An Experimental Component Index for the CPI: From Annual Computer Data to Monthly Data on Other Goods. (Preliminary and Incomplete)

Timothy Erickson
Bureau of Labor Statistics

Ariel Pakes*
Harvard University and NBER

June 11, 2007

Abstract

Until recently the Consumer Price Index consisted solely of "matched model" component indexes. The latter are constructed by BLS personnel who visit stores and compare prices of goods with the same set of characteristics over successive periods. This procedure is subject to a selection bias. Goods that were not on the shelves in the second period, and hence whose price comparisons were discarded, were disproportionately goods which were obsoleted over the period, and consequently represented goods whose prices were falling. Pakes (2003) provided an analytic framework for analyzing this selection effect and showed that it could be partially corrected using a particular hedonic technique. Using personal computer data he showed that the hedonic correction could be substantial. The BLS staff has recently increased the rate at which they incorporate techniques to correct for selection effects in their component indexes. However their work and the work of other researchers shows *very little* difference between hedonic and matched model indices for other (non computer) components of the CPI. This paper explores why.

We look carefully at the data on the component index for TV's and show that differences between the TV and computer markets, together with the fact that the BLS data are high frequency, imply that to obtain an effective selection correction we need to use a more general hedonic procedure than has been used to date. The computer market is special in having both well defined cardinal measures of the major product characteristics, and exiting goods with relatively low values for them. In markets where such measures are absent and where turnover can be at the high quality end, we need to allow for selection on unmeasured, as well as measured, characteristics. We develop a hedonic selection correction that accounts for these phenomena and show that when applied to TVs it yields much larger selection corrections. In particular we find that matched model techniques underestimate the rate of price decline by over 20%. Moreover the BLS staff's recent successful push to modernize their data gathering procedures has made it possible to compute our index within the BLS's time constraints, making it a "real time" alternative to current procedures.

*We would like to thank Teague Ruder and Aylin Kumcu for excellent research assistance, Paul Liegey for his patience in explaining the BLS's algorithm for constructing the TV component index to us, Dale Jorgenson and Jack Triplett for helpful comments, and the BLS NSF and Tolouse Network for Information Technology for the support that allowed us to undertake this project. All errors and omissions are those of the authors and not of any of these institutions. In particular Table 8 contains empty entries because we have not yet finished all the calculations.

1 Introduction

This paper reports recent progress on improving the hedonic procedures used in the construction of price indexes. It modifies the approach developed and used on annual data on desktop computers by Pakes (2003) in a way that allows us to obtain tighter bounds to the compensating variations underlying the BLS's component indexes. The modification requires an extension of the hedonic adjustment techniques used in the earlier paper. Along the way we explain why the "biases" in both matched model and in prior hedonic indices seem to differ; (i) across component indexes and (ii) with the time interval between successive price observations.

Pakes (2003) used a model of a differentiated-product market as a framework for clarifying the implications of hedonic regressions that are relevant for the construction of price indexes. It showed that such regressions do not identify either utility or cost parameters. Nevertheless under the conditions supplied in that article the regressions can be used to bound the transfer needed to compensate consumers for changes in their choice sets (for the "compensating variation"). The bound is typically tighter than that given by the matched-model index because it takes partial account of the selection bias in the matched model indices caused by the exit of goods. Goods that exit, and hence whose price changes are not included in the index, are disproportionately goods whose charactristics have been obsoleted, and hence whose prices have declined. So omitting these goods removes price changes from the left tail of the distribution of price changes, and this causes an upward bias in the estimate of the average price increase.

Hedonic indexes partially correct for this bias by using a hedonic predictions for the exiting period prices of the goods that exit. In order to reflect the current relationship between prices and characteristics, the regressions used to predict the exiting good's price should include all relevant characteristics, be updated every period, and have no cross-period constraints. Subject to these requirements, any sufficiently rich functional form can be used.

The relationship between the hedonic prediction for the price of exiting goods, and that implicit in the matched model index, clarifies the difference between the indexes. The matched-model index implicitly imputes its own value, which is an average of the values for all continuing goods, as the predicted price relative for every exiting good. The hedonic prediction weights more heavily the predicted prices of continuing goods that have characteristics closer to those of the exiting good. So a matched-model index takes the index weight intended for an exiting good and redistributes it to the continuing goods proportional to their index weights, whereas our indexes implicitly redistribute an exit's weight more towards those continuing goods with similar characteristics. In Pakes' (2003) computer application the use of the hedonic rather than the matched model prediction changed the index rather dramatically. The reason was that the value of the observed tuples of characteristics that were similar to those of exiting computers fell rather dramatically – largely in response to the entry of newer machines that obsoleted them.

Until very recently hedonic predictions that were based on regression functions that were updated every period were difficult, if not impossible, to do within the BLS's monthly time constraints. The fact that the BLS has modernized its data gathering procedures by providing their data gatherers with hand held computers and instructing them to download their data nightly onto a central BLS data management system has changed this situation dramatically. After some preparatory work it should now be possible to substitute our hedonic index for the current index and still meet the BLS's production schedule.

However, as is reported by the National Academy of Sciences (2004), when standard hedonic procedures are tried on most of the BLS's component groups the resultant indexes are not much different from the matched model indexes for those groups. Take our TV example. Despite the twin facts that on average over 20% of our sample turns over during the bimonthly sampling interval, and that there is ample evidence indicating that the goods that exit have prices that fall disproportionately, the matched model and hedonic indices for TV's produce almost identical rates of deflation (see below). We begin by exploring the reasons for this phenomena. We then show that those reasons suggest use of hedonic indexes of a different kind then those used to date and that, at least in the TV market, the difference has marked implications on the results.

The reason that the standard hedonic produces results which are similar to the matched model index is not that there is no selection bias, rather it is because standard hedonic procedures do not correct this bias. The TV market is different from the computer market in that it does not have sharp cardinal measures of most of the characteristics that consumers value. Instead most of our TV characteristics are dummy variables indicating the presence or absence of advanced features (see Appendix 3). Moreover exit is disproportionately of high priced goods that have most of these features. They exit because they are obsoleted by newer high priced goods with higher quality versions of the same features, and we do not have good quality indexes for those features. As a result in the TV market, and we suspect in many other markets, selection is partly based on characteristics the analysts cannot condition on, i.e., on what an econometrician would call "unobservables".

Standard hedonic predictions for the prices of goods that exit do not account for the price differences generated by characteristics the analyst does not condition on. One alternative is to augment the standard hedonic with a good-specific "fixed effect" to account for the unobserved characteristics of the good, and then use the coefficients from a regression for the differences of prices of continuing goods on observed characteristics to predict the change in price of the exiting goods. We show that though this procedure does move the index in the expected direction, it only corrects for a small part of the problem (about 5% of our correction in the TV application).

This is consistent with the theoretical framework in Pakes (2003) which emphasythes that we should expect the hedonic evaluations of different characteristics to vary across periods. That is differential rates of entry, exit, demand growth, etc., in different parts of the product space typically generate hedonic pricing functions in which the relationship of both the observed *and* the unobserved characterisics to price changes over time. We provide rather convincing evidence that this is true in our data. Indeed we show that, perhaps not surprisingly, the contribution of the unobserved characteristics to price of exiting goods are falling at a disporportionate rate. As a result even if we used a fixed effect model which allowed the contributions of the observed characteristics to price to change over time, we would not pick up the impact of changes in the contribution of the unobserved product characteristics to price[1].

There is a second reason why standard hedonic procedures do not adequately control for selection and it has to do with the relatively high frequency of BLS data (most prices in this data are resampled at two month intervals)[2]. At this frequency prices are often "sticky", i.e. they do not change between successive readings. In fact on average only 40% of the prices change over a two-month interval. The regression function of prices on characteristics for goods which change their prices is different than that regression for goods that do not change their prices. If exiting and continuing goods prices were equally sticky, this would have no impact on our predictions for exiting goods prices in the period they exit. However, as we shall see below, prices of goods that are similar to the exiting goods are, perhaps not surprisingly given that they are in a changing part of the market, systematically less sticky than most. This fact will actually help us develop imporved correction procedures below.

This paper develops hedonic indexes that partially account for these phenomena. The procedure we adopt abides by the general conditions discussed in Pakes (2003), but differs in that our predictions for the prices of goods which leave the sample makes an explicit correction for the expected value of the disturbance from the hedonic regression for those goods. The prediction only provides an upper bound to the expected contribution of the disturbance, and hence is not "tight". It still, however, makes a substantial difference to the index.

There are two other features of our procedure that turn out to be important. First our procedure enables the BLS analyst to use only a small number of "easy-to-clean" product characteristics in the hedonic regression. Given the recent computerization of the data gathering process, labor-intensive

[1]Further this fact, together with any reasonable model for the goods that exit, will imply that the fixed effect estimator also generates biased estimates of the contribution of the observed characteristics to price. We note that the fact that selection is based on unobserved as well as observed characteristics also rules out the use of familiar statistical sample selection correction procedures like the propensity score.

[2]About 75% of the BLS TV sample comes from alternating bimonthly subsamples. The bimonthly samples we use are obtained by apportioning the CPI monthly subsample of prices between the CPI's two bimonthly samples. The annual rates implied by monthly indexes differ slightly due to the splicing operations used to construct a monthly index from bimonthly indexes.

cleaning of a large number of characteristics is now the main reason the BLS analyst can not fully implement Pakes'(2003) suggestion and run different hedonic regressions every period. Currently the BLS runs only one TV hedonic regression a year (see Moulton, Lafleur and Moses,1998, for more detail on the current CPI TV component index), and a hedonic index based on hedonic prediction functions which are not updated every period does not provide a bound to the compensating variation. A procedure which only uses a small number of easy to clean characteristic should enable the analyst to remedy this situation.

A second feature of our procedure that is helpful is that it uses local-linear nonparametric regression to insure that our price predictions are based on the price movements of products that have similar charactristics to the characteristics of the product that exited. The price variation in the TV market is enormous; from \$66 to over \$10,000, reflecting the large differences in products that the BLS includes in this commodity group. The entry and exit of particular TVs in this market, as in many markets, tends to disproportionately influence, and be disproportionately influenced by, prices of close competitors. Use of the local-linear nonparametric estimator insures that the hedonic predictions for one good are not overly sensitive to goods which are in very different parts of the product space (and movements in price of very different types of goods are not highly correlated).

A Summary of the Empirical Results. The matched model's estimate of the average annual rate of price change in the TV market is -9.93%. If we apply a standard linear in logs hedonic prediction technique for the exiting goods prices using a set of twenty four characteristics that are similar to the set of characteristics used in the annual BLS hedonic regression for their TV component index we get an average rate of change of -10.21; a selection correction of under 3%. However, were the BLS to use a characteristic set this large, the amount of data cleaning needed would imply that they could not produce an index which used a new hedonic regression every period. So we also consider a nine variable characteristic set which does not require extensive cleaning and hence could be used in a production setting. When we use standard hedonic procedures with these nine characteristics we get a hedonic index falls at a *much slower pace* then the matched model index.

