Title	International Comparison of Productivity Levels Consumer Preferences in Industrial Countries
Author(s)	Movshuk, Oleksandr
Citation	大阪大学, 1999, 博士論文
Version Type	VoR
URL	https://doi.org/10.11501/3155105
rights	
Note	

International Comparison of Productivity Levels and Consumer Preferences in Industrial Countries.

Oleksandr Movshuk

1

International Comparison of Productivity Levels and Consumer Preferences in Industrial Countries.

by

Oleksandr Movshuk

A Thesis submitted to the Department of Economics at Osaka University in partial fulfillment of the requirements for the degree of Doctor of Economics

February 1999

Table of Contents.

Introduction.

The primary task of this Ph.D. thesis is to provide a comprehensive comparison of international economic differences from both the supply-side and demand-side, focusing mainly on industrialized countries. The thesis deals with international differences of productivity and consumer preferences, extending the available literature in several theoretical and applied aspects.

The thesis contains six chapters. Chapter 1 is a general introduction to data transformations used in international economic comparisons. Chapters 2, 3, and 4 focus on international comparisons at the supply side, while chapters 5 and 6 deal with international comparisons at the demand side.

More specifically, chapter 1 deals with the problem of choosing appropriate price converters, or purchasing power parities, to transform original data from national currencies into internationally comparable units. After introducing two major approaches for calculating purchasing power parities, this chapter considers one of the most frequent problems in applying purchasing power parities, when the aggregation pattern of national data do not correspond to available aggregation pattern of purchasing power parities.

Chapter 2 presents results of estimating levels of total factor productivity among nine industrial countries. Presently, the estimation of productivity levels remains one of the most daunting tasks in economic research, with relatively few studies that deal with the problem. Moreover, even when TFP levels have been estimated, almost all studies took into account only one or two factors of production (such as labor and capital), repeatedly ignoring intermediate materials, though the latter frequently has much larger share in total revenues compared with the shares of capital and labor. To incorporate intermediate inputs into productivity estimates, one has to merge national account data and input-output tables (both expressed at purchasing power parities). Due to numerous national peculiarities in compiling national accounts, and especially – input-output tables, this task has proved almost impossible, and has been achieved so far only in a few Japan-US productivity studies, but has never been attempted for a larger sample of countries.

Fortunately, recent efforts at OECD, UNIDO and other international organizations have produced several new databases that contain internationally comparable national accounts and input-output data. Using these databases, chapter 2 reports new estimates of three-input productivity levels for nine industrialised countries (Australia, Canada, Germany, France, Japan, Netherlands, Sweden, the United Kingdom and the United States) at the level of twenty-three manufacturing sectors and the total manufacturing. This chapter also compares these three-input productivity estimates with less comprehensive (but easier to calculate) productivity indexes, estimating, in particular, inconsistencies in country productivity rankings due to the omission of intermediate materials and capital inputs. Quite surprisingly, in most cases there were little changes in country rankings if simpler versions of productivity measurements were used.

In chapters 3 and 4 we apply productivity estimates from chapter 2 to two problems that recently have received considerable attention in the economic literature. In chapter 3 we consider whether our estimates of productivity levels support the frequently expressed idea of productivity convergence. Despite the large number of previous studies on convergence, the vast majority of them dealt with productivity convergence either for the whole economy or for the total manufacturing, but, apparently, never – at more disaggregated levels. Besides, most of available convergence studies relied on very simple productivity measures such as labor productivity, rarely using estimates of total factor productivity.

In chapter 3 we estimated σ-convergence/divergence in productivity with the same sample of countries and manufacturing sectors as in chapter 2. The major conclusion is that productivity convergence was not automatic, taking place only in about the half of considered manufacturing sectors. Second, the identification of convergence was found very sensitive to the concept of productivity, used in convergence estimates, with most inconsistencies when we used the conventional measure of labor productivity.

In chapter 4 we use productivity estimates from chapter 2 to address the perennial problem of the international economics: which of two major theories of international trade – the Ricardian or the Heckscher-Ohlin – provides a better account of actual trade flows among countries. After generalizing the conventional one-input Ricardian model of relative comparative costs to the more realistic case of multiple inputs, the chapter identifies country-pairs when the Ricardian theory has good predictive ability. The same approach is used for verifying the Heckscher-Ohlin theory of international trade. On the whole, the Ricardian theory tends to match real trade flows better that the Heckscher-Ohlin one, though in the majority of considered trade flows neither of the competing theories provided correct predictions.

Chapters 5 and 6 deal with the demand side and consider international differences in consumer preferences for a large number of non-durable goods and services. While essentially all related studies compared international preferences by using time-series data for highly aggregated goods and services, the present thesis develops a new approach to analyze the heterogeneity of cross-section data. The approach aims at identifying individual countries where consumer preferences may be unusual (outlying) compared with the regular consumption pattern in the majority of analyzed countries. After introducing the approach in chapter 5, we apply it in chapter 6 to international consumption data for 22 OECD countries in 1990 and 35 OECD and a few East-European countries in 1993. While most conventional approaches have often been limited to international comparisons for fewer than a dozen goods and services, the new approach allowed to analyze consumer preferences in 73 and 61 categories of goods and services. Interestingly, the majority of identified national peculiarities corresponded to prior expectations about national idiosyncrasies in consumer preferences.

Acknowledgments.

Working on the thesis, I benefited tremendously from the highly useful suggestions of my thesis supervisor Professor Kanemi Ban. I also gratefully acknowledge helpful comments during discussions with Professors Shuntaro Shishido, Charles Yuji Horioka, Yasuo Maeda and Kenzo Abe. Any remaining errors remain my sole responsibility. I also would like to gratefully acknowledge the financial support of the Nomura Cultural Foundation.

Charter 1. The problem of choosing appropriate converters from national currencies.

The first problem encountered in international comparisons is how to convert national currencies into a single comparable unit. The proper choice of converting unit is highly important, because there is often a considerable spread among possible currency converters, making results of international comparisons highly dependent on the initial choice of currency converters.

The most straightforward way to convert national currencies into a common unit is to use their exchange rates. However, the conversion by exchange rates frequently may produce highly misleading results, especially in the regime of floating exchange rates, in which swings in exchange rates reflect not only relative international price differentials, but also general perceptions among market players about the current conditions and future development of specific economies. Such perceptions are highly volatile, frequently subject to pure speculative movements, often producing exchange rates that have little to do with relative price differences.

Consider, for example, recent changes in the yen/dollar exchange rate since 1992. Over this period, the Japan-US relative price level has changed only a little (uniformly decreasing at a 2-3% annual rate). However, the yen/dollar exchange rate has been much more volatile, with the yen appreciating from around 126 ¥/$ rate in 1992 to 80 ¥/$ rate in 1995, with an average annual rate of change 12%.

Evidently, the exchange rate swing reflected market pessimism about the American economy in the wake of the Mexican crisis, and optimism about the future prospects of the Japanese economy, mostly due to the substantial Japanese surplus in

Japan-US trade. However, as the effect of the Mexican crisis gradually waned, and more news about troubles in the Japanese financial system penetrated the market, the yen appreciation came to a halt, with a subsequent sharp drop in the yen's value to more than 140 ¥/$ in 1998 (or 16% annually), again substantially overshooting the change in Japan-USA relative price levels.

While a number of early international comparisons relied on currency exchange rates (for example, the classical paper by Arrow *et. al,* 1961), the vast majority of more recent studies on international comparison use more stable price converters which are derived from direct international price comparisons for specific outputs, commodities and services. Hereafter, we refer to such direct estimates of relative prices as purchasing power parities, or PPP for short.

This chapter considers two major approaches to the calculation of PPP, which essentially reflect two ways of subdividing gross domestic product into its constituent parts.

The first approach makes comparisons from the production side of GDP, subdividing GDP into separate industries. To estimate PPP, this approach uses information from censuses of production or official industrial surveys. Since this approach is essentially comparing producer prices for specific industries, it is frequently called as 'industry of origin', or 'production approach'. The production approach is particularly well suited for international comparisons of output and productivity.

The second approach makes international comparisons in terms of the major macroeconomic components of GDP (consumption, government, and capital formation), and their components at less aggregated levels (such as food, clothing, and housing as components of consumption). To estimate PPP, this approach compares prices when goods and services are purchased in the domestic market. Due to the focus on particular categories of gross domestic expenditures, this approach is conventionally defined as the 'expenditure approach'. It has been most frequently applied in international comparisons of consumer preferences, but it can also be applied to international comparisons of output and productivity (once a few adjustments are made in the original PPP). The estimation of PPP by the production and expenditure approaches will be discussed in more detail in sections 1.1 and 1.2.

Once original PPP estimates are obtained, one has to solve a different problem, which is common to both production and expenditure approaches. In most cases PPP estimates are available only from external sources, and very frequently these data are too disaggregated, requiring an aggregation of original PPP estimates. After discussing the merits and drawbacks of available methods of PPP aggregation, section 1.3 describes the aggregation methods used in this study.

Section 1.1. Purchasing power parities based on the production approach, and their use for productivity studies.

The first study that used PPP estimates, derived by the production approach, appears to be Paige and Bombach (1959). Before this seminal study, many international comparisons of productivity (for example, Rostas (1948), or Maddison (1952)) dealt with output in physical units, which greatly limited their scope, and excluded the possibility of productivity comparisons in industries with heterogeneous output.

Paige and Bombach suggested converting the value of output to a common currency, using ex-factory prices – or unit values – by dividing the value of sectoral sales (in national currency) by the corresponding quantity of produced commodities. Subsequently, this approach to calculate PPP was applied by West (1971), Smith, Hitchens and Davies (1982), and recently, – in a number of studies under the aegis of the International Comparisons of Output and Productivity (ICOP) of Groningen University (see Ark and Pilat, 1993; Ark, 1995; Pilat and Rao, 1996; Pilat, 1996).

Unfortunately, except for a number of industries that produce relatively homogenous products (such as tobacco or iron), it is not always possible to match the majority of produced commodities. Table 1-1 shows coverage ratios of matched commodities in several ICOP estimations of international price ratios. Matched products often cover less than 50% of all sectoral output, raising justifiable doubts about how representative the ICOP estimates are. For example, the coverage ratio for the manufacturing sector as a whole is only slightly above 10% in the France/UK comparison, and around 20% in the USA/UK comparison, and deviates little from zero for a number of industries (such as electrical engineering).

Table 1-1. Coverage ratios of selected PPP estimates by the production approach.

	France/UK (1984)		USA/UK (1975)	
	France	UK	USA	UK
Food products	[b]	[b]	14.6	25.8
Beverages	[b]	[b]	28.7	47.8
Tobacco	[b]	[b]	74.6	93.8
Textiles	[b]	[b]	43.5	23.1
Apparel	18.7	20.1	8.5	7.6
Leather products	46.7	50.2	[a]	[a]
Wood products	10.1	6.1	11.4	11.0
Paper products	40.4	24.4	22.3	29.9
Chemical products	15.9	14.0	15.2	18.9
Petroleum refining	[b]	[b]	76.1	76.5
Stone, clay, and glass products	13.2	9.5	8.2	16.5
Basic metals and metal products	13.3	10.1	10.7	10.9
Electric engineering	2.5	2.6	4.9	10.7
Machinery and transport equipment	20.4	11.1	25.1	15.3
Total manufacturing	11.6	11.6	20.4	23.1

Source: van Ark (1990), table 2.
[a]Leather and footwear were included in chemicals
[b]PPP was not estimated

Unfortunately, such low coverage ratios are typical in PPP estimations by the production approach. For example, more recent Germany/US price comparisons from 1987 represented around 25% of total manufacturing sales in both countries (van Ark, 1996, p.26). The coverage was even lower for the France/US comparison, covering slightly above 10% of total manufacturing sales (ibid.).

In cases when few or no matching categories can be found, the production approach extends PPP estimates for available subcategories to missing ones, or assigns to the latter PPP estimates at a higher level of aggregation. Obviously, this is a very crude procedure, but in most cases there is no other way to estimate unit value ratios, mostly due to the confidentiality of price information for disaggregated manufacturing products.

However, the actual bias in the production approach appears to be not so large. For example, the PPP estimates by the ICOP for Japan/USA and Germany/USA (reported in van Ark and Pilat (1993)) were subsequently checked by independent work at McKinsey (1993). McKinsey had better access to industry-specific price information, and also could utilize expert opinion about differences in the product mix and quality in specific manufacturing sectors. As a result, McKinsey made several substantial corrections in the original ICOP estimates for investment goods (primarily, various types of machinery equipment). Nevertheless, for the rest of the manufacturing sector, the majority of original PPPs estimated by the ICOP changed little (see Pilat, 1996, p. 5; also van Ark, 1996, p. 31-32).

The lack of comprehensive price information on machinery is unfortunate, because the machinery sectors are often associated with high-tech production methods where productivity comparisons are the most interesting. However, PPP estimates from the alternative, expenditure approach can be useful, because they usually include price comparisons for a wide variety of investments, and among them durable goods, with a good correspondence to the conventional division of machinery sectors into general, electrical and transportation equipment.

A more serious drawback of PPP estimates by the production approach is that they cover bilateral price comparisons, derived from pair-wise comparisons, when two measures of relative prices (with domestic and foreign quantity weights) are averaged by applying the 'ideal' Fisher index.

Unfortunately, the Fisher index does not produce *multilateral* PPP estimates. Consequently, using the ICOP estimates of PPPs, it is not possible to derive a correct index for countries A and B from A's and B's comparison with a country C (Pilat and Rao, 1996). For example, ICOP purchasing power parities for France-USA and Germany-USA in metal products are 7.52 FF/$ and 2.20 DM/$, respectively (van Ark, 1996, p. 28-29), so that the *derived* PPP for France-Germany is $7.52/2.20 = 3.42$ FF/DM. However, a *direct* estimation of France-Germany's PPP by Freudenberg and

Unal-Kesenci (1996) produced a different conversion factor – 3.19 FF/DM (p. 56).

So far multilateral PPPs from bilateral ICOP estimates have been calculated only for the total manufacturing PPP (see Pilat and Rao, 1996), and no similar multilateral price parities are available at less aggregated levels. As a result, international comparisons that use PPPs from the production approach can be made only with reference to country-pairs for which the original PPP is available.

Section 1.2. Purchasing power parities based on the expenditure approach.

The expenditure approach goes back to the seminal study by Gilbert and Kravis (1954). This study was based on the national accounting framework, with separate estimates of PPP for consumption, capital formation, government expenditures, and total GNP for France, Germany, Italy, UK (with the USA as the base country) in 1950. In the 1960s, Kravis and his associates established the International Comparison Project (ICP) at the University of Pennsylvania. In the late 1960s the importance of the ICP work was recognized by a number of international organizations, including the United Nations, the World Bank, OECD and Eurostat. These organizations provided sufficient financial support for the extension of the ICP both in terms of country coverage and in terms of the level of desegregation of goods and services.

Starting in 1970, the ICP work proceeded at regular intervals, usually making benchmark price comparisons every 5 years. The commodity coverage was also standardized. With minor modifications, it usually includes 110 categories of consumption goods, 32 categories of capital formation, and 5 categories of government expenditures. These PPP categories at the least disaggregated level will be defined as the basic heading level.

Up to the present, there have been seven benchmark studies of the ICP. The first benchmark study was still relatively limited, producing PPP estimates for only six countries in 1967. This study was the first to estimate PPP not only for developed countries, but also for a number of less developed countries, for which the discrepancy between exchange rates and estimated PPPs was found to be especially pronounced.

In the second benchmark study the country coverage was further extended to 10 countries with PPP estimates for the year 1970 (see Kravis, Kenessey, Heston and Summers (1975)). The study made price comparisons using highly standardized descriptions of goods and services with a large number of explicitly specified characteristics (such as physical identity, equivalence in quality and use, etc.), ensuing a relatively high quality of price comparisons. Due to the backing of the United Nations and other international organizations, the ICP work was conducted in close co-operation with national statistical offices, so that the ICP had access to usually confidential data, used in the calculation of domestic consumer price indexes.

The ICP became a truly global project during its third benchmark study in 1975, when as many as 35 countries joined the project. Kravis, Heston and Summers (1982) provide a highly readable account of major results of this benchmark study. Due to the large country coverage, Kravis et. *al* could use the ICP results to conduct an extensive cross-country study of consumer preferences by running cross-country regressions for specific goods and services.

The fourth ICP study took place in 1980 and has been so far the most extensive benchmark study in terms of covered countries (60 in total). The results of this study are reported in United Nations-Eurostat (1987) joint report.

All subsequent benchmark studies of the ICP were conducted on a regional basis. For example, the fifth ICP study consisted of independent PPP estimates for OECD countries (by OECD and Eurostat), for Asian countries (by the United Nations), for African nations (by Eurostat), for East European countries (by the Economic Commission for Europe), and for the Caribbean nations (by Eurostat). The UN was assigned with the final task of 'globalizing' these intermediate PPP estimates, but, due to the ongoing financial crises at the UN, the task has not yet been completed. Except for the PPP estimates for OECD countries, available in OECD (1987), all other regional results from this benchmark study still remain publicly unavailable. Furthermore, even the published results for the OECD countries remain incomplete, because they include only PPP estimates at aggregated levels (covering, for example, only 31 consumption categories, derived from more than one hundred basic heading data for consumption).

The sixth benchmark study of the ICP was even less extensive, covering only some OECD members (22 countries in total), and its results were only partially published in OECD (1992), again with the omission of results at the basic heading level.

The seventh ICP study took place in 1993, and consisted of regional estimates for OECD, African, Asian and East European countries. A novel feature of the latest ICP study is the extension of country coverage to a wider sample of East European countries (13 in total) compared with at most 3 countries in previous ICP studies. Results for the subset of OECD countries has been published in OECD (1995), but again containing only PPP estimates for aggregated categories of goods and services.

The comprehensive coverage of the ICP estimates allows us to apply them not only to the conversion of consumer expenditures, but also to the price conversion of various types of investment goods, most importantly producer durables for which, as we have already noted, the application of the production approach is the most problematic.

Besides, the ICP estimates contain a number of consumer goods that can be used for output conversion in other manufacturing sectors as well (such as food, beverages, tobacco, textiles, footwear, furniture, pharmaceutical products, etc.). However, there are several conceptual differences between the production and expenditure approaches that should be accounted for when the ICP data are substituted for PPP estimates based on the production approach.

First of all, the expenditure estimates of PPP are based on price comparisons at the retail (for consumer goods) or wholesale level (for investment goods), thus containing distribution and transport margins. Second, indirect taxes (such as VAT) and subsidies to producers also affect PPP estimates based on the expenditure approach. Finally, while the production approach produces PPP estimates for domestically produced commodities, the expenditure approach takes into account not only domestic but also imported goods.

It is not particularly difficult to eliminate the first two inconsistencies between the production and expenditure approaches to PPP estimation because sector-specific distribution and transport margins as well as ratios of indirect taxes and subsidies can be extracted from input-output tables, which are usually expressed in producer prices. However, it is much more difficult to account for the bias in the expenditure approach due to the inclusion of imported goods. As a result, this adjustment was not attempted in most studies that 'mapped' available ICP estimates to separate manufacturing sectors, such as Jorgenson, Kuroda (1992) and Kuroda (1996). The same approach will be followed in this study whenever purchasing power parities from the ICP are 'mapped' into currency converters for specific outputs.

Section 1.3. Methods for the aggregation of individual purchasing power parities.

Quite often currency converters do not accord with corresponding data, expressed in national currency. Cases, when data in national currency are more disaggregated than available PPPs present no problem, because the correspondence can be achieved by a simple summation of the data in national currency. A more problematic case involves the opposite case, when there are PPP at more disaggregated level compared with data in national currency.

For example, the output of non-electrical machinery (in constant domestic prices) is usually available at 3-digit ISIC level (ISIC 382). On the other hand, PPP estimates from the ICOP often refer to a pair of sectors at 4-digit level of aggregation – ISIC 3822 (computers and office equipment) and the rest of ISIC 382, such as in Pilat (1996, table A3). Due to the well-known difficulty in estimating 'hedonic' price indexes for computers and office equipment, real output time series for ISIC 3822 are usually available in very few countries, thus making necessary the aggregation of the two available PPPs into a single PPP estimate for the whole non-electric machinery.

Another case when the aggregation of original PPP was necessary in this study involved ICP estimates from the latest benchmark study in 1993. Available data included PPP estimates only at the basic heading level. On the other hand, the estimation of the Almost Ideal Demand System (Deaton, Muelbauer, 1980) involved the total budget (expenditure) term, which could be calculated by the division of consumer expenditures in national currencies by PPP, aggregated from the original PPP estimates at the basic heading level.

The first case of PPP aggregation is conceptually simple, because the lack of transitivity in the ICOP data allows to use simple aggregation methods for binary comparisons. The most conventional choice is to apply the 'ideal' Fisher index. Consider two countries, denoted by C and U. The index U refers to the base country. In accordance with common practice, let the United States be the base country. Suppose it is necessary to aggregate PPPs for m detailed categories of output (or consumer expenditures). First, these PPPs for detailed categories are aggregated, using US expenditure weights and domestic expenditure weights, using the following formulas:

$$\bar{\rho}^{US} = \sum_{i=1}^{m}\left(\frac{P_i^C}{P_i^{US}}\right)\omega_i^{US} \tag{1.1}$$

$$\bar{\rho}^{C} = \frac{1}{\sum_{i=1}^{m}\left(\frac{P_i^{US}}{P_i^{C}}\right)\omega_i^{C}} \tag{1.2}$$

where P_i^C is price level in country C on i^{th} good, weights given by $\varpi_i^{US} = e_i^{US} \Big/ \sum_{i=1}^{m} e_i^{US}$

and $\varpi_i^{C} = e_i^{C} \Big/ \sum_{i=1}^{m} e_i^{C}$, where e - expenditures in domestic currencies.

Geometric average of $\bar{\rho}_C$ and $\bar{\rho}_{US}$ yields the Fisher index $\bar{\rho}_{C\text{-}US}$ for aggregated PPP over m disaggregated categories:

$$\bar{F}_{C\text{-}US} = \sqrt{\bar{\rho}_C \times \bar{\rho}_{US}} \tag{1.3}$$

The application of Fisher index for PPP aggregation is illustrated in Table 1-2.

Table 1-2. Aggregation of PPP for electrical machinery (ISIC 383) in 1987.

	Relative prices, USA=1				Domestic expenditure shares			
	JPN	DEU	FRA	USA	JPN	DEU	FRA	USA
TV, radio, comm. equip.	138.8	2.930	8.198	1.000	0.582	0.409	0.563	0.553
Electrical equip., nes	148.5	2.669	9.556	1.000	0.418	0.591	0.437	0.447
Domestic weight	142.7	2.770	8.740	1.000				
USA weight	143.1	2.814	8.805	1.000				
Fisher index	142.9	2.792	8.773	1.000				

Sources: relative prices are taken from Pilat, table A.3. Expenditure shares in national currencies are calculated from gross output values for the year 1987, using STAN database (OECD, 1995).

Though the example in Table 1-2 is very simple, it nevertheless produced a few interesting results. First, there is a clear 'Gerschenkron effect', when index numbers, weighed by foreign weights, exceed index numbers, weighted by domestic weights. Second, compared with almost identical composition of ISIC 383 among Japan, France and the United States (as can be seen in their expenditure shares), Germany shows an unusual pattern in expenditure weights. Consequently, the aggregation in the Germany-US pair produced the largest spread between PPP, aggregated by domestic and base-country weights.

Finally, data in table 1-2 can be used to illustrate the lack of transitivity in the Fisher 'ideal' index. Consider Japan-US and Germany-US estimates of the aggregated PPP. From these estimates one can derive Japan-Germany PPP as equal $\overline{F}_{JP-DE}=\dfrac{\overline{F}_{JP-US}}{\overline{F}_{DE-US}}=\dfrac{142.9}{2.792}=51.19$ ¥/DM. However, a direct use of Germany as the base country (instead of the United States) produces a different yen-mark conversion rate, which equals 50.66 ¥/DM.

To achieve the base-country invariance, there are several alternative aggregation methods. The most widely used are Geary-Khamis (GH), Elteto-Koves-Szulc (EKS), and Walsh aggregation methods. Each of these methods have a number of strong and weak points which, and there is not yet a general consensus which is the best aggregation method among the three alternatives.

Let n represent the number of countries, and m – the number of disaggregated categories. The GH method calculates aggregated PPP by solving the following system of equations.

$$\pi_i = \sum_{j=1}^{n} \frac{ppp_{ij}}{PPP_j} \left[\frac{q_{ij}}{\sum_{j=1}^{n} q_{ij}} \right], \qquad i = 1,\ldots,m \tag{1.4}$$

$$PPP_j = \frac{\sum_{i=1}^{m} ppp_{ij} q_{ij}}{\sum_{j=1}^{m} \pi_i q_{ij}}, \qquad j = 1,\ldots,n \tag{1.5}$$

where ppp_{ij} denoted the original disaggregated PPP estimates, which are used to calculate the aggregated PPP_j of country j. In the first subsystem, the international price π_i of i^{th} category of expenditures is the quantity-weighted average of PPP-adjusted price of i^{th} category, calculated across n countries. In the second subsystem, the aggregated PPP of country j equals the ratio of total expenditures in country j (evaluated in national prices) to its total expenditures (evaluated at international prices). To get aggregated PPP_j for country j, it is necessary to estimate (1.4) and (1.5) simultaneously.

One distinctive feature of the GH aggregation method is additivity, which means that that aggregates, converted with PPP from the GH method, will be equal to the sum of converted values of their components. This is not the case with both the EKS and Walsh methods. As a result, the GH method is preferable for the analysis of industrial or expenditure *structure* across countries, such as the share of investment expenditures in GDP, of the share of expenditures on food in the total consumption expenditures.

On the other hand, since the GH method treats all countries as a common group, its estimate of international price π_i may be too much influenced by the price structure of countries that have a large share in the group's expenditure, as evident in equation (1.4). In contrast, the EKS aggregation method treats members of the group as independent units, assigning them equal weight. The aggregation by the EKS method can be divided into two major steps. First, PPPs are aggregated for all country-pairs by applying equations (1.1) – (1.3). Second, these binary PPPs are 'multilateralized' by the following 'bridge-country' formula:

$$EKS_{C,U} = \left[\overline{F}_{C,U}^{2} \prod_{\substack{l=1 \\ l \neq C,U}}^{n} \frac{\overline{F}_{C,l}^{2}}{\overline{F}_{U,l}^{2}} \right]^{\frac{1}{n}} \tag{1.6}$$

where l denotes countries other that C and U.

The EKS aggregation method is transitive, because the relative price aggregate between C and U makes use of PPPs in all other countries in the group. Since the original country-pair gets weight 2, while all remaining bridge-country comparisons get the same weight 1, the method allows to preserve specific characteristics of price structure in countries C and U.

It also can be shown that the EKS provides multilateral PPP aggregates that have the least deviation from the aggregated PPP, calculated if prices in each pair of countries are compared in a separate bilateral comparison. The last property makes the EKS aggregation method well suited for the comparison across countries of the volumes and prices of individual aggregates of GDP, preserving to the high degree national peculiarities in price and expenditure patterns.

One disadvantage of the EKS method is that its routine application is quite difficult, since the method requires calculations of a large number of 'bridge comparisons' $\frac{\overline{F}^2_{C,I}}{\overline{F}^2_{U,I}}$. In contrast, the computation burden is significantly lower in the Walsh aggregation method (Walsh, 1901). The Walsh method achieves transitivity by using weights that are common to all compared countries. These common weights are calculated as cross-country geometric averages:

$$\omega_i^{Walsh} = \prod_{j=1}^{m}\left[\frac{e_{ij}}{\sum_{i=1}^{n} e_{ij}}\right]^{\frac{1}{m}} \tag{1.7}$$

With the Walsh aggregation procedure for PPP given by:

$$PPP_j^{Walsh} = \prod_{i=1}^{n} ppp_{ij}^{\varpi_i^{Walsh}} \tag{1.8}$$

After averaging expenditure weights across m countries in the comparison, the same input weights are applied in the PPP aggregation of inputs, thus achieving transitivity. Diewert (1996) showed that among indexes with averaged weights, the Walsh geometric mean index is the most successful in passing tests for index numbers (p. 14). Besides, Diewert (1981) in theorem 21 proved that the Walsh index is exact to the Cobb-Douglas aggregation function *f*, defined as $f(x) = \alpha_0 \prod_{i=1}^{n} x_i^{\alpha_i}$, where $\alpha_0 > 0$, α_1 ... $\alpha_n > 0$, and $\sum_{i=1}^{n} \alpha_i = 1$.

