Assessing Efficiency of Elementary Indices with Monte Carlo Simulations

(EKS, EKS-STAR, EKS-S, CPD and CPD-WEIGHTED)

Yuri Dikhanov (World Bank)

This paper compares response of various elementary aggregators such as EKS, EKS-S, EKS-STAR, CPD and CPD-WEIGHTED to stochastically generated inputs. The inputs were designed to simulate cases happening in reality, with widely varied densities of price and representativity matrices. When price matrix is dense and representativity of prices is close to 100%, all indices behave in a similar way. In general, CPD-WEIGHTED index is found to be the most robust, especially with sparse price and representativity matrices. Additionally, all EKS-style indices, if unmodified, had high failure rates.

Introduction

Recently there have been discussions on what elementary aggregator to use at the below basic heading level. The current paper differs from them in method. The paper attacks the issue of index efficiency via a Monte Carlo simulation.

In short, two inputs are randomly generated - Price matrix $P=[p^i_j]$, and Weight matrix $\Omega=[\omega^i_j]$ - and run through various aggregators.

Both weighted and unweighted indices are considered. Two unweighted indices – the CPD and EKS are employed. For these indices weight matrix predictably does not have any effect.
Three weighted indices are used alongside: CPD-WEIGHTED[1], EKS-STAR and EKS-S [Sergeev]. Those five indices cover most of the variety of the below basic heading aggregators proposed for use in international comparisons. In addition, some modifications for the EKS indices are suggested to reduce their failure rate and increase their applicable range.

Equivalence of two CPD forms

CPD index[2] can be presented in two equivalent forms – with the intercept and without. First, the regression equation for the CPD can be written as

$$\ln p_{cp} = y_{cp} = x_{cp}\beta + \varepsilon_{cp} \qquad \textbf{(1)}$$

where p_{cp} - price of product *p* in country *c*;

Dc_j and Dp_i - country and product dummies;

Np and *Nc* – number of products and countries, respectively;

$$x_{cp} = \left[Dc_2 \ldots Dc_{Nc} Dp_1 Dp_2 \ldots Dp_{Np} \right]$$
$$\beta = \left[\alpha_2 \ldots \alpha_{Nc} \gamma_1 \gamma_2 \ldots \gamma_{Np} \right]^T \qquad \textbf{(2)}$$

In matrix notation, by stacking individual observations, this can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \textbf{(3)}$$

Note that the first country dummy is dropped from the system because matrix X is of rank *(Np+Nc-1)*[3] [in fact, we can drop any variable from the system, dropping the first country's dummy simply makes it the base country].

The solution is given (under the conditions of independently and identically distributed random disturbances) by

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \qquad \textbf{(4)}$$

In addition, we can drop one product variable (let's say the first product dummy) and introduce the intercept. This is the second form of the CPD. In this case[4],

[1] Introduced by Cuthbert, called elsewhere CPRD [R for representativity]. Can be easily generalized to include any values for weights [then it could be called CPWD, with W for weight].

[2] Introduced by Summers (1973). For a thorough discussion see Rao (2004) and Diewert (2004).

[3] The sum of country dummies equals the sum of product dummies, so one dummy has to be dropped.

[4] Note that the sign **(′)** does not mean transpose. Sign **(^{T})** is used for that purpose.

$$x'_{cp} = \left[Dc_2 \ldots Dc_{Nc} \; 1 \; Dp_2 \ldots Dp_{Np} \right] \quad \textbf{(5)}$$
$$\beta' = \left[\alpha'_2 \ldots \alpha'_{Nc} \; c_{\text{intercept}} \; \gamma'_2 \ldots \gamma'_{Np} \right]^T$$
$$\mathbf{y} = \mathbf{X}'\boldsymbol{\beta}' + \boldsymbol{\varepsilon} \quad \textbf{(6)}$$

The solutions of systems (4) and (6) we are interested in - the country and product price relatives - are identical up to a scalar in these two cases. In the case with intercept,

$$\alpha_j = \alpha'_j$$
$$\gamma_1 = c_{\text{intercept}}$$
$$\gamma_i = \gamma'_i + c_{\text{intercept}},$$

for $i = 2 \ldots Np,\; j = 2 \ldots Nc$

γ'_i and α'_j are the product coefficients for product *i* and country *j*, respectively, in the case with intercept.

Fast CPD (unweighted)

Using some basic properties of block-structured matrices it is possible to omit many computations necessary in solving for general linear systems.

First, in the case of full price matrix, matrix $(\mathbf{X}^T\mathbf{X})$ becomes[5]

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} Np & 0 & . & 0 & 1 & 1 & . & 1 \\ 0 & Np & . & 0 & 1 & 1 & . & 1 \\ . & . & . & . & . & . & . & . \\ 0 & 0 & . & Np & 1 & 1 & . & 1 \\ 1 & 1 & . & 1 & Nc & 0 & . & 0 \\ 1 & 1 & . & 1 & 0 & Nc & . & 0 \\ . & . & . & . & . & . & . & . \\ 1 & 1 & . & 1 & 0 & 0 & 0 & Nc \end{bmatrix} =$$

$$= \begin{bmatrix} Np\mathbf{I}_{Nc-1} & \mathbf{J}_{(Nc-1)\times Np} \\ \mathbf{J}_{Np\times(Nc-1)} & Nc\mathbf{I}_{Np} \end{bmatrix} \quad \textbf{(7)}$$

With obvious notation of $\mathbf{I}$ as an identity matrix and $\mathbf{J}$ as a rectangular matrix whose elements are equal to 1.

Using Graybill Theorem matrix $(\mathbf{X}^T\mathbf{X})^{-1}$ becomes[6]

[5] See Rao (2004).

$$(\mathbf{X}^T\mathbf{X})^{-1} = \quad \textbf{(8)}$$
$$= \begin{bmatrix} \frac{1}{Np}\mathbf{I}_{Nc-1} + \frac{1}{Np}\mathbf{J}_{(Nc-1)\times(Nc-1)} & -\frac{1}{Np}\mathbf{J}_{(Nc-1)\times Np} \\ -\frac{1}{Np}\mathbf{J}_{Np\times(Nc-1)} & \frac{1}{Nc}\mathbf{I}_{Np} + \frac{Nc-1}{NpNc}\mathbf{J}_{Np\times Np} \end{bmatrix}$$

Unfortunately, equations (7) and (8) describe the trivial CPD case with a full price matrix, where there is no need to conduct any matrix computation whatsoever, as the EKS produces identical results. In a general case with missing price observations, expression (8) becomes intractable. However, computation of matrix $(\mathbf{X}^T\mathbf{X})$ in the general case with a sparse matrix can still be simplified. Moreover, in solving for $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, there is no need to invert matrices, so it is only important to simplify $(\mathbf{X}^T\mathbf{X})$ in order to achieve a significant speed up in computations.

In the case with missing price observation, expression (7) becomes

$$\mathbf{X}^T\mathbf{X} =$$
$$\begin{bmatrix} Np(c_2) & 0 & 0 & \delta_2^1 & \delta_2^i & \delta_2^{Np} \\ 0 & Np(c_j) & 0 & \delta_j^1 & \delta_j^i & \delta_j^{Np} \\ 0 & 0 & Np(c_{Nc}) & \delta_{Nc}^1 & \delta_{Nc}^i & \delta_{Nc}^{Np} \\ \delta_2^1 & \delta_j^1 & \delta_{Nc}^1 & Nc(p_1) & 0 & 0 \\ \delta_2^i & \delta_j^i & \delta_{Nc}^i & 0 & Nc(p_i) & 0 \\ \delta_2^{Np} & \delta_j^{Np} & \delta_{Nc}^{Np} & 0 & 0 & Nc(p_{Np}) \end{bmatrix} \quad \textbf{(9)}$$

where $Np(c_j)$ is the number of products that country *j* priced;

$Nc(p_i)$ is the number of countries that priced product *i*;

δ_j^i is equal to one if price of product *i* is observed in country *j*, otherwise it is zero.

[6] There seems to be a typo in Rao (2004), p. 6, where he describes the upper right element of the matrix as $\frac{1}{NcNp-(Nc+Np)}\mathbf{J}_{(Nc-1)\times Np}$ instead of $-\frac{1}{Np}\mathbf{J}_{(Nc-1)\times Np}$. Similar observation is applicable for the lower left element as well.

When not using special properties of such matrices, computing matrix $\left(\mathbf{X}^T\mathbf{X}\right)$ takes the lion's share of the time, so expression (9) is a significant shortcut.

