Editor's Overview

Welcome to the inaugural issue of the *International Productivity Monitor*, a new publication produced by the Ottawa-based Centre for the Study of Living Standards (CSLS). The objective of the *Monitor* is to focus atttention on the importance of productivity for improving living standards and quality of life. We will be publishing twice a year articles from leading researchers on productivity issues, trends and developments in Canada and other countries. The *Monitor* will also serve as a vehicle for an international exchange of ideas and information on productivity topics.

The first issue of the *Monitor* contains eight articles on a wide range of productivity topics. There are the new economy and trend productivity growth in Canada, the renaissance of service sector productivity in the United States, a regional comparison of U.S-Canada standards of living, the postwar productivity convergence experience among OECD countries, price cap regulation and productivity growth, and finally a symposium of three articles on the measurement and interpretation of total factor productivity.

In addition to the hard-copy version of the Monitor, which is available in English and French, all articles are available on-line at the CSLS website (www.csls.ca). Moreover, unabridged versions of many of the articles are also posted.

The first article by Andrew Sharpe and Leila Gharani from the Centre for the Study of Living Standards examines the factors behind slow productivity growth in Canada in the second half of the 1990s, in marked contrast to the acceleration of productivity in the United States, and discusses the prospects for trend productivity in Canada over the next decade or two. It concludes that the balance of evidence now favours an acceleration of trend labour productivity growth to the 2-2.5 per cent per year range, as the factors accounting for the U.S. productivity boom finally spill over into Canada.

The second article, which is closely related to the first article, is also by Andrew Sharpe of the Centre for the Study of Living Standards. It points out that there now appears to be a renaissance in productivity growth in the U.S. service sector, with output per worker growing five times faster in the 1995-98 period than in the 1981-95 period. This development appears to reflect the impact of the massive investments in information technologies, which finally now seem to be producing large productivity gains in a wide range of service industries.

The third article by Raynauld Letourneau and Martine Lajoie of Industry Canada provides a detailed regional analysis of levels of living standards, measured as output per capita, and productivity (output per worker) for the 1995-97 period, the most recent data currently available. They find that all regions and provinces of Canada trail the U.S. average in both living standards and productivity and that the productivity gap is the main factor behind the living standard gap.

The fourth article by Edward N. Wolff of New York University examines trends in convergence in OECD countries toward U.S. productivity levels during the postwar period and finds strong evidence for this phenomenon up to 1990, with rapid growth in investment, education, and R&D in OECD countries accounting for the catch-up. The process of convergence seems to have ended in the 1990s, reflecting slower growth in OECD countries, a diminishing of the forces behind the convergence process given the narrowing of the productivity gap with the United States, and perhaps most important, the acceleration of productivity in the United States.

The fifth article by Jeffrey I. Bernstein of Carleton University and the NBER discusses the use of total factor productivity for price setting in regulated industries. He argues that the long-term, industry-wide productivity experience that is not subject to strategic manipulation by regulated firms should be the productivity growth rate used to develop the appropriate offset factor for price cap regulation.

The final three articles are a symposium of total factor productivity. In the first of the articles, Timothy C. Sargent and Edgard R. Rodriquez of Finance Canada discuss the issue of the choice between labour and total factor productivity. They conclude that both measures have uses. For periods of less than a decade, labour productivity is the preferred measure, but for longer periods total factor productivity is superior. When capital stock estimates are of poor quality, it is better to use labour productivity.

In the second article in the symposium, Richard G. Lipsey of Simon Fraser University and Kenneth Carlaw of the University of Canterbury in New Zealand provide a trenchant critique of the concept of total factor productivity. They conclude that "the degree of confusion surrounding TFP, particularly the assumption that low TFP numbers imply a low degree of technological dynamism, would seem to us to justify dropping the measure completely from all discussions of long term economic growth".

In the third article in the symposium, Erwin Diewert of the University of British Columbia provides a comprehensive discussion of what is needed to develop reliable measures of total factor productivity in terms of output and the different classes of inputs. He concludes that the current system of industry statistics in all advanced countries has not kept up with the evolution of the economy from primary and manufacturing production to the production of services and that if we want accurate measures of total factor productivity at the industry level, statistical agencies must be given additional resources.

The New Economy and Trend Productivity Growth in Canada

Andrew Sharpe*
Leila Gharani
Centre for the Study of Living Standards

The United States has experienced a very strong productivity revival since 1995. Output per hour in the business sector advanced at a 2.7 per cent average annual rate in the 1995-99 period, up from 1.2 per cent in the 1989-95 period. In contrast, Statistics Canada data show that growth in output per hour in the business sector in Canada actually decelerated in the second half of the 1990s, falling to 1.0 per cent per year in the 1995-99 period from 1.3 per cent in 1989-95. Has the new economy, defined as higher trend productivity growth, pasted Canada by?

This article examines the Canadian productivity experience in the 1990s and asks the question why trend productivity has not picked up. It then discusses the likelihood of Canada entering the "new economy" of higher trend productivity.

Productivity Trends in Canada

The Canadian economy certainly picked up steam in the second half of the 1990s. Real GDP accelerated 1.9 percentage points to a 3.4 per cent average annual growth rate between 1995 and 1999 from only a 1.5 per cent rate in the 1989-95 period. But the increase in output has almost entirely been accounted for by increased employment, not productivity gains. Employment growth accelerated 1.7 percentage points between periods, while productivity growth (GDP per worker) was up only 0.3 points (GDP per hour growth, like business sector per hour growth, fell). This development in itself is not necessarily bad (some might say it is even positive) as employment growth is highly desirable as it reduces the unemployment rate and labour market slack and has marvelous effects of governments' fiscal position. But it does raise the question of why productivity growth was so poor, particularly in contrast to the U.S. experience.

Factors Behind the U.S. Productivity Revival

To understand the failure of productivity to revive in Canada in the second half of the 1990s, and the prospects for productivity in the first decade of the new millenium, it is crucial to understand the U.S. productivity experience. The factors that account for the U.S. productivity revival were largely not present in Canada. These factors are outlined below.

- The United States has a very large high tech sector and this information technology (IT)-producing sector (largely concentrated in manufacturing) has enjoyed extremely rapid productivity growth.

- Machinery and equipment investment, particularly in IT, has been very strong since the early 1990s and since 1995 appears to have given service sector productivity a major boost (see the following article in the *Monitor* on the renaissance of U.S. service sector productivity).
- The unemployment rate has fallen to very low levels in recent years, giving employers an additional incentive to substitute capital for labour to the benefit of labour productivity and resulting in the full utilization of human resources, a recipe for strong productivity growth.
- The Federal Reserve Board policy of probing the limits of the non-accelerating inflation rate of unemployment (NAIRU) has resulted in robust demand growth and allowed the increase in potential output arising from the IT revolution to manifest itself as actual output. In the short-to-medium term tradeoff between low inflation and economic growth, the Fed has given significant weight to robust growth.
- Changes by the Bureau of Economic Analysis in the methodology used to construct the national accounts, especially the decision to treat software as an investment good, have led to significant upward revisions in real output growth (e.g. 0.5 percentage points per year in the 1995-98 period).

