BEA Working Paper Series, WP2024-2

An Application of the Oaxaca-Blinder Decomposition to the Price Deflation Problem

Authors[*] Ana Aizcorbe, U.S. Bureau of Economic Analysis

Jan de Haan, Economic Statistics Centre of Excellence, London, UK

Contact Ana.Aizcorbe@bea.gov

Date June 2024

Abstract We apply the Oaxaca-Blinder decomposition method used in the labor literature to split changes in average prices into inflation and quality components. The inflation measure is a full imputation Törnqvist price index. Using this index to deflate nominal spending properly allocates changes in the quality of goods to changes in real spending, not inflation.

Keywords National accounts, inflation price index, quality index

JEL Code C43, E31

* Declaration of interests: none.

1. Introduction

The fundamental role of price indexes in the national accounts is to deflate spending. That is, to decompose changes in nominal spending into a piece that measures changes in prices—called inflation—and another that measures changes in "quantities" or real spending. In a static economy with a fixed set of goods whose quality remains constant, the inflation piece of this decomposition would measure changes in the prices of identical goods, and the implied real GDP component would measure only changes in the number of goods produced. However, product innovation that brings new and better goods to market is the rule rather than the exception, and in this dynamic setting, the price deflator must properly allocate any changes in quality to real spending, not inflation.

In the labor literature, decompositions have been done to explain differences in average wages across groups of workers. We apply the well-known Oaxaca-Blinder decomposition to the deflation problem. In our context, the method decomposes changes in average prices into a price index and a quality index. We show that the particular price index implied by the method is a (full) imputation Törnqvist price index. The decomposition also provides an explicit expression for the implied quality index and allows us to show that using the full imputation Törnqvist index to deflate nominal spending yields a measure of real spending that includes changes in quality in a sensible way.

We are not aware of other studies that apply the Oaxaca-Blinder decomposition to the deflation problem.

2. Oaxaca-Blinder Decomposition

We apply a decomposition method widely used in the labor literature to explain differences in average wages across groups of workers (Altonji and Blank 1999). The workhorse in this literature is the Oaxaca-Blinder decomposition method (Oaxaca 1973; Blinder 1973). It uses regressions of wages as a function of workers' characteristics for each group. The estimated coefficients are then used to decompose the difference in the average wage received by the two groups into an "explained" component (that measures how much of the difference comes from differences in the observed characteristics) and an "unexplained" component (that measures what is left).

Applied to the price index question, the "groups" are (two) time periods, hedonic regressions are estimated for each period, and the estimated coefficients are used to measure inflation as the difference between the average price of goods sold at time *t* to what the average price would have been for those same goods at time 0. We interpret the remaining changes in average prices from changes in the characteristics as a quality index.

The hedonic specification used in the Oaxaca-Blinder method is a log-linear regression estimated for each time period. In our context, the regression model is:

$$ln\, p_i^t = \alpha^t + \sum_{k=1}^{K} \beta_k^t z_{ik} + \varepsilon_i^t \quad (1)$$

where p_i^t denotes the price of product *i* in period t $(t = 0,1)$, z_{ik} is (the quantity of) characteristic *k* $(k = 1, \ldots, K)$ for product *i*, β_k^t is the coefficient for *k*, α^t is the intercept term, and ε_i^t is an error term with mean 0. The characteristics define the quality of the goods; the coefficients measure the market's valuation of those characteristics.

In the labor literature, this equation is interpreted as a reduced form whose role is to provide price predictions for time *t* prices; this is consistent with the interpretation used in Pakes (2003), Bajari et al. (2023), and Ehrlich et al. (2023). In addition, it is typically assumed that all relevant characteristics are observed, an assumption that some might question in a price context (Erickson and Pakes 2011).

