

## WEIGHTED COUNTRY PRODUCT DUMMY VARIABLE REGRESSIONS AND INDEX NUMBER FORMULAE

BY ERWIN DIEWERT\*

*Department of Economics, University of British Columbia*

The article considers a very simple type of hedonic regression model where the only characteristic of a commodity is the commodity itself. This regression model is known as the country product dummy method for calculating country price parities in the context of making international comparisons. The paper considers only the two country or two period case and introduces value or quantity weights into the regression. The resulting measures of overall price change between the two countries or time periods are compared to traditional bilateral index number formulae. It is shown how the Geary Khamis, Walsh and Törnqvist price indexes can be obtained as special cases of this framework.

### 1. INTRODUCTION

The Country Product Dummy (CPD) method for making international comparisons of prices is a very simple type of hedonic regression model that was proposed by Robert Summers (1973) where the only characteristic of the commodity is the commodity itself. In this article, we will consider how this simple hedonic regression model can be modified for the two country or two period case when quantity or expenditure weights are available in addition to the country (or “model”) prices. We will then relate these weighted versions of the CPD regression model to traditional index number formulae.<sup>1</sup>

In Section 2, we consider the case where *expenditure weights* for the two periods or countries are available, while Section 3 considers the case where *quantity weights* are available.

### 2. THE CPD METHOD WHEN EXPENDITURE WEIGHTS ARE AVAILABLE

If there are C countries in the comparison<sup>2</sup> and N products, the relationship of the prices between the various countries using the CPD assumptions is approximately given by the following model:

*Note:* The author is indebted to Bert Balk, Yuri Dikhanov and D. S. Prasada Rao for helpful comments and to the Social Sciences and Humanities Research Council of Canada for financial support.

\*Correspondence to: Erwin Diewert, Department of Economics, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z1 (diewert@econ.ubc.ca).

<sup>1</sup>Triplett and McDonald (1977, p. 150), Triplett (2000, p. 39) and Diewert (2001) noted how certain hedonic regression models could be related to various unweighted matched model elementary indexes. Thus this note is an extension of this analysis to cover the case where quantity or expenditure weights are also available. Part of the present paper can also be viewed as a specialization of Rao (2005) to the two country case.

<sup>2</sup>Alternatively, there could be C time periods.

$$(1) \quad p_n^c \approx \alpha_c \beta_n; \quad c = 1, \dots, C; \quad n = 1, \dots, N;$$

$$(2) \quad \alpha_1 = 1$$

where  $p_n^c$  is the price (in domestic currency) of commodity  $n$  in country  $c$ . Quantities for each commodity in each country are assumed to be measured in the same units. Equation (2) is an identifying normalization; i.e. we measure the price level of each country relative to the price level in country 1. Note that there are  $CN$  prices in the model and there are  $C - 1 + N$  parameters to “explain” these prices. Note also that the basic hypothesis that is implied by (1) is that commodity prices are approximately proportional between the two countries. Taking logarithms of both sides of (1) and adding error terms leads to the following CPD regression model:

$$(3) \quad \ln p_n^c = \ln \alpha_c + \ln \beta_n + \varepsilon_n^c; \quad c = 1, \dots, C; \quad n = 1, \dots, N.$$

The main advantage of the CPD method for comparing prices across countries over traditional index number methods is that we can obtain *standard errors* for the country price levels  $\alpha_2, \alpha_3, \dots, \alpha_C$ . This advantage of the stochastic approach to index number theory was stressed by Summers (1973) and more recently by Selvanathan and Rao (1994).

