Journal of Official Statistics, Vol. 21, No. 3, 2005, pp. 463–487

Purchasing Power Parity Measurement and Bias from Loose Item Specifications in Matched Samples: An Analytical Model and Empirical Study

Mick Silver[1] *and Saeed Heravi*[2]

This article considers the trade-off between two types of bias in the compilation of purchasing power parity (PPP) indices. The first is from the poor coverage of the items compared, or *out-of-sample bias*, when relatively tight specifications are used to define the items compared. The second is inappropriate quality comparisons, or *in-sample bias*, when loosely defined specifications are used to allow for greater coverage. An analytical framework is provided, which establishes the nature and extent of bias from poor coverage. Scanner data for three countries are used in this study to investigate the bias. Such data allow matching to be undertaken to different degrees of "tightness" of item specification and the resulting bias evaluated. Hedonic regression indices are argued to be a useful approach to dealing with out-of-sample bias. Such indices do not have to be based on matched specifications, but can extend to a representative sample of all prices, the differences in quality being "controlled" for by the regression as opposed to the matching and its restrictive effect on the sample.

JEL classification: C43, C81, C82, E43, I32, and O47.

Key words: Purchasing power parity; index numbers; scanner data; hedonic regression.

1. Introduction

Inter-country price comparisons require comparisons of prices between countries of similarly described items. Detailed specifications need to be drawn up for the selection of each item so that the prices of "like" in one country are compared with "like" in another country. The concern of this article is with the trade-off between tight and loose item descriptors. *Tight specifications* lead to less quality differential bias, but increase the chance of items not being matched resulting in a poorer coverage of the index. *Loose specifications* increase the chance of items being matched, but lead to price comparisons of dissimilar items. There are therefore two types of bias: the first is *in-sample bias* due to loose item specifications resulting in noncomparable items being compared; the second is *out-of-sample bias* due to low coverage from tight item specifications and unrepresentative comparisons (see Triplett 2004 for the use of such terms in the context of price measurement over time).

[1] Statistics Department, International Monetary Fund, 700 19th Street NW, Washington, DC 20431, USA. Email: msilver@imf.org
[2] Cardiff Business School, Cardiff University, Colum Drive, Cardiff CF10 3EU, UK. Email: heravis@cardiff.ac.uk
Acknowledgments: We acknowledge the support and help of the World Bank. Helpful comments were received on an earlier draft at a conference on the International Comparisons Program, World Bank, Washington D.C., March 2002 from Bart van Ark (University of Groningen), Yonas Biru (World Bank), Jim Cuthbert, Ed Dean (George Washington University), James Mead and Michael Ward. Erwin Diewert (University of British Columbia) helped us with the notation and a more formal derivation of the analytical framework in Equation (2) and anonymous referees provided useful comments, particularly the reference to selectivity bias in Section 2c. GfK Marketing Services Ltd. provided the data for research purposes. The usual disclaimers apply.

In Section 2 the article provides an analytical framework for *out-of-sample bias* and Section 3 examines the use of hedonic indices in this context. The empirical analysis in Sections 4 to 6 is based on scanner data for television sets from the bar-code readers of retail outlets in three countries: United Kingdom, France, and the Netherlands. It is noted that what is being done in this article is to use scanner data to *simulate* the coverage bias that might arise from using tight item descriptors for matching models. Since scanner data cover all transactions it allows the tightness of specifications for matching to be increased and the resulting bias from the poorer coverage to be estimated. The use of scanner data for PPP measurement was the subject of Heravi et al. (2003) and its potential advantages are noted in the ICP Handbook 2004 (World Bank 2004).

A description of the data is provided in Section 4. Section 5 provides the empirical results on how coverage varies with the tightness of the specification. Coverage is reduced with the fall in the number of matches as the item descriptors become increasingly tighter. The extent of this is shown for bilateral comparisons between the three countries. A fall in the coverage of items matched is shown in the analytical model in Section 2 to be only one necessary condition for bias. For bias to occur it is also required that matched and unmatched quality-adjusted prices differ, and this is the subject of Section 6. The empirical analysis concludes in Section 7 with comparisons of the parity adjustments (for different item specifications) using matched data, and the whole data set and conclusions are drawn in Section 8. Scanner data are shown to be useful for providing insights into this methodological issue which PPP data, being constrained to the sample collected, as opposed to the universe of transactions, cannot identify.

2. Inter-Country Price Comparisons: An Analytical Framework

2.1. Methods of inter-country price comparisons

Consider a sample/universe of models of a product sold whose quality characteristics differ between models within and between countries. Some of the models sold in one country will be sold in another, i.e., they will be matched models, while unmatched models are sold in one country but not the other. Consider the hedonic dummy variable regression for models sold in countries A and B:

$$\ln p_{cm} = \alpha_A + \alpha_D D_B + \sum_{k=1}^{K} \beta_k z_{mck} + \varepsilon_{cm} \quad m \in S(c); \quad c = A, B \tag{1}$$

where $S(c)$ is the set of models available in country c, p_{cm} is the country c price of model m; D_B is a dummy variable that is 1 if the observation is for Country B and is 0 otherwise; z_{mck} is the amount of characteristic k model m in country c possesses; and ε_{cm} is an error term with the usual desirable properties. Let the number of models available in country c be $N(c)$; i.e., there are $N(c)$ models in the set $S(c)$ for each c. The estimated coefficient $\hat{\alpha}_D = (\alpha_B - \alpha_A)$ is an estimate of the change in the (the logarithm of) "price" between countries B and A, having controlled for the effects of variation in quality (via$\sum_{k=1}^{K} \beta_k z_{mck}$). The semi-logarithmic functional form is equivalent to a comparison of quality-adjusted geometric mean prices in the two countries, as explained in Triplett (2004) and Diewert (2003). Note that the estimated change in "price" is made up of two elements. The first is the quality-adjusted price ratio between the two countries and the second is the exchange rate, since the prices are measured in different currencies. Such "prices" are effectively the parity adjustments for this product which, together with other such parity adjustments, aggregate up to an estimate of purchasing power parity (World Bank 2004). The $\hat{\beta}_k$ are estimates of the marginal value of each characteristic. Equation (1) is a basic hedonic regression model akin to those used for temporal studies, examples of which are Triplett (1987), Griliches (1990), Berndt et al. (1995), and Silver and Heravi (2001). The hedonic formulation provides a basis for price comparisons between countries when the quality of items purchased differs – see, for example, Aten (1999) and, for U.S. inter-area comparisons, Kokoski, Moulton, and Zieschang (1999).

There is an alternative regression formulation, the Country-Product-Dummy (CPD) procedure developed by Robert Summers. In Summers's formulation missing "unmatched" prices are assumed to be randomly distributed (Summers 1973). It was employed in the calculations for the initial studies of the World Bank's International Comparisons Program. Instead of using the z_{mck} characteristics of models sold, dummy variables are included for each (matched) model, i.e.:

$$\ln p_{cm} = \alpha_A + \alpha_D D_B + \sum_{m=2}^{M} \beta_m D_m + \upsilon_{cm} \tag{2}$$

where D_B (and D_m) are dummy variables that are 1 if the observation is for country B (and model m, in either country A or B) and zero otherwise and $\alpha_D = (\alpha_B - \alpha_A)$ (and β_m) the parameters to be estimated and υ_{cm} a residual with usual desirable properties. The method assumes that models are matched across countries. The formulation used in this study is that of Equation (1) as opposed to Equation (2). First, it will be shown in the empirical section that a substantial proportion of the models were not matched and given the availability of data on their quality characteristics, Equation (1) provides a suitable framework for such analysis. Second, the purpose of the article is to examine the effects of loose and tight specifications on inter-country comparisons and, in the next section, the formulation in (1) is seen to serve us well in this respect since it allows adjustments to be made to prices for quality differences. There are of course other approaches, the main one being the EKS method.