The remainder of the analysis was done with the nine variables characteristic set. First we moved to predictions based on hedonic regressions for the log of price differences, regressions which difference out the impacts of unobservables whose effects on price are constant over time. In these regressions we allowed the coefficients of observables to vary freely from period to period. Allowing for unobserved fixed effect in this way caused the index then to jump back up to -10.33%, now surpassing the matched model index, but only by 4%.

We now move to our correction procedures. They can all be justified in terms of upper bounds to the required compensating variation. To correct for the selection generated by exit we need a rule for when exit occurs. We assume that exit occurs when the price a product would sell for is lower than some threshold. The threshold itself is modeled as a nonparametric function of the good's observed *and* its unobserved characteristics, and the functional form for the threshold is free to change from period to period. We try three correction procedures and introduce them in order of their data requirements[3].

The first correction procedure uses the fact that the exit rule implies that the observed conditional expectation of the change in the unobservable of the goods that continue provides an upper bound for the conditional expectation of the unobservable for the goods that exit. Here the conditioning set includes the observed characteristics and the intial value of the contribution of the unobserved characteristics to price. This correction, then, only requires the BLS's two month sampling interval data on goods that *continue*. It produces an annual average rate of inflation of -11.02%. That is our simplest selection correction produces an index which falls at a 11% faster pace than the matched model index; a correction which is much more than three times the correction obtained from the standard hedonic.

About a quarter of the BLS sample has a one month, rather than the standard two month, sampling interval. This subsample contains information on the first month price changes of goods

[3]Each time we add data we have to strengthen our exit rule slightly to insure that use of the added data does not violate our bound on the compensating variation.

which exited the sample in the second month of the bimonthly sampling interval. The first of our two other procedures augments the data in the bimonthly sample with; (i) the rates at which the goods that exited in the second month of the sampling interval changed their price in the first month of that interval, and (ii) with the two month price changes of the *continuing* goods who changed and those that did not change their price in the first month of the sampling period. The rates of price change for goods which exit in the second month are noticeably larger than for goods which continue, and the two month price changes of goods that continue and changed their price in the first month of the two month sampling interval are much more negative than the price changes for goods that continue and do not change their price in the first month of the sampling interval. Moreover under our assumptions an upper bound to the conditional expectations of the change in unobservable for the goods that exit and change their price in the first period is the conditional expectation of the unobservalbe of the goods that do not exit and change their price in the first period (and similarly for goods that do not change their price in the first month). This procedure produces an index with an annual rate of change of -11.29%; a rate of price decline 14% larger than that of the matched model index and an adjustment for selection over four times as large as the adjustment provided by the standard hedonic.

The final correction adds the information on; (i) the first month price changes of the goods that exit in the second month of the sampling interval, (ii) and the expected price change for *continuing* goods conditional on their prices at the end of the first month of the sampling period as well as on their observed and unobserved characteristics upon enterring the two month sampling period. The first month price declines of goods which exit in the second month are much larger the price declines of goods with similar characteristics which continue over both months. Moreover our model indicates that the conditional expectation of the second month price decline for continuing goods is larger than that for goods which exit in the second month. As a result if we substitute the conditional expectation of the price change in the second month for continuing goods for the conditional expectation of the second month for exiting goods we maintain our bound. When we use this procedure the average value of the index now falls sharply, to -12.48%. This is a rate of price decline which is more than 25% larger than the decline of the matched model index and implies an adjustment which is over eight times the adjustment provided by the standard hedonic.

Our hedonic indexes differ systematically from matched-model indexes because the characteristics (both observed and unobserved) of exiting TVs, and their sticky price rates, do not look like those of a random sample from the distribution of the characteristics of continuing TVs. The exiting goods characteristics *do*, however, look much more similar to the characteristics of a subset of the continuing TVs: those that will exit in the next period (see Appendix 3). So an alternative index we could construct is to use the price changes for the about to exit TV's for the price change of TV's that do exit. Unfortunately this would not be possible in a production setting, since we would not know which goods would exit in the following period when the index is constructed. However an alternative that could be used in a production setting and has a similar intuitive justification, is to use the price falls of the goods that do exit in the period prior to exiting as an estimate of the price falls of those goods in the period they do exit.

Though there may be shocks to the price surface in a period which makes the price change in the period preceeding the period the good exits different than in the period the good does exit, we would expect that on average the price changes in the period before a good exits to be similar, though somehat smaller (in absolute value), than in the period they do exit. "Similar" because actual data gatherring dates vary over the two month interval for different goods in an arbitrary way, and smaller because both a priori reasoning and the monthly data subsample indicate that the price falls in the exiting period are larger than in the period prior to exit. Substituting the prior period's price change of the good for the price change in the period they exit leads to an average annual index of between -11.8% and -12.1% (depending on details to be explained below). These rates of decline are higher than the rates obtained when we do not use the rates of price decline in the first month of the two month sampling period, but lower than those obtained when we do use this data. That is the prices of about to exit goods fall faster than those for continuing goods but slower than those for goods which actually exit, just as we expected. Since this result is gotten from an entirely different, though perhaps equally intuitive, procedure it gives us reason to believe that our result that the selection correction for the annual rate of price decline for TV's is over 20% is robust to using different correction techniques. Moreover we think similar procedures could be applied to a large fraction of the CPI component indices (just how large is a topic of our ongoing research).

Organization of the Paper. The remainder of this paper is organized as follows. Section 2 describes our data, Section 3 describes the formulas for our research indexes given a set of price prediction equations. Section 4 explains and discusses our methods for predicting price. Section 5 reports our empirical results. The rest of the sections are not written yet, but we working on them. We intend to have a Section 6 which compares our approach to a research index that mimics the hedonic method currently used in the CPI. We also intend to have a Section 7 which redoes the analysis on date that is "out of sample" (data that starts after the end of our sample period). This should give us an idea of how well our technique would do in a production setting in which there is no time to experiment with alternatives before picking an appropriate estimator.

2 Background: Properties of the Data and Biases in the Index.

CPI price quotes from March 2000 to January 2003 are used. A "cleaned characteristics" subset of each period's July and August data was prepared by the CPI industry analyst for use in their current hedonic procedure. We assigned the cleaned characteristics to all months by matching model numbers. The resulting 35-month data set contains 8,195 prices, or 79.9% of all prices.[4] These range from \$66 to \$10,079. The average monthly sample contains 234 prices and has mean, median, minimum, and maximum prices equal to \$725, \$366, \$81, and \$7836 respectively.

Just over three quarters of the CPI price quotes are collected at 2-month intervals from odd

[4]Comparing, where possible, statistics for the full and cleaned data sets shows that the latter data is very similar to the full data. Noteworthy departures are slightly lower entry and exit rates (making our problem harder) and a mean price that is about \$40 higher than that for the full data.

and even numbered month subsamples (these are regionally defined). The other one quarter of the quotes are collected at one month intervals (these are from NY, LA, and Chicago). As a result we focus on price relatives, exits, etc. over two month intervals, though all the sample observations available for the two months period are used (whether from the one month or one of the two month subsamples).

On average, 22.5% of the TVs present in any period $t-2$ are not present in t, with 19.7% of these being permanent exits. That is some of the goods that are not available in the current period are expected to be back (and in many cases are back) on the shelf in a future period. Similarly, 24.0% of TVs in t are not present in $t-2$, with 17.0% of these being substitutes (the good that was to be sampled for comparison period prices was not present at the outlet so another good had to be substituted for it) and 4.1% being scheduled additions to the sample (goods that were scheduled to be rotated out of the sample). An average of 2.9% of the exits are temporary, while 2.9% of entering TVs are returning from temporary absence.[5]

Price relatives for different subsets of the data play a key role in this paper. For any two periods $t-2$ and t there are TVs in our sample in both periods, and for these we have relatives. Between July 2000 and November 2002 the average number of $t-2$ to t relatives was 183.45. Some of their characteristics are listed in Table 1. [6]

The average number of $t-2$ to t relatives that were for TVs that would exit before $t+2$ was 40.21 (which is 22.37% of those relatives). We say these TVs are "about to exit" and denote them by "a-exit." We believe the behavior of their price relatives will be more like that of goods which exit the sample between t and $t+2$ than that of a randomly drawn price relative, and we will show that what evidence exists lends strong support to this belief. As a result we will use this subsample

[5]Note that a good that is temporarily off the shelf may not be off the shelf because it has been obsoleted but rather because of a stock-out caused by unexpectedly high demand. *** how is this dealt with in our framework **** Also the numbers above come from slightly different series. The exit rates are computed on a series that excludes the last 4 months from each bimonthly subsample, the deleted months used to determine which exits eventually return. Computation of the entry rates exclude the earliest months from each subsample for analogous reasons.

[6]This 29-month span is derived from our full 35-month data set as follows: One of the 18 odd months is used up to make relatives. Two more of the 18 are used up to determine those relatives that will exit before $t+2$ and those that entered in $t-2$. Three of the 17 even months are used up in the same way.

for clues as to the unobserved price relatives for goods that exit before their t+2 price data was gatherred. The average number of $t-2$ to t relatives that were for TVs that *entered* in $t-2$ was 46.03 (which is 25.52% of the relatives.) These TVs have "recently entered" from the standpoint of t, and we denote them by "r-new."

Table 1 provides some summary statistics on the price relatives for these subsets. We note that 61.55% of all $t-2$ to t relatives equal 1; that is there are a lot of "sticky" prices. Thus we provide summary statistics for the subsample of non-sticky prices as well as for the overall sample.

We begin with the data on the goods that are about to exit. The first point to note is that these goods have a faster rate of price decline from $t-2$ to t then continuing goods. The about to exit goods decline at about *twice* the average rate of decline and the difference is highly significant (with a t-ratio of about six). Second the about to exit goods have a significantly lower fraction of sticky prices (the standard errors for these fractions vary from .006 to .015). Moreover if we look just among non-sticky prices the absolute difference between the mean price relative for goods about to exit and the goods that continue is even more pronounced. That is among prices that do change, the prices of the goods that are about to exit fall substantially more than a randomly chosen price change.[7] If goods that are about to exit have prices that behave more similar to the prices of goods that do exit, then these numbers reinforce the belief that, by throwing out the goods that exit, matched model procedures overestimate inflation.

The last panel of this table uses the data from the quarter of the sample with monthly observations. By calculating the first month sticky price and price relative decline rates of the goods that exit in the second month of the two month sampling interval we get an idea of these rates for exiting goods in the period in which they exit. We then compare these rates to the rates for the same goods in the period in which they are classified as about to exit. The two month rate of decline of the price relatives for the sample with monthly observations is similar to that of the overall sample, while the rate of decline for the about to exit relatives is a bit larger (though this

[7] The about to exit goods also have higher price variance then other goods, though most (though not all) of this increased variance is because they have a larger fraction of non-sticky price.

difference is not statistically significant).