Due to the relative simplicity of the Walsh aggregation method, along with its correspondence to the Cobb-Douglas aggregation function, the method will be the primary approach for aggregating too much disaggregated original index numbers.

Section 1.4. Conclusions.

In this chapter we dealt with specific problems in international comparative studies which has to be solved during the preliminary stage, when the original data in national currencies are converted to internationally comparable units.

We have noted two sources of the potential bias in reported results due to inaccurate original data in common currency. First is the lack of correct price converters. Currently there are few studies that use the plain conversion with the market exchange rates, so that this bias may seem to be minor. However, a more careful look at the majority of international comparative studies (especially ones on relative productivity levels) still often apply a single price converter (usually - for the total GDP) to various disaggregated categories of output, thus neglecting the variability of relative prices of GDP components across countries. Due to the wide disparities in relative price levels, it is important to avoid using not only the market exchange rates, but also purchasing power parities for the total GDP.

The second source of bias during the original data compilation stage is due to incorrect use of aggregation methods for available purchasing power parities. The extend of this bias is less clear, but still it may become quiet significant in comparing countries that have different size (and, consequently, different weights in equation (1.4)). Moreover, the choice between the GH and EKS aggregation formulas may be impractical in the most researchers, since both require prohibitively large amount of calculations which may be feasible only for large institutions such as OECD or the World Bank (often requiring 5-7 years to produce global estimates).

Given these difficulties, we emphasized in this chapter the usefulness of generally neglected Walsh aggregation method. In addition to a sound theoretical basis (such as its correspondence to the Cobb-Douglas production function in productivity studies), the Walsh method can be easily implemented even with large samples of data. In the next chapter we will illustrate the application of the Walsh method in aggregating too much disaggregated purchasing power parities to the level of major machinery sectors. In sum, both theoretically and practically the Walsh aggregation method appears to be a reasonable compromise between unnecessary complexity and sufficient precision of estimates, and as such the Walsh method deserves further use in international comparative studies.

Chapter 2. International comparison of sectoral productivity levels in major OECD countries.

Estimation of international differences in productivity levels may be among the most interesting and useful topics of economic research. As no other economic indicator, comparative levels of productivity may express national achievements in economic development, distinguishing countries that managed to alleviate the problem of scarce economic resources by increasing productivity of their use. High productivity levels may be a reflection of superior management skills, accomplishments in technological development, etc. On the other hand, failures to achieve international productivity advantage almost certainly predestine nations to sagging relative living standards.

Estimates of productivity levels are also in the "wanted lists" in many fields of economic science itself, ranging from the ongoing debate on productivity convergence to the determinants of international trade flows. Finally, little is known about the magnitude of productivity gaps among nations and how these gaps can be explained.

Yet, the demand for estimates about international productivity levels failed to induce comparable supply of economic studies on this topic. Especially rare have been studies that dealt with the productivity of multiple inputs (or total factor productivity, TFP for short), not just the productivity of labor. Besides, there have been few studies that made productivity comparisons for more than two countries.

The scant supply of multy-country studies of TFP can be explained by both theoretical and practical difficulties that researches have to solve. First, in order to calculate TFP, one needs to aggregate multiple inputs into a single input index, and there is no standard way yet how to choose the most pertinent weights in this aggregation, especially in the multilateral setting.

Even more daunting is the lack of relevant data for TFP calculations. The most notorious is paucity of data for capital inputs, and especially – for intermediate materials. The problem of scarce data becomes even more serious at more disaggregated levels of analysis.

The estimation of international TFP levels requires output and input measurement in internationally comparable units, which can be obtained only if appropriate PPPs are available. It is noteworthy that facing numerous difficulties in estimating sector-specific PPPs, the Bureau of Labor Statistics (USA) decided not to release estimates of TFP levels, producing only growth rates productivity, which do not require the conversion into a common currency.

The chapter addresses and attempts to solve some of the above-mentioned difficulties in the productivity calculations. Section 2-1 reviews recent major approaches in the calculation of productivity levels. Section 2-2 focuses on the productivity calculation by the index number approach, dealing with the problem which index number may be the most appropriate approach to TFP calculations, given the bilateral nature of available PPPs from the ICOP. Section 2-3 describes data sources of productivity estimates. Section 2-4 reports TFP estimates for 3-input TFP index (intermediate materials, capital, and labor), and other, less comprehensive index numbers, such as capital and labor, and labor only. Given the substantial amount of extra efforts for the extension of inputs beyond labor, it is interesting how much inconsistency in national productivity rankings may be created by the omission of capital and intermediate materials. Though the omission of these inputs is a widespread practice, there appears to be no study that compared productivity estimates with different number of included inputs.

Section 2.1. Review of available approaches to estimate international productivity levels.

Since the mid-1970s, considerable research efforts in productivity analysis has produced a variety of methods that estimate productivity levels. Following the classification by Diewert (1981), three general approaches in productivity estimations can be distinguished.

First, productivity levels can be calculated econometrically, either by fixed effect approach in panel estimation, or by the stochastic production frontier method. Green (1997) provides a comprehensive review of both approaches. Second, productivity levels can be estimated by the data envelopment analysis (DEA), which is a linear programming counterpart to the stochastic production frontier method. Charnes *et. al* (1995) extensively reviewed recent developments in the DEA approach. Finally, productivity levels can be calculated by various index numbers in either bilateral or multilateral international comparisons. An up-to-date review of the index number approach is given in Good *et. al* (1997).

Consider advantages and disadvantages of each of these. An important advantage of both econometric methods is that they may exclude statistical noise from productivity estimates. On the other hand, econometric methods may require a large number of observations, especially for the estimation of flexible production functions, such as the translog.

Among remaining approaches, the DEA may seem to be more attractive than the index number approach, since the DEA, as a non-parametric approach, does not impose a priori any structure on the production function (Greene, 1997, p. 97-98), whereas the index number assumes the constant returns to scale and perfect markets for inputs. However, the DEA is primarily designed for productivity estimations from cross-section data, such as ones, derived from a multilateral international comparison for a specific industry. Since the sample size of the present study will contain only nine developed countries, the application of the DEA was not feasible as well due to the small number of the remaining degrees of freedom.

As a result, though it may be very interesting to compare major approaches to productivity estimation after they are applied to the same dataset, we finally opted for the index number approach, chiefly because it does not require pooling a large number of countries into a single cross-section to derive their productivity levels.

Though the theory of index numbers was already well-developed in the early 1920s (Diewert, 1993), the application of index numbers to international productivity comparisons can be traced only to the early 1980s, particularly, to the seminal papers by Caves and Christensen (1980) and Caves, Christensen and Diewert (1982).

Despite the relatively short history, the research field has already gone through at least three distinct stages. At each of these stages, the measurement of productivity was improving either due to the use of more sophisticated theoretical concepts of productivity or due to a better quality and coverage of data.

As shown in an early survey of productivity studies (Kravis, 1976), until the early 1980s the majority of productivity studies dealt with bilateral estimates of labor productivity, rarely venturing in the measurement of joint productivity of several inputs. For example, this survey mentioned just one truly multilateral estimate of total factor productivity (p. 24-25) by Denison (1967) for the whole economy of several developed countries, deriving TFP indexes with uniform American weights for inputs.

Apparently, the first attempt to estimate international TFP levels by the index number approach can be traced to Christensen *et. al* (1982). The study covered nine countries (Canada, France, Germany, Italy, Japan, Korea, the Netherlands, the United Kingdom, and the United States), and estimated TFP levels with respect to labor and capital. Besides, the study was among the first applications of a multilateral version of the Tornquist-Theil (T-T) index of productivity, developed in Caves and Christensen (1980), and, in more general form - by Caves, Christensen and Diewert (1982).

While TFP comparisons in Christensen *et. al* (1982) were limited to the level of the whole economy, at the next stage of international productivity studies the scope of TFP estimates was extended to specific industries, starting from the seminal paper by Jorgenson and Kuroda (1987). The study dealt with Japan-US productivity levels, and covered thirty industrial sectors. Similarly to Christensen *et. al* (1982), factors of production included capital and labor. PPPs were calculated for specific industry, by "mapping" benchmark ICP estimates to relevant industrial sector, with subsequent "peeling off" transportation and trade margins. Unfortunately, the study contained surprisingly limited information about estimated differences in productivity levels, mentioning only years when technological gap was closed between Japan and the United States, without exact figures about the extend of relative productivity gap for specific years.

Subsequently, sectoral TFP levels with reference to both capital and labor were calculated in a number of studies, covering Japan, Germany and the United States (Ark and Pilat, 1993), French and Germany (Freudenberg and Unal-Kesenci, 1996), Australia, Canada, Japan, Germany, France, the Netherlands, Sweden, the United Kingdom, and the United States (Pilat, 1996), and, most extensively –Belgium, Canada, Germany, Denmark, France, Italy, Japan, Sweden, the United Kingdom and the United States in Harrigan (1996).

The third phase in recent estimates of international TFP levels was originated by Jorgenson and Kuroda (1990), who extended the number of inputs to three, including intermediate materials along with capital and labor. The study estimated TFP levels in Japan-US bilateral comparison. Due to paucity of data for intermediate inputs, the inclusion of intermediate inputs in Jorgenson and Kuroda (1990) still remains unique among current productivity studies. To the best of our knowledge, there has been no attempt to replicate its approach with respect to other countries (the same observation was made by Ark (1996, p. 36)). In this chapter, we will attempt to fill in the gap, calculating TFP estimates for nine OECD countries, using the same country coverage as in Pilat (1996), but with inputs represented by intermediate materials, capital, and labor.

The next section deals specifically with the index number approach to TFP estimation, focusing, in particular, on merits and drawbacks of index numbers that have been suggested for international comparisons of TFP.

Section 2.2. The problem of choosing appropriate index number in international productivity comparisons.

The index number approach usually defines TFP as a ratio of output to an aggregated measure of inputs. To derive relative TFP levels, one can express national output per unit of aggregated inputs (i.e., input index) for country *C* and then make comparison with the same measure for the base country, say the United States (*US*):

$$TFP^{C,US} = \frac{Y_{i,t}^{C}}{I^{C}} \Big/ \frac{Y_{i,t}^{US}}{I^{US}} \qquad (2\text{-}1)$$

with *Y* denoting output, and *I* representing an index of aggregated inputs.

There is a large number of ways to aggregate inputs into the denominator of (2-1). The choice of the most appropriate index number to aggregate inputs should satisfy a number of desirable properties, such as index number 'tests', originally formulated by Fisher (1927):

(1) Identity test. If quantities and weights do not change, the index number should equal to unity.

(2) Proportionality test. If quantities are scaled up by some constant, the value of index should change by the scaling factor.

(3) Commodity reversal change. The value of index remains the same regardless permutations in the order of inputs.

(4) Commensurability test. The value of index is not sensitive to the units of input measurements.

(5) Determinateness test. When some (but not all) individual inputs are zero, the value of index should not become zero.

(6) Country-reversal test. For countries A and B, the test implies that $I^{A,B} = 1/I^{B,A}$.

(7) Transitivity test. For countries A, B, and C, the test implies that the following inequality holds $I^{A,B} = I^{A,C} / I^{B,C}$.

Next two desirable properties are due to Diewert (1976):

(8) Exact indexes. Index numbers are exact, if they can be derived from some transformation function, such as production, utility, cost, profit function, etc.

(9) Superlative indexes. If index numbers are exact to a flexible functional form (defined as a second-order local approximation to arbitrary functional form), such index numbers are superlative.

These tests of index numbers are convenient yardsticks for the evaluation of alternative index numbers. Consider, for example, the popular Tornquist-Theil (T-T) index, which relates aggregated inputs in C and US as follows:

$$\ln \frac{I_{T-T}^{C}}{I_{T-T}^{US}} = \frac{1}{2}\left(W_U^C + W_U^{US}\right)\left(\ln U - \ln U^{US}\right) + \frac{1}{2}\left(W_K^C + W_K^{US}\right)\left(\ln K^C - \ln K^{US}\right)$$
$$+ \frac{1}{2}\left(W_L^C + W_L^{US}\right)\left(\ln L - \ln L^{US}\right) \qquad \text{(2-2)}$$

where U, K, and L denote intermediate materials, capital and labor, and W denotes input shares in the total output.

Combining (2-1) and (2-2), relative productivity $TFP_{T-T}^{C,US}$ can be expressed

$$\ln TFP_{T-T}^{C,US} = \ln \frac{Y^C}{Y^{US}} - \ln \frac{I_{T-T}^{C}}{I_{T-T}^{US}} \qquad \text{(2-3)}$$

It can be shown (see Diewert, 1996) that the T-T index passes tests (1)-(6) and (8)-(9) (specifically, the T-T index is exact to the translog production function, which is a flexible function). However, similarly to the Fisher ideal index, the T-T index fails the transitivity test (7).

The lack of transitivity in the T-T index means that its TFP estimates are not base-country invariant, and productivity rankings across countries may be different, depending on which country serves as base-country. As a result, the T-T index, as given by (2-2) and (2-3), can be applied to only bilateral comparisons of TFP.

The lack of transitivity of the T-T index is due to different weights, used in aggregation of inputs in countries A, B, and C. One solution to this problem is to use weights that are not specific to individual countries. When such uniform weights are calculated as geometric averages across countries, we get the Wash aggregation method (Walsh, 1901), already discussed in the previous chapter in the context of price aggregation, but presently applied to input aggregation as follows:

$$\ln TFP_W^{C,US} = \ln Y^C - \ln Y^{US} - \begin{bmatrix} \overline{W}_U^{wa} \cdot \left(\ln U^C - \ln U^{US}\right) + \\ \overline{W}_K^{wa} \cdot \left(\ln K^C - \ln K^{US}\right) + \\ \overline{W}_L^{wa} \left(\ln L^C - \ln L^{US}\right) \end{bmatrix} \quad (2\text{-}5)$$

where, for example, the common weight for capital input across n countries is given by

$$\overline{W}_K^{wa} = \frac{\left[\prod_{i=1}^{n} W_K^n\right]^{1/n}}{\left[\prod_{i=1}^{n} W_U^n\right]^{1/n} + \left[\prod_{i=1}^{n} W_K^n\right]^{1/n} + \left[\prod_{i=1}^{n} W_L^n\right]^{1/n}} \quad (2\text{-}6)$$

The average labor weight is defined similarly to capital, with the weight for intermediate inputs calculated as residual $\overline{W}_U^{wa} = 1 - \overline{W}_K^{wa} - \overline{W}_L^{wa}$ (due to constant returns to scale, assumed in productivity indexes).

The Walsh index (2-5) passes all Fisher's tests for index numbers. The index is exact to the Cobb-Douglas production function with unitary elasticity of substitution among inputs, thus satisfying the Diewert's test (8) as well. However, since the Cobb-

Douglas function is not flexible, the Walsh index is not superlative.

When the number of inputs is limited to capital and labor, the Walsh index is identical to the most widely-used measure of relative TFP, which can derived from the Cobb-Douglas production function $Y = AK^{1-\alpha}L^{\alpha}$ after its weights $1-\alpha$ and α are substituted for $\overline{W}_K^{wa}$ and $\overline{W}_L^{wa}$ in (2-5):

$$\ln TFP_{CD}^{C,US} = \frac{\ln A^C}{\ln A^{US}} = \\ = \left[\ln Y^C - \alpha \ln K^C - (1-\alpha)\ln L^C\right] - \left[\ln Y^{US} - \alpha \ln K^{US} - (1-\alpha)\ln L^{US}\right] \qquad (2\text{-}7)$$

The only difference of the Cobb-Douglas TFP index (2-7) from the Walsh TFP index (2-5) is that the former does not specify how the common weight α is derived from country-specific weights, while the latter requires to use geometric mean of cross-country weights (2-6).

Despite the similarity between (2-5) and (2-7), the majority of multy-country TFP studies estimated productivity differences by (2-7), with no reference to the Walsh index (2-5). For example, Mahony (1993) estimated "relative joint factor productivity" (RJFP) in four countries at the total industry and total manufacturing levels. The RJFP index is equivalent to the Cobb-Douglas productivity index (2-7). However, since Mahony used factor weight that are "averages over the number of countries from OECD data" (p.114), her TFP index can be more correctly classified as a Walsh index with arithmetic average weights.

Similarly, Bernard and Jones (1996) calculated TFP levels with arithmetic averages of input weights over 14 OECD countries for 6 broad industrial sectors. Using a less precise approach, Pilat (1996) simply used "feasible" (but nevertheless arbitrary) weights 0.3 and 0.7 for capital and labor. However, this simplification allowed him to made TFP estimates at apparently the most disaggregated level so far (mostly 3-digit ISIC sectors, with several 4-digit ones).

The only index number that passes both Fisher's and Diewert's requirements of satisfactory index numbers is due to Caves and Christensen (1980) and Caves, Christensen and Diewert (1982). The index (to be referred as CCD) is closely related to the bilateral T-T in equation (2-3), with one additional advantage that the CCD is a transitive index.

To achieve transitivity of the T-T index, Caves, Christensen and Diewert (1982) suggested to make bilateral comparisons through a "hypothetical representative country" country, denoted as *H*. Instead of direct productivity comparison between *C* and *US* by (2-5), the CCD index first compares *C* with *H*, then – compares *US* with the same hypothetical base *H*. Finally, the productivity index between C and US is obtained as the ratio between *C-H* and *US-H* indexes. Inputs of the representative country *H* are equal to geometric average of inputs for all countries in the sample, and its input shares are equal to arithmetic average for all countries in the sample. Formally, the TFP comparison between *C* and *H* at the first stage, equals

$$TFP_{T\text{-}T}^{C,H} = \ln Y^C - \ln Y^H - \begin{bmatrix} \frac{1}{2}\left(W_U^C + \widetilde{W}_U^H\right)\cdot\left(\ln U^C - \ln \widetilde{U}^H\right) + \\ \frac{1}{2}\left(W_K^C + \widetilde{W}_K^H\right)\left(\ln K^C - \ln \widetilde{K}^H\right) + \\ \frac{1}{2}\left(W_L^C + \widetilde{W}_L^H\right)\left(\ln L^C - \ln \widetilde{L}^H\right) \end{bmatrix} \qquad (2\text{-}8)$$

where, for example, capital weigh in the hypothetical country H equals

$$\widetilde{W}_K^H = \frac{1}{n}\sum_{i=1}^{n} W_K^i \qquad (2\text{-}8a),$$

and its capital stock is

$$\widetilde{K}^{H} = \left[\prod_{i=1}^{n} K^{n}\right]^{\frac{1}{n}} \qquad (2\text{-}8b),$$

with average weights and levels of intermediate inputs and labor defined similarly.

Combining $TFP_{T-T}^{C,H}$ and $TFP_{T-T}^{US,H}$, the CCD index between C and US equals:

$$\ln TFP_{CCD}^{C,US} = \ln TFP_{T-T}^{C,H} - \ln TFP_{T-T}^{US,H} = \ln Y^{C} - \ln Y^{US} -$$

$$\begin{aligned}
&-\frac{1}{2}\left(W_{U}^{C} + \widetilde{W}_{U}^{H}\right)\left(\ln U^{C} - \ln\widetilde{U}^{H}\right) + \frac{1}{2}\left(W_{U}^{US} + \widetilde{W}_{U}^{H}\right)\left(\ln U^{US} - \ln\widetilde{U}^{H}\right) - \\
&-\frac{1}{2}\left(W_{K}^{C} + \widetilde{W}_{K}^{H}\right)\left(\ln K^{C} - \ln\widetilde{K}^{H}\right) + \frac{1}{2}\left(W_{K}^{US} + \widetilde{W}_{K}^{H}\right)\left(\ln K^{US} - \ln\widetilde{K}^{H}\right) - \qquad (2\text{-}9)\\
&-\frac{1}{2}\left(W_{L}^{C} + \widetilde{W}_{L}^{H}\right)\left(\ln L^{C} - \ln\widetilde{L}^{H}\right) + \frac{1}{2}\left(W_{L}^{US} + \widetilde{W}_{L}^{H}\right)\left(\ln L^{US} - \ln\widetilde{L}^{H}\right)
\end{aligned}$$

With the addition of common base country H, the CCD index (2-9) becomes transitive, making irrelevant which country is chosen as the basis for productivity comparisons. Also note that though CCD weights take into account economic conditions in all countries, included in the multilateral comparison, at the same time more than half of each weight is specific to compared countries *C* and *US*, as noted by Caves, Christensen and Thetheway (1981, p. 50).

The CCD index is related not only to the T-T productivity index (2-3), but also to the Cobb-Douglas index of TFP (2-7). To show that (2-7) is a special case of (2-9), consider a simple case of two inputs, such as capital and labor, disregarding terms that refer to intermediate materials *U* in (2-9).

Note that under unitary elasticity of substitution (assumed in the Cobb-Douglas production function) input shares must be the same across countries (that is, unitary), yielding cross-country identities $W_{K}^{C} = W_{K}^{US} = \widetilde{W}_{K}^{C} = 1-\alpha$ and $W_{L}^{C} = W_{L}^{US} = \widetilde{W}_{L}^{C} = \alpha$. Once α and $1-\alpha$ are substituted into (2-9), the CCD index simplifies to the Cobb-Douglas TFP index (2-7).

A recent application of the CCD index to international productivity studies was made in Harrigan (1996). Harrigan calculated TFP levels from a simplified version of (2-9), omitting intermediate materials from inputs. The study covered 9 two-digit ISIC manufacturing sectors in 12 OECD countries. Though Harrigan used multilateral PPP estimates, it is important to note that a single national PPP (for the total GNP) was applied to various manufacturing sectors, thus ignoring the diversity of relative prices among disaggregated manufacturing sectors. The variance may be quite substantial, as will be demonstrated in table 2-6, where the coefficient of variation for disaggregated PPPs exceeded 25% in Australia, Germany, France, the United Kingdom, reaching in the case of Japan as much as 42.3%.

Among recent studies of international TFP, there are a few studies that attempted to develop new productivity measures, but, eventually, ended up with conceptually inferior ways to measure TFP, once they are compared with the T-T or CCD indexes of TFP.

For example, Dollar and Wolff (1993) suggested the following alternative measure of TFP:

$$TFP^{C} = \frac{Y^{C}}{(1-\alpha^{C})K^{C} + \alpha^{C}L^{C}} \tag{2-10}$$

so that relative TFP equals

$$TFP_{DW}^{C,US} = \frac{Y^{C}}{Y^{US}} \bigg/ \frac{(1-\alpha^{C})K^{C} + \alpha^{C}L^{C}}{(1-\alpha^{US})K^{US} + \alpha^{US}L^{US}} \tag{2-11}$$

where α^{C} and α^{US} are "individual country averages of the ratio of wages to value added over the full period of this study" (p.199).

Though Dollar and Wolff claim that (2-10) is "the most intuitive formulation" of TFP, the measure has a few significant drawbacks. First, as Bernard and Jones (1996b) already noted, the measure is sensitive to change in units of input measurements, failing the commensurability test (4).

To show this, express labor inputs in larger units (i.e., in thousands instead of millions). Then (2-11) will be approaching index of labor productivity, eventually "blending" with the later if units of capital measurement become indefinitely small compared with units of labor measurement.

Second, inputs in (2-11) are aggregated, using country-specific weights, thus violating the "fixed weight" principle of index numbers that requires weighting nominator and denominator by *the same* set of weights.

Another unconventional productivity measure was suggested recently by Bernard and Jones (1996b). They went further than Dollar and Wolff in disregarding the "fixed weight" principle, claiming that joint productivity of capital and labor should vary not only with the "*A* term" of Cobb-Douglas function, but also with factor exponents, or input shares α and $1-\alpha$ (p. 1231).

Justifying the importance of α along with more conventional "*A* term", Bernard and Jones give quite an interesting explanation. "Suppose that two countries have exactly the same inputs ... as well as the same level of *A*, but they have different α's. Clearly, these two countries will produce different quantities of output. The problem with *A* in this case is that it is incomplete: the technology of production varies with the parameter as well as well as with the A's, and the simple Hicks-neutral measure does not take this into account" (ibid.). However, if one follows this approach in measuring, say, price inflation, then "complete" aggregate price index I_p

would equal to $I_p = \dfrac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0}$, and the index will be able to register price changes even if p_I remain the same, as soon as price weights q_I and q_I happen to be different.

Dissatisfied with available productivity measures, Bernard and Jones (1996) proposed an alternative, called "total technological productivity" (TTP). Fixing capital and labor across time and country sectors at K_0 and L_0, TTP for country C equals:

$$\ln TTP_{i,t}^{C} = \ln A_{i,t}^{C} + (1-\alpha_{i,t}^{C})\ln K_0 + \alpha_{i,t}^{C} L_0 \tag{2-12}$$

with $\ln A_{i,t}^{C}$ is defined as

$$\ln A_{i,t}^{C} = (1-\alpha_{i,t}^{C})\ln\left(\frac{Y_{i,t}^{C}}{K_{i,t}^{C}}\right) + {}_{i,t}^{C} \ln\left(\frac{Y_{i,t}^{C}}{L_{i,t}^{C}}\right) \tag{2-13}$$

Thus, the TTP measure varies both as "*A* term" (2-13) and with input shares, weighted by some fixed K_0 and L_0. Since Bernard and Jones assume Cobb-Douglas production function, their formula for TTP can be further simplified. Opening parenthesis and collecting terms in (2-13), we get:

$$\begin{aligned} \ln A_{i,t}^{C} &= \ln Y_{i,t}^{C} - \ln K_{i,t}^{C} - \alpha_{i,t}^{C}\ln Y_{i,t}^{C} + \alpha_{i,t}^{C}\ln K + \alpha_{i,t}^{C}\ln Y_{i,t}^{C} - \alpha_{i,t}^{C}\ln L_{i,t}^{C} = \\ &= \ln Y_{i,t}^{C} - (1-\alpha_{i,t}^{C})\ln K_{i,t}^{C} - \alpha_{i,t}^{C}\ln L_{i,t}^{C} \end{aligned} \tag{2-14}$$

Substitution of (2-14) into (2-12) yields

$$\ln TTP_{i,t}^{C} = \ln Y_{i,t}^{C} - (1-\alpha_{i,t}^{C})(\ln K_{i,t}^{C} - \ln K_0) - \alpha_{i,t}^{C}(\ln L_{i,t}^{C} - \ln L_0) \tag{2-15}$$

Then TTP of country C compared with base-country US becomes

$$\begin{aligned} TTP_{i,t}^{C,US} &= \ln Y_{i,t}^{C} - \ln Y_{i,t}^{US} - \\ &- (1-\alpha_{i,t}^{C})(\ln K_{i,t}^{C} - \ln K_0) + (1-\alpha_{i,t}^{US})(\ln K_{i,t}^{US} - \ln K_0) - \\ &- \alpha_{i,t}^{C}(\ln L_{i,t}^{C} - \ln L_0) + \alpha_{i,t}^{US}(\ln L_{i,t}^{US} - \ln L_0) \end{aligned} \tag{2-16}$$

Among various options for K_0 and L_0, Bernard and Jones suggested to use cross-country medians at the beginning of sample period, hereafter denoted as $\overline{K}_0$ and $\overline{L}_0$. Moreover, Bernard and Jones applied the TTP formula only once (at the beginning of sample period), and subsequently updated the benchmark estimate by the T-T index of productivity change[1]. Denoting capital and labor shares as ω_K and ω_L, and dropping subscripts for sector and time (now they are identical in every term of (2-16)), we get the final formula for TTP for the benchmark TTP estimate:

$$\begin{aligned} TTP^{C,US} &= \ln Y^C - \ln Y^{US} - \\ &- \omega_K^C(\ln K^C - \ln \overline{K}\ ) + \omega_K^{US}(\ln K^{US} - \ln \overline{K}\ ) - \\ &- \omega_L^C(\ln L^C - \ln \overline{L}\ ) + \omega_L^{US}(\ln L^{US} - \ln \overline{L}\ ) \end{aligned} \qquad (2\text{-}17)$$

Comparison of the final formulation of TTP (2-17) and the CCD index of TTP in (2-9) shows a surprising degree of similarity, once, to make the comparison more direct, intermediate materials are omitted from (2-9). In both indexes inputs of *C* and *US* are compared with some average inputs, geometric mean and median respectively, but difference is insignificant, since Bernard and Jones could have opted for geometric averages as well.