Frequency-adjusted CPD

The original paper by Summers (1973) discussed frequency-weighted CPD, which we will call frequency-adjasted CPD in order to avoid confusion over product weights (as well as representativity). The paper suggested adjusting for country-related biases, as the countries that collect more prices would have a greater influence on the results than other countries under the regular CPD. The adjustment would be proportional to the reciprocal of the number of price observations for the country. Hence, each country's observations received equal weight.

We can note that this approach is equivalent to the Generalized (or Weighted) Least Squares. For that reason, we use $\mathbf{\Omega}$ to denote the matrix of adjustments.

Matrix $\mathbf{\Omega}$ with diagonal elements ω_s :

$$\omega_s = \frac{Nc}{N_{obs}} Np(c_{j(s)})$$

where N_{obs} is total number of observations,

$Np(c_{j(s)})$ is the number of products for country *j(s)* which collected observation *s*.

$$\mathbf{X}^T\Omega^{-1}\mathbf{X} = \frac{N_{obs}}{Nc}\begin{bmatrix} 1 & 0 & 0 & \frac{\delta_2^1}{Np(c_2)} & \frac{\delta_2^i}{Np(c_2)} & \frac{\delta_2^{Np}}{Np(c_2)} \\ 0 & 1 & 0 & \frac{\delta_j^1}{Np(c_j)} & \frac{\delta_j^i}{Np(c_j)} & \frac{\delta_j^{Np}}{Np(c_j)} \\ 0 & 0 & 1 & \frac{\delta_{Nc}^1}{Np(c_{Nc})} & \frac{\delta_{Nc}^i}{Np(c_{Nc})} & \frac{\delta_{Nc}^{Np}}{Np(c_{Nc})} \\ \frac{\delta_2^1}{Np(c_2)} & \frac{\delta_j^1}{Np(c_j)} & \frac{\delta_{Nc}^1}{Np(c_{Nc})} & \sum\limits_{\forall s\in S(p_1)}\frac{1}{\omega_s} & 0 & 0 \\ \frac{\delta_2^i}{Np(c_2)} & \frac{\delta_j^i}{Np(c_j)} & \frac{\delta_{Nc}^i}{Np(c_{Nc})} & 0 & \sum\limits_{\forall s\in S(p_i)}\frac{1}{\omega_s} & 0 \\ \frac{\delta_2^{Np}}{Np(c_2)} & \frac{\delta_j^{Np}}{Np(c_j)} & \frac{\delta_{Nc}^{Np}}{Np(c_{Nc})} & 0 & 0 & \sum\limits_{\forall s\in S(p_{Np})}\frac{1}{\omega_s} \end{bmatrix} \quad \textbf{(10)}$$

Where $S(p_i)$ is a set of observations for product *i*.

Expressions (9) and (10) allow significantly speed up CPD computations. For example, computation of $\mathbf{X}^T\mathbf{X}$ is of a [(Nc*Np)*(Nc+Np)²] complexity, whereas solving $\hat{\boldsymbol{\beta}} = \mathbf{Ay}$ (the final step of computations) is only of a (Nc+Np)³/6 complexity (using the Cholesky decomposion, for example). Thus, in the case of a 100x100 matrix, computing $\mathbf{X}^T\mathbf{X}$ would require 400 mln operations vs. 1.3 mln for solving $\hat{\boldsymbol{\beta}} = \mathbf{Ay}$. Using expression (9) would cut the number of operations for computing $\mathbf{X}^T\mathbf{X}$ to 40 thousands only, speeding up the whole computation by the factor of 300(!).

Some important properties of the indices [to be extended]

1. IDENTITY: All five indices produce identical results if the price matrix is full and weight matrix is filled with some identical values.
2. IDENTITY: CPD, EKS and CPD-WEIGHTED produce the same results if the price matrix is full and weight columns sum up the same value for each country – i.e., $\sum_i \omega_j^i = const$, with weights being any number, including negative ones. It also means that if $[\omega_j^i]$ are expenditure weights summing up to 1 by country, then the results of those three indices would be identical. For the special case of representativity weights $[\omega_j^i = \{1,0\}]$ this implies that these indices will produce identical results if the number of representative products by country is the same for all countries[7].
3. WEAK SEPARABILITY: If [*n*-1] countries have full matrices ***P*** and ***W***, then for any inputs for the remaining country $p_1^i \geq 0 \wedge \omega_1^i = \{0,1\}$, the geometric mean for those [*n*-1] countries [or the remaining country's position relative to [*n*-1] countries, estimated as an unweighted geomean] when using CPD-WEIGHTED is equal to EKS-S, and when using CPD is equal to EKS.
4. STRONG SEPARABILITY: If [*n*-1] countries have full matrices ***P*** and ***W***, then for the remaining country if $p_1^i > 0 \wedge \omega_1^i = 1$, otherwise $p_1^i = 0$, [1] CPD-WEIGHTED is equal to CPD, and EKS is equal to EKS-S and EKS-STAR for all countries, and [2] that country's position [estimated as an unweighted

[7] This implies that if one can expect the price matrix to be full and the number of representative items by country will be the same, then one can use any of these indices, with EKS being preferable, of course, due to its simplicity.

geomean] relative to [*n*-1] countries does not change, or the geometric mean for those [*n*-1] countries is identical *for all five indices*.

The *separabity* conditions are important in understanding the effect of one country in a comparison on the remaining countries.

Principles of data simulation

To test the sensitivity of indices to input data, the following data were used:

1. A random matrix ***P*** is generated. With log(p)=U[0,1], where U[0,1] is uniform distribution on [0,1]. In this case expected price levels are equal for each country.
2. To simulate sparseness of the price matrix, price p_j^i is removed from ***P*** with probability $1-E_P$.
3. To simulate the effect of weights [or representativity in our special case], the following operation has been done: A product (i,j) with price p_j^i out of remaining prices is removed with probability $1-E_\Omega$ and a substitute [non-representative] with price p_j^i* is put in its place, with the assumption that log(p*)=log(p)-W, where W=1. This logic, of course, can be generalized to cover any values, not only 0 or 1. In Annex V we simulate W to be a random number as well [generated by U[0,1]].
4. The reference computation is done based on the original full price matrix ***P***.
5. The five indices are run and compared against the reference computation.
6. This operation is repeated 3,000 times for each of the [{50,100},{50,100}] (*price matrix density* and *matrix of representativity* space) grid points with step 10. The results are presented in tabular as well as in graphic form [see Figures 1-5 and Table 1 from each of Annexes]. The Figures show equi-altitude lines of the *results plane* computed from the grid points using cubic splines[8].

SIMULATION RESULTS

Simulation with the original EKSs

CPD-WEIGHTED displayed the most consistent results. CPD and EKS were close to CPD-STAR in terms of predictive power, with surprisingly little difference between CPD and EKS across the whole domain. [See Figures 2 and 5 of each of Annexes]. EKS-STAR proved to be mostly better than CPD and EKS for representativity range of [60%-90%], and mostly worse for other areas, but the difference was not significant. This result is not obvious given the fact that both CPD and EKS are unweighted.

However, this result is somewhat explained by the data model we used – i.e., we assumed that representativity incidence is independent of particular countries, so, on average, each country would have approximately the same number of representative items. In Annexes II and IV a different model is utilized with one country having 50% less representative product than the rest of the countries.

High failure rate for original EKS-STAR and EKS-S indices

These indices under strict formulation have high failure rate when the price matrix density drops and the number of representative prices decreases. It is necessary that for any country at least one representative item common with one other country existed, otherwise that country's Laspeyres fails, and the whole index becomes undefined[9]. The failure rate is presented in Figures 1.6 and 3.6[10] (for matrices of size [4, 10].

For areas around {50, 50} in *price matrix density* and *matrix of representativity* space [where possibly many or even most of the non-OECD cases would fall to], the failure rate would approach 80%. If we allow for asymmetric inputs, when countries face different probabilities E_Ω and E_P, the failure rate goes even higher. The original EKS indices can be modified to disregard dead links. However, in this case we will get systematically biased results [e.g., the countries with higher rates of Laspeyres failure – which may happen to be the least developed countries - would have artificially high

[8] This is similar to the presentation of altitude in topography.

[9] It is possible to reformulate the EKS-STAR and EKS-S to fall back to one leg only: if one of the binary indices [Laspeyres or Paasche] for countries A and B does not exist – then use another one. That would somewhat defeat the purpose of accounting for representativity, as such index would produce biased results, as those two countries will be linked through one country's non-representative items only.