This regression is typically estimated for each time period using weighted least squares using expenditure shares as weights: $s_i^t = p_i^t q_i^t / \sum_{i \in S^t} p_i^t q_i^t$, where the set of products sold in period *t* $(t = 0,1)$ is denoted by S^t, and p_i^t and q_i^t are the price and quantity sold of good *i*. Then, the weighted residuals sum to zero in each period, and the weighted average of (logged) prices equals the weighted averages of the predicted (logged) prices:

$$\overline{ln\, p^t} = \sum_{i \in S^t} s_i^t\, ln\, p_i^t = \sum_{i \in S^t} s_i^t\, \widehat{ln\, p}_i^t \ (t = 0,1) \quad (2)$$

Substituting in the predicted values from the hedonic regression in (1) allows us to write a simple difference in the average prices at *t=1* and *t=0* ($\overline{ln\,p^1} - \overline{ln\,p^0}$) in terms of the estimated hedonic parameters, $\hat{\alpha}^t$ and $\hat{\beta}_k^t$ $(k = 1, \ldots, K)$, and the expenditure-share weighted average characteristics, $\bar{z}_k^t = \sum_{i \in S^t} s_i^t z_{ik}$.

$$\overline{ln\,p^1} - \overline{ln\,p^0} = \alpha^1 + \sum_{k=1}^{K} \hat{\beta}_k^1 \overline{z_k^1} - \alpha^0 - \sum_{k=1}^{K} \hat{\beta}_k^0 \overline{z_k^0} \tag{3}$$

However, this literal read of the change in average logged prices comingles the effects of changes in the coefficients and those of the characteristics.

To separately parse out these effects, the Oaxaca-Blinder method considers two counterfactuals. The first asks how much of the difference in the average prices can be attributed to differences in the mean characteristics *evaluated using t=1 price structure* (i.e., the *t=1* coefficients):

$$\overline{ln\,p^1} - \overline{ln\,p^0} = [(\hat{\alpha}^1 - \hat{\alpha}^0) + \sum_{k=1}^{K} \bar{z}_k^0 \left(\hat{\beta}_k^1 - \hat{\beta}_k^0\right)] + \sum_{k=1}^{K} \hat{\beta}_k^1 \left(\bar{z}_k^1 - \bar{z}_k^0\right) \tag{4}$$

The first two terms (in brackets) measure the effect of changes in the hedonic coefficients at the *t*=0 level for the characteristics ($\bar{z}_k^0$). Similarly, the last term measures the effect of changes in the characteristics using the *t*=1 level for the hedonic coefficients ($\hat{\beta}_k^1$).

However, one could just as easily do a counterfactual using the time *t=0 price structure:*[1]

$$\overline{ln\,p^1} - \overline{ln\,p^0} = [\,(\hat{\alpha}^1 - \hat{\alpha}^0) + \sum_{k=1}^{K} \bar{z}_k^1 \left(\hat{\beta}_k^1 - \hat{\beta}_k^0\right)] + \sum_{k=1}^{K} \hat{\beta}_k^0 \left(\bar{z}_k^1 - \bar{z}_k^0\right) \tag{5}$$

[1] Auno et al (2023) use this counterfactual to justify indirectly calculating a price index by subtracting a quality adjustment based on a *t*=0 hedonic from the average prices on the left-hand side of (5).

Thus, the Oaxaca-Blinder approach provides two decompositions that use different reference points, neither superior to the other. Oaxaca's paper provided both estimates as bounds on what the "true" answer might be. Of the subsequent studies that sought optimal ways to combine the counterfactuals, Reimers (1983) took a simple average of the two counterfactuals:[2]

$$\overline{ln\,p^1} - \overline{ln\,p^0} \;=\; [\,(\hat{\alpha}^1 - \hat{\alpha}^0) + \sum_{k=1}^{K} \frac{(\bar{z}_k^0 + \bar{z}_k^1)}{2} \left(\hat{\beta}_k^1 - \hat{\beta}_k^0\right)] + \sum_{k=1}^{K} \frac{(\hat{\beta}_k^0 + \hat{\beta}_k^1)}{2} (\bar{z}_k^1 - \bar{z}_k^0) \qquad (6)$$

We interpret the term in brackets as a price index that measures the effect of changes in the hedonic coefficients on average prices. The index uses the average level for the characteristics as a reference point and evaluates the effect of changes in the coefficients on average prices at that point. This is what it means to hold quality constant.