More importantly, the indexes differ in input weights, but the CCD weights seem preferable, as they correspond to country-specific and average inputs in the brackets[2]. However, in most cases the CCD weight $\frac{1}{2}\left(W^C + \widetilde{W}^H\right)$ and $\frac{1}{2}\left(W^{US} + \widetilde{W}^H\right)$ should not differ substantially from W^C and W^{US}, so that in general the CCD and TTP

---

[1] Using virtually the same index as (2) and (3), with time subscript *t* and *t-1* being substituted for country superscripts *C* and *US*.

[2] For example, W_K^C corresponds to K^C, and $\widetilde{W}_K^H$ corresponds to $\widetilde{K}^H$.

indexes should yield similar productivity estimates, contrary to the alleged novelty of the TTP measure.

To sum up, the CCD remains the only productivity measure which does not violate the fixed weight principle of index numbers, and, besides, satisfies all desirable properties of index numbers (1)-(9), making the CCD currently the best choice to estimate international differences in productivity.

However, due to the limitations of available PPPs for output (discussed in Chapter 1), it is not currently feasible to apply the CCD index for the estimation of transitive TFP levels. To see why it is so, consider again the basic formula for calculating relative TFP levels, such as (2-3). By applying the CCD index, we aggregate inputs into a transitive measure of multiple *inputs*, given by $\ln\left(I_{CCD}^{C} / I_{CCD}^{US}\right)$. On the other hand, the measure of relative output $\ln\left(Y^{C} / Y^{US}\right)$ is not transitive, because it is derived by using *intransitive* output PPPs from the ICOP.

As a result, although the application of the CCD index creates the transitivity of aggregate input indexes, this transitivity of relative input is essentially redundant for productivity comparisons, because it is not matched by the transitivity of converted output data. Consequently, relative TFP levels can be only defined in bilateral comparisons with the base country in the ICOP estimates of PPPs (most often –the United States), for which the bilateral T-T index (2-5) is a sufficient choice.

Section 2.3. Description of data sources and methods for international productivity comparisons.

Most of the data come from two OECD databases — the International Sectoral Database (ISDB) and Structural Analysis Database (STAN). The study covers nine developed countries: Australia, Canada, Germany, France, Japan, the Netherlands, Sweden, the United Kingdom, and the United States over 1975-1993. Sectoral productivity is measured in most manufacturing sectors at 3-digit ISIC level[3]. Specific details of data sources are given below.

2.3.1. Output.

The STAN database contains data for nominal gross output. To calculate real output at 1990 prices $Y_{i,t}^{1990}$, we used sectoral deflators from various issues of OECD's *Indicators of Industrial Activity*. Then intermediate input in constant prices was derived by subtracting real value added (available in STAN) from $Y_{i,t}^{1990}$.

The resulting measure of intermediate input is not satisfactory, since it is subject to double counting when sectoral output is counted twice, first as output, and then as input within the same sector (in so called intra-industry transactions). Using national input-output tables from the OECD Input-Output database (OECD, 1996), sectoral shares of intra-industry transactions were calculated for each country $k_i = U_{i,i} \Big/ \sum_i U_{i,j}$, where, for example, $U_{i,j}$ is intermediate input, produced in sector i and subsequently used in sector j. Then *net* real intermediate input was derived as

[3] However, due to persistent data problems, a few sectors were excluded from our database, such as tobacco (ISIC 314), petroleum refineries (ISIC 353), petroleum and coal products (ISIC 354), professional goods (ISIC 385), and other manufacturing (ISIC 390).

$U_i^{NET} = (1 - k_i)U_i$. Finally, the real net output was calculated as the sum of real U_i^{NET} and real value added.

The adjustment made output equal to the expenditures on goods and services, excluding transactions within the sector. The exclusion of intra-industry sales was originally suggested by Domar (1961), and was recently adopted by the US Bureau of Labor Statistics in its 3-input indexes of TFP, as discussed in Gullickson (1995).

For some countries, the OECD input-output database contains several input-output tables (Table 2-1). The database does not contain input-output data for Sweden, so, as a substitute, we used intra-sectoral shares from Danish input-output tables.

Table 2-1. The coverage of OECD Input-Output Database.

	Mid/late-1970s	early-1980s	mid-1980s	1990
Australia	1974		1986	1989
Canada	1976	1981	1986	1990
Denmark	1977	1980	1985	1990
France	1977	1980	1985	1990
Germany	1978		1986,1988	1990
Japan	1975	1980	1985	1990
Netherlands	1977	1981	1986	
United Kingdom	1979		1984	1990
United States	1977	1982	1985	1990

To convert net sectoral output to internationally comparable units, we used sectoral-specific PPPs from Pilat (1996). These currency converters came from a number of country-specific studies for different years, and were updated by Pilat to the same year (1987) by using price deflators for value added from the STAN database.

For a number of manufacturing sectors, Pilat provides PPPs at too much level of aggregation. For example, instead of a single PPP for ISIC 384 (transportation machinery), there are separate PPPs for shipbuilding (ISIC 3841), railroad equipment (ISIC 3842), motor vehicles (ISIC 3843), motorcycles (ISIC 3844), aircraft (ISIC 3845), and other transportation equipment (ISIC 3849). In cases when only disaggregated PPPs were available, we used the Walsh aggregation method, given by (1-7) and (1-8), using expenditure weights in national currencies. These aggregated PPPs are reported in Table 2-2, along with other manufacturing sectors for which we will estimate international differences in TFP.

Table 2-2. Purchasing Power Parities for Manufacturing Industries, Major OECD Economies, 1987 (US$=1).

ISIC	AUA	CAN	DEU	FRA	JPN	NET	SWE	UK
311 Food products	1.34	1.50	2.07	7.40	266	2.44	9.75	0.794
313 Beverages	1.22	1.78	2.38	8.74	221	2.19	9.69	0.586
321 Textiles	1.70	1.42	2.61	7.09	182	2.32	10.53	0.683
322 Clothing	1.58	1.41	2.91	10.17	179	2.53	9.72	0.691
323 Leather products	1.82	1.22	2.22	6.71	209	1.95	8.39	0.574
324 Footwear	1.30	1.23	2.81	6.71	209	1.95	8.39	0.574
331 Wood	2.07	1.41	2.69	6.48	471	2.78	10.15	0.918
332 Furniture	1.65	1.23	3.39	11.84	564	4.02	8.35	0.942
341 Paper products	1.80	1.35	2.26	7.46	188	2.29	7.16	1.044
342 Printing, publishing	1.34	1.65	4.24	9.73	248	5.08	11.95	0.645
351 Industrial chemicals	1.33	1.29	2.56	8.41	267	2.06	7.76	0.632
352 Chemical products	1.33	1.29	2.20	8.41	210	2.06	7.76	0.632
355 Rubber products	1.26	1.24	2.32	5.86	125	2.06	6.53	0.548
356 Plastic products	1.33	1.29	2.56	8.41	267	2.06	7.76	0.548
361 Pottery, china, etc.	1.48	1.32	1.99	5.71	189	1.85	8.58	0.648
362 Glass products	1.48	1.32	2.45	5.71	189	1.85	8.58	0.648
369 Mineral products	1.48	1.32	1.99	5.71	189	1.85	8.58	0.648
371 Iron & Steel	1.50	1.29	1.88	7.52	146	2.89	7.05	0.634
372 Non-ferrous metals	1.71	1.44	2.21	7.64	233	2.56	8.24	0.742
381 Metal products	1.57	1.31	2.28	8.68	140	2.84	7.13	0.667
382 General Machinery	1.08	1.40	2.43	7.13	149	3.31	9.06	0.609
383 Electrical Machinery	1.99	1.63	2.79	8.77	143	2.62	8.64	0.737
384 Transport equipment	1.34	1.32	2.22	9.57	145	2.74	9.30	0.762
3 Total manufacturing	1.44	1.39	2.31	7.54	176	2.48	8.51	0.690

Source: Pilat (1996), table A3. PPPs for ISIC 352, 382, 383, 384 are aggregated by the Walsh method, defined by equations (1-7) and (1-8).

Since net real output is expressed in 1990 constant prices, it was necessary to update Pilat's PPPs forward from 1987 to 1990. Using national price indexes from OECD's *Indicators of Industrial Activity*, sectoral PPP for country C were updated to 1990 by the following formula:

$$PPP_{1990} = PPP_{1987} * \left[\frac{P^{C}_{1990}}{P^{C}_{1987}} \bigg/ \frac{P^{US}_{1990}}{P^{US}_{1987}} \right] \qquad (2\text{-}18)$$

Some sectoral price indexes for gross output were missing in the *Indicators of*

Industrial Activity, and we used instead deflators for value added from the STAN database. The updated sectoral PPPs are listed in Table 2-3.

Table 2-3. Updated PPP for Manufacturing Industries, Major OECD Economies, 1990 (US$=1).

ISIC	AUA	CAN	DEU	FRA	JPN	NET	SWE	UK
311 Food products	1.30	1.42	2.01	6.92	249	2.18	10.51	0.799
313 Beverages	1.37	1.87	2.32	9.89	195	2.14	12.08	0.633
321 Textiles	1.87	1.40	2.54	6.97	171	2.22	11.49	0.739
322 Clothing	1.78	1.40	2.83	9.69	173	2.46	11.46	0.725
323 Leather products	1.45	1.08	1.81	5.97	183	1.67	8.19	0.523
324 Footwear	1.37	1.22	2.58	6.04	210	1.98	10.43	0.570
331 Wood	2.25	1.29	2.61	6.27	448	2.62	11.12	0.941
332 Furniture	1.80	1.13	3.29	11.45	536	3.80	9.15	0.966
341 Paper products	1.39	1.30	2.10	7.11	167	2.10	7.20	1.027
342 Printing, publishing	1.26	1.62	3.91	9.16	231	4.70	13.95	0.624
351 Industrial chemicals	1.02	1.21	2.41	7.55	229	1.62	7.76	0.601
352 Chemical products	1.14	1.21	2.07	7.55	180	1.62	7.76	0.601
355 Rubber products	1.23	1.31	2.47	5.33	121	1.70	7.76	0.528
356 Plastic products	1.19	1.57	2.57	8.84	270	2.17	7.79	0.593
361 Pottery, china, etc.	1.16	0.85	2.19	6.26	180	1.81	12.06	0.785
362 Glass products	1.54	1.38	2.57	6.08	189	1.85	10.51	0.737
369 Mineral products	1.62	1.59	2.22	6.48	199	2.12	10.85	0.777
371 Iron & Steel	1.31	1.22	1.78	6.92	140	2.82	7.68	0.627
372 Non-ferrous metals	2.16	1.22	2.14	7.52	221	2.41	8.26	0.721
381 Metal products	1.69	1.23	2.13	8.37	131	2.68	7.61	0.684
382 General Machinery	1.28	1.57	2.64	7.66	158	3.54	10.70	0.749
383 Electrical Machinery	2.40	1.66	2.74	8.60	124	2.50	6.58	0.787
384 Transport equipment	1.48	1.19	2.20	10.09	128	2.77	9.98	0.867
3 Total manufacturing	1.55	1.31	2.19	7.16	161	2.32	9.02	0.715

Note: PPPs were updated according to (2-18).

After Pilat's PPPs were updated to 1990, they were applied to net real output data to get output time series in internationally comparable units.

2.3.2. Intermediate input.

The preceding section has already described the derivation of net (sectoral) intermediate input in national currency. A much more demanding task was to obtain estimates of PPPs for intermediate input. Such currency converters are not available either from the ICP or ICOP. Jorgenson and Kuroda (1990) remains the sole study that addressed the problem in their Japan-US productivity comparison. The lack of PPP estimates for intermediate inputs is due to the requirement to use integrated data from national accounts and input-output table, which is a highly demanding task. Fortunately, the STAN database and OECD input-output tables (as well as Pilat's PPPs for final output) have been developed by following internationally-comparable principles of data compilation, making them a suitable choice for the calculation of PPPs for intermediate inputs.

Following Jorgenson and Kuroda (1990, p. 33-34), sectoral PPPs for U_i were calculated by aggregating sectoral PPPs for output, using shares of intermediate input deliveries to sector i from j[4]:

$$SU_{ij} = \frac{U_{ij}}{\sum_i U_{ij}} \tag{2-19}$$

To get PPP for intermediate input in sector i, its PPP for output and similar output PPPs in other sectors were weighted and aggregated by the Walsh aggregation method, defined by (1-7) and (1-8).

Since part of intermediate input for manufacturing sectors comes from non-manufacturing sectors (most importantly – services), the approach requires PPPs not only for manufacturing sectors from table (2-3), but for the rest of economy as well.

Input-output tables in the OECD database allow calculating input value shares for major non-manufacturing sectors. Besides, results of the ICP benchmark studies contain a number of relevant PPP estimates for services, but generally they refer to the basic heading level and need to be aggregated. Table 2-4 lists these original

---

[4] In the extreme case, when no intermediate input came from the other sectors, the PPP for M_i would coincide with the PPP for output. Otherwise, the PPP for M_i depends on the composition of intermediate input flows across the economy.

subcategories along with respective ISIC sectors from input-output tables from the OECD's input-output database.

After 'mapping' basic heading PPPs from the ICP to relevant service sectors in available I-O tables, these PPPs were aggregated by the Walsh method. The aggregated PPPs still refer to the expenditure side. To be compatible with PPPs from the production approach (for manufacturing sectors), these PPPs from the ICP need to be 'peeled out' of distribution and transportation margins.

Input-output tables from OECD are expressed in producer prices (with distribution and transportation margins assigned to trade and transportation), so that the 'peeling off' procedure can be easily implemented. Relevant formulas can be found in Pilat (1996, p. 6).

Original basic heading PPPs came from ICP's 1990-benchmark study. It was not possible to find a match in ICP data for 'wholesale and retail trade', so we used international ratios of transportation margins (with the USA as the base country) from OECD's I-O tables. Final aggregated PPPs for service sectors are reported in table 2-5.

Table 2-4. Concordance between sectors in OECD I-O tables and ICP basic headings.

ISIC	*I-O sector*	*ICP basic heading*		
4	Electricity, gas, water		4-1.	Electricity
			4-2.	Gas
5	Construction		5-1.	Family dwellings
			5-2.	Multifamily dwellings
			5-3.	Agricultural buildings
			5-4.	Industrial buildings
			5-5.	Buildings for market services
			5-6.	Buildings for non-market services
			5-7.	Routes, roads, bridges, tunnels
			5-8.	Other transport utility
			5-9.	Other civil engineering
63	Restaurants, cafe		63-1.	Workers' cafeterias
			63-2.	Restaurants, cafes, etc.
			63-3.	Hotels, lodgings
71	Transport and storage		71-1.	Local taxis
			71-2.	Local buses, trams & the like
			71-3.	Railway transport
			71-4.	Air transport
			71-5.	Other long distance transport
72	Communication		72-1.	Postal communication
			72-2.	Telephone, telegraph
81+ 82	Financial services	8-4	Financial services (bank, etc.)	
83	Real estate		83-1.	Rents of tenants
			83-2.	Imputed rents of owner-occupiers
9	Private and public services		9-1.	Clothing, rental and repair
			9-2.	Repairs to footwear
			9-3.	Repair & maintenance of houses
			9-4.	Sanitary services & water charges
			9-5.	Repairs to furniture, fixture, etc.
			9-6.	Repairs to household textiles etc.
			9-7.	Repairs to major household appliances
			9-8.	Repairs to glassware, tableware, etc.
			9-9.	Domestic services
			9-10.	Household services
			9-11.	Services of gen. practitioners
			9-12.	Services of specialists
			9-13.	Services of dentists
			9-14.	Services of nurses
			9-15.	Other medical services
			9-16.	Medical personnel
			9-17.	Non-medical personnel
			9-18.	Repair of transportation equip.
			9-19.	Repair to equipment & accessories
			9-20.	Cinema, theatre, sports ground, etc.
			9-21.	TV & radio license; hire of equip.
			9-22.	Others: religious, cultural, etc.
			9-23.	Total education expenditures
			9-24.	Barber and beauty shops
			9-25.	Services, nes
			9-26.	Welfare services
			9-27.	Compensation of employees
			9-28.	Commodities, goods & services.

Note: codes for ICP basic heading are taken from Kravis et. al. (1982), pp. 60-66.

Table 2-5. PPP estimates for service sectors for major OECD countries in 1990.

ISIC	AUA	CAN	DEU	FRA	JPN	NET	SWE	UK
4 Electricity, gas & water	1.41	1.15	3.91	10.43	290	3.79	9.42	1.046
5 Construction	1.28	1.06	2.61	6.45	220	3.00	11.18	0.933
61/2 Wholesale & retail trade	1.42	1.47	2.36	6.39	209	2.21	9.86	0.576
63 Restaurants & hotels	2.00	1.66	2.46	9.70	352	2.93	14.74	0.874
71 Transport & storage	1.10	1.11	1.95	6.18	129	2.42	7.00	0.604
72 Communication	1.02	1.34	2.52	5.35	183	1.45	5.37	0.737
81/2 Finance & insurance	0.92	1.14	1.85	5.40	101	1.68	7.45	0.370
83 Real estate & business services	1.42	1.41	1.85	4.39	191	1.94	7.04	0.356
9 Private and public services	1.14	0.67	1.68	5.55	126	1.79	8.52	0.488

Note: PPPs were aggregated from original basic heading data from the ICP, using concordance in table 2-4 and Walsh aggregation procedure, given by (1-7) and (1-8).

The last problem to be solved in the calculation of PPPs for intermediate input was incompatibility between calculated output PPPs for 3-digit manufacturing sectors (table 2-3) and manufacturing sectors in OECD input-output tables, which were in some cases only 2-digit sectors. After aggregating incomparable PPPs for output by the Walsh method, PPPs for intermediate input *PPPU* were obtained by weighing sectoral PPPs for output (*PPPY*) with shares of intermediate input deliveries from corresponding sectors (*SU*), defined by (2-19).

$$\ln PPPU_i^{C-USA} = \sum_j \overline{SU}_{ij} \times \ln PPPY_j^{C-USA} \tag{2-20}$$

with bar over *SU* denoting the geometric average nine countries in the study.

Final PPPs for intermediate inputs, along with corresponding PPPs for final output, are reported in table 2-6. To check the plausibility of these PPPs for intermediate input, we compared variation of PPPY and PPPU in specific countries. Each country shows a substantial variation in relative prices of output, with coefficient of variation ranging from 17.1% in Canada to 44.2% in Japan (see the last row of table 2-6). Given such variation, we hypothesized that sectoral output with relatively high prices will be purchased as intermediate input in lower proportions than cheaper intermediate inputs. As a result, relatively expensive intermediate input will correspond to smaller weights SU in the calculation of PPPU by (2-20), producing less volatile estimates of PPP for intermediate inputs (compared with corresponding PPP for final output). In fact, this regularity occurred in every country in the study, with the coefficient of variation for PPPU in almost all lower than 10% (except Japan, where it is 10.1%, still 4 times lower than the corresponding variation in PPPY).

2.3.3. Capital stock data.

Official capital stock estimates are available in the ISDB database, but only for 2-digit ISIC manufacturing sectors. On the other hand, the STAN database contains investment time series at 3-digit ISIC level, but no capital stock data.

Capital stock was calculated according to the perpetual inventory method, as outlined, for example, Jorgensen and Nishimizu (1978). Capital stock KS_t^C for country C at time T equals to the sum of past investment flows $I_{t-\tau}^C$, weighted by their relative efficiency level d_τ^C:

$$KS_t^C = \sum_{\tau=0} d_\tau^C I_{t-\tau}^C \qquad (2\text{-}21)$$

The relative efficiency of capital is assumed to decline geometrically with the age of capital stock:

$$d_\tau^C = (1-\delta)^\tau \qquad (2\text{-}22)$$

where δ is the rate of depreciation of capital stock.

Noting that for $\tau = 0$ (the present period) $d_\tau^C = 1$, KS_t^C in (2-21) can be expressed as

$$KS_t^C = I_t^C + \delta KS_{t-1}^C \qquad (2\text{-}23)$$

Table 2-6. PPP estimates for intermediate inputs in major OECD countries, 1990.

ISIC	Sector	AUA		CAN		DEU		FRA		JPN		NET		SWE		UK	
		PPPU	PPPY	PPPU	PPPY	PPPU	PPPY	PPPU	PPPY	PPPU	PPPY	PPPU	PPPY	PPPU	PPPY	PPPU	PPPY
1	Agriculture, forestry & fishing	1.37	1.38	1.26	1.44	2.21	2.51	6.77	7.37	183	194	2.32	2.31	9.09	10.13	0.612	0.627
2	Mining & quarrying	1.40	1.47	1.13	1.43	2.65	4.29	6.19	6.52	162	164	2.24	2.63	8.17	10.41	0.595	0.621
31	Food, beverages & tobacco	1.37	1.34	1.32	1.42	2.22	1.84	6.86	7.08	194	220	2.38	2.51	9.39	9.63	0.619	0.721
32	Textiles, apparel & leather	1.45	1.67	1.34	1.49	2.44	2.78	7.02	8.17	187	186	2.20	2.35	9.30	11.23	0.625	0.763
33	Wood products & furniture	1.53	2.08	1.26	1.24	2.42	2.89	6.97	8.01	229	481	2.39	3.10	9.07	10.59	0.663	0.951
34	Paper products & printing	1.39	1.51	1.26	1.43	2.46	2.87	6.96	8.35	192	195	2.67	3.49	8.75	9.64	0.601	0.793
351/2	Industrial chemicals	1.28	1.05	1.19	1.22	2.21	2.06	6.35	6.79	185	199	2.07	1.54	7.70	6.46	0.557	0.538
353/4	Petroleum & coal products	1.49	3.96	1.33	1.11	2.68	1.54	6.02	4.74	151	99	2.34	2.03	9.54	7.15	0.608	0.638
355/6	Rubber & plastic products	1.25	1.20	1.21	1.47	2.25	2.55	6.38	7.59	195	221	2.08	2.06	7.80	7.83	0.572	0.574
36	Non-metallic mineral products	1.44	1.80	1.20	1.35	2.33	2.13	6.42	5.80	175	188	2.27	1.87	8.80	10.24	0.627	0.736
371	Iron & steel	1.45	1.31	1.18	1.22	2.16	1.78	6.87	6.92	160	140	2.53	2.82	8.62	7.68	0.630	0.627
372	Non-ferrous metals	1.61	2.16	1.25	1.22	2.34	2.14	7.22	7.52	195	221	2.41	2.41	8.63	8.26	0.643	0.721
381	Metal products	1.47	1.69	1.21	1.23	2.13	2.13	6.91	8.37	166	131	2.49	2.68	8.16	7.61	0.604	0.684
382	Non-electrical machinery	1.43	1.28	1.29	1.57	2.30	2.64	6.62	7.66	161	158	2.54	3.54	8.44	10.70	0.619	0.749
383	Electrical apparatus, nes	1.62	2.40	1.33	1.66	2.33	2.74	6.85	8.60	158	124	2.32	2.50	7.83	6.58	0.612	0.787
384	Transportation equipment	1.45	1.48	1.23	1.19	2.24	2.20	6.83	7.07	149	128	2.56	2.77	8.65	9.98	0.657	0.867
385	Professional goods	1.43	1.53	1.24	1.00	2.28	2.51	6.55	11.59	162	115	2.40	3.25	8.04	8.66	0.571	0.387
39	Other manufacturing	1.44	1.71	1.26	1.46	2.26	3.08	6.65	11.68	192	234	2.39	3.28	8.85	13.45	0.600	0.833
4	Electricity, gas & water	1.48	1.41	1.13	1.15	2.86	3.91	6.60	10.43	168	290	2.60	3.79	9.12	9.42	0.752	1.046
5	Construction	1.57	1.28	1.28	1.06	2.22	2.61	6.14	6.45	182	220	2.42	3.00	8.33	11.18	0.659	0.933
61/62	Wholesale & retail trade	1.37	1.42	1.16	1.47	2.10	2.36	5.78	6.39	176	209	2.21	2.21	7.90	9.86	0.528	0.576
63	Restaurants & hotels	1.34	2.00	1.26	1.66	2.06	2.46	6.37	9.70	195	352	2.33	2.93	8.91	14.74	0.593	0.874
71	Transport & storage	1.44	1.10	1.14	1.11	2.05	1.95	5.83	6.18	151	129	2.21	2.42	7.73	7.00	0.564	0.604
72	Communication	1.33	1.02	1.23	1.34	2.21	2.52	5.91	5.35	173	183	2.21	1.45	8.45	5.37	0.570	0.737
81/2	Finance & insurance	1.19	0.92	1.15	1.14	1.98	1.85	5.33	5.40	150	101	1.88	1.68	7.49	7.45	0.450	0.370
83	Real estate & business serv.	1.31	1.42	1.16	1.41	2.08	1.85	5.53	4.39	172	191	2.17	1.94	8.40	7.04	0.488	0.356
9	Services	1.33	1.14	1.19	0.67	2.18	1.68	6.12	5.55	181	126	2.18	1.79	8.28	8.52	0.545	0.488
	Coefficient of variation	7.2	37.6	5.1	16.8	8.8	25.8	7.6	25.2	10.6	42.3	7.7	24.9	6.6	23.5	9.7	25.4

Note: PPP for intermediate inputs are calculated by formula (2-20), using PPP estimates for final output from tables 2-3 and 2-5. Weights for sectoral deliveries of intermediate inputs are taken from OECD Input-Output database, using input-output tables, listed in the last column of table 2-1.

Data on $I_{t-\tau}^{C}$ are available from the STAN database. The only problem is how to estimate the benchmark capital stock KS_{t-1}^{C} and depreciation rate δ .

To choose δ , we assumed that there is a uniform asset life of 15 years. The corresponding rate of depreciation δ , commensurate with the double declining balance method, is 0.133.

The calculation of capital stock proceeded as follows. First, we used capital stock estimates from the ISDB database for 2-digit SIC manufacturing sectors, and calculated benchmark capital stocks in 3-digit sectors in 1970, assuming the same capital intensity among 3-digit sectors. For example, capital stock in paper products (SIC341) was derived from available capital stock for SIC34 (paper products *and* publishing) $KS_{341} = \frac{KS_{34}}{LI_{34}} LI_{341}$.

Then capital stock in 1971 was calculated by formula (2-23), using the benchmark estimate for 1970, and sectoral investment flow for 1971, taken from the STAN database.

The original benchmark estimate for 1970 is arguably quite rough due to the assumption of the same capital intensity within 2-digit ISIC industries. However, the extents of potential measurement error of capital stock becomes lower and lower as capital estimates move away from the original benchmark estimate in 1970. Finally, relatively imprecise capital estimates for 1970-1974 were discarded, so that capital data, actually used in the study, start since 1975.

The resulting estimates of capital stock are expressed in national currency (at 1990 prices), so it was necessary to convert them to a common currency by PPPs for gross fixed capital formation, taken from the ICP benchmark study in 1990.

2.3.4. Labor input data.

In the majority of productivity studies, labor input is given as sectoral employment or as the product of sectoral employment *EMPL* and hours worked *HW*, producing $LI_i = EMPL_i \times HW_i$. However, the conventional approach still does account for international differences in the value of labor input (in contrast to capital stock).

To account for the international differences in the value of labor, we used the same approach as in Jorgenson and Kuroda (1990), and compared national wages per hour (say, in country *C*) with ones in the base country (the United States). Then we normalized the relative wage rate by relative domestic prices between *C* and the United States $P^C_{CONS} / P^{US}_{CONS}$, or, more exactly, the purchasing power parity for total consumption expenditures:

$$PPP^L_{C-USA,i} = \frac{WH_i^C}{WH_i^{USA}} \Big/ \frac{P^C}{P^{USA}} = \frac{WH_i^C}{WH_i^{USA}} \Big/ PPP^{C,US}_{CONS}$$

As a result, the updated measure of labor input becomes

$$LI_i = EMPL_i \times HW_i \times \frac{WH_i^C}{WH_i^{USA}} \Big/ PPP^{C,US}_{CONS}$$

The most readily available is sectoral employment that comes from the STAN database. Sectoral hours worked are available in *Labor Force Statistics,* published by the International Labor Organization. In cases, when the source did not contain data at 3-digit level, we used data from relevant 3-digit or 2-digit manufacturing sectors. Similarly to hours worked, hourly wage rates came from the *Labor Force Statistics*. PPPs for consumption were taken from the ICP benchmark study in 1990.

