Understanding Dennard Scaling*, https://www.rambus.com/blogs/understanding-dennard-scaling-2/?nabe=4857318206603264:1,6583178454368256:0

the challenge of increasing power usage per area,[7] it is clear the feature-size reduction no longer delivers the same decreases in cost and increases in performance[8] as it did previously.

While the slowdown of Moore's Law is debatable, there is no question that microprocessors have changed significantly since 2009, both from a technological and pricing viewpoint. These changes have led to discussions about the PPI match-model methodology for the microprocessors index and whether it is showing a slower rate of quality increase than a hedonic approach would. Researchers have used regression modeling methods, or hedonic modeling methods, to generate estimates of quality adjusted price[9] declines. A hedonic model uses a regression to estimate the price of a good by breaking it down to its component parts. Price is the dependent variable and the independent variables are the various characteristics of the product. A recent example of a hedonic regression based critique of the PPI microprocessors index is a paper by Byrne, Oliner, and Sichel, henceforth referred to as BOS[10]. The BOS paper uses a hedonic regression to construct a quality adjusted index of price change for Intel desktop microprocessors, which shows a 42 percent price decline per year from 2009-2013[11]. This suggests that the decline of under 10 percent for the PPI over this period may be quite biased.

Data and Methods

To understand the 42 percent price change, we attempt to reconstruct the model presented in the BOS paper. We use Intel processors and prices to recreate a comparable data set to do this analysis[12]. BOS uses a time dummy hedonic model in their analysis of microprocessor prices. A time dummy hedonic model utilizes a dataset that consists of a panel of two time periods. The BOS paper uses two adjacent overlapping years for each of their panels. For instance, two of their panels are 2009-2010 and 2010-2011. Their specification uses a time dummy and one other regressor, the log of a performance benchmark (SPEC speed) with log price as the dependent variable. The coefficient on the time dummy variable shows the price change between the two time periods in the panel that is not explained by the other independent variables and it can be used to calculate an annual inflation rate. The benchmark

---

[7] Peter Van Zant, *Microchip Fabrication: A Practical Guide to Semiconductor Processing*, Sixth Edition, (New York: McGraw Hill, 2014), p. 434.
[8] Performance implicitly includes power usage.
[9] A quality adjusted price accounts for technological change in a product.
[10] David M. Byrne, Stephen D. Oliner, and Daniel E. Sichel (2017). "How Fast are Semiconductor Prices Falling?", The Review of Income and Wealth, 2017.
[11] Ibid, p.18.
[12] Any reference to Intel in this paper is for this analysis only.

(SPEC speed) is a measure of performance of a microprocessor by calculating how long it takes a microprocessor to run a suite of software[13].

In putting together their dataset, BOS identified two possible problems. The first possible problem is that Intel's "posted prices do not represent true transactions prices because Intel offers progressively larger discounts to selected purchasers as models age"[14]. The second possible problem is that Intel's prices are unweighted which "would put too much weight on price observations for which there were few transactions", especially for older processors.[15] They contend that a newly introduced microprocessor will generate much more revenue than a microprocessor that is several years old, but an unweighted dataset will give them equal importance. Their solution for both of these possible problems was to "use the first four quarterly prices for each model (or fewer prices if the model is in the market for less than a year), and refer to this as the 'early-price' hedonic regression".[16] They aggregate these quarterly prices into yearly panels. By doing this, they will exclude any microprocessor introduced before the first year in a two year panel. We believe that this truncation is too abrupt because the PPI tracks the entire production of companies, not just newly introduced products. We implement a different interval which is explained in a later section.

After recreating the BOS results, we explore the use of additional characteristics. The data we use in our following research is derived from Intel's publicly available price sheets. Detail on processor characteristics were obtained from Intel's ARK website. In addition to SPEC "speed" (and SPEC "rate", which we discuss in the next section), we look at:

- Cores – the number of physical processing units, which read and execute program instructions.
- Threads – Intel's Hyper-threading technology allows each core to run more than one thread, essentially doubling the amount of programmed instructions/commands that each core can process.
- Base Frequency – measures the processor's speed in clock cycles per second. The base frequency is the fastest a processor can run without manual overclocking[17].

---

[13] See Appendix A for more information on the SPEC benchmarks.
[14] David M. Byrne, Stephen D. Oliner, and Daniel E. Sichel (2017). "How Fast are Semiconductor Prices Falling?", The Review of Income and Wealth, 2017, p. 11.
[15] Ibid, p. 11.
[16] Ibid, p. 12.
[17] https://www.pugetsystems.com/labs/articles/Is-CPU-Base-Frequency-Still-a-Relevant-Spec-512/

- Turbo Frequency – the highest possible speed a processor can run with manual overclocking. Intel's Turbo Boost Technology allows processor cores to run faster than their base frequency if they are operating below power, current, and temperature specification limits[18].

- Cache – Central Processing Unit (CPU) caches are small pools of memory that store information the CPU is most likely to need next. The information loaded into cache depends on sophisticated algorithms and certain assumptions about programming code. The goal of the cache system is to ensure that the CPU has the next bit of data it will need already loaded into cache by the time it goes looking for it[19].

- TDP – Thermal Design Power, is the maximum amount of heat generated by the processor that the cooling system is designed to dissipate[20]. Typically, when comparing processors with the same architecture, the one with lower TDP will consume less power, while the one with higher TDP will have a better performance.

- Graphics – Microprocessors can have an integrated graphics processing unit (GPU)

All regressions in this paper have log price as the dependent variable and include a time dummy variable.

Criticisms of PPI Microprocessors Index

The 42-percent decline per year results from the BOS are the result of using only a single regressor, the log of the SPEC speed performance benchmark. When additional characteristics are added to the regression, the rate of price decline becomes much smaller.

We are able to replicate the BOS result using publicly available Intel data from 2009 to 2015, and applying their specification on "early prices". We obtain an average annual price decline of 46.50 percent with an average adjusted R^2 of 0.7518. This result is comparable to the result obtained by BOS. It serves as a check on data compatibility and is shown in Table 1 below.

---

[18] http://www.intel.com/content/www/us/en/architecture-and-technology/turbo-boost/turbo-boost-technology.html

[19] https://www.extremetech.com/extreme/188776-how-l1-and-l2-cpu-caches-work-and-why-theyre-an-essential-part-of-modern-chips

[20] https://ark.intel.com/ - Thermal Design Power

Table 1: log Performance ("speed") Model on "Early Prices"

	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15
Year Dummy	-0.6936*	-1.1810*	-0.4715*	-0.3011*	-0.3705*	-0.7348*
	(0.1368)	(0.0926)	(0.0554)	(0.0361)	(0.0462)	(0.1364)
log Performance ("speed")	3.0977*	3.3130*	2.9087*	3.0155*	3.4819*	3.7546*
	(0.2962)	(0.1949)	(0.1132)	(0.1127)	(0.1651)	(0.2686)
Observations (Year 1, Year 2)	68 (18, 50)	143 (50, 93)	166 (93,73)	150 (73, 77)	141 (77, 64)	93 (64, 29)
Adjusted R^2	0.6645	0.6267	0.6772	0.8604	0.8479	0.8342

*Significant at the 5-percent level

The SPEC speed benchmark only measures the performance of a single core of a microprocessor. Since microprocessors have multiple cores, a better measure of performance is the log of the SPEC "rate" benchmark that measures *multi*-core performance. The difference between the speed and rate tests is clear. For example, consider the following three processors, all with speed scores in the low-to mid-70s:

Table 2: SPEC Speed and SPEC Rate Benchmark Comparison

Processor Model	i3-6100	i7-4790K	i7-5960X
SPEC Speed benchmark	73	71	72
SPEC Rate benchmark	132	183	328
Price	117	339	999

As we can see in Table 2, the SPEC speed benchmarks are similar for these three processors, while the SPEC rate score better reflects the increase in performance with the increase in price from the i3-6100 to the i7-5960X. The disparity in rate scores is due mostly to different numbers of physical cores, which are two, four, and eight respectively. The cache is another factor, increasing at even greater rate than the core count: three, eight, and 20 megabytes. Intel, of course, charges for these features, as the large price spread illustrates. Running a regression with only the log of the SPEC rate benchmark drops the average annual price decline to 28.91 percent with an average adjusted R^2 of 0.8891. The results are shown in Table 3 below.

