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

March 1, 2007

Abstract

Until recently the Consumer Price Index consisted solely of "matched model" component indexes. The latter were constructed by data gatherers who visited stores and compared 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 components of the CPI. This paper explores why.

We look carefully at 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, make it necessary to use a more general hedonic correction 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. Also in high frequency data we need to correct for differential "sticky price" rates among different goods. 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%. When we apply the BLS's correction algorithm to our data we find that it does generate a substantial correction to the matched model index, but one of only 7.8%. 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 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.

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). The modifications take into account both the properties of the monthly and bimonthly data actually used in the construction of the CPI, and characteristics of the non-computer components of that data. We show that this 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 a hedonic for a matched model 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 matched model indexes for those groups. For example despite the twin facts that on average over 20% of our TV 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).[1] We begin by exploring the reasons for this phenomena. We then shows that those reasons

[1] We note that some of the attrition in the sample is *temporary*. That is some 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. A good that

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. This is for two reasons. First 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 the new 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 significantly in the expected direction, it only controls for a small part of the problem. This is for two reasons. First, as stressed in the earlier work, the hedonic regression function exhibits differences over adjacent periods, and as a result the contribution of the unobserved characteristics of a given good to price changes also. This change is not accounted for by the fixed effect procedure and also changes the coefficients of the observed $x's$ in that regression. Second, the goods that exit have selected values for the unobserved, as well as for the observed, characteristics, a fact we want to incorporate in our indexes. The fact

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. Temporary exits account, on average, for 2.9% of the sample.

that selection is based on unobserved as well as observed characteristics also rules out the use of familiar sample selection correction procedures like the propensity score.

The second reason why standard hedonic procedures do not adequately control for selection has to do with the relatively high frequency of BLS data (where prices are usually 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. If all prices were equally sticky we would not have to adjust our predictions for this fact. 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 paper develops hedonic indexes that partially account for these two phenomena. The procedure we adopt abides by the general conditions discussed in Pakes (2003), but differs in three ways. First 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 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. Second, we write the fitted regression as a function of the rate of price "stickiness" so that we can condition our predictions on estimates of this rate for different subsets of TVs.

Third, we use 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

[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.

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).

There is one other important characteristic of our procedure; it 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 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.

The empirical results show that our procedures make a substantial difference to the TV component index. The matched model's estimate of the average annual rate of price change in the TV market is -10.1%. 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 identical index value (replicating ealier results for such comparison on other component indexes). 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. The index then jumps back up to -10.65%, now surpassing the matched model index by a noticeable amount. When we implement our correction for unobserved characteristics but make no adjustment for the differential sticky price rate of exiting goods, the index jumps further to -11.1%.

We now move to the adjustment for sticky prices. The sticky price rate to use in this adjustment is not obvious. We both expect, and will present evidence that, the sticky price rates for the goods that exit is less than the sticky price rate for goods in the period prior to the period they exit. This is because the exit period is likely to be a period where large changes occur in the valuations of products with characteristics similar to those of the exiting goods. So if we use the sticky price rate of goods just prior to exit it should give us a conservative index. When we use that sticky price rate the index jumps to -11.5%. If we assume lower sticky price rates for exiting goods the index increases monotonically going all the way to between -14.5% and -15.5% when we assume that none of the goods that exit are goods with sticky prices (depending on functional form assumptions).

About a quarter of the BLS sample has a one month, rather than the standard two month, sampling period. This subsample contains both the one month sticky price rates, and the one month rate of price change for the non sticky prices, of the goods that exit by $t+2$ but are still in the sample by $t+1$. These rates can then be adjusted to get estimates of the same rates for exiting goods over our two month sampling interval. As expected the sticky price rate obtained from these calculations is lower than that for the about to exit goods and the rate of price decline is greater (though the latter difference is not statistically significant). Just adjusting the prior results for the new sticky price estimate produces an index of -12.1 to -12.0% (again depending on functional form assumptions). So we conclude that corrections to the index would reduce the rate of inflation by over 20%.