52% of the monthly observations that exit over the two month sampling intervals have observed prices after the first month. The average price relative of these goods for the first month is .9756. If the average price decline in the second month of the sampling period were distributed independently of the average price decline in the first month, then the implied two month average price relative would be $(.9756)^2 = .9518$, with a standard error of .0136. That is, under the independence assumption, the rate of price decline of the goods that exit over the two month period in which they exit is larger (in absolute value) then in the two month period when these same goods were about to exit (though the difference is not statistically significant). Under this same independence assumption the rate of price stickiness of these goods in the period in which they exit is lower than that rate for the same goods in the period before they exit (.39 vs. .57) and this difference is both large and precisely estimated (its standard error is under .02). Below we weaken the independence assumption and use these data to get lower bounds to the rate of price decline and the price change rate for exiting goods. However the "back of the envelope results" presented here are illustrative of the more detailed results below. From an economic point of view this should not be surprising. All it is saying is that the periods in which goods exit tend to be periods in which they are under increased price pressure.

Note also that goods that are recently introduced also have price relatives that on average fall at a faster pace than continuing goods, though the difference is not nearly as striking as it is for about to exit goods (it is only 1/4 to 1/5 the differential rate of decrease of goods that are on the verge of exit, and the difference with continuing goods is not statistically significant). Sill this finding has interesting implications for price index construction procedures. As noted by Pakes (2003) introducing new goods earlier into the index will only ameliorate new goods biases if prices fall in their introductory periods. It seems that early introduction of new goods would indeed ameliorate new goods biases in TVs. Moreover the tendency for new goods prices to fall is more pronounced among new goods whose prices do change.

Finaly we note that the results in Table 1 go a long way towards explaining the difference in results for matched model indices based on different intervals of time. Compare, for example, the average of the matched model indices with a two month sampling interval with that from a four month sampling interval. The latter omits price changes of two types of goods that are included in the two month interval data; (i) goods that are "about to exit" in the first two month interval, and (ii) goods that are "recently new" in the second two month interval. Both these subgroups of goods have prices that fall at a faster rate then a randomly drawn continuing good. So the four month interval index misses two groups of prices changes whose prices are falling disproportionately.

The fact that the longer sampling interval data omits price changes of about to exit goods accentuates the selection bias we study here. The fact that it omits initial price changes of recently enterred goods, will, in markets where initially prices fall, accentuate a bias we do not attempt to correct for in this paper. This is the bias caused by the fact that the index does not attempt to capture the inframarginal rents which accrue to individuals who would have bought the new good at a price higher than the highest price at which the new good enterred the index (see Pakes, 2003, for further discussion). To get some indication of how these biases increases with the length of the sampling interval, we used our data to calculate the matched model indexes when we assumed two, four, and twelve month sampling intervals. The annualized rate of deflation for the three intervals were, respectively, -10.59%, -8.99%, and -6.48%. So going from a two month to an annual interval increases the matched model's estimate of inflation by about 40%.

Prices of Enterring and Exiting Goods. The next table summarizes information on the prices of enterring and about to exit goods which will help with an understanding of the role of selection in this market. It has coefficients and t-values from regressions of log prices on a constant and two dummies, one for the goods that just enterred and one for goods that are about to exit. This provides an indication of the level of prices, and hence the "type" of goods, that just entered and/or are about to exit. The regressions are done differently for odd and even numberred periods as the BLS samples different cities in those periods.

Table 1: **Price Relatives.**

Variable	Full Sample.	a-exit	r-new	contin.	exit-cont	new-cont
mean	.9849	.9729	.9844	.9881	-.0152	-.0037
(s.d. of mean)	(.0010)	(.0024)	(.0019)	(.0014)	(.0028)	(.0023)
cross-section s.d.	.0677	.0778	.0606	.0646	n.r.	n.r.
Fraction of Subsample With Relatives						
Equal 1 (or "sticky")	.6155	.5390	.6203	.6380	-.0990	-.0176
Greater than 1	.1166	.1097	.1142	.1213	n.r.	n.r.
Less than 1	.2679	.3513	.2655	.2407	n.r.	n.r.
# of obs.	5320	1167	1335	2818	n.r.	n.r.
Among Price Relatives Not Equal to 1 (i.e. not "sticky").						
mean	.9622	.9460	.9608	.9682	-.0222	-.0074
(s.d. of mean)	(.0024)	(.0056)	(.0049)	(.0034)	(.0063)	(.0058)
cross-section s.d.	.1039	.1083	.0920	.1024	.0059	-.0104
# of obs.	2017	549	514	1067	n.r.	n.r.

Using One Quarter of Sample with Monthly Price Quotes				
variable	All Monthly Data 2-month	a-exit 2-month	month 1 (goods that exit by month 2)	(month 1 exit by month 2)^2
mean price relative	.9835	.9679	.9756	.9518
(s.d. of mean)	(.0016)	(.0036)	(.0068)	(.0136)
sticky price rate	.6569	.5776	.6270	.3931
# of obs.	1428	334	207	207

The point made by this table is that both the newly enterring goods and the about to exit goods have prices that are *higher* than those of continuing goods. This is not surprising for newly enterring goods as it simply means that new goods typically enter at the high quality end of spectrum. What is somewhat surprising is that this is also true for goods that are about to exit. This differentiates the TV market from the market for computers where almost all exits are from the low end of the quality spectrum in the period before they exit. Like in computers, in TV's most improvements have been at the high end. However in TV's the exitors that are displaced by the new entrants are also typically high end goods. The "low-end" products in the TV market do not turnover nearly as much.

We will see that though our characteristics can differentiate between high and low quality TV's, they have more difficulty with distinguishing between two high quality TV's one of which is based

Table 2: **Characteristics of Entering and Exiting goods.**

Specification	Constrained OLS		Minimum Distance	
	exit	new	exit	new
1. S0 (Odd)	.106	.161	.075	.146
	(2.66)	(4.14)	(1.94)	(3.86)
2. S0 (Even)	.121	.133	.097	.130
	(3.17)	(3.53)	(2.61)	(3.51)

S0 has a constant and two dummies, one for goods about to exit and one for goods that just enterred. Odd and Even number periods done separately as they sample different cities. The constrained OLS and minimum distance estimates differ in that the latter weights with the covariance matrix across periods.

on older technology and hence has been obsoleted. For example we know which TV's have liquid crystal display, but we do not have a good measure of the improvements that have occured in sharpness of the liquid display over time. This is a second feature which differentiates the TV market from the computer market. In the computer market the major characteristics that are improving over time (e.g., speed, RAM, hardrive capacity, ...) have natural cardinal measures which make them easy to compare across products .

3 Index Formulas

To begin as simply as possible we start with indexes that are linear in the logs of price relatives. This makes the indexes linear in the regression error from the logarithmic hedonic regressions we and others have used, and this in turn makes the relationship of our results to "quality change" bias transparent. We intend to come back to more complex indexes that work directly with this regression error at a later date, as they have a larger role to play in other indexes[8].

[8]In particular to construct the Laspeyre's index we need to exponentiate the logs and hence exponentiate the hedonic regression error. Since the Laspeyre's index is the only index that has an intepretation in terms of a bound on compensating variation, there are good reasons for thinking the indices that deal directly with the regression error might be important.

3.1 Bimonthly indexes

All indexes are versions of

$$G_t = \sum_{q \in S_{t-2}} w_{q,t-2} y_{qt} \tag{1}$$

where q denotes a quote, w_{qt} is period-t weight, $y_{qt} = \log(p_{qt}/p_{q,t-2})$ is an actual or imputed log-relative, and S_{t-2} is a subset of all quotes active in period $t-2$.[9] This is the log of a geometric mean index, which approximates the average proportionate change in prices. The weights $w_{q,t-2}$ are obtained by dividing each period-$(t-2)$ regional TV expenditure-share equally among all the quotes for that region and then renormalizing them so that $\sum_{q \in S_{t-2}} w_{q,t-2} = 1$. The regional expenditure shares are estimates from the CPI data base.[10]

Explicitly denoting an estimate of a log-relative as $\hat{y}_{qt}$, hedonic and matched-model indexes can be written as

$$G_t^{hed} = \sum_{q \in A_{t-2}} w_{q,t-2}^{hed} \hat{y}_{qt} \tag{2}$$

$$G_t^{mm} = \sum_{q \in C_{t-2}} w_{q,t-2}^{mm} y_{qt}, \tag{3}$$

where A_{t-2} is the set of quotes for which prices were successfully collected in period $t-2$, and $C_{t-2} = A_{t-2} \cap A_t$. That is, the matched model indexes average the price relatives for goods for which price information was collected in *both* periods, while the hedonic averages predicted price relatives for *all* goods whose prices were collected in period $t-2$[11].

[9] The use of t and $t-2$ is a result of the bimonthly sampling procedure which implies the basic indexes are bimonthly. Monthly indexes are derived from these by a linear splicing procedure.

[10] Past values of the CPI subindex for TVs are used in making the estimate for any period t. We take these estimates as *given;* we do not prepare our own estimates based on past values of any of our research indexes.

[11] An early version of this paper also computed the hybrid indexes introduced in Pakes (2003). These impute relatives only for TVs that exit between $t-2$ and t, and use actual price relatives for goods that were available in both periods, i.e. $G_t^{hyb} = \sum_{q \in C_{t-2}} w_{q,t-2}^{hed} y_{qt} + \sum_{q \in A_{t-2} - C_{t-2}} w_{q,t-2}^{hed} \hat{y}_{qt}$. The attraction of hybrids is that they have no estimation error in their price relatives for the continuing goods, and they eliminate much of the selection bias in the matched model index by using hedonic predictions for the goods that exit. On the other hand they treat the error from the hedonic regression differently for the two types of goods, and this can cause a (different) selection bias. We decided that the tradeoff between bias and variance was not an issue we wanted to deal with in this paper and hence omitted the hybrids. The actual values of the hybrids we calculated for our sampling period are, however, available on request from the authors.

4 Hedonic Predictions.

For simplicity we begin with a linear regression model for the (log) price levels of goods in a given period (we present non-parametric results directly thereafter). Let Z_t be the $n \times K$ matrix of characteristics of those TVs for which prices were collected in period t and p_t be the corresponding $n \times 1$ vector of log prices. Then a typical period-t hedonic regression coefficient is given by

$$d_t = \left(Z_t' Z_t\right)^{-1} Z_t' p_t, \tag{4}$$

and the prediction for log price is $\hat{p}_t = Z_t d_t$. As noted above there are no restrictions on these coefficients and there is no necessary relationship between the coefficient vectors estimated in different periods.

We fit this regression to every month in each bimonthly sample, using each of three different sets of regressors for Z, all of which include a column of ones. The three sets of regressors, to be denoted by $S4$, $S9$, and $S24$ are:

- $S4$: log of screensize in inches, a dummy indicator for projection TVs, the interaction between these two variables, and the square of log-screensize.
- $S9$: the variables in $S4$ plus dummy indicators for picture-in-picture, flat-screen CRT display, HDTV-ready, a high-quality reputation Brand A, and a low-quality reputation Brand Z.
- $S24$: the variables in $S9$ plus the additional variables listed in the notes to Table A1 at the end of the paper.