This price index turns out to be a full imputation Törnqvist, an index with several desirable properties (Diewert 2023). As with other superlative indexes, the Törnqvist formula allows expenditure shares to change each period, thus better reflecting the composition of goods sold. Second, imputing prices is thought to allow a better way to handle the entry and exit of goods. Indeed, imputation indexes have been used in recent hedonic studies for markets characterized with a high degree of turnover (Bajari et al. 2023; Ehrlich et al. 2023).

To see that (6) is a full imputation Törnqvist, consider the usual formula for the index:

$$ln\widehat{PI}_{FIT}^{01} \equiv \sum_{i \in S^0 \cup S^1} \frac{(s_i^1 + s_i^0)}{2} (ln\,\hat{p}_i^1 - \ln \hat{p}_i^0) \qquad (7)$$

Using the hedonic regression in (1), one can restate the index in terms of the hedonic arguments:

$$ln\widehat{PI}_{FIT}^{01} \equiv (\hat{\alpha}^1 - \hat{\alpha}^0) + \sum_{k=1}^{K} \frac{(\bar{z}_k^0 + \bar{z}_k^1)}{2} \left(\hat{\beta}_k^1 - \hat{\beta}_k^0\right) \qquad (8)$$

[2] Equations like this have been used to make comparisons of different hedonic-based price index methods. For example, Diewert, Silver, and Heravi (2009) compare the expression in (6) for the full imputation Törnqvist and a similar one for the time dummy hedonic index to show conditions under which the two indexes are equivalent.

which gives the bracketed term in the Oaxaca-Blinder decomposition in (6).

We interpret the last term in (6) as a quality index. It measures the effect of changes in goods' characteristics on average prices using the average level of the hedonic coefficients as a reference point. As such, it could be called a "constant price quality index," in that it chooses a fixed level for the coefficients to value the change in quality:

$$ln\widehat{QI}_{FIT}^{01}=\sum_{k=1}^{K}\frac{(\hat{\beta}_k^1+\hat{\beta}_k^0)}{2}(\bar{z}_k^1-\bar{z}_k^0) \quad (9)$$

The quality index reflects changes to the average characteristics from two important sources of quality change: (1) turnover (as new, better goods enter the market and older vintages exit), and (2) shifts in the incumbent goods (as diffusion towards higher-quality goods plays out).

Summing up, the Oaxaca-Blinder decomposition provides a way to separate out the inflation and quality components of average prices using a full imputation Törnqvist price index and its attendant quality index:

$$\overline{ln\,p^1}-\overline{ln\,p^0} \;=\; ln\widehat{PI}_{FIT}^{01}+ln\widehat{QI}_{FIT}^{01} \quad (10)$$

3. Deflation

Changes in spending can arise from changes in average prices and/or changes in the quantities purchased:

$$\ln(EXP^1/EXP^0) = \ln(\overline{p^1}/\overline{p^0}) + \ln(Q^1/Q^0) \quad (11)$$

where $EXP^t = \sum_{i\in S^t} p_i^t q_i^t$ is spending (expenditure) at time t, $\overline{p^t} = \frac{\sum_{i\in S^t} p_i^1 q_i^1}{\sum_{i\in S^t} q_i^1}$ the average price, and $Q^t = \sum_{i\in S^t} q_i^t$ the total number of units sold at time t.

To link the indexes implied by the Oaxaca-Blinder method to this identity, we note that the functional form for average prices in (11) is arithmetic (so that the change in average prices is $\ln\overline{p^1} - ln\overline{p^0}$), whereas that used in the Oaxaca-Blinder decomposition in (10) is geometric (so the change in average prices is $\overline{ln\,p^1} - \overline{ln\,p^0}$).