[10] It is fair to say that with an increase in matrix size, the failure rate of the EKS-style indices will diminish. However, it will still be below that for the CPD indices.

price levels], which is something we would like to avoid. Asymmetric inputs are dealt with in Annexes III and IV. Extended EKSs are used in Annexes II, IV and V.

Extension of EKS indices

As we said earlier, the EKS indices can be amended to be defined in a greater domain than the original EKS indices. Thus, the EKS [unweighted] can be defined with only one binary index in existence. So, if a Laspeyres is undefined, then the Paasche can be used as a sole component in index computation. A greater problem presents itself when neither Paasche nor Laspeyres exist. However, in this case it is possible to follow the logic of spanning trees, dropping the dead links and routing computational path to the existing ones. Thus amended, the EKS will have the range comparable to that of the CPD. EKS-S and EKS-STAR can be amended in a similar manner: if one of the binary links does not exist, then use the existing one(s), and dead links can be dropped.

Such an extension greatly enhances success rates of the EKS indices, making them comparable in that respect to the CPDs. However, such an extension results in introduction of significant new biases in the computation, with performance of all EKS-style indices decreasing across the board.

Extension of EKS-style indices

EKS-STAR is computed on the basis of two bilateral sub-indices, [geometric Lapeyres and Paasche], using only representative items. Failure rate of such an index attains 80% under symmetric probability assumptions on price and representativity matrices with (50, 50) parameters. If we allow for asymmetric inputs [some countries have relatively worse data than others] failure rate rises even more.

For example, for these matrices

$$\boldsymbol{P} = \begin{bmatrix} 1 & & & 5 \\ & 3 & 4 & 4 \\ & 1 & 6 & 3 \\ 1 & 2 & & 7 \\ & 4 & 2 & 1 \end{bmatrix}$$

$$\boldsymbol{W} = \begin{bmatrix} 0 & & & 0 \\ & 0 & 0 & 1 \\ & 0 & 1 & 1 \\ 1 & 1 & & 0 \\ & 0 & 1 & 1 \end{bmatrix}$$

price matrix is 70% full, with 50% of items being representative. However, bilateral indices for pairs (1, 3), (2, 3), (3, 1), (4,1) and, as a result, EKS-STAR index fail.

We can redefine the EKS-STAR index to use only *one* lateral link in such cases, even though it will introduce a systematic bias in the estimate[11] [first modification]. That will ease some of the failure rate concerns. However, when both binary indices fail, as in pairs (1, 3) and (3, 1), the index will still be undefined. In this case, a second modification needs to be introduced: if one Fischer fails, the computation will use valid links only.

EKS-S is extended in a similar fashion. EKS is more robust to start with and its modification requires only the extension with valid links [second modification] as it does not use weights.

The resulting Fischer matrices for EKS-style indices with these modifications are provided below[12]:

EKS-STAR

$$\begin{bmatrix} 1.000 & 2.000 & & 7.000 \\ 0.500 & 1.000 & 1.732 & 1.871 \\ & 0.577 & 1.000 & 0.561 \\ 0.143 & 0.535 & 1.782 & 1.000 \end{bmatrix}$$

EKS

$$\begin{bmatrix} 1.000 & 2.000 & & 5.916 \\ 0.500 & 1.000 & 1.587 & 1.368 \\ & 0.630 & 1.000 & 0.630 \\ 0.169 & 0.731 & 1.587 & 1.000 \end{bmatrix}$$

EKS-S

$$\begin{bmatrix} 1.000 & 2.000 & & 7.000 \\ 0.500 & 1.000 & 1.732 & 1.871 \\ & 0.577 & 1.000 & 0.500 \\ 0.143 & 0.535 & 2.000 & 1.000 \end{bmatrix}$$

The results of using the extended EKS-STAR, EKS and EKS-S are:

EKS-STAR	1.000	2.136	3.509	3.968
EKS	1.000	2.232	3.101	3.424

[11] Countries with lower representativity will have an elevated price level.

[12] For EKS-S, a weighted geometric mean of three sub-indices.

EKS-S	1.000	2.136	3.647	3.855

As a result of these modifications, EKS-style indices become defined in the range comparable to that of CPD.

Simulation with asymmetric inputs

In real world it is unrealistic to expect similarly filled price and weight matrices across the board: some countries do better than others. In this paper we simulated the case when one country has consistently lower representativity of its items relative to other countries [by 50%]. The results are presented in Annexes III and IV.

One can note that in this case the performance numbers of CPD-style indices have deteriorated as well, though not significantly. This is explained by the fact that the cases which could not be processed by the original EKSs at all would lead to lower precision results for CPDs.

Another interesting result is that EKS-S started to make more sense around 90% price matrix density, where it bested both EKS-STAR and EKS. However, CPD-WEIGHTED was still the highest performing index, and in the areas below 70-80% in the representativity and price matrix density terms, EKS-S was surpassed by even unweighted CPD, EKS and EKS-STAR.

More skewed data

In this run we estimate effect of lowering the number of products for one country by 25%, and simultaneously decreasing representativity by 25% as well with respect to other [*n*-1] countries. In Annex VI we show how this asymmetry affects that country's position relative to the other [*n*-1] countries.

The results of the simulation show that the only unbiased estimator in this situation is CPD-WEIGHTED, all other indices being affected to various degrees. CPD and EKS behave almost identically. EKS-S is definitely preferable to all other indices except for CPD-WEIGHTED. When matrices ***P*** and ***W*** for the [*n*-1] countries are relatively full, EKS-S becomes an unbiased estimator as well.

The *separability* properties state that some of the data effects on one country can be neutral for that country[13] and simultaneously not neutral to other countries. This property illustrates the fact that the effects of

Simulation with random price penalty for representativity

Finally, in Annex V we discuss the case when W is generated by U[0, 1] process. The results are quite similar to those from Annex II, thus we can make an important conclusion that the data model used is quite robust with respect to the choice of the price differential between representative and non-representative products.

CONCLUSIONS

As a result of simulations, the CPD-WEIGHTED was found to be superior to the other CPD and EKS-style indices. In addition, the EKSs without modifications exhibited a very high failure rate around the critical {50, 50} point in *price matrix density* and *matrix of representativity* space, where many countries outside the OECD area are expected to fall. By modifying the EKS indices, the their failure rate can be lowered, but it resulted in a loss of precision.

[13] We define the separability of data effects in one country in terms of the ratio of the result for that country relative to the geomean of results for other [n-1] countries. At the same time, those [n-1] countries can be affected individually. For example, consider the following price matrix ***P***, with matrix ***W*** being a full matrix with ***w***=1:

$$\begin{bmatrix} 1 & 2 & 5 & 5 \\ 1 & 3 & 4 & 4 \\ 1 & 1 & 6 & 8 \\ 1 & 2 & 7 & 7 \\ & 4 & 2 & 1 \end{bmatrix}$$

The results of index computation are the following:

	countries				
	1	*2*	*3*	*4*	*Geomean for countries [2-4]*
CPD-W	1.000	2.475	5.040	4.647	3.870
CPD	1.000	2.475	5.040	4.647	3.870
EKS	1.000	2.305	5.124	4.909	3.870
EKS-STAR	1.000	2.305	5.124	4.909	3.870
EKS-S	1.000	2.305	5.124	4.909	3.870

As we can see, the position of the first country does not change relative to the geomean of other countries [1 to 3.870] no matter which index we use. However, results for countries [2-4] do vary depending on the index used.

