Agricultural Research Policy

International Quantitative Perspectives

Edited by

Philip G. Pardey, Johannes Roseboom and Jock R. Anderson

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011–4011, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1991

Printed in Great Britain at the University Press, Cambridge

A catalogue record of this book is available from the British Library

Library of Congress cataloguing in publication data available

ISBN 0 521 40009 0 hardback

PART II

Measuring Agricultural Research and Economic Development

Chapter 5

Internationally Comparable Growth, Development, and Research Measures

Barbara J. Craig, Philip G. Pardey, and Johannes Roseboom

Constructing measures of real economic activity for the purpose of making international comparisons is a useful but tricky exercise. Unless data are collected specifically for the problem at hand, the resulting series may only poorly measure the variable of interest. Available data may not provide uniform coverage of the countries or periods of interest or may be too broadly or narrowly defined. In the analysis of agricultural research and development patterns, problems of data availability and quality are compounded by the need to recast value and volume aggregates into units that can be meaningfully compared over time and across countries.

When the data on a series of interest are aggregate values measured in local currency units, the aggregates must typically be deflated to take account of changes in the local price level and converted to a common currency in such a way as to provide an accurate picture of their real value or volume. Both the choice of appropriate converters and deflators and the order in which these two operations are performed matter and thus will, in many instances, have substantial effects on the interpretation of the resulting real-value series.

Data series that are reported directly as quantities or volumes often appear to relieve analysts of the problems of turning nominal values into real ones; nevertheless, subtle but substantive issues of comparability are likely to persist. With volume aggregates, some unweighted and perhaps undesirable aggregation is implicit in the data collection process.

In this chapter the strengths and weaknesses of data used throughout this volume are discussed, and the systematic approach taken to enhance comparability is described. Since there are some insurmountable obstacles in these data sets, evidence on the effects of using less than ideal aggregation procedures is presented. This evidence will both aid the interpretation of imperfect data and provide some sense of the boundaries within which the truth lies.

In section 5.1 we describe ideal aggregation procedures. Some practical options for translating preaggregated data into real value or volume measures are discussed in section 5.2 along with selected evidence on the consequences of using different conversion, deflation, and scaling procedures. In section 5.3 we discuss the specific concepts and measurement issues that underpin the agricultural statistics used throughout this volume and conclude in section 5.4 with a brief review of the definitional, measurement, and practical issues involved in constructing agricultural research indicators.

5.1 AGGREGATION

In a multidimensional framework, the construction of comparable aggregates measuring real economic activity always involves two distinct steps. Index number theory tells us to begin with disaggregated data on prices and quantities in each country to calculate directly a real quantity index.[1] To translate the resulting index into an aggregate that can be compared over time and across countries, some scaling factor — based again on disaggregated prices and quantities — must be applied to the base country and/or time period. One of the major problems of making international comparisons lies in the shortage of systematically disaggregated data.

Index number theory, informed by neoclassical models of economic behavior, argues for aggregating real quantities using price weights that are most specific to the economic activity and agents whose behavior is being summarized (Drechsler 1973). For constructing indices, representative or characteristic price vectors need not replicate local *absolute* price levels, but they should reflect local *relative* prices. Otherwise, one may fail to distinguish between changes in the size of the real commodity basket and changes in the composition of the basket. Even when analyzing sectors or entire countries with badly distorted prices — whether due to trade restrictions, price controls, subsidies, or the like — it is important to use the prices actually faced by economic agents when forming the real aggregate. When constructing comparable international aggregates, it is still desirable to use value weights that are representative but, in this context, accounting for differences in absolute price levels is necessary as well.

5.1.1 Temporal Indices

Contemporary international data sets span years of high price volatility, so the pitfalls of using value aggregates denominated in current local currency units are obvious. To compare commodity baskets produced in different periods, index number theory provides arguments for using timely local prices as weights in the construction of indices that have changing rather than fixed price weights. Changing weights allow one to capture shifts over time in local relative prices, which influence changes in the composition of local commodity baskets. Consequently, discrete approximations of the Divisia index (Divisia 1928) are to be preferred to the more commonly used fixed-weight Laspeyres index; they are less likely

---

[1] For a useful discussion of index number issues in the context of international comparisons, see Caves, Christensen, and Diewert (1982) or Craig and Pardey (1990a).

to confound changes in the size of the commodity basket with changes in its composition.

There are several possible discrete approximations of the Divisia index. The most commonly used are the Laspeyres and the Törnqvist-Theil approximations:

$$\text{Laspeyres:}\quad I_t^{DL} = I_{t-1}^{DL}\left[\frac{1 + P'_{t-1}(Q_t - Q_{t-1})}{P'_{t-1}Q_{t-1}}\right] = I_{t-1}^{DL}\frac{P'_{t-1}Q_t}{P'_{t-1}Q_{t-1}} \tag{5.1a}$$

$$\text{Törnqvist-Theil:}\quad I_t^{DT} = I_{t-1}^{DT}\prod_{i=1}^{m}\left[\frac{Q_{it}}{Q_{it-1}}\right]^{\overline{w}_i} \tag{5.1b}$$

where

$$\overline{w}_i = ½\left(\frac{P_{it}Q_{it}}{P'_tQ_t} + \frac{P_{it-1}Q_{it-1}}{P'_{t-1}Q_{t-1}}\right)$$

Here P is an m-dimensional vector of commodity prices and Q is an m-dimensional vector of the corresponding quantities. The transpose of a vector is indicated by a prime, so that $P'Q$ is the sum of the products of the respective elements of P and Q. The t subscripts indicate the time period. The choice between alternative approximations of the Divisia index depends on the nature of the data on hand and the functional form deemed most appropriate for aggregating the quantities of interest (Diewert 1978; Craig and Pardey 1990a).

5.1.2 Spatial Indices

Economic theory gives us less guidance in constructing indices of real aggregates in the cross-sectional dimension since there is no single vector of price weights that is representative for all countries to the extent that international markets for goods and factors are not entirely integrated. As argued elsewhere (Craig and Pardey 1990a), a chained index is of less use in cross section because there is not the same behavioral notion that prices and output evolve over space, i.e., across countries, as they do over time.

If one is forced to resort to fixed-weight indices, attention is focused on the construction of value weights that can be used to calculate real aggregates expressed in common units. The two options most frequently used in international comparisons are the conversion of commodity prices to common currency units or the conversion of all commodities to a common physical unit such as wheat equivalents (Hayami and Ruttan 1985).[2]

---

[2] The problem of currency conversion can be avoided if one uses the Törnqvist-Theil approximation of the Divisia index to construct multilateral indices. In these indices, local prices only enter the calculation in the construction of local value *shares*. Since it is only local and base-country value shares that are averaged, one need never employ an exchange rate. If only a single cross section is being considered, this index method has a lot to recommend it. However, with panel data, i.e., cross sections for several years, the

When choosing an exchange rate series to convert local currency to a common or numeraire currency, the goal is to find a converter that correctly translates the purchasing power of the local currency in the particular sector of the economy being analyzed. This is typically not the same problem as searching for an equilibrium exchange rate. Market-determined exchange rates reflect the relative purchasing power of a currency in trade and are thus influenced by a fairly narrow set of real and financial transactions that may or may not be directly related to the aggregates of interest. The managed or fixed exchange rates common in less-developed countries may be even less useful for translating real purchasing power. There is ample empirical evidence that neither market nor managed exchange rates vary in the short run in a way that reflects differences in average price levels across countries (Levich 1985), yet this is exactly what the ideal converter would do.

World Bank staff developed one converter, the Atlas exchange rate, that uses both official exchange rates and short-run changes in relative price levels (World Bank 1983).[3] For some countries, trade restrictions, government exchange rate policies, and the like cause official market exchange rates to deviate flagrantly from the rate that applies to the foreign transactions effectively taking place. In such cases, the Atlas exchange rate is adjusted using secondary data concerning the nature and estimated impact of these distortions.

The International Comparisons Project (Kravis, Heston, and Summers 1982) has generated an alternative series of synthetic exchange rates called purchasing power parities (PPPs) using the Geary-Khamis procedure.[4] These PPPs are an attempt to get a broader measure of relative currency values by comparing the relative costs in local currencies of a detailed basket of traded and nontraded goods and services. One feature of the Geary-Khamis procedure is that it actually performs two steps at once. The set of n country PPPs are calculated at the same time as the m-dimensional "international" price vector by solving a system of $m+n-1$ equations:

$$\Pi_i = \sum_{j=1}^{n} \frac{P_{ij} Q_{ij}}{PPP_j \sum_{k=1}^{n} Q_{ik}} \qquad i = 1, \ldots, m \tag{5.2a}$$

$$PPP_j = (P_j' Q_j) / (\Pi' Q_j) \qquad j = 1, \ldots, n-1 \tag{5.2b}$$

---

implied time series for each country in the cross section will not be calculated using only local prices and so may yield a biased picture of real local growth rates.

3 World Bank (1983) gives details of two earlier versions of the Atlas method, while *World Development Report 1985* (p. 244) describes the current Atlas method, which uses a simple average of the official market exchange rate for the current year and two predicted exchange rates for the current year that are based on observed exchange rates and relative inflation rates of the two previous years. Specifically, $e_t^* = 1/3\ [e_{t-2} (P_t/P_{t-2})/(\$P_t/\$P_{t-2}) + e_{t-1} (P_t/P_{t-1})/(\$P_t/\$P_{t-1}) + e_t]$, where e_{t-j}, P_{t-j}, and $\$P_{t-j}$ are, respectively, the official market exchange rate, a local general price index, and the US general price index in year $t-j$.

4 For a comprehensive discussion of PPP indices, see Kravis et al. (1975), Kravis, Heston, and Summers (1978, 1982), Summers and Heston (1984, 1988), Kravis (1986), and EUROSTAT (1982).

The PPP for the numeraire country, say n, is set to unity by definition:

$$PPP_n = (P_n' Q_n) / (\Pi' Q_n) = 1 \tag{5.2c}$$

In these formulas, Q_{ij} is the quantity of commodity i produced in country j. The international price of commodity i, Π_i , in equation set 5.2a is the weighted average price of the n country-specific prices, P_{ij}, where country prices are converted to a common currency using implicit exchange rates and then weighted by the physical share of country j in the total quantity of commodity i. The implicit exchange rate or purchasing power parity for country j, PPP_j, is defined in equation 5.2b as the value of its commodity bundle evaluated at international prices relative to that same bundle's value when evaluated at domestic prices.

There is empirical evidence that official exchange rates vary from PPPs in a significant and systematic manner (Heston and Summers 1988). A ratio of annual, average, official exchange rates to PPPs is generally greater than unity for low-income countries and often slightly less than unity for high-income countries. This pattern is due in large measure to differences across countries in the relative prices and quantities of tradable versus nontradable goods and services. Nontradables are generally more labor intensive than tradables, and productivity differences between low- and high-income countries tend to be lower in nontradables. When combined with the fact that labor is relatively cheap in low-income countries, these structural factors lead to lower relative prices of nontradables in low- versus high-income countries.

One advantage of PPPs is that they are not unduly influenced by policy shifts in exchange rates or by sudden swings in financial transactions. They may also be constructed to reflect differences in average prices for a very specific segment of an economy and for a particular set of countries. If the aggregates of interest are dominated by nontraded goods, the PPPs are likely to be more accurate converters than official exchange rates. A practical disadvantage of PPPs is the need to collect detailed data on local prices and comparable quantities in all countries and years in the sample.

Official exchange rates, Atlas exchange rates, and PPPs are converters that have been used in a variety of ways to construct cross-sectional indices. If we use P_j^* to represent the price vector of country j which has been converted to a common currency, say dollars, and let Q_j represent the corresponding quantity vector of country j, then one possible cross-sectional index with base country b is given by

$$I_j^{CS} = (P_j^{*\prime} Q_j) / (P_b^{*\prime} Q_b) \tag{5.3}$$

In this index each country's quantities are aggregated using corresponding prices expressed in a common numeraire which maintains the *local* relative price structure. A more commonly used cross-sectional index formula applies an identical set of value weights to aggregate quantities in all countries, using

$$I_j^{\overline{CS}} = (\overline{P}' Q_j) / (\overline{P}' Q_b) \tag{5.4}$$

The single price vector $\overline{P}$ may be the price vector of the base country or of an arbitrarily chosen third country, a simple average of sample price vectors, or a weighted average of sample price vectors as in the Geary-Khamis procedure.

If the relative price structure differs across countries, each of these different ways of defining the common price vector used in (5.4) will typically result in different indices. If the converters have, in fact, translated all local prices into comparable currency units, there is no obvious need to tamper with the local *relative* price structures when constructing cross-sectional indices. Lack of data on individual country prices may force one to use (5.4), but (5.3) comes closer to the ideal of using relative prices that represent those faced by local agents when summarizing the economic outcomes that are the consequences of their actions.

