Online Appendix to

Resource Misallocation in Cross-country Differences in Manufacturing Productivity

Robert Inklaar, Addisu A. Lashitew and Marcel P. Timmer

Estimating industry output prices

The challenge to accurately estimating manufacturing output prices can best be illustrated in a supply and use framework. Suppose that there are $i = 1, \dots, N$ manufactured goods that can be used for final consumption and investment or as intermediate inputs. Furthermore, there is a set of countries $j = 1, \dots, C$. Our aim is to compare the price of manufacturing output in country j relative to another country k. If we would have data on the output value and prices of individual products i, an estimate of the relative price of manufacturing output would only require aggregating over the relative prices of the individual products, denoted by p_{ij}^{y}. The Törnqvist index is such an aggregator function and a flexible one (Diewert, 1976; Caves et al. 1982b):[1]

$$\log(p_j^y) - \log(p_k^y) = \sum_{i=1}^{N} \frac{1}{2}(s_{ij}^y + s_{ik}^y)(\log(p_{ij}^y) - \log(p_{ik}^y)), \tag{A1}$$

where $s_{ij}^y = \frac{p_{ij}^y y_{ij}}{\sum_i p_{ij}^y y_{ij}}$. (A1) states that the log relative price of manufacturing output is equal to the weighted-average relative output price of individual products, where the weight is the share of each product in overall output, averaged across the two countries under comparison. When the comparison is across more than two countries, the final

[1] A flexible aggregate function is a second-order approximation to an arbitrary twice-differentiable linearly homogenous function.

index would depend on the choice of base country k. To avoid this, Caves, Christensen and Diewert (1982a) proposed comparing each country not to an actual country but to a (synthetic) average country:

$$\log\left(p_j^y\right) - \overline{\log(p^y)} = \sum_{i=1}^{N} \frac{1}{2}\left(s_{ij}^y + \bar{s}_i^y\right)\left(\log\left(p_{ij}^y\right) - \overline{\log\left(p_i^y\right)}\right). \tag{A2}$$

where the upper bar indicates an arithmetic average across countries. This approach is typically referred as the GEKS method and is used by the OECD, Eurostat and World Bank in their relative price computations.[2]

The problem in implementing (A2) is that we do not have reliable data on relative industry output prices for a large sample of countries, and especially not for developing economies. To see how results based solely on the commonly-used expenditure prices are related to the relative output prices, consider the equality between the value of supply and use:[3]

$$p_{ij}^q q_{ij} + p_{ij}^x x_{ij} + p_{ij}^z z_{ij} = p_{ij}^m m_{ij} + p_{ij}^y y_{ij}, \forall i. \tag{A3}$$

Here q denotes domestic final demand, x is exports, z is intermediate demand, y is output and m is imports. As expressed in (A3), the value of the supply of each product (shown on the right-hand side) should equal to the value of its demand (on the left-hand side). Next consider prices $p_{ij}^q, p_{ij}^x, p_{ij}^z, p_{ij}^y$ and p_{ij}^m for goods i in each country j. In this general setting, we allow the price to differ according to each source of supply or use destination. Next we sum across all products and rearrange:

$$\sum_{i=1}^{N} p_{ij}^y y_{ij} = \sum_{i=1}^{N}\left(p_{ij}^q q_{ij} + p_{ij}^x x_{ij} + p_{ij}^z z_{ij} - p_{ij}^m m_{ij}\right). \tag{A4}$$

---

[2] The only difference is that those organizations would use a Fisher, rather than a Törnqvist index.
[3] Ignoring net taxes on products, which should be added to the right-hand side.

The left-hand side shows the total value of manufacturing output, which consists of the sum of domestic final demand, domestic intermediate demand and exports, and subtracts imports. (A4) implies that the relative price of manufacturing output can be either measured directly, as in (A2), or indirectly using prices of domestic final and intermediate demand and export and import prices. The indirect alternative to the direct approach of (A2) can be expressed as:

$$\log(p_j^y) - \overline{\log(p^y)} = \sum_{i=1}^{N} \frac{1}{2}(s_{ij}^y + \bar{s}_i^y)\left[\widehat{v_{ij}^q}\left(\log(p_{ij}^q) - \overline{\log(p_i^q)}\right) + \widehat{v_{ij}^z}\left(\log(p_{ij}^z) - \overline{\log(p_i^z)}\right) + \widehat{v_{ij}^x}\left(\log(p_{ij}^x) - \overline{\log(p_i^x)}\right) - \widehat{v_{ij}^m}\left(\log(p_{ij}^m) - \overline{\log(p_i^m)}\right)\right], \quad \text{(A5)}$$

where $\widehat{v_{ij}^q} = \frac{1}{2}\left(\frac{p_{ij}^q q_{ij}}{p_{ij}^y y_{ij}} + \overline{\frac{p_i^q q_i}{p_i^y y_i}}\right)$, the share of domestic final expenditure in the value of output of each product, averaged between country *j* and the arithmetic mean of shares across all countries, analogous to the definition of $\overline{s_i^y}$. The other $\hat{v}$'s are defined analogously.