The values for the variables in $S4$ and $S9$ can be verified with minimal effort on the part of CPI staff, and therefore can be used to fit an up-to-date hedonic regression at the time each index is prepared in a production setting. This is not so for the additional variables in $S24$. The current hedonic procedure of the CPI index for TVs uses a different but similarly lengthy list of regressors, most of which have values that are difficult to verify in the short period of time during which each index's production. This is why the current method fits a regression no more than once a year.[12]

The first three rows of Table 3 show that any of the three sets of characteristics does quite a good job of accounting for variance in the traditional dependent variable of hedonic regression, log-price. Even S4 has very high R^2's. It is not unusual to get high R^2's in hedonic regressions on differentiated product markets, indeed it is a major reason for the increased use of characteristic models in demand estimation. However these R^2's are higher than usual, which probably attests to the quality of the BLS data.

Note that there is a noticeable improvement in fit in moving from $S4$ to $S9$ but not much further improvement in adding the 15 characteristics needed for $S24$. The fourth panel of the table provides fits from a non-parametric estimate of the hedonic surface. The method used is local linear kernel regression with a cross validated bandwidth. Appendix 1 provides the formulae used in the non-parametric analysis. The small improvement in the fit of the $S24$ regression relative to the $S9$ regression is similar to the improvement obtained when we substitute the non-parametric (NP) regression for the linear regression; and the NP regressions use only the same 7 characteristics of the $S9$ specification. Moreover the NP regression is also easy to compute in a production setting, as it involves only running a pre-programmed algorithm.

Table 3: **Hedonic Regressions: Dependent Variable is Log-Price**

Regressors	mean R^2	mean adj R^2	min R^2	min adj R^2	max R^2	max adj R^2
S4	.8927	.8908	.8698	.8676	.9127	.9111
S9	.9552	.9533	.9413	.9369	.9663	.9649
S24	.9707	.9672	.9585	.9530	.9777	.9753
NP	.9641	.9626	.9236	.9179	.9730	.9719

Table gives summary statistics from log-price regressions run on each of the 35 months from March 2000 to January 2003. The "degrees of freedom" used for the calculation of the adjusted R^2's from the non-parametric regression was assumed to be 24 (the same as in S24).

[12] Indeed when we move to out-of-sample predictions, see our "to do" list at the end of this report, we will assume that we only have the $S9$ variables to work with.

5 Unobserved Characteristics and Hedonic Bounds.

Under standard assumptions on consumer behavior the prices of two goods with identical characteristics should be the same. So if we observed all relevant product characteristics, we should be able to predict the prices of goods that exit the sample from the prices of goods with similar characteristics that remain in sample[13]. This prediction problem, however, gets more complicated when there are characteristics of the goods that consumers value but Econometricians do not observe (and hence can not condition on). Recall that in the TV market exit is largely a result of high quality goods obsoleting older high quality goods, and that we do not have good cardinal measures which capture the differences between the different generations of high quality goods. As a result we might want to pay particular attention to unobserved product characteristics when correcting for selection bias in this market (and in other markets with similar features).

Part of the impact of the unobserved product characteristics on price will be capturred by the relationship between unobserved and observed characteristics, but the rest will appear as the residual from the hedonic regression function. If the relationship of the residual to the observed characteristic were no different for exiting goods than for a randomly drawn good, then we could obtain an unbiased estimate for the price of a good that exited the sample between two periods from the hedonic regression coefficients in the second period (a regression which only has price observations for the continuing goods and the new entrants), and the characterisitcs of the good that exited. However as we now show there are solid economic arguments that lead us to believe that the relationship between the unobserved and observed characteristics is different in the selected sample of exiting goods. Moreover the predictions these arguments make are born out by the data.

For simplicity assume the true hedonic function is linear and let η measure the contribution of unobserved characteristics to price, so that

[13]Strictly speaking this assumes that there is a rich set of products offered, so that the characteristic space is densely populated. For a statement of this property, and a demand estimation algorithm that makes intensive use of it, see Bajari and Benkard,2005. For justification of hedonic indices when their conditions are not satisfied, see Pakes,2003

$$p = z\beta + \eta, \tag{5}$$

where we have normalized the coefficient of η to be one. Our hedonic equation is obtained from a regression of p on z. To analyze its properties we need the properties of the regression of η on z.

If we let $j = x$ denote exiting goods, $j = n$ denote new entrants, and $j = c$ denote continuing goods, then

$$E[\eta|z] = \Sigma_{j=\{c,x,n\}} P\{j|z\} E[\eta|z, j].$$

Though the theory that tells us that goods with the same characteristic should sell for the same prices implies the coefficients on z in equation (5) should not differ between enterring, exiting and continuing goods, it says nothing about whether $E[\eta|z, j]$ differs by j. Moreover a standard selection argument would lead us to believe this regression function differs by j.

To see this we need a model for which goods exit. Temporarily assume that a product exits if its price falls below its marginal cost, denoted $m(z, \eta)$, and that $\partial m(z, \eta)/\partial \eta < 1$ (everywhere). Our normalization implies that price increases one to one with η, so we are assuming that price increases more than marginal cost when unobserved quality increases. Then there is a function $\underline{m}(z)$, such that the good exits if and only if $\eta \leq \underline{m}(z) - z\beta$, and

$$E[\eta|z, j = x] = E[\eta|\eta \leq \underline{m}(z) - z\beta] \leq E[\eta|z].$$

In particular when the good's observed characterisics lead to a small $\underline{m}(z) - z\beta$ then the goods that continue will all have values of η which are very high, while if $\underline{m}(z) - z\beta$ is large, goods will continue even if they have low values of η. So the distribution of η conditional on z (its support, its mean,....) will be different for the continuing than for the exiting goods.

To see whether such logic leads to a significant differences in the relationship between z and η for exiting, continuing and newly enterred goods in our data set, we estimated hedonic regressions for each period which allowed each of the three groups of goods to have different z-coefficients. Using the $S9$ regressor set of the last subsection, we then tested whether these coefficients differed from each other. The results are presented in Table 4. They clearly reject the null that the new and exiting good interactions are all zero.

Table 4: **Testing for Exit and New Good Interaction Terms.**

Test	$j = x$; F-test	$j = n$; F-test	$j = x$; Wald-test	$j = n$; Wald-test
Fraction Significant At Different α Levels				
$\alpha = .01$	.14	.11	.50	.54
$\alpha = .05$	.29	.21	.71	.71
$\alpha = .10$	.46	.29	.79	.75

F-test assumes homoscedastic variance-covariance,
Wald-test allows for heteroscedastic consistent covariance matrix.

The data contains at least two pieces of more detailed evidence on the difference between the unobserved characteristics of the exiting and continuing goods. First we can compare the estimates of the unobserved characteristics (i.e. the residuals from the hedonic regression) of exiting goods to those of continuing goods in the period prior to exit. Second we can compare the change in the estimate of the $\eta's$ of the exiting goods over the period immediately preceeding the period in which they exit to the change in estimate of the η's of the continuing goods over that period.

The change in η results also throw light on the appropriateness of an alternate procedure for correcting for selection bias; one based on the assumption that the contribution of unobserved characteristics to price does not change over time. Recall that if selection was based only on observed characteristics and a *time invariant* unobserved characteristic, or a "fixed effect", the average of $\eta_{t+1} - \eta_t$ should not differ between exiting and continuing goods. So under the fixed effect assumption we can form an unbiased prediction for exiting goods prices by regressing the log of the price differences (or of price relatives) of the *continuing* goods onto their characteristics, and then using that regression function to predict the price change for the exiting goods. This because the fixed effect assumption guarantees both that; (i) the contribution of the unobserved characteristics to price does not change over time, and (ii) that the log of price differences regression provides an unbiased estimate of the changes in prices caused by the market's re-evaluation of observed characteristics. Note the arbitrary difference in the way the fixed effect assumption treats the unobserved and the observed characteristics; it calculates its price change predictions from differences in the contribution of the observed characteristics to price over time but assumes, *a priori*, that the contribution of unobserved characteristics to price never changes.

The average difference between the η's of exiting goods and those of continuing goods in the period prior to exit was negative, but only slightly so, and the difference was not statistically significant. Table 5 presents the results from splitting the data into three groups – the goods that are about to exit (they exit during the next sampling interval), those recently new (they enter in period t), and the remaining goods – and then calculating the average change in the residual for each. About to exit goods have an average change in residual which is very negative, more than five times the absolute value of the change in the residual for the continuing goods, and highly significant, with a t-value more than five. That is though the values of the unobserved characteristic of the exiting goods in the year prior to exit are only marginally lower than that of the continuing goods, the values of η of the exiting goods are falling at a dramatically faster pace than those of the continuing goods.

There are a number of implications of this table that are worth noting. First since the contribution of the unobserved characteristic to price is falling at a rather strking rate just prior to exit, it is likely to be falling during the exiting period (and probably at a faster rate, as the exiting period is the period in which the changes in the environment actually caused exit). That is the assumption that the unobserved characteristic's contribution to price is constant over time seems inconsistent with the data. Second, note that the average change in the residual of all the continuing goods is also negative (though not nearly as negative as that for continuing goods) and this result is also significant (though only marginally so). It indicates that the new goods that enter have unobserved characteristics that, on average, have larger values then do those of continuing goods (which, given the above discussion, should not be surprising). It also has the implication that we would improve on standard hedonic correction for selection, a correction which ignores unobserved characteristics,

Table 5: **Hedonic Disturbances for About to Exit, Recently Entered, Goods.**

Variable	All Continuing	a-Exit	r-New	Remaining Goods.
Using the S9 Specification for the Hedonic Regression[1].				
mean	-.0028	-.0150	-.0050	-.0021
s.d. of mean	.0017	.0028	.0025	.0021
s.d.(across months)	.0091	.0151	.0132	.0113
percent < 0	.6207	.8621	.5517	.6552
Using a Local Linear Kernel Regression for the Hedonic[1].				
mean	-.0023	-.0133	-.0026	-.0025
s.d. of mean	.0015	.0023	.0024	.0017
s.d.(across months)	.0081	.0126	.0130	.0093
percent < 0	.6897	.7931	.6552	.6552

[1] See the description of the S9 specification and the local linear regression in the text.

by making an adjustment for the change in the market's value of the unobserved characteristics of the exiting goods equal to the measured change in the evaluation of the unobserved characteristics of the *continuing* goods – a point we come back to in a more formal way below.

The results in this section make it doubtful that we can get an adequate correction for the selection bias in exiting goods by simply re-evaluating the observed characteristics of those goods in the period in which they exit. We now consider alternative adjustments for selectin.

5.1 Hedonic Bounds With Unobserved Characteristics.

Adding an i subscript to differentiate goods and a t subscript to differentiate time periods, our hedonic equaton (equation 5) becomes

$$p_{i,t} = z_i \beta_t + \eta_{i,t}. \tag{6}$$

Though the text maintains the linearity assumption in equation (6), we shall also present indices based on non-parametric (local linear kernel) estimates of the needed hedonic functions, so nothing will depend on linearity. However we do maintain the assumption that the disturbance in both specifications is additive and mean independent of z_i. So now our interpretation of $\eta_{i,t}$ is not that it is the unoberved characteristic per se, but rather that it is the residual from projecting the price onto the observed characteristics. Also, at some risk of confusion, from now on we index the sampling intervals by $t = 1, 2, \ldots$, so the difference between $t = 2$ and $t = 1$ is actually two months.