The EKS method is not regression-based but is based on matched models comparisons. For bilateral comparisons the Fisher index number formula can be used at the "Basic Heading"-level (BH) where weights are available. Below this level an implicit Fisher-type weighting system is used whereby the weights are a crude division into "representative" and "unrepresentative" products. A Laspeyres-type index below the BH level uses representative models in Country A, while a Paasche-type index below the BH level uses representative models in Country B. The Fisher index is the geometric mean of the two, giving twice the weight to models which are representative in both countries. The method is not readily amenable to the required analytical framework motivating this article (see Balk 2001, and Rao and Timmer 2003, for details).

2.2. *Specification issues for the hedonic regression*

The functional form used for the hedonic index in (1) is a semi-logarithmic one. Diewert (2002) considers the axiomatic properties of a linear form versus semi- or double-logarithmic forms and finds against a linear form. The semi-logarithmic form is used here instead of a double-logarithmic form since many of the characteristic variables on the right-hand side are dummy variables precluding the taking of logarithms of such variables.

The estimated coefficients from hedonic regressions are subject to omitted variable bias if relevant quality characteristics are excluded. The extent of the bias on an included variable depends on the product of the coefficient on the omitted variable (its importance in a price-determining sense), and the coefficient on an included variable from an auxiliary regression of the omitted on all included variables (the extent to which the included is related to the omitted variable) (Berndt 1990). The variable set used is relatively extensive and an $\overline{R}^2$ of over 0.9 reflects this, as does the inability to reject the specification test in Table 3. We nonetheless note that omitted variable bias in general can be problematic in such studies, especially insofar as the measures of quality used may not be measures that consumers perceive to be important (Berndt 1990).

A particular feature of the data is that the estimation is across models available in one country, but not in another. There may be, for example, something in the preferences of the French that renders the supply of a model available in the UK unavailable in France. It can thus be argued that there is a selectivity bias in that the data sampled do not include observations relating to models in a country which have no price. While the use of a Heckman two-stage estimator may merit some attention, this is a subject for further research.

2.3. *The analytical framework*

We borrow now on the valuable analytical framework by Aizcorbe et al. (2000) and, following the analysis for CPIs in Silver and Heravi (2002), generalise it for unmatched comparisons between Countries A and B. We relax the matched model restriction, but still assume that there are only two countries in the hedonic regression model defined by (1). Some additional notation is required in order to model this case. Define the following sets of models: $S(A \cap B) \equiv S(A) \cap S(B)$, $S(A \neg B) \equiv S(A) \neg S(B)$ and $S(B \neg A) \equiv S(B) \neg S(A)$.

Thus $S(A \cap B)$ is the set of models that are present in both Countries A and B, $S(A \neg B)$ is the set of models that are present in Country A but not Country B, and $S(B \neg A)$ is the set of models that are present in Country B but not Country A. Let the number of models in the sets $S(A \cap B)$, $S(A \neg B)$ and $S(B \neg A)$ be denoted by $N(A \cap B)$, $N(A \neg B)$ and $N(B \neg A)$, respectively. Relating our new notation to the total number of models in Countries A and B, $N(A)$ and $N(B)$, respectively, it can be seen that: $N(A) = N(A \cap B) + N(A \neg B)$ and $N(B) = N(A \cap B) + N(B \neg A)$.

Using $z_{mAk} = z_{mBk}$ for the common models $m \in S(A \cap B)$ leads to the following formula for the log of the *hedonic price index* for matched and unmatched models:

$$
\begin{aligned}
1\text{n}\ P_B/P_A = & [1/N(A \cap B)] \sum_{m \in S(A \cap B)} \ln\ [p_{mB}/p_{mA}] \\
& + [l/N(A \cap B)] \sum_{m \in S(B-A)} \lfloor \ln\ p_{mB} - \sum_{k=1}^{K} z_{mBk}\hat{\beta}_k - \hat{\alpha}_B \rfloor \\
& - [l/N(A \cap B)] \sum_{m \in S(A-B)} \lfloor \ln\ p_{mA} - \sum_{k=1}^{K} z_{mAk}\hat{\beta}_k - \hat{\alpha}_B \rfloor
\end{aligned}
\quad (3)
$$

where the ˆ denotes estimated values for the parameters, i.e., $\hat{\beta}_k, \hat{\alpha}_A$, and $\hat{\alpha}_D$ estimated from Equation (1) where $\hat{\alpha}_D = (\hat{\alpha}_A - \hat{\alpha}_B)$ and thus $\hat{\alpha}_B = \hat{\alpha}_A - \hat{\alpha}_D$.

The first set of terms on the right-hand side of (3) is the *matched model contribution* to the overall index, $\ln P_B/P_A$. If the second set of terms, $+[1/N(A \cap B)]\sum_{m \in S(B-A)} \lfloor \ln p_{mB} - \sum_{k=1}^{K} z_{mBk}\hat{\beta}_k - \hat{\alpha}_B \rfloor$ is positive, then the matched model price index is too low and must be adjusted upward. A typical term in this sum of terms is $\ln p_{mB} - \sum_{k=1}^{K} z_{mBk}\hat{\beta}_k - \hat{\alpha}_B$ and this is equal to the logarithm of a Country B model price p_{mB} (where the model was not present in Country A) less a quality adjustment for its characteristics $\sum_{k=1}^{K} z_{mBk}\hat{\beta}_k$ less the Country B price level $\hat{\alpha}_B$. Thus the price of a new model m in Country B will raise the overall price index if its quality-adjusted price is higher than the Country B price level that is estimated by the hedonic model, which is a very sensible result (see also Triplett and McDonald 1977). The effect this expression has depends on the number of models in Country B but not in A, as a proportion of the total number of matched models, again an intuitive result. There is a corresponding interpretation for the last set of terms on the right-hand side of (3).

Matched methods indices can be seen to ignore the last two lines in Equation (3) and will thus differ from the hedonic dummy variable approach using all of the data. If, for example, unmatched quality-adjusted prices in Country B are the same on average as unmatched quality-adjusted prices in Country A, there can be seen from the last two lines of Equation (3) to be no bias. If, however, unmatched quality-adjusted prices in Country B are lower than their matched counterparts, and if unmatched quality-adjusted prices in Country A are above their matched counterparts, the bias will be compounded as the negative contribution of the first term in the last line of Equation (3) is combined with the negative contribution of the second term. The index will understate price differences between Countries B and A. If the signs are reversed, unmatched quality-adjusted prices in Country B are on average above matched quality-adjusted prices in B and quality-adjusted unmatched prices in A are on average below quality-adjusted matched prices in A–the bias is in the other direction. The biases will offset each other if quality-adjusted unmatched prices both in country A or country B are either above or below their matched counterparts, as again is apparent from the last line of Equation (3). The extent of any of this bias depends on the weights in Equation (3). These are the relative numbers of observations in each country. Ideally the weights should relate to sales (Diewert 2002) and not the number of observations, and the sales weights should correspond to a superlative formulation as discussed in Diewert (2002) and Silver (2002).