Factors Behind Canadian Productivity Stagnation

The five factors enumerated above as contributing to robust productivity growth in the United States in the second half of the 1990s did not obtain in Canada.

- Canada has a much smaller high tech sector than the United States and has paid the price with much weaker productivity growth in manufacturing.
- Canada's machinery and equipment investment was much weaker than investment in the United States in the first half of the 1990s (only 2.1 per cent per year in real terms) and this lackluster performance failed to produce, with a lag, a revival of service sector productivity in the second half of the 1990s.
- The unemployment rate in Canada has remained much higher, giving employers less incentive to substitute capital for labour and thwarting the positive productivity effects of full utilization of resources.
- The Bank of Canada has been less aggressive than the Federal Reserve Board in testing the limits of the NAIRU, with the result that demand growth has been somewhat weaker in Canada since 1995. The economic growth objective appears to receive lower weight relative to the low inflation objective in the conduct of monetary policy in Canada than in the United States.
- Statistics Canada has not adopted the methodology of treating software as part of investment, so measured productivity growth has not had a boost from statistical revision.

Canadian Productivity Prospects for the Next Decade

A strong case can be made that the new economy characterized by strong trend productivity growth is finally arriving in Canada, occasioned by a reversal of most of the factors that have impeded productivity growth in the second half of the 1990s.

- The high technology sector is now enjoying very rapid growth and this growth is expected to continue for the foreseeable future.

Indeed, high tech industries are fueling rapid growth in many urban centres such as Ottawa and Kitchner-Waterloo.

- Real machinery and equipment investment in Canada skyrocketed in the second half of the 1990s, advancing at a 14.3 per cent average annual rate. The productivity payoff from this investment will be felt in coming years throughout the economy.
- Canada's unemployment rate in the first half of 2000 has fallen below 7 per cent and could go significantly lower if the economic growth remains robust. This will foster more productive use of labour.
- Bank of Canada policy remains the wild card as it is not clear it will adopt the successful U.S. policy of aggressively pushing down the unemployment rate until inflation accelerates. A recession or significant growth slowdown precipitated by the Bank of Canada's preoccupation with the attainment of price stability could nip incipient productivity gains in the bud.
- Statistics Canada plans to follow the U.S. lead in the treatment of software as an investment, which would increase both past and future measured productivity growth.

Is A New Productivity Era Upon Us?

No one can be certain about future productivity trends. Rosy scenarios of productivity growth projected by many economists for the 1970s, 1980s, and 1990s have gone down in smoke. It is very possible that the future productivity path will be a continuation of the anemic productivity growth experienced over the last 25 years. Yet economic history moves in cycles, often associated with waves of technological innovation. The wave of innovation associated with information technology finally has increased productivity growth in the United States, although there is of course no assurance that this pace (2.5-3.0 per cent per year for business sector labour productivity) will continue for any length of time. The productivity-enhancing effects of the current technological wave may be less potent that earlier technological waves.

In our view, the balance of evidence now suggests that Canada will see a significant pick-up in productivity growth (to at least the 2-2.5 per cent range for business sector output per hour) over the next decade if not for two decades. Very preliminary evidence of such a revival of productivity growth in Canada is found in the productivity figures for the first half of 2000, showing GDP per worker rising at a 2.2 per cent annual rate.

But the dominant reason for adopting an optimistic view on trend productivity growth is the recent U.S. experience. What happens in the United States spills over to Canada, although often with a lag. Our productivity growth in the past has tracked or even exceeded U.S. growth as the same forces are at play in the two countries. The factors that have produced an acceleration of measured productivity growth in the United States since 1995 (innovation in IT, strong machinery and equipment investment, low unemployment, changes in statistical methodologies) are now beginning to operate in Canada.

Notes

* This article summarizes sections from a much longer paper entitled "The New Economy and Trend Productivity." This paper is available at www.csls.ca under the International Productivity Monitor. Email: csls@csls.ca.

The Productivity Renaissance in the U.S. Service Sector

Andrew Sharpe*
Centre for the Study of Living Standards

Since 1995 productivity growth has accelerated in the United States. According to the most widely used aggregate productivity measure, the series on non-farm business sector output per hour produced by the Bureau of Labor Statistics (BLS), productivity advanced at a 2.9 per cent average annual rate from 1995 to 1999 and at a 4.3 per cent annual rate in the first half of 2000.

Unfortunately, the BLS does not provide data on productivity growth by industry so it is not possible to identify from this source which sectors are fueling the productivity revival. Such data can, however, be constructed from the real output and labor input series compiled by the Bureau of Economic Analysis (BEA) and publicly available on the BEA website (although this data series has only recently been updated to 1998). This article uses the BEA data to provide estimates of labour productivity growth on an industry basis in the post-1995 period, the period when the new economy appears to have flowered.

This article's key message is that after many decades of stagnant growth, there now appears to be a renaissance in service sector productivity. Real (1996$) value added per person employed in the broadly defined service sector (transportation and public utilities, wholesale trade, retail trade, finance, insurance and real estate, services, and government) advanced at a 2.4 per cent average annual pace in the 1995-98 period, up nearly five-fold from the 0.5 per cent rate of the 1981-89 and 1989-95 periods (see Table 1). It now appears that the service sector productivity drought is over, at least for the second half of the 1990s, and possibly into the future.

This development may come as a surprise. Economists have long deplored lagging productivity growth in the service sector, and have advanced numerous explanations, including measurement problems, and an intrinsic lack of dynamism in many service industries.[1]

Service Sector Productivity Growth by Industry

The productivity renaissance in the service sector is broadly based, with four of the six basic service sector industries showing at least a one percentage point increase in labour productivity growth between the 1989-95 and 1995-98 periods. The growth rate of output per worker in wholesale trade accelerated 6.4 points, in retail trade 4.8 points, in finance, insurance and real estate 1.3 points and in services (personal, busi-

Table 1: Value Added per Worker Employed, U.S. Estimates of GDP per employed worker in constant 1996 dollars

Industry Title	1981-89	1989-95	1995-98	(1995-98)-(1989-95)
	% Average compound growth rates			
Total Economy	**1.38**	**1.11**	**1.85**	**0.74**
Goods Sector	**3.18**	**2.20**	**2.79**	**0.59**
Agriculture, forestry, and fishing	3.60	0.01	5.53	5.52
Mining	8.02	4.71	3.23	-1.48
Construction	0.64	-0.13	0.03	0.16
Manufacturing	3.74	3.14	3.50	0.36
Service Sector	**0.48**	**0.54**	**2.41**	**1.88**
Transportation and public utilities	2.21	2.59	2.03	-0.56
Wholesale trade	3.37	2.85	9.20	6.35
Retail trade	1.61	0.91	5.74	4.83
Finance, insurance, and real estate	-0.12	1.64	2.89	1.26
Services	-0.16	-0.79	0.19	0.99
Government	0.33	0.28	0.58	0.30

Source: Data for GDP and employment are obtained from the Bureau of Economic Analysis, 2000. Release date: June 2000. http://www.bea.doc.gov/bea/uguide.htm#_1_14

Note: Because of the use of non-additive chain indices for real output, industries totals do not sum to the total economy total. This explains why the total economy productivity growth rate in the 1995-98 period is less than both the goods sector and service sector productivity growth rates.

ness and other services) 1.0 points. Even government enjoyed improved productivity growth, up 0.3 points, although the estimates of real output for government are not appropriate for productivity calculations as they are largely estimated on the basis of inputs. The only service sector industry that did not enjoy faster productivity growth after 1995 was transportation and public utilities, experiencing a 0.6 point fall-off.