LITERATURE [to be added]

ANNEX I. RESULTS OF THE SIMULATION W/ SYMMETRIC REPRESENTATIVITY AND ORIGINAL EKS INDICES

Figure 1.1. Precision of CPD-WEIGHTED index in price matrix density and representativity space

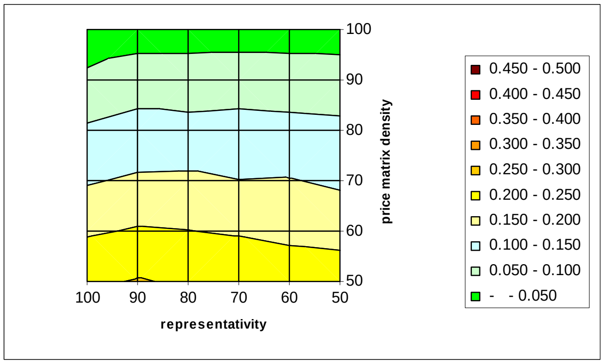

Figure 1.2. Precision of CPD index in price matrix density and representativity space

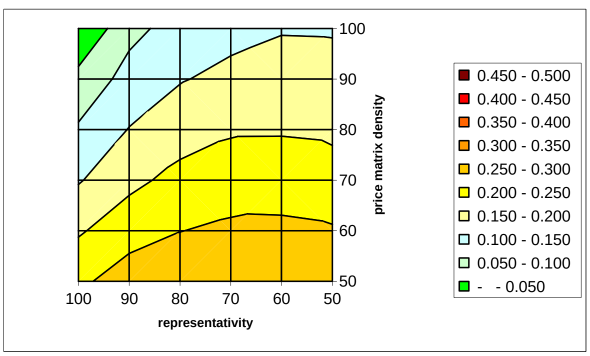

Figure 1.3. Precision of EKS-STAR index in price matrix density and representativity space

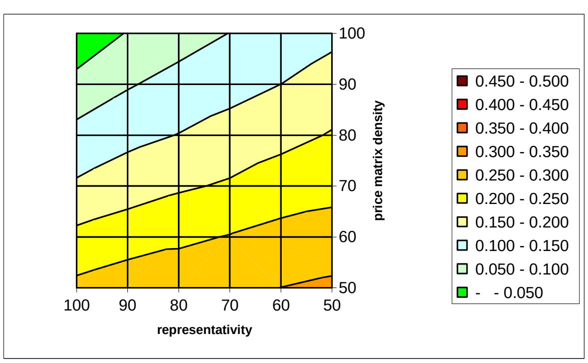

Figure 1.4. Precision of EKS-S index in price matrix density and representativity space

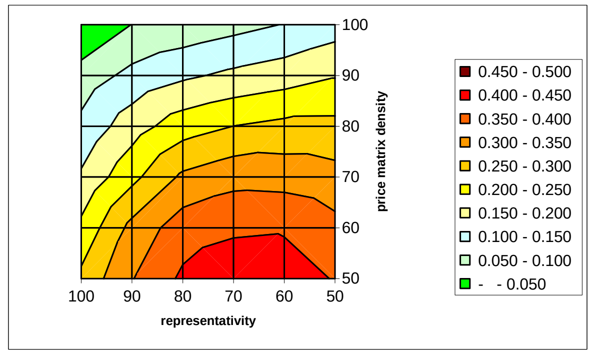

Figure 1.5. Precision of EKS index in price matrix density and representativity space

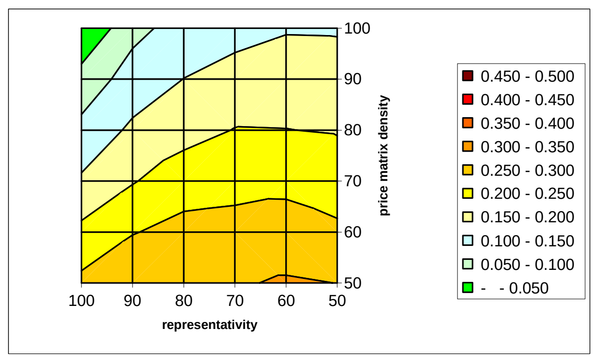

Figure 1.6. Success rates of EKS-STAR and EKS-S procedures

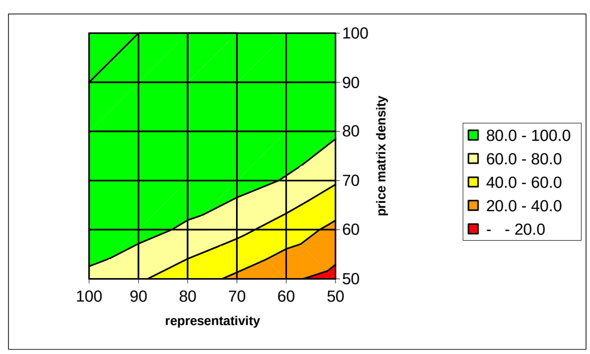

Table 1.1. Precision of indices in price matrix density (vertical) and representativity (horizontal) space

CPD-WEIGHTED	50	60	70	80	90	100
50	**0.228**	**0.231**	**0.237**	**0.243**	**0.254**	**0.241**
60	0.183	0.187	0.196	**0.201**	**0.204**	0.194
70	0.142	0.152	0.151	0.158	0.157	0.146
80	0.110	0.113	0.116	0.114	0.117	0.106
90	0.075	0.076	0.079	0.076	0.076	0.066
100	0.026	0.027	0.027	0.026	0.027	0.000

CPD	50	60	70	80	90	100
50	**0.293**	**0.296**	**0.282**	**0.284**	**0.273**	**0.241**
60	**0.254**	**0.258**	**0.259**	**0.249**	**0.231**	0.194
70	**0.218**	**0.231**	**0.224**	**0.216**	0.186	0.146
80	0.192	0.195	0.195	0.177	0.152	0.106
90	0.169	0.173	0.162	0.147	0.117	0.066
100	0.146	0.146	0.136	0.118	0.087	0.000

EKS-STAR	50	60	70	80	90	100
50	0.309	0.300	**0.285**	**0.278**	**0.276**	**0.262**
60	**0.271**	**0.265**	**0.252**	**0.242**	**0.228**	**0.212**
70	**0.235**	**0.225**	**0.206**	0.194	0.176	0.157
80	**0.203**	0.185	0.170	0.151	0.137	0.113
90	0.172	0.150	0.132	0.117	0.096	0.071
100	0.137	0.118	0.101	0.079	0.054	0.000

EKS-S	50	60	70	80	90	100
50	0.397	0.422	0.436	0.409	0.347	**0.262**
60	0.365	0.395	0.391	0.377	0.315	**0.212**
70	0.318	0.330	0.334	0.309	**0.236**	0.157
80	**0.263**	**0.263**	**0.250**	**0.227**	0.177	0.113
90	0.196	0.176	0.161	0.141	0.114	0.071
100	0.126	0.102	0.083	0.066	0.051	0.000

EKS	50	60	70	80	90	100
50	**0.299**	0.306	**0.295**	**0.298**	**0.290**	**0.262**
60	**0.260**	**0.269**	**0.270**	**0.266**	**0.248**	**0.212**
70	**0.224**	**0.239**	**0.232**	**0.227**	0.197	0.157
80	0.197	**0.201**	**0.202**	0.182	0.159	0.113
90	0.172	0.175	0.165	0.151	0.120	0.071
100	0.146	0.146	0.136	0.118	0.087	0.000

Notes:

1. All countries have equal probability of a product to be representative.

ANNEX II. RESULTS OF THE SIMULATION W/ SYMMETRIC REPRESENTATIVITY AND EXTENDED EKS INDICES

Figure 2.1. Precision of CPD-WEIGHTED index in price matrix density and representativity space

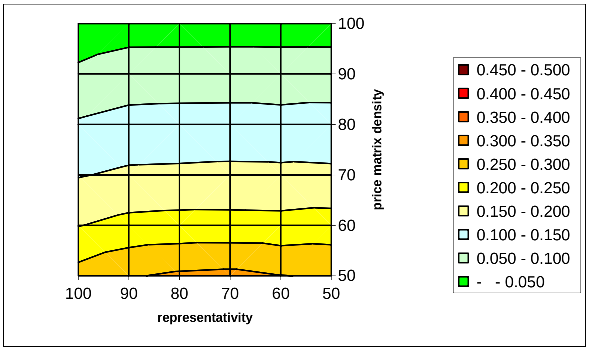

Figure 2.2. Precision of CPD index in price matrix density and representativity space

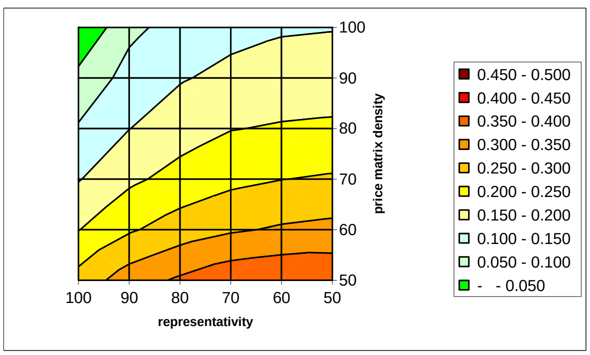

Figure 2.3. Precision of EKS-STAR+ index in price matrix density and representativity space

