New Hedonic Quality Adjustment Approaches for the PPI

Timothy Erickson Steven Sawyer*

May 2016

The Producer Price Index of the Bureau of Labor Statistics adjusts price indexes for quality change by using hedonic coefficient estimates to value the change in measured product characteristics between a product that is exiting the index and the product that is substituted for it. The quality adjustment, or "QA," is the value-weighted sum of the characteristic-by-characteristic changes, the weights being the coefficient estimates from a hedonic regression. Over time, perhaps 20-25 different product characteristics are required to compute the QAs for a product class, but for any one QA, that is, for any one instance of the substitution of a new item for an exiting item, it is typical that no more than 5 characteristics are involved, frequently even less. Nevertheless, PPI hedonic regressions are estimated with the full complement of at least 20 regressors. For this reason, only products with relatively large sample sizes are quality-adjusted.

We propose here a small-sample method that would allow extending hedonic quality adjustment to many more PPI products. Specifically, we propose that rather than fit a single large regression to the available data, PPI staff fit a different regression for each individual QA. The regression for a specific QA would include those characteristics that changed between the exit and the substitute, and any additional characteristics that improved the fit. By "fit" we mean estimates of out-of-sample performance provided by cross-validation. Our algorithms for choosing additional characteristics are those from the machine-learning/data-analytic literature. These algorithms can be "automated": the complete list of 20–35 characteristics are inputed, along with the characteristics required for the specific QA, and the output is the hedonic specification that optimizes out-of-sample performance subject to the restriction that the regression include the QA-specific characteristics. In addition to shortening the regressor list by starting with the minimum number of characteristics needed for the specific QA, our variable selection algorithms are known to penalize lengthy regressor lists more than in-sample measures such as adjusted R-squares. The resulting sparsely parameterized hedonic regressions could be applied to smaller sample sizes than PPI currently thinks are the minimum necessary for hedonic regression.

The performance of the QA-specific regressions can be measured in various ways. First, is the index that results from using QAs obtained in this way better than not quality-adjusting at all? Second, how close are the QAs obtained from QA-specific hedonic regressions to QAs obtained in the current PPI fashion ? Third, simulations and resampling methods can be used to evaluate performance. A benefit of our proposed approach is that the machine-learning programs replace valuable human labor and training hours. Another benefit is that the "black box" of hedonic specification-searching currently used in PPI production is replaced by a transparent and reproducible regression methodology.

* All authors are employees of the Bureau of Labor Statistics. All of the analysis, views, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the Bureau of Labor Statistics.