Our problem is that we do not observe the value of $p_{i,t+1}$ for the goods that exit between t and $t+1$. This section introduces method for predicting $p_{i,t+1} - p_{i,t}$ conditional on $(z_i, \eta_{i,t})$ which, given our assumptions, maintains the hedonic bound in the sense that the resultant predictor for $p_{i,t+1} - p_{i,t}$ will have an expectation which is larger than the expectation of $p_{i,t+1} - p_{i,t}$ conditional on z_i and $\eta_{i,t}$.

We develop three bounds. The first only uses the information in the bimonthly sample. The second and third add information from the monthly sample on prices at the end of the first month of the bimonthly sampling period for the goods that exited during the second month of that period. The second only adds the information on the probability of these goods changing prices in the first month, whereas the second also uses the information on the prices these goods changed to. We conclude by noting that there is another computation which, at least intutitively, should provide a lower bound to the average price adjustment over the entire sample period, and, at the very least, it can be used to check the robustness of our results.

5.1.1 Hedonic Bounds From the Bimonthly Data.

From equation (6) we have

$$E[p_{i,t+1} - p_{i,t} | z_i, \eta_{i,t}] = z_i[\beta_{t+1} - \beta_t] + E[\eta_{i,t+1} - \eta_{i,t} | z_i, \eta_{i,t}], \tag{7}$$

where

$$E[\eta_{i,t+1} - \eta_{i,t} | z_i, \eta_{i,t}] = \sum_{j_{i,t+1}} E[\eta_{i,t+1} | j_{i,t+1}, z_i, \eta_{i,t}] Pr\{j_{i,t+1} | z_i, \eta_{i,t}\} - \eta_{i,t}.$$

We can estimate the probability of continuing and the expected change in η when the good continues (or $j_{i,t} = c$). However to get an upper bound for the price index we also need an upper bound for the conditional expectation when $j_{i,t} = x$. To obtain this bound we need a rule for when the good exits. The rule we use is contained in the following assumption.

Assumption 1 (Exit Rule.)

$$j_{i,t} = x \quad \Leftrightarrow \quad \eta_{i,t+1} \leq \underline{\eta}_{t+1}(z_i). \qquad \spadesuit$$

Assumption 1 implies that

$$E[\eta_{i,t+1} - \eta_{i,t} \mid j_{i,t} = x, z_i, \eta_{i,t}] = E[\eta_{i,t+1} - \eta_{i,t} \mid \eta_{i,t+1} \leq \underline{\eta}_{i,t+1}(z_i), \eta_{i,t}, z_i] \leq \tag{8}$$

$$E[\eta_{i,t+1} - \eta_{i,t} \mid \eta_{i,t+1} \geq \underline{\eta}_{t+1}(z_i), \eta_{i,t}, z_i] = E[\eta_{i,t+1} - \eta_{i,t} \mid j_{i,t} = c, z_i, \eta_{i,t}] \equiv gb(z_i, \eta_{i,t+1}).$$

That is the assumption guarantees that the conditional expectation of $\eta_{i,t+1} - \eta_{i,t}$ for continuing goods provides a bound for the conditional expectation of $\eta_{i,t+1} - \eta_{i,t}$ for exiting goods.

As noted above Assumption 1 would be true were we to assume that goods exit when their price is lower than their marginal cost. Moreover assuming exiting goods are more similar to about to exit goods than continuing goods, the assumption is also consistent with the table 1 fact that price changes of about to exit goods are lower than price changes of continuing goods. We place no restrictions on $\underline{\eta}_{t+1}(z_i)$ and estimate separate non-parametric functions for the needed conditional expectation in every period (see below)[14].

Note that there is a very real sense in which (8) is unlikely to be a very "tight" bound. Recall from the last subsection that the data indicate that the unobservables for exiting goods had: (i) systematically lower values of $\eta_{i,t}$ and (ii) systematically lower values of $\eta_{i,t+1} - \eta_{i,t}$ given $\eta_{i,t}$. The bound from equation (8) will make a correction for the lower values of $\eta_{i,t}$ of exiting goods, and for the negative trend in $\eta_{i,t+1} - \eta_{i,t}$ of continuing goods (see table 5). However it does not acount for the fact that the $\eta_{i,t+1} - \eta_{i,t}$ for exiting goods tends to be less than that for continuing goods (again see table 5).

This source of the upward bias in the bound in equation (8) can, at least in principal, be corrected if we are willing to make one more assumption; that the stochastic process generating

[14]The underlying assumption here is that there is a single index of unobserved characteristics which can summarize their impact on the exit decision. The obvious generalization is to allow for multiple indices of unobserved characteristics and relate the values of those indices to the exit decision. This would require a significantly more complex set of assumptions and computational algoirthms; see for e.g. Bajari and Benkard (2005a).

Table 6: **Predicting $\eta_{t+1} - \eta_t$ for Continuing Goods in the Bimonthly Sample**.

Condition on	z		(z, η_t)		(z, η_t), r-New.	
Goods/Mean	R^2	Adj. R^2	R^2	Adj. R^2	R^2	Adj. R^2
all continuing	.15	.10	.27	.18	.28	.19
nonsticky-only	.16	.04	.43	.20	.47	.21

η is Markov and *independent* of z. Recall that our prediction for price conditional on z_i is a regression function, so each period's $\eta_{i,t}$ is mean independent of z_i by construction. The additional assumption we need for the tighter bound corresponds to the movement from mean independence to full independence. Appendix 2 shows that with the Markov assumption a procedure analogous to that used to correct for selection in production functions by Olley and Pakes (1994) can be used to tighten our bound (this without restricting the Markov process in any way). However when we tried to implement this procedure we found that the estimates we obtained were quite unstable. There are two possible reasons. First the additional assumption could be inappropriate. Second, as we explain in appendix 2, the tighter bounds, even if appropriate, are quite sensitive to estimation error. Since our intention is to produce a bound which is both robust and can be automated for use by BLS analysts, we ignore the Olley-Pakes bound below. This does however leaves us with a bound which we know is not tight, and is part of the motivation for turning to the monthly data in the next subsection.

Table 6 presents the R^2's from regressing $\eta_{t+1} - \eta_t$ on a polynomical in η and z for continuing goods on the bimonthly data. The η's used here are the residuals from non-parametric cross sectional hedonic regressions done separately for each period. As a result the η's from the full sample are mean independent of the z's by construction. So if selection were not partially based on the "unobserved" characteristics (our η), we would expect the first set of regressions to have adjusted R^2's of zero. In fact they are highly significant is evidence that the selection into continuing goods is at least partly based on our unobservables, as is modelled above.

Formally we can replace the expectations in equation (8) with expectations conditional on anything in the information set in period t that help predict $\eta_{t+1} - \eta_t$ and the inequality still holds. So we want to condition on anything that will help in that prediction. As noted earlier the fixed effect assumption, i.e. that the $\eta_{i,t}$ are time invariant, is at odds with the data. Accordingly when we use η_t as a predictor we get a significant increase in the fit of our regression, so we include it below. Also we noted earlier that the regression function for $\eta_{i,t+1} - \eta_{i,t}$ for newly enterred goods might be different than that for other continuing goods. So we did the $\eta_{i,t+1} - \eta_{i,t}$ regression once using a dummy for newly enterred goods and once not. We get a small improvement in fit with this dummy, and hence use estimates of that allow this dummy in what follows (though we get very similar results when the prediction without this dummy are used).

Note that once we include η_t the fit of the regression for $\eta_{t+1} - \eta_t$ for the continuing goods whose prices change is noticeably better than the fit of that same regression for all continuing goods. The value of $\eta_{t+1} - \eta_t$ for the goods whose prices do not change is, by definition, $p_{i,t} - z_i\beta_{t+1} - \eta_{i,t}$ (or its non-parametric analogue). So instead of estimating $gb(\cdot) = E[\eta_{i,t+1} - \eta_{i,t} \mid z_i, \eta_{i,t}, j_{i,t} = c]$ directly, we can make use of the information in the price change variable, and obtain what is likely to be a more precise estimate of the $\eta_{t+1} - \eta_t$. That is, if we let $q \in \{\Delta, s\}$ indicate whether a good's price changes ($q = \Delta$) or stays the same ($q = s$), then our bound can be rewritten as

$$gb(z_i, \eta_{i,t}) = \sum_{q \in \{\Delta, s\}} E[\eta_{i,t+1} - \eta_{i,t} \mid q, j_{i,t} = c, z_i, \eta_{i,t}] Pr\{q \mid j_{i,t} = c, z_i, \eta_{i,t}\} \tag{9}$$

$$= \Big(E[\eta_{i,t+1} - \eta_{i,t} \mid q = \Delta, j_{i,t} = c, z_i, \eta_{i,t}] - [p_t - z_i\beta_{t+1} - \eta_{i,t}] \Big) Pr\{q = \Delta \mid j_{i,t} = c, z_i, \eta_{i,t}\} + [p_t - z_i\beta_{t+1} - \eta_{i,t}].$$

Our estimated bounds are found by substituting nonparametric estimates of $E[\eta_{i,t+1} - \eta_{i,t} \mid q = \Delta, j_{i,t} = c, z_i, \eta_{i,t}]$ and of $Pr\{q = \Delta \mid j_{i,t} = c, z_i, \eta_{i,t}\}$ for their true values in equation (9).

5.1.2 Hedonic Bounds Using The Information in the Monthly Data.

The monthly data contains the probability of a price change and the actual changes in the first month of the two month sampling period for about half of the exiting goods (the half that exited during the second month of that period). In using that information we again apply Assumption 1, but this time to the first month of the sampling period. That is we assume that the expected price change over the two month sampling period of goods which continue into the second month of that period is greater than or equal to the expected price change of goods which exited in the first month of the sampling period. Formally if we let $j^- = x$ ($j^+ = x$) denote the event that the good had exited by the end of the first (second) month of the sampling period we assume that

$$E[\eta_{i,t+1} - \eta_{i,t} \mid j^+_{i,t} = x, j^-_{i,t} = c, z_i, \eta_{i,t}] \geq E[\eta_{i,t+1} - \eta_{i,t} \mid j^+_{i,t} = x, j^-_{i,t} = x, z_i, \eta_{i,t}],$$

and then develop bounds for the left hand side of this inequality.