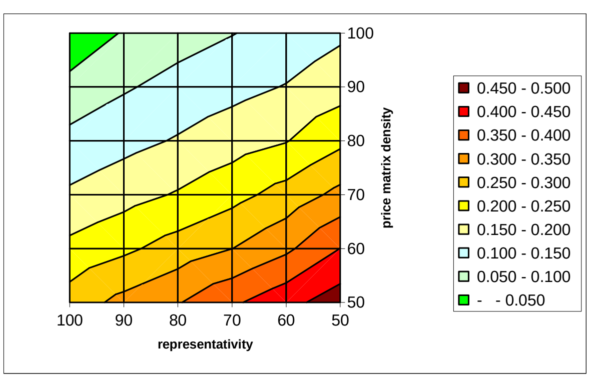

Figure 2.4. Precision of EKS-S+ index in price matrix density and representativity space

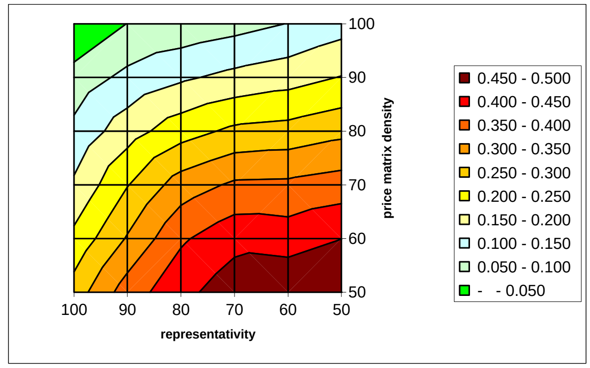

Figure 2.5. Precision of EKS+ index in price matrix density and representativity space

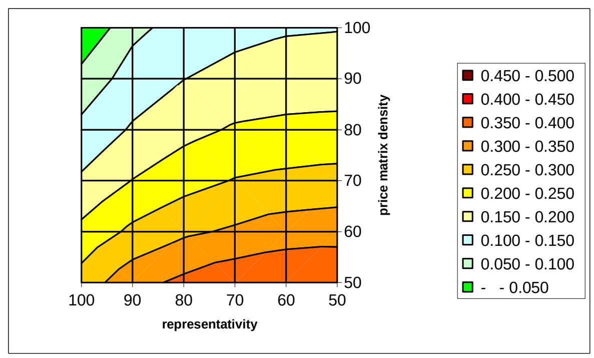

Table 2.1. Precision of indices in price matrix density (vertical) and representativity (horizontal) space

CPD-WEIGHTED	50	60	70	80	90	100
50	**0.297**	0.301	0.313	0.308	**0.296**	**0.269**
60	**0.221**	**0.216**	**0.217**	**0.217**	**0.214**	0.198
70	0.159	0.161	0.162	0.160	0.158	0.147
80	0.118	0.115	0.118	0.117	0.115	0.105
90	0.076	0.077	0.076	0.076	0.076	0.065
100	0.027	0.027	0.027	0.027	0.027	0.000

CPD	50	60	70	80	90	100
50	0.392	0.394	0.385	0.357	0.326	**0.269**
60	0.313	0.306	**0.294**	**0.274**	**0.244**	0.198
70	**0.255**	**0.249**	**0.238**	**0.217**	0.190	0.147
80	**0.208**	**0.205**	0.198	0.178	0.149	0.105
90	0.172	0.170	0.162	0.146	0.117	0.065
100	0.148	0.146	0.136	0.117	0.089	0.000

EKS-STAR+	50	60	70	80	90	100
50	0.476	0.434	0.391	0.347	0.315	**0.273**
60	0.401	0.340	**0.299**	**0.272**	**0.240**	**0.213**
70	0.314	**0.269**	**0.233**	**0.204**	0.181	0.158
80	**0.238**	0.198	0.177	0.154	0.134	0.113
90	0.179	0.152	0.134	0.116	0.095	0.070
100	0.142	0.118	0.098	0.080	0.055	0.000

EKS-S+	50	60	70	80	90	100
50	0.504	0.494	0.480	0.434	0.375	**0.273**
60	0.449	0.426	0.434	0.393	0.305	**0.213**
70	0.374	0.361	0.358	0.324	**0.247**	0.158
80	**0.287**	**0.268**	**0.261**	**0.229**	0.178	0.113
90	**0.202**	0.179	0.164	0.142	0.113	0.070
100	0.129	0.101	0.081	0.065	0.051	0.000

EKS+	50	60	70	80	90	100
50	0.397	0.398	0.387	0.362	0.332	**0.273**
60	0.330	0.324	0.307	**0.291**	**0.261**	**0.213**
70	**0.268**	**0.262**	**0.253**	**0.231**	**0.201**	0.158
80	**0.215**	**0.212**	**0.205**	0.185	0.156	0.113
90	0.175	0.173	0.166	0.149	0.120	0.070
100	0.148	0.146	0.136	0.117	0.089	0.000

Notes:

1. All countries have equal probability of a product to be representative.
2. EKS-STAR, EKS-S are extended to be defined when either Paasche or Laspeyres fails. Additionally, for these indices and EKS, when an entire link fails, calculations are rerouted via valid links.

ANNEX III. RESULTS OF THE SIMULATION W/ SKEWED REPRESENTATIVITY AND ORIGINAL EKS INDICES

Figure 3.1. Precision of CPD-WEIGHTED index in price matrix density and representativity space

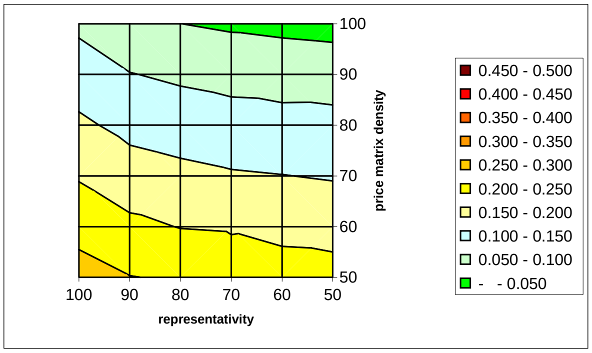

Figure 3.2. Precision of CPD index in price matrix density and representativity space

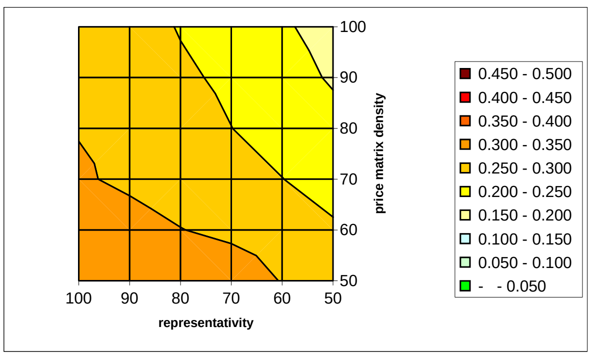

Figure 3.3. Precision of EKS-STAR index in price matrix density and representativity space

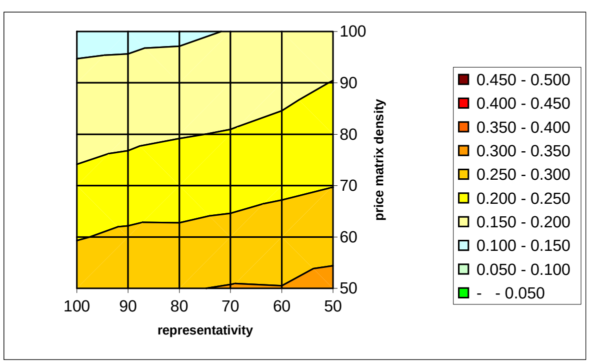

Figure 3.4. Precision of EKS-S index in price matrix density and representativity space

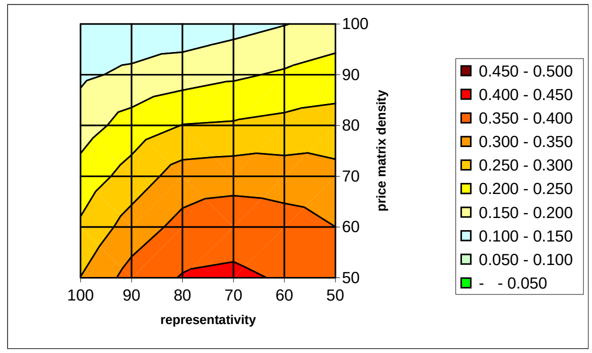

Figure 3.5. Precision of EKS index in price matrix density and representativity space