As noted we believe that it would be possible for the BLS to use our procedure to produce components indexes for the CPI. However, we have not yet analyzed just what effect that would have on the CPI. Since we think the effects are likely to be large enough to have some impact on quantities of importance to the economy, we intend to pursue this question in later versions of the paper. Before doing that, however, we wanted to consider whether our figures are consistent with the other information available to us.

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 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 just over -12%.[3] This reinforces our belief that the correct correction for the annual rate of price decline for TV's is over 20% (i.e. from -10 to -12%).

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

[3] A small fraction of the goods that exit do not have a price change in the period prior to exit, because they are also recent entrants (see below). For these goods we use the prediction developed in the paper for the price change.

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: Charactersitics of the Data.

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 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. 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

[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.

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 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

[5]These numbers 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.

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 only is based on the data from the quarter of the sample with monthly observations. By calculating the sticky price and price relative decline rates between t and t+1 for goods that exit between t+1 and t+2 in this data we get an idea of the rates for goods which do exit over our sampling period during the period 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 difference is not statistically significant).

52% of the monthly observations in period t that exit before t+2 have observed prices in period t+1. The average price decline over the one-month period is .9756. This translates into a two month average price decline of $(.9756)^2$ =.9518 with a standard error of .0136. That is the rate of price decline of the goods that exit over the time period in which they exit seems larger (in absolute value) then it was when these same goods were on the verge of exit; though the difference is not statistically significant. In addition 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). These findings should not be surprising since the period in which these goods exit is likely to be the period in which they are under increased price pressure from competitors. Still it does seem that the absolute values of the price declines of the about to exit goods and of their rates of price

[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.

stickiness are *lower* bounds to the price declines and sticky price rates in the period in which they do exit. Moreover we have a pretty precise estimate of the difference in the sticky price rates.

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 from t to $t+2$ and from $t+2$ to $t+4$ to the matched model index from t to $t+4$. The latter does not contain the price changes of the goods that are on the "verge of exit" in period t, and are "recently new" in period $t+2$. Both these subgroups of goods have prices that fall at a faster rate then a randomly drawn continuing good. That is the t to $t+4$ index misses two groups of prices changes whose prices are falling disproportionately. So the longer the sampling interval the larger the upward biased expected in matched model indexes. To get some indication of how the bias 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

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 t to t+2	v-exit t to t+2	t to t+1 & exit by t+2	implied t to t+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

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.

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

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 differe in that the latter weights with the covariance matrix across periods.

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 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].

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 $t-2$.

The hybrid indexes introduced in Pakes (2003) impute relatives only for TVs that exit between

[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.

[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.

$t-2$ and t, and use actual price relatives for goods that were available in both periods, i.e.

$$G_t^{hyb} = \sum_{q \in C_{t-2}} w_{q,t-2}^{hed} y_{qt} + \sum_{q \in A_{t-2} - C_{t-2}} w_{q,t-2}^{hed} \hat{y}_{qt}. \tag{4}$$

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.

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

$$d_t = \left(Z_t' Z_t\right)^{-1} Z_t' p_t, \tag{5}$$

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.[11]

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 inovolves only running a pre-programmed algorithm.

[11]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.

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

Regressors	mean R^2	mean adj R^2	min R^2	min adj R^2	mean 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.

4.1 Unobserved Characteristics and Hedonic Regression Functions.

Under standard assumptions on consumer behavior the prices of two goods with identical characteristics should be the same[12]. Thus 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. 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 in correcting for the selection problems induced by exit one might want to pay particular attention to unobserved product characteristics.

Part of the impact of the unobserved product characteristics on price will be capturred by their relationship to 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 and the characterisitcs of the good that exited. However as we now show their are solid economic arguments to lead us to believe that the relationship between the unobserved

[12]For a statement of this property, and a demand estimation algorithm that makes intensive use of it, see Bajari and Benkard, forthcoming

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 regression function is linear and let η measure the contribution of unobserved characteristics to price. Then the hedonic regression function is

$$E[p|z,\eta] = z\beta + \eta, \tag{6}$$

where we have normalized the coefficient of η to be one. The regression function we estimate is the regression of p only on z. To analyze its properties we need the properties of the regression of η on z.