The use of a common numeraire commodity, rather than a common currency, requires converting each real quantity into units of the numeraire using relative prices to make the translation. For example, a wheat equivalent index can be formed using

$$I_j^{WE} = (R_j' Q_j)/(R_b' Q_b) \tag{5.5}$$

where R_j is the vector of relative prices in country j, or

$$I_j^{WE} = (\overline{R}' Q_j)/(\overline{R}' Q_b) \tag{5.6}$$

where $\overline{R}$ represents a *common* vector of wheat relativities applied to aggregate quantities of both the base country b and country j.[5]

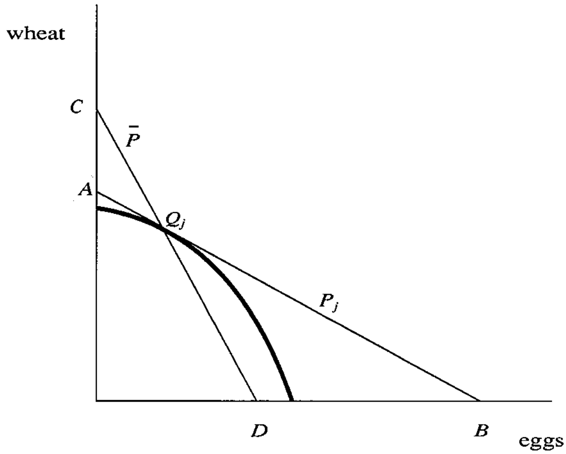
Equation 5.6 can be criticized for the same reasons given for (5.4). The use of a common vector of price relativities, however they are chosen or constructed, amounts to imposing an artificial relative price structure on all or most of the countries in the sample. If the units of both the base and comparison country aggregates can, in fact, be converted to comparable units of wheat using local prices, there is no need to impose a synthetic or nonrepresentative set of value weights on either aggregate.

The problem with both of these wheat equivalent indices is that the choice of the numeraire commodity is critical. As shown in figure 5.1, the value of the commodity bundle represented by output point Q_j can be measured in either tons of wheat (on the vertical axis) or dozen eggs (on the horizontal axis). If one uses the local relative price vector P_j the value of country j's output is either A tons of wheat or B dozen eggs. If an alternative relative price vector such as $\overline{P}$ is used, the total value of the country's output as measured in wheat rises from A to C or alternatively falls from B to D when output is measured in eggs. Cardinal and even ordinal rankings of countries may be altered by the choice of the numeraire commodity.

---

[5] Each element i of relative price vector R is the ratio P_{ij}/P_{wj} where P_{ij} is the price of commodity i in country j and P_{wj} is the local price of wheat. Each price is expressed in units of the currency of country j.

Figure 5.1: *Wheat- versus egg-equivalent output measures*



In practice, wheat equivalent aggregates have been manipulated even further. For instance, in Hayami and Ruttan's (1971, 1985) work, the total value of output as measured in wheat is calculated for each country using several *nonlocal* relative price vectors. A geometric average of the resulting aggregates is then taken to be the final measure of the aggregate. This procedure mimics the bilateral Fisher ideal index in a multilateral context.[6] Since different relative price vectors imply different aggregate volumes, the geometric average will tend to provide some ad hoc smoothing of these differences. Referring again to figure 5.1, the geometric average of aggregates would fall somewhere between points *C* and *A* if wheat is the numeraire commodity, or between *D* and *B* if eggs are the numeraire.

5.1.3 Multidimensional Indices and Comparable Aggregates

When we want to compare aggregates *both* across countries *and* over time, sticking to aggregation procedures that use representative relative price weights is still a guiding principle. An effective way to accomplish this is to produce chained time-series indices for each country using local prices, and then scale the resulting series for each country with volume aggregates that have comparable units for all countries in the base year.

As it happens, the construction of these cross-sectional scaling factors involves precisely the same issues discussed above for spatial indices. One need only calculate a comparably measured aggregate for each country in the base year and use it to multiply each observation in a country's time series. The numerators of the spatial indices in equations 5.3, 5.4, 5.5, and 5.6, calculated with base-year data, all provide reasonable scaling factors

---

6 Hayami and Inagi's (1969) wheat-equivalent procedure does not yield a true Fisher ideal index unless the price relativities employed include a local price relativity for each country.

if the objective is obtaining comparable *aggregates*. These same spatial indices can be used to recalibrate country-specific time-series indices if the object is a multidimensional *index*. Once again the desirability of using representative prices points to the use of scaling factors (equations 5.3 and 5.5) that preserve local relative price structures.

We find this procedure more appealing than other procedures for multilateral international indices that have been advocated. Caves, Christensen, and Diewert (1982) have suggested using the Törnqvist-Theil approximation of the Divisia index (equation 5.1b) that compares all country observations to one specific (perhaps synthetic) country and time period. It is, however, more natural to link prices and quantities in neighboring periods of time than to blend value weights from disconnected periods and countries. Khamis (1988) suggests constructing a single set of international prices that are averages of prices over all years and countries in the sample. This may impose a large computational burden while remaining in essence an index method based on fixed, nonrepresentative value weights.

5.2 AGGREGATION IN INTERNATIONAL DATA SETS

We have discussed direct calculations of comparable quantity aggregates, but most international data sets include *preaggregated* data. In many cases, aggregates are reported in total local currency units, so we can only hope to deflate such measures to arrive at implicit volumes. In other cases, the volumes that are reported are unweighted totals of heterogeneous commodities or factors of production. Secondary data may be available to adjust such volumes, but comparable volumes may only be derived indirectly.

5.2.1 Value Aggregates

When confronted with value aggregates measured in a variety of local currency units, each must usually be deflated to reflect changes over time in each country's average price level and converted to arrive at aggregates in comparable real values or volumes.

The choice of an appropriate local price index entails some conceptual difficulties. Readily available price indices are typically general indices that may not reflect price developments in specific sectors of an economy such as agriculture. World Bank (1989) statistics indicate that implicit deflators of GDP and AgGDP are systematically different. Broadly speaking, AgGDP deflators indicate lower average rates of inflation in more-developed countries than do deflators defined over all sectors of the economy. The opposite holds true in most less-developed countries. Thus, using AgGDP deflators instead of GDP deflators will yield lower estimates of implied growth in real agricultural output for less-developed countries, as indicated in table 5.1.

Another problem is that price indices are commonly constructed using fixed quantity weights, as in a Laspeyres price index. The advantage of these measures is their ease of interpretation; they tell us how much the cost of purchasing exactly the same basket of goods has changed over time. Their disadvantage lies in the fact that they tend to overstate

Table 5.1: *Difference in Growth in Real Agricultural Output Using Alternative Deflators*

Region	1961-70	1971-80	1981-85	1961-85
	%	%	%	%
Sub-Saharan Africa (20)[a]	0.8	-0.8	-3.1	-0.7
China	-1.1	-1.2	-0.5	-1.7
Asia & Pacific, excl. China (12)	-0.7	1.0	-0.3	0.5
Latin America & Caribbean (21)	0.9	0.3	-2.1	0.1
West Asia & North Africa (7)	-1.1	1.3	-1.3	-0.1
Less-Developed Countries (61)	-0.3	0.3	-1.0	-0.2
More-Developed Countries (13)	0.5	0.7	2.9	0.9
Sample Total (74)	0.0	0.4	-0.2	0.1

Source: Implicit GDP and AgGDP deflators and AgGDP data primarily taken from World Bank (1989), and PPPs taken from Summers and Heston (1988).

Note: Percentages indicate absolute differences in compound annual rate of growth of "real" AgGDP deflated with implicit AgGDP less that deflated with implicit GDP deflators. A positive difference in growth means that agricultural prices have grown more slowly than the average price level.

[a] Bracketed figures indicate the number of countries in regional totals.

changes in the general price level by failing to allow for changes in the composition of the basket of goods produced or consumed that are likely to occur if there are changes in relative prices over the period being considered. The longer the time horizon of the study, the more likely are fixed-weight indices to understate the volume of economic activity by deflating with an index that fails to account for substitution. As argued in the index-number literature, the use of Divisia price indices would alleviate this last problem.[7] However, in an international context, these indices are so rarely constructed that they are currently not an option for international comparative analysis.

The problems of currency conversion have already been touched on, so the only new question is the order in which one employs deflators and converters. From the various algorithms available for translating values into comparable volumes, a practical alternative is to select a two-step procedure. One can first convert local currency values into a numeraire currency, such as US dollars, then apply an appropriate price index to account for price-level variability in the numeraire currency. The other option is to first deflate local currency values using local price indices then convert local prices into a numeraire currency using some base-year measure of relative currency values.

There are numerous deflators and currency converters that can be incorporated into either algorithm. Unfortunately, the choices matter. Since we have no independent measure of the truth, we are forced to proceed using some rules of thumb.

In choosing a price deflator, one should use the price index that most nearly reflects

7 See Diewert (1978). He demonstrates the quantitative differences between fixed weight, chained, implicit, and explicit quantity indices using time series data on Canadian consumption expenditures.

the composition of the aggregate value to be deflated. In multicountry studies, this rule of thumb will argue for an algorithm in which aggregates are deflated first with a local price index whenever adequate price indices are available for each country in the sample. The basket of goods covered in a local price index may be quite different from that of a numeraire country's index when living standards and local relative prices vary substantially across the countries in a sample. This cross-sectional variability would lead to biases in measurement whose direction and magnitude would be difficult to predict.

A more subtle problem is the combined choice of deflator and converter. If the values to be compared are the total values of a single uniform good, the two algorithms (deflation then conversion or conversion then deflation) yield the same result if and only if the deflator and converter are defined over the specific good. If the values to be compared are aggregates, the deflator and the converter must be defined over the specific basket of goods represented by the aggregate. General price indices, market and/or official exchange rates, and nonspecific PPPs all introduce biases to the extent that they reflect aggregates whose composition may differ from the aggregate of interest.

Even with properly defined deflators and converters, the problems of aggregation cannot be escaped. As demonstrated in Pardey, Roseboom, and Craig (forthcoming) the two algorithms will yield different volume series unless it is the scale and not the composition of the aggregates that varies over time and across countries. Both algorithms diverge from the desired volume measure as the composition of the aggregate changes across the sample. So, when using the convert-first procedure, the volume measure will be biased unless the composition of the numeraire country's aggregate is representative of all other countries in all years of the sample. The deflate-first procedure will generate biases in the volume measure whenever the base-year basket within each country is not representative of that country for the period being considered.

So, in a particular application, the choice of algorithm must be made on the basis of whether it is the temporal or cross-sectional composition of the aggregate that is likely to vary most. Researchers have shown a preference for converting local currencies to dollars first and then deflating using a US price index. However, in a data set that includes countries at diverse stages of development, it is quite likely that cross-country differences in the composition of the aggregates will dominate the temporal variability unless the data span several decades; hence, a deflate-first procedure would demand far less of the data.

Pardey, Roseboom, and Craig (forthcoming) contrast the results of applying the two procedures to data on agricultural research expenditures in a sample of 90 countries. Volume measures were constructed using the convert-first algorithm with annual average exchange rates and PPPs as alternative currency converters; both series were then deflated using the US implicit GDP deflator. These were contrasted with volume measures produced by deflating first with country-specific implicit GDP deflators and converting with each of the two base-year currency converters.

For this application, no price index covering the specific mix of labor, materials, and equipment peculiar to agricultural research was available in each country, so the GDP deflator was a practical compromise.[8] The annual average exchange rate used was the yearly official market rate, which generally corresponds to the IMF's *rf* or inverted *rh* rate. The PPP series, which was defined over GDP, represented another compromise. Published PPPs either cover too few countries or a basket of goods that is not particularly representative of agricultural research.[9] The commodity coverage of PPPs obtained from Summers and Heston (1988) did, at least, correspond closely to that of the implicit GDP deflators being used.

Table 5.2 reports the 1981-85 average annual volume of resources committed to agricultural research implied by each of four measures.[10] In each column the regional total is indexed on the total sample volume implied by the particular conversion method.

For the 1981-85 period, the regional shares exhibited nontrivial sensitivity to the choice of translation. Converting the series first with annual average exchange rates lowered the measured total research commitment by at least one billion dollars. In general terms, the differences between the estimates were more dramatic for the less-developed than for the more-developed countries. A difference of approximately 55% in the less-developed countries' share of global research expenditures arose simply from the choice of converters. In particular, using PPPs rather than exchange rates approximately doubled the Asia & Pacific region's share of total research resources from around 6.3% to more than 13%. This pattern can be traced to the fact that relative price levels in less-developed countries reflected in Summers and Heston's (1988) PPPs are much lower on average than those implied by market exchange rates.

The volume measure was somewhat sensitive to the order of deflation and conversion near the base year but over longer time periods the two algorithms produced more obviously divergent results — particulary when volumes were obtained using annual average exchange rates as converters. Figure 5.2a presents the percentage deviation of the deflate-first

---

8 A long-run agricultural research deflator for the US which takes account of annual variations in the mix of labor, capital, and materials used in agricultural research is given by Pardey, Craig, and Hallaway (1989). For additional discussion relating to R&D deflators, see also NSF (1970), Jaffe (1972), Mansfield, Romeo, and Switzer (1983), Mansfield (1987), and Bengston (1989).

9 MacDonald (1973) and OECD (1981) discuss the concept of a PPP for R&D at some length. However, MacDonald provides such series for only a very small set of more-developed countries. PPPs defined over subsectors of the economy differ substantially, as described in Kravis, Heston, and Summers (1982). They point out that, on average, currencies for less-developed countries have substantially less purchasing power over a basket of investment goods and services than over a more general basket of goods and services. For government goods and services, the converse is true. We chose to use the broadly based Geary-Khamis PPPs of Heston and Summers (1988) calculated over GDP rather than any of its subaggregates because Heston and Summers themselves were concerned about the robustness of these more specific PPPs.