De Haan (2023) derived the relationship between the two types of means as:

$$\overline{ln\, p^1} - \overline{ln\, p^0} \simeq \left[\ln \overline{p^1} - ln\overline{p^0}\right]\left[\frac{1}{2}((CV^1)^2 - (CV^0)^2)\right] \quad (12)$$

where CV^t $(t = 0{,}1)$ is the expenditure-share weighted coefficient of variation of the price observations (the weighted standard deviation divided by the weighted arithmetic mean). In a time series context, if *the coefficient of variation is relatively small with no trend,* the right-hand side of $\left[\frac{1}{2}((CV^1)^2 - (CV^0)^2)\right]$ will fluctuate around 1 with small bounds and the weighted geometric average prices used in the decomposition will provide a good approximation to the arithmetic average prices in (11).

Thus, the Oaxaca-Blinder decomposition provides an approximation to the decomposition in (11):

$$ln(EXP^1/EXP^0) \simeq \overline{ln\, p^1} - \overline{ln\, p^0} + ln\,(Q^1/Q^0) \quad (13)$$

Using (10) to substitute out the change in average prices and rearranging shows that using the full imputation Törnqvist as a deflator (i.e., $\ln(EXP^1/EXP^0) - ln\widehat{PI}_{FIT}^{01}$) implies a real spending measure that includes both changes in the units sold as well as their average quality:

$$\ln(EXP^1/EXP^0) - ln\widehat{PI}_{FIT}^{01} \simeq ln\widehat{QI}_{FIT}^{01} + \ln(Q^1/Q^0) \quad (14)$$

4. Conclusion

We have shown how the Oaxaca-Blinder decomposition can split changes in average prices into a full imputation Törnqvist price index and a quality index. Using the price index to deflate nominal spending properly allocates the quality index to real spending.

References

Altonji, Joseph, and Rebecca Blank. 1999. "Race and Gender in the Labor Market." In *Handbook of Labor Economics*, vol. 3, Part C, 3143–3259.

Auno, Ville, Henri Luomaranta-Helmivuo, Hannele Markkanen, Satu Montonen, Kristiina Nieminen, and Antti Suoperä. 2023. Paper presented at Meeting of the Group of Experts on Consumer Price Indices, June 7–9, Geneva.

Bajari, P., Z. Cen, V. Chernozhukov, M. Manukonda, J. Wang, R. Huerta, J. Li, L. Leng, G. George Monokroussos, S. Vijaykunar, and S. Wan. "Hedonic prices and quality adjusted price indices powered by AI." 2021. Technical report, The Institute for Fiscal Studies Department of Economics, UCL, CENMAP working paper CWP 04/21.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *The Journal of Human Resources* 8 (4): 436–55. https://doi.org/10.2307/144855.

Cotton, Jeremiah. 1985. "A Comparative Analysis of Black-White and Mexican-American-White Male Wage Differentials." *The Review of Black Political Economy* 13 (4): 51–69. https://doi.org/10.1007/bf02899890.

Diewert, W.E. 2023. "Consumer Price Index Theory: An Introduction." Vancouver School of Economics, Discussion Paper 23-01.

Diewert, W.E., S. Heravi and M. Silver. 2009. "Hedonic Imputation versus Time Dummy Hedonic Indexes," 87–116 in W.E. Diewert, J. Greenlees and C. Hulten (eds.), Price Index Concepts and Measurement, Studies in Income and Wealth, Vol. 70. University of Chicago Press, Chicago.

De Haan, Jan. 2023. "Constructing Stratified Price Indexes from Scanner Data with Limited Product Characteristics," unpublished manuscript.

Ehrlich, Gabriel, John Haltiwanger, Ron Jarmin, David Johnson, Ed Olivares, Luke Pardue, Matthew D. Shapiro, and Laura Yi Zhao. 2023. Quality Adjustment at Scale: Hedonic versus Exact Demand-Based Price Indices.

Erickson, Tim, and Ariel Pakes. 2011. "An Experimental Component Index for the CPI: From Annual Computer Data to Monthly Data on Other Goods." *American Economic Review* 101 (5): 1707–38.

Pakes, A. 2003. "A Reconsideration of Hedonic Price Indexes with an Application to PCs." *American Economic Review* 93(5): 1578–1596, December.

Reimers, Cordelia W. 1983. "Labor Market Discrimination against Hispanic and Black Men." *The Review of Economics and Statistics* 65 (4): 570. https://doi.org/10.2307/1935925.