We will consider a special case of the above model where there are only two countries but we will generalize the above model to allow for expenditure weights. Our reason for considering only the case of two countries is that we want to compare our weighted CPD estimator for  $\alpha_2$  with traditional bilateral index number formulae.<sup>3</sup> Thus let

$$(4) \quad e_n^c \equiv p_n^c q_n^c; \quad c = 1, 2; \quad n = 1, \dots, N$$

denote the expenditure on commodity  $n$  in country  $c$  and consider the following *weighted least squares model*:<sup>4</sup>

$$(5) \quad \min_{\gamma, \delta} \sum_{n=1}^N e_n^1 [\ln p_n^1 - \delta_n]^2 + \sum_{n=1}^N e_n^2 [\ln p_n^2 - \gamma - \delta_n]^2$$

where  $\gamma \equiv \ln \alpha_2$  and  $\delta_n \equiv \ln \beta_n$  for  $n = 1, \dots, N$ . In order to make the second summation of terms comparable to the first, prices should be measured in the same currency units.<sup>5</sup> In order to justify the model (5) in terms of a traditional stochastic specification, assume that  $\ln p_n^c$  is an independently distributed random variable with mean  $\delta_n$  and variance  $\sigma^2/e_n^1$  and  $\ln p_n^2$  is an independently distributed random variable with mean  $\delta_n + \gamma$  and variance  $\sigma^2/e_n^2$ . An alternative way for justifying the weighted model (5) is to argue that each logarithmic price  $\ln p_n^1$  should be weighted according to its economic importance; i.e. if consumers are spending  $e_n^c$  dollars on

<sup>3</sup>For extensive material on the various approaches to bilateral index number theory and the various index number formula that are “best” for each approach, see Chapters 15–19 in the ILO (2004).

<sup>4</sup>Rao (2002) considers the generalization of this model to the case of many countries.

<sup>5</sup>In the time series context, if there is high inflation going from period 1 to 2, then the weighted least squares model will give too much weight to the second set of terms. In this situation, the absolute expenditure weights (the  $e_n^1$  and  $e_n^2$ ) should be replaced by the period 1 and 2 expenditure shares,  $s_n^1$  and  $s_n^2$ . We consider this model later.

commodity  $n$  in country  $c$ , then  $\ln p_n^c$  should appear  $e_n^c$  times in the regression instead of only once.<sup>6</sup> This assumption also leads to (5).

The solution for  $\gamma$  which solves the above weighted least squares minimization problem is

$$(6) \quad \gamma^* = \sum_{n=1}^N S_n \ln(p_n^2/p_n^1)$$

where

$$(7) \quad S_n \equiv h(e_n^1, e_n^2) / \sum_{i=1}^N h(e_i^1, e_i^2); \quad n = 1, \dots, N \text{ and}$$

$$(8) \quad h(a, b) \equiv [(1/2)a^{-1} + (1/2)b^{-1}]^{-1} = 2ab/[a + b]$$

so that  $h(a, b)$  is the *harmonic mean* of the numbers  $a$  and  $b$ . Thus  $\gamma^*$  is a share weighted average of the logarithms of the price ratios  $p_n^2/p_n^1$ . If  $\gamma^*$  is exponentiated, then an estimator  $\alpha_2^*$  for  $\alpha_2$ , the price level of country 2 relative to country 1, is obtained.

It can be seen that if countries 1 and 2 are interchanged, then the corresponding weighted least squares estimator for the price level of country 1 relative to that of country 2 turns out to equal the reciprocal of  $\alpha_2^*$ , i.e. the exponential of  $\gamma^*$  defined by (6) satisfies the *time reversal test*.<sup>7</sup> This is a very desirable property for an index number formula.

Note that if country 2 is much bigger than country 1, then the expenditure weights of country 2, the  $e_n^2$ , will play a much larger role in the shares  $S_n$  defined by (7) than the expenditure weights of country 1, the  $e_n^1$ . Thus weighting by country expenditures leads to a *plutocratic* type index number formula where the bigger country plays a larger role in determining the relative price level. In order to eliminate this asymmetry and obtain a *democratic* type of index where each country has equal weighting no matter what its size is, we can replace the expenditure weights in the weighted least squares problem (5) by *expenditure shares*. Thus let

$$(9) \quad s_n^c \equiv p_n^c q_n^c / \sum_{i=1}^N p_i^c q_i^c; \quad c = 1, 2; \quad n = 1, \dots, N$$

denote the expenditure share for commodity  $n$  in country  $c$  and consider the following *weighted least squares model*:<sup>8</sup>

$$(10) \quad \min_{\gamma, \delta_n} \sum_{n=1}^N s_n^1 [\ln p_n^1 - \delta_n]^2 + \sum_{n=1}^N s_n^2 [\ln p_n^2 - \gamma - \delta_n]^2$$

where  $\gamma \equiv \ln \alpha_2$  and  $\delta_n \equiv \ln \beta_n$  for  $n = 1, \dots, N$ . In order to justify the model (10) in terms of a traditional stochastic specification, assume that  $\ln p_n^1$  is an independently distributed random variable with mean  $\delta_n$  and variance  $\sigma^2/s_n^1$  and  $\ln p_n^2$  is an independently distributed random variable with mean  $\delta_n + \gamma$  and variance  $\sigma^2/s_n^2$ .