Table 3: log Performance ("rate") Model on "Early Prices"

	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15
Year Dummy	-0.2548*	-0.7709*	-0.2965*	-0.1298*	-0.1902*	-0.4051*
	(0.1031)	(0.0515)	(0.0368)	(0.0267)	(0.0279)	(0.0411)
log Performance ("rate")	1.6022*	1.7785*	1.6729*	1.5878*	1.7195*	1.8554*
	(0.1068)	(0.0776)	(0.0595)	(0.0416)	(0.0381)	(0.0439)
Observations (Year 1, Year 2)	68 (18, 50)	143 (50, 93)	166 (93,73)	150 (73, 77)	141 (77, 64)	93 (64, 29)
Adjusted R^2	0.7899	0.8430	0.8667	0.9324	0.9433	0.9593

*Significant at the 5-percent level

However, there are reasons to not solely rely on SPEC benchmarks. SPEC does not register the graphical improvements to many Intel desktop processors in the last decade. In more recent years, Intel has been integrating a GPU onto their microprocessors. For this reason, a hedonic regression examining Intel CPUs should include controls for graphics. The performance of a GPU can be gauged by the number of execution units it has. The regressor we use for graphics is the log of the number of execution units; if it does not have an onboard GPU, a zero will be assigned.

Additional controls can also be added to distinguish processors with equivalent benchmark performances that differ in operating frequencies, power consumption, thread counts, and cache. The nine-regressor model seen in Table 4 includes; log Cores, log Threads, log Base Frequency, log Turbo Frequency, log Cache per Core, log TDP, log Graphical statistic along with the single and multi-core SPEC benchmarks. With these specifications, the average annual price decline slips to 14.56 percent and the average adjusted R^2 rises to 0.9775.

Table 4: Full Model on "Early Prices"

	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15
Year Dummy	-0.1378*	-0.1971*	-0.1563*	0.0193	-0.0991*	-0.3028*
	(0.0602)	(0.0824)	(0.0216)	(0.0184)	(0.0165)	(0.1089)
log Performance ("speed")	-3.0198*	-1.8465*	-2.2244*	-2.9583*	-1.3128*	0.0673
	(0.7128)	(0.3215)	(0.2688)	(0.3172)	(0.4113)	(0.6721)
log Performance ("rate")	5.1822*	1.7636*	1.7798*	2.7927*	1.3274*	0.9937*
	(0.7433)	(0.3694)	(0.3105)	(0.2657)	(0.3613)	(0.3409)
log Cores	-1.3143*	0.5202*	0.1979	-0.2068	0.7375*	0.7218*
	(0.3243)	(0.1745)	(0.1471)	(0.1321)	(0.1978)	(0.1894)
log Threads	-0.6364*	0.0134	0.1606*	0.0308	0.2559*	0.1997*
	(0.2168)	(0.0673)	(0.0627)	(0.0570)	(0.0737)	(0.0642)
log Base Frequency	5.1094*	1.8304*	0.8376*	0.8175*	1.9080*	1.1522*
	(0.5128)	(0.2583)	(0.1737)	(0.1611)	(0.2716)	(0.3806)
log Turbo Frequency	-0.0671	0.9584*	1.3078*	1.2467*	0.3412	-0.2960
	(0.5156)	(0.1338)	(0.1225)	(0.1381)	(0.2174)	(0.6015)
log (Cache/Cores)	-1.5895*	0.3799*	0.4650*	0.4033*	0.3521*	0.4113*
	(0.2362)	(0.1102)	(0.0637)	(0.0917)	(0.0599)	(0.0979)
log TDP	-0.7807*	-0.9251*	-0.5072*	-0.5062*	-0.6867*	-0.4931*
	(0.1718)	(0.0719)	(0.0435)	(0.0552)	(0.0524)	(0.0676)
log Graphics	-0.1869*	-0.1529*	-0.1448*	-0.0405*	-0.0827*	-0.0367
	(0.0300)	(0.0268)	(0.0159)	(0.0165)	(0.0120)	(0.0205)
Observations (Year 1, Year 2)	68 (18, 50)	143 (50, 93)	166 (93,73)	150 (73, 77)	141 (77, 64)	93 (64, 29)
Adjusted R^2	0.9657	0.9587	0.9836	0.9812	0.9895	0.9864

*Significant at the 5-percent level

We perform a standard F-test on the two nested models, the restricted model with just the time dummy and SPEC speed and the unrestricted model which includes all our additional characteristics:

$$\log price = \beta_0 + \beta_1 \, Year \, Dummy + \beta_2 \log Performance \, ("speed") + \varepsilon$$

$$\log price = \beta_0 + \beta_1 \, Year \, Dummy + \beta_2 \log Performance \, ("speed") + \beta_3 \log Performance \, ("rate") \\ + \beta_4 \log cores + \beta_5 \log Threads + \beta_6 \log Base \, Frequency + \beta_7 \log Turbo \, Frequency \\ + \beta_8 \log cache/cores + \beta_9 \log TDP + \beta_{10} \log Graphics + \varepsilon$$

For every time panel 2009-10 through 2014-15 we test the null hypothesis that our restricted parameters are equal to zero against the alternative hypothesis that at least one parameter does not equal zero.

$$H_0: \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$$

$$H_A: at \, least \, one \, \beta \neq 0$$

The F-statistic measures the reduction of the residual sum-of-squares per additional parameter in the unrestricted model.

$$F_0 = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n - (k + 1))}$$

Where RSS_R and RSS_{UR} are the residual sum-of-squares of the restricted and unrestricted models respectively. The variable n is the number of observations, k is the number of independent variables and q is the number of restrictions. Blackwell notes that "if the residuals are larger in the restricted model, then our F-statistic will also be large. When the residuals are large we know that the fit of the regression is worse; therefore, the F-statistic is larger when the restrictions reduce the fit of the model. If these variables had no predictive power, then removing them should not affect the residuals"[21].

Table 5: F-statistic for all time periods

	2009-10	2010-11	2011-2012	2012-2013	2013-14	2014-15
F-Statistic	72.31	141.61	382.63	119.34	234.26	126.37

From the above results the null hypothesis is rejected in all the time periods even at the one percent level. This result confirms our regression outputs in Table 4 where most of our additional characteristics are significant at the five percent level from 2009 to 2015.