10 These four measures all involved deflating with implicit GDP deflators. This contrasts with the method used by Evenson and Kislev (1975a), Judd, Boyce, and Evenson (1986), and Mergen et al. (1988). The clearest description of the translation procedure used in these studies appears to be in Judd, Boyce, and Evenson (1983, p. 3) where it is stated that "[research] expenditures were converted to US dollars using official exchange rates and were then inflated to 1980 dollars using a general wholesale price index."

Table 5.2: *Alternative Measures of the Volume of Agricultural Research Resources, 1981-85 Average*

Region	Convert-first		Deflate-first	
	AAER	PPP	AAER	PPP
	%	%	%	%
Sub-Saharan Africa (31)[a]	4.1	4.7	4.6	4.7
Asia & Pacific, excl. China (11)	6.3	13.3	6.3	13.4
Latin America & Caribbean (17)	6.7	9.0	5.8	8.9
West Asia & North Africa (8)	2.1	2.9	2.1	2.8
Less-Developed Countries (68)	*19.2*	*29.8*	*18.8*	*29.8*
MDCs other than US (21)	55.4	49.2	59.7	49.7
United States (1)	25.4	21.0	21.5	20.5
More-Developed Countries (22)	*80.8*	*70.2*	*81.2*	*70.2*
Total Sample (90)	100.0	100.0	100.0	100.0
Total Sample Volume [b]	*5491*	*6646*	*6493*	*6821*

Source: Annual average exchange rates and implicit GDP deflators are primarily taken from World Bank (1989), PPPs from Summers and Heston (1988), and agricultural research expenditure data from Pardey and Roseboom (1989a).

Note: Translation procedures involved deflating with either US or local implicit GDP deflators and converting with either annual average exchange rates (AAER) or purchasing power parities (PPP) over GDP. Figures represent regional shares of a 90-country total. Data may not add up exactly because of rounding.

[a] Figures in brackets indicate the number of countries in regional totals.
[b] Millions of 1980 US Dollars.

versus the convert-first volume measures when annual average exchange rates and implicit GDP deflators are used to derive the respective volume measures. In figure 5.2b the same graph is presented for the volume series which used PPP exchange rates and GDP deflators.

When annual average exchange rates are used, the deflate-first algorithm led to a consistently larger volume measure than that obtained when expenditures were converted first. This suggests that, ceteris paribus, either the US dollar was undervalued with respect to virtually every country's currency in 1980, or that movements in local price levels were imperfectly translated by changes in the official annual exchange rates. The difference between these two volume measures is most pronounced in the Bretton Woods years when all exchange rates were essentially fixed. This gives further credence to the idea that official exchange rates may carry little or no information about changes in the relative purchasing power of different currencies, and so will be inappropriate converters for the purposes of international comparisons of long time series.

The temporal pattern of deviations of the PPP-converted measures in figure 5.2b is far less dramatic than those in figure 5.2a. By construction, changes in PPPs over time should do a better job of capturing changes in relative price levels between countries. In contrast

Figure 5.2a: *Percentage deviation of convert-first from deflate-first formula using annual average exchange rate convertors and implicit GDP deflators (Base-year = 1980)*

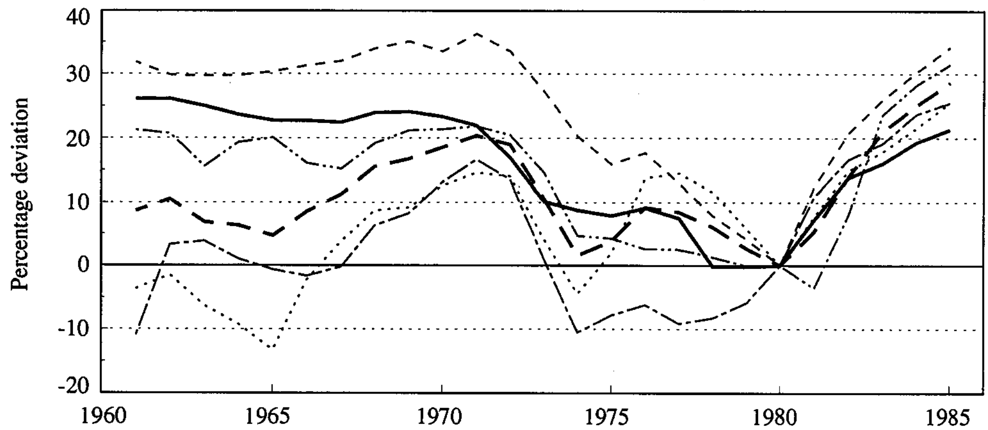
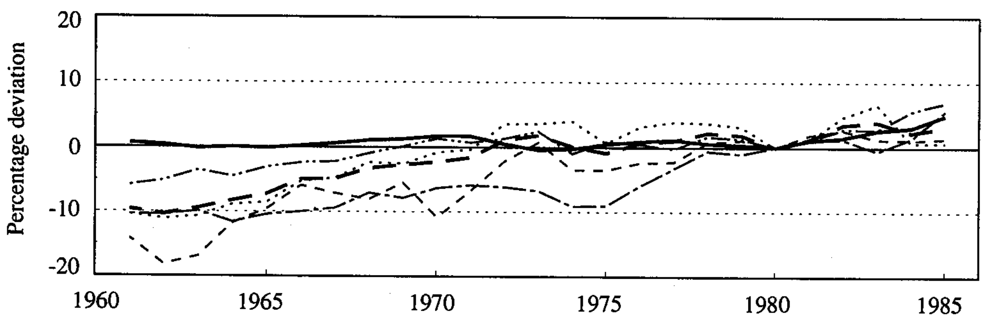


Figure 5.2b: *Percentage deviation of convert-first from deflate-first formula using PPP convertors and implicit GDP deflators (Base-year = 1980)*



Legend: [a]

- - - ·	Sub-Saharan Africa (32) [b]	——··	West Asia & North Africa (8)
······	Asia & Pacific (11)	– —	Less-Developed Countries (68)
··—···	Latin America & Caribbean (17)	——	More-Developed Countries (22)

Source: See table 5.2.

[a] Legend applies to both figures. Regional averages weighted by proportion of the 1981-85 average of agricultural research expenditures (expressed in 1980 PPPs) for each group accounted for by each country.
[b] Figures in brackets indicate the number of countries.

to the measures with exchange rate conversions, there appear to be no systematic differences between the convert- and deflate-first methods for the more-developed countries in any particular subperiod and for most less-developed country regions in the post Bretton Woods years. With these data, the convert-first procedure generates a larger volume measure than the deflate-first method for many less-developed country groupings during the Bretton Woods years. It is difficult to make too much of this trend as pre-1975 PPPs for many of the less-developed countries were derived using so-called short-cut extrapolation methods based, among other things, on market exchange rates without the benefit of local price measures based on benchmark survey data (Summers and Heston 1988).

5.2.2 Noncomparable Volumes

Statistics on many important agricultural inputs such as land, labor, tractors, and fertilizer are published as real totals or volumes. This would appear to make the job of international comparisons easier; however, the totals do not always count strictly comparable units.

The hectares of land in agriculture are far from homogeneous even when broken down into categories of cropland, pastureland, and rangeland. In an aggregate counting of just cropland hectares, one hectare of cropland that receives ten centimeters of rain per year may well have been added to another hectare of cropland that receives ten centimeters of rain per month. Moreover, some croplands are irrigated while others are not. While the aggregation of heterogeneous cropland hectares is less likely within a small region, it will almost invariably be the case when forming totals within a country, and it is certainly a problem for comparisons of cropland totals in a large international sample.

The problem of aggregating heterogeneous cropland hectares is compounded by the fact that we often want a measure of total land in agriculture that combines hectares of cropland with even more dissimilar hectares of pastureland, rangeland, and so on. If local values or rents for different land types can be observed, we have some direct way of measuring the relative productivity of different hectares. The use of local relative rents to reflect actual quality differences rests on the implicit assumption that local rents reflect the marginal value of a hectare of land in agriculture.

With a series of local rents, we could construct a direct "quality-adjusted" Divisia index for land volumes using methods described in section 5.1. The Divisia index would then provide a measure of real changes in land in agriculture over time that could then be scaled to get a volume aggregate measured in any type of numeraire hectare. For example, if we designate cropland as the numeraire or representative hectare, then when one hectare of pastureland rents for a fraction of a hectare of cropland, its relative rent would lead us to count it as that same fraction of a hectare of cropland. To arrive at internationally comparable cropland totals, cross-sectional scaling of the resulting time series indices for individual countries would be necessary. This would require the most difficult calculation; namely, constructing real estate converters that translate local cropland into an international cropland hectare of constant quality.

An additional problem with using unweighted total hectares of land, especially in international comparisons, is the fact that it is difficult to find measures on the intensity of land use. If a hectare of cropland is used for several crops in one year in Asia, but is rarely used for more than one planting in other parts of the world, simple counts of total hectares will give a distorted view of the cross-sectional variation in the flow of land services. This could be solved if land were measured in service flow units such as hectare-plantings rather than as a stock measure of hectares.[11]

As with the land input, the ideal measure of labor inputs in agriculture would reflect the service flow from labor and not merely the stock of workers available for the sector. Moreover, a human-capital perspective on labor gives rise to an analagous aggregation issue for labor variables. An hour worked by a farmer with no experience and a primary education is quite likely to be less effective than an hour worked by an experienced farmer with the same or higher level of education. In addition to differences in the human capital characteristics of farmers, there are differences in the effectiveness of hours worked by farmers, family members, and hired workers.

Once again, indices of the quantity of a quality-adjusted labor input could be constructed using the Divisia methods outlined in section 5.1 if wages and hours of different worker types were available. Scaling the index to get local aggregates measured in consistent units over time would require the use of base-year local relative wages. To make international comparisons, one also needs converters to capture the cross-sectional differences in the human-capital characteristics of the representative or numeraire worker.

5.3 CONSTRUCTING AGRICULTURAL DEVELOPMENT MEASURES

The ideal aggregation methods discussed above can hardly ever be implemented when using international data sets because detailed information on prices and quantities is typically not available. In addition, a distinction is rarely made between stocks of inputs and the service flows from those inputs. In this section, the compromises required to analyze agricultural development or productivity measures are discussed along with the likely biases inherent in using preaggregated data.

5.3.1 Agricultural Output Measures

In measuring agricultural output, one would often like an output measure that represents gross production less *inside* inputs, i.e., those inputs produced and reused within agriculture (Star 1974). In other words, products such as seeds or eggs which are required as inputs in their own production or feeds such as hay or milk which will be used as inputs in livestock should be deducted from gross production to avoid double counting. It is possible to start from FAO statistics[12] on "gross-gross output" and deduct inside inputs to get a final output aggregate, but comparable and reliable data on these inside inputs are difficult to come by for all countries and all time periods. FAO also publishes agricultural production indices

[11] A comparative study of Asian agriculture factored in multiple cropping levels to distinguish total cropped area versus cultivated land area. It gave land utilization rates ranging from 189% for Taiwan in 1966 to 93% for Thailand during 1980-81 (APO 1987, p. 17).

[12] See FAO (1974) for a discussion on output concepts used by FAO.

which are simply Laspeyres indices of final agricultural output. But, using this index for international comparisons of aggregates requires a great deal of data in order to calculate cross-sectional scaling factors.[13]

An alternative is to use value-added figures of AgGDP from national accounts data. By construction, these have the advantage of comparability with broader measures of economic activity such as GDP. However, they may present some problems for international comparisons if their calculation does not result in strictly comparable output aggregates. The problem does not arise because there is something intrinsically wrong with netting out intermediate inputs whether purchased from *outside* or produced *inside* the agricultural sector. Rather, the problem stems from asymmetric treatment of these two types of inputs. For example, in more-developed economies, fertilizer inputs are more likely to be purchased inputs than in less-developed economies where the same services may be provided largely from inside inputs such as manure. If the inside and outside inputs are not treated symmetrically in arriving at value-added figures, this will introduce some biases when using value-added measures in comparisons of levels and growth rates of agricultural output in cross-country studies.

The ratio of value-added to final agricultural output differs across countries. As one would expect, this ratio is much higher in less- than in more-developed countries. In 1975, this ratio ranged from 39% in Switzerland to 96% in Thailand (FAO 1986b). Within most countries, changes in this ratio over the past two decades are much less pronounced than cross-country differences (ECE and FAO 1981, 1989; FAO 1986b). However, there are shifts in this ratio. Changes in the structure of agriculture — in particular degrees of specialization — or changes in relative prices which lead to substitution between inside and outside inputs are likely to affect the ratio of value-added to gross output even in the absence of changes in technology and productivity.

The measures of agricultural output that are used throughout this volume are time series of AgGDP in current local currency units extracted primarily from World Bank (1989b). AgGDP measures were chosen over the alternative of scaling the FAO's agricultural production index (*FAO Production Yearbook*). The FAO production index excludes forestry and fishery outputs, which introduces problems of mismatched coverage in research and in most of our conventional input variables. Moreover, direct comparisons of AgGDP and non-AgGDP were deemed important for contrasting the development of agriculture with the rest of the economy.

