

DATA FOR R&D SPILLOVER PROJECT

Data have been gathered for two groups of countries. These roughly correspond to the set of industrial countries used in Coe and Helpman (1995), for which R&D data exist from about 1970, and the set of developing and transition countries used in Coe, Helpman, and Hoffmaister (1997).

Industrial countries

The basic **R&D data** for most countries are from the OECD and published in *Main Science and Technology Indicators*. Sharon Standish of the Directorate of Science, Technology, and Industry (DSTI) provided a file with annual data for total R&D expenditures (GERD) and for R&D done by business (BERD), both in millions of current PPP dollars. R&D data for these countries are available from the 1970s or earlier. The R&D data for Germany refer to West Germany from 1964-1990 and to unified Germany thereafter. The start and end years, which represent the maximum period the data are available from the OECD, are shown in Table 1. The countries in Table 1 include all of the countries used in Coe and Helpman (1995) except Israel,¹ as well as Iceland and Korea.

The following transformation converts the data from nominal to real (GERDV, BERDV) in 2000 prices measured at 2000 PPP exchange rates:

$$\begin{aligned} \text{BERDV} &= [(\text{BERD} \cdot \text{PPP}) / \text{PGDP}] / \text{PPP}_{00} \\ \text{GERDV} &= [(\text{GERD} \cdot \text{PPP}) / \text{PGDP}] / \text{PPP}_{00} \end{aligned}$$

Where PPP is the purchasing power parity exchange rate in local currency per US dollar; PPP_{00} is PPP in 2000; and PGDP is the GDP price deflator, 2000=1.² For Korea, R&D data from the Ministry of Science and Technology are available from 1976 whereas the data from the OECD only start in 1995. These data for Korea are in the national currency (BERDNC, GERDNC), and hence

$$\begin{aligned} \text{BERDV} &= [(\text{BERDNC}) / \text{PGDP}] / \text{PPP}_{00} \\ \text{GERDV} &= [(\text{GERDNC}) / \text{PGDP}] / \text{PPP}_{00} \end{aligned}$$

¹ It may be possible to reconcile the 1970-89 data for Israel used by CH with more recently published data for 1989-04.

² The lack of consistent data for wages prevented deflating nominal R&D expenditure by an average of PGDP and an index of wages, as was done in Coe and Helpman (1995).

Most countries do not have data for all years from 1970-2004, either because the data do not start until after 1970 or because there are gaps in the data, which were only collected every other year or even more sporadically. The number of missing years for BERD and GERD from 1970-2004 are indicated in the table. For some countries, DSTI also publishes data on R&D expenditure by industry in the STAN database, and for eight of these countries the data for total industry (IN) are available for more years than are the data for BERD. For these eight countries the correlation between IN and BERD is 0.99 except for Australia where it is 0.72. The ratios of BERD/IN for these eight countries, which tended to be relatively stable and close to 1, were interpolated and then used to estimate missing values of BERD for those observations where only IN data were available. This substantially reduced the number of missing observations for a number of countries, which can be seen by a comparison of the columns labeled BERD and BERD/STAN, where the latter indicates the number of missing years after this interpolation.³

For most countries, data for GERD and BERD are available for the same years. For Austria, however, there are many more observations for GERD, which is available for all years from 1981-2004, than for BERD. The ratio BERD/GERD for Austria is relatively stable at about 0.55 from 1981-93 and then rises monotonically to 0.67 in 2004. This ratio was linearly interpolated and used to estimate the missing values for BERD for 1981-2004. For the full 1970-2004 period, this interpolation left only 8 missing years for BERD for Austria.

For BERDV, the remaining missing observations were estimated using the predicted values from OLS regressions of $\log \text{BERDV}$ on $\log \text{GDPBV}$ and $\log \text{IBV}$, where GDPBV is real value-added in the business sector and IBV is real non-residential private investment (for a few countries it was necessary to use real GDP and real total investment). If significant a time trend (T) and T^2 were also included. The change in the predicted values from these regressions,⁴ which typically had R^2 s of 0.95-0.99, were used fill in the remaining missing observations for BERDV.

With the original OECD source data supplemented, where necessary, by estimates for the missing observations, data/estimates of real R&D expenditure in the business sector are available for the 23 countries (excluding Israel) in Table 1 for the full 1970-2004 period. While the interpolations using the IN data in STAN (and the GERD data for Austria) are based on actual data on R&D expenditure, the interpolations based on the predicted values from the regressions are not. Of the 23 countries, this latter interpolation was needed for about half or more of the total of 35 observations from 1970-2004 for Greece, New Zealand,

³ Coe and Helpman (1995) were not aware of the availability of information in the STAN database useful for estimating missing values of the BERD data.

⁴ Using the change in (rather than the level of) the predicted values effectively assumes that the prediction error remains constant (rather than goes to zero).

and Switzerland; and for about one-third of the observations for Iceland. For the remaining 19 countries only a few observations – usually less than three – had to be estimated in this way. To avoid confusion with the original data series, the data series containing the full set of data/estimates of real R&D expenditure in the business sector measured in 2000 prices at 2000 PPP exchange rates from 1970-2004 are labelled RB.

The missing observations for GERDV were estimated using the predicted values from OLS regressions of $\log \text{GERDV}$ on $\log \text{RB}$, $\log \text{GDPV}$ and $\log \text{ITV}$, where RB is real R&D expenditure in the business sector (the largest subcategory of GERDV), GDPV is real value-added, and ITV is real private investment. If significant a time trend was also included. The change in the predicted values from these regressions, which typically had R^2 s of about 0.99, were used to fill in the missing observations for GERDV. The data series containing the full set of these data/estimates are labelled RG.