Given the statistical significance of variables excluded by BOS, it seems that their model may be subject to omitted variable bias. The BOS estimate of inflation is not robust to changes in specification. Indeed, inflation varies greatly with specification changes. This raises the question of how different model specifications perform with respect to one another and with respect to the time dummy coefficient that represents the annual price decline for microprocessors. To provide perspective on this question, we create every possible subset of the nine regressors, which gives 512 different linear regression specifications. We then take these 512 specifications and estimate them for all six of the two-year panels. Finally, for every one of the 512 specifications, we calculate its average annual rate of inflation and average adjusted R^2 across the six two-year panels. We get a mean inflation rate of -18.74 percent. As mentioned earlier, the model with only the log of SPEC speed as a regressor has an inflation rate of negative 42.

---

[21] Blackwell. "Multiple Hypothesis Testing: The F-test", p. 4.

Figure 1: Six-year Average Annual Inflation from all 512 models for Overlapping 2-year Time Dummy Models, 2009-2015
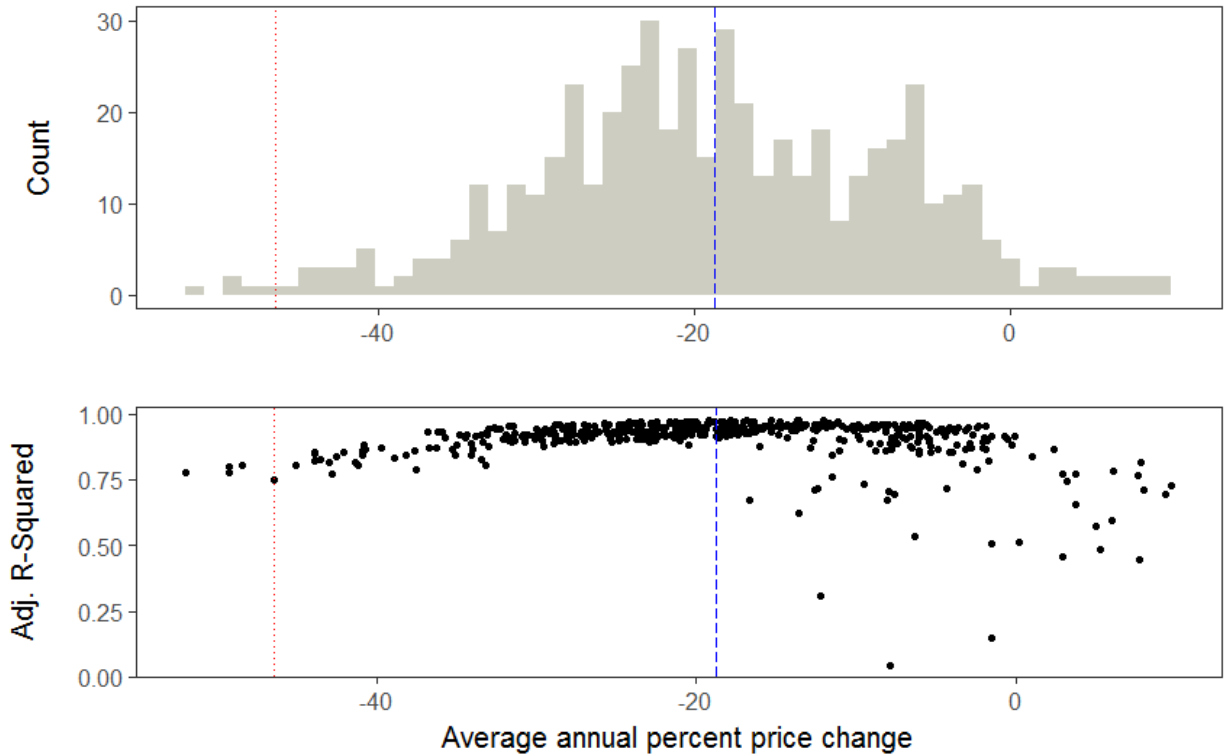
In Figure 1, the blue dashed line shows the average price change for all 512 models. The dotted red line shows the price change using the BOS specification, which is a clearly larger rate of decline than the vast majority of possible models. In fact, only 4 other models, less than 1 percent, show a sharper rate of annual price decline than the BOS result. This lends evidence that the BOS specification does not include variables that are important and that it may be an outlier. The bottom section of Figure 1 arrays the 512 models by average adjusted R^2 and average annual price decline. Most cluster above the average adjusted R^2 of 0.7518 of the BOS specification. More significantly is the range of differences in average annual price declines relative to the BOS result, which appears to be an outlier. This is even more clearly seen in the top section histogram of Figure 1.

The annual deflation results obtained from our data are in line with the anecdotal evidence reported by industry observers. For instance, an article in *Anandtech*, detailing 6th generation processors, states that there was around a 25-percent increase, *in total*, in performance between 2011 and 2015[22].

From our findings and the results of real world performance testing from well-known technology publications, we find little empirical support that the 42-percent annual price decline from BOS is an accurate representation of the microprocessor industry. However, the BOS criticism of the PPI microprocessors index has sparked an evaluation of PPI's semiconductor index methodology and an examination of alternative quality adjustment methods that PPI could implement.

Constructing a Hedonic Model for PPI Microprocessors

PPI has studied quality adjusting microprocessors in the past, most notably in a paper by Michael Holdway[23], but none of the methods examined were feasible for implementation. Despite our findings, we think the BOS paper lays out a useful framework for developing a hedonic model appropriate for use in the PPI.

Before we look at estimating a model for use in the PPI microprocessors index, we need to put together a different dataset than the one used by BOS. First, we only look at microprocessors from 2015 and 2016. Since we are interested in seeing what the impact of a hedonic model would be on the microprocessors index currently (2017), we see little value in building datasets that contain observations from more than two years ago. We are also unable to calculate counterfactual indexes for years prior to 2015, which is an important method for evaluating models that is shown later in this section. We estimate models with overlapping two-period quarterly data. PPI indexes are published monthly which means models need to be updated more often than annually.

As stated earlier, BOS use "early prices" to address two possible problems. The first possible problem, posted prices not equaling transaction prices for older microprocessors, has been examined by Flamm (2017). Flamm found that for the retail microprocessors market, which is 20 percent[24] of Intel's microprocessors sales by volume, there was "no evidence to support the suggestion that there was some structural change after 2006 in the relationship between observed Intel list price and observed

---

[22] Cutress, Ian. "The Intel 6th Gen Skylake: Core i7-6700K and i5-6600K Tested". *Anandtech*, August 5, 2015. http://www.anandtech.com/show/9483/intel-skylake-review-6700k-6600k-ddr4-ddr3-ipc-6th-generation

[23] Michael Holdway, "An Alternative Methodology: Valuing Quality Change for Microprocessors in the PPI", Bureau of Labor Statistics, https://www.bea.gov/papers/pdf/mpuvqa.pdf.

[24] Kenneth Flamm, "Has Moore's Law Been Repealed? Empirical Analysis of Innovation in Semiconductors", p. 24.

retail market prices"[25]. This gives proof that *at least* some of Intel's sales for older microprocessors are being accurately represented by its posted price list, while solid evidence that Intel's posted prices for older microprocessors are unrepresentative is lacking. On that basis, we do not think that truncating the sample from the full dataset as the BOS "early prices" does is appropriate for the PPI.