To get comparable volume measures, these nominal output aggregates were first deflated using implicit AgGDP deflators based in 1980 (World Bank 1989). They were then

---

[13] Hayami et al. (1971) and Hayami and Ruttan (1971, 1985) first constructed country-specific estimates of final agricultural output, averaged over 1957-62 and measured in wheat-equivalent units. They then extrapolated these country-level estimates using FAO's agricultural production indices. Potential biases in using wheat-equivalent measures are discussed in section 5.1, while biases from inferring output growth rates when using fixed weight indices such as the FAO production indices are discussed in section 5.2 and also in Craig and Pardey (1990a).

converted to US dollars in the base year using PPPs defined over gross agricultural output (FAO 1986). We were able to use the preferred order of translation as well as representative deflators and converters for the bulk of the sample. An even more representative converter would have been PPPs defined over value added in agriculture, but no such converter was available for the whole sample.[14]

5.3.2 Agricultural Input Measures

Data on total hectares in agriculture used in this volume are unfortunately aggregates of stocks of heterogeneous land types. Information is sparse on multiple cropping, so one is also forced to use stocks of land instead of service flows. An annual breakdown of land types was available but not the local rent or value data that would have allowed calculation of economically meaningful aggregates.

Table A5.1 gives some idea of the heterogeneity of agricultural lands by indicating the percentage of total agricultural land accounted for by permanent pasture, or by arable and permanently cropped land. The latter is further disaggregated to indicate the percentage of land under irrigation. The differences across countries in the types of land and the extent of irrigation are quite dramatic. In China, for example, a low percentage of total agricultural land is either arable or permanently cropped, but almost half of that land is irrigated. No country except Japan irrigates as large a percentage.

The percentage of agricultural land in permanent pasture is dictated more by agroecological characteristics than by stage of development. Countries in Asia have a significantly lower percentage of land in pastures than do countries anywhere else in the world, and they irrigate their arable and permanent cropland more intensively than do countries in any other region. At the other end of the spectrum, agricultural land in sub-Saharan Africa is predominantly pastureland, and a very small percentage of arable land is irrigated.

An international index of land quality has been calculated by Peterson (1984) using an hedonic approach to valuing the cross-sectional differences in agricultural land characteristics. First, value weights for different land characteristics are derived by regressing a cross section of US land values on the differing characteristics of land in agriculture in the US. These weights are then used to place a relative value — and therefore a measure of relative quality — on hectares of agricultural land in different countries. The indices of relative quality of total agricultural land range from a regional average of 67 for West Asia & North Africa to an average of 161 for Asia & Pacific. A group of 83 less-developed countries had an average index of 101, while the average for the 21 more-developed countries in his study was 81.

These land-quality indices could be used to scale up or down the unadjusted total hectares in agriculture instead of implementing the more demanding Divisia input indices,

---

14 Terluin (1990) provides a recent attempt at constructing such a converter for 10 EEC countries.

but they have some shortcomings. First of all, the index fails to account for changes over time in the quality of the average hectare in agriculture. Given the increased irrigation usage and the changing mix of land types evident in table A5.1, it is clear that the index will not fully account for changing quality differentials. More subtle problems of the correct weights and the relevant land characteristics lead us to think that these indices are useful but not completely satisfactory indicators of cross-sectional land quality.

Labor aggregates available for international studies of agriculture are as inadequate as those available for land. There are no broadly based studies that allow one to distinguish between stocks of labor available to agriculture and actual hours worked in agriculture. Instead, the available aggregate counts the economically active agricultural population, whether actually engaged or seeking employment in agriculture, forestry, hunting, and fishing.

In addition, the information on labor force characteristics is so difficult to come by that it is virtually impossible to construct Divisia indices using current international data sets. Even hedonic procedures analogous to Peterson's land-quality index are difficult to implement for a very large set of countries in the absence of country-specific information on age, education, and income profiles of agricultural workers.

What information is available indicates that the educational attainment of workers has varied dramatically, both across countries and over time (table 5.3). In the past decade, the secondary school enrollment ratio in less-developed countries has ranged from a low of 19% in sub-Saharan Africa to 45% in Latin America & Caribbean. For the same period, this ratio averaged 89% across more-developed countries. Over the past two decades, the ratio of primary enrollment has doubled in sub-Saharan Africa, and secondary school enrollment has more than quadrupled. These relatively recent changes in human capital investments may not have shown up yet in the labor input to agriculture in Africa but they are suggestive of the impact of development on the quality of the labor force.

5.3.3 Productivity Measures

To assess the development of the agricultural sector and, in particular, the sectoral rate of productivity change, one needs detailed price and quantity information on outputs and the whole range of inputs such as land, labor, energy, fertilizer, pesticides, and capital. With such data, it would be possible to construct total factor productivity indices (TFPs) which seek to separate out that part of output growth that can be attributed to increased or altered input usage from that which is interpreted as a pure productivity change (Capalbo and Antle 1988). The construction of such TFPs for a large international sample makes such demands on the currently available data that it is difficult to interpret the unexplained changes in output as productivity changes in the face of so many potential measurement errors.

While meaningful TFPs may be difficult to construct, we can learn a great deal about the patterns of development in agriculture with a judicious use of partial productivity indices. The interpretation of all productivity measures requires some care. Some, if not all, of the change in a particular factor's productivity may be attributable to increased usage of

Table 5.3: *Primary and Secondary School Enrollment Ratios*

Region	1961-65	1966-70	1971-75	1976-80	1981-85
			Primary school enrollment ratio		
	%	%	%	%	%
Sub-Saharan Africa (38)[a]	38	43	50	67	74
China	95	89	95	100	100
Asia & Pacific, excl. China (15)	67	73	76	80	85
Latin America & Caribbean (25)	91	95	92	95	98
West Asia & North Africa (18)	66	75	80	86	91
Less-Developed Countries (97)	76	78	81	87	90
More-Developed Countries (21)	99	99	99	99	100
Total (118)	81	82	85	89	92
			Secondary school enrollment ratio		
	%	%	%	%	%
Sub-Saharan Africa (38)[a]	4	5	8	12	19
China	21	24	36	47	38
Asia & Pacific, excl. China (15)	20	24	25	29	34
Latin America & Caribbean (25)	17	24	30	39	45
West Asia & North Africa (18)	16	23	30	38	44
Less-Developed Countries (97)	18	22	28	34	36
More-Developed Countries (21)	66	74	82	85	89
Total (118)	29	33	39	44	45

Source: World Bank (1989) and *UNESCO Statistical Yearbook 1983 and 1987.*

Note: Primary school enrollment ratio represents enrollment of students of all ages at primary level as a percentage of primary age students. Secondary school enrollment ratio is calculated in the same way. Definitional inconsistancies make it possible to obtain school enrollment ratios greater than 100%. Where this has occurred we have rounded to 100%.

[a] Bracketed figures indicate the number of countries in regional totals.

nonmeasured inputs. Nevertheless, the measured changes (especially when used in conjunction with data on other inputs) still provide useful information on development patterns.

Measurement errors are of concern in both total and partial productivity measures. In assessing changes in output per hectare, it is as important to have a consistently defined denominator in the fraction as it is to define output in a uniform way across countries and over time. Accurate representations of output per worker in agriculture likewise require one to use labor inputs in units that are comparable over time and space. Because the input aggregates one is forced to use do not have the same composition or average characteristics across countries, they make the comparison of productivity measures problematic. However, it is far easier to anticipate the magnitude and direction of biases from mismeasuring individual inputs in partial productivity indices than it is to disentangle multiple sources of measurement error in TFPs.

Using state-level data on the continental US, the implications of systematically adjusting land and labor totals for cross-sectional quality differences are clear (Craig and Pardey 1990b). When land rents are used to account for differences in the quality of pastureland, nonirrigated cropland, and irrigated cropland, measured levels and growth rates of output per acre are systematically changed. The implied growth rates of land productivity are reduced somewhat when one accounts for the fact that there has been some improvement over time in the average quality of land in US agriculture. As arid western cropland is turned into irrigated cropland, total unadjusted acres do not increase even though the average quality of land in agriculture does. Once this increase in input quality is accounted for, part of the likely increase in output will be attributed to increased quality of land and not to increased productivity of a quality-constant acre of land.

The most dramatic effect of accounting for quality differentials in land comes when rescaling implied *levels* of output per acre. The spread in measured regional differences in land productivity is greatly reduced with quality adjustment of land. Since there is a wide range of land quality in the US, these results can be used to anticipate the problems of using unweighted total hectares in international productivity studies.

Craig and Pardey (1990b) have also constructed a quality-adjusted index of labor for the US using actual hours worked by hired and family workers, human capital characteristics of farm operators, and data reflecting the shift from full- to part-time farming by operators. Not surprisingly, quality adjustment has a significant effect on measured labor productivity. Measured growth rates of output per hour are reduced when the increases in the average age and educational attainment of farm workers are taken into account. Regional differences in measured levels of output per hour are also reduced when the labor input is quality-adjusted. While regional differences in agricultural labor markets in the US have become less pronounced, differences across states in the mix of labor types and the average quality of labor have been historically important in accounting for part of the measured cross-sectional differences in levels and growth rates of agricultural output per hour.

Figures 5.3a and 5.3b illustrate the combined effect of quality adjustments on land and labor productivity measures for the US. Each path represents the average level of agricultural output per acre and output per hour in a different region in the US.[15] The more

[15] The regions depicted in figures 5.3a and 5.3b correspond to the 10 USDA production regions with one exception. The states in the USDA's Northeast region were split into two groups to distinguish between states whose agricultural growth has slowed dramatically or stopped in recent years and those that have continued to grow at rates more typical of the Corn Belt and Lake States. The regions comprise Northeast 1 (Maine, New Hampshire, Vermont, Connecticut, Rhode Island, Massachusetts, New Jersey), Northeast 2 (Delaware, Maryland, New York, Pennsylvania), Corn Belt (Illinois, Indiana, Iowa, Missouri, Ohio), Lake States (Michigan, Minnesota, Wisconsin), Northern Plains (Kansas, Nebraska, North Dakota, South Dakota), Appalachian (Kentucky, North Carolina, Tennessee, Virginia, West Virginia), Southeast (Alabama, Florida, Georgia, South Carolina), Delta States (Arkansas, Louisiana, Mississippi), Southern Plains (Oklahoma, Texas), Mountain (Arizona, Nevada, New Mexico, Idaho, Colorado, Montana, Utah, Wyoming), and Pacific (California, Oregon, Washington).

Figure 5.3a: *Quality-unadjusted land and labor productivity paths for the US, 1949-85*

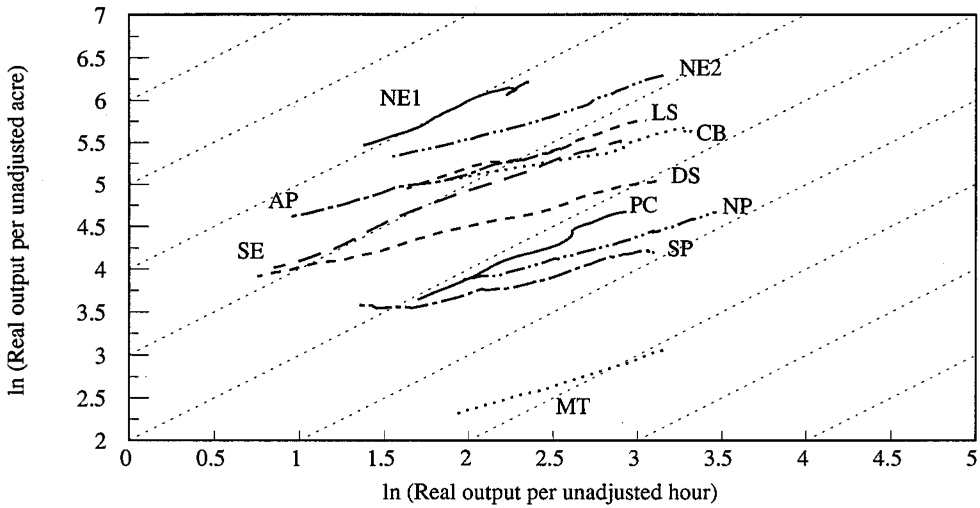
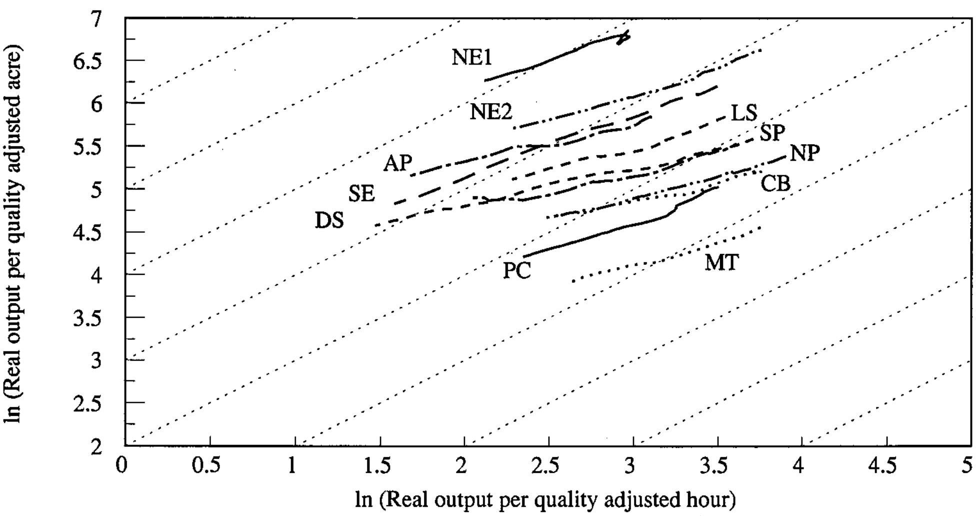


Figure 5.3b: *Quality-adjusted land and labor productivity paths for the US, 1949-85*



Legend: NE1=Northeast 1; NE2=Northeast 2; CB=Corn Belt; LS=Lake States; NP=Northern Plains; AP=Appalachian; SE=Southeast; DS=Delta States; SP=Southern Plains; MT=Mountain; PC=Pacific.