2.3.5. Factor share data.

Conventionally, factor input shares were calculated as shares of intermediate materials (total output minus value added), operating surplus and wages in sectoral total output. All data came from the same source (the STAN database).

Original factor shares displayed high volatility, mostly due to cyclical variation in profits (with sectoral profits turning in some cases negative). To remove such short-term variation in input shares, we used the Hodrick-Prescott filter, with λ set to 100. Nevertheless, even after extracting trend from the original data, there remained a number of cases when labor shares exceeded 100% (due to negative profits). In these cases we used average labor and capital shares, calculated over 1975-1993.

In addition to factor shares, cyclical factors may affect the variation of input and output levels, as noted by Harrigan (1996, p. 12), producing TFP estimates that may reflect international differences in busyness cycles rather than in productivity *per se*. Moreover, even if business cycles were synchronized across analyzed countries, differences in labor hoarding or in the ability to attract extra capital *during the same phase of business cycle* would affect our productivity estimates. To eliminate the impact of cyclical factors, we applied the Hodrick-Prescott filter to real net output and inputs as well. Hence, our TFP estimates mostly refer to secular levels of productivity that, unlike conventional productivity measures, are less affected by transitory variations in outputs, inputs and factor shares.

Eventually, the compiled database covered 23 manufacturing sectors and total manufacturing in 1975-1993 in 9 OECD countries. The next section reports results of TFP calculations from the compiled database.

Section 2.4. Comparison of international productivity measurements by alternative index numbers.

The index number approach allows making 'point' estimates of TFP, producing substantial number of bilateral TFP index comparisons with the United States as the base country. For a single measure of TFP, there are 8 country-pairs with the United States, 19 annual observations for 24 manufacturing sectors, producing in total 3,648 bilateral indexes.

Table 2-7 reports international differences in TFP that take into account three factor inputs (intermediate materials, capital, and labor). These TFP differences were calculated by bilateral T-T TFP indexes (2-2) and (2-3). Table 2-7 reports only a small part of calculated 3-input TFP indexes, which refer to TFP estimations in 1975 and 1990, so it is possible to compare TFP levels both across countries in these two years, and also the change in productivity leadership over 1975-1990.

First we consider estimated TFP levels for the total manufacturing. The United States remained productivity leader both in 1975 and 1990. However, since 1975 several countries have managed to catch up with the American productivity levels. The catching up was especially pronounced in Japan and the United Kingdom, where the gap with the United States decreased from 72 to 84%, and 58 to 70%, respectively. Catching up in productivity in the Netherlands and France was slightly lower, increasing from 74 to 84% and 66 to 73%, respectively. Another interesting result is that Canada consistently keeps the second place in relative productivity, falling short of the American level by around 7%.

Estimated productivity levels at lower levels of aggregation are especially interesting. On the whole, table 2-7 shows a high degree of similarity in TFP levels, with only a small number of substantial gaps in productivity. The gap with the United States is especially pronounced in wood products (ISIC 331), where Australia, Japan and the United Kingdom had 2 or 3-fold gaps with the United States.

As for the productivity performance of separate countries, it was also quite dissimilar. For example, Australia consistently excelled in chemical industries. In contrast, it had relatively low productivity in machinery sectors.

Canada had exceptionally high productivity in non-metal products and furniture, and came close to the American level in other manufacturing sectors (except beverages, footwear, and rubber products, where it trailed the United States by around 30%).

Germany had unusually low productivity estimate in printing and publishing. Over time, it managed to substantially catch up with the United States in iron and steel and non-ferrous metals.

France also managed to catch up with the United States in these two manufacturing sectors. It especially excelled in leather products, but in the rest of manufacturing sectors it trailed a lot behind the United States, especially in machinery sectors.

In contrast, Japan demonstrated high productivity performance in machinery, surpassing the United States as the productivity leader in electrical machinery and sharing the leadership with the United States and Canada in transportation machinery. In general machinery, the gap with the United States noticeably decreased from 77% in 1975 to 92% in 1990. On the other hand, Japan had very low relative productivity in wood products and furniture. One quite unexpected result is the high Japanese productivity in footwear and leather products.

The Netherlands showed good productivity performance in a large number of manufacturing sectors, especially in 1990, including leather products and several chemical sectors, such as industrial chemicals and rubber products.

Sweden approached the American productivity in few manufacturing sectors, such as leather products, paper products, plastic products and electrical machinery. In the rest of manufacturing sectors, Swedish productivity rarely exceeded 80% of American level.

A distinct feature of productivity estimates in the United Kingdom is the large number of drastic changes in relative productivity. The United Kingdom substantially improved its relative standing in printing and publishing, glass products, iron and steel, non-ferrous metals, general machinery and transportation equipment. On the other hand, there were large drops in relative productivity in such sectors as furniture and rubber products.

Though these TFP estimates, due to their account for intermediate materials, are more comprehensive measures of productivity compared with other related estimates of international TFP levels, they at the same time require considerably more efforts in data preparation, compared with more traditional productivity measures which include fewer inputs. Previously, the lack of three-input TFP estimates did not allow addressing the question whether the omission of some inputs in productivity studies led to a significant bias in productivity estimates with limited number of inputs. One may argue that if there is almost no difference between TFP estimates with three and two inputs (such as capital and labor), then there is little point in the extension of the number of inputs beyond capital and labor.

Table 2-7. Relative TFP levels for 3-input case, major OECD countries (US TFP = 100).

ISIC Manufacturing Sector	AUA		CAN		DEU		FRA		JPN		NET		SWE		UK	
	1975	1990	1975	1990	1975	1990	1975	1990	1975	1990	1975	1990	1975	1990	1975	1990
311 Food products	96	95	102	96	99	92	86	80	86	66	86	88	81	76	68	67
313 Beverages	95	92	95	75	78	71	88	72	107	84	116	105	70	64	84	87
321 Textiles	76	73	100	98	82	81	86	82	79	66	101	102	76	66	69	66
322 Clothing	88	75	107	97	82	70	61	57	123	95	78	72	96	66	63	63
323 Leather products	101	86	118	105	96	98	118	107	133	119	102	124	115	92	101	94
324 Footwear	110	97	56	69	76	72	87	81	112	109	132	97	78	63	93	90
331 Wood	47	51	87	94	70	71	71	76	28	31	75	75	63	69	50	49
332 Furniture	95	71	131	120	82	69	48	46	47	42	57	57	89	84	69	55
341 Paper products	60	85	97	93	82	90	65	75	77	84	74	101	77	92	40	43
342 Printing, publishing	48	90	67	85	34	44	49	57	51	66	33	43	34	44	65	87
351 Industrial chemicals	96	111	81	87	79	72	59	67	68	66	96	110	81	79	72	82
352 Chemical products, nes	78	90	83	94	71	70	45	58	97	81	74	84	73	65	89	79
355 Rubber products	107	108	77	73	92	81	90	83	150	136	84	117	88	87	89	96
356 Plastic products	151	103	103	86	82	80	69	64	75	64	107	102	101	95	123	89
361 Pottery, china, etc.	75	100	176	184	68	76	61	74	60	73	115	114	58	63	54	55
362 Glass products	69	85	91	92	60	75	69	92	88	84	112	91	55	72	49	66
369 Non-metallic prod., nes.	69	82	102	88	97	96	93	98	68	79	95	100	69	70	75	76
371 Iron & Steel	63	91	85	92	76	93	51	70	100	115	67	76	60	86	45	68
372 Non-ferrous metals	75	69	81	102	70	89	63	74	68	78	83	101	65	88	47	64
381 Metal products	72	69	100	97	82	84	60	61	87	102	67	72	82	82	66	67
382 General Machinery	84	86	78	80	81	70	76	70	77	92	63	55	69	59	67	55
383 Electrical Machinery	63	58	92	91	76	79	59	67	60	118	79	87	74	97	55	61
384 Transport equipment	71	78	99	99	84	86	42	50	96	100	62	69	60	63	47	57
300 Total manufacturing	70	77	93	94	82	83	66	73	72	84	74	84	67	73	58	70

Note: relative productivity levels were calculated by the bilateral version of T-T productivity index, given by (2-2) and (2-3).

To address the question, we compared country rankings according to the 3-input TFP index (hereby denoted as 3TFP) with country rankings, when TFP was measured by productivity indexes that accounted for capital and labor (2TFP) and for labor (LP).

To compare country rankings according to these alternative measures of productivity, we used the Spearman rank correlation. Specifically, after estimating simpler productivity measures 2TFP and LP, we calculated the Spearman rank correlation between 3TFP and 2TFP, and then – between 3TFP and LP. For 24 sectors and 19 years, there were $24 \times 19 = 456$ correlation coefficients each pair of indexes, again making necessary a summarized presentation of estimated correlation coefficients.

Table 2-8 reports the share of cases when there was a significant correlation between 3TFP and its alternatives 2TFP and LP over 1975-1993. Quite surprisingly, there is a close correspondence between county rankings according to 3TFP and its simpler alternatives, indicating that the inclusion of intermediate materials may not lead to large biases in relative productivity estimates. The correspondence is especially close with 2TFP, producing statistically significant correlation in all, but two sectors (food products and non-ferrous metals). Table 2-8 also reports that the Spearman rank correlation was statistically significant in 94.7% of the total in food products. Since the total number of cross-sections over 1975-1993 was 19, it means that there was statistically close correspondence for 17 annual cross-sections.

Table 2-8. The share of correlated comparisons of alternative measures of productivity.

ISIC	Manufacturing sector	3TFP vs. 2TFP	3TFP vs. LP
311	Food products	94.7	100.0
313	Beverages	100.0	100.0
321	Textiles	100.0	100.0
322	Clothing	100.0	100.0
323	Leather products	100.0	100.0
324	Footwear	100.0	100.0
331	Wood	100.0	100.0
332	Furniture	100.0	100.0
341	Paper products	100.0	100.0
342	Printing, publishing	100.0	100.0
351	Industrial chemicals	100.0	31.6
352	Chemical products, nes	100.0	84.2
355	Rubber products	100.0	100.0
356	Plastic products	100.0	100.0
361	Pottery, china, etc.	100.0	100.0
362	Glass products	100.0	100.0
369	Non-metallic prod., nes.	100.0	68.4
371	Iron & Steel	100.0	100.0
372	Non-ferrous metals	89.5	84.2
381	Metal products	100.0	100.0
382	General Machinery	100.0	78.9
383	Electrical Machinery	100.0	100.0
384	Transport equipment	100.0	100.0
3	Total manufacturing	100.0	100.0

As can be expected, the omission of one more input (capital) produces poorer correspondence between 3TFP and LP. There is especially pronounced conflict in industrial chemicals (ISIC 351) and non-ferrous metals (ISIC 372), where similar results were obtained in 31.6 and 68.4% of all considered cross-sections.

Still, there are a large number of cases when there was no conflict between 3TFP and LP (to be exact, in as many as 18 manufacturing sectors), and the average number of significantly correlated rankings across all 24 sectors was 90.4%. Consequently, simple productivity measures, including even labor productivity, appear be a good approximation for productivity measures with larger number of inputs, though the degree to correspondence varies substantially among manufacturing sectors.

Section 2.5. Conclusions.

In this chapter we presented results of international comparison of sectoral productivity that take into account intermediate inputs, capital and labor. After reviewing available approaches to estimate productivity levels in section 2.1, we opted for the index number approach. The choice was primarily due to the relatively small number of countries in our database, which precluded the use of conventional econometric or linear programming methods, such as the stochastic frontier approach or the DEA. In section 2.2 we introduced standard productivity indexes along with less familiar Walsh index which combines the ease of calculation with the important exactness property to the Cobb-Douglas production function.

In section 2.3 we deal with the major stumbling block in estimating productivity levels – the compilation of database with specific PPPs for output and three inputs in 3-digit manufacturing sectors. To derive purchasing power parities for intermediate materials, one has to merge national account data with input-output tables. Currently, this task has been solved only for US-Japan comparison, while we extended the coverage of countries to nine major OECD countries. Another distinctive feature of this study is the removal of cyclical movements from the original time series, so that the "residual" productivity estimates are obtained from secular levels of output, inputs, and factor shares.

Applying a three-input index of TFP, we found that the United States remained the productivity leader for the total manufacturing, though in some cases the productivity gap has decreased since 1975, especially with Japan and the United Kingdom, indicating the high extend of productivity convergence. There was also a considerable variation in sectoral productivity levels, where the productivity advantage often belonged to countries other than the United States.

Since there are very few studies that attempted to estimate productivity levels by comprehensive productivity measures with multiple inputs, in section 2.4 we compared our productivity estimates by the three-input TFP index with much simpler, but potentially biased measures of two-input TFP and labor productivity. Surprisingly, there were very few cases when there was substantial incompatibility in country rankings according to the primary three-input TFP index and alternative, much simpler productivity measures. Consequently, the current practice of generally avoiding complex productivity measure may not lead to substantial distortions in reported productivity results[5].

---

[5] However, this optimistic conclusion will be reversed in the next chapter, where we will show that there may be substantial inconsistencies in reported results in studying productivity convergence if inappropriate productivity measures are used. Therefore, the high degree of consistency among productivity measures, which we identified in this chapter, should not be considered as unqualified recommendation to avoid comprehensive productivity measures with multiple inputs.

Charter 3. Productivity convergence for specific manufacturing sectors and the total manufacturing in major OECD countries.

In this chapter we will apply estimates of TFP from chapter 2 to examine productivity convergence over 1975-1993 for nine industrialized countries. In section 3-1 we briefly introduce the concept of σ-convergence. The application of this convergence concept has relied primarily on the visual inspection of cross-sectional variance of productivity, without a statistical test to ascertain the significance of observed trends in relative productivity levels. To fill in the gap in the literature, we introduce in section 3-2 a simple test of σ-convergence/divergence, which is well known in the statistical literature but appears to be little used in economic studies. Section 3-3 contains major results of applying the suggested test for productivity convergence/divergence.

Section 3.1. Previous studies on convergence in productivity.

The vast majority of previous studies of productivity convergence relied on labor productivity, which was usually estimated only for the whole economy, as in Baumol, 1986) and Mankiw *et. al* (1992). As argued by Bernard and Jones (1996, p. 1218), the use of labor productivity may seriously distorts the reported results, primary because labor productivity fails to identify the separate impact of capital accumulation on output. This is not desirable due to the important role of capital in the neoclassical growth theories on which the convergence concept is based.

After noting the conceptual limitations of labor productivity, Bernard and Jones (1996) analyzed international productivity convergence, using estimates of total factor productivity for 14 industrial countries over the period 1970-1987. The study also dealt with more disaggregated data compared with previous studies, analyzing six major industrial sectors (agriculture, mining, manufacturing, electricity, construction and services).

However, in addition to a number of shortcomings in constructing productivity indexes (such as their TTP index, already discussed in section 2-2), this study used hardly adequate methods in compiling the original data on output and inputs. First, similarly to Harrigan (1996), Bernd and Jones applied the same purchasing power for the total GNP to output conversions in disaggregated industrial sectors, thus creating a potential bias in output measurement if relative prices across industrial sectors were not the same. Second, the study calculated labor input by the total employment, without taking into account the varying degree of hours worked. Besides, there was no adjustment for the varying value of labor input across countries, as previously discussed in section 2.3.4. Due to these limitations in compiling the original data and in the use of TFP indexes, results, subsequently reported in this chapter are not conceptually comparable to estimates in Bernard and Jones (ibid.).

The only case where results of Bernard, Jones (1996) can be compared with convergence estimates in our study refers to the total manufacturing. Interestingly, it is precisely the sector where Bernard and Jones obtained their most unexpected result, failing to identify productivity convergence in contrast to the rest of analyzed industrial sectors. Therefore, we will pay specific attention to the total manufacturing, and show that the puzzling result in Bernard-Jones study can be largely attributed to their failure to extract cyclical components from the original data. Once cyclical movements are removed from their original data (as discussed in subsection 2.3.5), the resulting productivity estimates exhibit a distinct convergence in productivity, though cyclically-unadjusted data do not indicate a clear-cut productivity convergence.

There are two primary approaches to measure the tendency of countries to converge/diverge in productivity. The most widely used is the β-convergence, which is essentially a cross-country regression of productivity growth on the initial productivity levels. Barro and Sala-i-Martin (1995, pp. 383-387) provide detailed discussion of this concept of convergence. To apply the β-convergence, one needs sufficiently large cross sectional data, which is problematic with the sample of only nine countries in the present study.

This limited sample of countries is better suited for calculating the σ-convergence in productivity, which is essentially the dispersion of productivity levels, measured by the standard deviation of productivity across countries for a specific year. The approach identifies convergence by the reduction in the dispersion of productivity levels *over time*, which accords with a longer time span (1975-1993) of productivity estimates in our study compared with its less extensive country coverage.

One potential drawback of the σ-convergence is that there has been no statistical procedure to verify the significance of observed changes in the productivity variance. For example, Bernard and Jones (ibid., p. 1228-1229) rely exclusively on the graphical presentation of calculated standard deviations in international productivity levels. The same graphical approach is used by Barro and Sala-i-Martin (1995, pp. 392-393, 397-398, 400-401). In the next section we propose a simple procedure to fill in the gap in the literature.

Section 3.2. Simple test of σ-convergence.

The test is based on the observation that σ-convergence is identified by the decreasing time trend in the dispersion of productivity. Thus, the presence of σ-convergence can be verified by applying a tests for trend to the standard deviation σ in international productivity levels.

Specifically, one can use either the Cox-Stuart test or the Daniels test, both of which are well known in the statistical literature (see, for example, Conover (1971)). The Daniels test is especially attractive due to its relatively high power and relative ease of use.

The Daniels test verifies the presence of trend in some random variable $X = \{X_1, X_2, ..., X_n\}$ by pairing X with the trend variable $T = \{1,2,...,n\}$ and calculating the Spearman's rank correlation ρ between X and T. The null hypothesis of the test is that $X_1, X_2, ..., X_n$ and the trend variable T are independent. The alternative hypothesis is that $X_1, X_2, ..., X_n$ are related to the trend function, so that if the time go on, $X_1, X_2, ..., X_n$ are either increasing or decreasing (if the two-tailed version of the test is applied). Assuming that $X_1, X_2, ..., X_n$ are independent random variables, the test Daniels produces p-values that measures the statistical significance how X violates the null hypothesis. One more attractive property of the Daniels test is that it is a non-parametric test, so that no assumption of normality is required for its implementation.

To verify the σ-convergence by the Daniels test, one proceeds as follows:

1) Using productivity estimates for cross-sectional data, the standard deviation of productivity (σ) is calculated for a specific manufacturing sector i and time period t, producing the time series of $\sigma_{i,t} = \{\sigma_{i,1}, \sigma_{i,2}, ..., \sigma_{i,n}\}$ for the i^{th} manufacturing sector.
2) The Spearman's rank correlation ρ is calculated between $\sigma_{i,t} = \{\sigma_{i,1}, \sigma_{i,2}, ..., \sigma_{i,n}\}$ and $T = \{1,2,...,n\}$.

3) If the p-value for ρ is not statistically significant (such as above the conventional 0.05 threshold), the null hypothesis of the test is accepted. Conversely, if p-value for the Spearman rank correlation ρ is statistically significant and $\rho < 0$ ($\rho > 0$), the test indicates the presence of convergence (divergence), since productivity variance has tendency to decrease (increase) over time.

Section 3.3. Results of applying the test for σ-convergence.

Using productivity estimates from chapter 2, we tested the hypothesis of σ-convergence in international productivity levels in 23 manufacturing sectors and the total manufacturing. This appears to be the finest level of desegregation compared with available studies of σ-convergence.

To investigate the sensitivity of testing for σ-convergence to alternative estimates of productivity, we used 4 different productivity measures. The primary productivity measure was three-input total factor productivity, which was calculated with *trended* data, so that it is not affected by cyclical movements in output and inputs. Since the productivity measure is less noisy, one can expect that it will produce more clear-cut picture of σ-convergence compared with unadjusted productivity measures. To verify the claim, we also calculated σ-convergence for 3-input total factor productivity, when cyclical movements in output and inputs were not extracted from original time series. Two remaining productivity measures were two-input and one-input productivity measures, calculated with capital and labor, and labor only, and with removed cyclical factors. Hereafter we will refer to these 4 alternative productivity measures as 3TFP, 3TFP-nt, 2TFP and LP.

Results of testing for σ-convergence are presented in Table 3-1. We first consider the test results for the primary 3TFP measure of productivity. The null hypothesis of the test was not rejected in 3 manufacturing sectors: wood products, pottery and china products, and non-ferrous metals, indicating that relative productivity levels changed little over time in these sectors.

On the other hand, the Spearman rank correlation with time trend was negative and statistically significant in 12 sectors, such as beverages, apparel, footwear, furniture, paper products and other sectors.

Interestingly, the test identified the σ-convergence in the total manufacturing (and this result holds for each productivity measure that we considered), in contrast to Bernard and Jones (1996). The latter study claimed that "manufacturing shows little or no evidence of convergence for either measure of productivity [two-factor TFP and labor productivity] and, in particular, shows divergence during the 1980s" (p. 1230). We shall return to this discrepancy shortly.

Table 3-1 also shows that convergence in 3TFP is nothing like automatic phenomenon, since in as many as 8 manufacturing sectors the Spearman's ρ is statistically significant and positive, very close, indeed, to its maximum level. Interestingly, the σ-divergence is found in two machinery sectors (general and electrical machinery), indicating that the diversity in productivity levels increased consistently (note that in these sectors the Spearman's ρ are essentially unity, while p-values for the ρ are zero).

Table 3-1. Results of testing for σ-convergence/divergence in productivity levels among industrialized countries.

	3TFP		3TFP-nt		2TFP		LP	
	Spearman rho	p-level	Spearman rho	p-level	Spearman rho	p-level	Spearman rho	p-level
Food	1.00	0.000	0.93	0.000	-1.00	0.000	-1.00	0.000
Beverages	-0.61	0.005	-0.27	0.257	-0.99	0.000	-1.00	0.000
Textiles	0.49	0.033	0.34	0.154	0.32	0.188	0.62	0.005
Wearing Apparel	-1.00	0.000	-0.84	0.000	-1.00	0.000	-0.98	0.000
Leather & Products	0.51	0.026	0.09	0.721	-0.62	0.005	0.42	0.075
Footwear	-0.98	0.000	-0.90	0.000	-1.00	0.000	-1.00	0.000
Wood Products	0.00	0.989	-0.10	0.684	0.04	0.858	1.00	0.000
Furniture & Fixtures	-0.76	0.000	-0.57	0.011	-0.78	0.000	-1.00	0.000
Paper & Products	-0.66	0.002	-0.54	0.017	-0.99	0.000	-0.98	0.000
Printing & Publishing	0.94	0.000	0.63	0.004	0.90	0.000	-0.99	0.000
Industrial Chemicals	1.00	0.000	0.93	0.000	0.99	0.000	1.00	0.000
Chemicals, nes	-1.00	0.000	-0.78	0.000	-1.00	0.000	1.00	0.000
Rubber Products	-0.72	0.000	-0.15	0.533	-0.65	0.003	0.13	0.606
Plastic Products, nec	-0.99	0.000	-0.86	0.000	-1.00	0.000	-1.00	0.000
Pottery, China etc	0.20	0.403	-0.11	0.642	0.98	0.000	0.98	0.000
Glass & Products	-1.00	0.000	-0.98	0.000	-1.00	0.000	-1.00	0.000
Non-Metal Products	-1.00	0.000	-0.84	0.000	-1.00	0.000	-1.00	0.000
Iron & Steel	-1.00	0.000	-0.89	0.000	-0.67	0.002	-0.42	0.074
Non-Ferrous Metals	-0.33	0.166	-0.06	0.797	-0.01	0.977	1.00	0.000
Metal Products	0.89	0.000	0.52	0.023	-1.00	0.000	-0.93	0.000
General Machinery	1.00	0.000	0.95	0.000	1.00	0.000	0.99	0.000
Electrical Machinery	0.97	0.000	0.99	0.000	-0.10	0.684	0.54	0.017
Transport Equipment	-1.00	0.000	-0.92	0.000	-1.00	0.000	0.69	0.001
Total Manufacturing	-1.00	0.000	-0.96	0.000	-1.00	0.000	-1.00	0.000
Summary[a]								
Converging relative productivity		12		10		15		11
Diverging relative productivity		8		6		4		9
Stable relative productivity		3		7		4		3

[a]Excluding total manufacturing

When we applied the convergence test to data, unadjusted for cyclical factors, as expected, there was not so strong evidence in favor of divergence or convergence in productivity. Though the test results for 3TFP and 3TFP-nt are generally quite close, still there are two fewer cases when both convergence and divergence in productivity could be identified in unadjusted data. Specifically, while 3TFP data demonstrated convergence in beverages and rubber products, 3TFP data do not show any clear trend in these manufacturing sectors. Similarly, textile and leather products indicated divergence in productivity levels, but with unadjusted productivity data neither convergence nor divergence could be detected.

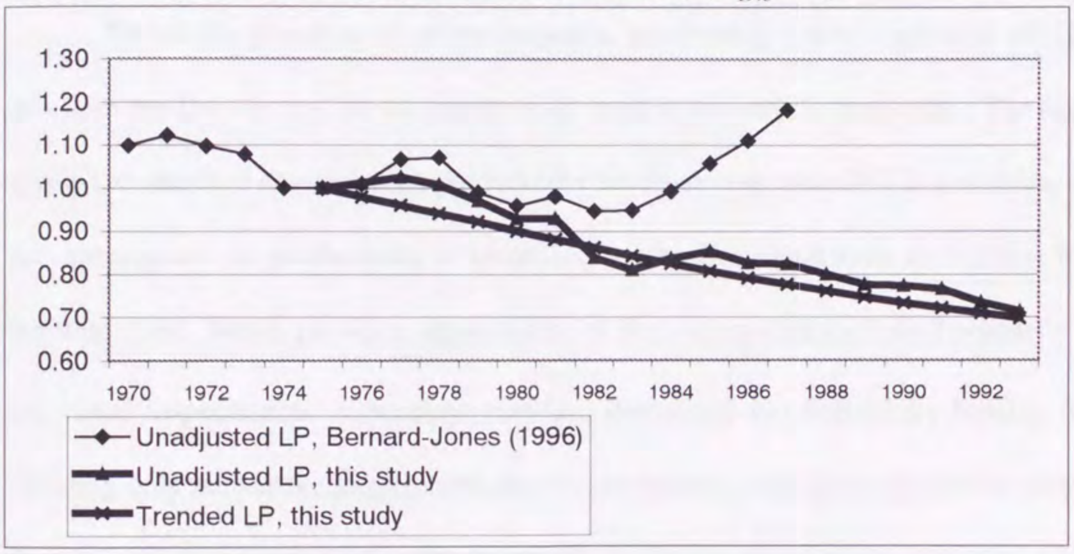
When the test was applied to two-input productivity estimates, the difference with estimates from 3TFP has become even more substantial, producing results that are more favorable to the σ-convergence, as shown at the bottom of Table 3-1. Besides, while there were only 4 inconsistencies between test tests with 3TFP and 3TFP-nt, the number increases to 6 when 3TFP and 2TFP are compared[6].

The inconsistency of the test results is even more pronounced when productivity is measure by labor productivity. Superficially, the total number of convergent, divergent and stable productivity sectors is similar for 3TFP and LP (12-8-3 and 11-9-3, respectively). Yet, there are as many as 11 inconsistencies for individual manufacturing sectors (such as for food products, where divergence was identified with 3TFP, but convergence was found with LP), demonstrating the importance of accounting for as many inputs as possible in the study of productivity convergence.

Finally, we return to the previously mentioned discrepancy of our result in table 3-1 for the total manufacturing compared with Bernard and Jones (1996) study. To make the comparison as close as possible, we will consider convergence results, derived from labor productivity, since they are free from idiosyncratic features of the Bernd-Jones' TTP index.

As chart 3-1 shows, there was quite a close correspondence between standard deviations of labor productivity from the two datasets for the time span when the comparison is possible (that is, from 1975 to 1987). The agreement is especially close if our original data are not adjusted for short-term variations. Then both measures show an upward shift in σ after the first oil shock. However, starting from the end of 1970s, manufacturing productivity resumes its convergence, interrupted by the second oil shock at the beginning of 1980s. As a result, our unadjusted data indicate the short-term occurrence of the productivity divergence until 1985, while estimates with Bernard-Jones data extend the period until 1987, which is the end of their sample.