The second problem that BOS cite, that of the sales of microprocessors falling over time, we fully recognize, but we do not think that "early prices" are the best way to address it. For example, in a regression for 2012-2013, a processor introduced in the fourth quarter of 2011 would not be included even though it would likely have sales similar to processors introduced in the first quarter of 2012. We think this cutoff is too abrupt. Intel will continue to sell a processor even after the introduction of a newer, more technologically advanced version of it is introduced. As mentioned earlier, the PPI tracks the entire production of companies, not just newly introduced products. Furthermore, when Intel introduces a microprocessor, shipments typically start off at a low level and increase several months before peaking. Figure 2 below shows this pattern.

Figure 2: Intel Microprocessor Shipments[26]

---
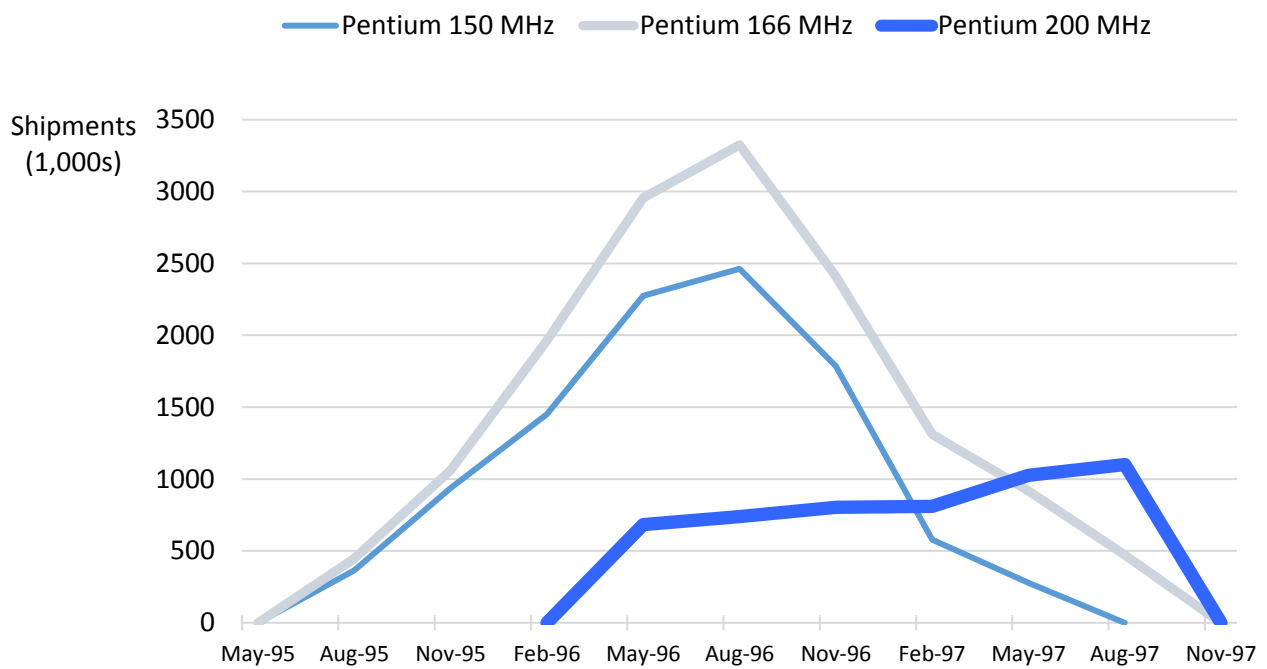
[25] Ibid, p. 26.
[26] Thanks to Ana Aizcorbe for this data which was first used in Aizcorbe, Corrado, and Doms (2000)

A better approach would be to include all microprocessors Intel is selling that were introduced within a certain interval from a given quarter. This approach should better be able to capture the pattern of microprocessor shipments shown in Figure 2 and yield a more representative dataset. For some microprocessors geared towards businesses, Intel guarantees that these microprocessors will be available and supported for 15 months[27]. It is reasonable to assume that if Intel is actively supporting a processor, it should be still be selling at a significant volume. We use this 15-month interval when building our dataset. For example, for the first quarter of 2015, we include all microprocessors that were introduced from the fourth quarter of 2013 or later that are still being sold in the first quarter of 2015.

Instead of the SPEC performance benchmarks, we use the PassMark[28] CPU performance benchmark. While SPEC is a well-known benchmarking tool, it is primarily used by manufacturers and vendors to performance test their server or workstation computer systems. Many power-efficient versions of the i3, i5, and i7 chipsets do not have SPEC benchmark scores; low-end desktop processors such as the Pentium and Celeron chipsets are also missing scores from SPEC. While compiling the quarterly 2015 and 2016 data we noticed that only eight processors did not have a PassMark score, while the SPEC score was missing for 72 processors. PassMark also is able to capture the performance gain from multicore microprocessors, as Table 6 shows below.

Table 6: SPEC Speed, SPEC Rate, and PassMark Benchmark Comparison

| Processor Model	i3-6100	i7-4790K	i7-5960X
SPEC Speed benchmark	73	71	72
SPEC Rate benchmark	132	183	328
PassMark benchmark	5454	11184	15972
Price	117	339	999

In the first section of this article, we questioned the BOS model because it had a different inflation rate than most of the models with much higher adjusted R^2s. A first step in evaluating our quarterly data is reproducing Figure 1 with it. Please note that we constrain the log PassMark variable to be in every model we estimate with our quarterly data because we know for certain that microprocessor performance is increasing over time. We also think that the PassMark benchmark may be able to

---

[27] Intel. "Intel Stable Image Platform Program [Intel SIPP]. http://www.intel.com/content/www/us/en/computer-upgrades/pc-upgrades/sipp-intel-stable-image-platform-program.html
[28] See Appendix A for more information on PassMark

account for improvements to microprocessors that are not associated with changes in any of the characteristic variables. Microprocessors are complex products with many attributes so it is not possible to estimate a model that includes their every aspect. Since the PassMark benchmark shows total microprocessor performance, it may be able to show quality improvements to microprocessors caused by changes to the characteristics we are unable to include in our models.