3. Hedonic Indices and Hedonic Adjustments to Replacement Items

The measurement of price comparisons unaffected by quality differences is primarily achieved by matching models. When a matched model of comparable quality is unavailable, an imputation may be used, say based upon the price comparison of other matched models in the product category. Alternatively, with a sufficiently loose specification, the comparison can be made with an item which is broadly similar and the coefficients from hedonic regressions used to adjust either of the prices for the quality differences, if they are considered to be significant.

We thus distinguish between the use of hedonic regressions to make explicit adjustments for quality differences when noncomparable replacements are used, and their use in their own right as *hedonic price indices,* as in the estimated $\hat{\alpha}_D = (\hat{\alpha}_B - \hat{\alpha}_A)$ in (1), which is an estimate of the inter-country quality-adjusted price parity difference between countries A and B. Hedonic price indices are to be preferred over matched model indices when the scale of unmatched models is substantial. However, unmatched items can only be included in the sample as long as price-determining characteristics are also collected, say by means of checklists, as advised by Zieschang, Armknecht, and Smith (2001). This allows out-of-(matched) sample data to be included, and controlled for by means of the hedonic regression results.

4. Empirical Study: Variables and Data

The empirical work utilises scanner, bar-code data for television sets covering all transactions in June and July 1998 in three countries: UK, France, and the Netherlands. The data are before European Monetary Union in which a common currency was adopted for France and the Netherlands. It is worth noting that even with common currencies price comparisons are still required if GDP or its subaggregates are to be compared across countries in real terms (World Bank 2004). The data were provided by GfK Marketing Ltd. who compile scanner data from retailers for domestic electrical consumer durable goods for European countries. Scanner data provide, for each model, information on price (unit value), sales (through aggregation of transactions), characteristics of the product and type of outlet. Since data are available on the characteristics of each item the potential exists to allow for quality-adjustments in the measurement of price comparisons between countries. The advantages and disadvantages of the practical use of such data for CPI compilation have been discussed in Fenwick et al. (2003). Since scanner data cover just about all transactions they are used here to simulate the use of different sample/item selection criteria, since their selection can be evaluated against patterns for the data as a whole.

The observations are for a model of the product for which there was a transaction in a country in a particular outlet type. For example, an observation in the data set for June and July 1998 includes the unit value (£284.52), volume (3,686 transactions) and quality characteristics (including possession of Nicam stereo and fastext text retrieval facilities) of the (Panasonic TX21MD3 21-inch screen) television set sold in electrical chain store multiples only in the UK. For June and July 1998 there were 4,827 observations: 1,186 for the Netherlands, 2,146 for France and 1,495 for the UK, representing over a million transactions over June and July 1998: about 0.2, 0.6, and 0.3 million transactions in each of the Netherlands, France and the UK, respectively. While observations may be for the same model in different outlet-types, models were, on average, sold in 1.37 different outlet-types in the Netherlands, but more so for France at 1.97 and the UK at 2.2. In total there were 2,425 different models and 4,827 different models *when also differentiated by the outlet-type* in which they were sold: models being sold in 1.84 outlet-types on average.

The variable set is quite extensive and includes: *price,* which is the unit value of a model in a month/outlet-type across all transactions, i.e., sales value/volume; *volume,* which is the sum of the transactions during the period; *vintage,* which is the year in which the first transaction of the model took place, and a *characteristics set* which includes:

(i) manufacturer (make) – dummy variables for about 36 makes; (ii) size of screen; dummy variables for possession of: (iii) Nicam stereo; (iv) tube type – flat and square tube/Trinitron; (v) tuner – Pal, Pal/Secam, Pal/Secam/NTSC, Pal plus varieties; (vi) satellite; (vii) text retrieval system – fastext/TOP, teletext; (viii) dolby system; (ix) wide screen; (x) s-vhs socket; (xi) digital and finally, (xii) the outlet type.

The country and outlet classifications for the Netherlands (NL), France (Fr) and the UK were: NL Departmental and Catalogue; NL Electrical multiples (chain stores); NL Photographers; Fr Departmental; Fr Electrical multiples; Fr Hypermarkets; Fr Specialist (independents); Fr Catalogue; UK Mass merchandisers (departmental); UK Electrical multiples; UK Renters and others n.e.c.; UK Independents; UK Catalogue. These were combined in four groups for each country: multiples, mass merchandisers (department stores), catalogue and independents, with NL = Departmental and Catalogue being allocated to mass merchandisers.

5. Empirical Results: Coverage and Tightness of the Specifications

Tables 1 to 3 provide results on the coverage of matched comparisons for all data while allowing the tightness of item specification to vary. Table 1 presents the degree of coverage between France and the UK, Table 2 between France and the Netherlands, and Table 3 between the Netherlands and the UK. Table 1 clearly shows how matching by model number using scanner data does not work. Only about 5% of sales value was covered in France and the UK when matched by model number. This may be the curiosity of the scanner coding by the data compilation agency. It may be that very slightly differentiated models, involving say changes in style, packaging or to meet country-specific safety/technical specifications, when sold in different countries have different model numbers.

When matching is by the 20 screen sizes available, all UK models and 99.4% of French models are covered. The coverage of these matched models by value of transactions is seen to be similarly very high. As the item specification expands, to say the 40 cells: screen size with and without stereo; or the 80 cells: screen size, stereo and flat and square tubes, the coverage can be seen to remain very high in terms both of value and of number of models matched. What is more problematic is if a specific *brand* of television is required to be matched. Matching the 80 brand names alone reduces the coverage of French (sales) values matched to about 80%. If the specification were tighter, whereby *within each brand* a specific screen size, stereo, and flat and square tube were required to be matched, the possible $80 \times 20 \times 2 \times 2 = 6,400$ cells would reduce the coverage to under one-third of all French models and three-quarters by value. Thus it is not just the number of cells used for the disaggregation what matters but also the variables used for it.