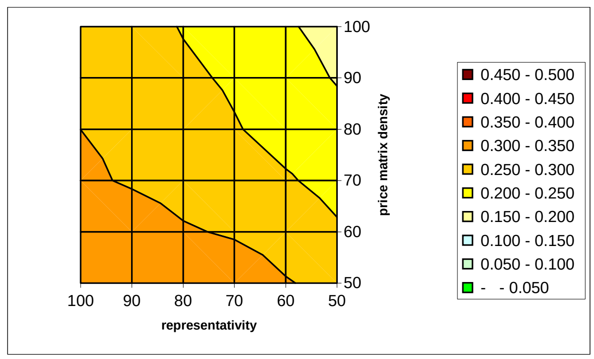

Figure 3.6. Success rates of EKS-STAR and EKS-S procedures

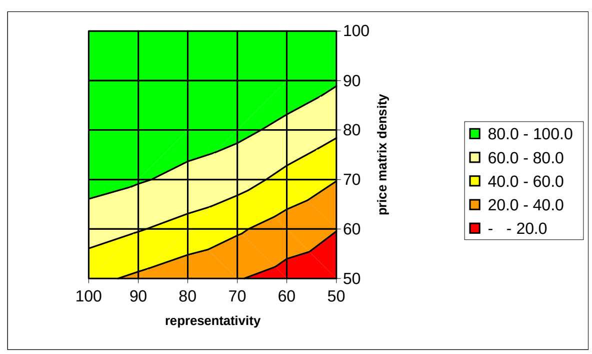

Table 3.1. Precision of indices in price matrix density (vertical) and representativity (horizontal) space

CPD-WEIGHTED	50	60	70	80	90	100
50	**0.222**	**0.230**	**0.225**	**0.245**	**0.251**	**0.270**
60	0.178	0.181	0.195	0.198	**0.211**	**0.234**
70	0.147	0.151	0.155	0.162	0.170	0.196
80	0.114	0.114	0.119	0.128	0.137	0.158
90	0.079	0.083	0.085	0.092	0.102	0.126
100	0.033	0.037	0.043	0.050	0.063	0.090

CPD	50	60	70	80	90	100
50	**0.291**	**0.298**	0.326	0.318	0.317	0.335
60	**0.256**	**0.273**	**0.290**	0.301	0.315	0.310
70	**0.231**	**0.251**	**0.265**	**0.279**	**0.292**	0.305
80	**0.215**	**0.230**	**0.250**	**0.268**	**0.290**	**0.298**
90	0.195	**0.219**	**0.242**	**0.257**	**0.276**	**0.296**
100	0.183	**0.206**	**0.229**	**0.247**	**0.270**	**0.291**

EKS-STAR	50	60	70	80	90	100
50	0.317	0.301	0.303	**0.297**	**0.286**	**0.291**
60	**0.278**	**0.272**	**0.266**	**0.259**	**0.259**	**0.247**
70	**0.249**	**0.241**	**0.231**	**0.227**	**0.217**	**0.213**
80	**0.228**	**0.212**	**0.202**	0.198	0.192	0.181
90	**0.201**	0.186	0.176	0.170	0.162	0.161
100	0.174	0.160	0.152	0.142	0.141	0.138

EKS-S	50	60	70	80	90	100
50	0.387	0.396	0.408	0.403	0.370	0.301
60	0.350	0.373	0.383	0.366	0.322	**0.258**
70	0.314	0.324	0.329	0.323	**0.271**	**0.219**
80	**0.271**	**0.265**	**0.256**	**0.251**	**0.221**	0.177
90	**0.222**	**0.207**	0.192	0.177	0.161	0.141
100	0.170	0.148	0.131	0.116	0.110	0.102

EKS	50	60	70	80	90	100
50	**0.286**	0.303	0.326	0.322	0.324	0.340
60	**0.256**	**0.279**	**0.295**	0.304	0.320	0.312
70	**0.235**	**0.255**	**0.268**	**0.284**	**0.296**	0.306
80	**0.218**	**0.233**	**0.253**	**0.271**	**0.291**	**0.300**
90	0.197	**0.221**	**0.243**	**0.259**	**0.277**	**0.298**
100	0.183	**0.206**	**0.229**	**0.247**	**0.270**	**0.291**

Notes:

1. One country has probability of a product to be representative 50% lower than the rest of the countries.

ANNEX IV. RESULTS OF THE SIMULATION W/ SKEWED REPRESENTATIVITY AND EXTENDED EKS INDICES

Figure 4.1. Precision of CPD-WEIGHTED index in price matrix density and representativity space

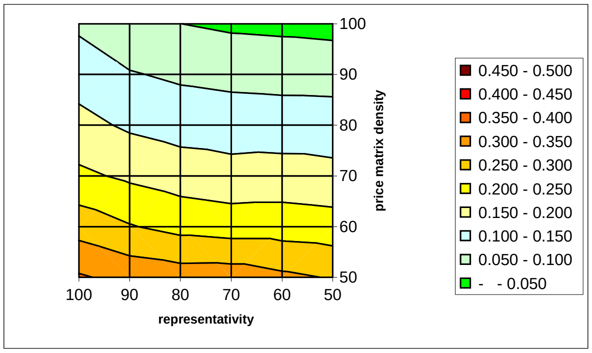

Figure 4.2. Precision of CPD index in price matrix density and representativity space

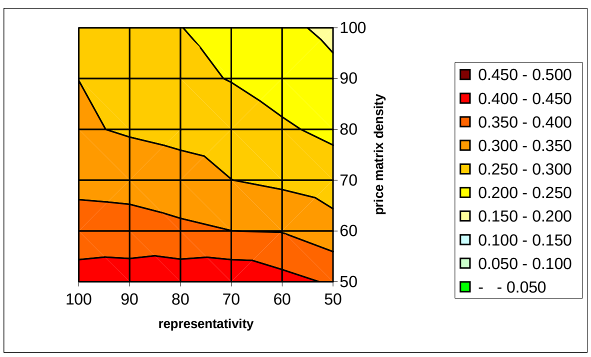

Figure 4.3. Precision of EKS-STAR+ index in price matrix density and representativity space

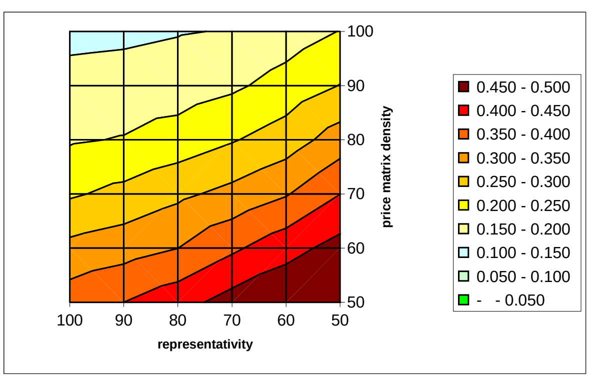

Figure 4.4. Precision of EKS-S+ index in price matrix density and representativity space

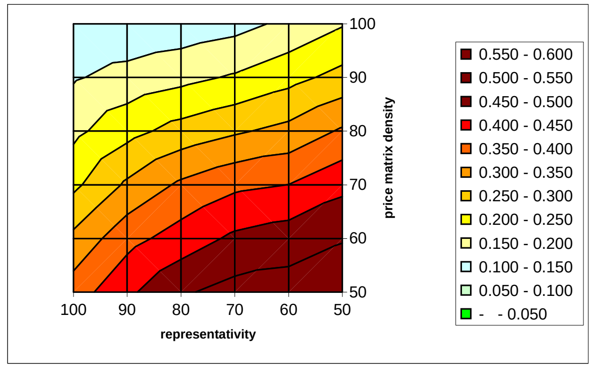

Figure 4.5. Precision of EKS+ index in price matrix density and representativity space

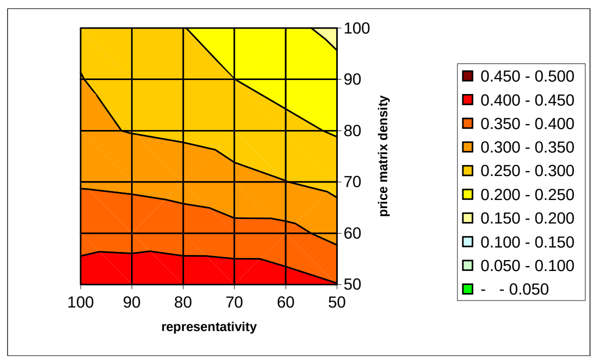

Table 4.1. Precision of indices in price matrix density (vertical) and representativity (horizontal) space