The services industry accounts for 30 per cent of total employment. A more disaggregated analysis of this industry shows that the productivity growth acceleration was particularly strong in miscellaneous professional services (up 4.1 points), personal services (2.0 points), amusement and recreation services (1.6 points), and motion pictures (1.5 points).

The service sector productivity renaissance refers to the post-1995 acceleration of labour productivity growth, not the overall pace of this growth in the service sector. Indeed, productivity growth in the goods sector continues to outperform that in the service sector at 2.8 per cent versus 2.4 per cent per year in the 1995-98 period. But goods sector productivity did not pick up after 1995 from its robust pace in the 1989-95 period, due to strong productivity growth in manufacturing and mining during the first half of the 1990s.

Not all service sector industries have experienced better productivity growth since 1995. Indeed, between the 1989-95 and 1995-98 periods, productivity growth actually fell 3.6 points in hotels and other lodging places, 2.8 points in social services and membership organizations, 1.3 points in educational services, and 1.0 points in auto repair, services and parking. Moreover, productivity growth rates continue to be negative in the 1995-98 period in the four industries mentioned above as well as for insurance carriers; insurance agents, brokers and service; miscellaneous repair services; motion pictures; health services; and legal services. This suggests they may be still room for significant productivity improvement in a number of service industries.

What Explains the Renaissance?

A full explanation of what appears to be a productivity renaissance in the U.S. service sector is well beyond the scope of this article. One explanation is that better output measures are now capturing real output gains that eluded government statisticians in the past. And output gains translate on a one-for-one basis into productivity gains.

A second explanation is that the massive investment in information technology (IT) made in the service sector throughout 1990s is finally paying off in terms of increasing output. The Solow productivity paradox seems to be resolved as we are now seeing the impact of computers in the productivity statistics. The lags between IT investment and productivity appear to have ended as firms and workers have now learned to use these new technologies in an effective manner. The large IT investment in wholesale and retail trade and the very strong increases in productivity in these two industries support the IT story.

Of course, the two explanations outlined above are not mutually exclusive and both have probably contributed to the improvement in service sector productivity growth.

Implications for the New Economy Debate

It is now widely recognized that trend labour productivity growth has experienced a significant upward shift since 1995, and that this shift is more than just a cyclical development as it reflects the impact of the massive IT investment in the 1990s. Until recently, it was believed that most of the productivity gains were taking place in the IT-producing sector and that the productivity-enhancing impact of IT was not spreading to the IT-using sectors. With the renaissance of productivity growth in these IT-using service industries such as wholesale and retail trade, it now appears that the acceleration of productivity growth is broadly based.

The jury is still out on whether the silicon chip represents a general purpose technology (GPT) of the importance of past GPTs such as the steam engine, the internal combustion engine, and electricity. It is unclear whether productivity gains of the second half of the 1990s represent a transitory phenomenon or whether they will have some legs and continue for two or more decades. Certainly, such long-term productivity gains are within the realm of possibility and in my view the broadening of productivity gains to the IT-using service sector since 1995 augurs well for this view.

Notes

* This article is drawn from a longer article with the same title available under the International Productivity Monitor at www.csls.ca. Email: csls@csls.ca.

1 See, for example, the April 1999 special issue of the *Canadian Journal of Economics* edited by Erwin Diewert, Alice Nakamura, and Andrew Sharpe on service sector productivity and the productivity paradox. The papers are posted at www.csls.ca under publications.

A Regional Perspective on the Canada-US Standard of Living Comparison

Raynald Létourneau
and Martine Lajoie
Micro-Economic Analysis
Micro-Economic Policy Analysis Branch
Industry Canada

Introduction

This article presents a comparison of standards of living between Canadian provinces and US States. Most comparisons with the United States focus on the national perspective while provincial analyses are essentially restricted to the domestic context. This study extends the scope of the exercise to the regional level since the relative performance of the provinces varies significantly and, therefore, the challenges raised by the greater integration of the North-American market are also likely to differ. The comparison focuses on standards of living with a special emphasis on labour productivity.

The article is divided as follows. First, we present our framework of analysis and discuss issues related to the comparison of productivity and standard of living at regional levels between the two countries. We then move to the discussion of standard of living and productivity. Each of these sections presents separately an analysis of US states, Canadian provinces and a comparison of both provinces and states. The paper concludes with a brief review of the main results.

1. Framework of Analysis and Empirical Issues

1.a Framework of Analysis

Standard of living is best measured through real GDP per capita as it encompasses all earnings accrued to residents of a country. The standard of living can be expressed as:

(1) GDP/POP = GDP/E * E/POP

where

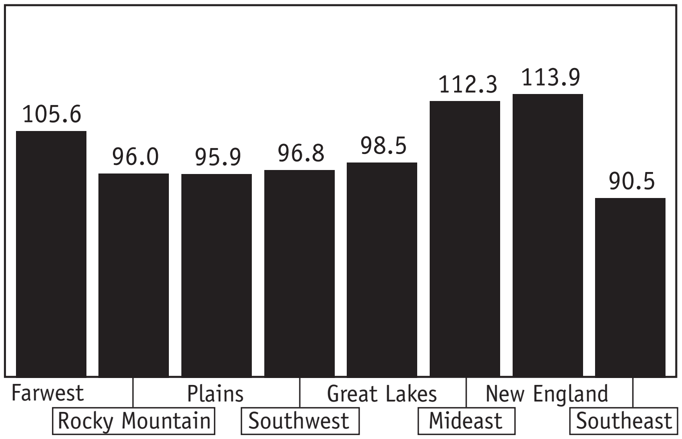
GDP/POP: Real GDP per Capita or Standard of Living

GDP/E: Labour Productivity (Real GDP per Worker)

E/POP: Employment Rate or the proportion of the population that is working

The framework of analysis is relatively simple and states that real income per capita is determined by the productivity of workers as well as the proportion of the population at work. A high level of productivity and a large proportion of the population at work will result in a high standard of living.

Figure 1:
Standard of Living, US regions, 1995-1997
(US=100)



Source: Bureau of Economic Analysis, Bureau of Labor Statistics and Statistics Canada

1.b Empirical issues

Comparing standards of living and labour productivity for the 10 provinces and 50 states raises serious challenges in terms of data requirements. These measures are based on three variables: real gross domestic product, population and employment. Standard of living is measured through real GDP per capita and labour productivity through real GDP per worker.

In order to make valid comparisons with US data, real output is measured as GDP at market prices, which, because of data constraints in Canada related to the market price measure, restricts the comparison exercise to the 1992-1997 period. US data have been obtained from three different sources. Gross State Product (GSP) in constant 1992 dollars were taken from the Bureau of Economic Analysis while population and employment data were obtained from the Bureau of the Census and the Bureau of Labor Statistics respectively. Canadian data are derived from Statistics Canada's Provincial Economic Accounts, Labour Force Survey and Population estimates. US real GDP is expressed in Canadian dollar terms by using the Purchasing Power Parity value of the exchange rate for 1992 ($1.23 according to Statistics Canada).