<sup>6</sup>Thus we are using the same type of reasoning used by Theil (1967, pp. 136–8) to justify his weighted stochastic approach to index number theory. Rao (2002) also used this weighting by economic importance approach. See the ILO (2004, pp. 299–304) for an introduction to the stochastic approach to index number theory.

<sup>7</sup>See Fisher (1922, p. 64) or the ILO (2004, p. 306) for a formal definition of this test.

<sup>8</sup>It is no longer necessary for prices to be defined in a common international currency. However, we still assume that the quantity units of measurement are the same in both countries.

The solution for  $\gamma$  which solves the above weighted least squares minimization problem is

$$(11) \quad \gamma^{**} = \sum_{n=1}^N S_n^* \ln(p_n^2 / p_n^1)$$

where

$$(12) \quad S_n^* \equiv h(s_n^1, s_n^2) / \sum_{i=1}^N h(s_i^1, s_i^2); \quad n = 1, \dots, N.$$

Thus  $\gamma^{**}$  is another share weighted average of the logarithms of the price ratios  $p_n^2/p_n^1$ . If  $\gamma^{**}$  is exponentiated, then an estimator  $\alpha_2^{**}$  for  $\alpha_2$ , the price level of country 2 relative to country 1, is obtained. Let us call the bilateral index  $\alpha_2^{**}$  the *harmonic share weights log price change index* or more briefly, the *harmonic share weights index*.

It can be shown<sup>9</sup> that  $\alpha_2^{**}$  approximates the *Törnqvist index* to the second order around an equal price and quantity point; i.e. for most data sets, the harmonic share weights index will be very close to the Törnqvist index.<sup>10</sup> Thus the weighted hedonic regression model defined by (10) leads to a bilateral index number formula which will be reasonably close to a superlative index number formula.<sup>11</sup>

Instead of weighting the period 1 logarithmic prices by the corresponding period 1 shares and the period 2 prices by the period 2 shares, we could weight both prices by the arithmetic average of the two expenditure shares. Thus consider the following *weighted least squares model*:<sup>12</sup>

$$(13) \quad \min_{\gamma, \delta_n} \sum_{n=1}^N (1/2)(s_n^1 + s_n^2)[\ln p_n^1 - \delta_n]^2 + \sum_{n=1}^N (1/2)(s_n^1 + s_n^2)[\ln p_n^2 - \gamma - \delta_n]^2$$

where as usual,  $\gamma \equiv \ln \alpha_2$  and  $\delta_n \equiv \ln \beta_n$  for  $n = 1, \dots, N$ . In order to justify the model (13) in terms of a traditional stochastic specification, assume that  $\ln p_n^1$  is an independently distributed random variable with mean  $\delta_n$  and variance  $\sigma^2/(1/2)(s_n^1 + s_n^2)$  and  $\ln p_n^2$  is an independently distributed random variable with mean  $\delta_n + \gamma$  and variance  $\sigma^2/(1/2)(s_n^1 + s_n^2)$ .<sup>13</sup> The solution for  $\gamma$  which solves the above weighted least squares minimization problem turns out to be:

<sup>9</sup>Use the techniques in Diewert (1978).

<sup>10</sup>The logarithm of the Törnqvist index is defined by (11) except the shares defined by (12) use *arithmetic* means of the expenditure shares instead of *harmonic* means. The Törnqvist index is Fisher's (1922, p. 473) formula number 123. The Törnqvist and Fisher ideal indexes are examples of superlative indexes; see the ILO (2004, pp. 318–24) for material on superlative indexes.