CPD-WEIGHTED	50	60	70	80	90	100
50	**0.297**	0.310	0.326	0.325	0.334	0.356
60	**0.221**	**0.226**	**0.227**	**0.235**	**0.253**	**0.279**
70	0.165	0.171	0.168	0.176	0.191	**0.210**
80	0.122	0.123	0.126	0.130	0.143	0.165
90	0.083	0.084	0.086	0.092	0.103	0.129
100	0.034	0.038	0.042	0.050	0.064	0.091

CPD	50	60	70	80	90	100
50	0.394	0.417	0.438	0.432	0.426	0.421
60	0.320	0.347	0.350	0.359	0.368	0.372
70	**0.274**	**0.289**	0.300	0.322	0.333	0.336
80	**0.239**	**0.256**	**0.275**	**0.285**	**0.294**	0.307
90	**0.209**	**0.230**	**0.248**	**0.262**	**0.282**	**0.300**
100	0.191	**0.209**	**0.226**	**0.251**	**0.267**	**0.292**

EKS-STAR+	50	60	70	80	90	100
50	0.522	0.494	0.471	0.431	0.400	0.376
60	0.468	0.431	0.391	0.350	0.329	0.314
70	0.399	0.346	0.314	**0.289**	**0.264**	**0.244**
80	0.324	**0.274**	**0.246**	**0.221**	**0.203**	0.195
90	**0.251**	**0.219**	0.191	0.176	0.169	0.165
100	**0.201**	0.175	0.153	0.147	0.141	0.138

EKS-S+	50	60	70	80	90	100
50	0.536	0.522	0.517	0.494	0.440	0.375
60	0.497	0.476	0.460	0.422	0.383	0.312
70	0.437	0.400	0.390	0.359	0.308	**0.239**
80	0.356	0.315	**0.293**	**0.269**	**0.233**	0.187
90	**0.266**	**0.234**	**0.206**	0.186	0.168	0.145
100	0.196	0.162	0.133	0.119	0.109	0.103

EKS+	50	60	70	80	90	100
50	0.401	0.419	0.435	0.435	0.429	0.421
60	0.334	0.365	0.366	0.372	0.381	0.384
70	**0.285**	0.301	0.312	0.334	0.340	0.345
80	**0.245**	**0.263**	**0.280**	**0.290**	**0.298**	0.310
90	**0.212**	**0.233**	**0.250**	**0.263**	**0.284**	0.301
100	0.191	**0.209**	**0.226**	**0.251**	**0.267**	**0.292**

Notes:

1. One country has probability of a product to be representative 50% lower than the rest of the countries.
2. EKS-STAR, EKS-S are extended to be defined when either Paasche or Laspeyres fails. Additionally, for these indices and EKS, when an entire link fails, calculations are rerouted via valid links.

ANNEX V. RESULTS OF THE SIMULATION W/ SYMMETRIC REPRESENTATIVITY AND EXTENDED EKS INDICES, VARIABLE Ω

Figure 5.1. Precision of CPD-WEIGHTED index in price matrix density and representativity space

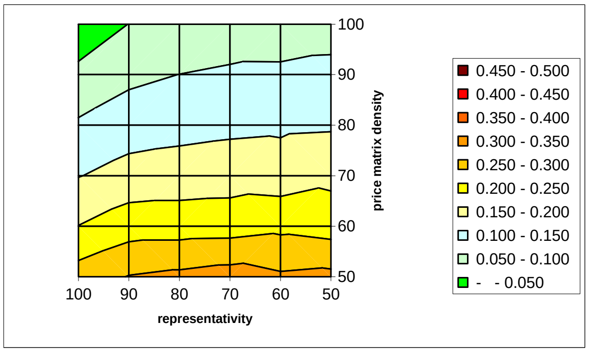

Figure 5.2. Precision of CPD index in price matrix density and representativity space

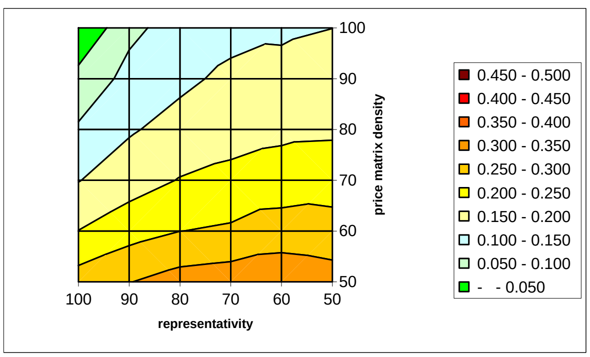

Figure 5.3. Precision of EKS-STAR+ index in price matrix density and representativity space

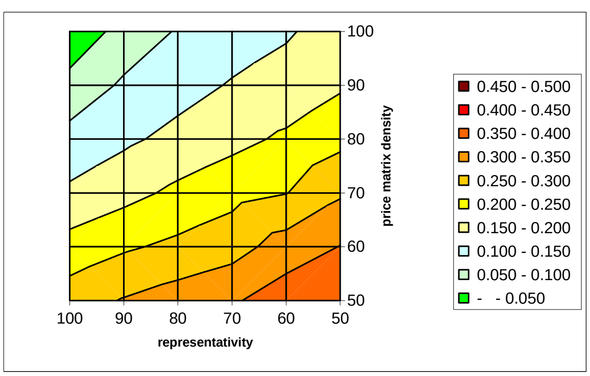

Figure 5.4. Precision of EKS-S+ index in price matrix density and representativity space

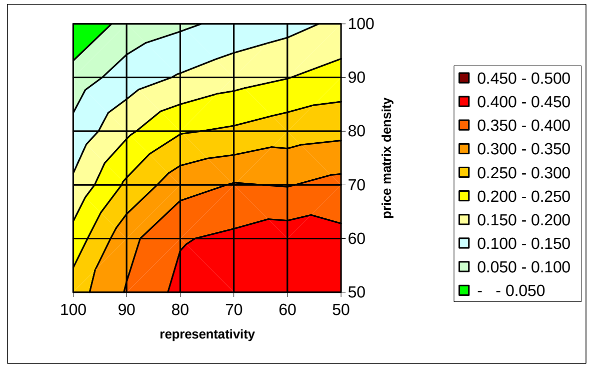

Figure 5.5. Precision of EKS+ index in price matrix density and representativity space

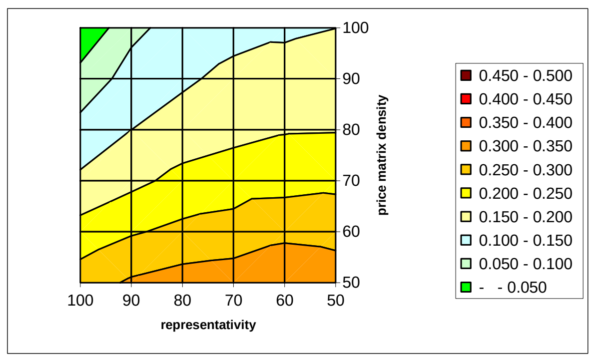

Table 5.1. Precision of indices in price matrix density (vertical) and representativity (horizontal) space

CPD-WEIGHTED	50	60	70	80	90	100
50	0.313	0.307	0.322	0.312	0.302	**0.273**
60	**0.228**	**0.238**	**0.228**	**0.227**	**0.227**	**0.201**
70	0.188	0.174	0.178	0.174	0.169	0.148
80	0.144	0.142	0.139	0.133	0.125	0.106
90	0.112	0.107	0.107	0.100	0.089	0.068
100	0.081	0.080	0.072	0.064	0.051	0.000

CPD	50	60	70	80	90	100
50	0.323	0.333	0.328	0.321	**0.298**	**0.273**
60	**0.270**	**0.276**	**0.257**	**0.249**	**0.230**	**0.201**
70	**0.228**	**0.220**	**0.215**	**0.202**	0.178	0.148
80	0.193	0.191	0.177	0.168	0.145	0.106
90	0.167	0.159	0.161	0.139	0.114	0.068
100	0.150	0.145	0.134	0.118	0.089	0.000

EKS-STAR+	50	60	70	80	90	100
50	0.382	0.377	0.344	0.324	0.304	**0.276**
60	0.351	0.323	**0.279**	**0.261**	**0.243**	**0.219**
70	**0.294**	**0.248**	**0.234**	**0.211**	0.184	0.160
80	**0.236**	**0.208**	0.185	0.164	0.141	0.114
90	0.194	0.168	0.154	0.132	0.106	0.073
100	0.171	0.145	0.127	0.103	0.075	0.000

EKS-S+	50	60	70	80	90	100
50	0.425	0.446	0.432	0.414	0.354	**0.276**
60	0.413	0.426	0.410	0.396	0.334	**0.219**
70	0.366	0.347	0.353	0.330	**0.259**	0.160
80	**0.286**	**0.278**	**0.258**	**0.245**	0.190	0.114
90	**0.220**	0.198	0.180	0.155	0.123	0.073
100	0.162	0.133	0.114	0.091	0.069	0.000

EKS+	50	60	70	80	90	100
50	0.327	0.338	0.329	0.322	0.307	**0.276**
60	**0.284**	**0.289**	**0.268**	**0.262**	**0.245**	**0.219**
70	**0.238**	**0.231**	**0.228**	**0.215**	0.187	0.160
80	0.198	0.197	0.185	0.172	0.150	0.114
90	0.168	0.162	0.163	0.142	0.117	0.073
100	0.150	0.145	0.134	0.118	0.089	0.000

Notes:

1. All countries have equal probability of a product to be representative.
2. EKS-STAR, EKS-S are extended to be defined when either Paasche or Laspeyres fails. Additionally, for these indices and EKS, when an entire link fails, calculations are rerouted via valid links.
3. Adjustment for non-representativity is uniformly distributed on [0, 1].