If we let x denote exiting goods, n denote new entrants, and 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 (6) 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 $\underline{p}(z)$. Then

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

In particular when the good's observed characterisics lead to a small $\underline{p}(z) - z\beta$ then the goods that continue will all have values of η which are very high, while if $\underline{p}(z) - z\beta$ 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.

Let period $t + 1$ be the comparison period; the period for which we want to predict exiting good's prices. If selection into continuing goods is a function of unobserved as well as observed characteristics, the prices of continuing goods in $t + 1$ conditional on their characteristics will differ from those of goods that exit in that period for at least two reasons. First the unobserved characteristics of exiting goods in the year prior to exit will differ systematically from those of the continuing goods (in our notation the values of η_t will differ). Second the increment in the value of the unobserved characteristics will differ systematically between the two groups ($\eta_{t+1} - \eta_t$ will differ).

To see whether the first effect is likely to be empirically important we calculated the average difference between the residuals of goods which continued and those that exited in the year prior to exit. It was $-.017$ with a standard error of .007; clearly negative and both statistically and economically significant. So if η_{t+1} is positively correlated with η_t a hedonic prediction for the price of exiting goods which did not account for the contribuion of unobserved characteristics would be upwardly biased.

We now check to see if there is reason to believe the increment in the unobserved characteristics differs between continuing and exiting goods. Note 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. That is under the fixed effect assumptions we can form an unbiased prediction for exiting goods prices by regressing the log of the price differences (or of price relatives) for the continuing goods onto their characteristics, and then using that regression function to predict price change for the exiting goods. This because the assumption guarantees 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, and there is no systematic difference in the change in contribution of unobserved characteristics to price.

Note however that this procedure depends critically on the assumption that the unboserved characteristics that influences price has exactly the same impact on price in the two periods; a condition which we know is not true for observed characteristics. To see whether, nevertheless, this assumption is likely to be appropriate for our data, we split all continuing goods into three groups; the goods about to exit (they exit during the next sampling interval), those recently new (they enter in period t), and the remaining goods. We then calculated the average change in the residual for each group in each period.

The results are presented in table 5. Interestingly the average change in the residual of all the continuing goods is slightly negative, though this result is only marginally signficant. This indicates that the new goods that enter in period $t + 1$ have unobserved characteristics that, on average, make a larger contribution to price then do those of the continuing goods (which, given the above discussion, should not be surprising). 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. Since the contribution of the unobserved characteristic to price is falling just prior to exit, we expect it to fall during the exiting period (and as noted previously, probably at a faster rate). That is the assumption that the unobserved characteristic's contribution to price is constant over time seems inconsistent with

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.

the data.

Note also that the results in this section mreinforce the selection argument that lead us to worry about matched model indices in the first place. I.e. we have shown that the market's evaluation of the unobserved characteristics of products that are about to exit are more negative then those of other goods and are falling at a disproportionate rate.

4.2 Hedonic Bounds in the Presence of Unobserved Characteristics.

At the risk of a slight abuse of notation, we write our hedonic equation for evaluating the observed characteristics of all goods marketed in each period as

$$p_{i,t} = z_i \beta_t + \eta_{i,t} \tag{7}$$

Note that $\eta_{i,t}$ is now not the unoberved characteristic per se, but rather it is the residual from projecting the unobserved characteristic onto the observed characteristics (this is the source of notational abuse).

Our problem is that we do not observed the value of $p_{i,t+1}$ for goods that exit between t and $t+1$ and this makes it difficult to get an estimate of $\eta_{i,t+1}$. This section will introduce a methodology for predicting $\eta_{i,t+1}$ which 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 $z_i, \eta_{i,t}$ *and* the fact that the good exited between $t+1$ and t. That is if we let $j_{i,t} = x$ denote the event that a good exits between t and $t+1$, what we require is an upper bound for

$$E[\eta_{i,t+1} - \eta_{i,t} | z_i, \eta_{i,t}, j_{i,t} = x]. \tag{8}$$

To evaluate this expression we need a model for exit. Letting $j_{i,t} = c$ denote the event that a product continues in operation, we will assume that