Let $q^- = \Delta$, ($q^+ = \Delta$) denote the event that their was a price change in the first (second) month of the sampling period. Then

$$E[\eta_{i,t+1} - \eta_{i,t} \mid j^+_{i,t} = x, j^-_{i,t} = c, z_i, \eta_{i,t}] = \tag{10}$$

$$E[\eta_{i,t+1} - \eta_{i,t} \mid j^+_{i,t} = x, q^-_{i,t} = \Delta, j^-_{i,t} = c, z_i, \eta_{i,t}] P[q^-_{i,t} = \Delta \mid j^+_{i,t} = x, j^-_{i,t} = c, z_i, \eta_{i,t}] +$$

$$E[\eta_{i,t+1} - \eta_{i,t} \mid j^+_{i,t} = x, q^-_{i,t} = s, j^-_{i,t} = c, z_i, \eta_{i,t}] P[q^-_{i,t} = s \mid j^+_{i,t} = x, j^-_{i,t} = c,\ z_i, \eta_{i,t}].$$

There are now two ways of proceeding. First note that the probabilities in equation (10) can be consistently estimated, while our assumptions bound the two conditional expectations by the same conditional expectations for goods that continue, that is by

$$\left(E[\eta_{i,t+1} - \eta_{i,t} \mid j^+_{i,t} = c, q^-_{i,t} = \Delta, j^-_{i,t} = c, z_i, \eta_{i,t}],\ E[\eta_{i,t+1} - \eta_{i,t} \mid j^+_{i,t} = c, q^-_{i,t} = s, j^-_{i,t} = c, z_i, \eta_{i,t}] \right). \tag{11}$$

These two conditional expectations can be consistently estimated. So we can substitute the estimates of both the conditional expectations in equation (11) and the needed probabilities into equation (10) and compute the resultant bound.

This is our method 1 for computing bounds using the monthly sample. Row 2 of Table 7 shows that the probability of a price change in the first month of the sampling period of goods that exit during the second month is almost double that probability for goods that continue, while row 1 shows that the average two-month price relative of continuing goods that do change price in the

Table 7: **Summary Statistics from The Monthly Sample.**

Two-Month Price Relatives for Continuing Goods.		
$q^- =$	Δ	s
1. Average Price Relative	.969	.988
One-Month Price Relatives for First Month.		
	goods exiting in second month	continuing goods
2. Fraction Changing Price	.416	.244
3. Average Price Relative	.973	.993
4. Av. Rel. when $q^- = \Delta$	.933	.974

first month is significantly lower then the average price relative of all continuing goods. So we should expect method 1 to be rather effective at tightening the bound. Note also that, as before

$$E[\eta_{i,t+1} - \eta_{i,t} \mid j^+_{i,t} = c, q^-_{i,t} = s, j^-_{i,t} = c, z_i, \eta_{i,t}] =$$

$$E[\eta_{i,t+1} - \eta_{i,t} \mid q^+_{i,t} = \Delta, j^+_{i,t} = c, q^-_{i,t} = s, j^-_{i,t} = c, z_i, \eta_{i,t}]P[q^+_{i,t} = \Delta| \; j^+_{i,t} = c, q^-_{i,t} = s, j^-_{i,t} = c, z_i, \eta_{i,t}] +$$

$$[p_t - z_i\beta_{t+1} - \eta_{i,t}]\big(1 - P[q^+_{i,t} = \Delta| \; q^-_{i,t} = s, j^+_{i,t} = c, j^-_{i,t} = c, z_i, \eta_{i,t}]\big),$$

so we use the empirical analogue of this equation to compute $E[\eta_{i,t+1} - \eta_{i,t} \mid j^+_{i,t} = c, q^-_{i,t} = s, j^-_{i,t} = c, z_i, \eta_{i,t}]$.

Our method 1 does not use the information available on the actual first month price changes of the goods which exited in the second month and changed their prices in the first month. Moreover rows 3 and 4 of table 7 show that the exiting goods who change price in their first month have price relatives quite a bit lower than the price relatives of goods who change price in the first period but continue after the second period. Our second method for the monthly data is designed to use this information.

Given z_i and the hedonic function at the end of the first month, conditioning on the price at the end of the first month is equivalent to conditioning on the value of η at the end of the first month, which we will denote by $\eta^+_{i,t}$. Note that

$$E[\eta_{i,t+1} - \eta_{i,t} \mid j^+_{i,t} = x, \eta^+_{i,t}, j^-_{i,t} = c, z_i, \eta_{i,t}] \tag{12}$$

$$= E[\eta_{i,t+1} - \eta^+_{i,t} \,|\jmath^+_{i,t} = x, \eta^+_{i,t}, j^-_{i,t} = c, z_i, \eta_{i,t}] \; + \; [\eta^+_{i,t} - \eta_{i,t}].$$

$$\leq \; E[\eta_{i,t+1} - \eta^+_{i,t} \,|\jmath^+_{i,t} = c, \eta^+_{i,t}, j^-_{i,t} = c, z_i, \eta_{i,t}] \; + \; [\eta^+_{i,t} - \eta_{i,t}].$$

Since the inequality holds conditional on each $\eta^+_{i,t}$, it holds when we average over the distribution of $\eta^+_{i,t}$ conditional on $(j^+_{i,t} = x, j^-_{i,t} = c, z_i, \eta_{i,t})$, a distribution which is available in the data. So we form our estimates of $E[\eta_{i,t+1} - \eta_{i,t} \,|\jmath^+_{i,t} = x, j^-_{i,t} = c, z_i, \eta_{i,t}]$ from these averages.

In estimating the expectation, $E[\eta_{i,t+1} - \eta^+_{i,t} \,|\jmath^+_{i,t} = c, \eta^+_{i,t}, j^-_{i,t} = c, z_i, \eta_{i,t}]$, we want to take explicit account of the sales phenomena in the monthly data, as we are worried that if we do not we might inadvertently register price changes which were too large. We define a sale to occur when a price fall in the first month is followed by a price increase of equal magnitude in the second month. Over all 4.2% of continuing goods experience a sale, and this is 17.4% of continuing goods whose prices change in the intial month of the two month sampling period. To insure that we take appropriate accout of sales we let $r = S$ indicate a sale and $r = NS$ indicates no sale, and write

$$E[\eta_{i,t+1} - \eta^+_{i,t} \,|\jmath^+_{i,t} = c, \eta^+_{i,t}, j^-_{i,t} = c, z_i, \eta_{i,t}] =$$

$$\sum_{r \in \{S,NS\}} E[\eta_{i,t+1} - \eta^+_{i,t} \,|\jmath^+_{i,t} = c, \eta^+_{i,t}, j^-_{i,t} = c, z_i, \eta_{i,t}, r] Pr[r|\jmath^+_{i,t} = c, \eta^+_{i,t}, j^-_{i,t} = c, z_i, \eta_{i,t}, r].$$

We then estimate both the probability of a sale and the expectation of a price change given no sale separately[15].

5.1.3 A Robustness Check.

Partly as a result of the fact that we know that our bounds are not tight, and partly to check the robustness of our procedures, we also consider an alternative intuitive bound for the price relatives of exiting goods. The alternative assumes that the change in the exiting good's price in the period in which it exits is, on average, at least as negative as it was for the same good in the period prior to them exiting. We expect exiting goods to be goods which are being obsoleted. The intutition

[15] Econometrically the conditional distribution of η_{t+1} has a mass point at $p_t - z_i\beta_{t+1}$, so we estimate both the probability of the mass point, and the expectation conditional on not being at the mass point.

for this bound is that the period in which exit occurs is likely to be the period when the impact of changes in market conditions on the price of the good that exited was particularly sharp.

If we accept this (informal) reasoning we can simply use the price change between periods t and $t-1$ for the unobserved price change for the goods that exit between $t+1$ to t. Of course we can only do this for the goods that exited between t and $t+1$ but *were present* in period $t-1$. This is about 85% of the goods that exit between t and $t+1$. The other 15% enterred between $t-1$ and t and then exited before $t+1$. For this latter group of goods we use one of our other bounds.

5.1.4 A Recap of Our Procedures.

We conclude by summarizing the steps used in producing the bounds used for our indexes.

- In the first step we use all of the data to estimate unrestricted hedonic regression functions for each period.
- The second step we use the residuals from first step regressions to construct one of our three bounds for $E[\eta_{i,t+1} - \eta_{i,t}|j_{i,t}, z_i, \eta_{i,t}]$.
- The third step subtracts the estimated change in residuals from the change in log prices and regresses the result on observed characteristics to obtain estimates of the change in value of those characteristics[16].
- The final step uses the estimated coefficient vectors, the observed z_i, and the bound on $E[\eta_{i,t+1} - \eta_{i,t}|j_{i,t}, z_i, \eta_{i,t}]$ to construct the bound on the price relative of each good in the period t basket and then averages these price relatives.

6 Geomean Indexes for TVs: Empirical Results

Table 8 provides the results from computing alternative indexes with the TV data. It is divided into panels, each of which corresponds to a different procedure for calculating the index. Panel A

[16]In the linear regression case this would produce results which are identical to the results had we used the coefficient estimated in the first step. However this need not be the case with the local linear non-parametric procedures.

uses coefficients from a hedonic regression of the log of price levels on observed characteristics to construct predicted price relatives for both continuing and exiting goods. This panel does not make any correction for unobserved characteristics, so it corresponds to what we have been calling the standard hedonic index. Panel B uses an index constructed by first regressing the change in the log of the price levels (i.e. the log of the price relatives) of the continuing goods on their observed characteristics, and then constructing predicted price relatives for both exiting and continuing goods from these regression coefficients. So this is the panel which provides the results from what we have been calling the fixed effect model.

The next three panels use the three different procedures for correcting for the expected change in the valuation of the unobserved characteristic described in the last subsection (for the second step of our procedure). Panel C is the method which uses only the bimonthly data. It uses a correction for the change in unobserved characteristics obtained from regressing the residuals of the *continuing* goods on their (observed and unobserved) characteristics . Panel D is our method 1 for the monthly data. It uses the probabilities of price change during the first month of the two month sampling period of the goods that exit in the second month, combined with the expected change of the residual over the two month period for the goods that continue conditional on changing (or not changing) their price in the first month. Panel E adds the data on the actual price change in the first month of the two month sampling period of the goods that exit in the second month of that period. We return to panel F below, as it provides our robustness tests.

Unless otherwise indicated the column labelled "hedonic" refers to indexes whose predictions are from linear regressions with the $S9$ regressor while the column labelled hedNP uses local-linear nonparametric regression with the same seven product characteristics as in $S9$. All reported indices are the mean of 29 monthly indexes, multiplied by 1200 to give the implied percentage annual inflation rate.