US real GSP is available on a 1992-chained dollars basis and components are not strictly additive, especially for years far away from the base period.[1] For the period under consideration (1992-1997) which is close to the 1992 base year, however, GSP estimates are nearly additive. This allows us to calculate the relative performance of individual states in comparison to the US national average. Finally, comparisons, at both national and international levels, are calculated using an average of the three most recent years (1995 to 1997) in order to obtain more robust estimates.

Data constraints at the regional/state levels and the presence of significant cost-of-living differences across US states[2] and provinces impose some limitations on the interpretation of the results. Since US nominal GSP is deflated using producers prices rather than some expenditures-based deflators and since differences between production and consumption measures can be large at the state level, productivity comparisons are less affected by these considerations than standard of living.

2. Standards of Living

US Regions

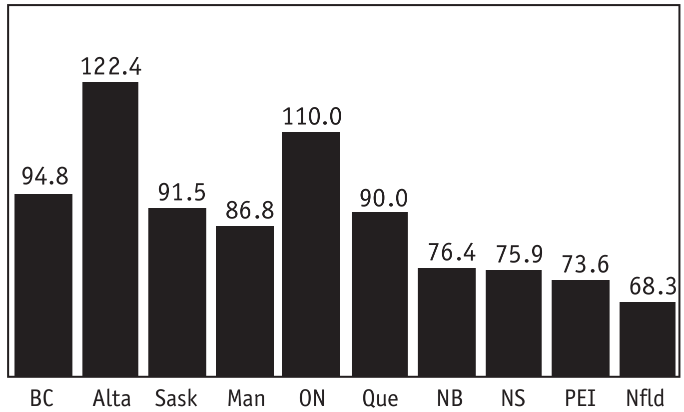
Real GDP per capita varies significantly across states: real GDP per capita in Delaware, the highest income state, is nearly twice that of Mississippi, the lowest income state. Figure 1 shows the relative level of regional real GDP per capita (US national average =100) over the period 1995-1997 using the US Bureau of Economic Analysis' classification. Three out of eight regions recorded above-average standards of living: New England, Mideast and Farwest, all regions that have extensive trade links with Canada. The Great Lakes, Plains and Rocky Mountain regions' standards of living were slightly below-average, while it was as much as 10% below-average in the Southeast.

Real GDP per capita also varies significantly within US regions. The relative standing of a region is determined by its concentration of high and low income states. For example, in New England, real GDP per capita ranges from 20% below the national average in Maine to 32% above in Connecticut. High income regions include a greater number of states with standards of living well above the US average. Connecticut, Massachusetts and New Hampshire account for New England's strong performance, more than offsetting the weakness of Maine and Vermont. The Mideast's standard of living is supported by Delaware, New York and New Jersey while California is raising the Farwest's standard of living, followed by Nevada and Hawaii.[3]

Regions with average standards of living below that of the US still conceal a few high income states. In the Great Lakes, the strong standing of Illinois, with real GDP per capita nearly 10% above the national average, is offset by under performers such as Michigan and Indiana while Minnesota is driving up the Plains' standard of living. In Rocky Mountain states, the weak performance of Montana, Idaho and Utah is offsetting that of Wyoming and Colorado.

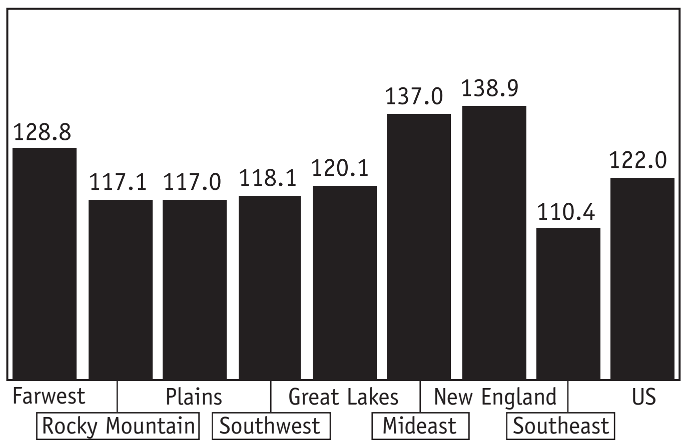
Texas is the only Southwest state to post above-average real GDP per capita. Mississippi, West Virginia, Arkansas, Alabama and South Carolina are behind the Southeast's low standard of living, largely offsetting the positive effect of Virginia and Georgia.

Figure 2:
Standard of Living, Canadian provinces, 1995-1997
(Canada = 100)



Source: Statistics Canada

Figure 3:
Standard of Living, 1995-1997 average
(Canada = 100)



Source: Bureau of Economic Analysis, Bureau of Labor Statistics and Statistics Canada

Canadian Provinces

Standards of living vary less across provinces than among US states. Although this reflects different economic profiles between provinces and states, it is also related to the presence of federal transfers to the provinces, such as the equalization program, which tend to reduce regional disparities.

Standards of living are generally highest in provinces west of Quebec. These provinces tend to be more productive and also have a higher proportion of their population at work. Alberta ranks first with real GDP per capita more than 20% above the national average, followed by Ontario. The standard of living is next highest in British Columbia, Saskatchewan and Quebec. It is lowest in Newfoundland, at about 30% below the national average, behind Prince Edward Island, Nova Scotia and New Brunswick.

Table 1 Rankings of Canadian Provinces and US States Average Standard of Living, 1995-1997, Canada=100

1	Delaware	169.8	31	Michigan	113.4
2	Alaska	167.9	32	New Mexico	113.0
3	Connecticut	161.1	33	Tennessee	112.1
4	Wyoming	154.6	34	Rhode Island	111.9
5	New Jersey	146.9	35	Kansas	110.8
6	Massachusetts	144.5	36	South Dakota	110.6
7	New York	144.0	**37**	*Ontario*	110.0
8	Nevada	137.5	38	Arizona	108.3
9	Illinois	133.8	39	Vermont	106.9
10	Hawaii	132.3	40	Kentucky	106.2
11	New Hampshire	130.9	41	Utah	105.8
12	California	129.6	42	Florida	104.7
13	Colorado	129.4	43	Idaho	102.6
14	Minnesota	127.4	44	South Carolina	101.8
15	Virginia	127.1	45	North Dakota	101.2
16	Texas	125.1	46	Alabama	97.4
17	Georgia	123.9	47	Maine	97.3
18	*Alberta*	122.4	48	Arkansas	95.2
19	North Carolina	122.3	**49**	*British Columbia*	94.8
20	Washington	121.7	50	Oklahoma	93.8
21	Oregon	120.4	**51**	*Saskatchewan*	91.5
22	Nebraska	120.2	**52**	*Quebec*	90.0
23	Maryland	119.7	53	Montana	88.8
24	Ohio	116.8	54	West Virginia	88.5
25	Wisconsin	115.6	55	Mississippi	88.0
26	Iowa	115.0	**56**	*Manitoba*	86.8
27	Louisiana	114.7	**57**	*New Brunswick*	76.4
28	Pennsylvania	114.4	**58**	*Nova Scotia*	75.9
29	Missouri	113.9	**59**	*Prince Edward Island*	73.6
30	Indiana	113.7	**60**	*Newfoundland*	68.3

US-Canada Standards of Living Comparison

When US real GDP per capita is expressed in Canadian currency, using the 1992 PPP, the US standard of living is, on average, 22% higher than that of Canada.[4] This aggregate number conceals, however, a few important facts. First, Figure 3 shows that all US regions post standards of living well above the Canadian average. Second, the gap with respect to the highest income region, New England, reaches up to 40%. Third, the lowest US region, the Southeast, has a standard of living still 10% above the Canadian average.