A more natural starting point in item specification might be the outlet-type. Outlet-types do not seem to specialise in different screen sizes/stereo/flat and square tubes in different countries, the coverage remaining high (Table 1). If it was expected that different outlet-types specialised in different brands in different countries, for example multiple chain stores in France sold only a few brands while those in the UK sold many, then the brand matching in outlet-types would be worse than by brand alone. There is some evidence of this. For example, the average value of matched UK sales covered was about 87% for

Table 1. Coverage of matched comparisons: France/UK

	Number of cells for matching	Percent of UK matched	Percent of French models matched	Percent of UK value matched	Percent of French value matched
All data – matching by**:**					
Model number	2,146 × 1,495	8.20	5.50	5.00	5.20
Screen sizes **(S)**	20	100.00	99.40	100.00	99.70
S × Stereo **(ST)**	20 × 2	100.00	98.10	100.00	99.00
S × ST × Flat & square tube **(F)**	20 × 2 × 2	95.20	97.20	93.80	98.00
Brand **(B)**	80	81.30	75.40	90.70	82.80
B × S	80 × 20	76.50	67.70	87.40	76.80
B × S × ST	80 × 20 × 2	72.40	64.50	85.70	75.80
B × S × ST × F	80 × 20 × 2 × 2	66.80	62.70	78.30	74.40
Outlet-type **(OT)** × S	3 × 20	99.80	99.10	99.80	99.60
OT × S × ST	3 × 20 × 2	99.70	97.80	99.80	98.90
OT × S × ST × F	3 × 20 × 2 × 2	94.90	96.90	93.60	98.00
OT × B	3 × 80	76.20	72.30	88.50	80.00
OT × B × S	3 × 80 × 20	67.10	59.00	81.40	69.30
OT × B × S × ST	3 × 80 × 20 × 2	63.30	55.00	79.30	67.30
OT × B × S × ST × F	3 × 80 × 20 × 2 × 2	58.10	53.40	72.30	66.10
Volume >30					
Screen sizes **(S)**		99.80	99.30	99.80	99.70
S × Stereo **(ST)**		99.80	98.70	99.80	99.00
S × ST × Flat and square tube **(F)**		95.10	97.60	93.60	98.00
Brand **(B)**		80.80	74.50	90.60	79.80
B × S		72.60	67.20	85.30	73.30
B × S × ST		70.80	64.60	84.30	72.30
B × S × ST × F		64.60	62.30	76.90	70.00
Outlet-type **(OT)** × S		99.70	99.10	99.70	99.60
OT × S × ST		99.70	98.40	99.70	98.90

Table 1. Continued

	Number of cells for matching	Percent of UK matched	Percent of French models matched	Percent of UK value matched	Percent of French value matched
OT × S × ST × F		95.10	97.30	93.60	97.90
OT × B		76.50	70.60	88.50	77.50
OT × B × S		64.60	56.40	78.80	64.90
OT × B × S × ST		62.30	53.50	77.60	62.70
OT × B × S × ST × F		56.80	51.20	70.60	61.30
Volume >100					
Screen sizes **(S)**		99.80	99.50	99.80	99.70
S × Stereo **(ST)**		99.80	98.90	99.80	99.00
S × ST × Flat and square tube **(F)**		95.20	97.90	93.70	98.00
Brand **(B)**		77.40	73.90	88.50	79.10
B × S		71.50	56.50	82.60	62.70
B × S × ST		67.20	53.40	76.90	60.20
B × S × ST × F		60.00	51.10	69.40	58.60
Outlet-type **(OT)** × S		98.80	98.60	98.90	99.00
OT × S × ST		98.80	98.00	98.90	98.30
OT × S × ST × F		93.70	96.90	92.20	97.50
OT × B		72.50	69.40	84.00	76.60
OT × B × S		58.50	48.70	72.40	53.50
OT × B × S × ST		55.90	45.50	68.40	51.10
OT × B × S × ST × F		49.90	43.90	61.60	50.10
Volume >1,000					
Screen sizes **(S)**		100.00	98.10	100.00	96.00
S × Stereo **(ST)**		97.10	96.90	97.80	95.20
S × ST × Flat and square tube **(F)**		97.10	96.30	97.80	95.00

Table 1. Continued

	Number of cells for matching	Percent of UK matched	Percent of French models matched	Percent of UK value matched	Percent of French value matched
Brand **(B)**		80.20	58.90	90.50	66.50
B × S		60.50	29.40	66.00	29.10
B × S × ST		53.50	24.50	56.90	25.80
B × S × ST × F		53.50	24.50	56.90	25.80
Outlet-type **(OT)** × S		95.70	80.90	97.90	76.00
OT × S × ST		92.90	80.30	95.80	75.70
OT × S × ST × F		92.90	79.70	95.80	75.40
OT × B		64.70	44.10	74.00	51.80
OT × B × S		40.80	20.80	43.70	18.30
OT × B × S × ST		36.60	17.10	38.30	16.20
OT × B × S × ST × F		36.60	17.10	38.30	16.20

Table 2. Coverage of matched comparisons: France/Netherlands

	Number of cells for matching	Percent of Netherlands models matched	Percent of French models matched	Percent of Netherlands value matched	Percent of French value matched
All data – matching by:					
Model number	2,146 × 1,186	0.42	0.27	0.01	0.01
Screen sizes **(S)**	20	99.90	99.80	99.90	99.90
S × Stereo **(ST)**	20 × 2	97.00	99.50	97.20	99.90
S × ST × Flat and square tube **(F)**	20×2×2	96.90	98.60	97.20	99.40
Brand **(B)**	80	86.70	78.50	93.50	82.20
B × S	80 × 20	77.10	71.80	90.00	77.50
B × S × ST	80 × 20 × 2	66.20	61.20	80.40	69.50
B × S × ST × F	80 × 20 × 2 × 2	66.20	60.90	80.40	69.40
Outlet-type **(OT)** × S	3 × 20	99.80	97.60	99.90	98.60
OT × S × ST	3 × 20 × 2	96.40	94.70	96.90	96.90
OT × S × ST × F	3 × 20 × 2 × 2	96.20	94.20	96.90	96.50
OT × B	3 × 80	79.40	63.60	89.40	72.50
OT × B × S	3 × 80 × 20	70.20	51.20	80.90	62.30
OT × B × S × ST	3 × 80 × 20 × 2	52.70	39.80	65.80	52.30
OT × B × S × ST × F	3 × 80 × 20 × 2 × 2	52.70	39.70	65.80	52.30
Vol > 30					
Screen sizes **(S)**		99.80	99.60	99.90	99.50
S × Stereo **(ST)**		97.30	96.10	97.00	96.70
S × ST × Flat and square tube **(F)**		97.30	95.40	97.00	96.20
Brand **(B)**		79.80	74.20	88.30	78.20
B × S		70.50	65.50	83.80	72.40
B × S × ST		54.40	53.80	66.10	63.40
B × S × ST × F		54.40	53.50	66.10	63.30
Outlet-type **(OT)** × S		99.80	91.90	99.90	90.00
OT × S × ST		95.70	79.10	96.10	79.70

Table 2. Continued

	Number of cells for matching	Percent of Netherlands models matched	Percent of French models matched	Percent of Netherlands value matched	Percent of French value matched
OT × S × ST × F		95.70	78.40	96.10	79.30
OT × B		68.40	54.00	79.90	62.70
OT × B × S		58.50	36.90	70.40	45.20
OT × B × S × ST		42.30	25.10	49.30	33.00
OT × B × S × ST × F		42.30	25.00	49.30	33.00
Vol > 100					
Screen sizes **(S)**		100.00	96.50	100.00	96.30
S × Stereo **(ST)**		96.90	95.40	96.90	96.00
S × ST × Flat and square tube **(F)**		96.90	94.80	96.90	95.60
Brand **(B)**		78.20	67.20	88.10	73.50
B × S		67.60	51.40	82.00	56.50
B × S × ST		52.70	42.60	65.30	49.90
B × S × ST × F		52.70	42.50	65.30	49.80
Outlet-type **(OT)** × S		99.60	82.50	99.80	76.90
OT × S × ST		96.50	62.50	96.90	58.60
OT × S × ST × F		96.50	61.40	96.90	57.90
OT × B		65.90	45.00	74.30	55.30
OT × B × S		57.40	24.20	70.90	29.50
OT × B × S × ST		42.10	17.40	50.10	22.20
OT × B × S × ST × F		42.10	17.40	50.10	22.20
Vol > 1,000					
Screen sizes **(S)**		93.40	98.10	93.30	99.00
S × Stereo **(ST)**		80.40	90.10	78.70	93.20
S × ST × Flat and square tube **(F)**		80.40	89.50	78.70	92.90