Assumption 1 (Exit Rule.)

$$j_{i,t} = c \;\; \Leftrightarrow \;\; \eta_{i,t+1} \geq \underline{\eta}_{t+1}(z_i). \quad \spadesuit$$

That is a good with observed characteristics z exits only if $\eta_{i,t+1} \leq \underline{\eta}_{t+1}(z_i)$. We place no restrictions on $\underline{\eta}_{t+1}(z_i)$, and then estimate it non-parameterically (using a different non-parametric function in each period). So this exit rule is consistent with all exit models we are aware of.

Thia assumption implies that

$$E[\eta_{i,t+1} - \eta_{i,t} \mid z_i, \eta_{i,t}, j_{i,t} = x] = E[\eta_{i,t+1} - \eta_{i,t} \mid \eta_{i,t+1} \leq \underline{\eta}_{i,t+1}(z_i), \eta_{i,t}, z_i] \tag{9}$$

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

$$\equiv g_{t+1}\Big(z_i, \eta_{i,t}\Big).$$

The $g(\cdot)$ function can be estimated non-parametrically by regressing the estimates of $\eta_{i,t+1} - \eta_{i,t}$ for continuing goods on z_i and $\eta_{i,t}$.

It follows from equation (9) that

$$E[p_{i,t+1} \mid z_i, \eta_{i,t}, j_{i,t}] \leq \hat{p}_{i,t+1} \equiv z_i \beta_{t+1} + g_{t+1}\Big(z_i, \eta_{i,t}\Big) + \eta_{i,t}, \tag{10}$$

whether $j_{i,t} = x$ or $j_{i,t} = c$. That is equation (10) gives us a bound on expected price for *both* continuing and exiting goods. We estimate β_{t+1}, $\eta_{i,t}$ and $g(\cdot)$, substitute the estimates into equation (10), and use the resulting equation for our hedonic price predictions.

To seee how equation (10) relates to the previous discussion 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 prediction from equation (10) will make a correction for the lower values of $\eta_{i,t}$ of exiting goods, but it does not take into account the lower values of $\eta_{i,t+1} - \eta_{i,t}$ conditional on $\eta_{i,t}$.

This source of the upward bias in the bound in equation (10) can, at least in principal, be corrected if we are willing to make one more assumption; that the stochastic process generating η is Markov and *independent* of z. If for simplicity we assume it is a first order Markov process, the formal statement of the additional assumption would be that the stochastic process generating $\{\eta\}_t$ is given by the family of probability distributions

$$\mathbf{F}_\eta = \{F(\ \eta_{t+1} \mid \eta_t\ );\ \eta_t \in \mathcal{R}\}. \tag{11}$$

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. This implies that the additional assumption we need for the tighter bound corresponds to the movement from mean independence to full independence. No further restrictions on the functional form of the Markov process are required.

With this assumption a procedure similar to the procedure used to correct for selection in production functions by Olley and Pakes (1994) can be used to tighten our bound. Olley and Pakes develop an expression for the expected value of the disturbance conditional on continuing. Appendix 2 shows how to use the same type of analysis to produce the expected value of the disturbance conditional on exiting and then generate a three step estimator for this expecation which, modulus estimation error, should generate a sharper bound for the the expectation of $\eta_{i,t+1}$ when $j_{i,t} = x$ then the one given in equation (10).

However when we tried to implement the procedure developed in Appendix 2 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 the the BLS analysts, we shall ignore the tighter bound in what follows.

A Robustness Check. Partly as a result of the fact that we know that the bound in equation (10) is not tight, and partly to check the robustness of our procedure, we also consider an alternative predicition for obtaining the prices of exiting goods. The alternative simply 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 average of the same goods in the period prior to them exiting. In this case we simply use the price change between periods t and $t-1$ for the 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 always use the inequality in equation (10).