Only seven states (Table 1) recorded standards of living below the Canadian average. Except for Maine, they all come from low-income regions, particularly the Southern States. The standard of living is more than 25% higher than the Canadian average in a third of the states — and is more than 50% higher in Delaware, Alaska, Connecticut and Wyoming.

Relative to their US counterparts, Canadian provinces tend to rank at the lower end of the spectrum. Alberta records the best performance, in 18th place, followed by Ontario (37th), British Columbia, 49th, Saskatchewan, 51th and Quebec, 52th. All other provinces rank below Mississippi, the state with the lowest standard of living in the US.

Table 2 Rankings of Canadian Provinces and US States Productivity, 1995-1997, Canada=100

1	Alaska	163.8	31	Arizona	106.4
2	Delaware	156.2	**32**	*Ontario*	106.2
3	Connecticut	149.5	33	Florida	105.4
4	New York	148.9	34	Nebraska	103.9
5	Wyoming	141.6	35	Indiana	103.2
6	New Jersey	139.7	36	Missouri	102.9
7	Massachusetts	133.0	37	Kansas	102.3
8	Hawaii	131.4	38	Utah	101.2
9	California	131.2	39	West Virginia	100.7
10	Nevada	127.2	40	South Dakota	99.6
11	Illinois	126.1	41	Iowa	99.4
12	Louisiana	123.8	42	Wisconsin	98.5
13	Texas	120.1	43	South Carolina	98.4
14	New Mexico	119.3	44	Alabama	96.8
15	Virginia	119.3	45	Idaho	96.3
16	Georgia	117.3	46	Arkansas	95.0
17	New Hampshire	114.9	47	Oklahoma	94.8
18	North Carolina	114.7	**48**	*British Columbia*	94.2
19	Washington	114.0	**49**	*Quebec*	93.8
20	Pennsylvania	113.8	50	Vermont	93.5
21	Colorado	112.4	51	Mississippi	93.1
22	Ohio	112.0	**52**	*Saskatchewan*	92.6
23	*Alberta*	111.4	**53**	*Newfoundland*	91.7
24	Oregon	110.6	54	North Dakota	90.3
25	Rhode Island	109.7	55	Maine	89.7
26	Michigan	109.4	**56**	*Manitoba*	86.0
27	Minnesota	108.8	57	Montana	85.1
28	Kentucky	107.1	**58**	*New Brunswick*	84.5
29	Tennessee	106.7	**59**	*Nova Scotia*	84.4
30	Maryland	106.6	**60**	*Prince Edward Island*	77.1

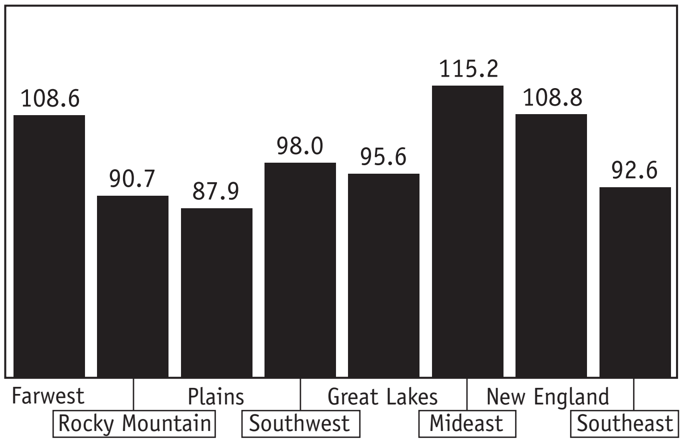
3. Productivity

US Regions

As for standards of living, there are significant differences in productivity[5] levels across US states. Figure 4 shows that three out of eight regions record above-average productivity: Mideast, New England and Farwest. Not surprisingly, these regions also posted the highest standards of living, indicating that productivity is the main driving force behind the standard of living.[6] In other regions, productivity is below the US average — it falls more than 10% below-average in the Plains region.

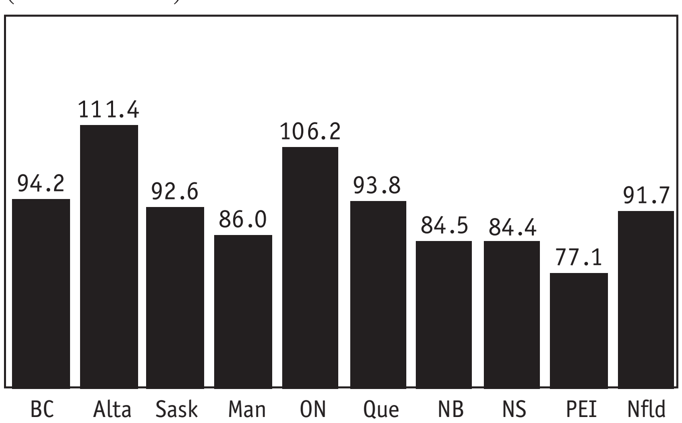
The high-productivity regions, as in the case of standard of living, are comprised of highly productive states which more than offset the weaker performance of a few low-productivity states. The most productive US region — the Mideast — records a productivity level 15% above the national average, supported by Delaware, New York and New Jersey. New England's high productivity standing lies on the strong performances of Connecticut and Massachusetts. High productivity levels are widespread across Farwest states, with only Oregon and Washington posting productivity below the region's average.

Figure 4:
Productivity*, US regions, 1995-1997
(US=100)



* Labour Productivity
Source: Bureau of Economic Analysis, Bureau of Labor Statistics

Figure 5:
Productivity*, Canadian provinces, 1995-1997
(Canada = 100)



* Labour Productivity
Source: Statistics Canada

Regions with productivity levels below the US average have a greater concentration of low-productivity states although they all conceal a few highly productive ones. Southwest's productivity is supported by Texas and New Mexico. In the Southeast, Louisiana and Virginia are raising the region's average, partly offsetting the impact of low productivity states such as Mississippi, Arkansas and Alabama. In the Great Lakes, Illinois is holding up the regional average while Wyoming and Colorado are the driving force of the Rocky Mountain region. Low productivity is widespread across Plains states.