Source: Based on data reported in Craig and Pardey (1990b).

Note: All partial productivities expressed in natural log terms. For the sake of clarity, productivity paths represent five-year moving averages. Regional groupings of states correspond with USDA's farm production regions except for NE1 and NE2 which together form the USDA's Northeast region. See footnote 15 for details. Diagonal lines represent constant land/labor ratios.

careful quality-adjusted aggregation procedures broadly reduce measured differences in land and labor productivity. Since land is completely immobile across states and is not nearly as mobile intersectorally as labor, it is not surprising that land productivity still remains variable even within one country. If the nonagricultural uses of arid lands in the Mountain states are not pressing for its removal from agriculture, there is no economic necessity for its marginal product to match that of the Corn Belt or Lake States. Similarly, given the economic alternatives to agriculture in the Northeast, only land with a high marginal product is likely to remain in agriculture.

Because the regional labor markets are increasingly integrated within the US, we expect that, over time, significant differences in the marginal product of labor will be eroded. Workers in all regions will leave agriculture in response to opportunities outside agriculture (Kislev and Peterson 1982). With an increasingly national market for most labor skills, this should tend to eliminate spatial differences in returns to agricultural workers. As is evident from these figures, once we account for the difference in the labor mix across states, the spatial dispersion in the average productivity of labor appears to narrow more rapidly over time. Because there have been substantial cross-sectional differences in labor inputs to US agriculture over the post-war period, the effects described here of quality adjustment on US partial labor productivity measures can help us anticipate problems in comparing international labor productivity measures.

5.4 CONSTRUCTING AGRICULTURAL RESEARCH INDICATORS

The primary source for the public-sector agricultural research personnel and expenditure data used in this volume is the Indicator Series country files reported in Pardey and Roseboom (1989a). The Indicator Series represents a fully documented and sourced compilation of benchmark survey data plus information from approximately 1000 additional data sources for NARSs in 154 more- and less-developed countries, where possible, for the 27 years 1960 through 1986.[16] Extensive efforts went into achieving completeness and commensurability in the series. However, the unavoidably disparate nature of the data sources, plus the subject of the data series itself, means that these statistics should be considered indicative rather than definitive. Nevertheless, the series represents a major effort to consolidate and restructure previously available data compilations.

The following three sections briefly describe the statistical concept of a NARS used to compile the series, as well as some measurement issues that are germane to constructing comparative measures of agricultural research activity. While the scope of the series, in terms of country and time-period coverage plus number of indicators, constitutes a substan-

---

[16] The Indicator Series reports data on a calendar-year basis whenever possible. However, in numerous instances data were recorded on a fiscal- or academic-year basis. The procedure adopted in such cases was to place the observation in the calander year that overlaps most with the respective fiscal or academic year. Consequently, a fiscal year running from April 1, 1980 to March 31, 1981 was placed in calendar year 1980. A fiscal year running from July 1, 1980 to June 30, 1981 was placed in the calendar year 1981.

tial extension of or addition to global compilations, there remains a problem with missing observations that limits the use of these data for purposes of policy analysis. The procedures used to tackle this problem are detailed in section 5.4.4, while section 5.4.5 compares prior data compilations with those presented in this volume.

5.4.1 Defining a NARS

There is no universal agreement as to what constitutes a NARS, and while the concept of a NARS has value as an analytical tool, it is difficult to operationalize for statistical purposes. A useful beginning was to split the concept of a NARS into three dimensions, namely, (a) national, (b) agricultural, and (c) research, and to consider each of these dimensions separately.

National

The notion of what constitutes a "national" set of statistics on agricultural research is open to many interpretations. One option is to adopt a geographic interpretation and include all agricultural research performed within the boundaries of a country. Another possibility is to pursue a sectoral approach and include domestically targeted research activities funded and/or executed by the public sector of a particular country.[17] This latter approach was adopted for the Indicator Series, which attempts to include all agricultural research activities that are financed and/or executed by the public sector, including private, nonprofit, agricultural research. It explicitly excludies private, for-profit, agricultural research. This sectoral coverage corresponds to that adopted by the OECD (1981, pp. 83-91) and includes the government, private nonprofit, and higher-education sectors but excludes the business-enterprise sector.

The government sector was taken to include those federal or central government agencies, as well as provincial or state and local government agencies, that undertake agricultural research and development (R&D). One must be careful to avoid double-counting federal resources that fund agricultural research at the state or provincial level and to ensure that nonresearch activities are excluded. This is a particular problem for research performed by government agencies at the state and local level, which, in many instances, also deliver nonresearch services such as rural extension.

The private nonprofit sector generally includes only a small number of institutions,

---

17 Classifying by source of funds is known as a "funder-based" system of classification as opposed to a "performer-based" system, which classifies according to the nature of those institutions that actually execute the research. Clearly, these classification systems can give rise to different measures of research capacity, and a preferred approach would be to classify research activity by one or the other method. However, at a practical level, when attempting to construct a global database of agricultural research statistics, we were forced to adopt an eclectic approach and use an ad hoc combination of both procedures to arrive at a set of statistics.

which are nevertheless very important for some countries. Some commodity research in less-developed countries, particularly that concerned with export-oriented estate crops such as tea, coffee, and rubber, is often financed wholly or in part by (industry-enforced) export or production levies and performed by private or semiprivate nonprofit research institutions. These institutions often operate as pseudo-public-sector research agencies or, at the very least, substitute directly for such agencies, so it was appropriate to include them in our measure of public agricultural research.

The higher-education sector is fairly readily identified but does present special problems when agricultural research statistics are compiled. Care was taken in constructing the Indicator Series to isolate research from nonresearch activities (e.g., teaching and extension) and to prorate personnel and expenditure data accordingly.

The national agricultural research statistics reported in the Indicator Series exclude the activities of research institutions with an international or regional mandate, such as CIMMYT, IRRI, and WARDA, along with bilateral institutions such as ORSTOM and CIRAD. The research operations for many of these multilateral agencies are quantified and discussed in chapter 9. While their research output may often have substantial impact on the agricultural sectors of their host countries, their mandates direct their research activities towards international and regional, rather than national, applications of their findings. However, all foreign research activities (including those associated with organizations such as those noted above) that are either funded or executed in collaboration with the national research agencies (or administered by them) were included in the series.

Agricultural

When measuring science indicators by socioeconomic objective, the OECD (1981, p. 113) recognizes that two approaches to classification are possible. They can be classified

(a) according to the *purpose* of an R&D program or project;
(b) according to the general *content* of the R&D program or project.

For example, a research project to improve the fuel efficiency of farm machinery could be placed under "agriculture" if classified by purpose, but "energy" if classifies by R&D content. The Indicator Series adopted the procedure used by the OECD and classifies research by purpose rather than content, as it is generally the purpose for which research is undertaken that has the greatest relevance for policy.

The notion of agricultural research used for the Indicator Series includes research in primary agriculture (crops, livestock, plus factor-oriented topics) as well as forestry and fisheries.[18] In general terms, this corresponds with the coverage used by both OECD (1981)

---

[18] Prior compilations of NARS indicators, e.g., Evenson and Kislev (1971), Boyce and Evenson (1975), and Judd, Boyce, and Evenson (1983, 1986), have sought to limit their coverage by excluding forestry, fisheries, and sometimes veterinary research. A substantive argument in favor of adopting the wider

and UNESCO (1984). For policy and analytical purposes, it would be desirable to differentiate agricultural research among commodities, but for many systems this is practically impossible, particularly on a time-series basis going back to 1960. Quite a few countries only report data on national research expenditures that are not differentiated according to socioeconomic objectives, even at this rather aggregate level. For these systems, it was simply not possible, given currently available data, to generate plausible time series at this level. Nevertheless, for a sample of 83 less-developed countries, we constructed preliminary estimates of research personnel stratified by four commodity aggregates (namely, crops, livestock, forestry, and fisheries) for the post-1980 period. Our findings are reported in chapter 8.

A further difficulty is that a significant amount of agricultural research has an effect at the postharvest stage, while the technology is embodied in inputs that are applied at the farm level. Take, for example, the efforts of plant breeders to improve the storage life of horticultural crops or to alter the baking quality of cereals. These characteristics are embodied in new crop varieties that are adopted by farmers. Furthermore, there is a lack of uniformity in the way research that is applied directly at the postharvest stage is currently reported. The OECD (1981, p. 115) classification omits "... R&D in favor of the food processing and packaging industries" from their socioeconomic objective of *agriculture, forestry and fisheries,*[19] while UNESCO (1984, p. 64) includes "... R&D on the processing of food and beverages, their storage and distribution." The Indicator Series sought to implement a variant of these approaches, excluding, where possible, research applied directly at the postharvest stage. Omitting research on food processing and packaging improves the compatibility of these statistics with value-added measures such as agricultural GDP and the like. Nevertheless, public-sector research targeted directly to food and beverage storage (and in some cases, processing) may in practice be included in this series, although this is more likely to be true of advanced systems in the more-developed countries.

A final difficulty was to obtain statistics for the higher-education sector, classified by purpose or "socioeconomic objective." The more general case is to find personnel and, possibly, expenditure data classified by field of science, where the basis of classification is the nature rather than the purpose or objective of the research activity itself.[20] In those cases where it was necessary to rely on field-of-science data, the series attempted to follow the UNESCO (1984, p. 77) procedure and consider agronomy, animal husbandry, fisheries, forestry, horticulture, veterinary medicine, and other allied sciences, such as agricultural

---

definition of agricultural research, as reflected in the statistics reported in this volume, is that the resulting series is then consistent with the agricultural aggregates of GDP, population, and so on, as published by the World Bank and United Nations organizations

19 OECD (1981) includes it instead under the socioeconomic objective of "promotion of industrial development."

20 Classifying research on the basis of the nature of the R&D activity itself, rather than its principal economic objective, is called a "functional" approach (OECD 1981, p. 53).

sciences, thereby excluding fields such as bacteriology, biochemistry, biology, botany, chemistry, entomology, geology, meteorology, and zoology. These latter fields are more appropriately classified as natural sciences, although in some cases the classification is a little hazy. It was therefore necessary to apply a "purpose or objective test" to some of these so-called natural science disciplines and to include in the series research undertaken in these areas when the ultimate purpose or objective of that research could have a direct impact on the agricultural sector.

Research

It is possible to identify a continuum of basic, or upstream, research to applied, or downstream, research. Much agricultural research has been characterized as mission-oriented in the sense that it is problem-solving in orientation, whether or not the solution to the problem requires basic or applied research. OECD (1981, p. 28) states that "... the basic criterion for distinguishing R&D from related activities is the presence in R&D of an appreciable element of novelty." For instance, simply monitoring the incidence of plant and animal diseases in and of itself is not considered research and may only be undertaken to enforce quarantine regulations or the like. But, using this information to study the causes or control mechanisms associated with a particular disease is considered research. Of course, some screening of the literature, newly available plant and animal material, and alternative production practices should be included as part of measured research activity, given its importance in the many countries that are undertaking substantial efforts to adapt existing agricultural technology to their local conditions.

Agricultural research also includes a significant amount of maintenance research that attempts to renovate or replace any deterioration in gains from previous research.[21] Gains in output are often subject to biological degradation as pests and pathogens adapt to research-conferred resistance and control mechanisms. The role of maintenance research is substantial not only in many more-developed countries where current production practices employ technologies that are biologically intensive, but also in many less-developed countries, particularly those situated in the tropics where relatively rapid rates of pest and pathogen adaptation tend to shorten the life of research-induced gains.

The difficulties of differentiating research from nonresearch activities is especially pertinent in the case of agricultural research, given the dual role of many public-sector agencies charged with agricultural research responsibilities. It is common to find such agencies involved in additional nonresearch activities such as teaching; extension services; certification, multiplication, and distribution of seeds; monitoring and eradication of plant and animal diseases; health maintenance, including veterinary medicine; and analysis and certification of fertilizers. In general, it is separating the research component from the joint

---

21 For additional discussions on maintenance research see Ruttan (1982), Miranowski and Carlson (1986), Plucknett and Smith (1986), and Adusei and Norton (1990).

teaching-research activities (in the case of universities) and the joint extension-research activities (of ministerial or department-based agencies) that is most difficult. If direct measures of expenditure and personnel data were not available at the functional level, then secondary data were often used to estimate the appropriate breakdown of aggregate figures into their research versus nonresearch components.

