

## Description of Dataset Constructed from Other Sources

The dataset used in this paper ('data.dta') contains data on bilateral trade flows, import penetration rates, productivity levels, R&D expenditure levels, and 'PPP quality' levels. It was constructed from the sources described here. In all cases the data correspond to a cross-section of 21 OECD countries, 13 manufacturing industries, all in the year 1997, as described in the paper.

### Bilateral trade flows data:

This data was obtained from the OECD STAN Bilateral Trade Database (2008 Edition). Variables in 'data.dta' based on this source comprise:

- *exports*: value of exports in 1997 US dollars.

### Import penetration rate data:

This data was obtained from the OECD STAN, Industry Indicators database (2008 Edition). Variables in 'data.dta' based on this source comprise:

- *ipr*: import penetration ratio.

### Productivity levels data:

This data was obtained from the Groningen Growth and Development Centre's (GGDC) Productivity Level Database (at <http://www.ggdc.net/databases/levels.htm>). We use the 'benchmark 1997.xls' file. Variables in 'data.dta' based on this source comprise:

- *Gro\_PPP\_SO\_97*: producer price index based on value of gross output.
- *Gro\_PPP\_II\_97*: producer price index based on value of intermediate inputs.
- *Gro\_PPP\_LAB\_97*: producer price index based on value of labor inputs.
- *Gro\_PPP\_CAP\_97*: producer price index based on value of capital inputs.
- *sh\_ii\_Gro*: share of payments to intermediate inputs in gross output.
- *sh\_lab\_Gro*: share of payments to labor in gross output.
- *sh\_cap\_Gro*: share of payments to capital in gross output.
- *lps*: gross output per hour worked.
- *mfpso*: multi-factor (based on intermediate inputs, capital, labor) productivity of gross output.

### R&D expenditure levels data:

This data was obtained from the OECD ANBERD database. Variables in 'data.dta' based on this source comprise:

- *rdexpend*: value of R&D expenditure in 1997 US dollars.

### 'PPP quality' levels data:

This data was created based on Appendix Table 3 of "PPPs for Industry Output: A New Dataset for International Comparisons", Timmer, Ypma and van Ark, 2007 (GGDC Research Memorandum GD-82). Variables in 'data.dta' based on this source comprise:

- *pct\_covered\_total*: percentage of underlying output that is covered in reported observation (for industry-year) of *Gro\_PPP\_SO\_97*.
- *pct\_covered\_exp*: percentage of underlying output that is covered in reported observation (for industry-year) of *Gro\_PPP\_SO\_97* that is based on expenditure (rather than producer) prices.