4.2.1 Preliminary Results on Equation (9).

Table 6 presents the R^2's from regressing $\eta_{t+1} - \eta_t$ on a polynomical in η and z for continuing goods; that is it provides the R^2's from regressions used to form $g(\cdot)$ in equation (9). Recall that since the η's are from the hedonic regression, the full sample of η's are mean independent of the z's by the properties of that regression function. So if there were no selection problem we would expect the first set of regressions to have adjusted R^2's of zero. In fact they are highly significant which is evidence that the selection into continuing goods is at least partly based on the unobservables, as is modelled above. The prediction for $\eta_{t+1} - \eta_t$ improves noticeably when we use η_t as a predictor, so our ability to condition on the η_t prior to exit will be helpful in forming our predictions.

Two smaller points about this table are worth noting. First the regressions are meant to control for selection, and the appropriate selection correction for newly enterred goods might be different than for other continuing goods. So we did this once using a dummy for newly enterred goods and once not. We get a small improvement in fit with the dummy, and hence will use the estimate

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

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

of $g(\cdot)$ that includes this dummy in what follows (though we get very similar results when the prediction without this dummy are used). The second point to note is that the selection correction fit is better for goods in the sample of goods that actually change their prices than in the overall sample (i.e. the sample which includes goods with sticky prices). We come back to this presently.

We conclude by summarizing the steps used in producing an index based on the predictions in equation (10).

- First we use all of the data to estimate unrestricted regression functions for each period.
- In the second step we use the residuals from those regressions to estimate equation (9) non-parametrically. This can only be done for the continuing goods and it gives us our estimate of $g(\cdot)$.
- To obtain our estimate of the coefficient in the comparison period we take the continuing goods and regress the change in their (log) price minus our prediction for their change in η on their characteristics[13].
- Finaly use the estimated change in β and $g(\cdot)$ to construct the price predictions for exiting goods.

[13]One could also get the change in coefficients from the level price regressions. If all regression functions are estimated using OLS, the two procedures produce exactly the same result. This is not the case, however, when we use non-parametric estimates, in which case they can differ depending on the details of the non-parameteric procedures used.

4.3 Conditioning on the sticky-price rate

We noted that many of the prices are sticky, i.e. they do not change in value across two periods. As a result the hedonic price relative equation (and consequntly the hedonic "η-relative" equation), are quite different for the sticky and non-sticky price goods. As seen in Table 1 the prices for about to exit goods in the two month sample, and that of the exiting goods in the month before they exit in the one-month sample, are both noticeably less sticky then prices of a randomly drawn good. This, together with our a priori reasons for expecting exiting goods to be less sticky, leads us to an adjustement to $g(\cdot)$ which accounts for differences in sticky price rates between exiting and continuing goods.

Consider a polynomial regression function, and partition the continuing observations so that $\eta = (\eta_1', \eta_2')'$ where η_1 is the $n_1 \times 1$ vector associated with nonsticky prices, while η_2 is the $n_2 \times 1$ vector for the sticky. Let Z be the polynomial terms formed from (z, η). If we sort and partition Z so that it is conformable to η, we can write the estimated coefficients for $g(\cdot)$ as

$$d = \left[Z_1'Z_1 + Z_2'Z_2\right]^{-1} \left[Z_1'Z_1 d_1 + Z_2'Z_2 d_2\right], \tag{12}$$

where

$$d_j = \left(Z_j'Z_j\right)^{-1} Z_j'\eta_j, \quad \text{for} \ \ d = 1, 2. \tag{13}$$

Equation (12) implies that the coefficient vector used to build $g(\cdot)$, i.e. our d, is a matrix-weighted average of d_1 and d_2.

Using d for prediction of η changes is appropriate for imputing the relative of a good that is randomly selected from the distribution of all continuing TVs. If instead we know that a good is randomly selected from the distribution for continuing TVs that have non-sticky prices, then we would want impute its relative using d_1. There is, however, a question of what should we do if we know only that the good is randomly selected from the distribution of exiting TVs. The fact that our prior results indicate that exiting goods prices are less sticky than those of continuing goods, suggests we should make price predictions for exiting goods somewhere "between" d_1 and d_2, but we need to be more precise than that.