Chart 3-1. Normalized standard deviations of (log) labor productivity in the total manufacturing for alternative measures of productivity ($\sigma_{1975} = 1$).



Most importantly, once the sample is extended to 1993, *and short-term cyclical movements are removed from the original data*, there is a continuous convergence in manufacturing productivity that is not interrupted by the two oil shocks, in contrast to σ, derived from unadjusted data. Consequently, the sharp discrepancy with the Bernard-Jones study can be mostly attributed to our use of more recent data that are not subject to short-term fluctuations.

---

[6]Such as in food, textile and leather products, which were identified as diverging with 3TFP estimates, but become either converging or stable sectors when 2TFP was used

Section 3.4. Conclusions.

In this chapter we considered the verification of convergence hypothesis at the level of disaggregated manufacturing sectors. While the majority of studies on productivity convergence deals with productivity levels for the whole economy or for the whole manufacturing, this chapter reports estimates of σ-convergence for twenty three 3-digit manufacturing sectors, revealing that the productivity convergence can be detected in only about the half manufacturing sectors.

To test the presence of σ-convergence, we develop a new approach, which is based on the Daniels test for the presence of trend in univariate time series. The test is applied to standard deviations of productivity levels among nine OECD countries, and the convergence in productivity is identified by the Spearman rank correlation with the time trend. While previous applications of the σ-convergence relied primarily on the visual inspection of σ (namely, standard deviations in productivity levels), thus allowing only subjective judgements about convergence, our approach in this chapter is more objective, producing estimates of the statistical significance (p-values) about the hypothesized convergence in productivity.

We also showed the sensitivity of reported results to various measures of productivity. For example, when the Daniels test is applied to time series of σ, derived from data with cyclical movements in output, inputs and input shares, the distinction between convergence and divergence in productivity became more ambiguous. Since cyclical factors are generally not synchronized across countries, they contributed to additional volatility in σ, making it less correlated with the time trend.

On the other hand, the verification of σ-convergence with productivity measures that account for fewer inputs produced a large number in inconsistencies about the direction of productivity differences, often yielding significant convergence while the three-input productivity index produced the opposite result. The inconsistency was especially pronounced with labor productivity, when conflicting results were obtained in about half of all considered manufacturing sectors, thus demonstrating limitations of using simple productivity measures.

The large number of manufacturing sectors with no significant convergence point at the existence of persistent gaps in productivity that fail to disappear over time. One interesting question that we have not addressed in this chapter is the casual explanation of these technological gaps. Possible causal factors may include the intensity of competition from domestic and foreign producers, the role of the state in the dissemination of technology and other institutional factors.

To investigate the impact of these factors, one needs to collect internationally compatible data that are currently mostly unavailable for large numbers of countries at disaggregated manufacturing level. Besides, the interplay between these causal factors may be too subtle to be identified in a simple theoretical framework. For example, manufacturing sectors where Japan managed to achieve good performance in international productivity consist of sectors with vastly different industrial structure, such as auto parts with strong presence of vertical keiretsu, machine tools with mostly independent producers, and textile machinery with a mixed industrial structure, as shown by Okada (1997, p. 43). Similar lack of clear-cut picture emerges in industries that were less successful in technological advance (ibid.). As a consequence, attempting to find a simple correlation between Japanese industrial structures and its productivity performance would produce hardly any sensible result[7]. Nevertheless, the task of explaining away the productivity differences deserves further investigation.

[7] In fact, we have attempted to estimate the link with a larger sample of five countries in our sample (Japan, USA, France, Germany and UK), but obtained largely inconclusive results.

Charter 4. The impact of international differences in productivity on comparative advantages in international trade.

This chapter continues the application of TFP estimates from chapter 2, but uses them in the context of international trade theory. In particular, in this chapter we will attempt to identify circumstances when comparative advantages in productivity can explain trade flows relatively well (if at all, skeptics may add) in the tradition of the Ricardian theory of international trade.

Simultaneously, the chapter considers an alternative theory of international trade that assumes away international differences in productivity (the Heckscher-Ohlin theory, or H-O for short). This joint test of alternative trade theories (using the same trade data) has not been common in the empirical work in international economics. The latter commonly has dealt with just one theory of international trade, thus leaving open the crucial question whether theoretical alternatives may perform better.

Estimating comparative strength of Ricardian and H-O theories, we follow a recent call of Edward Leamer (1992) to "estimate, not test" trade theories. Leamer emphasized a repeatedly ignored fact that theoretical models "are only tools, each of which is appropriate in some circumstances and inappropriate in others. Empirical enterprises should therefore not attempt to test the validity of the theories" (p.7). To have more influence on intellectual life of international economists, "empirical work might identify circumstances under which each of the [theoretical] tools is most appropriate, or measure the "amount" of trade that is due to each of the resources. Neither of these tasks has been accomplished or even attempted " (ibid.).

To the best of our knowledge, there seems to be just one study that dealt with simultaneous verification of trade theories, using the same data set. McGilvray and Simpson (1973) investigated Irish trade with the United Kingdom and considered both the Ricardian and H-O models of trade. In contrast, this chapter examines the Ricardian and H-O theories, using a more extensive database that covers American trade with eight other developed countries over 1988-1992[8].

Second, unlike most of previous verifications of the Ricardian theory, the chapter adopts its multi-factor version (Woodland, 1982) instead of dealing with labor only[9].

Third, unlike conventional studies of "factor contents of trade" as an approach to verify the H-O theory, we employ a different extension of the H-O theory to multiple goods, and estimates whether there is a significant link between trade flows and the chain of relative factor intensities across manufacturing sectors, as was first suggested by Jones (1956-1957) and subsequently refined by Bhagwati (1970) and Deardorff (1979).

The chapter is organized as follows. Section 4-1 formulates theoretical version of the Ricardian and the H-O theories that will be subsequently applied to trade flows between the United States and other developed countries. Section 4-2 describes dataset that was used in the study. Section 4-3 reports results of the study, indicating in particular country-pairs when theoretical predictions of the Ricardian and H-O theories corresponded to actual trade flows.

---

[8] The use of bilateral trade data is due to the bilateral nature of productivity estimates in Chapter 2.

[9] Persuasive criticism of "labor only" versions of the Ricardian theory was given by Bhagwati (1964, 1972a)

Section 4.1. Relative productivity and trade in the Ricardian model.

Originally the Ricardian model was formulated in terms of relative differences in labor productivity. Subsequent refinements has modified the Ricardian theory so much that now little, but the original Ricardian emphasis on the technological differences, links its current version with the original one-factor, two-commodities world of Ricardo.

First of all, it was subsequently realized that differences in wages may invalidate the theory, if countries with high relative labor productivity may also have relatively high wages, thus offsetting their cost advantage due to superior productivity. As a consequence, the Ricardian theory was reformulated in terms of relative labor costs, thus incorporating both relative labor productivity and relative wages (for early references, see Viner, 1964, p. 493-500).

Second, Dornbusch, Fisher and Samuelson (1977) extended the theory to a more realistic many-commodity (but still one-factor) case. They suggested indexing commodities by their relative unit labor requirements at home and abroad for n commodities, $(a_1,...,a_n)$ and $(a_1^*,...,a_n^*)$ respectively. Then commodities are ranked in order of diminishing Home comparative advantage:

$$a_1^*/a_1 > \ldots > \ldots > a_i^*/a_i > \ldots > a_n^*/a_n \tag{4-1}$$

To determine which commodities are produced at Home and which – abroad, define w and w^* as domestic and foreign wages, measured in some common unit (using the exchange rate between countries e).

Dornbusch *et. al* also showed that Home will specialize in the production of commodities for which domestic unit labor cost is less or equal to foreign unit labor cost $a_i \frac{w}{e} \le a_i^* w^*$, or $\frac{a_i^*}{a_i} \ge \frac{w/e}{w^*}$, with the ratio of real wages $\varpi = \frac{w/e}{w^*}$ determining the pattern of specialization between home and abroad[10]. Home country will produce and export goods for which Home's relative unit labor cost $c_i = \frac{a_i}{a_i^*} \Big/ \frac{w^*}{w/e}$ is less than unity, and import goods for which it is more than unity.

Dornbusch *et. al* (1987) still assumed the perfect mobility of labor across sectors that equalizes wages within countries, so that there is a single national wage rate in their model. Golub (1994) suggested to introduce sectoral disparities across sectors in the definition of c_i to account for differences in human capital and imperfections in the labor market.

These modifications of the original Ricardian still retain the one-factor assumption of the Ricardian theory that, as Bhagwati (1972, p. 134) noted, is totally arbitrary, unless one supports the labor theory of labor (as Ricardo did). Instead, Bhagwati advocated to consider the Ricardian theory in the multi-factor world, with trade pattern reflecting Hicks-neutral differences in total factor productivity.

Such an approach was formalized by Woodland (1982, p. 187-190). He eliminated the original assumption that labor was the only factor of production and considered a model with two factors (capital and labor), with trade between countries with similar total factor endowments and demand functions, but different total factor productivity.

---

[10] The generalization to multi-commodity case is simple, and no wonder that it had been already suggested at least in the classical test of the Ricardian theory by MacDougall (1952), using British and American pattern of trade. MacDougall noted that American wages were on average twice as high as British ones, so that USA must have possessed unit labor cost advantage as long as American labor

Originally consider the case of just two commodities (subsequently the assumption will be relaxed).

Woodland assumed the following production function for output Y:

$$Y_i = \gamma_i f^i(X_i) \tag{4-2}$$

with X_i denoting vector of inputs (capital and labor) and $i = 1,2$.

In the model, f^i is the same in both countries, but their "efficiency parameter" γ_i can be different between them. Woodland assumes that countries have same efficiency in producing good 1, but that the home country has higher efficiency in producing good 2 compared with abroad, so that

$$\gamma_2/\gamma_1 > \gamma_2^*/\gamma_1^* \tag{4-3}$$

Woodland proves that from (4-3) it follows that in autarky good 2 should be relatively cheap at home, so that under free trade the home country will export good 2 and import good 1.

Woodland's model can be easily generalized to multi-commodity case with n commodities. Similarly to Dornbusch et. al (1977), define the following chain of relative TFP in producing commodities from *1* to *n*, with home relatively efficient at producing commodities, placed at the left side of the productivity chain:

$$\gamma_1/\gamma_1^* > \ldots > \ldots > \gamma_i/\gamma_i^* > \ldots > \gamma_n/\gamma_n^* \tag{4-4}$$

Then let μ_i be an index of *total relative* factor cost of producing commodity *i*, where labor cost ϖ_i and capital cost π aggregated by the following Törnquist index formula:

---

productivity was more than twice as high as British one. And MacDougall found that out of 24 considered sectors, 20 ones corresponded to this prediction in terms of unit labor cost.

$$\ln \mu_i = \frac{1}{2}\left(SK_i + SK_i^*\right)\left(\ln \pi - \ln \pi^*\right) + \frac{1}{2}\left(SL_i + SL_i^*\right)\left(\ln \varpi_i - \ln \varpi_i^*\right) \tag{4-5}$$

with SK_i (SL_i) denoting capital (labor) cost shares in producing commodity *i*. As before, asterisk denotes abroad. Note that the assumption of different sectoral wages within countries is retained, but capital is assumed to be mobile across sectors, with an equalized capital cost π across sectors.

Normalize μ_i by exchange rate *e*. Then the home country will produce those commodities for which its relative efficiency advantage γ_i / γ_i^* exceeds home's total cost μ_i / e compared with abroad. In other words, the home country will specialize in these commodities for which relative unit total cost $\rho_i = \frac{\mu_i}{e} \Big/ \frac{\gamma_i}{\gamma_i^*} \leq 1$, and import commodities for which it has relative unit total cost disadvantage, with $\rho_i \geq 1$.

The formulation of the Ricardian theory in terms of unit total cost blends the extensions by Dornbusch *et. al* (1977) and Woodland (1982) into a model that allows both multi-factor and multi-commodity verification of the Ricardian theory, thus extending the scope of previous empirical studies that related trade and differences in productivity[11].

[11] In addition to already mentioned MacDougall (1952) and Golub (194), also Stern (1962), Balassa (1965), and Torstensson (1996) verified the Ricardian theory of comparative costs, but all of them considered labor productivity only. We are not aware of any empirical study that incorporated Hicks-neutral differences in TFP to the analyses of international trade.

Section 4.2. Technology-neutral effects on trade in the H-O model.

Following the celebrated Leontieff paradox and subsequent extension of the H-O theory by Vanek (1968) to multi-factor case, most empirical studies of the theory calculated factor contents of net trade. Then they checked whether trade in factor services corresponded to national relative endowment with factors of production (for the most comprehensive study, see Bowen, Leamer, Sveiskauskas (1987)).

Quite surprisingly, a different extension of the H-O theory to multi-commodity case by Ronald Jones (1956-1957) has been generally neglected, though the model has been significantly improved by Bhagwati (1972a) and Deardorff (1979).

Consider two-country (home country and abroad), two-factor (capital and labor) and multiple-commodity case. Rank commodities by their relative factor intensities, so that relatively capital-intensive commodities (compared with labor) are placed at the left side of the chain:

$$K_1/L_1 > \ldots > \ldots > K_i/L_i > \ldots K_n/L_n \tag{4-6}$$

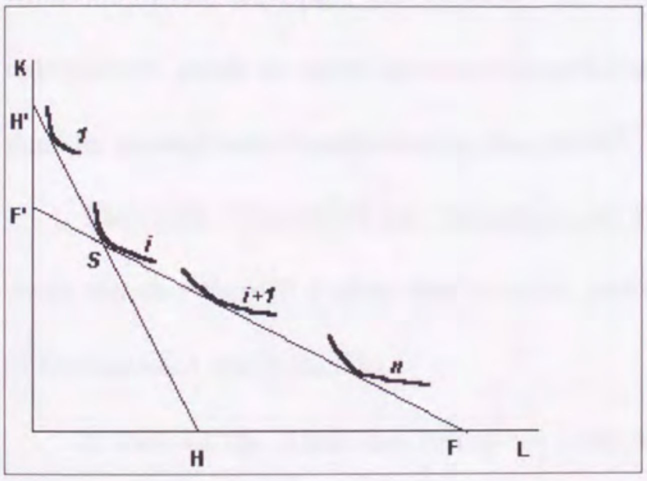
Suppose that the home country is relatively capital abundant than abroad, and that usual assumptions of the H-O hold (no factor reversals, same productivity across sectors and countries, common aggregate demand function between countries and so on). Then, due to the relative abundance of capital in the home country, the home country has higher wage-rental ratio compared with abroad[12].

---

[12] Note that the Jones model, in contrast to Vanek model, does not requires the stringent assumption of factor price equalization, and actually the Jones model breaks down with equal factor prices (Bhagwati, 1972b), which is hardly a drawback given the remote possibility of factor price equalization in the real world. One other relative advantage of the Jones' formulation of the H-O model is that it does not require application of input-output coefficients to calculation of factor services – quite a difficult task, especially in calculating factor services of imported commodities that ideally (but never in practice!)

Similarly to Deardorff (1979), draw Lerner diagram that depicts isocost lines for home HH' and abroad FF', and capital-labor composition in producing *n* commodities at home (Figure 4-1). Since home has higher wage–rental ratio, HH' is steeper than FF'. The diagram also shows unit-value isoquants in producing commodities *1, i, i+1,* and *n.* It is straightforward to show that if product markets are competitive, isoquant lines should be tangent to isocost lines[13], with segment *H'SF* defining possible combination of capital and labor that can be employed at home and abroad.

Figure 4-1. Illustration of multy-commodity version of the H-O model



requires that input-output tables for trading partners are applied. However, the Jones's model has also a liability that it is formulated in terms in two factors of production, and unlike Vanek's model, could not easily extended to more than two factors of production.

[13] Specifically, if isoquant line is above unit-value isocost line for some commodity, such a commodity will not be produced, since the same unit value can be obtained from producing other commodities, employing less capital and labor. On the other hand, if isoquant lines are located below isocost lines, smaller bundles of capital and labor now produce one unit of value, thus yielding profits that will eventually disappear, with the isoquant line becoming tangent to unit-value isocost.

More importantly, non-negative profit condition stipulates that the capital-abundant home country will produce commodity *1* (which employs more capital than labor), and possibly – the boundary good *i,* while goods *i+1* and *n* will be produced abroad. Note that if domestic and foreign prices are equalized, so that there is just one isocost line, say FF', commodity *1* will be produced in neither of countries, and moreover, no unambiguous ranking of commodities in terms of comparative advantage is possible (for example, the home country may produce both commodities *i* and *n*).

Deardorff (1979) showed that the chain version of the H-O theory can incorporate either impediments to trade or intermediate products, but not both of them. Due to the latter limitation, we opted for retaining two primary factors, since the sacrifice of intermediate goods seems to be more tolerable (at least for international economists!) than the assumption of impediments-free trade[14].

The chain versions of the Ricardian and H-O theories allows straightforward check whether there is a close link between relative factor intensities and the pattern of international specialization.

In case of the Ricardian theory we calculated sectoral ρ_i and run Spearman rank correlation between relative unit total costs and net exports from eight major

---

[14]Our original intention was to include intermediate materials along with capital and labor in estimating total unit cost ratios ρ_i in the estimation of the Ricardian theory, but we had to drop the idea in order to make results for the Ricardian estimation comparable with the chain version of the H-O theory that does not allow for intermediate materials as easily as the Ricardian theory does. The major drawback was the estimation of the stock of aggregate intermediate materials in addition to capital and labor inputs. Besides, the inclusion of one more factor of production considerably complicates country ranking in relative factor abundance.

industrial countries to the United States for which we assembled necessary output, input, and trade data.

Efficiency ratios γ_i/γ_i^* (or total factor productivity of capital and labor) were calculated by the bilateral Theil-Törnquist index, as described in Chapter 2. Denoting capital K_i and labor L_i as X_i (so that $X_i = \{K_i, L_i\}$), and factor value shares as SX_i, TFP index between country C and the United States (the base) equals

$$\ln TFP_i^{C-USA} = \ln Y_i^C - \ln Y_i^{USA} - \sum \frac{1}{2}\left(SX_i^C + SX_i^{USA}\right)\left(\ln(X_i^C) - \ln(X_i^{USA})\right) \qquad (4\text{-}7)$$

Due to the bilateral nature of original data for productivity estimations, only 8 possible trade flows with the United States can be analyzed[15]. Since we calculated ρ_i with the United States as the base country, relatively large levels of ρ_i indicate comparative total cost disadvantage for trade partners of the United States, and consequently, negative rank correlation between ρ_i (normalized by exchange rate) and net exports to the United States.

The chain version of the H-O theory was estimates in two steps. First, we calculated total factor endowment in capital and labor of the United States and its eight trading partners (using the same country sample and time span as in the estimation of the Ricardian theory). In the vast majority of cases the United States proved relatively more abundant in capital, thus suggesting that American advantage in producing and exporting commodities that employ relatively high ratios of capital

---

[15]But as many as 36 country-pairs might have been studied *if* the data allowed multilateral comparison among countries in the assembled dataset.

to labor. Conversely, other developed countries must have had comparative advantage in producing commodities with relatively intensive use of labor.

In the second step we run Spearman rank correlation between relative capital intensity and trade across manufacturing sectors. Since we specified trade as net export to the United States, the chain version of the H-O theory predicts negative correlation between sectoral trade and capital-labor ratios in labor-abundant trading partners of the United States[16].

[16] However, note that the correct sign of correlation becomes positive in a few cases when other developed countries, namely Canada and Japan, surpassed the United States in relative capital abundance.

Section 4.3. Data sources.

Data sources for productivity calculations were already described in Chapter 2. This section describes only data sources that were not discussed in section 2.3.

Total factor endowments in the estimation of the H-O theory was calculated from aggregate investment series in the World Bank's 1995 "World Tables". Due to the start of most time series since only 1970, the service life of capital goods was assumed to be 17 years. We applied the delayed linear scrapping rule. It assumes that new investments add to capital stock for a period of P years, and only afterwards, a constant proportion of past investments is scrapped every year. In particular, we assumed that P equals seven, with $\frac{1}{17-7} = 10\%$ scrapping proportion every year afterwards, so for eight year old investments 10% is scrapped, for nine year old investments 20% is scrapped, and eventually, 100% is scrapped at the year 17. Formally, the following formula was used:

$$KS_{i,t} = \sum_{n=1}^{P} INV_{i,t-n} + \sum_{m=P+1}^{P+S} INV_{i,t-m} \cdot \left(1 - \frac{m-s}{m}\right) \qquad (4\text{-}8)$$

Total employment data were taken from "Yearbook of Labor Statistics".

Sectoral trade data were obtained from CD-ROM "International Trade by Commodity, 1988-1995", prepared by OECD. Original data are provided in 5-digit SITC classification, and required the conversion to 3-digit ISIC manufacturing sectors. We used conversion matrixes, kindly provided by Michael Ward of the OECD.

Similarly to several other studies, e.g. Balassa (1986), Torstensson (1996), we specified sectoral trade variable as net exports between country C and the United States, normalized by the total trade between these countries in sector i:

$$NE_i^{C-US} = \left(E_i^{C-US} - M_i^{C-US}\right) / \left(E_i^{C-US} + M_i^{C-US}\right) \quad (4\text{-}9)$$

Since our trade data mostly included widely traded commodities, we followed the conventional practice and did not convert them by purchasing power parities, thus retaining the original data in US dollars.

Section 4.4. Estimation results.

Table 4-1 reports results of comparing the Ricardian chain of unit total costs with trade patterns between major developed countries and the United States.

Table 4-1. Spearman rank correlation between relative unit total cost ρ and sectoral trade

	Australia – USA	Canada-USA	Germany-USA	Great Britain -USA	France - USA	Japan-USA	Netherlands - USA	Sweden - USA
1988	0.190	0.300	-0.130	-0.382[b]	0.025	-0.350[b]	-0.064	-0.243
1989	0.169	0.174	0.017	-0.184	0.087	-0.388[b]	-0.077	-0.240
1990	0.293	0.143	-0.135	-0.106	0.073	-0.363[b]	0.029	-0.323
1991	0.310	0.086	-0.036	-0.069	-0.017	-0.286	-0.104	-0.353[b]
1992	0.272	-0.123	0.025	-0.085	0.027	-0.267	-0.128	-0.347[b]

Note: relative unit total cost ρ was calculated as explained in the text, using 1) *relative total cost index (5), and 2)* total factor productivity index (7). Sectoral trade is specified by (12).
[a]Significant at 5% level [b]Significant at 10% level

Among considered country-pairs, Japanese bilateral trade with the United States most consistently reveals close link with relative cost advantage due to relative total factor productivity, and has theoretically correct minus sign in 1988-1990, but afterwards falls short of the 10% threshold (when Spearman rank correlation equals 0.344).

Sweden-USA trade demonstrates two 10% significant case in 1991-1992, and UK-USA shows similarly close correlation between trade and relative unit total cost in 1988. Finally, other countries do not correspond to the Ricardian trade pattern, especially Australia that consistently shows theoretically incorrect sign of correlation.

It is possible to argue that the partially poor correspondence between trade and relative costs is due to the lack of time lag between achieving relative productivity advantage and *subsequent* trade pattern. To check the validity of such an assertion, we run correlation between sectoral trade and relative unit total cost, lagging the latter by one year. Table 4-2 shows results of this adjustment.

Table 4-2. Spearman rank correlation between relative unit total cost ρ (lagged one year) and sectoral net exports.

	Australia - USA	Canada-USA	Germany-USA	Great Britain -USA	France - USA	Japan-USA	Netherlands - USA	Sweden - USA
1988	0.228	0.226	-0.096	-0.423[a]	0.054	-0.357[b]	-0.060	-0.317
1989	0.117	0.340	0.005	-0.236	0.071	-0.359[b]	0.031	-0.239
1990	0.228	0.165	-0.122	-0.191	0.024	-0.339	-0.045	-0.292
1991	0.233	0.118	-0.113	-0.154	0.065	-0.326	-0.113	-0.292
1992	0.283	0.110	-0.058	-0.086	-0.036	-0.305	-0.150	-0.350[b]

Note: relative unit total cost ρ was calculated as explained in the text, using 1) *relative total cost index (5), and 2)* total factor productivity index (7). Sectoral trade is specified by (12).
[a]Significant at 5% level [b]Significant at 10% level

With lagged relative costs, the Ricardian theory fits trade data worse (for example, now Japan-US and Sweden-US reveals fewer cases, significant at 10% level). In contrast, the support for the Ricardian theory in the UK (in 1988) is now significant at 5% level. Results for other countries do not show closer link between trade specialization and comparative costs compared with the previous specification.

In sum, there seems to be just one robust Ricardian case of bilateral trade (Japan – USA) and two country-pairs that occasionally fit the Ricardian model of trade (Great-Britain-USA, and Sweden - USA), and as many as five consistently non-Ricardian cases (Australia-USA, Canada–USA, Germany-USA, France-USA, and the Netherlands-USA).

This conclusion is quite close to results in other recent studies of the Ricardian model, despite differences in considered country-pairs, and the reliance on labor as the only factor of production in other studies. For example, Golub (1994) dealt with trade flows of G7 countries and also noted that "Japan's trading pattern is better explained by the Ricardian model than that of many other countries" (p. 308). Similarly, Torstensson (1996) also found support for the Ricardian model, analyzing trade among Scandinavian countries, and we also found Swedish trade pattern quite close to Ricardian pattern of trade.

Now we consider results of estimating the chain version of the H-O theory. Table 4-3 ranks countries by their relative capital-labor ratio (with USA = 1). In most cases the United States is revealed to be abundant in capital than in labor compared with other developed countries, so that the theoretical sign of correlation between relative capital intensity and net trade to the United States should be negative in these cases. Note, however, that Canada in 1991-1992 and Japan in 1992 surpassed the United States in relative capital endowment, thus changing the expected sign of correlation to positive.

Table 4-3. Aggregate capital-labor ratio in major developed countries, USA = 1.

	1987	1988	1989	1990	1991	1992
Australia	0.987	0.973	0.959	0.951	0.960	0.965
Canada	0.919	0.930	0.955	0.979	**1.008**	**1.044**
Germany	0.900	0.891	0.882	0.846	0.809	0.818
Great Britain	0.530	0.523	0.523	0.524	0.543	0.563
France	0.947	0.942	0.939	0.929	0.922	0.938
Japan	0.931	0.944	0.961	0.969	0.972	**1.006**
Netherlands	0.825	0.805	0.793	0.765	0.735	0.776
Sweden	0.711	0.711	0.711	0.713	0.725	0.765

Note: capital and labor endowments were calculated as described in section 4-2.

Table 4-4 contains results of estimating the H-O theory in its chain version. Compared with predictive ability of the Ricardian theory, the H-O theory performs even worse. Only Canada shows some close correspondence between relative factor intensity and predicted trade flows, but note that the correlation has wrong sign in 1991-1992 when Canada turned into relatively capital abundant country[17].

[17] Similarly to the estimation of the Ricardian theory, we also experimented with lagging capital-labor ratio by one year. However, this modification did not change result substantially (for example, there was no new significant correlation, and in the case Canada sign and statistical significance of correlation remained the same).

Table 4-4. Spearman rank correlation between relative sectoral capital-labor ratios and sectoral net exports.

	Australia - USA	Canada-USA	Germany-USA	Great Britain -USA	France - USA	Japan-USA	Netherlands - USA	Sweden – USA
1988	0.012	0.298	-0.180	0.004	-0.049	0.009	0.103	0.168
1989	-0.057	0.400[b]	-0.261	-0.145	-0.070	-0.023	0.192	0.152
1990	-0.046	0.376[b]	-0.177	-0.190	-0.034	0.041	0.181	0.160
1991	-0.063	0.419[a]	-0.137	-0.085	-0.037	0.089	0.236	0.311
1992	-0.004	0.428[a]	-0.137	-0.053	-0.142	0.141	0.332	0.286

Note: Capital stock was calculated by (8) and (9) and as described in section 2. Labor endowment is measured by sectoral employment. Sectoral trade is specified by (12).
[a]Significant at 5% level [b]Significant at 10% level

In sum, the H-O theory showed much lower predictive ability, fitting data only for Canadian trade, but producing the correct sign only in the half of all statistically significant cases. Such unsatisfactory conclusion is not new for the H-O theory, is evident in the seminal study by Bowen *et. al* (1987). It is interesting that similarly to our study, Brecher and Choudhri (1993) found support for the factor-proportions theory in the US-Canadian study, analyzing their sectoral production pattern (but not their trade).