Canadian Provinces

Productivity rankings in Canada are also very similar to those for standard of living, highlighting the importance of the level of productivity as a fundamental determinant of the standard of living. Moreover, the gap among provinces is somewhat smaller for productivity compared to standard of living, reflecting that high-productive provinces tend also to be advantaged by a higher employment-population ratio. Newfoundland perhaps is the only exception to this general trend as the deterioration of its ranking from productivity (sixth) to standard of living (tenth) reflects a very low employment-population ratio.

Alberta records the best productivity performance, followed by Ontario. Productivity is next highest in British Columbia, Quebec and Saskatchewan. Manitoba's productivity is the weakest among western provinces, at more than 10% below the Canadian average. Atlantic provinces post productivity levels below the national average: the productivity level in PEI is more than 20% below the Canadian average.

US-Canada Productivity Comparison

Overall, US states are about 18% more productive than their Canadian counterparts (Figure 6), slightly below the 22% gap in the case of standard of living.[7] All US regions recorded productivity levels above the Canadian average over the 1995-1997 period with the gap ranging from a low of 3% in the Plains to a high approaching 40% in the Mideast. The gap is particularly high compared to Mideast, New England and Farwest, all regions that have strong traditional links with Canada.

Table 2 shows the relative ranking of Canadian provinces and US states. Only thirteen states, located largely in Southern and Rocky Mountain areas, registered productivity below the Canadian average in 1995-1997. At the exception of Alberta and Ontario, which record again the best performance within Canada, (23th and 32th place respectively), other provinces ranked at the lower end of the spectrum. This table also emphasizes the large differences in productivity levels among US states. For example, productivity in Alaska and Delaware is nearly double that of Montana, the least productive state.

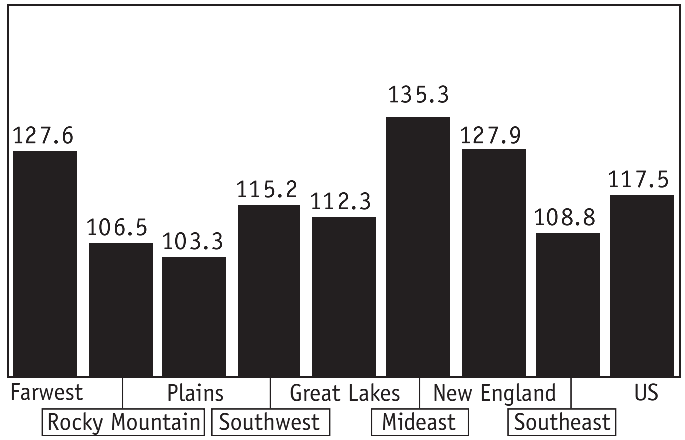
We tested the ranking of productivity with that of standard of living to assess the importance of productivity in determining standard of living in a North-American context (Table 1 versus Table 2). Not surprisingly, we found the Spearman rank correlation coefficient to be high (0.92) and significant, indicating a very strong relationship between the two variables for all jurisdictions in North America.

Conclusion

This comparative exercise has revealed that Canada's regions all face, albeit to varying degrees, income and productivity gaps vis-à-vis the US. Standards of living of Canadian provinces are well behind those of US states. In fact, the best Canadian performer, Alberta, ranks 18th among the 60 states and provinces, followed by Ontario in 37th place. Most Canadian provinces are concentrated at the bottom of the list.

Our results also show that productivity is the predominant factor explaining income gaps among provinces and states, a conclusion that supports findings at the national level. Differences in employment rates play a limited role in explaining these gaps and, therefore, do not influence final rankings. A similar picture to that of standard of living emerges from the productivity comparison: except for Alberta and Ontario, provinces are ranked at the bottom end.

Figure 6:
Productivity*, 1995-1997 average
(Canada=100)



* Labour Productivity
Source: Bureau of Economic Analysis, Bureau of Labor Statistics and Statistics Canada

Notes

* This article is based on "A Regional Perspective on the Canada-US Standard of Living," Occasional Paper Number 22, Industry Canada, February 2000. The authors wish to thank Richard G. Harris, Someshawr Rao and Shane Williamson for helpful comments. The views expressed in this paper reflect those of the authors and should not be attributed to Industry Canada. Martine Lajoie has joined the Privy Council Office. Email address: letourneau.raynald @ic.gc.ca

1 For more details on the estimation of chained-1992 dollars GSP estimates, see Friedenberg and Beemiller (1997).

2 For example, see Engel C. and J. H. Rogers, Violating the Law of One Price: Should We Make a Federal Case Out of It, NBER, Working Paper 7242, July 1999. Nevertheless, the authors found that law of one price deviations are not as important for locations within the US as compared to deviations among countries.

3 Although Alaska's standard of living is the highest among Farwest states, its contribution to the region's average standard of living is marginal because it only accounts for about 2% of the region's output. California, however, accounts for more than 70% of the Farwest output, contributing largely to the region's high standard of living.

4 Canada-US comparisons are based on the 1992 PPP value of US$1.00= $1.23 Canadian calculated by Statistics Canada.

5 Productivity refers here to real output per employee.

6 The variability of the employment ratio (E/POP) among regions is relatively small so productivity is the key determinant of each region's standard of living. However, higher-than-average standard of living in some states such as, Minnesota, Colorado, North Carolina, New Hampshire and Georgia is also the result of a greater-than-average share of population at work. For more details see Appendix C.

7 As in the case of standard of living, US real GDP per employee is expressed in Canadian dollars using the 1992 PPP.

References

Beemiller, R.M. and G.K. Downey, "Gross State Product by Industry, 1977-96" in *Survey of Current Business*, June 1998, p.15-37

Friedenberg, H.L. and R.M. Beemiller, "Comprehensive Revision of Gross State Product by Industry, 1977-94" in *Survey of Current Business*, June 1997, p.15-41

Lum, S.K.S. and B.C. Moyer, "Gross Product by Industry" in *Survey of Current Business*, November 1998, p.20-40

Seskin, E.P. and R.P.Parker, "A Guide to the NIPA'S", in *Survey of Current Business*, March 1998, p.26-68

Statistics Canada, National Income and Expenditures Accounts, volume three, publication 13-549

Statistics Canada, Gross Domestic Product by Industry, publication 15-001

Statistics Canada, Provincial Gross Domestic Product by Industry, 1984-1997, publication 15-203

Productivity Convergence among OECD Countries: The Postwar Experience

Edward N. Wolff,
Department of Economics,
New York University

1. Introduction

The "catch-up hypothesis" rests on the view that the historical course of technological progress operates through a mechanism which enables countries whose standard of productivity performance is reasonably close to that of the leader(s) to catch up. Through the constant transfer of new technology, leader countries and those most closely in their van learn the latest productive techniques from one another. Virtually by definition, the follower countries have more to learn from the leaders than the leaders have to learn from them (the so-called "advantages of backwardness"). This mechanism has two implications: First, it means that those countries that lag somewhat behind the leaders can be expected systematically to move toward the level of achievement of the leaders. Second, the mechanism undermines itself automatically as follower countries gradually eliminate the difference between their own performance and that of the countries that were ahead of them — that is, the very fact of convergence means that the differential learning opportunities that are the source of these advantages of (slight) backwardness will exhaust themselves. On an analytical level, this hypothesis would imply faster productivity growth for the (initially) more backward economies relative to the more advanced ones, but the gap in productivity performance would gradually diminish over time as convergence was achieved.