So we introduce a family of coefficient estimators indexed by s, a "sticky-price" rate, defined as

$$d(s) = \left[(1-s)\left(\frac{Z_1'Z_1}{n_1}\right) + s\left(\frac{Z_2'Z_2}{n_2}\right)\right]^{-1}\left[(1-s)\left(\frac{Z_1'Z_1}{n_1}\right)d_1 + s\left(\frac{Z_2'Z_2}{n_2}\right)d_2\right] \tag{14}$$

where s ranges from 0 to

$$s_0 = \frac{n_2}{n_1+n_2},$$

the sticky-price rate for TVs that continue from $t-2$ to t. Note that $d(s_0) = d_1$ and $d(s=0) = d_2$.

Neither we nor the CPI personnel observe a sticky price rate for exiting goods, so there is no obvious way to chose an s. Consequently our choices for s in the empirical section below will be pragmatic and conservative. One obvious choice for s is the sticky-price rate for "about-to–exit" TVs. Since the date for dividing months is arbitrary, exiting TVs should much more closely resemble about-to-exit TVs than they do other continuing TVs. Another is the sticky price rate during the first month of the one month sample for the corresponding two month period. We will provide the index that corresponds to s_0, the about-to-exit s, and rates lower than that. This will give us a chance to see how sensitive our results are to the choice of s.

5 Geomean Indexes for TVs: Empirical Results

Table 7 gives the results of fitting alternative indexes to the TV data. It is divided into panels, each of which corresponds to a different procedure for calculating the index. Panel A uses a traditional hedonic regression in the log of price levels to construct the hedonic indices it displays. Panel B uses the change in log levels, or the price relative regression, to construct its indices. Panel C uses a combination of the prior period log price changes where available and our prediction (i.e equation 10) where not. Finaly Panel D uses our prediction equation with adjustments for the sticky price rate as explained in section 4.3. The panels differ somewhat in structure, as is needed to convey what we have actually learned from the data.

Table 7: **Alternative Monthly Indexes for TV**[1]

SummaryStat	matched model	hed	hyb	hedNP	hybNP
Panel A: Using Log-Price Regression Fit to All Observations					
hed with S24[2]	-10.11	-10.10	-10.10	n.c.	n.c.
S24 % l.t. mm		.52	.52	n.c.	n.c.
hed with S9	-10.11	-8.85	-9.68	-9.08	-9.71
s.d.	5.75	8.36	5.72	7.88	6.10
S9 % l.t. mm		.39	.36	.39	.39
Panel B: Using Log-Relative Regression Fit to Continuing TVs Only					
mean	-10.11	-10.63	-10.25	-10.68	-10.30
s.d.	5.75	6.27	5.92	6.39	5.98
% l.t. mm		.77	.68	.74	.65
Panel C: Using A-exit Price Changes Where Available, and Equation (10) Where Not[3].					
mean	-10.11	-11.97	-11.48	-12.00	-11.52
use pre-exit s	-10.11	-12.02	-11.53	-12.07	-11.59
s.d.	5.75	6.21	6.08	6.20	6.04
% l.t. mm		.87	.68	.87	.71
Panel D: Using Equation (10) With Different Sticky-Price Rates s[4]					
$s = s_0$	-10.11	-11.05	-10.64	-10.99	-10.61
s.d.	5.75	6.38	6.14	6.38	6.10
%l.t.mm		.74	.74	.68	.68
pre-exit	-10.11	-11.42	-11.01	-11.36	-10.99
$s = 0.5$	-10.11	-11.58	-11.16	-11.46	-11.08
$s = 0.4$	-10.11	-12.12	-11.71	-12.01	-11.63
$s = 0.3$	-10.11	-12.68	-12.27	-12.58	-12.20
$s = 0.2$	-10.11	-13.28	-12.87	-13.18	-12.80
$s = 0.1$	-10.11	-13.97	-13.56	-13.82	-13.45
$s = 0.0$	-10.11	-15.52	-15.54	-14.54	-14.16

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, $(.793, .171, .036)$.