The preceding analysis makes possible to identify circumstances in which the Ricardian and H-O theories predicted trade flows well.

First, the multi-commodity, multi-factor version of the Ricardian theory generally predicted Japanese trade with the United States rather well. The Ricardian model can claim some success also in case of Swedish trade in 1991-1992 and British trade in 1988 (especially with lagged unit total cost). Seconds, the H-O theory fitted trade flows relatively well only for Canadian trade, and only in 1989-1990. Finally, Australian, German, French and Dutch trade patterns with the United States remained unexplained by both considered trade theories, thus warranting attention to some other factors of trade in addition to relative TFP and factor endowments.

Section 4.5. Conclusions.

In this chapter we suggested a number of testable extensions of the Ricardian and the H-O theory of international trade to the case of multiple factors of production (inputs). Though in principle these extensions permit the inclusion of several inputs, we limited the present study to the conventional pair of labor and capital, mostly due to difficulties in measuring the stock of intermediate inputs.

Applying the Leamer's recommendation to find circumstances when trade theories perform well (in contrast to simultaneously considered theoretical alternative), we analyzed trade flows between the United States and eight other major developed countries, using the same sample of countries and manufacturing industries an in chapters 2 and 3. However, due to the availability of trade statistic, the time span in this chapter was shortened to 1988-1992.

Quite in accordance with the Leamer's assertion, neither of major theories could comprehensively account for all considered trade patterns, though in relative terms the Ricardian theory performed better then the H-O theory. Most importantly, neither of these two theories was able to explain American trade with more than half of considered country-pairs. This result, in addition to quite possible data measurement errors, may suggest that some other factors of trade (apart from considered differences in technology and factor endowments) may be relevant, indicating that supply-side factors may be *not* the only feasible explanations of international trade. Consequently, the complementary impact of demand-side factors appears to deserve additional scrutiny. The following two chapter will deal with the diversity in consumer preferences in more details.

Charter 5. A new approach to estimate the diversity of cross-section data, based on outlier diagnostic.

This chapter attempts to fill the gap in current studies of consumer preferences by introducing a new approach that estimates the diversity of consumer preferences in a cross-section framework. After estimating a demand function in cross-section of countries, this approach attempts to identify countries-outliers from the estimated international consumption pattern, once the impact of economic factors of consumption is accounted for. Then, for example, the absence of national outliers in the estimation of demand functions indicates a general international homogeneity of consumer preferences for a specific good or service. Conversely, identified outliers with respect to the estimated consumption function point at specific countries that do not conform to consumer preferences in other countries.

The chapter is organized as follows. Section 5.1 introduces the concept of regression outliers on which our study of diversity of consumer preferences will be based in chapter 6. Section 5-2 discussed the available approaches in the statistical and econometric literature on the identification of regression outliers. Finally, section 5-3 contains the algorithm of the new outlier test.

Section 5.1. Evaluation of the international diversity of consumer preferences by outlier diagnostic in the linear regression model.

Conventionally, econometric studies of international consumption preferences placed major emphasis on point estimates of income and price elasticities with little, if any at all, attention on possible deviations in consumer preferences in separate countries. In this chapter we will introduce an opposite approach. It is based on verifying the stability assumption of the linear regression model that the same set of regression parameters is applicable to all observations in the analyzed data sample, with the absence of outlying observations from the estimated functional form.

To define regression outliers, consider the conventional linear regression model:

$$Y = X\beta + \varepsilon \tag{5-1}$$

where Y is $n \times 1$ matrix of dependent variables, X is $n \times k$ matrix of independent variables, and ε is the $n \times 1$ matrix of the disturbance (error) term. The stability assumption states that the mean of (unobservable) error term ε equals zero[18] for each observation i:

$$E(\varepsilon_i) = 0 \qquad \text{for each } i = 1,\ldots,n \tag{5-2}$$

The error term is unobservable, but the basic assumption (5-4) can be easily extended to observable least squares residuals, defined as $e_i = Y_i - X_i\hat{\beta}$, where $\hat{\beta} = (X'X)^{-1}X'Y$. From $E(\varepsilon_i) = 0$ it follows that $E(Y_i) = E(X_i\beta + \varepsilon_i) = X_i\beta$, which implies that $E(e) = E(Y - X\hat{\beta}) = X\beta - X\beta = 0$.

Hereafter an observation, for which $E(e) = 0$ does not hold, will be referred as regression outlier[19]. To illustrate the definition, it is useful to compare distributional properties of the outlying observation with the rest of data, which, having zero mean

---

[18]Other basic assumptions of the linear regression model (such as the full rank k of matrix X, the lack of serial correlation and heteroskedasticity, the independence between ε and X) are assumed to hold through.

[19] Note that the definition can encompass not only individual observations, but also sets of multiple observations as well.

of the error term, do not violate the basic assumption (5-2). Then, if error terms for homogenous (not-outlying) observations correspond to the normal distribution $N(0,\sigma^2)$, outlying observations are those that follow the normal distribution with the same variance σ^2, but with non-zero (shifted) mean μ.

The definition of regression outliers by the "shifted" mean of individual ε_i has been originally suggested by Srikantan (1961, p 252-253), and subsequently has become the most widely used approach to model regression outliers, as reviewed in Barnett and Lewis (1994).

Generally, deviation of regression outliers from the estimated majority pattern $E(Y) = X\beta$ may be due to three sources. First, outliers may be indicators that the estimated regression specification is insufficient, and omits some important variables. Second, outliers may be due to gross recording or measurement errors. Finally, the presence of outliers may indicate that regression estimation is applied to intrinsically diverse data that will continue to exist unless outlying observations are dropped from the data, or their impact on regression estimates is neutralized by suitable dummy variables.

The third case may illustrated the estimation of saving behavior, using a pool that mostly includes prudent households with a high saving rate and a few wealthy households with extremely extravagant spending habits far beyond their means. Another example can be the study of alcohol consumption, using data mostly from secular counties and a few Muslim countries where alcohol consumption is prohibited by law.

Consider the last example in more detail. Suppose that the original regression specification contains only general economic variables, such as price and income terms, failing to account for unusual legal features in Muslim countries. Quite often, the impact of included atypical observations (when compared with the majority of data) can be so large that the final regression estimates are representative of only atypical data, having little in common with the majority of analyzed countries.

In the following sections we will develop a new approach to identify observations that substantially deviate from the estimated regression pattern, formed by the majority of data. Once such regression outliers are identified, at the next step we will study whether these outlying observations can provide important information about relevant omitted variables that can meaningfully enrich the original regression specification.

Section 5.2. Conventional approaches to the identification of regression outliers.

Despite the possibility that regression outliers may point out important deficiencies in the original regression specification, the analyses of outliers has been itself an outlying subject in applied economics. This can be attributed the general neglect of specification testing, especially during early stages in the development of econometric methods that focused primarily on better estimation methods, implicitly assuming that the model is correctly specified from the outset.

However, the recent surge of interest in the specification testing has also drawn attention of economists to the problem of regression outliers, and currently the problem of regression outlier is sometimes discussed in econometric literature (examples include Belsley *et. al* (1980), Cook and Weisberg (1982), Donald and Maddala (1993)).

Another possible reason for the paucity of interest among econometricians in outlier detection is the lack of outlier detection procedures that might have been routinely applied in econometric work. For example, available tests of regression outliers are primarily designed for a limited case of a single regression outlier, as was demonstrated in Monte Carlo studies by Antille, Ritschard (1990) and Kianifard, Swallow (1990). With several outliers present in the data, OLS estimates will be biased by leftover outliers in the "benchmark" subset. If such leftover outliers resemble the single excluded outlier, the former may make the latter undetectable (creating so-called masking effect, when a testing procedure is unable to detect even a single outlier in the presence of several atypical observations).

Before presenting a new procedure for detecting regression outliers, we briefly discuss major procedures for outlier detection. Most of them include two distinctive steps. First, data are partitioned into a subset that contains 'the majority pattern' and a subset of potential outliers from the 'majority pattern'. At the second step, statistical significance of potential outliers is estimated. Advantages of available outlier tests can be evaluated with respect to these two stages.

Gentleman and Wilk (1975) developed what seems to be the first procedure for detection of multiple outliers. Being first, the procedure was highly awkward. Its major drawback was the requirement that the number of possible outliers be specified in advance. Suppose that the number is known and equals to *m*. Gentleman and Wilk suggested to look for such a combination of *m* observations that produces the largest decrease in the sum of squared residuals of the OLS fit, after these observations are excluded from the full dataset. If the total number of observations is *n*, finding "the most-likely outlier subset" requires the study of as many as $\binom{n}{m}$ combinations of data.

Suppose that the combination has been, indeed, found and we can proceed to the second step in outlier identification. However, Gentleman and Wilk suggested no statistical test for evaluating statistical significance of the found subset of potential outliers. They mentioning only the calculation of significance level by Monte Carlo study of "null cases" (that is, ones, that does not belong to outliers). Obviously, if such "null cases" were known in advance (as the Gentleman and Wilk's Monte Carlo procedure requires), then who would have trouble with calculation of $\binom{n}{m}$ statistics in the first place?! Indeed, hardly anybody did.

Marasinghe (1985) suggested a simplification to the Gentleman and Wilk procedure. The decrease in the sum of squared residuals after deleting a single observation is equal to squared adjusted residual $d_i = e_i / \sqrt{1-h_i}$, where e_i is OLS residual, h_i is a diagonal element of "hat" matrix $H = X(X'X)^{-1}X'$. Marasinghe proposed to form "the most-likely outlier subset" by subsequent deletion of observations that have the largest adjusted residual d_i.

However, this simplification still required that the number of outliers be specified before running the test. Kianifard and Swallow (1991), after a series of Monte Carlo tests, found that the Marasinghe's test had high power in detecting multiple outliers when the pre-specified number of outliers exactly coincided with the actual number of planted outliers. On the other hand, the test lost power dramatically when the equivalence between pre-specified and actual number of outliers was lacking, which is, indeed, the most common case in practice.

Paul and Fung (1991) suggested to form the "most-likely outlier subset" by ranking observations according to a statistics that can distinguish the presence of severe outliers. They proposed to rank observations first by studentized residual $t = e_i / s(i)\sqrt{1-h_i}$, where $s(i)$ is standard error of OLS regression with deleted observation i, and on the next step – by the Cook's D statistic, selecting around 20% of potential outliers in each step (this share of potential outliers was arbitrary). Subsequently, Hadi and Simonoff (1992) noted that Paul and Fung did not adequately solve the problem, because both studentized residual and the Cook's D would have little power to distinguish multiple outliers at high leverage points.

Paul and Fung also mentioned that robust regressions, such as least median of squares (Rousseeuw, 1984), can be used to flag possible outliers, but they eventually rejected the option, motivating the decision that "LMS is not on everyone's desk and it requires substantially more computations" (p.345).

Subsequently, Hadi and Simonoff (1993) suggested their own outlier test in linear regression. A novel feature of their test was the use of t-statistic to test the significance of potential outliers at the second, confirmatory part, after the list of potential outliers was determined[20]. However, the original formation of the subset of clean data in the first part of test still relied on the OLS estimation, which may be very susceptible to the influence of even a single regression outlier, since the OLS fit involves the summing of squared residual for *all* observations. Still, the Hadi-Simonoff test remains, evidently, the most powerful among available alternative methods of detecting multiple outliers in linear regression that can be a good benchmark for the evaluation of the new detection procedure of multiple outliers.

---

[20] In Marasinghe (1985) and Paul and Fung, critical values had to be derived from simulations, and were generally available only for the case of simple regression with a constant and a single independent variable.

Section 5.3. Test algorithm.

Similarly to the outlined approaches to outlier identification, the new diagnostic of multiple outliers consists of two stages. First, the subset of most likely outliers is *identified,* along with the benchmark subset of data. Second, the statistical significance of potential outliers is *verified.*

During the first stage the test involves a robust regression method which, unlike the OLS, is less sensitive to outlying observations. However, this robust regression estimator does not currently have an analytically tractable distribution of its residuals. Consequently, the distribution of the test statistic is approximated by repeating the test procedure many times, producing artificial test statistics, from which p-values for the actual test statistic are obtained.

To choose the most appropriate robust regression for the identification stage, the concept of breakdown point of regression (Donoho, Huber, 1983) is useful. The breakdown point of regression is the smallest share of data contamination by outlying observations that can make regression estimates infinitely large.

For the OLS, even a single deviant observation can substantially change coefficient estimates, so the breakdown point of the OLS is $1/n$. Rousseeuw and Leroy (1987) showed that the maximal breakdown point for any possible regression equals $([(n-k)/2]+1)/n$, where k is the number of independent variables, and $[q]$ denotes the integer of q.

The maximal breakdown value is currently attained by three robust regressions – the repeated median method (Siegel, 1982), the least median squares (LMS) and the least trimmed squares (LTS), the last two are due to the pioneering paper of Rousseeuw (1984). The repeated median estimator has disadvantage that it is not affine equariant[21]. The LMS and LTS are both affine equariant, but the LTS is slightly preferable. The LTS has a better convergence rate, and its objective function[22] is also smoother than the LMS (Rousseeuw, Zomeren, 1990, p. 650).

One possible disadvantage of using the high-breakdown regression at the first stage is that these regressions may require a prohibitively large amount of computations, as alleged by Paul and Fung (1991). Generally, robust regressions involve calculation of residuals for $\binom{n}{k}$ subsets of data. Still, due to constant improvements in computer hardware, the computational burden has become not so severe as in the past, and currently the amount of required calculations could be feasible in many cases.

Also note that, unlike the Gentleman and Wilk's procedure, $\binom{n}{k}$ does not depend on the number of possible outliers p, but only on the total size of dataset n and the number of independent variables k. For $n = 50$ and $k = 3$, the LTS has to consider 230,300 subsets of residuals. The number may seem to be too large, but the calculation took 86 seconds on a standard personal computer.

Another common objection to robust regressions – that they are "not on everyone's desk" – has also become largely irrelevant, since both LMS and LTS are

---

[21] Affine equariance holds when linear transformations of X change regression estimator accordingly, so that predicted values of dependent variable (and regression residuals as well) are the same in various coordinate systems.

[22] The sum of h smallest squared residuals, where the optimal h with the highest breakdown point is $h_{opt} = [(n + k + 1)/2]$.

currently included in several statistical and econometric software packages, such as S-PLUS, SAS, or TSP (we used the LMS routine in TSP).

There remains the problem of the number of potential outlier that should be retained after performing the robust fit by the LTS. Before discussing the problem, the following useful assumption is invoked:

Assumption 1. If the *least* deviant observation in the subset of potential outliers is found to be statistically different from the benchmark subset, than all the remaining (and more extreme) observations in the subset of potential outliers are also adjudged to be outliers.

The primary use of this assumption is that it allows to apply a single-case diagnostic (such as the standardized prediction residuals, to be defined shortly) for inferences about subsets of *multiple* observations. Now we can introduce the complete test algorithm, which follows the following steps:

1. Apply a robust regression estimator (such as LTS) to all data.
2. Sort observations by absolute values of their residuals from the robust fit. Let the relabeled sequence of absolute residuals from the smallest to the largest be denoted by $e_1 \le e_2 \le e_3 \le \ldots \le e_{n-2} \le e_{n-1} \le e_n$.
3. Form the benchmark subset with the smallest where k is the number of independent variables (including the intercept),
4. Partition data into the original benchmark subset B with the smallest $k+1$ absolute residuals and the subset P of $n-(k+1)$ potential outliers.
5. Estimate OLS regression with the benchmark subset, and then calculate standardized prediction residuals for all $n-(k+1)$ potential outliers:

$$t_i = \frac{y_i - x_i \hat{\beta}(i)}{\hat{s}(i)\sqrt{1 + x_i^T \left(X^T(i)X(i)\right)^{-1} x_i}}, i \in P \tag{5-3}$$

where $\hat{\beta}(i)$ and $\hat{s}(i)$ denote OLS estimates of regression coefficients and the standard error of regression, calculated with the benchmark observations only (that is, excluding observations, indexed by *i*, where $i \in P$. Similarly, *X(i)* denotes $(k+1) \times k$ matrix of independent variables in the benchmark subset B.

6. Find an observation with the smallest studentized prediction residual, and record it as $t_{k+1}^{\min}$.

7. Include the least outlying observation into the benchmark subset *B,* and repeat steps 1-6 until the subset P contains just one observation. Record the last sequential test statistic $t_n^{\min}$.

After repeating the test algorithm $n-(k+1)$ times, one obtains the sequence of $t_{k+1}^{\min}$, $t_{k+2}^{\min}, \ldots, t_n^{\min}$ test statistics for progressively decreasing subsets of potential outliers. Since the test assumes no knowledge about the number of regression outliers and about the direction of outlying observations, one possible criteria to identify the most likely subset of regression outliers by the largest $\left|t_j^{\min}\right|, j = k+1, \ldots, n$.

Under the null hypothesis of no regression outliers, each $t_j^{\min}$ is approximately distributed as t-distribution with $j-k-1$ degrees of freedom (see, for example, Cook and Weisberg (1981, p. 33-34)).

However, note that the null distribution of sequential studentized prediction residuals depends on the varying degrees of freedom, so that the sequence of $t_{k+1}^{\min}$, $t_{k+2}^{\min}, \ldots, t_n^{\min}$ test statistics is not directly comparable. Yet, one may transform the sequence of $t_{k+1}^{\min}$, $t_{k+2}^{\min}, \ldots,$ $t_n^{\min}$ into standard normal deviates *N(0,1)*, as proposed by Hawkins (1991, p. 223). Hawkins also suggested using the following normalizing transformation, originally due to Wallace (1959):

$$u_j^{\min} = \frac{8v+1}{8v+3}\left(v \ln\left[1+\frac{(t_j^{\min})^2}{v}\right]\right)^{1/2} \tag{5-4}$$

where v denotes the number of degrees of freedom $j-k-1$.

After transforming the $t_{k+1}^{\min}$, $t_{k+2}^{\min}, \ldots,$ $t_n^{\min}$ into $u_{k+1}^{\min}$, $u_{k+2}^{\min}, \ldots, u_n^{\min}$ statistics, the most-likely outlying subset is identified by $u^* = \max\left|u_j^{\min}\right|$, $j = k+1, \ldots, n$.

Once an observation with the test statistic u^* is found, we move to the second step of the outlier test, when we evaluate the statistical significance of u^*. Since the distribution of u^* is non-standard (after all, the preliminary ordering of data by absolute LTS residuals invalidates the basic assumption that $u_{k+1}^{\min}$, $u_{k+2}^{\min}, \ldots, u_n^{\min}$ are independently distributed), p-values for the test statistic u^* are calculated by a parametric bootstrap test. The test procedure is repeated a large number of times with data, artificially generated data according to the null hypothesis $H_0 : \varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, and then counting the number of times when the actual test statistic exceeds test statistics from simulated data. In th common cases when the direction of regression outliers is not known, the test becomes two-sided, so that both actual and simulated test statistics should be expressed in absolute values.

The algorithm for the bootstrap calculation of p-values is the same as in Diebold and Chen (1996, p. 234). To save space, it will not be discussed here. We only note that both the LTS and OLS residuals are invariant to linear transformations in regression parameters and in the error term variance. Therefore, without the loss of generality, one can generate artificial error terms from, say, $N(0,1)$, and then apply an arbitrary vector of regression coefficients to calculate simulated dependent variables. Once the vector of artificial dependent variables is generated, the test procedure is applied to the artificial data (using actual matrix of independent variables).

To study critical values of the suggested outlier test, we performed a small Monte Carlo experiment. Using random number generator in TSP, we generated samples of independent variables X_1 and X_2 such that $X_1 \sim \mathrm{U}(0,15)$ and $X_2 \sim \mathrm{U}(0,10)$. The error term ε was generated as $N(0,1)$. Without loss of generality, each coefficient β_i was assumed to equal one, so that dependent variable Y was calculated as $Y = 1 + X_1 + X_2 + \varepsilon$.

The testing procedure was again applied to 1000 sets of artificial data with sample size that ranged from 20 to 100 observations. Absolute values of the test statistics u_r^* *(r = 1,...,1000)* were saved and then sorted from the smallest to the largest test statistic. 5% critical value of the test statistic was obtained as the arithmetic average of 950^{th} and 951^{st} test statistics, so that exactly 5% of generated artificial samples produced test statistic above the calculated empirical critical value.

Table 5-1 reports the calculated small-sample critical values for the outlier diagnostics. The table also contains nominal critical values of the test, calculated as follows. Since the test involves the preliminary sorting of n observations by absolute values of their LTS residuals (step 2 of the test algorithm), according to the Bonferroni inequality, the significance level of the test should not exceed α/n. Correspondingly, the nominal critical value is given by $t_v^{\alpha/n}$, where v defines the number of degrees of freedom $v = n - k - 1$.

Table 5-1. Simulated critical values u^* for the outlier test, with $k=3$.

Sample size	5% critical values	
	Artificially simulated with 1000 Monte Carlo replications	*Nominal critical values* $t_\nu^{\alpha/n}$
20	3.14	3.02
40	3.19	3.23
60	3.28	3.34
80	3.37	3.42
100	3.46	3.58

It is important to note that the simulated critical values in table 5-1 in the strict sense are relevant only to artificially generated matrix X, which we assumed to be uniformly distributed, and can yield a completely different set of critical values with a different choice of the matrix X.

Nevertheless, in a number of cases there is a surprisingly close agreement between the simulated and nominal critical values for the test, especially for large n, when the computational burden of the LTS estimations and subsequent Monte Carlo replications is particularly heavy. Therefore, it appears that for n exceeding 60 observations, the approximation by nominal critical values $t_\nu^{\alpha/n}$ can be quite a close substitute for computationally demanding exact critical values, derived from multiple Monte Carlo replications.

The power of the suggested outlier diagnostic was compared with other available tests of regression outliers. They included the maximum studentized residual $t_i = e_i / \left[s(i)\sqrt{1-h_i} \right]$ and the Hadi-Simonoff outlier test. The test with the studentized residual is a typical example of a single outlier diagnostic. It was expected to perform well in the presence of a single outlier, but break down in the case of multiple regression outliers. The Hadi-Simonoff test is a good example of sequential testing procedure that does not use a robust regression method during the first identification stage.

The power study of the new and available outlier diagnostics was performed with artificial data, used in the preceding Monte Carlo experiments. The sample size was fixed at 40. Data generation process for the clean data was $Y = 1 + X_1 + X_2 + \varepsilon$, where $X_1 \sim U(0,15)$, $X_2 \sim U(0,10)$, and $\varepsilon \sim N(0,1)$. To create unambiguous outliers, we assumed that outlier had error terms that were shifted by 5 standard deviations from the homogenous data, generating regression outliers as $Y = 1 + X_1 + X_2 + [5 + \varepsilon]$.

We considered two cases with a single and five outliers. It is well-known that the power of outlier diagnostics may be greatly influenced by the presence of high leverage points (defined as observations with large diagonal elements h_i of the projection matrix $H = X(X'X)^{-1}X'$), as demonstrated by the Monte Carlo experiment in Hadi and Simonoff (1993). To differentiate such cases, we separately considered regression outliers, planted at low leverage and high leverage points. Other details how these leverage points were generated can be found in Hadi and Simonoff (1993, p. 1269).

The test power was calculated by the number of cases when the test rejected the null hypothesis that data contain no outliers. Rejection frequencies for 1000 test runs are reported in table 5-2.

Table 5-2. The power of suggested and available outlier diagnostics for multiple regression with n=40 and k=2.

Type of generated Outliers	*Outlier diagnostic by*		
	Largest studen-tized residual	*Hadi-Simonoff sequential test*	*Suggested outlier test*
One outlier at low leverage point	0.987	0.898	0.968
One outlier at high leverage point	0.776	0.546	0.708
Five outliers at low leverage points	0.129	0.898	0.943
Five outliers at high leverage points	0.046	0.104	0.375

In accordance with prior expectations, the test, using the studentized residuals $t_i = e_i / \left[s(i)\sqrt{1-h_i} \right]$ performed well in its 'designed' case when the number of planted outlier was just one (in fact, the test achieved the best performance among considered diagnostics). However, studentized residuals fared miserably in the case of multiple outliers, having power no more than to the nominal level with multiple outliers at high leverage points.

The performance of Hadi-Simonoff test was satisfactory with outliers at low leverage points, including the case of multiple outliers at low leverage points. However, the presence of high leverage points greatly affected the power of Hadi-Simonoff test, especially in the case of multiple outliers, indicating that the test is not robust to the masking effect (apparently, due to its reliance on the OLS estimation during the first, identification stage).

Finally, the new outlier diagnostic was the only testing procedure that did not break down in the most challenging case of multiple outliers at high leverage points. The new test only slightly fall short of the best performance of studentized residuals in the case of a single regression outlier. Besides, it regularly outperformed the Hadi-Simonoff test as well, especially in the most difficult case of multiple regression outliers, planted at high leverage points.

Chapter 6. The application of the suggested outlier test to international consumption data.

Compared with studies of international productivity levels, analyses of international consumer preferences have two distinctive features. First of all, while most productivity studies have utilized index numbers based on just two production functions (Cobb-Douglas and translog), studies of consumer preferences have applied a wider specter of demand functions (Deaton and Muellbauer, 1980), such as the linear expenditure system, the Rotterdam model, and the Almost Ideal Demand System.

Second, unlike productivity studies, the study of international consumer preferences are usually "blessed" with better quality data. For example, compared with the persistent lack of good capital data in productivity studies, national accounts provide essentially every sort of data in time series on per capita consumption expenditures and deflators for major consumption categories that, I additional, are generally well-standardized across countries.

The coverage of data may be even better in the case of cross-section analysis of consumer preferences. Due to the longer history of the ICP compared with the ICOP, the ICP data are already collected using the same methodology and classification scheme, whereas these tasks have not been achieved yet in the ICOP. As a result, studies of international consumption preferences can use data from the same source, covering a wider range of goods and services. Even more importantly, the ICP data are generally expressed in the multilateral form, facilitating the multilateral comparisons of consumer preferences.

However, the good quality of the ICP data has not often been utilized in past studies of international consumer preferences. Until the early 1980s, the vast majority of them used time series data from national account statistics, which are limited to consumption categories at a highly aggregated, level (usually, no more than 10 categories). The first analysis of international consumer preferences using the ICP data appears to be Kravis, Summer and Heston (1982). Subsequently, with apparently the sole exception of Dowrick and Quiggin (1994), the ICP data has again been neglected in economic studies of international consumer preferences.

The oversight is especially difficult to understand if we compare the limited usage of the ICP data with the huge popularity of Penn World Tables (PWT), which are nothing more than an extrapolation in time of ICP's benchmark studies, using the most aggregated categories[23]. A comparison with the PWT data may shed some light on the enigma of limited use of ICP data. While the PWT are panel data that can be subject to extensive analysis by panel estimation methods, data from ICP's benchmark studies can be only applied in cross-section analysis due to the lack of appropriate deflators for highly disaggregated consumption categories for the basic heading categories of the ICP. It appears that the small number of applications of the *original* ICP data may ultimately be due to the lack of a theoretical approach (other than the revealed preference approach, used by Dowrick and Quiggin (1994)) that might have been extracted some interesting results from these cross-section data.

This chapter we will apply the new outlier test from chapter 5 to the real data on international consumption expenditures. Before moving to the 'real' task, in section 6.1 we review conventional approaches to estimate the diversity of

[23]Heston and Summers (1996) noted that by some estimates, more that 20,000 regressions were performed on the PWT dataset!

international consumer preferences. Section 6.2 discusses data sources, while the choice of appropriate regression specification is considered in section 6.3. Section 6.4 contains major results.

Section 6.1. Review of previous studies of international differences in consumer preferences.

Previous studies of international consumer preferences can be classified into two major groups: studies that analyzed time series data on major consumption categories as a part of national accounts, and studies that relied on cross-section data of the ICP.

Houthakker (1965) pioneered the analysis of time-series data. He examined per capita expenditures on 5 categories of consumption in 13 OECD countries, using the following log-linear regression specification:

$$\ln q_{it} = \alpha_i + \beta_i \ln C_t + \delta_i \ln \frac{P_{it}}{P_t} + \theta_i t + \varepsilon_{it} \tag{6-1}$$

where q_{it} is per capita expenditures (in constant prices) on commodity i in year t, c_t is total consumption expenditures per capita in year t, p_{it} is implicit price deflator on commodity i in year t, p_t is deflator for total consumption expenditures in year t, and ε_{it} is a conventional error term, i.i.d. as $N(0, \sigma^2)$.