(A5) allows us to relate the standard approach, which relies solely on relative prices of domestic final expenditure q to this more comprehensive approach. The standard approach is only valid if either all relative prices are equal to each other or if the share of domestic final expenditure in total output is equal to one. In the more-common case where the share is less than one, another potential bias is when the share of a product in domestic final expenditure, $p_{ij}^q q_{ij}/\sum_i p_{ij}^q q_{ij}$, is used rather than the share in output, s_i^y.

For a broad group of countries, we have implemented a modified version of (A5). Specifically, we use data on relative prices of domestic final expenditure from the 2005 ICP round, which covers 146 countries. We supplement that with data on relative prices of exports and imports from Feenstra and Romalis (2014). Data on prices of domestic intermediate demand are not separately available and we deal with this in two ways. First, part of domestic intermediate demand is supplied from foreign sources, i.e. imports. As imports of intermediate inputs have no (direct) bearing on output prices of domestic producers, imported intermediates can be excluded from the set of intermediate products and imports in (A5). Second, we assume that the price that producers charge to domestic final users is equal to the price charged to domestic intermediate users, so $p^q = p^z$. Also, we assume that relative *purchaser* prices of domestic final expenditure are equal to relative *producer* prices. In other words, we assume that the trade and transportation margins are equal across countries. Relaxing this stringent assumption would require detailed input-output tables, which are missing for many of the countries we analyze (see also the discussion below). These assumptions lead to the following modified version of (A5):

$$\log(p_j^y) - \overline{\log(p^y)} = \sum_{i=1}^{N} \frac{1}{2}\left(s_{ij}^y + \bar{s}_i^y\right)\Big[\left(\widehat{v_{ij}^q} + \widetilde{v_{ij}^z}\right)\left(\log(p_{ij}^q) - \overline{\log(p_i^q)}\right) + \widehat{v_{ij}^x}\left(\log(p_{ij}^x) - \overline{\log(p_i^x)}\right) - \widetilde{v_{ij}^m}\left(\log(p_{ij}^m) - \overline{\log(p_i^m)}\right)\Big], \tag{A6}$$

where $\widetilde{v_{ij}^z}$ denotes the share of *domestic* intermediate demand in output and $\widetilde{v_{ij}^m}$ the imports of products for final demand. By implementing (A6), we resolve an issue in the literature that has long been know, but never resolved in a satisfactory manner (see e.g. Hooper, 1996). As discussed in Herrendorf and Valentinyi (2012, 330), the price of final demand, p^q reflects the price of domestically produced goods and of imports, which are produced using 'world market' technology. The fact that certain products are imported rather than domestically produced suggests that domestic technology is at least no better than world market technology. This effect would imply that the variation in p^q will be lower than the variation in p^y. At the same time, following the logic of the Melitz (2003) model, only the most-productive firms in an economy will export and their prices would have to be competitive in world markets. Which of these effects dominates is hard to say ex ante, so it requires implementing (A6).

This implementation requires not only data on relative prices, but also on output, domestic (final and intermediate) demand, exports and imports (of final products) by manufacturing product. For most advanced economies and a growing number of emerging economies, such information is available from input-output tables. However, such data is not available for many of the countries we analyze here. We therefore constructed a dataset by combining industry output data from UNIDO, export and import data from Comtrade and domestic final expenditure data from ICP. This requires detailed matching across different product classifications, dealing with missing data and reconciling conflicting data, all of which is discussed in more detail below.

Input-output and employment data construction

Implementing (A6) requires data for the left-hand side and right-hand side of (A6). This means we need data on the value of (gross) output for individual products and total manufacturing and information on domestic demand, export demand and imports of final products for the same products. With sufficiently detailed input-output tables for each country analyze, this would be fairly straightforward. However, those are not available for the large majority of countries so we combine and reconcile the data sources that are available. For output data we use the UNIDO INDSTAT databases, for domestic final demand we use the ICP basic heading expenditure data and for exports and imports we use the UN Comtrade database. Note that we have no independent information on domestic intermediate demand, so we compute it as a residual.