Table 2. Continued

	Number of cells for matching	Percent of Netherlands models matched	Percent of French models matched	Percent of Netherlands value matched	Percent of French value matched
Brand (**B**)		78.20	41.10	88.70	48.70
B × S		65.20	35.50	78.60	43.90
B × S × ST		52.10	30.00	59.70	36.70
B × S × ST × F		52.10	30.00	59.70	36.70
Outlet-type (**OT**) × S		91.30	39.80	88.70	33.90
OT × S × ST		78.20	36.80	74.10	32.10
OT × S × ST × F		78.20	36.10	74.10	31.80
OT × B		65.20	10.40	75.80	9.20
OT × B × S		43.40	8.50	48.80	7.50
OT × B × S × ST		36.90	8.00	39.00	7.00
OT × B × S × ST × F		36.90	8.00	39.00	7.00
Screen sizes (**S**)					

Table 3. Coverage of matched comparisons: Netherlands/UK

	Number of cells for matching	Percent of UK models matched	Percent of Netherlands models matched	Percent of UK value matched	Percent of Netherlands value matched
All data					
Matching by:					
Model number	1,495 × 1,186	1.87	1.60	0.15	0.03
Screen sizes **(S)**	20	99.60	99.60	99.80	99.20
S × Stereo **(ST)**	20 × 2	99.60	88.60	99.80	91.40
S × ST × Flat and square tube **(F)**	20 × 2 × 2	94.20	88.40	92.70	91.20
Brand **(B)**	80	80.70	74.40	90.70	87.30
B × S	80 × 20	69.80	68.80	82.10	85.00
B × S × ST	80 × 20 × 2	55.10	47.50	59.00	62.30
B × S × ST × F	80 × 20 × 2 × 2	51.10	46.20	52.40	61.30
Outlet-type **(OT)** × S	3 × 20	98.30	99.40	98.20	99.90
OT × S × ST	3 × 20 × 2	95.80	88.50	96.20	91.40
OT × S × ST × F	3 × 20 × 2 × 2	90.50	88.10	89.70	91.20
OT × B	3 × 80	60.50	70.70	73.80	84.30
OT × B × S	3 × 80 × 20	48.90	60.70	60.70	79.10
OT × B × S × ST	3 × 80 × 20 × 2	38.00	39.40	41.00	57.70
OT × B × S × ST × F	3 × 80 × 20 × 2 × 2	35.00	38.50	37.50	56.80
Vol > 30					
Matching by:					
Screen sizes **(S)**		99.30	99.40	98.10	99.90
S × Stereo **(ST)**		97.20	91.80	95.30	91.70
S × ST × Flat and square tube **(F)**		91.40	91.60	88.10	91.60
Brand **(B)**		66.00	69.80	73.40	85.40
B × S		57.80	61.20	68.30	79.70
B × S × ST		47.30	43.60	53.50	59.10

Table 3. Continued

	Number of cells for matching	Percent of UK models matched	Percent of Netherlands models matched	Percent of UK value matched	Percent of Netherlands value matched
B × S × ST × F		42.90	42.10	47.00	57.70
Outlet-type **(OT)** × S		93.20	99.20	89.70	99.90
OT × S × ST		77.20	91.60	74.30	91.70
OT × S × ST × F		72.60	91.20	69.60	91.40
OT × B		44.00	61.70	52.60	78.00
OT × B × S		31.60	50.30	43.10	71.00
OT × B × S × ST		21.90	34.80	28.60	50.40
OT × B × S × ST × F		19.30	33.90	25.80	49.70
Vol > 100					
Matching by:					
Screen sizes **(S)**		97.90	100.00	95.00	100.00
S × Stereo **(ST)**		97.10	93.10	94.90	92.20
S × ST × Flat and square tube **(F)**		91.70	93.10	88.30	92.20
Brand **(B)**		60.80	70.00	69.30	84.40
B × S		51.00	56.80	63.20	74.10
B × S × ST		41.80	42.50	49.20	57.70
B × S × ST × F		37.60	41.80	43.30	57.20
Outlet-type **(OT)** × S		79.60	97.20	73.20	95.00
OT × S × ST		61.90	92.10	58.50	90.50
OT × S × ST × F		57.40	91.80	53.80	90.30
OT × B		37.20	63.90	47.40	80.00
OT × B × S		23.90	52.40	32.10	70.60
OT × B × S × ST		18.00	39.10	22.90	52.70
OT × B × S × ST × F		15.80	38.40	21.10	52.30

Table 3. Continued

	Number of cells for matching	Percent of UK models matched	Percent of Netherlands models matched	Percent of UK value matched	Percent of Netherlands value matched
Vol > 1,000					
Matching by:					
Screen sizes **(S)**		97.10	91.30	98.50	88.70
S × Stereo **(ST)**		81.70	82.60	81.30	76.90
S × ST × Flat and square tube **(F)**		81.70	82.60	81.30	76.90
Brand **(B)**		57.70	76.00	71.60	88.20
B × S		46.40	43.40	52.10	38.20
B × S × ST		26.70	28.20	22.90	19.20
B × S × ST × F		26.70	28.20	22.90	19.20
Outlet-type **(OT)** × S		42.20	84.70	47.70	86.20
OT × S × ST		35.20	67.30	40.20	64.20
OT × S × ST × F		35.20	67.30	40.20	64.20
OT × B		29.50	76.00	39.20	88.20
OT × B × S		21.10	32.60	26.00	21.90
OT × B × S × ST		11.20	26.00	10.50	15.40
OT × B × S × ST × F		11.20	26.00	10.50	

brand and screen–size, which fell to about 81% when outlet-type, brand and screen size were used.

Price collectors may find a match for items, but these may be unrepresentative and have relatively low sales. They may be items, for example, at the end of their life cycle and be unusually priced given their characteristics. Thus Table 1 replicated the analysis to only include items which are more popular, with at least sales of 30 in the period. The coverage can be seen to remain quite similar to, if a little lower, than that based on no such restriction. The exercise was further replicated to only include items with sales over 100, and then over 1,000. Again if the selection was by screen size, stereo and flat and square tubes, the coverage remained high. When the outlet-type was included in the specification, the coverage remained high for quantities larger than 100, but was reduced to 75% by (sales) value for France for quatities larger than 1,000. Restrictions to the more popular models have a serious effect on coverage when brand is used. For outlet-type by brand, screen size, and stereo the coverage by value for French matching with the UK was reduced to 50% for volume restricted to 100, falling to 16% for volume restricted to 1,000.