However, being backward does not itself guarantee that a nation will catch up. Other factors must be present, such as strong investment, an educated and well trained work force, research and development activity, developed trading relations with advanced countries, a receptive political structure, low population growth, and the like. Indeed, Abramovitz (1986 and 1994) has summarized this group of characteristics under the rubric of *social capability*. This process has also been referred to as "conditional convergence" by Mankiw, Romer, and Weil (1992).

A. Investment.

It seems generally agreed that there are two prime ingredients in the growth of labor productivity: technological innovation, and the accumulation of capital through saving (and the subsequent investment of those savings). Innovation and the international transfer of its products play a prime role in the converging productivity levels of a number of relatively successful industrialized economies. But even if technological innovation is the more important factor in the scenario (which is by no means certain), substantial capital accumulation very likely would have been required to put the inventions into practice and to effect their widespread employment. Moreover, if saving and investment play a primary role of their own, it becomes all the more important to explore the nature of that role, recognizing that because of unavoidable interactions between the rates of innovation and investment, any attempt to separate the two may prove to be artificial, if not ultimately unworkable (see Abramovitz and David, 1973, for an extremely illuminating analysis of the data and the theoretical issues, as well as some references to other discussions by economic historians).

B. Education

Another critical factor appears to be education. It will be seen that the statistical evidence is consistent with the hypothesis that the quantity of education provided by an economy to its inhabitants is one of the major influences determining whether productivity in that society is growing rapidly enough to narrow the gap with productivity in the more prosperous economies. This is important for policy because it suggests that a country can do a great deal to improve its performance in the convergence arena by increasing the resources it devotes to education. It is at the secondary school level, and to an even greater degree in higher education, that large differences persist.

C. Science and Technology

The role of science and technology, which are in the catch-up process, should also be considered. There is a vast literature which supports the view that research and development (R&D) is positively associated with productivity growth. This has been demonstrated on the aggregate (national) level, the industry level, and the firm level (see, for example, Griliches, 1979, for a review of the literature).

D. Foreign Trade

Another factor that may be directly relevant to the international transfer of technology is the extent of international trade and the pattern of trade. It is generally argued that trade is a mechanism for the transmission of information concerning new technologies and products. For example, imports of computers may revolutionize the production technology of importing industries. Also, the exposure to new products may induce local competitors to imitate. The argument on the export side is weaker. Competition in export markets may lead to the exposure to new foreign products; it may also lead to more rapid developments of new technology in industries competing in export markets.

2. Comparative Statistics among OECD Countries

I begin with measures of real GDP (1985 dollar equivalents) per worker (RGDPW). Computations for 1950 to 1990 are based on the variable RGDPW from Penn World Table (PWT) Mark 5.6 (see Summers and Heston, 1991, for a description of the database). Computations for 1996 are based on the variables GDPD (Gross Value Added in US Dollars) and ET (Total Employment) from the OECD ISDB (International Intersectoral) Database for 14 OECD countries. The PWT figures for 1990 are used as benchmarks, and the figures are updated to 1996 on the basis of the growth rate of labor productivity for 1990-1996 computed from the ISDB data.

The now familiar convergence story is evident. The coefficient of variation (the ratio of standard deviation to mean) among the 24 OECD countries listed in Table 1 declines by more than half between 1950 and 1990. Results are also shown for a sample of Industrial Market Economies (IMEs), as classified by the World Bank, which consists of all OECD countries except Greece, Portugal, and Turkey. Convergence is even stronger among this group, with the coefficient of variation falling by almost two-thirds. The 14 ISDB countries are by and large the biggest OECD economies, excluding countries such as Austria, Greece, Iceland, Ireland, New Zealand, Portugal, and Turkey. As a result, the convergence results for the ISDB countries are very similar to those for the 21 IMEs. However, after 1980, the rate of convergence in RGDPW slows markedly in all three samples. Indeed, among the 14 ISDB countries, there is virtually no convergence in labor productivity levels during the 1990s.

Catch-up is also evident, as indicated by the correlation coefficient between the 1950 RGDPW level and the rate of growth of RGDPW after 1950. The correlation coefficient is -0.93 among all OECD countries, -0.92 among the IMEs, and -0.93 among the ISDB sample. The results indicate that the countries with the lowest productivity levels in 1950 experienced the fastest increase in labor productivity.

Table 2 shows the results of a regression analysis, where other pertinent factors beside the catch-up effect are included to explain country-level productivity growth. The convergence hypothesis predicts that the coefficient of initial productivity will be negative (that is, countries further behind near the beginning of the period will show more rapid increases in GDP per worker). The coefficients of the investment rate, R&D intensity, and education should be positive. Results are shown in Table 2 for all OECD countries over the 1950-1990 period and for a variety of educational measures.

The initial RGDPW level of the country relative to the U.S. level is by far the most powerful explanatory variable in accounting for differences in labor productivity growth among OECD countries. By itself, the catch-up variable explains 74 percent of the variation in RGDPW growth over the 1950-1990 period. The coefficient of the investment rate (INVRATE) is positive and significant at the 5 percent level or greater. The average investment rate, together with the catch-up variable, explains 80 percent of the variation in RGDPW growth. R&D intensity (RDGNP) is significant at the 10 percent in all cases and, together with the other two variables, explains 83 percent of the variation in labor productivity growth. The secondary school enrollment rate (SCND-ENRL) has a positive coefficient that is significant at the ten percent level. Together with the other three variables, it explains 86 percent of the variation in productivity growth among the OECD countries. The international trade variables are not statistically significant.

Table 1 Real GDP per Worker (RGDPW) in OECD Countries, 1950-1996
(Figures are in 1,000s, 1985 Dollar Equivalents)