Product/industry classification and correspondences

The product/industry classification that we use distinguishes 14 manufacturing industries that together comprise all of manufacturing and is based on the ISIC revision 3 classification system, see Table A1. This is also the classification used in the World Input-Output Database (WIOD) and represents a compromise between a detailed view of manufacturing and limits to data availability.

Table A1, Product/industry classification

Product/Industry	ISIC rev. 3 code
Food, beverages and tobacco	15-16
Textiles, textile products and wearing apparel	17-18
Leather and leather products	19
Wood, paper, printing and publishing	20-22
Petroleum and coal products	23
Chemicals, rubber and plastics	24-25
Non-metallic mineral products	26
Basic and fabricated metal products	27-28
Machinery	29
Electronic and optical equipment	30-33
Transport equipment	34-35

Miscellaneous manufacturing	36-37

Much of the gross output data is already available in the ISIC rev. 3 classification. Where the previous, revision 2, system is used, the official correspondence table is used.[4] The export and import data are collected according to the SITC revision 2 system. The correspondence between SITC and ISIC (both rev. 2) is from Muendler (2009).

The ICP basic heading expenditure data are allocated to manufacturing industries based on category names. So, for example, all food products (rice, fresh milk, sugar, etc.) are allocated to the 'Food, beverages and tobacco' industry. Given that the ICP categories are organized by consumption or investment purpose, this means that the correspondence is not precise. The main problem with precision is the investment category 'metal products and equipment', which includes investment in metal products (27-28), machinery (29) and electronic and optical equipment (30-33). To avoid a biased allocation, we use the share of imported investment goods to split up this expenditure category.

Gross output data

As mentioned earlier, we rely on the UNIDO INDSTAT database for data on industry gross output. Where available, we use data for 2005 from the 2012 INDSTAT4 database (based on ISIC rev. 3). However, this covers only 29 of the 52 countries. For a further 16 countries, there is data in either the 2012 INDSTAT4 or the 2006 INDSTAT3 (based on ISIC rev. 2) database but for an earlier or later year. Mostly, the data are for a year in the 2000s, but in a few cases we have to go back further. We use information on value added

[4] See: http://unstats.un.org/unsd/class/.

in total manufacturing from the UN National Accounts Main Aggregates Database to put the data on a comparable 2005 basis.[5]

For the 7 countries that have never been covered in UNIDO, we use the following estimation procedure. For most industries, the output share in total manufacturing does not systematically vary with income level and for these we start of with the median cross-country output share. For 5 industries – food, metal, machinery, electronics and transport equipment – there is such a relationship, with the importance of the food industry declining with (the log of) GDP per capita and the other 4 increasing. For these 5 industries we compute the predicted share given the income level. The shares are then normalized to sum to one. The shares are then multiplied by total manufacturing output, which is based on value added from the UN National Accounts Main Aggregates Database and the median value added to gross output ratio across countries.

Import data

As discussed in the main text, imports should only cover imports of products for final demand. To make this distinction, we use the Broad Economic Classification (BEC), which groups traded products by final use. This allows us to exclude BEC categories that are typically used as intermediates: materials, parts, etc. We apply the distinction used in the World Input-Output Database (WIOD) and classify BEC categories 111, 121, 21, 22, 31, 322, 42 and 53 as intermediate products and exclude these from the import data (see the UN classification registry for details on the individual codes).

---

[5] This assumes that the shares of each industry in manufacturing output is unchanged and that the ratio of manufacturing value added to gross output is unchanged.

Balancing input-output data

We have data on gross output from UNIDO, exports and imports from Comtrade and domestic final expenditure from ICP and ideally these would be internally consistent without further adjustments. However, it turns out that often imports would exceed domestic final expenditure or that output is smaller than exports plus domestic final expenditure minus imports. Inconsistencies when mixing sources is not uncommon when compiling National Accounts (see e.g. Heston, 1994, or Lequiler and Blades, 2006) and can be due to measurement error, incorrect product correspondence and differing concepts. As an example of the latter issue, UNIDO's gross output refers only to the formal manufacturing sector, but domestic final expenditure also covers consumption from informal firms. Similarly, domestic final expenditure is valued at purchaser prices, which includes product taxes, trade and transportation margins; gross output is at basic or producer prices; exports is valued fob (free on board) and imports are cif (cost, insurance, freight). Especially the inclusion of product taxes, trade and transportation margins in domestic final expenditure overestimates the size of domestic final expenditure relative to the other flows. Country-specific input-output tables would (again) be needed to fully resolve this, but in their absence we use information from the US input-output tables. Those tables indicate that expenditure on manufacturing products at producer prices is approximately half of expenditure a purchaser prices, with cross-industry variation between about 40 and 60 percent.