4. s_0 is the current period's sticky price rate. "pre-exit" sets s equal to the sticky-price rate for the preceding period's about-to-exit TVs (the current period's exited TVs). "avg pre-exit ratio" sets s equal to s_0 times the long-run average value (from the relevant bimonthly subsample) of sx_t/s_t, where s_t and sx_t are the unconditional and conditional-on-about-to-exit sticky price rates for period t. 4. The bandwidth used for the upper right entry in the third panel differs from that used for the bottom right entry in the bottom panel.

Unless otherwise indicate *hed* and *hyb* denote hedonic and hybrid indexes using linear regression with the $S9$ regressor set. $hedNP$ and $hybNP$ use local-linear nonparametric regression with the same seven product characteristics as in $S9$. All reported indices are the mean of 31 monthly indexes, multiplied by 1200 to give the implied percentage annual inflation rate.

Panel A allows us to make a 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 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 virtually identical to those for the matched model index. 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. This difference in results from the two regressor sets reflects the larger role for omitted characteristics when we use the S9 regressor set.

The remainder of the table uses only the S9 regressor set. The second panel 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. These alternative hedonic indexes all register small but appreciable reductions in inflation vis a vis both the matched model and the traditional hedonic indices (which implies much larger reductions versus the S9 level hedonics). Notice also that the hedonic indices exhibit only slightly more variance across months then does the matched model indices. As expected (see Pakes,2003) the hybrid indices are in the middle in this regard. We have not yet investigated whether the increased variance is "real" or is an artifact of estimation variance, but whatever the source it is not large enough to worry too much about. All indexes now have a .7 or greater probability of being less than the matched-model. In sum, moving to the price relative regression, that is accounting for constant differences in the contributions of unobservables to price, clearly moves us in the expected direction. What is not clear is whether it moves us far enough in that direction.

Panel C uses the price change in the period prior to the period in which the goods exit as a measure of the price change in the period in which they exit when that is avaialable (it is available for about 83% of the exits) and our prediction, or equation (10), when not. The logic is that the price change for goods that exits should be, if anything, more negative in the period they exit than in prior periods, so these price changes should provide us a lower bound to the rate of deflation. These results are strikingly different and indicate that that there is a bias of over 20% in the matched model indices. Again using the estimated index results in only a slight increase in variance across months over that in the matched model index.

Panel D uses our prediction, equation (10), for all predicted price relatives. The first three rows present the results without any correction for the fact that the prices of goods that are similar to the goods that exit are likely to be less sticky than those of a randomly chosen good. These results are in between the results in panels B and C, with standard deviations across months similar to those in Panel C. This should have been expected since table 1 showed that; (i) about to exit goods have prices that change more often than those of a randomly chosen good, and (ii) when an about to exit good does have a price change that price change is, on average, more negative than that of a randomly chosen good whose price changes.

The next two rows present estimates which partially correct for the first of these two phenomena. It assumes the sticky price rate for exiting goods is the same as it is for the goods that are about to exit (which is about .54). We do this twice; once for the actual pre-exit rate of change and once for the average of the pre-exit rates, but the results are virtually identical. These estimates show a fall in prices that is about .5% less than the estimates which use the about to exit price changes directly. If instead we use the sticky price rate for the first month of the two month sampling interval from the subset of the data that is sampled at one month intervals, which is a sticky price rate just above .4, we get a result which is almost identical to the result in panel C. Again both these results are likely to be lower bounds on the true rate of deflation; the one in this panel because of (ii) above, and the one in panel C because we expect prices to fall more in the period of exit than in the period prior to exit. Apparently the two biases are of a similar magnitude.

The movement of the price index from -10% to below -12% is a movement of more than 20%. This consitutes a major change in the TV component index. Of course the differences we would find in going to our index might well be different for different component indices, and one topic for future research is to get a rough idea of what fraction of the index, when corrected for exiting goods biases, are likely to result in changes of twenty per cent or more.

6 Comparison to the current CPI method of imputing relatives

To be written.

7 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.[14]

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.

[14]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.

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 "(forthcoming) *Journal of Political Economy.*

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.