<sup>11</sup>In fact, the harmonic share weights index is what Diewert (1978, p. 889) called a *pseudosuperlative index*; i.e. it approximates a superlative index to the second order around an equal price and quantity point. It turns out that all known superlative index number formulae approximate each other to the second order around an equal price and quantity point; see Diewert (1978, p. 884) or the ILO (2004, p. 324). Thus a pseudosuperlative index has the same level, first derivatives and second order partial derivatives as a superlative index when both indexes are evaluated at a point where the two price vectors are identical and the two quantity vectors are identical.

<sup>12</sup>It is no longer necessary for prices to be defined in a common international currency. However, we still assume that the quantity units of measurement are the same in both countries.

<sup>13</sup>Clements and Izan (1981, p. 746) derived the Törnqvist price index in a related stochastic setting where  $\ln(p_n^2/p_n^1)$  was taken to be the dependent variable.

$$(14) \quad \gamma^{***} = \sum_{n=1}^N (1/2)(s_n^1 + s_n^2) \ln(p_n^2/p_n^1).$$

It can be seen that  $\gamma^{***}$  is the logarithm of the Törnqvist (1936) price index and hence if  $\gamma^{***}$  is exponentiated, then an estimator  $\alpha_2^{***}$  for  $\alpha_2$ , the price level of country 2 relative to country 1, is obtained. This estimator is the Törnqvist price index. Given that the Törnqvist price index is a superlative index number formula, it appears that choosing the weights in a weighted CPD regression to be the arithmetic average of the expenditure shares is a reasonable choice, provided that there are positive expenditures on all  $N$  goods for both countries.

Finally, consider the following *unweighted version* of the least squares minimization problem (5), (10) or (13):

$$(15) \quad \min_{\gamma, \delta^s} \sum_{n=1}^N [\ln p_n^1 - \delta_n]^2 + \sum_{n=1}^N [\ln p_n^2 - \gamma - \delta_n]^2$$

which is the original CPD method specialized to 2 countries. Let  $\gamma^{****}$  solve (15). It can be shown that:

$$(16) \quad e^{\gamma^{****}} = P_J(p^1, p^2) \equiv \prod_{n=1}^N [p_n^2/p_n^1]^{1/N}$$

and  $P_J(p^1, p^2)$  is the *Jevons (1865) price index* between countries 1 and 2.

### 3. A MODIFIED CPD REGRESSION MODEL WHEN QUANTITY WEIGHTS ARE AVAILABLE

In the previous section, we suggested weighting prices in each country by their economic importance. But how exactly should “economic importance” be measured? In Section 2, we initially suggested that the country’s expenditure on each commodity was a reasonable indicator of economic importance. However, we eventually rejected this initial suggestion since it led to an asymmetry in weighting when a large country was being compared to a small country. Thus we eventually ended up suggesting country expenditure shares as reasonable weights that reflected the relative economic importance of each price in each country and this led to the harmonic share weights index. In this section, instead of using *expenditures* on a commodity as an indicator of the economic importance of the commodity, we start out by suggesting that perhaps the *quantities purchased* could serve as indicators of economic importance. Unfortunately, we shall find that this approach is ultimately not very successful although at the end, we will succeed in providing a (rather tortured) justification for the Geary (1958) Khamis (1970, 1972) bilateral index number formula using this “quantities as weights” approach.

Thus, instead of taking logarithms of prices and weighting by expenditure shares as in (10) or (13) above, we could consider the following *weighted least squares regression model where we weight by quantities sold*  $q_n^c$  in each country:<sup>14</sup>

$$(17) \quad \min_{\alpha, \beta^s} \sum_{n=1}^N q_n^1 [p_n^1 - \beta_n]^2 + \sum_{n=1}^N q_n^2 [p_n^2 - \alpha_2 \beta_n]^2.$$

<sup>14</sup>As in the previous section, we argue that each price  $p_n^c$  should be weighted according to its economic importance; i.e. if consumers in country  $c$  are purchasing  $q_n^c$  units of commodity  $n$  in country  $c$ , then  $p_n^c$  should appear  $q_n^c$  times in the regression instead of only once.