In balancing step 1, we multiply domestic final expenditure by one half. In step 2, we reduce imports to be no larger than domestic final expenditure. Data for 40 countries from the WIOD confirms that this constraint holds when input-output tables are available. This adjustment affects about 37 percent of the country/industry pairs in the countries we analyze. This is a substantial share of observations requiring adjustment, but if we follow the same procedure for WIOD countries, the share of imports in domestic supply shows a correlation of 0.54 with the actual input-output data, compared to a correlation of -0.03 when the adjustment is not made.[6]

In step 3, we ensure that industry gross output covers at least exports plus domestic final expenditure minus imports, i.e. domestic intermediate demand is equal to zero. This adjustment affects 35 percent of the country/industry pairs. We could assume that margins make up less than half of domestic final expenditure, which would lead to a smaller number of observations needing adjustment in step 2. However, that would lead to many more adjustments in step 3, so we struck this balance. More in general, this balancing procedure gives greatest weight to the data on domestic final expenditure as the composition of expenditure across industries is left intact. This implies that any differences between our preferred approach and the standard approach – aggregating domestic final expenditure prices using shares in domestic final expenditure – are not (artificially) driven by the balancing choices we make but instead by the differences in the prices of domestic final expenditure, exports and imports. A further reassuring result is that if WIOD data is used directly, rather than our constructed data, the final manufacturing output price levels never differ by more than 1 percent for the group of countries we consider here.

[6] Even for the WIOD economies, where data is of arguably higher quality in many cases, measurement error, classification mismatches, etc. lead to imports being larger than domestic final expenditure in more than 20 percent of country/industry pairs using the ICP and Comtrade data.

Price aggregation

Prices of domestic final expenditure, exports and imports are all given at a greater level of detail than the 14 industries we analyze. The same Törnqvist/GEKS procedure outlined in (A2) is used to aggregate the more detailed prices to the level of the 14 industries. At that point, using the balanced input-output data, (A6) can be applied to compute aggregate manufacturing relative price levels.

Employment data

To estimate the number of manufacturing workers in each country, we draw on a number of sources. For 10 of the 52 countries, the UN National Accounts, Official Country Data provides data on the number of workers (employees and self-employed) in manufacturing. For an additional 20 countries, the ILO publishes employment data. For a further 13 countries, the World Bank's World Development Indicators (WDI) publishes the share of workers in industry, a sector that includes workers in mining, utilities and construction in addition to manufacturing workers. We estimate the share of manufacturing in industry value added using UN National Accounts value data and apply this to estimate the share of manufacturing workers in industry. For the 9 countries where no direct employment data is available we regress the share of manufacturing workers in total employment on the share of manufacturing value added in GDP. We apply the predicted share from this regression to the remaining 9 countries. For all countries, we used total employment from the Penn World Table (PWT) version 8.0 as a control total.

Additional references

Caves, Douglas W., Laurits R. Christensen and W. Erwin Diewert (1982a) Multilateral comparisons of output, input and productivity using superlative index numbers. *Economic Journal* 92, 73-86.

Caves, Douglas W., Laurits R. Christensen and W. Erwin Diewert (1982b) The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica* 50, 1392-1414.

Diewert, W. Erwin (1976) Exact and superlative index numbers. *Journal of Econometrics* 4, 114-145.

Heston, Alan (1994) A brief review of some problems in using national accounts data in level of output comparisons and growth studies. *Journal of Development Economics* 44, 29-52.

Hooper, Peter (1996) Comparing manufacturing output levels among the major industrial countries. In OECD (ed.) *Industry Productivity – International Comparison and Measurement Issues*, pp. 263-292. Paris: OECD.

Lequiler, François and Derek Blades (2006) *Understanding National Accounts*. Paris: OECD.

Muendler, Marc-Andreas (2009) Converter from *SITC* to *ISIC.* downloadable at http://econ.ucsd.edu/muendler/html/resource.html.