Figure 3: Annual Inflation and Average Adj. R^2 from All 128[29] Models for Overlapping 2-Quarter Time Dummy Models, 2015-2016
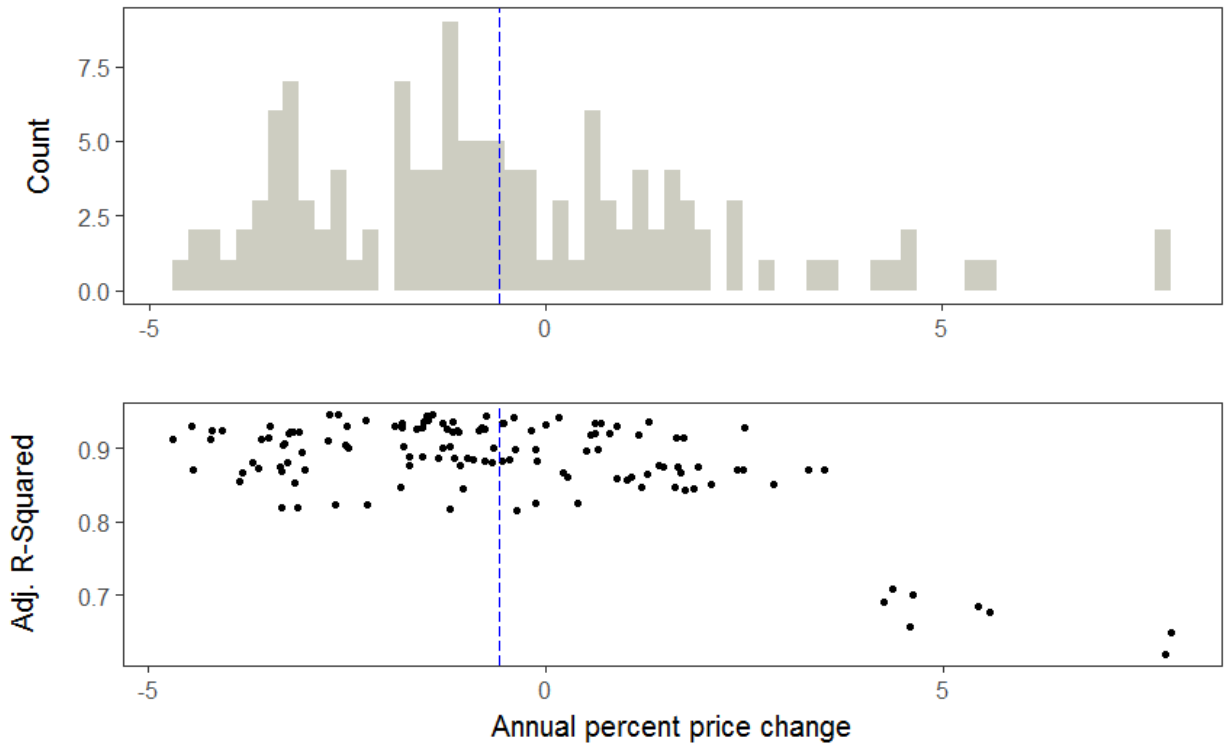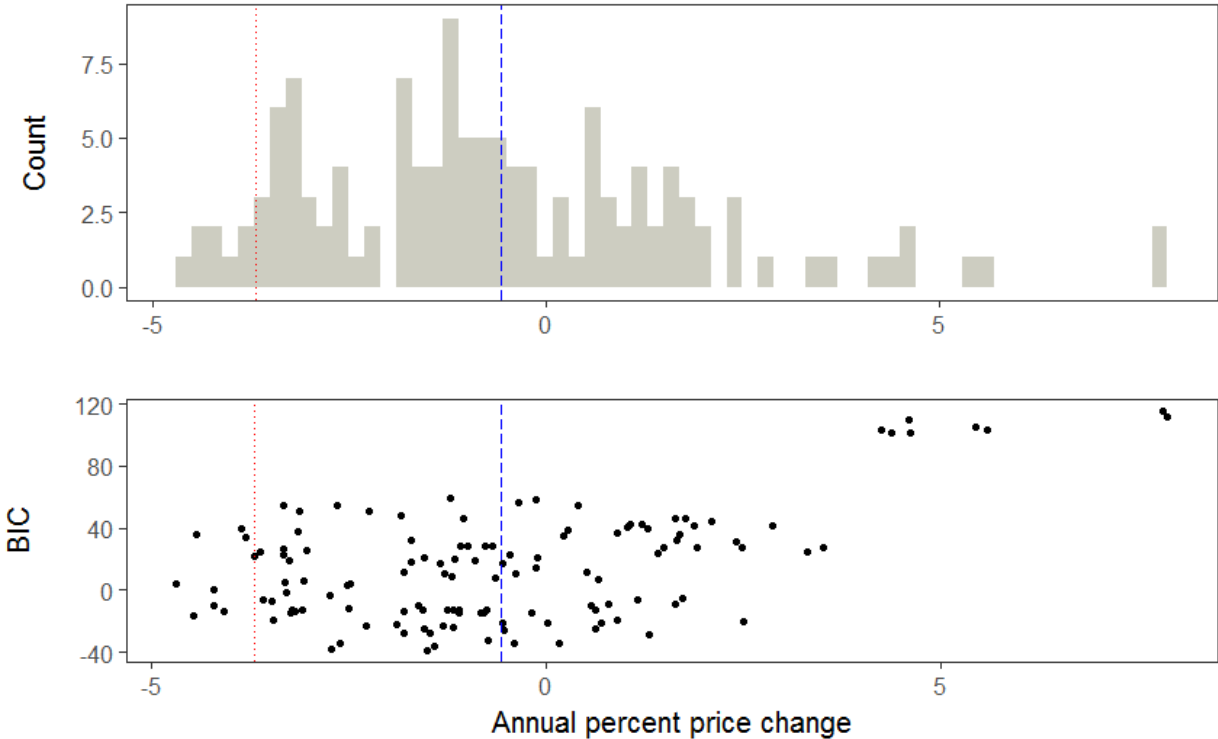

The blue dashed line in Figure 3 shows the average annual inflation of -0.58 percent.

Another method to evaluate models is to use an information criteria. We use the Bayesian information criterion (BIC) value. Information criteria, in general, can be defined as "choosing the model with the

---

[29] With our quarterly data, there are only 128 possible models since the PassMark benchmark replaces the two SPEC benchmarks and it is included in every model. This means that subset models of only the seven characteristics are being estimated.

best penalized log-likelihood".[30] For our purposes BIC is useful because it is a widely-known measure of model performance and it can be used for variable selection for both nested and non-nested models. The main risk of using BIC for model selection is underfitting[31], which would be selecting fewer characteristics than exist in the "true" model. When we calculate the BIC for all possible models for every two-quarter overlapping panel, and then select the model with the minimum BIC for every period, we get an average annual price decline of 3.68 percent for 2015-16.

Figure 4: Annual Inflation and Average BIC from All 128 Models for Overlapping 2-Quarter Time Dummy Models, 2015-2016

The average annual inflation rate is again shown by the blue dashed line. The dotted red line shows the annual inflation rate that is calculated from selecting the minimum BIC model from every two-quarter panel. Please note that the dotted red line in Figure 4 does not intersect the lowest BIC point because

---

[30] Dziak, John J., Donna L. Coffman, Stephanie T. Lanza, and Runze Li. "Sensitivity and specificity of information criteria". The Methodology Center, The Pennsylvania State University, https://methodology.psu.edu/media/techreports/12-119.pdf, p. 2.
[31] Ibid, 23.

the lowest BIC model changes from period to period.   Table 7 shows data resulting from the selected
models.