However, there are a number of problems with the  $\alpha_2$  solution to the above least squares problem. One of the biggest problems is that the resulting index number formula is not invariant to changes in the units of measurement. This is a fatal flaw.<sup>15</sup>

In order to deal with this invariance problem, we consider replacing the prices  $p_n^c$ , which are the dependent variables in the regression model, by a transformation of these prices. A very simple transformation is the power transformation. Thus let  $\rho$  be a nonzero number and replace  $p_n^1$  by  $(p_n^1)^\rho$ . Using (1) and (2),  $(p_n^1)^\rho$  is approximately equal to

$$(18) \quad (\beta_n)^\rho \equiv \delta_n; \quad n = 1, \dots, N.$$

The country 2 price for commodity  $n$ ,  $p_n^2$ , is replaced by  $(p_n^2)^\rho$ , and using (1),  $(p_n^2)^\rho$  is approximately equal to  $(\alpha_2 \beta_n)^\rho = \alpha_2^\rho \beta_n^\rho = \gamma \delta_n$ , where  $\gamma$  is defined as follows:

$$(19) \quad \gamma \equiv \alpha_2^\rho.$$

The weighted least squares regression model that is the counterpart to (17), except that prices  $p_n^1$  and  $p_n^2$  are replaced by transformed prices  $(p_n^1)^\rho$  and  $(p_n^2)^\rho$  respectively (and using definitions (18) and (19) to reparameterize the  $\beta_n$  and  $\alpha_2$ ), is:

$$(20) \quad \min_{\gamma, \delta_n} \sum_{n=1}^N q_n^1 [(p_n^1)^\rho - \delta_n]^2 + \sum_{n=1}^N q_n^2 [(p_n^2)^\rho - \gamma \delta_n]^2.$$

There are two problems with the weighted least squares problem (20):

- We cannot obtain a closed form solution for the key parameter  $\gamma$ .
- The price levels in the two countries could be quite different and hence the prices of the country with the higher price level could get too much weight in (20); put another way, there could be a heteroskedasticity problem.

Both of these problems can be solved if we divide the country 2 transformed prices  $(p_n^2)^\rho$  by the transformed index of country 2's prices relative to country 1's prices,  $\alpha_2^\rho$ , which is equal to  $\gamma$ . Performing this division, the weighted least squares problem (20) becomes:

$$(21) \quad \min_{\phi, \delta_n} \sum_{n=1}^N q_n^1 [(p_n^1)^\rho - \delta_n]^2 + \sum_{n=1}^N q_n^2 [(p_n^2)^\rho \phi - \delta_n]^2$$

where the new parameter  $\phi$  is defined as

$$(22) \quad \phi \equiv 1/\gamma = 1/\alpha_2^\rho.$$

In order to justify the weighted least square model (21) in terms of a traditional stochastic specification, assume that  $(p_n^1)^\rho$  is an independently distributed random variable with mean  $\delta_n$  and variance  $\sigma^2/q_n^1$  and  $(p_n^2)^\rho$  is an independently distributed random variable with mean  $\delta_n \gamma$  and variance  $\gamma^2 \sigma^2/q_n^2$ .

The solution for  $\phi$  which solves the above weighted least squares minimization problem is:

<sup>15</sup>Note that this problem did not arise with any of the four least squares problems defined in the previous section; i.e. in all four cases, the resulting index number formula was invariant to changes in the units of measurement. This invariance to changes in the units of measurement test and its history are discussed in the ILO (2004, p. 294).

$$(23) \quad \phi^* = \sum_{n=1}^N h(q_n^1, q_n^2)(p_n^1)^\rho (p_n^2)^\rho / \sum_{n=1}^N h(q_n^1, q_n^2)(p_n^1)^\rho (p_n^2)^\rho$$

where  $h(q_n^1, q_n^2)$  is the harmonic mean of the consumption of commodity  $n$  for country 1 and 2. It is evident that  $\phi^*$  is not invariant to changes in the units of measurement for each commodity unless  $\rho = 1/2$ . Hence, in what follows, we assume:

$$(24) \quad \rho = 1/2$$

Using assumption (24) and definition (19), the  $\phi^*$  defined by (23) translates into the following estimator for the price level of country 2 relative to that of country 1:

$$(25) \quad \alpha_2^* = \left[ \sum_{n=1}^N h(q_n^1, q_n^2) p_n^2 / \sum_{n=1}^N h(q_n^1, q_n^2)(p_n^1)^{1/2} (p_n^2)^{1/2} \right]^2.$$