Even in the case of those institutions whose mandate is ostensibly limited to research, there were problems in obtaining consistent coverage of research-related activities. For example, general overhead services, including administrative personnel or expenditures required to support research, can be excluded from reported figures for a variety of reasons. In some instances, the institutional relationship between a national research agency and the ministry within which it is located means that overhead services and the like are charged against the ministry and not the research agency. Alternatively, some research agencies report total personnel and expenditure statistics based on an aggregation of project-level rather than institution-level data. In such cases, administrative overheads cannot be allocated across projects and thus may be omitted entirely or in part from the agency-level statistics.

A further issue involved identifying the research component of the farm operations that are usually undertaken in support of agricultural research. To the extent that such farm operations are necessary to execute a program of research, it seems appropriate that they be included in a measure of the commitment of national resources to agricultural research. However, some systems undertake farm operations at levels well above those required to support research, with the surplus earnings from farm sales being siphoned off to support research and even various nonresearch activities. In some instances, including all the resources devoted to the farm operations of a NARS substantially overstates the level of support to agricultural research within the system.

There was also the need to make a clear distinction between economic development and experimental development. According to OECD (1981, p. 25), "... experimental development is systematic work, drawing on existing knowledge gained from research and/or practical experience that is directed to producing new materials, products, or devices, to installing new processes, systems, and services, or to improving substantially those already produced or installed." Experimental development is therefore concerned with applying new findings from formal and informal research activities. This contrasts with the notion of economic development, which in general terms, is concerned with improving the well-being or standard of living of members of a society in a particular country or region.

Clearly, while improvements in agricultural productivity that follow from experimental development contribute to the process of economic development, they represent only part of the story. Improvements in rural infrastructure, via investments in irrigation, transportation and communication facilities plus improved rural health and education services, also contribute to the economic development of the agricultural sector and, ultimately, to society as a whole (Antle 1983).

A problem arises when one attempts to compile statistics on agricultural research and experimental development activities in less-developed countries. This occurs when a substantial portion of R&D activity is financed and/or executed as part of an economic-development aid package. It is often difficult to identify the experimental versus economic-development component of an aid package, particularly given the project orientation of much development aid. For instance, development assistance to establish, upgrade, or rehabilitate irrigation facilities can often incorporate research to evaluate water quality and identify preferred crop varieties as well as agronomic and irrigation practices. However, including all of the project's resources in a measure of NARS capacity could seriously overestimate the level of resource commitment to agricultural research.

Another less obvious difficulty concerns the somewhat transient nature of some of the agricultural research funded through development projects, which tends to be of relatively short duration, often between one to five years. In some cases, it is undertaken largely by expatriates and is seldom a part of the existing national research infrastructure. This type of research presumably contributes to the overall level of national research activity and should be captured in a NARS indicator, particularly if one is concerned with measuring sources of growth or technical change within a country. However, to the extent that such research is not integrated into the existing national research infrastructure, it is not a good measure of the "institutionalized research capacity" of a national system. The strategy pursued in this case was to include such development-financed research only when the research component could be isolated from the nonresearch component with an acceptable level of precision, and when it appeared to be integrated into the existing agricultural research infrastructure within a country.

5.4.2 Research Personnel Indicators

One possibility for measuring the human resource commitment to a NARS is simply to report the total number of personnel employed within a research system. This personnel aggregate would not only sum together scientific staff regardless of their qualifications and skills, but would also include, in an unweighted fashion, research technicians and other support staff. Because support staff often substitute directly for other capital and operating expenses in the research process, such a series may be driven largely by differences in the relative cost of research labor and nonlabor inputs, resulting, for example, in quite volatile fluctuations in the ratio of researchers to nonresearchers. As a consequence, all-inclusive research personnel aggregates would not accurately reflect differences in the underlying scientific capacity that is relevant for many purposes and is our measurement objective here. Thus, the Indicator Series sought to include only research personnel, i.e., researchers engaged directly in the conception or creation of new knowledge, products, processes, methods, and systems.[22] The series attempted to exclude technicians as well as support and clerical staff who normally perform research and technical tasks under the supervision of a

---

[22] This corresponds to the OECD (1981, p. 67) definition of a researcher.

researcher.

A practical procedure for differentiating research from nonresearch staff was to rely principally on educational levels rather than occupational classes. While there are clearly substantial difficulties in standardizing educational levels on a global basis, an international standard classification of education (ISCED) has been developed and is in general use (UNESCO 1976). The Indicator Series sought to include only NARS personnel who held at least a third-level university degree (ISCED-level categories 6 and 7) as researchers.[23] This included holders of first and postgraduate degrees (or their equivalent) earned at bona fide universities or at specialized institutes of university status.

The series further attempted to classify national research personnel by degree status — PhD, MSc, BSc, or equivalent. This substantially improved our ability to use the personnel data as an indicator of the human capital or "quality-adjusted" research commitment to national agricultural research, as reported in chapter 8. There was also an attempt to differentiate between local and expatriate scientific staff in order to enhance the information contained in the personnel series. As discussed earlier, the series sought to include only those expatriates who were working directly on domestic issues in an integrated fashion with the national research system.

Personnel who were classified as research managers or administrators presented special problems. To the extent that they are engaged in the planning and management of the scientific and technical aspects of a researcher's work, they should be classified as researchers and included, at least on a prorated basis, in the series. They are usually of a rank equal to or above that of persons directly employed as researchers and are often former or part-time researchers (OECD 1981, p. 67). However, in many cases, it is not at all clear if research managers or administrators maintain any direct involvement with the scientific process itself.

The problems of dealing with data on research administrators are analogous to those of dealing with data on other NARS personnel who may hold dual research and nonresearch appointments. This is particularly important when including personnel from institutions in the ministry or department of agriculture, who perform, for example, a dual research-extension function, or from universities where personnel often hold joint research-teaching appointments. In all cases, an attempt was made to measure researchers in full-time equivalent (FTE) units. If direct measures in FTE units were not available, secondary data, which enabled total researcher figures to be plausibly prorated to FTE units, were used.

---

23 An alternative procedure (see OECD 1981, pp. 67-69) is first to classify researchers, technicians, and other supporting staff on the basis of the ILO (1986) classification scheme and then use ISCED procedures to classify researchers by educational level. Given the rather heavy reliance on secondary data in the Indicator Series, it was not possible to operationalize the ILO classification scheme.

5.4.3 Research Expenditure Indicators

There are several commonly accepted methods of measuring the (annual) commitment of financial resources to R&D (OECD 1981, pp. 72-82):

(a) *performer-based* reporting of the sum of funds received by all relevant R&D agencies for the performance of intramural R&D;
(b) *source-based* reporting of the funds supplied by all relevant agencies for the performance of extramural R&D;
(c) *total intramural* expenditures for R&D performed within a statistical unit or sector of the economy, whatever the source of funds.

The Indicator Series sought to report actual research expenditures, not simply budgeted funds, appropriations, or funds available, and so was based on method *c* wherever possible. A substantial number of the major discrepancies in prior compilations were due to large variations — sometimes upward of 30% to 50% — between funds budgeted or appropriated and funds actually spent. Some funds allocated to research at the beginning of a fiscal year, for example, may never materialize, especially if governments are forced to trim proposed outlays over the course of the year because of unforeseen budgetary shortfalls. Conversely, some research systems may actually receive more funds than are spent, and thus carry funds over to future budgetary periods. This is particularly true for systems experiencing substantial capital investments where funds are allocated initially in a lump-sum fashion and then drawn on over a period of time as needed.

The expenditures reported in the Indicator Series are total, inclusive of salary, operating, and capital expenses. While the series reports actual expenses, for some purposes it may be more appropriate to measure resources used rather than funds spent. This would involve explicitly separating capital from noncapital expenditures. Capital expenditures, which measure (gross) additions to the stock of capital invested in agricultural research, could then be converted into flow terms by estimating the future service flows derived from them.[24] These capital service flows could then be added back to noncapital expenditures to derive an overall measure of the resources actually used for research over time.

One of the major undertakings in compiling the Indicator Series was to collect all expenditure data in current local currency units. This allowed the standardized translation procedures described in section 5.2.1 to be applied to all countries. All expenditures were deflated using local implicit GDP deflators based in 1980. The series was then converted using PPPs defined over GDP.

---

24 See Pardey, Craig, and Hallaway (1989) for details. Unfortunately, there are simply not enough data available at the international level to construct a time series that differentiates research expenditures by factor type. However, in chapter 8 we do report our preliminary attempts to differentiate research expenditures by factor type in a sample of 43 less-developed countries for the post-1980 period.

5.4.4 Shortcut Estimation Methods

The data base that underpins the analysis reported in this volume substantially upgrades and extends previously available indicators on research personnel and expenditures. But the nontrivial number of missing observations that remain impedes our ability to form aggregates or undertake comparative analyses.

The practical and, by necessity, ad hoc approach of dealing with this problem has been to implement a hierarchical series of shortcut estimation procedures. The preferred and most data-demanding approach uses econometric procedures to estimate a series of reduced form equations first. These are then used to construct ordinary least squares (OLS) predictions of missing research personnel and expenditure observations. In the absence of suitable regressors, various non-econometric extrapolation and interpolation procedures have been applied directly to the country-level data.

Naturally the appropriate level of precision for any data set is a function of the uses to which it is to be put. In this instance, the objective was to construct aggregate measures that represent broad trends at the regional or subregional level over the 1961-85 period. With an underlying unit of analysis consisting of annual, national-level data, we opted first to construct simple, quinquennial averages of the country-level observations beginning with the 1961-65 period.[25] While this aggregation procedure may artificially dampen variability in data where there are strong trends, we argue that five-year averages offer a basis for more realistic global comparisons than the point estimates used by previous analysts.[26] Such aggregation also serves to minimize the influence of spurious variability as well as substantially reducing the number of observations to be estimated by shortcut methods. Specifically, the primary data matrix includes 151 countries over a 25-year period for a total of 3775 entries per indicator. Averaging over a five-year period reduces the size of this matrix to 755 entries, which in turn, given the available data, reduces the number of personnel data points to be estimated by short-cut methods from 67% to 30% and the number of expenditure data points from 65% to 45%.

The regression procedures used to derive shortcut research personnel and expenditure estimates do not presume any causal relationship between the set of right-hand-side (RHS) "explanatory" variables and the research indicators for which estimates are being sought. But, as Ahmad (1980) and Clague (1986) point out when using analogous shortcut procedures in a different context, an informed choice of candidate RHS variables should draw on some understanding of the likely partial correlations these variables may exhibit with research expenditure and personnel indicators.

On the presumption that research personnel and expenditure aggregates at the national level exhibit relatively stable trends over time, then one-period lags or leads of these

---

[25] The 1961-65 period averages were centered on 1963 and so on for later periods.

[26] For instance Boyce and Evenson (1975), Evenson and Kislev (1971, 1975a), Judd, Boyce, and Evenson (1983, 1986), and Oram and Bindlish (1981).

variables are credible regressors for our purposes.[27] With personnel expenditures accounting for an average of 56% of total research expenditures in less-developed countries and probably an even higher percentage for more-developed countries as US data suggest (section 8.2), current research expenditures were also deemed suitable for inclusion in the set of variables regressed against research personnel. Similarly, research personnel were used as a regressor in the equation used to predict research expenditures. Measures of the absolute (AgGDP) and relative (AgGDP/GDP) size of the agricultural sector also appear to be systematically related to the level of agricultural research activity and were added to the set of RHS variables used to generate shortcut estimates.[28] To capture the effects of a myriad of complex socioeconomic influences that would otherwise be difficult if not impossible to quantify, regional dummies were also incorporated into the estimating procedure. Finally, a series of Chow tests rejected the hypothesis that a set of time dummies added significant "explanatory" power to the regressions, and these dummies were omitted from the final specifications summarized in table 5.4.

Table 5.4: *Specification of Shortcut Estimating Equations*

	Research personnel (RP_t)						Research expenditures (RE_t)					
	specification number						specification number					
Regression	1	2	3	4	5	6	1	2	3	4	5	6
Constant	*	*	*	*	*	*	*	*	*	*	*	*
$AgGDP_t$	*	*	*	*	*	*	*	*	*	*	*	*
$(AgGDP/GDP)_t$	*	*	*	*	*	*	*	*	*	*	*	*
RP_{t-1}	*		*									
RP_t							*	*			*	
RP_{t+1}		*		*								
RE_{t-1}							*		*			
RE_t	*	*			*							
RE_{t+1}								*		*		
Regional dummies	*	*	*	*	*	*	*	*	*	*	*	*

Note: To minimize the influence of spurious observations these equations were estimated in double-log form.

Before fitting the six research personnel and six research expenditure specifications used to generate shortcut estimates, the data were stratified across five income classes, so that a total of 60 empirical relationships were estimated. This enabled the predictive influence of each RHS variable to be conditioned by stage-of-development considerations.

27 In some cases, data permitting, the research personnel and/or expenditure figures were predicted by averaging the estimates obtained from equations that involved regressors consisting of leads and lags of the respective RHS variable.

28 Evenson and Kislev (1975b) and Pardey, Kang, and Elliott (1989) present empirical support for this notion, while in chapter 1 some of the conceptual underpinnings of such a relationship are discussed.

Adjusted R^2s for the specifications reported in table 5.4 averaged 0.94 across the total of 60 personnel prediction equations and 0.90 across the 60 expenditure prediction equations that were estimated by OLS procedures, with the best fits being obtained for the high-income countries. In all but one instance, the current and/or lagged RHS research personnel and expenditure variables entered with positive signs, as, in general, did the coefficient on the AgGDP variable. The sign on the AgGDP/GDP variable was somewhat more volatile but was uniformly negative and significant for the low-income countries.