Table 7: Minimum BIC Selected Models Q12015 - Q42016

	15Q1-15Q2	15Q2-15Q3	15Q3-15Q4	15Q4-16Q1	16Q1-16Q2	16Q2-16Q3	16Q3-16Q4
Quarter Dummy	0.0059	0.0235	0.0084	-0.0556		-0.0005	-0.0474
	(0.0257)	(0.0273)	(0.0330)	(0.0352)		(0.0386)	(0.0373)
log PassMark	-0.2386	0.4295*	0.2986	0.1016		-0.1573	-0.2331
	(0.2489)	(0.1957)	(0.1859)	(0.1149)		(0.1437)	(0.1604)
log Cores	0.9721*	1.0864*	0.6907*	0.3721*		0.7310*	0.8177*
	(0.1019)	(0.1246)	(0.0706)	(0.1301)		(0.1448)	(0.1496)
log Threads	0.5248*	0.2830*	0.5136*	0.6222*		0.6368*	0.6438*
	(0.0930)	(0.0810)	(0.0979)	(0.0678)		(0.0648)	(0.0634)
log Base Frequency		0.7982*		-1.3323*		-0.9775*	-0.7959*
		(0.1556)		(0.3558)		(0.3276)	(0.3234)
log Turbo Frequency	1.2687*		0.6776*	2.2484*		2.2187*	2.0403*
	(0.2445)		(0.2696)	(0.4814)		(0.4100)	(0.3747)
log (Cache/Cores)	0.7386*	0.4538*				0.2829*	0.4079*
	(0.1338)	(0.1453)				(0.1299)	(0.1317)
log TDP	-0.2599*	-0.5620*	-0.2187*				
	(0.0761)	(0.0779)	(0.0674)				
log Graphics							
Observations	97	104	97	84		78	71
Adjusted R^2	0.9652	0.9546	0.9408	0.9245		0.9339	0.9550
BIC	-81.1238	-66.6948	-46.2510	-29.8551		-11.3687	-16.5014

*Significant at the 5 percent level

The first and second quarters of 2016 are composed of the same microprocessors with the same prices;
consequently, no model is estimated for this panel. Across quarters, there is a fair amount of
consistency in the variables chosen and their magnitude. Log Threads, log Turbo Frequency, and log
Cores are selected for every model. Interestingly, log Graphics is never selected. This implies that
graphics is not playing an important role in differentiating the price between microprocessors even
though Intel uses up to half the silicon for the integrated GPU on many microprocessors. Lastly, log TDP
was included and significant in half the models. For a given level of performance, a processor that uses
less power should sell at a premium compared to one that uses more. The negative sign on log TDP
supports this assumption.

Using BIC to select a model specification has several advantages for the PPI. Since it is a widely-known technique for model evaluation, it makes the process of model selection easy to explain to users of our data. Users of PPI indexes will have more confidence and trust in our data when our methods are clear and transparent. Using BIC for model specification selection is also quick and easy to implement. This is important if a model is going to be used in the PPI. A PPI industry analyst (IA) would need to re-estimate the model quarterly. IAs already have heavy workloads with the various activities needed to calculate PPI indexes every month. For it to be practical to implement a hedonic model operationally in the PPI, the development of the model needs to be efficient.

BIC is a traditional technique for model evaluation. Since there is no definitive method of specification selection, we will also use a technique that comes from the field of statistical learning. If we are able to estimate similar rates of inflation from specifications chosen using different techniques, this consistency would be a sign that we are selecting a reasonable model. The following from *Elements of Statistical Learning* gives a basic overview behind statistical learning methods:

> The *generalization* performance of a learning method relates to its prediction capability on independent test data. Assessment of this performance is extremely important in practice, since it guides the choice of learning method or model, and gives us a measure of quality of the ultimately chosen model[32].

It is important to emphasize that although statistical learning methods look at the predictive power of the model on the dependent variable, they are evaluating the overall quality of the model, including the independent variables. We are using these methods to perform model selection by "estimating the performance of different models in order to choose the best one"[33].

Statistical learning evaluates a model by its predictive ability on data not used to estimate the model. In other words, out of sample prediction. Models that are over-fitted may have a high adjusted R^2, but they have poor out of sample predictive performance. One way to test for over-fitting is to use a procedure known as validation. This involves splitting a dataset, estimating the model on one part of the dataset (the training set) and then predicting the dependent variable on the other part of the dataset (the test set) using the coefficients from the model estimated on the training set. The test set

---

[32] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, Springer, http://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf, p.219.
[33] Ibid, p. 222.

predictions of the dependent variable are then compared to the actual test set values of the dependent variable to give the prediction errors; the prediction errors are squared and then averaged which gives the mean squared error (MSE). The lower the MSE is, the better the predictive performance of the model.

The disadvantage of validation is that part of the dataset is not used to estimate the model, which increases the variability of the estimated model parameters. An alternative option is to use cross-validation. With k-fold cross-validation, the data set is split into k parts. Each of the k parts is held out in turn (the test set) and the model is estimated on the remaining data (the training set). Then, just as with validation, the MSE is calculated by making predictions on the test set by using the model estimated on the training set. This procedure yields k MSEs which are averaged together to produce an overall MSE.

Performing k-fold cross-validation on every possible model is computationally intensive. To reduce the number of models to evaluate, we can first prescreen the dataset. It is important to note that the prescreening is only performed on the training set, and not the full dataset. With our dataset, we have seven possible regressors to choose from (the time dummy and log PassMark are always included). We start by calculating the residual sum of squares (RSS) for every model that contains just one regressor (plus the time dummy and log PassMark) and the one-regressor model with the lowest RSS makes it through the prescreening. We then repeat this procedure for every model that contains two regressors (plus the time dummy and log PassMark), and the two-regressor model with the lowest RSS makes it through the prescreening. We continue this process, increasing the number of regressors by one each time, until we have prescreened a model that contains seven regressors (plus the time dummy and log PassMark). At the end of the prescreening procedure, we have seven models, which contain from one to seven regressors. We then use cross validation to calculate the MSE for each of the seven models[34]. Since we used 10-fold cross validation, we repeat this procedure (including prescreening) 10 times.

---

[34] We used code from *An Introduction to Statistical Learning* (*ISL)* to implement this procedure. See Appendix B for more detail.

When the dataset is split into 10 folds, it is done so randomly. To average the random variation from splitting the dataset, we perform the above procedure 500 times and take the average MSE for each of the seven models. The best model is the one with the minimum MSE[35].

Table 8: Minimum MSE Selected Models Q12015 - Q42016

	15Q1-15Q2	15Q2-15Q3	15Q3-15Q4	15Q4-16Q1	16Q1-16Q2	16Q2-16Q3	16Q3-16Q4
Quarter Dummy	0.0059	0.0260	0.0084	-0.0556		0.00268	-0.0443
	(0.0257)	(0.0269)	(0.0330)	(0.0352)		(0.0379)	(0.0374)
log PassMark	-0.2386	0.2787*	0.2986	0.1016		-0.2707	-0.3118
	(0.2489)	(0.2458)	(0.1859)	(0.1149)		(0.1637)	(0.1590)
log Cores	0.9721*	1.0173*	0.6907*	0.3721*		0.8597*	0.9135*
	(0.1019)	(0.1130)	(0.0706)	(0.1301)		(0.1576)	(0.1536)
log Threads	0.5248*	0.3536*	0.5136*	0.6222*		0.6745*	0.6762*
	(0.0930)	0.0951	(0.0979)	(0.0678)		(0.0706)	(0.0710)
log Base Frequency		0.4691		-1.3323*		-0.6968	-0.5853
		(0.2673)		(0.3558)		(0.3729)	(0.3668)
log Turbo Frequency	1.2687*	0.5461	0.6776*	2.2484*		1.9030*	1.7867*
	(0.2445)	(0.4002)	(0.2696)	(0.4814)		(0.4112)	(0.4089)
log (Cache/Cores)	0.7386*	0.4337*				0.3498*	0.4521*
	(0.1338)	(0.1323)				(0.1226)	(0.1207)
log TDP	-0.2599*	-0.4771*	-0.2187*				
	(0.0761)	(0.0795)	(0.0674)				
log Graphics						0.0431*	0.0299
						(0.0339)	(0.0302)
Observations	97	104	97	84		78	71
Adjusted R^2	0.9652	0.9554	0.9408	0.9245		0.9340	0.9547

*Significant at the 5 percent level

The average annual price decline for 2015-16 is 3.20 percent. For most of the two-quarter panels, log Cores, log Turbo Frequency and log (Cache/Cores) were selected and significant. This finding is consistent across the two years of data. Log Graphics was included in two of the models and significant in one of them.