The index number formula defined by (25) is a bit strange looking but it does have some of the usual properties of bilateral indexes, such as being homogeneous of degree one in the prices of country 2 and homogeneous of degree minus one in the prices of country 1.<sup>16</sup> Unfortunately, the index defined by (25) does not satisfy the important time reversal test.<sup>17</sup>

In order to remedy this last defect, we make use of Fisher's (1922, pp. 136–40) rectification procedure.<sup>18</sup> Instead of letting country 1 play the role of the base country, we can let country 2 be the base country; i.e. replace the normalization (2),  $\alpha_1 = 1$ , with  $\alpha_2 = 1$ . Repeating the above analysis with this change leads to the following estimator for the price level of country 1 relative to that of country 2:

$$(26) \quad \alpha_1^* = \left[ \sum_{n=1}^N h(q_n^1, q_n^2) p_n^1 / \sum_{n=1}^N h(q_n^1, q_n^2)(p_n^1)^{1/2} (p_n^2)^{1/2} \right]^2.$$

Now use  $1/\alpha_1^*$  as an estimator for the price level of country 2 relative to that of country 1 and finally, take the geometric mean of  $\alpha_2^*$  defined by (25) and  $1/\alpha_1^*$  where  $\alpha_1^*$  is defined by (26) to obtain our *final estimator of the price level of country 2 relative to that of country 1*:

$$(27) \quad \alpha^* = [\alpha_2^* / \alpha_1^*]^{1/2} = \sum_{n=1}^N h(q_n^1, q_n^2) p_n^2 / \sum_{n=1}^N h(q_n^1, q_n^2) p_n^1 \equiv P_{GK}(p^1, p^2, q^1, q^2)$$

where as usual,  $h(q_n^1, q_n^2)$  is the harmonic mean of the consumption of commodity  $n$  for country 1 and 2. It can be seen that the bilateral index number formula  $\alpha^*$  defined by (27) is actually the *Geary* (1958) *Khamis* (1970, 1972) *bilateral index number formula*,  $P_{GK}(p^1, p^2, q^1, q^2)$ . This formula was advocated by Geary and Khamis in the more general context of making multilateral price and quantity comparisons between many countries but their more general formula boils down to the above index in the case of two countries. This formula was also considered by Irving Fisher (1922, p. 485) as his index number formula 3153.

Recall that the *Marshall* (1887), *Edgeworth* (1925) and *Walsh* (1901) bilateral index number formulae can be defined as follows:

<sup>16</sup>See the ILO (2004, pp. 293–4) for a discussion of these homogeneity tests.

<sup>17</sup>See the ILO (2004, p. 295) for a discussion of this test and its early history.

<sup>18</sup>This technique is actually due to Walsh (1921, p. 524).

$$(28) \quad P_{ME}(p^1, p^2, q^1, q^2) \equiv \sum_{n=1}^N (1/2)(q_n^1 + q_n^2) p_n^2 / \sum_{n=1}^N (1/2)(q_n^1 + q_n^2) p_n^1;$$

$$(29) \quad P_W(p^1, p^2, q^1, q^2) \equiv \sum_{n=1}^N (q_n^1 q_n^2)^{1/2} p_n^2 / \sum_{n=1}^N (q_n^1 q_n^2)^{1/2} p_n^1.$$

Comparing (28) and (29) with the Geary Khamis index defined by (27), it can be seen that all three indexes are of the fixed basket type. The basket being compared is the harmonic, arithmetic and geometric mean of the baskets pertaining to the two countries respectively. All three indexes will approximate each other to the second order around an equal price and quantity point.<sup>19</sup> Thus while the Geary Khamis bilateral index number formula is not superlative, it will approximate a superlative index to the second order around an equal price and quantity point.<sup>20</sup>