Panel A provides the comparison between the matched model index and the traditional hedonic for hedonic indexes that are based on two different regressor sets; the S9 and the S24 sets of

Table 8: **Alternative Monthly Indexes for TV**[1]

Index Calculated	matched model	hedonic	hedNP
Panel A: Using Log-Price Regression Fit to All Observations			
hedonic uses S24[2]	-9.93	-10.22	n.c.
S24 % l.t. mm		.45	n.c.
hedonic uses S9	-9.93	-8.51	-8.61
standard deviation	5.89	8.18	7.88
S9 % l.t. mm		.38	.34
Panel B: Using Log-Relative Regression Fit to Continuing TVs Only			
mean	-9.93	-10.33	-10.40
standard deviation	5.89	6.33	6.43
% l.t. mm		.72	.66
Panel C: Using Only Bimonthly Data.			
mean	-9.93	-11.02	
standard deviation	5.89	6.14	
%l.t.mm		.79	
Panel D: Method 1 Using Monthly Data (on Probabilities).			
mean	-9.93	-11.29	
standard deviation	5.88	7.30	
%l.t.mm		.72	
Panel E: Method 2 Using Monthly Data (on Probabilities and Price Changes).			
mean	-9.93	-12.48	
standard deviation	5.89	9.22	
%l.t.mm		.66	
Panel F: A-Exit Price Changes When They Exist and Panel ? Otherwise[3].			
Panel C Otherwise.			
mean	-9.93	-11.76	
standard deviation	5.89	5.84	
% l.t. mm		.79	
Panel D Otherwise.			
mean	-9.93	-11.85	
standard deviation	5.89	6.11	
% l.t. mm		.83	
Panel E Otherwise.			
mean	-9.93	-12.07	
standard deviation	5.89	6.17	
% l.t. mm		.86	

Notes:
1. Above values are implied rates of percent annual change, obtained by multiplying the average monthly index by 1200. Averages are over 31 monthly indexes covering the period from June 2000 to January 2003. n.c. means not calculated.
2. S24 refers to the S24 regressor set. All other indices in this table are based on the S9 regressor set. n.c. means not calclulated because there were too few observations to use non-parametric regression with this many regressors.
3. The average (over all months) fractions of goods that are continuing, exiting-with-a-previous-relative, and exiting-without-a-relative are, respectively, (.793, .171, .036).

regressor sets described in section 4. When we use the S24 regressor set, which is the regressor set that is similar to the regressor set used in the once a year hedonic regressions done by the BLS analyst, the results for the hedonic index are quite close to those for the matched model index. The hedonic does estimate an increased rate of deflation, but the increase is only 3%. This accords with our earlier comments about the similarity of the two procedures.

However, as noted above, time constraints imply that the S24 regressor set could not be used in a hedonic procedure which updated the hedonic regression in each period, and it is only a procedure which does this updating that generates an index which is a bound to the compensating variation we are after. The S9 regressor set could be used in a procedure which updates the hedonic regression each period, but when we use that regressor set the matched-model index registers *more deflation* than any of the hedonic indexes; exactly the opposite what the hedonic selection correction is expected to do. That is when we use the S9 regressor set it seems that the value of the unobserved characteristics of the continuing goods are declining faster than the value of their observed characteristics are[17].

The remainder of the table uses only the S9 regressor set. Panel B reports an index based on a hedonic function estimated from the change in log prices, or the price relatives. This is the regression that would be appropriate if the contribution of unobserved characteristics to price were constant over time. This moves us in the expected direction, but only reports an increase of 1% on the rate of deflation obtained from the standard hedonic procedure. Indeed, as noted in the introduction, traditional hedonic procedures, as represented by either Panel A or Panel B, provide an upper bound to the compensating variation that is not very different from that provided by the matched model index.

Panel C uses uses the prediction in equation (9) to correct for the expected change of the unobservable for all predicted price relatives. This is the prediction that uses only information on the change in the residual of the continuing goods. Now we see a substantial change in the index;

[17]This comment assumes that the value of the observed characteristics of the continuing and exiting goods are declining at similar rates.

it registers a rate of price decline 11% faster than the rate of decline matched model index; an adjustment which is three and a half times the adjustment of the standard hedonic.

Next we move to Panel D. This panel uses the probability of price change in the first month of goods which exit in the second month, in conjunction with the expected change of prices for continuing goods conditional on whether they changed their price in the first month, to compute the adjustment for the expected change of the unobservable for all goods. Then the estimated rate of price decline makes another jump, this time to -11.29%, which is 14% larger than the rate of decline in the matched model index.

Panel E, provides the last of our indexes, the index which incorporates the information on the actual prices of the goods after the first month of the two month sampling period. Use of the price information generates a further sharp increase in the rate of decline in the index, this time to -12.48%. This is a rate of price decline of just over 25% per annum, more than eight times the rate of price decline of the matched model procedure. We note, however, that this index has considerably more variance than the other indexes we have presented, a point we return to below.

What seems clear is that once one uses one of our procedures for accounting for the change in value of unobserved characteristics, the use of hedonic procedures make a marked difference to the estimated index. We now ask whether the results from our procedures are consistent with the other information at our disposal. Panel F is designed to provide an answer to this question.

Panel F builds an its index by averaging; (i) matched model price relatives for continuing goods, (ii) the price relative in the period prior to the period in which goods exit for exiting goods for which there was an *observed* price relative in the period prior to exit, and (iii) estimates of price relatives from one of our other procedures for exiting goods which did not have an observed price change in the period prior to exit. The latter are goods which enterred and exited within the four month period and they account for 12.6% of all exits in our data. Note that this index only uses our assumptions for about 2.5% of the constructed price relatives, the rest of the price relatives are constructed from actual data. On the other hand it substitutes the assumption that the price change in the about to exit period is less than that in the period that the good actually exits. We would expect that, at least on average, the prices fall more sharply in the period in which the good exits as this was the period in which market circumstances combined to actually induce exit. On the other hand there may be period effects that make using the about to exit prices hazardous when constructing the index at the frequency used in the construction of the CPI.

All three of the estimates of the rate of deflation using this "robust" lower bound are between our panel D and panel E rates. That is we obtain rates of price change which are larger (in absolute value) then the indexes which only take account of the *probabilities* of change given in the monthly data, but not as large as the indexes which take account of the *actual price changes* in the first month of the actual period in which they exit (rather than the price change in the period before they exit, as is done in Panel F). This is exactly what we would expect if both our assumptions and the assumption that, on average, about to exit price declines are smaller than price declines in the period in which goods exit, are correct. We take this as strong evidence in support of our assumptions.

We note that substituting the price changes in the year prior to exit for the missing data on the price changes of the exiting goods would provide an easy way for the BLS to correct for the selection bias in exiting goods for most of the missing data. If this were combined with one of our procedures for the price changes of the goods that exit that do not have a registerred price change in the year before exit, it would tighten up the matched model bound significantly. For example if we used Panel E for for those goods we would get a rate of price decline 22% faster than that obtained from the matched model index (which is over seven times higher than the rate of price decline from the standard hedonic). Even if we simply deleted those goods which did not have an observed price change in the period prior to exit we would have gotten a price decline of -11.82% (19% larger than that of the matched model). Moreover the the panel F values also have the lowest standard deviations of any of our procedures. Indeed the standard deviations in Panel F are similar to those gotten from the matched model index. On the other hand if there are strong period effects this procedure, a procedure which mixes price changes from consecutive two month periods, might give a less timely indication of inflationary trends.

We want to emphasythe that throughout we have abided rather strictly to the conditions that insure that our index is an upper bound to the true price index. In particular

- no adjustment at all has been made for the inframarginal rents that accrue to consumers that would have bought a new good at the highest price observed for that good in our sample, and

- all adjustments for the exiting goods price changs used data on the price changes of continuing goods, and all our evidence indicates that continuing goods' prices fall less than exiting goods prices.

Even so we obtain rates of decline in the TV price index which are 20% or more larger than the rates of decline from the matched model index, an adjustment which is about seven times the adjustment from the standard hedonic indexes.

Of course the differences we would find by using our index might well be different for different component indices, and one topic for future research is an analysis what fraction of the component indexes, when corrected for exiting goods biases, are likely to result in changes of this magnitude. What seems clear, however, is that a finding that a standard hedonic index generates values that are similar to the matched model index is not sufficient to insure that the selection bias in the matched model index is small.

7 Comparison to the current CPI method of imputing relatives

To be written.

8 Other Future Research

There are a number of small details we need to examine and a few important tests we need to run before we make the suggestion that the BLS follow our, rather than prior, procedures for constructing this component index. First among the important tests is the need to do an out of sample prediction and compare the result to what the BLS obtained for the out of sample period. So we will write a program which automates the procedure for constructing the $S9$ and NP indices and then apply it the data that has been gatherred by the BLS since August 2001.

Second, we will extend our approach to regional geomean and Laspeyres indexes. The CPI computes 38 regional TV subindexes, and then treats each one as a price relative to be plugged into a national Laspeyres index. Currently the CPI uses geomean regional indexes for TVs, and there are good theoretical reasons for thinking the true index may differ across regions (see Pakes, 2003). We will construct a simple research index that mimics this two-stage procedure.

This will require exponentiating regional log-geomean indexes based on predicted log-relatives. Exponentiating log-geomean indexes converts a zero mean random prediction error in log levels into a residual which is not zero mean and hence must be accounted for in computing the index. This can done with a standard method if prediction errors are homoscedastic; see Pakes (2003). We have generalized this method to heteroscedastic errors, with initially promising results. One alternative here is to use local-linear (or some other non-parmateric method) on levels rather than log-levels. This does away with the need for exponentiating an error, and should be rich enough to give a reasonably accurate picture of the hedonic surface (this would not be the case for the S9 regressor set and linear in levels regressions).

We will also construct Laspeyres price indexes, which, if based on log level regressions also requires exponentiating individual predicted log-relatives. Our Laspeyres experimentation is based on the argument in Pakes (2003) that the upper bound it affords for a true COLI has a quality-change induced upward bias that can be reduced or eliminated by hedonic methods. We will also apply these arguments to a computation of the the first-order effect of quality change on a geomean index. We can do so because a Taylor expansion of a geomean index yields a Laspeyres index as its first-order term. The Laspeyres term has weights that depend only on the expenditure-share weights of the geomean and the vector of relatives around which the expansion is made, freeing us from the difficult problem of calculating true Laspeyres weights.[18]

Finaly we want to study the problem of how to choose the sticky-price parameter s more carefully. As noted we do have credible data-based values for a lower bound for this parameter, but we worry that they may be too conservative, resulting in a misleadingly insufficient correction of the quality-change bias of the matched-model index.

References

Berry, S., Levinsohn, J., and A. Pakes. "Automobile Prices in Market Equilibrium," *Econometrica,* 63, pp. 841-890.

Bajari, P. and L. Benkard. “Demand Estimation with Heterogenous Consumers and Unobsreved Product Characteristics: A Hedonic Approach “(2005) *Journal of Political Economy.*

Bajari, P. and L. Benkard, “Hedonic Price Indexes with Unobserved Product Characteristics, with an Application to PC’s” *Journal of Business and Economic Statistics*, 2005a.

Court, A. "Hedonic Price Indexes with Automotive Examples", in *The dynamics of automobile demand,* General Motors Corportation, 1939, pp. 99-117.

Erickson, T. "On Incorrect Signs in Hedonic Regressions", *mimeo,* 2004, Bureau of Labor Statistics.

Erickson, T., and P. Langohr. "Equilibrium Market Dynamics and the Measurement of Inflation and Quality Change, mimeo, 2006, Bureau of Labor Statistics.

Fan, J., and I. Gijbels. *Local Polynomial Modelling and Its Applications,* Chapman and Hall/CRC, 1996, Boca Raton, Florida.

Konus, A. "The Problem of the True Cost-of-Living Index" translated in *Econometrica 7*, 1924, January, p. 10-29.

[18]BLS interprets the regional geomean indexes as estimates of the true COLI for a representative cosumer with a Cobb-Douglas utility function, which implies expenditure shares that do not depend on price and therefore remain constant over time. In contrast, Laspeyres expenditure-share weights must be updated every period to reflect price change.

Lancaster, K. *Consumer Demand: A New Approach,* Columbia University Press, 1971, New York, New York.