Table 9 summarizes the inflation rates and the corresponding indexes for the models chosen using the minimum BIC method and the minimum MSE method.

---

[35] To be more precise, the best model is selected by determining the model with the smallest number of variables whose standard error is within range of the lowest MSE value. This is called the "*one-standard-error-rule*". See page 214 of *ISL* for more detail.

Table 9: Inflation Rates

	15Q1-15Q2	15Q2-15Q3	15Q3-15Q4	15Q4-16Q1	16Q1-16Q2	16Q2-16Q3	16Q3-16Q4
Min BIC inflation	0.0059	0.0238	0.0084	-0.0541	0.0000	-0.0005	-0.0463
Min MSE inflation	0.0059	0.0263	0.0084	-0.0541	0.0000	0.0027	-0.0433
Min BIC index	100.59	102.98	103.85	98.24	98.24	98.19	93.64
Min MSE index	100.59	103.24	104.11	98.48	98.48	98.74	94.46

Indexes start at 100 in 14Q4-15Q1

The minimum BIC index shows a greater decline than the minimum MSE index. In Figure 5, we compare

the actual PPI semiconductors index for primary products, 334413P, with counterfactual hedonic

indexes that adjust desktop microprocessor items with the inflation rates in Table 9.

Figure 5: Counterfactual Microprocessors Index Comparison
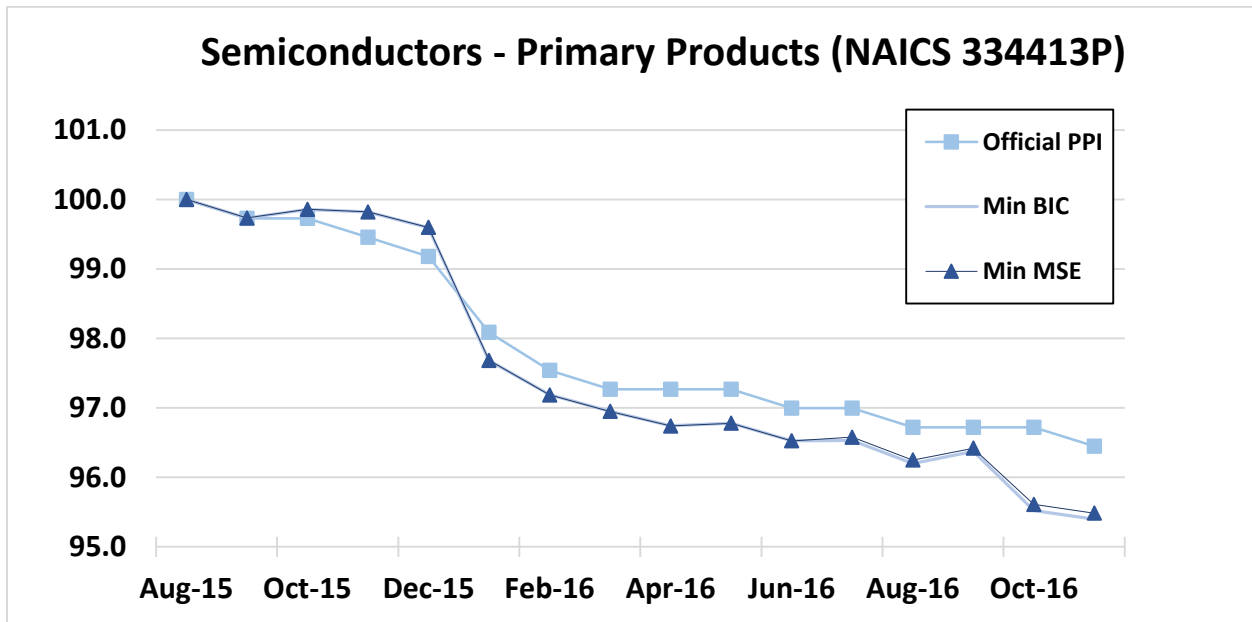
The hedonic indexes show an overall greater decline than the official PPI index. The magnitudes of

decline are modest, as would be expected from the time dummy coefficients, but they do suggest that

the matched-model index is not fully accounting for quality improvements in processors. The hedonic

indexes are also very similar, which suggests that our hedonic estimates of price change are consistent

even when selected using two different techniques. It is also likely that if the notebook and server

processors were also adjusted with hedonic time dummy models that the rate of decline for the hedonic

indexes would be greater, assuming, of course, that the time dummy coefficients on the time dummy models for the notebook and server models were similar to the desktop models.

Conclusion

While we do not agree that a time dummy hedonic model with a performance benchmark as its single regressor is suitable for the PPI, the work presented by BOS has provided a very good framework for us to expand upon. It has also sparked a discussion on how BLS can improve the PPI methodology for measuring the changing technology and pricing strategies in the microprocessor industry. After consulting the data evidence, we believe that the dramatic price decline shown by BOS is an artifact of excluding several significant processor product characteristics. The BOS result is an outlier when compared to all other possible models generated from the nine regressors we identify. Further, the BOS model has a comparatively poor fit.

The PassMark benchmark is also better suited for implementation in a model that the PPI could potentially use. The coverage across all low-end and high-end processors bolsters the already small sample sizes that we encounter when working with this data. Using this benchmark and quarterly data is necessary to ensure that we meet the operational goals of the PPI program.

We also believe that the methods we use to calculate any PPI index should be borne out by the data and not be *exclusively* determined by prior assumptions. It is vital that the data analysis we provide adheres to the BLS mission of providing objective products to the public.

The two techniques used to select the models we developed for the PPI index produced similar models, thus confirming that the data consistently support our estimations of inflation. We prefer to use statistical learning for model specification selection because it is intuitively easier to understand and explain to data users. Statistical learning techniques are continually improving which will allow us to refine our use of them over time.

The work in this paper offers a solution to the challenge of measuring quality change in the PPI microprocessor index and we hope that this research will lead to eventual implementation of new methods for quality adjustment. We also realize that the focus of this research is solely on desktop processors and developing a model for both mobile and server microprocessors is important to accurately measure all the outputs of the industry.

Acknowledgments

References

Aizcorbe, Ana, Carol Corrado and Mark Doms. "When Do Matched-Model and Hedonic Techniques Yield Similar Price Measures?", Presented at the 2000 NBER Summer Institute and updated in this condensed 2003 version, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=550421.

Blackwell. "Multiple Hypothesis Testing: The F-test"

David M. Byrne, Stephen D. Oliner, and Daniel E. Sichel (2017). "How Fast are Semiconductor Prices Falling?", The Review of Income and Wealth, 2017.

Cutress, Ian. "The Intel 6th Gen Skylake: Core i7-6700K and i5-6600K Tested". *Anandtech*, August 5, 2015. http://www.anandtech.com/show/9483/intel-skylake-review-6700k-6600k-ddr4-ddr3-ipc-6th-generation.