Though Houthakker claimed that the regression specification "remains without serious rivals in respect of goodness of fit, ease of estimation, and immediacy of interpretation" (p. 278), the specification still has one serious drawback – the lack of correspondence to any utility function. Besides, specification (6-1) does not allow the imposition of the adding-up restriction. On the other hand, one of the advantages of (6-1) was the ease of estimating price and income elasticities, reflecting the prevailing view in early demand studies that "the estimation of elasticities [was] the primary goal of empirical demand analysis", as noted by Deaton and Muelbauer (1980, p. 17).

Houthakker (1965) estimated specification (6-1) for each of 5 consumer expenditures and 13 countries separately, and found that international differences in elasticities were statistically significant for every consumption category (p.287). Subsequently, Goldberger and Gamaletsos (1970) analyzed a similar data on commodities and countries as in Houthakker (1965), but used regression specification, based on the linear expenditure system. The study also found that "elasticities show considerable variation across countries, uniformity being particularly lacking for price elasticities" (p. 385).

Parks and Barten (1973) attempted to explain away this international diversity in parameter estimates by differences in population composition. They regressed national parameter estimates of the linear expenditure system on several demographic characteristics. Lluch and Powell (1975) followed the same approach, using international differences in GDP per capita at purchasing power parities. Both these studies managed to explain away only a part of detected international differences in parameter estimates, concluding that international differences in demand systems are persistent, even after accounting for differences in demographic factors and levels in GNP per capita.

Curiously, the opposite conclusion was invariably reached in studies that analyzed cross-section data on consumer preferences, using the revealed-preference approach. The approach compares consumer choices at alternative sets of prices, from which (otherwise unobserved) consumer preferences can be derived.

When applied to time-series data, the approach can indicate whether a single (representative) consumer is consistent in his preferences (see Landsburg, 1981;

Chalfant, Alston, 1988). On the other hand, when applied to cross-sectional data, the approach can test whether there is a utility function that is shared by a pair of consumers, thus allowing tests on the existence of common tastes across countries. A distinctive feature of the revealed preference approach is that it deals with preferences not for a specific good or service, but their consumption bundles.

Originally, the revealed preference approach was applied to international consumption data by Kravis, Heston and Summers (1982). The study made comparisons of consumer preferences for 30 developed and developing countries that participated in the 1975 benchmark study of the ICP, analyzing 108 categories of goods and services that were aggregated to the total consumption bundle.

Quite surprisingly, the revealed-preference approach detected no country-pair that violated the hypothesis of common preferences for the total consumption bundle (p. 354-357). Subsequently, Dowrick and Quiggin (1994) analyzed data from the 1980 benchmark study of the International Comparison Program that covered 60 countries. With data on 10 broad categories of expenditures, the revealed-preference approach detected only two cases when the hypothesis of common preferences did not hold (Finland-Austria and Nigeria-Zimbabwe), though as many as 1,700 country-pairs were analyzed. Moreover, when Dowrick and Quiggin used more disaggregated data (38 categories of expenditures), the hypothesis of common tastes was supported by every country-pair.

In part, such an overwhelming rejection of common tastes in the latter study can be partially attributed to a rather unusual composition of consumption bundle, used by Dowrick and Quiggin. They considered not only consumption goods (as in the original study by Kravis et. *al*), but also other categories of GDP, most unusually –

investment expenditures (such as investments in non-residential buildings, producer durables, and inventory change).

Besides, both applications of the revealed-preference approach considered consumption bundles that included too many commodities, so that some international differences in tastes on *specific* commodities might have been remained undetected in the aggregated bundle.

In a further step beyond the inconclusive application of the revealed-preference approach to aggregate consumption bundles, *Kravis et. al* attempted to analyze cross-country data for separate commodities. They modified the log-linear regression specification from Houthakker (1965) to the cross-section data:

$$\ln q_{ij}^{PPP} = \beta_{1i} + \beta_{2i} \ln C_j^{PPP} + \beta_{3i} \ln \frac{PPP_{ij}}{PPP_j} + \varepsilon_{ij} \qquad (6\text{-}2)$$

where q_{ij}^{PPP} is consumption per capita of commodity i in country j (at purchasing power parity); C_j^{PPP} is total consumption per capita in country j (at purchasing power parity); PPP_{ij} and PPP_j are purchasing power parities in country j for commodity i and for total consumption, respectively; and ε_{ij} is the conventional error term.

Applying the varying parameter model to (6-2), Kravis et. *al* substituted the stochastic error term ε_{ij} in (6-2) with separate stochastic terms for regression parameters, as follows:

$$\ln q_{ij}^{PPP} = (\alpha_{1i} + \xi_{ij}) + (\alpha_{2i} + \varpi_{ij}) \ln C_j^{PPP} + (\alpha_{3i} + \psi_{ij}) \ln \frac{PPP_{ij}}{PPP_j}$$

where ξ_{ij}, ϖ_{ij}, and ψ_{ij} are i.i.d. as $N(0, \sigma_{\xi,i}^2)$, $N(0, \sigma_{\varpi,i}^2)$, and $N(0, \sigma_{\psi,i}^2)$, respectively.

The study associated international variation of tastes with heteroskedasticity of the stochastic terms ξ_{ij}, ϖ_{ij}, and ψ_{ij}. Kravis *et. al* noted that the magnitude of

$\sigma_{\xi,i}$, $\sigma_{\varpi,i}$, and $\sigma_{\psi,i}$ provides "basis for judging how close together price and income elasticities are around the world... The smaller they are, the more tenable becomes the hypothesis that tastes are common... with common tastes represented by $\sigma_{\xi,i} = \sigma_{\varpi,i} = \sigma_{\psi,i} = 0$" (p. 359-600). However, Kravis *et. al* failed to apply this model, as trial values of $\sigma^2_{\varpi,i}$ during the maximum likelihood estimation constantly entered negative range, so that the promising model was eventually dropped. It is noteworthy that the outlier test from chapter 5 also focuses on particular error terms ε_{ij}, but, on the other hand, it does not involve substantial computational complications.

Section 6.2. Data.

Our study of international diversity of consumer preferences was based on a single source of data – the ICP benchmark estimates for 1990 and 1993. The data were obtained by request from the World Bank, and were supplied on floppy disks. The ICP data included per capita expenditures on 110 categories of consumption goods and services, which were expressed in domestic currencies and at purchasing power parities.

The ICP data are arranged in international cross-sections with no linkage in time. This feature of ICP data precluded the modeling of consumer preferences for durables[24]. We also omitted several consumption categories that contained a large number of missing data. Finally, following the comment in Kravis *et. al* (1975, p. 49) that residual categories of ICP data (such as 'milk products, not else specified') may contain a high share of internationally incomparable data[25], the residual categories were also omitted from the final sample of analyzed commodities.

The final selection of cross-sectional data contained 73 and 61 categories of goods and services for 1990 and 1993-benchmark studies, respectively. The sample of countries included 22 OECD countries in 1990 and the extended sample of 35 countries in 1993 (see Table 6-1). The latter sample consisted of 24 OECD countries (with the addition of Iceland and Switzerland) and 11 former socialist countries which, with the exception of Hungary, Poland and Romania, participated in the ICP for the first time.

---

[24] Unlike non-durables and services, the consumption of durables is extended over a long period of time, and is affected by the past and future conditions in the economy.

[25] Since it is highly unlikely that national statistical offices interpret such categories in similar ways.

Table 6-1. Countries in 1990 and 1993 cross-sections.

	1990	1993
1	Australia	Australia
2	Austria	Austria
3	Belgium	Belourussia
4	Canada	Belgium
5	Denmark	Bulgaria
6	Finland	Canada
7	France	Croatia
8	Germany	Czech Republic
9	Greece	Denmark
10	Ireland	Finland
11	Italy	France
12	Japan	Germany
13	Luxembourg	Greece
14	Netherlands	Hungary
15	New Zealand	Iceland
16	Norway	Ireland
17	Portugal	Italy
18	Spain	Japan
19	Sweden	Luxembourg
20	Turkey	Netherlands
21	United Kingdom	New Zealand
22	United States	Norway
23		Poland
24		Portugal
25		Romania
26		Russian Federation
27		Slovak Republic
28		Slovenia
29		Spain
30		Sweden
31		Switzerland
32		Turkey
33		Ukraine
34		United Kingdom
35		United States

Due to high level of data desegregation of final 73 and 61 categories of non-durable goods and services, the data were rearranged according to the concept of multistage budgeting (or utility tree). Assuming the weak separability in preferences, the concept allows the determination of consumer preferences only by a subset of a few related commodities, ignoring the impact from less relevant ones.

Generally, the assumed structure of the utility tree was the following:

Aggregation level 1. Consumers allocate expenditures to the most aggregated categories of data, such as food, clothing, fuel and power, medical services, purchased transport and the like (see sub-system 0 in table 6-2).

Aggregation level 2. Subsequently, consumers subdivide consumption expenditures at the most aggregate level into lower levels of aggregation. For example, food is further subdivided into 11 less aggregated expenditures, such as bread and cereals, meat, fish, etc. (see sub-system 0.1 in table 6-2)

Aggregation level 3. Finally, expenditures on bread and cereals are further subdivided into expenditures at the lowest desegregation level, consisting of rice, flour, bread, bakery products, etc. (see sub-system 0.1.1 in table 6-2)

Table 6-2 shows the complete structure of consumer preferences during multy-stage budgeting for 73 categories of 1990 data, and 61 categories of 1993 data. Consumption categories in the multy-stage budgeting were formed in accordance with the standard ICP classification scheme, which, in turn, closely corresponds to the standard national account classification.

Table 6-2. The assumed structure of consumer preferences during multy-stage budgeting.

		1990	1993
Sub-system 0. Aggregate demand system			
1.0.0.	Food, beverages & tobacco		
2.0.0.	Clothing		
3.0.0.	Fuel and power		
4.0.0.	Total medical care & services		
5.0.0.	Purchased transport		
6.0.0.	Communication		
7.0.0.	Services for recreation		
8.0.0.	Total education expenditures		
9.0.0.	Restaurants, cafes & hotels		
Sub-system 0.1. Food, beverages & tobacco			
1.1.0.	Bread and cereals		
1.2.0.	Meat		
1.3.0.	Fish		
1.4.0.	Milk, cheese & eggs		
1.5.0.	Oils and fats		
1.6.0.	Fruits, vegetables & tubers		
1.7.0.	Coffee, tea & cocoa		
1.8.0.	Sugar and sweets		
1.9.0.	Non-alcoholic beverages		
1.10.0.	Alcoholic beverages		
1.11.0.	Tobacco		
Sub-system 0.2. Clothing			
2.1.0.	Men's clothing		
2.2.0.	Women's clothing		
2.3.0.	Children's clothing		
2.4.0.	Clothing accessories		Not estimated
2.5.0.	Clothing, rental and repair		
Sub-system 0.3. Fuel and power			
3.1.0.	Electricity		
3.2.0.	Gas		
3.3.0.	Liquid fuels		
Sub-system 0.5. Purchased transport			
5.1.0.	Local buses, trams & the like		Not estimated
5.2.0.	Long distance transport		Not estimated
Sub-system 0.6. Communication			
6.1.0.	Post services		
6.2.0.	Telephone		
Sub-system 0.1.1. Bread and cereals			
1.1.1.	Rice		
1.1.2.	Flour, other cereals		
1.1.3.	Bread		
1.1.4.	Bakery products, biscuits		
1.1.5.	Noodles, macaroni, spaghetti		
1.1.6.	Cereal preparations		Not estimated
Sub-system 0.1.2. Meat			
1.2.1.	Beef and veal		
1.2.2.	Pork		
1.2.3.	Poultry		
1.2.4.	Lamb, goat & mutton		
1.2.5.	Dried or processed meat, etc.		
Sub-system 0.1.3. Fish			
1.3.1.	Fish fresh/frozen		Not estimated
1.3.2.	Processed fish/seafood, canned, etc.		Not estimated
1.3.3.	Smoked or preserved fish		Not estimated
Sub-system 0.1.4. Milk, cheese & eggs			
1.4.1.	Milk fresh		
1.4.2.	Milk preserved		
1.4.4.	Cheese		
1.4.5.	Eggs & egg products		

(Continued on the next page)

Table 6-2. The assumed structure of consumer preferences during multy-stage budgeting.

		1990	1993
Sub-system 0.1.5. Oils and fats			
1.5.1.	Butter		
1.5.2.	Margarine, edible oils & lard		
Sub-system 0.1.6. Fruits, vegetables & tubers			
1.6.1.	Fresh fruits		
1.6.2.	Dried, frozen, preserved, juices		
1.6.3.	Fresh vegetables		
1.6.4.	Dried, frozen, preserved vegetables		
1.6.5.	Tubers, including potatoes		
Sub-system 0.1.7. Coffee, tea & cocoa			
1.7.1.	Coffee		
1.7.2.	Tea		
1.7.3.	Cocoa		
Sub-system 0.1.8. Sugar and sweets			
1.8.1.	Sugar		
1.8.2.	Jam		
1.8.3.	Chocolate		
1.8.4.	Condiments		Not estimated
Sub-system 0.1.9. Non-alcoholic beverages			
1.9.1.	Mineral water		
1.9.2.	Soft drinks		
Sub-system 0.1.10. Alcoholic beverages			
1.10.1.	Liquors & spirits		
1.10.2.	Wine, cider		
1.10.3.	Beer		
Sub-system 0.5.1. Local buses, trams & the like			
5.1.0.	Taxi		Not estimated
5.2.0.	Buses, trams, etc.		Not estimated
Sub-system 0.5.2 Long distance transport			
5.1.0.	Railway transport		Not estimated
5.2.0.	Air transport		Not estimated

Note: classified by the author, using detailed descriptions of the ICP's basic heading specifications in Kravis et. al. (1982), pp. 60-66.

Section 6.3. The choice of regression specification.

During the selection of the most appropriate demand function to model international consumption preferences, we considered three options to model consumer preferences: by the Stone's linear demand system, the Rotterdam model, and the Almost Ideal Demand System.

First was ruled out the applications of the Rotterdam system, because it can be used only with first-differenced time series data. The application of the linear expenditure system was also impractical due to its several implicit assumptions, such as the lack of inferior goods, making it applicable only for very broad aggregates of consumption such as food, clothing and the like (Phlips, 1983, p. 129).

This requirement contrasts with the high levels of desegregation of the ICP data. In principle, the original ICP data can be aggregated into the necessary highly aggregated categories, but the study of such aggregates may yield very few interesting results, similarly to the above-mentioned applications of the revealed-preference approach to the aggregated bundles of commodities.

As a result, the Almost Ideal Demand System (AIDS) turned out the only suitable choice for the highly disaggregated international cross-sections from the ICP. However, the choice of the demand system was still associated with one complication. Besides income and own-price terms, the conventional specification of the AIDS includes terms for cross-price elasticities.

The inclusion of the cross-price elasticities in sub-systems with large number of related commodities (such as the subsystem 0.1 with 11 related commodities) might leave too few degrees of freedom in 1990 sample of 22 OECD countries. Therefore, originally we decided to estimate a simplified version of the AIDS, which contains only the income terms (at the relevant level of the multy-stage budgeting), and the own price term[26].

$$w_i = \beta_0 + \beta_1 \ln\left(P_i/\overline{P}\right) + \ln\left(M/\overline{P}\right) \tag{6-3}$$

where w_i denotes a share in nominal expenditures (that is, expressed in national currency) of commodity i in the total expenditures at relevant consumption sub-system, P_i is purchasing power parity for commodity i, $\overline{P}$ is an aggregated purchasing power parity for the whole sub-system, and M is total consumption expenditures (in national currencies) on goods, included in relevant consumption sub-system.

Note that the simple specification (6-3) is essentially identical with the one, previously adopted in the cross-sectional demand analyses with the ICP data in Kravis *et. al* (1982, p. 357) and reproduced in equation (6-1), with the important change in the dependent variable. Instead of real quantities in (6-1), our specification includes nominal shares, thus incorporating the standard additivity property of utility functions. However, the specification of independent variables is essentially the same in (6-1) and (6-3).

---

[26] In the context of the present study, the restriction of cross-price elasticities to zero assumes away international differences in, say, substitution effects. However, given the persistent difficulties "in finding a robust classification of substitutes and complements" (Deaton, Muelbauer, 1980, p. 79) even at the most aggregated categories of consumption, it may be expedient in the present to deal with a relatively simple demand system.

Section 6.4. Identified countries-outliers from the estimated majority pattern in consumer preferences.

Table 6-3 reports results of applying the new outlier test to 1990 cross-sections of OECD countries. Statistically significant deviations from the estimated consumption pattern are identified by small p-values. The table contains standardized prediction errors for observations in identified most outlying subsets of data (normalized by the Wallace transformation), as well as the test statistic u^*, equal to the largest absolute value of reported standardized prediction residuals.

As tabulated in table 5-1, an approximate significance level for the test (at 5% significance level and 20 observations) is 3.14. However, we calculated the exact distribution of the test statistic with 500 replications for each cross-section, and the test p-values were derived by comparing the actual u^* with sorted absolute values of simulated test statistics.

At the 5% level of significance, regression outliers were identified in 17 cross-sections, representing only slightly above 20 percent of all cross-sections. One surprising result is that there is just one case of regression outliers for services (rent and repair of clothing in Japan). Moreover, in 12 cases there was just one outlying observation. Conversely, the largest number of national outlier took place in the demand for fish and pork, which were positive in as many as 9 countries.

A number of identified outliers in consumption corresponded to a priori expectations about national preferences for specific goods. Such cases included relatively distinct Japanese preferences for bread/cereals, fish, and, conversely – negative preferences for meat, milk and oils, the positive inclination of Italians of noodles, or American bent for juices (which are included in preserved fruits).

There is an unusually large value of standardized prediction error for tobacco in Luxembourg, but this outcome may have nothing with unusual consumption preferences of domestic residents. OECD (1994) clarified this odd result, attributing it largely to large tobacco purchases by non-residents. This example shows that large residuals may incorporate a variety of unusual factors that may reflect not only demand preferences, but also measurement errors, or peculiarities in market regulations for a specific product or service. Given such a wide-ranging interoperation of regression outliers in consumption, it is quite surprising that only in one-fifth of all considered cases we found countries that deviated significantly from the rest of the analyzed sample.

Table 6-3. Outlier test statistics after estimating the Almost Ideal Demand System in OECD countries, 1990.

	AUS	AUT	BEL	CAN	DNK	FIN	FRA	DEU	GRC	IRL	ITA	JPN	LUX	NLD	NZL	NOR	PRT	ESP	SWE	TUR	GBR	USA	u^*	p-value	Alt. p-value
Food									3.73														3.73	0.004	
Clothing																				2.55			2.55	0.392	
Fuel													2.97			2.99							2.99	0.130	
Medical services			2.90				4.04	2.86	-2.36			2.63		2.65		2.89			3.12				4.04	0.678	
Transportation															3.18						2.58		3.18	0.404	
Communication			-3.68		3.17			3.07	3.72		-2.81			2.68	4.65					3.70			4.65	0.300	
Recreation	2.61																						2.61	0.368	
Education					2.84																		2.84	0.178	
Restaurants, hotels																		2.91					2.91	0.132	
Bread and cereals												5.04											5.04	0.000	0.005
Meat												-3.35											3.35	0.020	0.121
Fish			4.17				3.68		4.76		4.17	6.64				4.80	6.36	5.62	3.87				6.64	0.006	0.080
Milk												-3.11											3.11	0.068	0.364
Oils												-3.07											3.07	0.092	0.040
Fruits and veget.								-2.56					-2.51							3.27			3.27	0.488	
Coffee								5.34															5.34	0.000	
Sugar and sweets													-2.29				-2.51						2.51	0.754	
Alcohol																	-2.24				2.41		2.41	0.822	
Non-alcoholic bever.	-2.48							2.49		-3.07										-2.75			3.07	0.520	
Tobacco													5.96										5.96	0.000	
Men's clothing	-4.13		-3.66					-3.11			4.70			-4.84	-4.75	2.71		-2.88		-3.35			4.84	0.262	
Women's clothing														3.36									3.36	0.028	
Children's clothing												2.49											2.49	0.526	
Accessories		3.02																					3.02	0.092	
Clothing repair												4.52											4.52	0.000	
Electricity											-2.23												2.23	0.822	
Gas														3.03									3.03	0.084	
Liquid fuels	-2.21											-2.17	2.29	-2.49	-2.89						-2.63		2.89	0.880	
Local transport													2.35				2.71						2.71	0.676	
Long-dist. Transport													-2.35				-2.71						2.71	0.676	
Post			2.88								2.78			2.52									2.88	0.500	
Telephone			-2.88								-2.78			-2.52									2.88	0.500	

Table 6-3 (continuation). Outlier test statistics after estimating the Almost Ideal Demand System in OECD countries, 1990.

	AUS	AUT	BEL	CAN	DNK	FIN	FRA	DEU	GRC	IRL	ITA	JPN	LUX	NLD	NZL	NOR	PRT	ESP	SWE	TUR	GBR	USA	u^*	p-value	Alt. p-value
Rice												5.95					4.87						5.95	0.000	
Flour, other cereals		3.60				3.69									3.47						3.95		3.95	0.014	
Bread				-3.87		-2.76			-3.98		-3.39	-3.61									-2.96	3.66	3.98	0.230	
Bakery products			2.77	3.14		2.85						-3.84		3.37			-3.42			-4.00		-4.34	4.34	0.204	
Noodles, macaroni											3.82												3.82	0.004	
Cereal prepar.												3.52											3.52	0.008	
Beef and veal									2.95											3.18			3.18	0.124	
Pork		4.01		4.37	4.78			3.26				4.22		4.65		5.03	4.16		4.41				5.03	0.044	0.086
Lamb, goat & mutton	3.09								3.66	2.58					2.57	3.52		2.73		4.26	2.72		4.26	0.422	
Poultry				2.44		-2.51		-2.29								-2.58	2.49		-2.88				2.88	0.776	
Processed meat									-2.94								-2.32			-2.75			2.94	0.720	
Fish fresh/												-2.16					-2.21		-2.23	2.98			2.98	0.890	
Processed fish					-2.23	2.60				-2.33	-2.75				2.30		-2.39		2.95			-2.61	2.95	0.826	
Preserved fish																						4.95	4.95	0.000	
Milk fresh																				-2.23			2.23	0.834	
Milk preserved									4.32														4.32	0.000	
Cheese	-2.98	-2.46								-3.19						-3.12		-3.17		4.05	-2.56	-3.73	4.05	0.546	
Eggs																		2.92					2.92	0.118	
Butter										2.31													2.31	0.714	
Margarine										-2.31													2.31	0.714	
Fresh fruits	-2.33			-2.89								-2.39		-2.26	-2.25						-2.05	-3.35	3.35	0.970	
Dried fruits, juices																						3.50	3.50	0.012	0.646
Fresh vegetables				-4.40		-4.00	3.46					4.39				-3.76						-3.76	4.40	0.020	0.065
Dried vegetables												2.93		2.99									2.99	0.120	0.313
Tubers, incl. potatoes				3.37						4.45		-3.70				-3.08	3.07			-2.91	3.72		4.45	0.122	0.566
Coffee												-2.88											2.88	0.162	0.334
Tea										5.43		5.39			4.07					5.96	4.84		5.96	0.000	0.000
Cocoa									3.18			3.36					3.69						3.69	0.054	0.015
Sugar																	3.04						3.04	0.102	
Jam											-2.75				3.08								3.08	0.246	
Chocolate	-2.39																						2.39	0.642	
Condiments	2.94											2.84											2.94	0.166	

Table 6-3 (continuation). Outlier test statistics after estimating the Almost Ideal Demand System in OECD countries, 1990.

	AUS	AUT	BEL	CAN	DNK	FIN	FRA	DEU	GRC	IRL	ITA	JPN	LUX	NLD	NZL	NOR	PRT	ESP	SWE	TUR	GBR	USA	u^*	p-value	Alt. p-value
Mineral water	2.90	3.58	3.56				4.58	4.15			3.35		4.64										4.64	0.172	
Soft drinks	-2.90	-3.58	-3.56				-4.58	-4.15			-3.35		-4.64										4.64	0.172	
Liquors & spirits	-2.65					3.20										3.08			2.98	-3.21			3.21	0.328	
Wine, cider		2.96		-2.95		-3.75	3.39					-3.25	3.90			-3.28	4.40		-3.06				4.40	0.136	
Beer	3.51			3.05			-3.33					3.42	-2.72		3.26		-3.94			3.12		2.80	3.94	0.242	
Taxi									-2.61														2.61	0.372	
Buses									2.61														2.61	0.372	
Railway transport												2.33											2.33	0.708	
Air transport												-2.33											2.33	0.708	

The sample coverage from 1993-benchmark study is more heterogeneous, so one may expect that there will be more identified cases of regression outliers. And true, the number of cross-sections with identified outliers at 5% significance level increased to 27, accounting for slightly less than half of all cross-sections.

There are several recurring peculiar cases from the 1990 cross-section, such as the strong preference for food in Greece, fish and rice – in Japan (and negative – for meat and bread), tobacco – in Luxembourg, noodles – in Italy, preserved fruits in the United States. The largest number of statistically significant outliers in a single cross-section is four (negative preferences for coffee Ireland, Japan, Russia, and Turkey, which mirrors their positive preferences for tea). Other interesting cases include positive preferences for dress in Russia and for alcoholic beverages (as compared with wine and beer) in all former Soviet republics in the study – Russia, Belorussia and Ukraine.

In addition to identifying commodities with diverse consumer preferences, we could also identify commodities that apparently exhibited unusually strong complimentary relationships. Some striking pairs of complimentary preferences are wine and beer in the 1990 sample, and coffee – tea in the 1993 sample. The complimentary relationship is revealed by opposite signs of estimated prediction residuals. For example, estimated preferences for wine were negative in Canada, Japan, Norway and Sweden, but simultaneously were positive in these countries for beer. The opposite pattern took place in France, Luxembourg and Portugal where wine was preferred at the expense of beer.

Similarly, there were negative prediction residuals at the cross-section for coffee in Ireland, Japan, Russia and Turkey, and at the same time – approximately the same positive prediction residuals at the cross-section for tea. There are similar interdependencies between related commodities for individual countries as well. For example, Japanese consumers demonstrate unusually strong positive preferences for bread and cereals (due to the inclusion of rice in this category) and fish, and negative preferences for meat and milk, and the pattern occurred both in 1990 and 1993 samples. Similarly, Turkish consumers in 1993 sample showed strong positive preferences for beef and lamb, and negative ones – for pork, poultry and dried meat. The frequent occurrence of such interdependencies between commodities indicates that our restricted specification of the AIDS demand system may miss some important factors in explaining international variations in demand preferences, since it includes only own-price effects, and assumes that all cross-price elasticities are negligible.

In it noteworthy that in many cases of apparent complimentary relationship among related commodities we simultaneously identified regression outliers. One may presume that once we account for additional factors of demand, such as cross-price elasticities, the number of unexplained peculiarities in consumer demand may decrease. To verify this conjecture, we estimated the extended version of the AIDS model, which, in addition to the basic specification (6-3), also included all cross-price elasticities from a relevant demand sub-system. For example, we augmented the original specification for bread and cereals with relative prices for meat, fish, milk, oils and so on (see the composition of subsystem 0.1 in table 6-2). The primary conjecture was that the extended specification with extra price terms may explain somehow the unusually pattern of Japanese demand for bread and cereals in the restricted original specification.

Table 6-4. Outlier test statistics after estimating the Almost Ideal Demand Systems in OECD and post-socialist countries, 1993.