Prior to incorporating these OLS estimates into the respective research personnel and expenditure series, a variety of screening procedures were used to ensure that they gave rise to plausible time series. The actual and first-differenced series for research expenditures, research personnel, and the implied expenditures per researcher for each country were jointly scrutinized for evidence of outliers that could not be accounted for by secondary information contained in various sources, including the country-level documentation of the Indicator Series. OLS estimates that were identified as outliers were then estimated by straightforward extrapolation and interpolation procedures.[29]

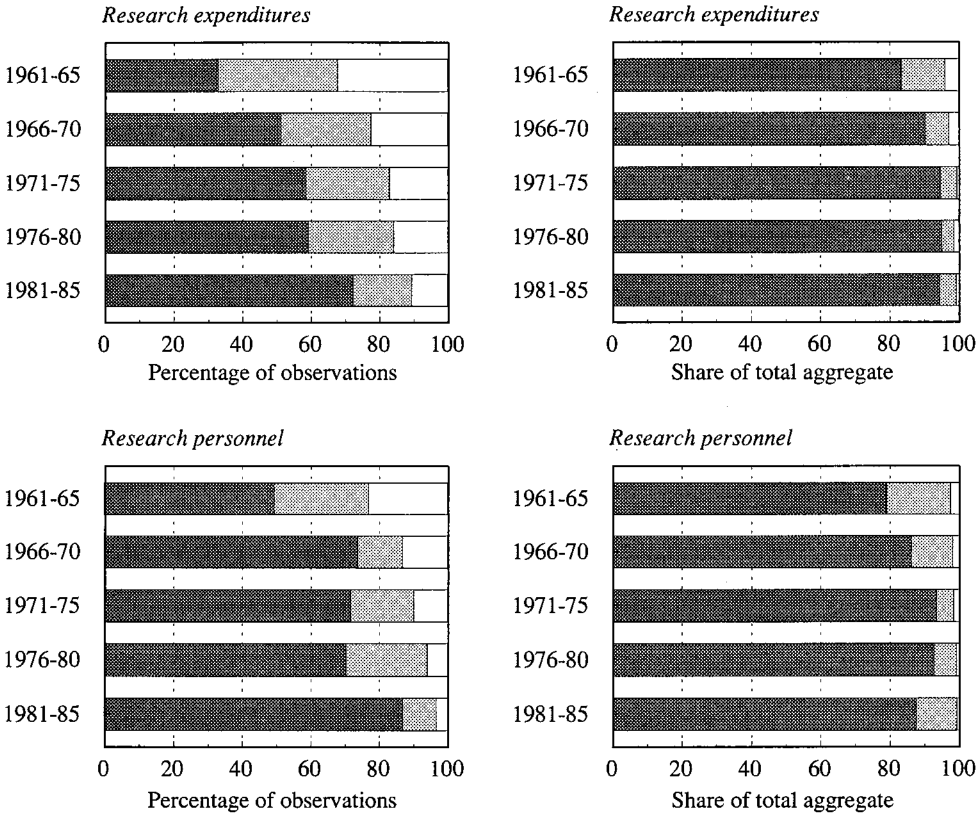
Figure 5.4 gives the percentage of observations and the share of research personnel and expenditure totals that were derived by various methods. Around 55% of the expenditure observations were accounted for by direct estimates, while a significantly higher proportion of research personnel observations (70%) were directly estimated. Regression-based procedures generated 26% of the research expenditure estimates and 19% of the research personnel figures. The remaining observations were derived using various non-econometric extrapolation and interpolation procedures.

For our purposes, it is significant that a substantially higher share of research expenditure and personnel totals were accounted for by direct observations because the problem of missing observations was concentrated in the group of small, less-developed NARSs. It must be emphasized, however, that a nontrivial share of the personnel and expenditure totals of the less-developed countries were constructed by shortcut procedures. One should bear this in mind when interpreting the various indicators presented elsewhere in this book.

---

29 A variety of procedures were employed that sought to make maximum use of the available data. Various country-specific ratios of expenditures per researcher were used in conjunction with expenditure-only or researcher-only figures for a particular year to infer corresponding researcher and expenditure data. In some instances ratios of expenditures per researcher for the region in which the country is located and/or expenditure ratios for countries of comparable size, stage of development, or time period were used in a similar manner. For a few, often small, countries where only recent observations were available, the series was backcast by assuming that the country's rate of growth for the indicator in question was approximated by the regional average.

Figure 5.4: *Proportion of estimates derived by direct and shortcut methods*



Legend: ▩ direct estimates; ▒ regression-based estimates; ☐ other estimates.

5.4.5 Comparison with Prior Compilations

Internationally comparable data sets of investments in NARSs are sparse and up to now have been based largely on the series constructed by Evenson and colleagues[30] supplemented by the efforts of Oram and Bindlish (1981). The data sources, NARS and variable definitions, data processing, and aggregation procedures used by Pardey and Roseboom to construct the compilation described in this chapter and detailed in the appendix to this book differ in substantive ways from those used in prior compilations. Rather than simply extending both the country and temporal coverage of these existing data sets, Pardey and Roseboom endeavored to rework the recent historical record concerning the global pattern of invest-

[30] See footnote 26 for list of references.

ments in public agricultural research. To permit a more informed use of these new data, the nontrivial and systematic differences between the Pardey and Roseboom series and the most recently published series by Judd, Boyce, and Evenson (1986) will be quantified in this section and placed in context.

The principal data sources for the Pardey and Roseboom series is the ISNAR Indicator Series (Pardey and Roseboom 1989a) supplemented by recently updated data on the US (Pardey, Eveleens, and Hallaway 1991) and China (Fan 1991b). In compiling the Indicator Series, Evenson-related sources (including the Judd, Boyce, and Evenson series) were directly used for only 11% of the personnel data and 6% of the expenditure data. In quite a few instances, however, they drew from the same primary sources used by Evenson and his colleagues, although these were often supplemented with information from additional primary sources that were considered more complete for our purposes. The multiplicity of sources that were used to compile the primary data underlying both the Pardey and Roseboom and the Judd, Boyce, and Evenson series compounds the difficulties of maintaining consistent coverage in several dimensions: namely, (a) over time within a country, (b) among countries, and (c) across the personnel and expenditure series. A key to improving consistency in the data underlying the Pardey and Roseboom series was to identify and track the institutional coverage of the available data, paying particular attention to dates of creation, organizational mergers or divisions, details of name or mandate changes, and the like. This involved gathering quantitative and qualitative data from as many documentable sources as possible, including several ISNAR benchmark surveys, then reconciling and synthesizing these multiple observations into a data series that represented as closely as possible the NARS concept identified earlier in this chapter. Boyce and Evenson (1975, p. iv) note that their "... data [may] appear to be in error simply because no attempt to determine how many agencies are involved in the support of research in the country has been made." Improvements over time in the institutional coverage of available data sources means that spurious, and in some cases substantial, growth in national research capacity can be introduced if issues related to institutional coverage are ignored.

The concepts of a NARS underpinning the two series differ substantially. The notion of agricultural research used in the Pardey and Roseboom series includes primary agricultural research (crops, livestock, plus factor-oriented areas) as well as forestry and fisheries, while the Judd, Boyce, and Evenson series tries to exclude forestry and fisheries research (Judd, Boyce, and Evenson 1986, pp. 79-80). There are several practical and conceptual issues which on balance, at least from our perspective, favor the broader definition. First, such a definition is consistent, in general, with the coverage used by both OECD (1981) and UNESCO (1984). Second, while for policy and analytical purposes it is desirable to differentiate agricultural research among commodities, for many systems this is practically impossible, particularly in time-series back to 1961. Certainly for a significant number of systems, disparate agencies are charged with the responsibility of funding and/or executing these various areas of research and care needs to be taken to ensure that they are included, where appropriate, in the aggregate measures. On the other hand, attempts to exclude forestry and fisheries research are confounded by the fact that quite a few countries only report preaggregated national research data that fail to differentiate among socioeconomic objectives. Finally, a substantive argument in favor of adopting the wider definition of agricultural research is that the resulting series is then generally consistent with the definitions of agricultural GDP, population, etc., commonly published by the World Bank, United Nations, and FAO.[31]

There are also significant and quantitatively important differences between the Pardey and Roseboom and the Judd, Boyce, and Evenson series concerning the manner in which research expenditures are translated into commensurable units. Research expenditures for the Pardey and Roseboom series were all compiled in current local currency units and then translated, in a standardized fashion, to constant 1980 US dollar aggregates. Our two-step translation procedure first deflates to base year 1980 using local implicit GDP deflators and then converts to US dollars using Summers and Heston's (1988) PPP measures. This is far from ideal, but in our judgment, it is likely to introduce less aggregation bias[32] than the approach used by Judd, Boyce, and Evenson that first converts using annual exchange rates and then deflates using a US wholesale price index.[33]

A final source of discrepancy between the Pardey and Roseboom and the Judd, Boyce, and Evenson compilations lies in the different regional and subregional aggregations they use. In brief, the principal differences are as follows: Judd, Boyce, and Evenson include Japan and China in their Asian aggregates while Pardey and Roseboom include Japan in the more-developed country grouping and, given the system's overwhelming size coupled with the preliminary nature of Chinese agricultural research statistics, list China as a stand-alone figure. Judd, Boyce, and Evenson include South Africa in their Southern Africa totals while the Pardey and Roseboom series omits South Africa from its sub-Saharan Africa aggregates in order to maintain consistency with existing World Bank and UN practice. Judd, Boyce, and Evenson do not report a West Asia & North Africa grouping. Judd, Boyce, and Evenson group Australia and New Zealand along with the Pacific islands into an Oceania aggregate while Pardey and Roseboom include Australia and New Zealand in their more-developed country totals and merge the Pacific islands into an Asia & Pacific aggregate. And, finally, Judd, Boyce, and Evenson include Ireland and the United Kingdom in a Northern Europe total, while Pardey and Roseboom group them under a Western Europe aggregate.

The cumulative effects of these different measurement and compilation procedures are captured in the comparative data on level and rate of growth presented in tables 5.5 and 5.6, where the Pardey and Roseboom data have been reaggregated to match the regional

---

31 An inconsistency that remains involves agricultural research or agricultural output (inclusive of forestry and, of particular concern here, fisheries) indexed over "agricultural" land.

32 See section 5.1.

33 In numerous cases Boyce and Evenson (1975) and Judd, Boyce, and Evenson (1983) recorded research expenditures directly in current or constant US dollars rather than current local currency units. This leaves their figures subject to the possibly capricious conversion methods of their source authors.

Table 5.5: *Level of Investment in Agricultural Research — Comparison of Pardey and Roseboom with Judd, Boyce, and Evenson Estimates, Percentage Differences*

Region[a]	Researchers[b]				Expenditures[b]				Expenditure per researcher[b]			
	1961-65	1966-70	1971-75	1976-80	1961-65	1966-70	1971-75	1976-80	1961-65	1966-70	1971-75	1976-80
	%	%	%	%	%	%	%	%	%	%	%	%
North Africa	14	44	49	40	36	41	48	46	26	–5	–2	10
West Africa	6	8	–13	–6	18	23	29	1	13	16	37	7
East Africa	31	17	–5	–34	55	50	35	29	35	40	38	47
Southern Africa	–8	–20	–29	–51	74	60	56	41	75	66	66	61
Africa	*14*	*28*	*21*	*12*	*36*	*38*	*37*	*20*	*25*	*14*	*21*	*9*
West Asia	37	16	11	18	52	65	59	53	25	59	54	44
South Asia	61	62	63	52	73	71	73	67	31	23	26	30
Southeast Asia	34	14	13	27	83	83	83	68	74	80	80	56
East Asia	29	14	2	–3	41	29	25	29	18	17	24	31
Asia (excl. China)	*36*	*27*	*21*	*21*	*55*	*50*	*48*	*47*	*29*	*32*	*34*	*33*
Caribbean and Central America	–2	–41	–25	–10	53	40	46	49	54	58	57	54
Temperate South America	9	4	1	–6	27	31	25	28	20	28	25	32
Tropical South America	8	2	–2	–20	50	48	44	46	46	48	45	55
Latin America	*7*	*–4*	*–5*	*–15*	*44*	*44*	*40*	*44*	*40*	*46*	*43*	*51*
Northern Europe	35	13	3	–25	45	41	39	29	15	32	37	43
Central Europe	18	7	0	–28	–5	–8	10	–7	–28	–16	10	16
Southern Europe	3	–10	–4	18	36	–6	–12	8	34	4	–8	–13
Western Europe	*21*	*6*	*1*	*–17*	*20*	*10*	*18*	*7*	*–2*	*4*	*17*	*20*
Oceania	23	19	16	42	16	–2	–10	–30	–8	–25	–31	–123
North America	47	45	42	38	17	15	8	17	–57	–53	–59	–33
North America & Oceania	*43*	*40*	*36*	*39*	*17*	*13*	*5*	*10*	*–46*	*–45*	*–50*	*–47*
Total	33	25	19	15	31	27	26	25	–3	3	8	12

Source: Judd, Boyce, and Evenson (1983) and the appendix to this book.