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

$$\hat{\delta}(h,q) = \left(Z'\Omega(h,q)Z\right)^{-1} Z'\Omega(h,q)y, \tag{15}$$

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,

$$\Omega_{ii}(h,q) \propto \prod_{j=2}^{K} \exp\left\{-\frac{1}{2}\left(\frac{z_{q,j}-z_{ij}}{h \times s_j}\right)^2\right\}, \tag{16}$$

where j indexes the columns of Z, and $s_j^2 = \sum_{i=1}^{n}\left(z_{i,j}-\bar{z}_j\right)^2/(n-1)$ 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,

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

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 (15) 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.

Local-linear regression conditional on the sticky-price rate

Linearity means there is a sticky-price version of (15) also:

$$\hat{\delta}(h,q,s) = \left[(1-s)\left(\frac{Z_1'\Omega_1(h,q)Z_1}{n_1}\right) + s\left(\frac{Z_2'\Omega_2(h,q)Z_2}{n_2}\right)\right]^{-1} \times \tag{17}$$

$$\left[(1-s)\left(\frac{Z_1'\Omega_1(h,q)Z_1}{n_1}\right)\hat{\delta}_1(h,q) + s\left(\frac{Z_2'\Omega_2(h,q)Z_2}{n_s}\right)\hat{\delta}_s(h,q)\right]$$

where s is defined as in the text

$$\Omega(h,q) = \begin{pmatrix} \Omega_1(h,q) & \mathbf{0} \\ \mathbf{0} & \Omega_2(h,q) \end{pmatrix},$$

and

$$\hat{\delta}_j(h,q) = \left(Z_j'\Omega_j(h,q)Z_j\right)^{-1} Z_j'\Omega_j(h,q)y_j. \tag{18}$$

Corresponding hybrid and hedonic indexes that are conditional on s are

$$G_t^{hyb}(s,h) = \sum_{q\in C_{t-2}} w_{q,t-2}^{hed} y_{qt} + \sum_{q\in A_{t-2}-C_{t-2}} w_{q,t-2}^{hed} x_q \hat{\delta}(h,q,s) \tag{19}$$

and

$$G_t^{hed}(s,h) = \sum_{q\in C_{t-2}} w_{q,t-2}^{hed} x_q \hat{\delta}(h,q,s) + \sum_{q\in A_{t-2}-C_{t-2}} w_{q,t-2}^{hed} x_q \hat{\delta}(h,q,s) \tag{20}$$

Appendix 2 "Tighter" Hedonic Bounds.

Using the Markov assumption in equation (11), and the exit rule in Assumption 1, the expectation of $\eta_{i,t+1} - \eta_{i,t}$ conditional on survival is given by

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

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 (9) to obtain

$$E[\eta_{i,t+1} - \eta_t \mid z_i, \eta_{i,t}, j_{i,t} = c] = g(\mathcal{F}_\eta^{-1}\Big[\mathcal{F}(\underline{\eta}_{t+1}(z_i), \eta_{i,t})\Big], \eta_{i,t}) \equiv h(\mathcal{F}_{i,t}, \eta_{i,t}), \tag{21}$$

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 (21) into equation (8) to obtain

$$E[p_{i,t+1} - p_{i,t} \mid z_i, \eta_{i,t}, j_{i,t} = c] = z_i(\beta_{t+1} - \beta_t) + h(\mathcal{F}_{i,t}, \eta_{i,t}). \tag{22}$$

This equation can be taken to data, and this would allows us to estimate both the function $h(\cdot)$, and $(\beta_{t+1} - \beta_t)$.[15]

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

[15] 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.

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

Consequently

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

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

$$E[p_{i,t+1} - p_{i,t} \mid z_i, \eta_{i,t}, j_{i,t} = x] = z_i(\beta_{t+1} - \beta_t) - \frac{\mathcal{F}_{i,t} h(\mathcal{F}_{i,t}, \eta_{i,t})}{[1 - \mathcal{F}_{i,t}]}. \quad (23)$$

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 (11) 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 8: **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.