Similar overall results apply for comparisons between France and the Netherlands and the Netherlands and the UK, as shown in Tables 2 and 3. However, the depletion in coverage is more severe. Disaggregation by screen-size, stereo and flat and square tube maintains a high coverage for comparisons between France and the Netherlands, though less so between the Netherlands and the UK, with the value matched in each country falling by nearly 10%. Item specification involving brand (OT × B × S × St × F) results in a fall-off to nearly 50% by value for France matched with the Netherlands, while the same item specification for the Netherlands/UK comparison results in a fall to 37.5% for the UK value matched with the Netherlands. The position worsens as only more popular models are sampled, the latter figure falling to 10% when sampling is for quantities larger than 1,000 and about 20% for quantities larger than 100.

Thus the criteria by which items are specified can have a major effect on the coverage of the matched comparisons. Matching by brand results in particularly severe degradation in the coverage of the prices sampled for the index. There can also be seen to be an absence of symmetry between loss in coverage, the screen-size and stereo matching restricting the UK coverage to 99.8% and the Netherlands coverage to 91.4% for a Netherlands/UK comparison. Increasing the selection to more popular items has a severe effect on coverage when brand is involved, though less so when only screen-size, stereo and flat and square tube is used. Indeed, the latter specification was negligible for the France/UK comparison, but much more so–a 20-25 percentage point fall-off–for other comparisons.

6. Empirical Results: Matched versus Unmatched Prices

It was seen from Equation (2) that bias depended on two things. First, was the proportion of observations covered, or if weighted least squares is used as the estimator of the hedonic CPD, the proportion by sales share of the sample covered. Second, was the difference between the quality-adjusted prices of the matched and unmatched comparisons covered. Hedonic regressions of the form outlined in Equation (1) above were run for each bilateral country comparison with dummy variables for the unmatched items. The diagnostic summary and test statistics are given in Table 4. The hedonic regression included about 75 variables (listed in Section 4), the model fitting well by criteria of $\overline{R}^2$ (above 0.9 for the semi-logarithmic form used), and the signs on the coefficients accorded with *a priori* expectations (details available from authors). The semi-logarithmic form was preferred over the linear form according to both the Bera-McAleer test (Maddala 1988, p. 179) and a Box-Cox test (Dougherty 1992, p. 132). The null hypotheses of homoskedasticity and normality of the residuals could not be rejected at a 5% level for the semi-logarithmic form by an LM test, based on regressing squared residuals on a constant and squared predicted values, and the Jarque-Bera test, respectively.

The equation was estimated with a weighted (by sales share) least squares estimator (WLS). Relative sales *values* as opposed to relative sales quantities have been shown to be preferable for weights for hedonic *indices* by Diewert (2002) on the grounds of representativity. Cheaper models are argued to have an unduly large weight if quantity weights are used. Hedonic indices using relative value weights have a close correspondence to superlative formulations (Diewert 2002). Silver (2002) also supports the use of relative value shares for weights for hedonic indices but draws attention to the possible disproportionate influence on the index of observations with unduly high leverage.

Matched and unmatched models were defined by the different combinations of specification of characteristics used in Tables 1 to 3. Consider, for example, the first row of the matching by "model number" for the France/UK comparisons in Table 5. A dummy variable was included in the hedonic regression if a model was sold in France and was not sold (unmatched) in the UK, and a further dummy variable was included if one was sold in the UK and was unmatched in France. The coefficients and *t*-statistics on these dummy variables from the hedonic regression are reported in the first row of the first four columns of figures in Table 5.

A French model unmatched in the UK can be seen from the first row of Table 5 to have no statistically significant difference ($t = 0.65$) in quality-adjusted prices on average from the benchmarked matched models, while a UK model unmatched in France is $[1 - \exp(0.07)] \times 100 = 7.3\%$ more expensive even allowing for hedonic, quality adjustments. When matching by model number it is not just that the sample degrades, it is also that quality-adjusted prices for unmatched French models are on average higher and, following Equation (2), further bias a matched models index.

Table 4. Diagnostic tests on regressions

Test	UK/N		UK/F		FR/N	
	Semi-log	Linear	Semi-log	Linear	Semi-log	Linear
$\overline{R}^2$	0.924	0.802	0.960	0.750	0.928	0.780
Bera- McAleer Test (*p*-values)	0.031	0.000	0.33	0.000	0.852	0.000
Box-Cox -χ^2 (*p*-values)		0.000		0.000		0.000
Hetroskedasticity – LM test (*p*-values)	0.68	0.000	0.27	0.000	0.000	0.000
Normality – Jarque-Bera (*p*-values)	0.000	0.000	0.000	0.000	0.000	0.000

Table 5. Comparison of matched and unmatched quality-adjusted prices

	Hedonic coefficient		Hedonic coefficient		Mean residuals			Hedonic indices	
	unmatched UK	*t*-stat-istic	unmatched France	*t*-stat-istic	unmatched UK	unmatched France	matched	matched data	all data
	UK/France								
Matching by:									
Model number	−0.02	−0.65	−0.07	−3.10	0.00	0.00	0.04	9.50	9.74
Screen sizes **(S)**	0.00	0.00	0.49	8.17	0.00	0.00	0.00	9.75	9.74
S × Stereo **(ST)**	0.00	0.00	0.08	1.12	0.00	0.05	0.00	9.77	9.74
S × ST × Flat & square tube **(F)**	0.04	1.46	0.08	1.57	0.02	0.05	0.00	9.79	9.74
Brand **(B)**	−0.14	−4.77	−0.34	−15.90	0.00	−0.09	0.01	10.01	9.74
B × S	0.06	1.73	−0.24	−14.00	0.01	−0.08	0.02	10.07	9.74
B × S × ST	0.01	0.39	−0.19	−11.50	0.00	−0.06	0.02	10.08	9.74
B × S × ST × F	0.00	−0.16	−0.17	−10.70	0.00	−0.06	0.02	10.04	9.74
Outlet-type **(OT)** × S	0.68	1.40	−0.03	−0.21	0.26	−0.01	0.00	9.77	9.74
OT × S × ST	0.68	1.40	0.06	0.97	0.26	0.04	0.00	9.79	9.74
OT × S × ST × F	0.07	1.88	0.07	1.43	0.03	0.04	0.00	9.81	9.74
OT × B	0.06	1.10	−0.28	−14.00	0.00	−0.08	0.02	10.05	9.74
OT × B × S	0.04	1.61	−0.15	−10.21	0.01	−0.05	0.02	10.12	9.74
OT × B × S × ST	0.04	1.90	−0.13	−9.20	0.01	−0.04	0.02	10.14	9.74
OT × B × S × ST × F	0.02	1.19	−0.12	−8.62	0.01	−0.04	0.02	10.10	9.74
	France/Netherlands								
Matching by:									
Model number	−0.04	−1.10	−0.06	−0.79	0.00	0.00	0.05	NA	3.16
Screen sizes **(S)**	0.00	0.00	1.32	4.52	0.00	0.00	0.00	3.16	3.16
S × Stereo **(ST)**	−0.04	−0.77	−0.08	−1.26	−0.03	−0.02	0.00	3.15	3.16
S × ST × Flat & square tube **(F)**	−0.08	−1.46	0.07	1.62	−0.05	0.05	0.00	3.15	3.16
Brand **(B)**	−0.14	−5.40	−0.20	−10.80	−0.04	−0.07	0.01	3.17	3.16