Dziak, John J., Donna L. Coffman, Stephanie T. Lanza, and Runze Li. "Sensitivity and specificity of information criteria". The Methodology Center, The Pennsylvania State University, https://methodology.psu.edu/media/techreports/12-119.pdf.

ExtremeTech, https://www.extremetech.com/extreme/188776-how-l1-and-l2-cpu-caches-work-and-why-theyre-an-essential-part-of-modern-chips.

Flamm, Kenneth, "Has Moore's Law Been Repealed? An Economist's Perspective", *Computing in Science and Engineering*, March/April 2017.

Flamm, Kenneth, "Has Moore's Law Been Repealed? Empirical Analysis of Innovation in Semiconductors", Presented at the March 2017 CRIW.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, Springer, http://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf.

Holdway, Mike, "An Alternative Methodology: Valuing Quality Change for Microprocessors in the PPI", Bureau of Labor Statistics, https://www.bea.gov/papers/pdf/mpuvqa.pdf.

Intel, http://www.intel.com/content/www/us/en/architecture-and-technology/turbo-boost/turbo-boost-technology.html.

Intel. "Intel Stable Image Platform Program [Intel SIPP].
http://www.intel.com/content/www/us/en/computer-upgrades/pc-upgrades/sipp-intel-stable-image-platform-program.html.

James, G., et al., *An Introduction to Statistical Learning: with Applications in R*, Springer Texts in Statistics 103, DOI 10.1007/978-1-4614-7_6, Springer Texts in Statistics+Business Media New York 2013.

Puget Systems, https://www.pugetsystems.com/labs/articles/Is-CPU-Base-Frequency-Still-a-Relevant-Spec-512/.

Rambus, *Understanding Dennard Scaling*, https://www.rambus.com/blogs/understanding-dennard-scaling-2/?nabe=4857318206603264:1,658317845436B256:0.

Van Zant, Peter, *Microchip Fabrication: A Practical Guide to Semiconductor Processing*, Sixth Edition, (New York: McGraw Hill, 2014).

Appendix A: SPEC and PassMark benchmarks

Our data set includes the SPEC CPU 2006 benchmark score for desktop microprocessors. SPEC benchmarks were obtained from their website in January 2016. Both Speed and Rate benchmarks were collected; the Speed metrics are used for comparing the ability of a computer to complete a single task. This metric measures the performance of the system using a single core of the system processor which is greatly affected by the clock speed of the processor and its cache size. The SPEC rate metric measures the throughput of a machine carrying out several simultaneous tasks. This metric provides a good overall measure of performance of modern multi-core processors. SPEC rate metrics are typically most affected by the number of processor cores on a system. We include the SPEC rate benchmark to our evaluation because a key price determining characteristic of today's processors is based on the number of cores it has[36]. The SPEC benchmark suite is used primarily by server and workstation manufacturers to performance test their products. Many low-end desktop CPUs do not have SPEC benchmark scores. To better reflect all the desktop CPUs that Intel produces we turn to PassMark.

PassMark CPU benchmark results are gathered from users' submission to the PassMark website as well as from PassMarks own internal testing. The PerfomanceTest software is available for purchase on the PassMark website, this benchmarking software conducts eight different tests and then averages the results together to determine the CPU Mark rating for a system. PassMark runs one simultaneous CPU test for every logical CPU physical CPU core or physical CPU package. The eight tests include an Integer Maths test, a Compression Test, a Prime Number Test, an Encryption Test, a Floating-Point Math Test, an Extended Instructions Test, a String Sorting test and a Physics Test[37].

---

[36] ftp://ftp.hp.com/pub/c-products/servers/benchmarks/SPEC_CPU2006_Overview_101907.pdf
[37] https://www.cpubenchmark.net/cpu_test_info.html

Appendix B: *ISL* Model Specification Selection

The basic steps we use for prescreening and cross-validation come from page 205 of *ISL*:

1. Let M_0 denote the *null* model, which contains no (p) predictors. This model simply predicts the sample mean for each observation.

2. For *k*=1,2,…*p*:

   (a) Fit all $\binom{p}{k}$ models that contain *k* predictors.

   (b) Pick the best among these $\binom{p}{k}$ models, and call it M_k. Here *best* is defined as having the smallest RSS, or equivalently R^2.

3. Select a single best model from among M_0, …, M_p using cross-validated prediction error, C_p (AIC), BIC, or adjusted R^2.

For step 3, we only use cross-validation (10 fold). We repeat the steps 500 times and we calculate the standard errors for each of the 1 through p models. The model with the smallest number of predictors whose standard error was within range of the lowest MSE value is selected.

The code we use to implement this technique is based on code from page 250 of *ISL*. We added code to repeat the process 500 times and to calculate the standard errors.

Appendix C: Calculating Inflation Rates from Time Dummy Coefficients

Let y_i denote the log of price for processor i. Let x_{ik}, $k = 1, \dots, K$ denote a set of regressors. Let $t = 1, \dots, T$ denote time periods. To describe the indexes we construct from our annual data, each period is one year and $T = 7$. Consider the two year time dummy model:

$$y_i = \alpha_t + \Delta_t d_i + \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + u_i \tag{1}$$

where

$$d_i = \begin{cases} 0 & \text{if } i \text{ was introduced in } t \\ 1 & \text{if } i \text{ was introduced in } t + 1 \end{cases}$$

Letting $\hat{\Delta}_t$ denote the OLS estimate of the time dummy coefficient, an index $\hat{\pi}_t$ of quality adjusted prices between t and $t + 1$ is defined by:

$$exp(\hat{\Delta}_t) = 1 + \hat{\pi}_t \tag{2}$$

Solving explicitly gives:

$$\hat{\pi}_t = exp(\hat{\Delta}_t) - 1 \tag{3}$$

Using overlapping two year regressions, $T - 1$ indexes like this can be formed from T years of data. An *average* annual rate of inflation $\hat{\pi}$ is defined by:

$$(1 + \hat{\pi}_1)(1 + \hat{\pi}_2) \dots (1 + \hat{\pi}_{T-1}) = (1 + \hat{\pi})^{T-1} \tag{4}$$

Solving for an explicit formula gives:

$$\hat{\pi} = exp(\bar{\Delta}) - 1 \tag{5}$$

where

$$\bar{\Delta} = \frac{\hat{\Delta}_1 + \hat{\Delta}_2 + \cdots + \hat{\Delta}_{T-1}}{T-1} \tag{6}$$

To describe the indexes we construct from our *quarterly* data, we redefine each period t to be a quarter, with $T = 8$. Noting that $\hat{\pi}_1, \dots, \hat{\pi}_{T-1}$ now denote quarterly estimates, we obtain an average annual index $\hat{\pi}$ by first rearranging equation (4) as

$$[(1 + \hat{\pi}_1)(1 + \hat{\pi}_2) \dots (1 + \hat{\pi}_{T-1})]^{\frac{1}{T-1}} = 1 + \hat{\pi}_Q \tag{7}$$

where $\hat{\pi}_Q$ is the average quarterly rate of inflation, and then solving

$$(1 + \hat{\pi}_Q)^4 = 1 + \hat{\pi} \tag{8}$$

which can be solved as

$$\hat{\pi} = (exp(\bar{\Delta}))^4 - 1 \tag{9}$$

where $\bar{\Delta}$ is now the average of quarterly estimates.