ANNEX VI. EFFECT OF ONE COUNTRY'S POOR QUALITY DATA [RESULTS W/ EXTENDED EKS INDICES]

Figure 6.1. Bias for the country with poor quality data [CPD-WEIGHTED index in *price matrix density* and *representativity* space]

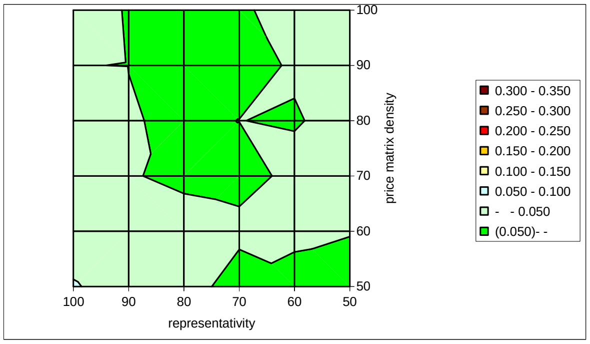

Figure 6.2. Bias for the country with poor quality data [CPD index in *price matrix density* and *representativity* space]

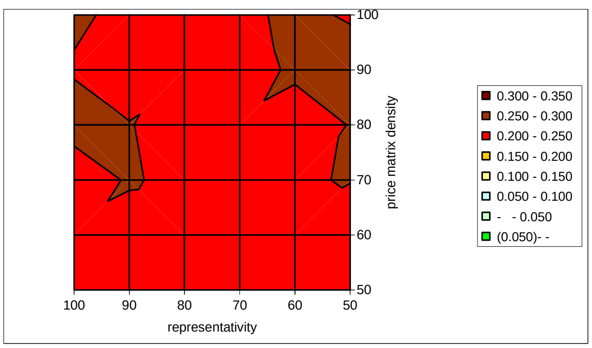

Figure 6.3. Bias for the country with poor quality data [EKS-STAR+ index in *price matrix density* and *representativity* space]

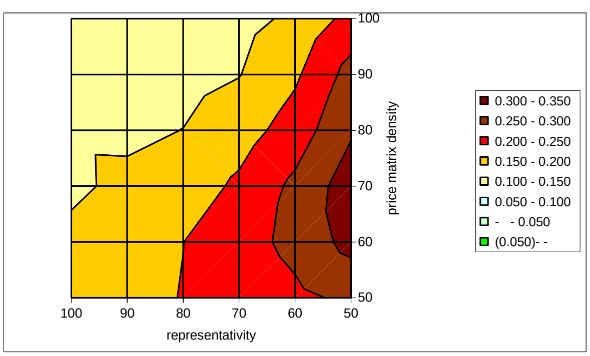

Figure 6.4. Bias for the country with poor quality data [EKS-S+ index in *price matrix density* and *representativity* space]

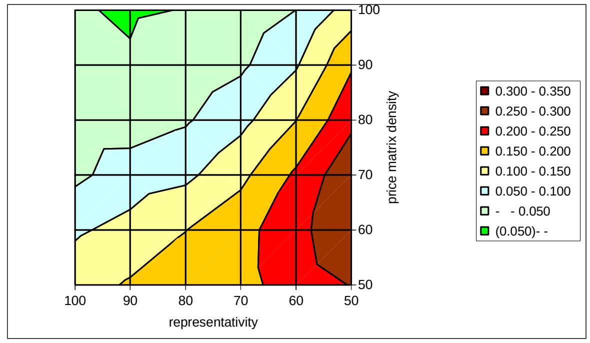

Figure 6.5. Bias for the country with poor quality data [EKS+ index in *price matrix density* and *representativity* space]

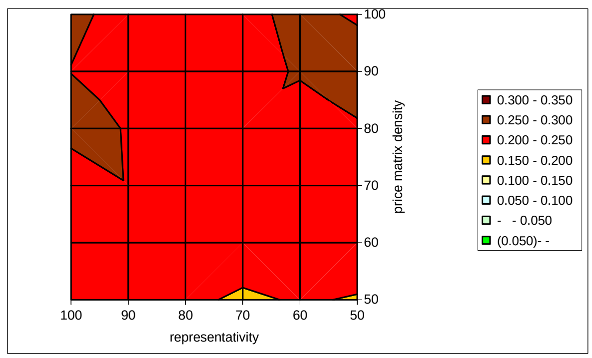

Table 6.1. Bias for the country with poor quality data
[various indices in *price matrix density* (vertical) and *representativity* (horizontal) space]

CPD-WEIGHTED	50	60	70	80	90	100
50	(0.004)	(0.003)	(0.008)	0.008	0.028	0.054
60	0.000	0.002	0.004	0.007	0.007	0.024
70	0.001	0.003	(0.005)	(0.003)	0.001	0.002
80	0.004	(0.001)	0.000	(0.002)	0.001	0.007
90	0.001	0.001	(0.004)	(0.004)	(0.000)	0.000
100	0.001	0.003	(0.001)	(0.001)	(0.000)	0.002

CPD	50	60	70	80	90	100
50	**0.217**	**0.224**	**0.212**	**0.234**	**0.237**	**0.231**
60	**0.241**	**0.247**	**0.241**	**0.243**	**0.244**	**0.248**
70	**0.251**	**0.249**	**0.250**	**0.246**	**0.251**	**0.242**
80	**0.250**	**0.246**	**0.249**	**0.245**	**0.250**	**0.255**
90	**0.255**	**0.252**	**0.246**	**0.248**	**0.244**	**0.249**
100	**0.249**	**0.252**	**0.248**	**0.246**	**0.247**	**0.252**

EKS-STAR+	50	60	70	80	90	100
50	**0.261**	**0.238**	**0.204**	**0.202**	0.186	0.166
60	0.315	**0.267**	**0.225**	0.200	0.177	0.165
70	0.327	**0.261**	**0.209**	0.173	0.164	0.139
80	**0.294**	**0.223**	0.177	0.151	0.138	0.139
90	**0.269**	0.192	0.148	0.133	0.123	0.126
100	**0.216**	0.160	0.133	0.124	0.123	0.127

EKS-S+	50	60	70	80	90	100
50	**0.252**	**0.222**	0.185	0.176	0.156	0.126
60	**0.296**	**0.233**	0.183	0.149	0.115	0.093
70	**0.289**	**0.208**	0.137	0.089	0.075	0.038
80	**0.237**	0.149	0.085	0.044	0.023	0.023
90	0.194	0.095	0.041	0.016	0.003	0.003
100	0.123	0.050	0.018	0.001	(0.003)	0.002

EKS+	50	60	70	80	90	100
50	0.197	**0.205**	0.191	**0.212**	**0.211**	**0.205**
60	**0.232**	**0.238**	**0.232**	**0.233**	**0.235**	**0.237**
70	**0.249**	**0.245**	**0.247**	**0.243**	**0.249**	**0.239**
80	**0.249**	**0.244**	**0.248**	**0.245**	**0.249**	**0.256**
90	**0.254**	**0.251**	**0.246**	**0.248**	**0.244**	**0.250**
100	**0.249**	**0.252**	**0.248**	**0.246**	**0.247**	**0.252**

Notes:

1. One country has 25% less of price matrix density and 25 % less of representative products.
2. EKS-STAR, EKS-S are extended to be defined when either Paasche or Laspeyres fails. Additionally, for these indices and EKS, when an entire link fails, calculations are rerouted via valid links.