Table 5. Continued

	Hedonic coefficient		Hedonic coefficient		Mean residuals			Hedonic indices	
	unmatched UK	*t*-statistic	unmatched France	*t*-statistic	unmatched UK	unmatched France	matched	matched data	all data
B × S	−0.12	−6.10	−0.18	−11.50	−0.04	−0.07	0.02	3.19	3.16
B × S × ST	−0.04	−3.00	−0.08	−5.40	−0.01	−0.02	0.01	3.18	3.16
B × S × ST × F	−0.04	−2.98	−0.08	−5.30	−0.01	−0.02	0.01	3.18	3.16
Outlet-type **(OT)** × S	0.00	0.00	0.09	2.10	0.00	0.05	0.00	3.16	3.16
OT × S × ST	0.01	0.33	0.05	1.65	0.01	0.02	0.00	3.16	3.16
OT × S × ST × F	−0.01	−0.20	0.07	2.39	−0.01	0.03	0.00	3.16	3.16
OT × B	−0.09	−4.10	−0.12	−8.30	−0.03	−0.04	0.02	3.19	3.16
OT × B × S	−0.09	−5.00	−0.12	−9.50	−0.03	−0.04	0.03	3.18	3.16
OT × B × S × ST	−0.02	−1.57	−0.04	−3.67	−0.01	−0.01	0.01	3.17	3.16
OT × B × S × ST × F	−0.02	−1.56	−0.04	−3.63	−0.01	−0.01	0.01	3.17	3.16
	Netherlands/UK								
Matching by:									
Model number	−0.04	−0.65	−0.49	−0.15	0.00	0.00	0.02	NA	3.16
Screen sizes **(S)**	2.29	68.10	−0.08	−0.71	0.00	0.00	0.00	3.16	3.16
S × Stereo **(ST)**	2.30	67.60	−0.04	−1.46	0.00	−0.02	0.00	3.17	3.16
S × ST × Flat & square tube **(F)**	0.15	4.70	−0.04	−1.52	0.07	−0.02	0.00	3.20	3.16
Brand **(B)**	0.10	1.93	−0.24	−8.67	0.00	−0.07	0.01	3.30	3.16
B × S	0.00	0.18	−0.16	−7.11	0.00	−0.06	0.01	3.30	3.16
B × S × ST	0.02	1.36	−0.08	−4.60	0.01	−0.02	0.01	3.26	3.16
B × S × ST × F	0.03	1.91	−0.07	−4.40	0.01	−0.02	0.01	3.38	3.16
Outlet-type **(OT)** × S	0.17	2.23	−0.51	−4.30	0.07	−0.12	0.00	3.17	3.16
OT × S × ST	0.12	2.69	−0.04	−1.49	0.06	−0.02	0.00	3.17	3.16

Table 5. Continued

	Hedonic coefficient		Hedonic coefficient		Mean residuals			Hedonic indices	
	unmatched UK	*t*-statistic	unmatched France	*t*-statistic	unmatched UK	unmatched France	matched	matched data	all data
OT × S × ST × F	0.14	5.30	−0.05	−1.79	0.07	−0.03	0.00	3.21	3.16
OT × B	0.03	1.35	−0.21	−8.19	0.01	−0.06	0.01	3.27	3.16
OT × B × S	0.04	2.15	−0.12	−6.22	0.01	−0.04	0.01	3.29	3.16
OT × B × S × ST	0.06	3.36	−0.07	−3.98	0.01	−0.02	0.00	3.29	3.16
OT × B × S × ST × F	0.06	3.45	−0.06	−4.04	0.01	−0.02	0.00	3.32	3.16

*The results equal those of France/Netherlands at two decimal places, but this is coincidental.

Table 5 includes results for matching by other combinations of specifications for the France/UK comparison. It is the *French unmatched* television sets that differ from the matched ones over most forms of matching and the differences are invariably negative, that is, even when quality-adjusted by the hedonic regression, unmatched prices in France are on average lower than their UK counterparts. Bear in mind that the regression in Equation (1) included a country dummy so the results are not an artefact of the currency exchange rates. Note, however, that the hedonic-adjusted prices of unmatched UK models only differ when brand is used in the item specification and their average quality-adjusted price is $[1 - \exp(-0.14] \times 100 = 13.1\%$ *lower* than that of the benchmark matched ones. Not only did specifying items by brand reduce coverage (Tables 1 to 3), but the excluded items, at least for France, have quite different quality-adjusted prices (Table 5).

A *single* regression of the form given in Equation (1) using WLS was also estimated using models in both countries compared; i.e., for the UK and France in the first part of Table 5. However, the mean residuals were separately calculated for the unmatched UK, unmatched French and matched (quality-adjusted) residual prices. Table 5 shows that both the UK and the French mean unmatched residuals were lower than the matched residuals when brands were involved in the item specification, the picture being mixed otherwise. Thus the use of brands for matching excludes unmatched models whose quality-adjusted residual prices are quite different from matched ones.

7. Empirical Results: Parity Measures

The final two columns of Table 5 are the hedonic indices using just matched data and using all the data. The former corresponds to a quality-adjusted geometric mean on the matched data. The results, as expected, show very little discrepancy when brand is not used as an item specification. For example, for screen size, stereo and flat and square tube the matched sample yields a parity estimate of 9.79 for the French *franc* (at the time of the study) and the UK *pound sterling*, while for all the data it is 9.74. The coverage in this case remained high and the unmatched quality-adjusted prices were similar to the matched ones. However, where brand was involved the discrepancies were larger. For example, for outlet-type and brand, the matched estimate and the estimate using all the data were 10.05 and 9.74 respectively, a larger difference but not a substantial one. The comparison between France and the Netherlands should be expected to give cause for concern since *both* unmatched Netherlands and unmatched France are lower than their matched counterparts (Table 5). But this is mainly for item specifications including brand and, furthermore, since they are both negative, Equation (2) shows that some of the bias cancels, even if the loss of coverage is substantial. Table 2 showed coverage by (Ot $\times$ B $\times$ S) to be reduced to under two-thirds of French sales value, and Table 5 found both French and the Netherlands unmatched mean residuals to be negative, the effects, to a large extent, cancelling each other. Parity estimates for the French *franc* and the Netherlands *guilder* (at the time of study) for the matched data and all of the data were very close at 3.18 and 3.16 respectively. Loose specifications with a low level of coverage and differential prices of unmatched items need not therefore lead to bias. It depends on the nature of the differences in prices.

However, for the Netherlands/UK comparison Table 5 found matched UK items to be priced above matched ones, and unmatched Netherlands prices to be below matched ones for a number of item specifications. In such instances, where the sample degradation was serious, the bias proved to be more marked. For example, item specification by (B × S × St × F) led to an estimate of the parity for the Netherlands *guilder* and the *pound sterling* using matched data of 3.38 compared with 3.16 for all data – not a substantial difference by the standard of these things, but larger than previously considered.

8. Conclusions

The subject of this study was the potential for bias in PPP indices when tight item descriptors/specifications are used for the matching of items between countries whose prices are to be compared. Tight item descriptors are desirable to ensure the prices of "like" are compared with "like," but undesirable in that the coverage of the matched sample is degraded.