	BLR	BEL	CAN	HRV	DNK	FRA	DEU	GRC	HUN	ISL	IRL	ITA	JPN	LUX	NLD	NZL	PRT	ROM	RUS	SVK	ESP	CHE	TUR	UKR	GBR	USA	u^*	p-value	Alt. p-value
Food								3.86																			3.86	0.005	0.090
Clothing																			3.90				3.34				3.90	0.020	0.859
Fuel														3.33													3.33	0.035	0.212
Medical services						2.66		-2.57									-2.42	-2.51									2.66	0.707	
Transportation								4.00								3.03		-2.90		-3.45							4.00	0.212	
Communication				2.63												2.41											2.63	0.717	
Recreation																											2.40	0.778	
Education								-3.12																			3.12	0.070	
Restaurants, hotels																					4.36						4.36	0.000	0.010
Bread and cereals				2.82			3.08	-3.00					3.17										3.47				3.47	0.283	
Meat													-3.15														3.15	0.045	0.060
Fish													5.03														5.03	0.000	0.000
Milk													-3.18														3.18	0.060	0.283
Oils																							3.47				3.47	0.020	0.152
Fruits and veget.																							2.56				2.56	0.525	
Coffee							3.11																				3.11	0.050	
Sugar and sweets					2.83					2.62																	2.83	0.515	
Non-alcoh. bever.										4.23			3.30													3.62	4.23	0.045	
Alcohol																							-2.80				2.80	0.293	
Tobacco														5.98													5.98	0.000	
Men's clothing												4.00			-4.26												4.26	0.000	
Women's clothing															4.05												4.05	0.000	
Children's clothing										2.95																	2.95	0.116	
Clothing repair													3.92														3.92	0.000	
Electricity												-2.47															2.47	0.606	
Gas															2.43												2.43	0.707	
Liquid fuels				3.55																							3.55	0.020	
Post		2.46										2.74			2.69		-2.85				-2.42		-3.21				3.21	0.657	
Telephone		-2.46										-2.74			-2.69		2.85				2.42		3.21				3.21	0.657	

Table 6-4(continuation). Outlier test statistics after estimating the Almost Ideal Demand Systems in OECD and post-socialist countries, 1993.

	BLR	BEL	CAN	HRV	DNK	FRA	DEU	GRC	HUN	ISL	IRL	ITA	JPN	LUX	NLD	NZL	PRT	ROM	RUS	SVK	ESP	CHE	TUR	UKR	GBR	USA	u^*	p-value	Alt. p-value
Rice													7.77														7.77	0.000	
Flour, other cereals																		3.06					2.76	3.02			3.06	0.313	
Bread													-4.29														4.29	0.000	
Bakery products		2.89																									2.89	0.182	
Noodles, macaroni												3.66															3.66	0.015	
Beef and veal																							3.42				3.42	0.010	0.055
Pork																							-2.48				2.48	0.616	
Lamb, goat										5.65													5.64				5.65	0.000	0.000
Poultry			3.61				-2.08		2.42		2.16		-2.76			2.21	2.74			2.42	2.08		-2.51		3.11		3.61	1.000	
Processed meat																							-3.41				3.41	0.015	0.323
Milk fresh																							-2.66				2.66	0.465	
Milk preserved								4.78																			4.78	0.000	
Cheese	-4.14					3.21		3.16	-3.23			4.40	-4.68	2.92			-3.57		-2.83			2.69		-3.79			4.68	0.384	
Eggs	4.47																										4.47	0.000	
Fresh fruits																											2.58	0.485	
Dried fruits, juices										3.52																3.36	3.52	0.020	0.242
Fresh vegetables													2.65						2.70								3.04	0.374	
Dried vegetables				3.42									3.29		3.75												3.75	0.020	0.162
Tubers & potatoes	3.31																			3.03				3.74			3.74	0.060	0.404
Coffee											-3.62		-3.95						-3.58				-4.05				4.05	0.010	0.016
Tea											4.57		3.61						3.95				4.64				4.64	0.005	0.046
Cocoa						3.41		3.50		2.50			4.17				4.57	3.26				2.98			2.54		4.57	0.697	
Sugar	3.70																										3.70	0.010	
Jam																2.65	3.07										3.07	0.919	
Chocolate	-2.65																-2.53										2.65	0.616	
Mineral water	3.02																										3.02	0.131	
Soft drinks	-3.02																										3.02	0.131	
Liquors & spirits	3.52																		3.68					3.29			3.68	0.045	0.596
Wine, cider																	2.58					2.59					2.59	0.556	
Beer	-4.02		2.84			-2.86							3.39				-3.49		-2.66			-2.93	2.67	-2.64		3.46	4.02	0.475	

Similar extensions were applied to several other commodities for which we originally identified outlying national consumption patterns. Specifically, in the 1990 cross-sections with extended regression specification included the demand for milk, fish, meat, oils, pork, preserved fruits, fresh vegetables, preserved vegetables, potatoes, coffee, tea and cocoa. In the 1993 cross-sections, we considered cross-price effects in the demand for aggregate food, clothing, fuel, restaurants, meat, fish, milk, oils, beef, lamb, dried meat, preserved fruits, preserved vegetables, potatoes, coffee, tea, liquors and spirits. P-values for the extended specification of the AIDS demand systems are given in the last column of tables 6-3 and 6-4.

First consider results for the 1990 sample. In a number of cases, the inclusion of extra price terms drastically increase p-values of the test. For example, though Japan remained the most outlying observation in the demand for milk, the test's p-value jumped from 0.068 to 0.364. Similarly, the p-value for preserved fruited (in both cases the single outlier was USA) increased from 0.012 to 0.646. Consequently, once we account for cross-price effects on milk and preserved fruits from related commodities, there is nothing exceptional in the demand for these commodities in Japan and USA, respectively. The same change in the test result took place in the demand for potatoes and coffee.

However, the significance of the test result remained generally unaffected in cases of bread/cereals (the sole outlier – once again Japan), fish (though the number of outliers decreased from nine to just four – Norway, Turkey, Portugal and Japan), oils (in both cases outlier – Japan), pork (but the number of outliers dropped from nine to just one – Norway) and fresh vegetables (with the same six outliers). Finally, there was still very low p-value in the case of tea (in both cases – less than 0.001, with five outlying countries), and cocoa, for which the p-value even further decreased in the extended regression (from 0.054 to 0.015), and one more country (France) entered the pool of outlying observations.

Results for the 1993 sample were broadly similar. The test's p-value increased sharply in cases of clothing (explaining away the previously identified exceptional results for Russia and Turkey), fuel (in both specifications the outlier was Luxembourg), milk (once again – Japan), oils (Turkey), dried meat (Turkey), preserved fruits (USA and Iceland), potatoes (with the number of outlying countries was reduced from Belourussia, Slovakia and Ukraine to only Ukraine), and spirits (where Russia remained the sole outlier). On the other hand, p-ratios remained largely unchanged in cases of aggregate food (with Romania substituting for previously identified Greece), restaurants (Spain), meat (Japan), fish (Japan), beef (Turkey), lamb (Island, Turkey), preserved vegetables (Croatia, Japan, and the Netherlands), and, finally, tea and coffee, with still opposite preferences for these goods in Ireland, Japan, Russia and Turkey.

In sum, the extension of the original restricted specification of the AIDS system helped to explain away a number of national peculiarities in the demand for milk, preserved fruits, potatoes and other commodities. Therefore, originally identified unusual demand patterns for these goods could be attributed to specific relative price effects rather than genuine peculiarities in consumer preferences.

Still, for a number of commodities, such as cereals and milk in Japan, or tea in Ireland, Japan and Turkey, even the extended set of economic variables failed to account for unusually large demand levels in these countries. Therefore, the demand for these commodities appears to be strongly affected by cultural, historical factors that could not be explained by conventional economic factors.

Section 6.5. Conclusions.

In this chapter we illustrated the application of the regression outlier test, introduced in chapter 5 with data on international consumption patters. An attractive feature of these data is that they allow one to verify the plausibility of reported regression outliers, since distinct consumption preferences may be generally well known, such as the relatively high preference in Japan for fish compared with meat, milk or oils.

After applying the restricted version of the Almost Ideal Demand System, we found mostly homogenous consumption patterns in the cross-section of 22 OECD countries in 1990. In only one-fifth of analyzed goods and services we found national observations that deviated significantly from the rest of the analyzed sample. Moreover, such distinctive consumption patterns were identified in just one country.

Conversely, when the analyzed sample was extended to 35 countries due to the inclusion of several East-European countries, the diversity of consumer preferences increased substantially, and in almost half of studied cross sections we found national observations that did not fit well with the estimated consumption pattern. In addition, there were a number of cases when identified outliers from the 1990 sample again appeared distinctive in 1993, indicating that the outliers were not due to transitory factors.

The application of the outlier test may not be limited to the study of consumer preferences. Due to the test's emphasis on residual for a particular observation, the test can be also applied for the estimation of technology from output data, once the impact of inputs is accounted for, as in the conventional association of technology with the 'Solow residual'.

Unfortunately, due to the small number of countries in our database for productivity estimation the additional application of the outlier test was not feasible.

Provided that data coverage is large enough, it may be interesting to compare productivity estimates by the index number approach and the outlier test, introduced in chapter 5.

References.

Antille Gerald, and Ritschard Gilbert (1990), "Robust and Classical Outlyingness Indicators: a Simulation Study". *Communications in Statistics – Simulations.* Vol. 19, 1990, p. 505-512.

Ark, B. Van (1990), "Comparative Levels of Manufacturing Productivity in Post-war Europe: Measurement and Comparisons", *Oxford Bulletin of Economics and Statistics*, 52, Number 4, pp. 343-374.

Ark, B. Van (1992), "Comparative Productivity in British and American Manufacturing", *National Institute Economic Review*, Number 142, November.

Ark, B. Van (1995), "Manufacturing prices, productivity and labor costs in five economies", *Monthly Labor Review*, Bureau of Labor Statistics, July, pp. 56-72.

Ark, B. Van, and D. Pilat. (1993). "Productivity Levels in Germany, Japan and the United States: Differences and Causes", *Brookings Papers on Economic Activity: Microeconomics*, 1-48.

Ark, Bart, (1996). "Productivity and Competitiveness in Manufacturing: A Comparison of Europe, Japan and the United States." In: Wagner, Karin and Ark B. Van (eds.), *International Productivity Differences: Measurement and Explanations.* Amsterdam; New York and Oxford: Elsevier Science, North-Holland, 23-52.

Arrow K., H. Chenery, B. Minhas, and R. Solow (1961). "Capital-Labor substitution and Economic Efficiency". *Review of Economics and Statistics*, 45, pp. 225-247.

Balassa, B. (1962), "An Empirical Demonstration of Classical Comparative Costs Theory", *Review of Economics and Statistics*, August, 45.

Barnett Vic, and Lewis Toby. (1994). *Outliers in statistical data.* 3rd edition, John Wiley.

Barro Robert, and Xavier Sala-i-Martin (1995). Economic Growth. McGraw-Hill.

Baumol W. J. (1986) "Productivity Growth, Convergence, Welfare: Whet the Long-Run Data Show". American Economic Review, vol. 76 (5), pp. 1072-1085.

Belsley David., Kuh Edwin, and Welsch Roy. Regression Diagnostics (1980). John Wiley.

Bernard, Andrew B., and Jones, Charles I. (1996). "Comparing Apples to Oranges: Productivity Convergence and Measurement across Industries and Countries." *American Economic Review*, 86, 1216-1238.

Bhagwati J. (1972a) "The Heckscher-Ohlin Theorem in Multi-Commodity Case", *Journal of Political Economy*, 80 (5), pp. 1062-1055.

Bhagwati J. (1972b), Comment on D. J. Daly, in D. J. Daly (ed.)., "*International Comparisons of Prices and Output*", Studies of Income and Wealth, Vol. 27, New York, Columbia University Press for NBER, pp. 131-137.

Bhagwati, J. (1964), "The Pure Theory of International Trade: A Survey", *Economic Journal*, March, 74, pp. 1-84.

Blades, D. (1993), "Comparing Capital Stocks", in: A. Szirmai, B. Van Ark and D. Pilat (eds.), *Explaining Economic Growth*, North Holland, Amsterdam, pp. 399-409.

Bowen, H.P., Leamer, E. E. and Sviekauskaus, L. (1987), "Multicountry, Multifactor Tests of the Factor Abundance Theory", *American Economic Review*, 77(5), December.

Caves Douglas W., Christensen Lauritis R., and Diewert Edwin W. (1982). "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers." *Economic Journal*, 92, p. 73-86.

Caves Douglas W., Christensen Lauritis R., Thetheway W. (1981). "U.S. Trunk Air Carriers, 1972-1977: A Multilateral Comparison of Total Factor Productivity." In Cowing Thomas, Stevenson Rodney E. (eds.), *Productivity Measurement in Regulated Industries*: Academic Press, pp. 47-76

Chalfant, James A., and Alston Julian M. "Accounting for Changes in Tastes." Journal of Political Economy, 1988, p. 391-410.

Charnes, A. W.W. Cooper, A. Y. Levin and Seiford (1995), "Data Envelopment Analysis: Theory, Methodology and Applications", Kluwer.

Christensen Lauritis R., Cummings D., Jorgenson Dale W. (1981). "Relative Productivity Levels, 1947-1973: An International Comparison." *European Economic Review*, 16, p. 61-94.

Conover W.J. (1971). *Practical Nonparametric Statistic*. New York: John Wiley.

Cook Dennis, and Weisberg. (1981). *Residuals and influence in regression*. New York, Chapman and Hall.

Deardorff A. (1979) "Weak Links in the Chain of Comparative Advantage", *Journal of International Economics*, 9, pp. 197-209.

Deaton, Angus and Muellbauer, John. (1980a). *Economics and consumer behavior*. Cambridge: Cambridge University Press.

Denison E. F. (1967). *Why Growth Rates Differ*. Washington, D.C., Brookings Institutions.

Diebold F., and Chen C. (1996). 'Testing structural stability with endogenous breakpoint: a size comparison of analytic and bootstrap procedures.' *Journal of Econometrics*, vol. 70, pp. 221-241.

Diewert Edwin (1996). "Axiomatic and Economic Approaches to International Comparisons." Discussion paper, University of British Columbia.

Diewert Edwin (1996). "Axiomatic and Economic Approaches to International Comparisons." Discussion paper, University of British Columbia.

Diewert Edwin. (1981). "The Economic Theory of Index Numbers." In Deaton A. (ed.) *"Essays in the Theory and Measurement of Consumer Behavior in Honor of Sir Richard Stone"*, Cambridge University Press, 163-208.

Diewert, Edwin (1993). "The Early History of Index Numbers." In Diewert E., and Nakamura A. (eds.) *"Essays in Index Numbers Theory."* Vol. 1., North Holland, 33-67.

Diewert, Edwin. (1976). "Exact and Superlative Index Numbers." *Journal of Econometrics*, Vol. 4, 15-145.

Dollar, David; Wolff, Edward N., "Capital Intensity and TFP Convergence by Industry in Manufacturing, 1963 — 1985", in Baumol William J., Nelson Richard R., and Wolff Edward N. (eds.) *"Convergence and Productivity: Cross National Studies and Historical Evidence"* , The Harvard University Press.

Domar, Evsey (1961), "On the measurement of Technical Change", *The Economic Journal*, pp. 709-729.

Donoho D. L., and P. J. Huber (1983). 'The notion of breakdown point'. In: P. Bickel, K. Doksum, and J. L. Hodges (ed.) *"A Festschrft for Erich Lechmann"*, Wadsworth, Belmont, CA.

Dornbusch R., Fisher S., and P.A. Samuelson (1977) "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods", *American Economic Review*, 67 (5), pp. 823-839.

Fisher , Irving (1927), "Making of Index Numbers", Boston: Houghton-Miffin.

Freudenberg M., and Unak-Kesenci D., "French and German Productivity Levels in Manufacturing". In: Wagner K. and B. Van Ark (eds.), *International Productivity Differences: Measurement and Explanations*, North Holland, Amsterdam.

Gentleman J., and Wilk M., "Detecting Outliers II: Supplementing the Direct Analysis of Residuals," Biometrics, vol. 31, 1975, pp. 387-410.

Gilbert M. and Kravis I. B. (1954). *An International Comparison of national Products and the Purchasing Power of Countries*, OECD, Paris.

Golub S. (1994) "Comparative Advantage, Exchange Rates, and Sectoral Trade Balances of Major Industrial Countries", *IMF Staff Papers*, 41 (2), pp. 286-313.

Good David, Nadiri Ishaq, and Sicles Robin (1997), "Index Number and Factor Demand Approaches to the Estimation of Productivity". In Pesaran Hashem, and Schmidt Peter (eds.) "Handbook of Applied Econometrics. Microeconomics". Blackwell, pp. 14-80.

Green William (1997), "Frontier Production Functions". In Pesaran Hashem, and Schmidt Peter (eds.) "Handbook of applied Econometrics. Microeconomics". Blackwell, pp. 81-166.

Gullickson, W. (1995). 'Measurement of Productivity Growth in U.S. Manufacturing', *Monthly Labor Review*, pp. 13-28.

Hadi Ali S., and Simonoff Jeffrey S. (1993). 'Procedures for the identification of multiple outliers in linear models'. *Journal of the American Statistical Association*, Vol. 88, p.1264-1271.

Hadi Ali S., and Simonoff Jeffrey S. "Comment on Paul and Fung (1991)". Technometrics, vol. 32, August 1992, pp. 373-374.

Harrigan James. (1996). "Technology, Factor Suppliers and International Specialization: Estimating the Neo-classical Model." *NBER Working Paper*, 5772.

Hawkins D.M (1991). 'Diagnostics for use with regression recursive residuals.' *Technometrics*, vol. 33, pp. 221-234.

Heston Alan and Summers Robert (1996) 'International Price and Quality Comparisons: Potentials and Pitfalls', *American Economic Review*, Vol. 86, p. 20-24.

Hooper, P. and E. Vrankovich (1995), "International Comparisons of the Levels of Unit Labor Costs in Manufacturing", *International Finance Discussion Papers*, No. 527, Board of Governors of the Federal Reserve System, Washington D.C.

Jones, R.(1957-1958) "Factor Proportions and the Heckscher-Ohlin Theory". *Review of Economic Studies*, 24, pp. 1-10.

Jorgenson, D. and Kuroda M. (1990), "Productivity and International Competitiveness in Japan and the United States, 1960-1985. In Hulten Ch. (Ed), *Productivity Growth in Japan and the United States*, The University of Chicago Press.

Jorgenson, D. W. and M. Kuroda (1992), "Productivity and International Competitiveness in Japan and the United States, 1960-1985", *Economic Studies Quarterly*, Vol. 43, December, pp. 313-325.

Kianifard, Farid, and Swallow William "A Monte-Carlo Comparison of Five Procedures for Identifying Outliers in Linear Regression". Communications in Statistics. Theory and Methods. Vol. 19, 1990, p. 1913-1938.

Kravis Irving, Heston Alan, and Summers Robert (1982), *World Product and Income: International Comparison of Real Gross Product*. The Johns Hopkins University Press, Baltimore and London (for the World Bank).

Kravis Irving, Heston, Alan and Summers, Robert. (1982). *World product and income: international comparisons of gross product*. Baltimore, MD. Johns Hopkins University Press.

Kravis Irving, Kenessey Zoltan, Heston Alan, and Summers Robert (1975), *A System of International Comparisons of Gross Product and Purchasing Power*. The Johns Hopkins University Press, Baltimore and London (for the World Bank).

Kravis, Irving (1976). "A Survey of International Comparisons of Productivity", *The Economic Journal*, No. 86, March.

Kuroda M. (1996). 'International Competitiveness and Japanese Industries, 1960-1985'. In: Wagner K. and B. Van Ark (eds.), *International Productivity Differences: Measurement and Explanations*, North Holland, Amsterdam.

Kuroda, M. (1996), "International Competitiveness and Japanese Industries, 1960-85", in: K. Wagner and B. Van Ark (eds.), *International Productivity Differences: Measurement and Explanations*, North Holland, Amsterdam, pp. 107-141.

Landsburg, Steven E. "Taste Change in the United Kingdom, 1900-1955." Journal of Political Economy, (February 1981), p. 92-104.

Leamer E. (1992) "Testing trade Theories", NBER Working Paper #3957, January.

MacDougall, G. D. A. (1951), "British and American Exports: A Study Suggested by the Theory of Comparative Costs, Part I", *Economic Journal*, December, 61, pp. 697-720.

MacGilvray J. and Simpson, D. (1973), "The Commodity Structure of Anglo-Irish Trade", *Review of Economics and Statistics*, November, 55, pp. 451 – 458.

Maddison A. (1952), 'Productivity in Canada, the United Kingdom, and the United States', *Oxford Economic Papers*, October.

Mankiw, N. Gregory; Romer David, and Weil David (1992). "A Contribution to the Empirics of Economic Growth." Quarterly Journal of Economics, vol. 107, pp. 407-438.

McKinsey (1993), *Manufacturing Productivity*, McKinsey Global Institute, Washington D.C.

Okada Y. (1997) 'Diversity in Technological Support: Industrial Differences'. In: *Japan's Industrial Technology Development', (*ed. By Okada Y.). Rakmo Press, Tokyo, pp. 41-60.

OECD (1987), *Purchasing Power Parities and Real Expenditures*, 1985, Paris.

OECD (1992), *Purchasing Power Parities and Real Expenditures*, 1990, Vol. 1, EKS Results, Paris.

OECD (1993), *Purchasing Power Parities and Real Expenditures*, 1990, Vol. 2, GK Results, Paris.

OECD (1995), *The OECD STAN Database for Industrial Analysis, 1974-1993*, Paris.

OECD (1996), *Purchasing Power Parities and Real Expenditures*, 1993, Vol. 1, EKS Results, Paris.

Mahony (1992), "Productivity Levels in British and German Manufacturing Industry", *National Institute Economic Review*, Number 139, February, pp. 46-63.

Paige, D. and G. Bombach (1959), *A Comparison of National Output and Productivity*, OEEC, Paris.

Pilat Dirk (1996), 'Labor Productivity Levels in OECD Countries: Estimates for manufacturing and Selected Service Sectors', OECD Working Paper N. 169.

Pilat, D. and D.S. Prasada Rao (1996), "Multilateral Comparisons of Output, Productivity and Purchasing Power Parities in Manufacturing", *Review of Income and Wealth*, Vol. 42, No. 2, pp. 1-18.

Pilat, Dirk (1996). "Labor Productivity Levels in OECD Countries: Estimates For Manufacturing and Selected Service Sectors." *OECD, Economics Department Working Paper*, No. 169.

Rostas, L. (1948). *Comparative Productivity in British and American Industry*, Cambridge.

Rousseeuw, Peter and Hubert M. (1997). Recent Development in PROGRESS. Manuscript, 1997.

Rousseeuw, Peter, and Leroy Annie. (1987). Robust regression and outlier detection. John Wiley.

Rousseeuw, Peter, and Zomeren, B.C. (1990).'Unmasking multivariate outliers and leverage points'. *Journal of the American Statistical Association*. Vol. 85, pp. 633-639.

Rousseeuw, Peter. (1984). 'Least median of squares regression', *Journal of the American Statistical Association*, Vol. 79, pp. 871-880.

Rousseeuw, Peter. (1997). 'Introduction to positive-breakdown methods'. In Maddala G. S., Rao C. R. (ed.) "*Robust Inference*", Elsevier, pp. 101-121.

Samuelson P.A. (1938), 'A note on the pure theory of consumer behavior', *Economica*, Vol. 5, pp. 61-71.

Smith A.D., Hitchens D. M. W. N. and Davies S. W. (1982). *International Industrial Productivity*, NIESR, Cambridge.

Srikantan, K. S. (1961). 'Testing for a single outlier in regression model'. *Sankhya*, Ser. A. Vol. 23, pp. 251-260.

Stern, R. M. (1962), "British and American Productivity and Comparative Costs in International Trade", *Oxford Economic Papers*, October, 14, pp. 275-296.

Torstensson, Johan (1996) "Technical Differences and Inter-Industry Trade in the Nordic Countries", *Scandinavian Journal of Economics*, 98 (1), pp. 93-110.

Vanek Y. (1968) "Factor Proportions Theory: The N-Factor Case" *Kyklos*, 21 (4), pp. 749-756.

Viner J. (1964) “Studies in the Theory of International Trade”, George Allen & Unwin.

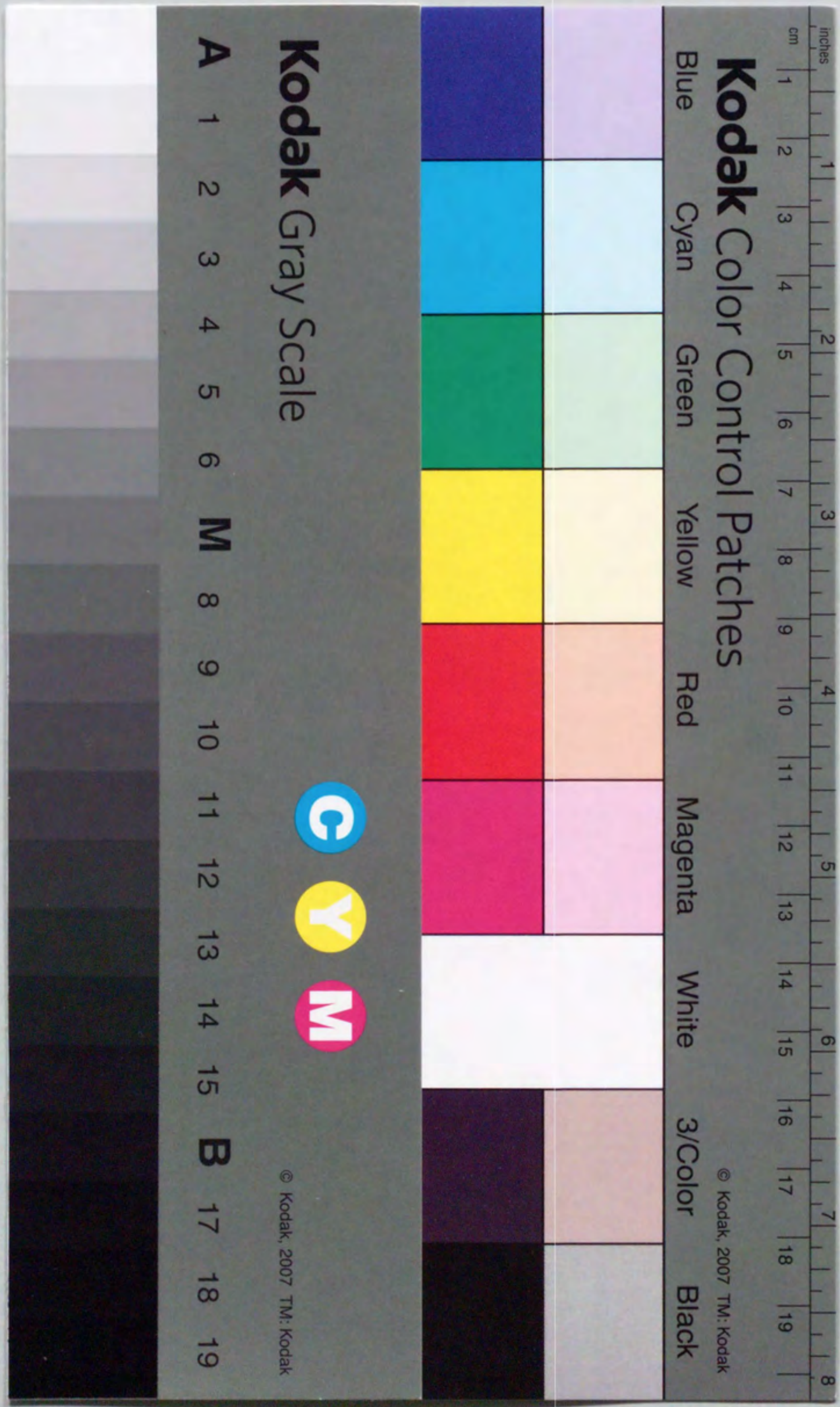
Wagner K. and B. Van Ark (eds.), *International Productivity Differences: Measurement and Explanations*, North Holland, Amsterdam.

Wallace D. L. (1959). “Bounds on Normal Approximations to Student’s and the Chi-Distributions”. *Annals of Mathematical Statistics*, vol. 30, pp. 1121-1130.

Walsh C.M. (1901). *The Measurement of General Exchange Value*. Macmillan and Co., New York.

West E. C. (1971). *Canada-United States Price and Productivity Differences in Manufacturing Industries, 1963*. Economic Council of Canada.

Woodland A. D. (1982) “International Trade and Resource Allocation”, North Holland.


Kodak Color Control Patches
Blue Cyan Green Yellow Red Magenta White 3/Color Black
© Kodak, 2007 TM: Kodak
Kodak Gray Scale
A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19
C Y M
© Kodak, 2007 TM: Kodak