R&D capital stocks in the business sector (SB), defined as end of year, are calculated using the perpetual inventory procedure as in CH (1965):

$$SB_t = (1-\delta)SB_{t-1} + RB_t$$

where the depreciation rate, δ , is assumed to be 0.05. The benchmarks are calculated as

$$SB_0 = RB_1 / (\delta + g)$$

where g is the annual average logarithmic growth rate from 1970-1985 (1963-1985 for the G7 countries). Because data on RB are available from the early 1960s for the G7 countries, the benchmarks for G7 countries are calculated for 1962 and the capital stocks from 1963. For the other countries the benchmarks are calculated for 1969 and the capital stocks from 1970. The estimates of total R&D capital stocks (SG) are calculated analogously.

For each of the 23 countries, an estimate of foreign R&D capital (SBFA, SGFA) is defined as the simple average of the domestic R&D capital stocks in the other 22 countries.

Table 1. Basic R&D Expenditure Data⁵

	<u>start-end</u>	<u>number of missing years from 1970 (70-04 is 35 years)</u>		
		<u>BERD</u>	<u>BERD/STAN (corr)⁶</u>	<u>GERD</u>
<u>C H (1995) countries:</u>				
US	63-04	0		0
Japan	63-04	0		0
Germany	64-05	1		1
France	63-05	0		0
Italy	63-06	0		0
UK	64-04	10	2 (0.99)	10
Canada	63-06	0		0
Australia	76-04	9	3 (0.72)	19
Austria	64-05	24 (8) ⁷		9
Belgium	67-05	6		9
Denmark	67-05	4	3 (0.99)	4
Finland	69-06	7	2 (0.99)	7
Greece	81-05	20		22
Ireland	63-05	6	2 (0.99)	6
Israel				
Netherlands	69-05	0		0
NZ	72-03	17		17
Norway	67-05	11	1 (0.99)	11
Portugal	71-05	7		7
Spain	69-05	1	0 (0.99)	1
Sweden	67-05	18	3 (0.99)	18
Switzerland	63-04	17		17
<u>Other & New OECD:</u>				
Iceland	71-03	12		10
Korea	76-05	6		6

⁵ Except for Israel and Korea, data are from the OECD.

⁶ Figures in parentheses are the correlation coefficients between the series for BERD and for IN from the STAN database, as discussed in the text.

⁷ Eight missing observations after interpolation based on GERD, as discussed in text.

In addition to the R&D data, **other data** collected for these countries include total factor productivity and measures of human capital and openness. The estimate of **total factor productivity** from the OECD's analytical database, labelled MFP, assumes Hick's neutral technological progress (i.e., it is equivalent to labour efficiency), is detrended using a Hodrik-Prescott filter, and refers to the total economy, as discussed in Beffy et al. (OECD Economics Working Paper ECO/WKP(2006)10). For Finland, Korea, Norway, and Portugal, estimates of MFP are missing from 1970-75 or 78, and for a few other countries estimates are missing for 1970 or 70-71. These missing values of MFP have been filled in based on an OLS regression of logMFP on logGDPV, T, T², which had R²s of 0.99-1.0 reflecting the detrending of MFP. The series including estimates of the missing values is labeled MFPhat.

A **second estimate of TFP** in the business sector has been constructed analogous to that used in CH:

$$TFP = Y / [K^{\beta} L^{(1-\beta)}]$$

where Y is real value added in the business sector, K is capital stock, L is labor input (defined as business sector employment times average annual hours worked), β is the average labor share, and all variables are indices with 2000=1. All data are from the OECD's analytical database. For Finland, France, Germany, Ireland, Italy, Japan, Netherlands, New Zealand, Portugal, Sweden, and the United States, the OECD only has capital stock estimates for the total economy; for the other countries the capital stock is for the business sector.⁸ For some countries missing data for business sector output or employment were filled in using the growth rate of the corresponding economy wide variable; missing data for the capital stock were filled in by assuming a constant scrapping rate and the perpetual inventory model (following the OECD's method); missing data for hours worked were filled in by assuming it remained constant at the last observation. These measures are labelled TFP.

Human capital is measured by the Barro-Lee (CID Working Paper No. 42, April 2000) measure of average years of schooling. The Barro-Lee data, which are reported for every five years to 2000, were interpolated to get annual series. These series available from 1970-2000 are labelled Hhat. A second series extrapolated to 2004 is labelled Hhatex.

Openness is measured as the ratio of imports of goods and services to GDP, and labelled M.

⁸ As discussed in Beffy et al. (2006) the OECD is moving from a focus on the business sector to using total economy.

Developing and Transition countries with R&D data from the 1980s or later

R&D expenditure data are available for 18 other countries. Except for Israel, Singapore, and Taiwan POC, these are all developing and transition countries. For India, data are available from 1971 on a fiscal year basis, and for 14 other countries data are available from the early 1990s or before, as shown in Table 2. Although data for South Africa are available from 1983-2004, there are 12 years in that period where data are missing.

Table 2. Basic R&D Expenditure Data for Other Countries⁹

	<u>Start-end (number missing)</u>	
	<u>BERD</u>	<u>GERD</u>
Argentina	96-04	96-04
Brazil	00-04	00-04
Chile	02-04	02-04
China	91-04	91-04
Czech Republic	91-05	91-05
Hungary	90-05	90-05
India	71-05	71-05
Israel	89-04	89-04
Mexico	89-03 (1)	93-03
Poland	92-05	90-05
Romania	93-04	91-04
Russian Federation	90-04	90-04
Singapore	90-05	90-05
Slovak Republic	90-05	90-05
Slovenia	93-04	93-04
South Africa	83-04 (12)	83-04 (12)
Taiwan POC	98-04	95-04
Turkey	90-04	90-04

⁹ Except for Brazil, Chile, India, Israel, and Singapore, data are from the OECD.