The trade-off between tight and loose (discretional) product specifications is thus of some practical concern. The article first outlined an analytical framework which demonstrated the factors leading to bias from using a matched sample. Two things were found to be important: first, the extent to which the sample coverage degraded as a result of constraining the sample to matched models, and second, the extent and nature of the difference between Country A average prices unmatched in Country B, and Country B prices unmatched in Country A, after hedonic adjustments. It is a subject not conducive to empirical research since data are generally only available on the matched sample collected. The use of scanner data allowed matching to take place at different degrees of tightness of specification and the (quality-adjusted) prices of out-of-sample items to be compared. It was found that:

- *Matching involving brand, especially for popular models, had a particularly severe degradation in coverage* (Tables 1–3). In other cases the coverage for this product was resilient to different matching criteria and, as a result, errors in parity estimates due to poor coverage for nonbrand matching were quite minor (Table 5).
- *The bias from poor coverage was compounded when average unmatched new/old (quality-adjusted) prices are asymmetric.* For example, if lower (quality-adjusted) priced local brands dominate the market in one country and higher (quality-adjusted) priced global brands dominate in another, then a price comparison of a, say, single "intermediate valued" brand will exclude the below and above average unmatched respective prices, compounding the bias via Equation (2). Price statisticians should attempt to increase the coverage of matched prices and/or make quality adjustments to prices for "brand effects" in such circumstances. In this study, for the France/UK comparison, the French unmatched average prices were on average lower than their British counterparts, even when quality-adjusted, and such differences were exacerbated when brands were part of the selection criteria.
- *Poor coverage need not always lead to bias*. The analytical framework in (2) and results showed that if the quality-adjusted unmatched prices are both lower (as with France and the Netherlands in Table 5) *or* both higher than the matched ones, the bias will tend to be offsetting. For example, if lower (quality-adjusted) priced local brands

dominate the markets in Countries A and B and a higher (quality-adjusted) priced global brand is compared, the bias from the exclusion of the unusually priced unmatched local brands will be offsetting.

Decisions as to which product specifications are to be used for matching in PPP studies, and thus which will be unmatched and excluded, therefore need to be based on the effect on coverage of the specifications used, especially for brand, and any priors as to the likely (quality-adjusted) price positioning of matched and unmatched models in the countries concerned.

9. References

Aizcorbe, A., Corrado, C., and Doms, M. (2000). Constructing Price and Quantity Indexes for High TechnoIogy Goods, Industrial Output Section, Division of Research and Statistics, Board of Governors of the Federal Reserve System, July.

Aten, B.H. (1999). Cities in Brazil: An Interarea Price Comparison. International and Interarea Comparisons of Prices, Output and Productivity, A. Heston and R. Lipsey (eds). 61. Chicago: University of Chicago Press, National Bureau of Economic Research, Committee on Research on Income and Wealth, 211–266.

Balk, B.M. (2001). Aggregation Methods in International Comparisons, What Have We Learned. Erasmus Institute of Management Report Series, June.

Berndt, E.R. (1990). The Practice of Econometrics: Classic and Contemporary. Reading, Mass: Addison-Wesley Publishing Company.

Berndt, E.R., Grilches, Z., and Rappaport, N.J. (1995). Econometric Estimates of Price Indices for Personal Computers in the 1990s. Journal of Econometrics, 68, 243–268.

Diewert, W.E. (2003). Hedonic Regressions: A Consumer Theory Approach. Scanner Data and Price Indexes, M. Shapiro and R. Feenstra (eds). 64. Chicago: University of Chicago Press, National Bureau of Economic Research, Studies in Income and Wealth, 317–348.

Diewert, W.E. (2002). Hedonic Regressions: A Review of Some Unresolved Issues. Mimeo, Department of Economics, University of British Columbia.

Dougherty, C. (1992). Introduction to Econometrics. Oxford: Oxford University Press.

Fenwick, D., Ball, A., Silver, M., and Morgan, P. (2003). Price Collection and Quality Assurance of Item Selection in the Retail Prices Index: How Can Scanner Data Help? Scanner Data and Price Indices, M. Shapiro and R. Feenstra (eds). 64. Chicago: University of Chicago Press, National Bureau of Economic Research, Studies in Income and Wealth, 67–88.

Griliches, Z. (1990). Hedonic Price Indices and the Measurement of Capital and Productivity: Some Historical Reflections. Fifty Years of Economic Measurement: The Jubilee Conference of Research in Income and Wealth 54. Chicago: University of Chicago Press, National Bureau of Economic Research, Studies in Income and Wealth, 185–202.

Heravi, S., Heston, A., and Silver, M. (2003). Using Scanner Data to Estimate Country Price Parities: An Exploratory Study. Review of Income and Wealth, 49, 1–22, March.

Kokoski, M.F., Moulton, B.R., and Zieschang, K.D. (1999). Interarea Price Comparisons for Heterogeneous Goods and Several Levels of Commodity Aggregation. In

International and Interarea Comparisons of Prices, Output and Productivity, A. Heston and R. Lipsey (eds). 61. Chicago: University of Chicago Press, National Bureau of Economic Research, Committee on Research on Income and Wealth, 123–166.

Maddala, G.S. (1988). Introduction to Econometrics. London: Maxwell-McMillan International Edition.

Rao, P.D.S. and Timmer, M.P. (2003). Purchasing Power Parities for Industry Comparisons Using Weighted Elteto-Koves-Szulc (EKS) Methods. Review of Income and Wealth, 49, 491–511.

Silver, M. (2002). The Use of Weights in Hedonic Regressions: The Measurement of Quality Adjusted Price Changes. Mimeo, Cardiff Business School, Cardiff University.

Silver, M. and Heravi, S. (2001). Scanner Data and the Measurement of Inflation. The Economic Journal, 111, 383–404.

Silver, M. and Heravi, S. (2005). A Failure in the Measurement of Inflation: Results from a Hedonic and Matched Experiment using Scanner Data. Journal of Business and Economic Statistics, 23, 269–281.

Summers, R. (1973). International Comparisons with Incomplete Data. The Review of Income and Wealth, March.

Triplett, J.E. (1987). Hedonic Functions and Hedonic Indices. The New Palgraves Dictionary of Economics, 630–634.

Triplett, J.E. (2004). Handbook on Quality Adjustment of Price Indexes for Information and Communication Technology Products. OECD Directorate for Science, Technology and Industry, Draft, OECD, Paris.

Triplett, J.E. and McDonald, R.J. (1977). Assessing the Quality Error in Output Measures: The Case of Refrigerators. The Review of Income and Wealth, 23, 137–156.

World Bank (2004). International Comparisons Program Handbook 2004. Washington D.C.: World Bank. Available by clicking on 'International Comparisons Program' on http://www.worldbank.org/data/

Zieschang, K.D., Armknecht, P.A., and Smith, D. (2001). Integrating Inter-Area and International Price Comparisons with Consumer Price Index Compilation. Joint World Bank-OECD Seminar on Purchasing Power Parities: Recent Advances in Methods and Applications. Washington D.C., 30 January-2 February. http://www.oecd.org/dataoecd/24/4/2424626.pdf

Received December 2003
Revised September 2004