Moulton B., T. Lafleur, and K. Moses, *Research On Improving Quality Adjustment in the CPI: The Case of Televisions*, Proceeding of the fourth meeting of the International Working Group on Price Indices, U.s. Department of Labor, April 1998, Bureau of Labor Statistics.

Olley, S. and A. Pakes. "The Dynamics of Productivity in the Telecommincation Equiptment Industry", *Econometrica*, 1994, pp.

Pakes, A. "A Reconsideration of Hedonic Price Indices with an Application to PC's." *American Economic Review* 93. 2003. pp. 1578-1596.

Pakes, A. and P. McGuire. "Computing Markov Perfect Equilibria; Numberical Implications of a Dynamic Differentiated Product Model" *RAND journal of Economics,* 25, , pp. 555-589.

Pollak, R. *The Theory of the Cost-of-Living Index,* 1989, Oxford University Press.

Triplett, Jack. *The OECD Hedonic Handbook,* forthcoming, The Brookings Institution.

Appendix 1: Local-linear Nonparametric Regression.

Local linear estimation predicts each individual value of the dependent variable with a separate weighted least squares regression. Specifically, the local linear WLS estimator for predicting the log-relative for the q^{th} value of the dependent variable is

$$\hat{\delta}(h,q) = \left(Z'\Omega(h,q)Z\right)^{-1} Z'\Omega(h,q)y, \tag{13}$$

where y is the vector of dependent variables, Z is the matrix of independent variables $\Omega(h,q)$ is a diagonal matrix whose diagonal elements are the weights assigned to each observation for the prediction $\hat{y}_{qt} = z_q\hat{\delta}(h,q)$, where z_q is the q^{th} row of Z. Note that it is indexed by the bandwidth parameter h as well as by q.

The i-th diagonal element of $\Omega(h,q)$ is a decreasing function of the distance between z_q and the z_i of the i-th observation. Specifically,

$$\Omega_{ii}(h,q) \propto \prod_{j=2}^{K} \exp\left\{-\frac{1}{2}\left(\frac{z_{q,j}-z_{ij}}{h \times s_j}\right)^2\right\}, \tag{14}$$

where j indexes the columns of Z, and $s_j^2 = \sum_{i=1}^{n}(z_{i,j}-\bar{z}_j)^2/(n-1)$ is the sample variance of column j, the column mean being $\bar{z}_j = \sum_{i=1}^{n} z_{i,j}/n$.

The bandwidth h determines the rate at which the weights decrease with distance. We let the data select h by cross validation,

$$h = \arg\min \sum_{i=1}^{n}\Big(y_i - z_i\hat{\delta}_{cv}(h,i)\Big)^2,$$

where $\hat{\delta}_{cv}(h,i)$ is defined for each i by deleting z_i from Z and y_i from y and then evaluating (13) with the remaining $n-1$ data points. We use the same h for all q. See Fan and Gijbels (1996) for details of this estimator.

Appendix 2 "Tighter" Hedonic Bounds.

This appendix only assumes $\{\eta_t\}$ follows a Markov process. If for simplicity we assume that process is first order, the formal statement of the additional assumption would be that the stochastic process generating $\{\eta\}_t$ is given by the family of probability distributions

$$\mathbf{F}_\eta = \{F(\ \eta_{t+1} \mid \eta_t\ );\ \eta_t \in \mathcal{R}\}. \tag{15}$$

This plus the exit rule in Assumption 1 implies that the expectation of $\eta_{i,t+1}-\eta_{i,t}$ conditional on survival is given by

$$E[\eta_{i,t+1}-\eta_{i,t} \mid z_i, \eta_{i,t}, j_{i,t}=c] = \frac{\int_{\underline{\eta}_{t+1}(z_i)}\Big[\eta_{i,t+1}-\eta_{i,t}\Big]\, dF(\eta_{i,t+1}\mid \eta_{i,t})}{F\Big(\underline{\eta}_{i,t+1}(z_i), \eta_{i,t}\Big)} \equiv g\Big(\underline{\eta}_{t+1}(z_i), \eta_{i,t}\Big).$$

We have an estimate of $\eta_{i,t}$ from the hedonic regression that uses *all* of the data. However we need an estimate of $\underline{\eta}_{t+1}(z_i)$.

As in Olley and Pakes (1994) the estimate of $\underline{\eta}_{t+1}(z_i)$ is obtained from the exit equation which is given by

$$Pr\{\eta_{i,t+1} \geq \underline{\eta}_{t+1}(z_i) \mid \eta_{i,t}\} = 1 - F\Big(\underline{\eta}_{t+1}(z_i) \mid \eta_{i,t}\Big) \equiv \mathcal{F}(\underline{\eta}_{t+1}(z_i), \eta_{i,t}).$$

The function $\mathcal{F}(\cdot)$ maps values of $(\underline{\eta}_{t+1}(z_i), \eta_{i,t})$ into the interval $(0,1)$ and, provided $F(\cdot|\eta_{i,t})$ has a density which is positive everywhere, is monotone decreasing in $\underline{\eta}_{t+1}(z_i)$ for any given value of $\eta_{i,t}$. This implies that for any $\eta_{i,t}$ there is an inverse which provides $\underline{\eta}_{t+1}(z_i)$ as a funciton of the value of $\mathcal{F}(\cdot)$ and $\eta_{i,t}$. Call that inverse $\mathcal{F}_{\eta}^{-1}$, so that

$$\underline{\eta}_{t+1}(z_i) = \mathcal{F}_{\eta}^{-1}\Big[\mathcal{F}(\underline{\eta}_{t+1}(z_i), \eta_{i,t})\Big],$$

and substitute it into equation (**??**) to obtain

$$E[\eta_{i,t+1} - \eta_t \mid z_i, \eta_{i,t}, j_{i,t} = c] = g(\mathcal{F}_{\eta}^{-1}\Big[\mathcal{F}(\underline{\eta}_{t+1}(z_i), \eta_{i,t})\Big], \eta_{i,t}) \equiv h(\mathcal{F}_{i,t}, \eta_{i,t}), \tag{16}$$

where $\mathcal{F}_{i,t} \equiv \mathcal{F}(\underline{\eta}_{t+1}(z_i), \eta_{i,t})$.

Both $\mathcal{F}_{i,t}$ and $\eta_{i,t}$ can be estimated, and hence, if we temporarily ignore estimation error, can be treated as observable. So we can substitute equation (16) into equation (**??**) to obtain

$$E[p_{i,t+1} - p_{i,t} \mid z_i, \eta_{i,t}, j_{i,t} = c] = z_i(\beta_{t+1} - \beta_t) + h(\mathcal{F}_{i,t}, \eta_{i,t}). \tag{17}$$

This equation can be taken to data, and this would allows us to estimate both the function $h(\cdot)$, and $(\beta_{t+1} - \beta_t)$.[19]

We now move to the prediction for *exiting* goods conditional on both observed and unobserved characteristics. First note that

$$0 = E[\eta_{i,t+1} - \eta_{i,t} \mid z_i] \equiv \mathcal{F}_t(z) E[\eta_{i,t+1} - \eta_{i,t} \mid \eta_{i,t+1} \geq \underline{\eta}_{t+1}(z_i), z_i] + [1 - \mathcal{F}_t(z_i)] E[\eta_{i,t+1} - \eta_{i,t} \mid \eta_{i,t+1} \leq \underline{\eta}_{t+1}(z_i), z_i].$$

Consequently

[19] Formally the estimator is a two-stage semiparameteric estimator. The non-parametric components are the functions $\mathcal{F}(\cdot)$ and $h(\cdot)$ and the parametric components are β_{t+1} and β_t. For econometric details see the review of semiparametric techniques by Newey( ) and the literature he cites.

$$E[\eta_{i,t+1}-\eta_{i,t} \mid \eta_{i,t+1} \leq \underline{\eta}_{t+1}(z_i), z_i] = -\frac{\mathcal{F}_{i,t} E[\eta_{i,t+1}-\eta_{i,t} \mid \eta_{i,t+1} \geq \underline{\eta}_{t+1}(z_i), z_i]}{[1-\mathcal{F}_{i,t}]} \equiv -\frac{\mathcal{F}_{i,t} h(\mathcal{F}_{i,t}, \eta_{i,t})}{[1-\mathcal{F}_{i,t}]}.$$

So the hedonic prediction for the price relatives of exiting goods conditional on both observed and unobserved characteristics could be obtained by

$$E[p_{i,t+1}-p_{i,t} \mid z_i, \eta_{i,t}, j_{i,t}=x] = z_i(\beta_{t+1}-\beta_t) - \frac{\mathcal{F}_{i,t} h(\mathcal{F}_{i,t}, \eta_{i,t})}{[1-\mathcal{F}_{i,t}]}. \tag{18}$$

We found that the estimates we obtained in this way to be quite imprecise and to vary a great deal with the way one estimates the non-paramteric function. Their are two possible reasons. First the independence assumption in equation (15) might be inappropriate. Second in the empirical work $\mathcal{F}_{i,t}$ must be estimated and if its true value of is near one even a small amount of estimation error will cause very imprecise estimates of the truncated expectation.

Appendix 3: Characteristic Data.

The next table defines the characteristics we use and gives their average values for different subsets of the data. All variables are 0-1 dummy variables except screen size and the number of dvd player inputs.

Table 9: **Average Characteristic Vectors for Subsets of TVs.**

characteristic	continue	exit	about to exit	enter
screen size (inches)	29.22	30.74	30.84	30.91
picture in picture	0.28	0.32	0.33	0.34
flat screen (not flat panel)	0.096	0.092	0.095	0.136
Projection TV (rear only)	0.148	0.181	0.188	0.185
High-definition ready (no tuner)	0.069	0.070	0.076	0.098
A prominent "quality" brand	0.232	0.202	0.205	0.209
A prominent "value" brand	0.142	0.145	0.149	0.141
1 extra video input	0.282	0.253	0.253	0.240
2 extra video inputs	0.288	0.310	0.304	0.273
3 extra video inputs	0.268	0.283	0.287	0.333
4 extra video inputs	0.046	0.047	0.049	0.069
No. dvd player inputs	0.442	0.481	0.491	0.613
A 3D comb filter	0.148	0.171	0.179	0.192
wide screen (16:9 aspect ratio)	0.023	0.031	0.035	0.037
mtx sur	0.394	0.410	0.409	0.427
store 1	0.159	0.155	0.153	0.161
store 2	0.205	0.192	0.191	0.206
store 3	0.118	0.114	0.112	0.112
store 4	0.099	0.063	0.065	0.069
New York City	0.105	0.112	0.115	0.107
Chicago	0.058	0.064	0.068	0.059
LA	0.105	0.092	0.095	0.108

Notes: 1. In the regressions the first characteristic is log-screensize; it is unlogged here. 2. Table is the average of the mean characteristic vectors for each of 29 bimonthly intervals t-2 to t: 15 from the odd-month subsample and 14 from the even-month subsample. "continue" indicates all TVs present in both t-2 and t. "exit" are those present in t-2 but not in t. "about to exit" are present in t-2 and t but not in t+2. "enter" refers to TVs present in t but not present in t-2.