It is possible to obtain the Walsh price index directly by considering weighted least squares regression models of the type defined by (20) above: replace the quantity weights  $q_n^1$  by  $q_n^{1*}$  and  $q_n^2$  by  $q_n^{2*}$  and repeat all of the steps (22) to (27) for these new weights. We end up with the following estimator for *the price level of country 2 relative to that of country 1*:

$$(30) \quad \alpha^{**} = \sum_{n=1}^N h(q_n^{1*}, q_n^{2*}) p_n^2 / \sum_{n=1}^N h(q_n^{1*}, q_n^{2*}) p_n^1$$

where as usual,  $h(q_n^{1*}, q_n^{2*})$  is the harmonic mean of  $q_n^{1*}$  and  $q_n^{2*}$ . Now choose the quantity weights  $q_n^{1*}$  and  $q_n^{2*}$  as follows:

$$(31) \quad q_n^{1*} \equiv [q_n^1 q_n^2]^{1/2}; \quad q_n^{2*} \equiv [q_n^1 q_n^2]^{1/2}.$$

Thus the weights for commodity  $n$  for both periods,  $q_n^{1*}$  and  $q_n^{2*}$ , are chosen to be the geometric mean of the period 1 and 2 quantities for commodity  $n$ ,  $[q_n^1 q_n^2]^{1/2}$ . If we substitute the weights  $q_n^{1*}$  and  $q_n^{2*}$  defined by (31) into (30), we find that  $\alpha^{**}$  is the Walsh price index  $P_W$  defined by (29).

Note that an unweighted version of the index defined by (27) above can be obtained if we consider an unweighted version of the least squares problem (21); i.e. if we set:

$$(32) \quad q_n^1 = q_n^2 = 1; \quad n = 1, 2, \dots, N.$$

Under these conditions,  $\alpha^*$  defined by (27) becomes:

$$(33) \quad P_D(p^1, p^2) \equiv \sum_{n=1}^N (1/N) p_n^2 / \sum_{n=1}^N (1/N) p_n^1$$

where  $P_D(p^1, p^2)$  is the *Dutot* (1738) *price index* between countries 1 and 2.<sup>21</sup>

<sup>19</sup>Hence all three indexes are pseudosuperlative.

<sup>20</sup>This follows from the fact that the Walsh index is superlative. The test properties of the Geary Khamis bilateral index are listed in Diewert (1999, pp. 27–8) who noted that this index satisfied 14 of the 20 tests collected in Diewert (1992, pp. 214–21).

<sup>21</sup>However, note that the Dutot index is not invariant to changes in the units of measurement. Thus it should be used only if commodities being aggregated have exactly the same units of measurement.



#### 4. CONCLUSION

At first glance, it seems that the Country Product Dummy method (and by extension, unweighted hedonic regression methods) for comparing prices between countries (or time periods) is totally unrelated to traditional index number methods for making price comparisons between two countries or time periods. However, in this note, we have shown that if *unweighted* Country Product Dummy (or hedonic) regressions are replaced by suitable *weighted* counterparts, then the resulting measures of price change are very closely related to traditional bilateral index number formulae. If expenditure share weights are used and the dependent variable is the logarithm of price, then the *harmonic share weights index*, the exponential of  $\gamma^{**}$  defined by (11) and (12), is the resulting weighted CPD estimator of the price level of country 2 relative to country 1. If the individual period shares are replaced by the arithmetic average of the shares for both sets of weights, then the exponential of  $\gamma^{***}$  defined by (14) turns out to be the Törnqvist price index. If quantity weights are used and the dependent variable is the square root of price, then after averaging two estimators, the Geary Khamis index defined by (27) is the resulting weighted CPD estimator of the price level of country 2 relative to country 1. If geometric average quantity weights are used as weights for both periods, then after averaging two estimators, the Walsh price index defined by (29) is the resulting estimator of the price level of country 2 relative to country 1.