[a] Regional groupings correspond to aggregations presented in Judd, Boyce, and Evenson (1983, 1986).
[b] Pardey and Roseboom minus Judd, Boyce, and Evenson estimates expressed as a percent of Pardey and Roseboom.

Table 5.6: *Growth of Investment in Agricultural Research — Comparison of Pardey and Roseboom with Judd, Boyce, and Evenson Estimates*

	Researchers		Expenditures	
	PR[a]	JBE[b]	PR[a]	JBE[b]
	%	%	%	%
North Africa	9.4	6.8	4.7	3.6
West Africa	8.1	9.0	6.8	8.1
East Africa	5.2	9.9	4.5	7.7
Southern Africa	2.5	4.9	6.4	12.2
Africa	*7.9*	*8.1*	*5.6*	*7.2*
West Asia	4.9	6.8	7.8	7.7
South Asia	6.8	8.2	8.0	9.4
Southeast Asia	8.8	9.5	6.4	10.8
East Asia	1.1	3.6	5.5	7.0
Asia (excl. China)	*3.6*	*5.1*	*6.5*	*7.6*
Caribbean and Central America	7.7	8.3	9.7	10.3
Temperate South America	4.2	5.3	3.7	3.6
Tropical South America	7.8	9.7	8.7	9.3
Latin America	*6.9*	*8.4*	*7.7*	*7.8*
Northern Europe	3.0	7.6	4.7	6.4
Central Europe	1.5	4.5	6.1	6.3
Southern Europe	3.4	2.3	5.2	7.8
Western Europe	*2.5*	*5.2*	*5.5*	*6.5*
Oceania	4.9	2.9	3.6	6.6
North America	1.0	2.1	3.3	3.3
North America & Oceania	*1.8*	*2.3*	*3.3*	*3.9*
Total	3.3	5.0	5.2	5.8

Source: See table 5.5.

[a] PR = Pardey and Roseboom; figures given here are compound annual rate of growth between 1961-65 and 1976-80 period averages.

[b] JBE = Judd, Boyce, and Evenson; figures given here are compound annual rate of growth between 1959 and 1980 estimates.

groupings used by Judd, Boyce, and Evenson. The Pardey and Roseboom series estimates an overall level of full-time equivalent researchers some 15% to 33% higher than the Judd, Boyce, and Evenson series, with the downward bias in the Judd, Boyce, and Evenson series being magnified in the earlier years, particularly for East Africa, Asia, Northern Europe, and North America. The systematic attempt to exclude forestry and fisheries research from the Judd, Boyce, and Evenson series, combined with the more extensive, but still not ideal, coverage of research performed by "secondary" public agencies, including universities, in the Pardey and Roseboom series, could account in large measure for these differences. Contrary to this general pattern, the Judd, Boyce, and Evenson series appears to substantially overestimate the number of research personnel for the 1976-80 period throughout Africa (excluding North Africa) and Latin America, plus Northern and Central Europe. A comparative examination of the underlying data for this period suggests that the Judd, Boyce, and Evenson series for these regions relies more heavily than the Pardey and Roseboom series on extrapolated rather than directly observed data, thereby (presumably erroneously) carrying forward the somewhat higher growth rates of earlier periods.

The Judd, Boyce, and Evenson series also reports a much lower level of real research expenditures than the Pardey and Roseboom series. The downward bias in the Judd, Boyce, and Evenson figures is much more dramatic for the less-developed countries than for the more-developed countries, presumably due to differences in coverage, compounded by substantial differences in currency translation procedures.

Although the Pardey and Roseboom series suggests that overall there are significantly more public resources devoted to agricultural research at the national level than the Judd, Boyce, and Evenson figures reveal, the rates of growth in research personnel and expenditures appear somewhat lower than has hitherto been reported.[34] It is to be hoped that the substantial efforts invested by Pardey and Roseboom (1989a) in tracking institutional coverage within a country over time has resulted in a series that is, in some senses, more consistent than the Judd, Boyce, and Evenson series. There is a tendency for more readily available, often preaggregated, data to improve implicitly in coverage over time, and it is particularly such aggregate data that were used by Judd, Boyce, and Evenson. Therefore, we believe the Pardey and Roseboom growth rates are more realistic than those reported by Judd, Boyce, and Evenson. The Pardey and Roseboom data form the basis for most of the analyses reported in subsequent chapters, particularly in chapters 7 and 8.

[34] Notwithstanding the fact that neither series adjusts for quality differences over time or across countries.

Table A5.1: *Agricultural Land*

Region	Total agricultural land[a]					Arable land and land under permanent crops: Percentage of total area					Arable land and land under permanent crops: Percentage irrigated[b]				
	61-65	66-70	61-75	76-80	81-85	61-65	66-70	71-75	76-80	81-85	61-65	66-70	71-75	76-80	81-85
	(millions of hectares)					%	%	%	%	%	%	%	%	%	%
Nigeria	48.5	49.3	50.5	51.1	51.6	60.6	60.3	59.3	59.2	59.4	2.7	2.7	2.7	2.7	2.7
Western Africa (17)	191.2	194.1	196.3	198.3	198.9	14.8	16.2	17.2	18.4	19.0	0.7	0.8	1.1	1.3	1.5
Central Africa (7)	38.5	39.0	39.5	39.9	40.2	25.3	26.5	27.7	28.9	29.7	0.1	0.2	0.4	0.5	0.7
South Africa (10)	206.1	212.4	216.0	217.2	217.6	8.2	9.0	9.6	9.8	10.0	2.4	2.4	2.8	4.1	5.2
East Africa (8)	210.3	211.9	213.5	214.4	215.0	16.8	17.6	18.3	18.8	19.1	5.1	5.1	5.1	5.3	5.5
Sub-Saharan Africa (43)	*694.6*	*706.9*	*715.8*	*720.9*	*723.4*	*17.2*	*18.1*	*18.8*	*19.4*	*19.8*	*2.7*	*2.7*	*2.8*	*3.1*	*3.4*
China	*422.7*	*421.1*	*419.7*	*418.7*	*419.1*	*24.5*	*24.2*	*24.0*	*23.8*	*23.9*	*30.2*	*35.1*	*40.1*	*44.6*	*44.3*
South Asia (8)	264.3	267.2	269.8	272.4	271.3	79.9	80.4	80.8	81.1	81.2	19.2	21.1	23.3	25.7	28.6
South-East Asia (9)	54.4	58.1	62.4	65.7	68.3	74.6	76.4	78.0	79.2	79.9	20.0	19.0	19.7	20.9	23.1
Pacific (10)	1.4	1.4	1.5	1.5	1.6	65.4	67.4	65.9	66.5	67.5	0.1	0.1	0.1	0.1	0.1
Asia & Pacific, excl. China (27)	*320.1*	*326.7*	*333.7*	*339.7*	*341.2*	*78.9*	*79.6*	*80.2*	*80.7*	*80.8*	*19.4*	*20.8*	*22.6*	*24.8*	*27.5*
Caribbean (17)	4.9	5.3	6.1	6.2	6.3	53.6	51.8	45.7	48.0	48.7	8.7	9.3	10.4	10.8	11.6
Central America (8)	114.5	114.7	115.9	117.5	119.4	25.9	25.7	25.4	25.8	26.3	11.1	12.3	14.9	17.4	17.4
South America (12)	502.2	531.6	560.5	582.6	599.6	16.1	19.3	20.9	22.2	23.2	6.0	5.3	5.3	5.5	5.7
Latin America & Caribbean (37)	*621.5*	*651.5*	*682.6*	*706.3*	*725.2*	*18.2*	*20.7*	*21.9*	*23.0*	*23.9*	*7.4*	*6.9*	*7.3*	*7.8*	*8.0*
North Africa (5)	86.5	87.2	88.9	89.2	85.8	26.4	26.9	27.5	27.9	29.4	16.9	17.9	18.0	17.3	17.6
West Asia (15)	228.6	229.6	230.9	229.1	227.3	24.6	25.5	26.0	25.6	25.3	15.7	16.3	18.0	18.4	19.5
West Asia & North Africa (20)	*315.1*	*316.8*	*319.7*	*318.3*	*313.1*	*25.1*	*25.9*	*26.4*	*26.2*	*26.4*	*16.1*	*16.8*	*18.0*	*18.1*	*18.9*
Less-Developed Countries (108)	*2374.1*	*2423.0*	*2471.5*	*2503.9*	*2522.0*	*28.1*	*29.2*	*29.8*	*30.3*	*30.7*	*15.6*	*16.5*	*17.8*	*19.0*	*19.9*

Table A5.1: *Agricultural Land (Contd.)*

	Total agricultural land[a]					Arable land and land under permanent crops									
						Percentage of total area					Percentage irrigated[b]				
Region	61-65	66-70	61-75	76-80	81-85	61-65	66-70	71-75	76-80	81-85	61-65	66-70	71-75	76-80	81-85
	(millions of hectares)					%	%	%	%	%	%	%	%	%	%
Japan	6.0	5.8	5.7	5.5	5.4	97.9	95.8	92.6	90.1	88.8	49.9	61.1	62.4	62.7	62.1
Australia & New Zealand (2)	493.0	503.2	511.8	509.3	504.9	6.9	7.9	8.2	8.5	9.3	3.5	3.7	3.8	3.8	4.0
Northern Europe (5)	13.1	12.8	12.5	12.4	12.3	72.7	72.1	71.8	71.5	72.2	1.0	1.6	2.8	5.6	7.1
Western Europe (8)	84.0	81.3	79.1	77.8	77.0	50.5	49.0	48.5	49.0	49.2	3.3	3.7	4.2	4.9	5.7
Southern Europe (4)	66.1	65.0	62.4	61.4	60.8	65.2	65.0	64.3	64.5	65.0	13.1	14.6	17.0	18.5	19.6
North America (2)	503.3	500.1	497.5	497.8	507.1	43.9	45.8	46.6	46.9	46.6	6.8	7.0	7.3	8.4	8.4
More-Developed Countries (22)	*1165.6*	*1168.2*	*1169.1*	*1164.3*	*1167.5*	*30.6*	*31.3*	*31.3*	*31.6*	*32.1*	*7.4*	*7.9*	*8.3*	*9.2*	*9.4*
Total (130)	3539.7	3591.2	3640.5	3668.1	3689.5	28.9	29.9	30.3	30.7	31.2	12.8	13.5	14.6	15.8	16.5

Source: Compiled from *FAO Production Yearbooks*.

[a] Agricultural land includes arable land, land under permanent crops and permanent pastures. It excludes forest and woodland areas. Variable levels and intensities of commercial exploitation of forest and woodland areas both over time within countries and between countries suggests that less distortion in the agricultural input-output relationship is induced by excluding rather than including forest and woodland areas in a measure of agricultural land.

[b] Percent of arable land and permanently cropped land under irrigation.

Table A5.2: *Economically Active Agricultural Population*

Region	1961-65	1966-70	1971-75	1976-80	1981-85
			(millions)		
Nigeria	14.2	16.0	18.1	20.7	23.3
Western Africa (17)	20.0	21.4	22.7	24.2	25.9
Central Africa (7)	11.5	12.0	12.6	13.4	14.4
Southern Africa (10)	13.1	14.2	15.6	17.6	19.5
Eastern Africa (8)	27.4	30.5	33.8	37.5	41.2
Sub-Saharan Africa (43)	*86.1*	*94.0*	*102.9*	*113.4*	*124.3*
China	*294.8*	*318.7*	*348.3*	*383.5*	*417.7*
South Asia (8)	190.3	202.0	215.9	231.9	250.2
South-East Asia (9)	59.7	63.8	67.8	71.4	75.2
Pacific (11)	1.2	1.2	1.3	1.3	1.4
Asia & Pacific, excl. China (28)	*251.1*	*267.0*	*285.0*	*304.7*	*326.8*
Caribbean (18)	2.9	2.9	2.9	3.0	3.1
Central America (8)	8.8	9.2	9.9	11.0	11.9
South America (12)	22.2	23.3	24.0	24.0	24.1
Latin America & Caribbean (38)	*33.9*	*35.4*	*36.8*	*38.0*	*39.1*
North Africa (5)	9.2	9.2	9.3	9.7	10.1
West Asia (15)	18.9	19.4	19.8	20.0	20.3
West Asia & North Africa (20)	*28.1*	*28.6*	*29.1*	*29.6*	*30.5*
Less-Developed Countries (130)	*694.0*	*743.6*	*802.2*	*869.2*	*938.5*
Japan	13.6	11.4	9.3	7.2	5.6
Australia & New Zealand (2)	0.6	0.6	0.6	0.6	0.6
Northern Europe (5)	1.5	1.3	1.1	1.0	0.8
Western Europe (8)	10.1	8.0	6.7	5.8	5.0
Southern Europe (4)	13.1	10.6	8.8	7.5	6.4
North America (2)	5.3	4.7	4.4	4.4	4.1
More-Developed Countries (22)	*44.3*	*36.5*	*30.9*	*26.5*	*22.4*
Total (152)	738.2	780.1	833.1	895.7	960.9

Source: Compiled from FAO (1987b).

Note: Economically active agricultural population includes all persons engaged or seeking employment in an economic activity related principally to agriculture, forestry, hunting, or fishing, whether as employees, own-account workers, salaried employees, or unpaid workers assisting in the operation of a family farm or business *FAO Production Yearbook 1987* (p. 4).