The corresponding unweighted estimators of the price level of country 2 relative to that of country 1 turn out to be the Jevons index defined by (16) and the Dutot index defined by (33). *These unweighted indexes can be very far from their weighted counterparts.* Thus the main conclusion that we draw from this note is that in running Country Product Dummy regressions or hedonic regressions in the time series context, *it is very important to run appropriately weighted versions of these regressions in order to obtain more accurate estimates of price levels.*

#### REFERENCES

- Clements, K. W. and H. Y. Izan, "A Note on Estimating Divisia Index Numbers," *International Economic Review*, 22, 745–7, 1981.
- Diewert, W. E., "Fisher Ideal Output, Input and Productivity Indexes Revisited," *Journal of Productivity Analysis*, 3, 211–48, 1992.
- , "Superlative Index Numbers and Consistency in Aggregation," *Econometrica*, 46, 883–900, 1978 (reprinted in W. E. Diewert and A. O. Nakamura (eds), *Essays in Index Number Theory, Volume 1*, North-Holland, Amsterdam, 253–73, 1993).
- , "Axiomatic and Economic Approaches to International Comparisons," in A. Heston and R. E. Lipsey (eds), *International and Interarea Comparisons of Income, Output and Prices*, Studies in Income and Wealth Volume 61, NBER, The University of Chicago Press, Chicago, 13–87, 1999.
- , "Hedonic Regressions: A Consumer Theory Approach," Discussion Paper 01-12, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1, forthcoming in R. Feenstra and M. Shapiro (eds), *Scanner Data and Price Indexes*, Studies in Income and Wealth, Volume 61, University of Chicago Press, Chicago, 2001.
- Dutot, Charles, *Réflexions politiques sur les finances et le commerce*, Volume 1, Les frères Vaillant et N. Prevost, La Haye, 1738.
- Edgeworth, F. Y., *Papers Relating to Political Economy*, Volume 1, Burt Franklin, New York, 1925.
- Fisher, I., *The Making of Index Numbers*, Houghton-Mifflin, Boston, 1922.
- Geary, R. G., "A Note on Comparisons of Exchange Rates and Purchasing Power between Countries," *Journal of the Royal Statistical Society*, A121, 97–9, 1958.
- ILO/IMF/OECD/UNECE/Eurostat/The World Bank, *Consumer Price Index Manual: Theory and Practice*, International Labour Office, Geneva, 2004.

- Jevons, W. S., "The Variation of Prices and the Value of the Currency since 1782," *Journal of the Statistical Society of London*, 28, 294–320, 1865 (reprinted in *Investigations in Currency and Finance*, Macmillan and Co., London, 119–50, 1884).
- Khamis, S. H., "Properties and Conditions for the Existence of a New Type of Index Number," *Sankhya*, B32, 81–98, 1970.
- , "A New System of Index Numbers for National and International Purposes," *Journal of the Royal Statistical Society*, A135, 96–121, 1972.
- Marshall, A., "Remedies for Fluctuations of General Prices", *Contemporary Review*, 51, 355–75, 1887.
- Rao, D. S. Prasada, "On the Equivalence of Weighted Country Product Dummy (CPD) Method and the Rao System for Multilateral Price Comparisons," *Review of Income and Wealth*, 51(3), 541–580, 2005.
- Selvanathan, E. A. and D. S. Prasada Rao, *Index Numbers: A Stochastic Approach*, The University of Michigan Press, Ann Arbor, 1994.
- Summers, R., "International Comparisons with Incomplete Data," *Review of Income and Wealth*, 29(1), 1–16, 1973.
- Theil, H., *Economics and Information Theory*, North-Holland, Amsterdam, 1967.
- Törnqvist, Leo, "The Bank of Finland's Consumption Price Index," *Bank of Finland Monthly Bulletin*, 10, 1–8, 1936.
- Triplet, J. D., *Handbook on Quality Adjustment of Price Indexes for Information and Communication Technology Products*, November 10 draft, OECD, Paris, 2000.
- Triplet, J. E. and R. J. McDonald, "Assessing the Quality Error in Output Measures: The Case of Refrigerators," *The Review of Income and Wealth*, 23(2), 137–56, 1977.
- Walsh, C. M., *The Measurement of General Exchange Value*, Macmillan and Co., New York, 1901.
- , "The Best Form of Index Number: Discussion," *Quarterly Publication of the American Statistical Association*, 17, March, 537–44, 1921.