							Annual Growth Rate (percent)	
	1950	1960	1970	1980	1990	1996	1950-90	1950-96
Australia	16.1	19.3	25.2	27.3	30.3	32.7	1.59	1.58
Austria	5.9	10.7	18.0	23.5	26.7		3.79	
Belgium	10.9	14.3	22.2	27.7	31.7	34.6	2.66	2.56
Canada	16.1	19.5	24.9	28.7	34.4	36.1	1.89	1.79
Denmark	10.8	14.8	20.0	21.5	25.0	27.5	2.09	2.07
Finland	7.0	11.6	17.0	21.8	27.4	32.4	3.41	3.41
France	8.8	13.5	21.6	26.8	30.4	32.1	3.10	2.88
Germany, West	7.3	13.9	21.3	27.3	29.5	31.9	3.48	3.27
Greece	3.4	5.2	10.9	15.5	17.7		4.09	
Iceland	8.7	12.6	15.6	22.5	25.0		2.64	
Ireland	6.3	8.4	13.2	18.4	24.1		3.34	
Italy	6.3	11.1	19.4	26.8	30.8	34.5	3.97	3.78
Japan	2.6	5.0	11.5	16.3	22.6	23.4	5.41	4.89
Luxembourg	14.2	18.8	25.4	28.5	37.9		2.46	
Netherlands	11.4	17.1	25.4	29.2	31.2	33.1	2.52	2.37
New Zealand	17.4	21.3	24.1	24.6	25.4		0.95	
Norway	10.2	14.3	19.4	25.3	29.2	34.4	2.64	2.71
Portugal	2.9	4.9	8.4	11.3	16.6		4.34	
Spain	5.0	8.2	16.6	21.4	26.4		4.16	
Sweden	13.2	17.4	23.2	24.9	28.4	32.8	1.91	2.02
Switzerland	15.0	20.1	27.2	29.5	32.8		1.96	
Turkey	1.9	3.2	4.8	6.7	8.6		3.85	
United Kingdom	11.7	14.8	18.6	21.2	26.8	28.8	2.08	2.01
United States	20.5	24.4	30.5	31.7	36.8	38.8	1.46	1.42
Summary Statistics: All 24 OECD Countries								
Mean	9.7	13.5	19.3	23.3	27.3			
Std. Deviation	5.0	5.6	6.2	6.0	6.3			
Coeff. of Var.	0.51	0.41	0.32	0.26	0.23			
Correlation with 1950 RGDPW							-0.93	
Summary Statistics: 21 Industrial Market Economies (All countries except Greece, Portugal, and Turkey)								
Mean	10.7	14.8	21.0	25.0	29.2			
Std. Deviation	4.5	4.7	4.6	3.9	4.0			
Coeff. of Var.	0.42	0.32	0.22	0.15	0.14			
Correlation with 1950 RGDPW							-0.92	
Summary Statistics: ISDB-14 Countries								
Mean	10.9	15.1	21.4	25.5	29.6	32.4		
Std. Deviation	4.4	4.4	4.3	3.9	3.5	3.7		
Coeff. of Var.	0.41	0.29	0.20	0.15	0.12	0.11		
Correlation with 1950 RGDPW								-0.93

Sources: own computations from the Penn World Table Mark 5.6 and the OECD ISDB (International Intersectoral) Database. See the text for details.

Table 2 Regressions of the Annual Growth in Real GDP per Worker (RGDPW) On Initial RGDPW, the Investment Rate, R&D Intensity, and the Secondary School Enrollment Rate, All OECD Countries, 1950-1990

Relative $RGDPW_0$	INVRATE	RDGNP	SCND-$ENRL_{65}$	R^2	Adjusted R^2	Standard Error	Sample Size
-0.017**				0.74	0.73	0.0056	24
-7.99							
-0.016**	0.063*			0.80	0.78	0.0051	24
-8.35	-2.38						
-0.018**	0.064*	0.314#		0.83	0.80	0.0050	24
-8.20	-2.51	-1.80					
-0.018**	0.070**	0.336#	0.018#	0.86	0.82	0.0047	23
-8.77	-2.88	-2.04	-1.90				

Note:

The sample consists of the 24 OECD countries shown in Table 1. The absolute value of t-ratios are shown in parentheses below the coefficient estimate.

Key:

- Dependent variable: ln(RGDPW90/RGDPW50)/40.
- $RGDPW_t$: GDP per worker in year t, measured in 1985 international prices (in units of $10,000). Source: Penn World Table Mark 5.6.
- Relative $RGDPW_0$: RGDPW level of the country relative to the RGDPW level of the U.S. in 1950. Source: Penn World Table Mark 5.6.
- INVRATE: Ratio of investment to GDP (both in 1985 dollar equivalents) averaged over the regression period. Source: Penn World Table Mark 5.6.
- RDGNP: the ratio of R&D expenditures to GNP averaged over the regression period. Source: UNESCO Statistical Yearbooks, various years.
- SCND-$ENRL_{65}$: Total enrollment of students of all ages in secondary school in 1965 as a proportion of the total population of the pertinent age group. Sources: World Bank, World Development Report 1988.

\# significant at the 10 percent level, 2-tail test.

* significant at the 5 percent level, 2-tail test.

** significant at the 1 percent level, 2-tail test.

3. Conclusion

This brief article has documented the gradual catch-up and convergence in productivity levels that has characterized the OECD countries over the postwar period. It has, moreover, shown that investment, education, and R&D investment have also played a role in this process. However, what is most surprising — if not disturbing — is that this process of convergence appears to have ended in the 1990s. Since that time, at least as far as the evidence shows, there has been almost no additional convergence in productivity levels. This is partly a result of the sluggish growth performance of both the European countries and Japan in the 1990s. Moreover, this lack of convergence has occurred even before the recent acceleration of productivity growth in the United States over the last three or four years.

Another reason for the slow convergence in the 1990s is that as countries catch up, the forces of convergence themselves diminish. This is particularly true for the catch-up term, which accounted for most of the productivity convergence among OECD countries. It is also germane for underlying factors such as investment, education, and R&D, which have converged over time among OECD countries as differences in their productivity levels have narrowed.

References

Abramovitz, Moses, 1986. "Catching Up, Forging Ahead, and Falling Behind." *Journal of Economic History*, 46: 385-406.

——————-, 1994. "Catch-Up and Convergence in the Postwar Growth Boom and After," in *Convergence of Productivity: Cross-National Studies and Historical Evidence*, edited by William J. Baumol, Richard R. Nelson, and Edward N. Wolff. New York, NY: Oxford University Press, 86-125.

Abramovitz, Moses and Paul A. David (1973), "Reinterpreting Economic Growth: Parables and Realities," *American Economic Review*, Vol. 63, No. 2, May, pp. 428-439.

Griliches, Zvi (1979), "Issues in Assessing the Contribution of Research and Development to Productivity Growth," *Bell Journal of Economics*, Vol. 10, No. 1, Spring, pp. 92-116.

Mankiw, N. Gregory, David Romer, and David N. Weil, 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107: 407-38.

Summers, Robert, and Alan Heston (1991), "The Penn World Table (Mark V): An Expanded Set of International Comparisons," *Quarterly Journal of Economics*, Vol. 106, pp. 327-336.

Price Cap Regulation and Productivity Growth

Jeffrey I. Bernstein*
Carleton University
and NBER

1. Introduction

Price cap regulation has become a popular form of regulation in many industries in the US, Canada, and the UK. In the U.S., for example, since 1999 at least 36 states were operating under some form of price cap regulation to govern the intrastate activities of their telecommunications suppliers. Since January 1998 Canadian telecommunications regulation has operated under a price cap regime. Price cap regulation typically specifies a minimum average rate at which the prices that a regulated firm charges for its services must decline, after adjusting for inflation. This rate is called the offset, or X factor. Downward price adjustments arise, in part, from productivity improvements, based generally on historical experience, considered achievable during the price cap period.

The proper choice of an X factor is critical for the long-term viability of any price cap regulation plan. If too small an X factor is imposed, the regulated firm will earn excessive profit, if too large an X factor is imposed, the viability of the regulated firm can be threatened. The essence of price cap regulation is to select an appropriate X factor that challenges the regulated firm, and that promises gains for customers. This paper describes the role of total factor productivity (TFP) growth in the implementation of current typical price cap regulation.

In section 2, we characterize the role of productivity growth in the guideline to X factor determination adopted by many regulatory authorities in Canada, the US, and UK. This basic guideline provides that the X factor should reflect the extent to which the regulated industry has; i) historically achieved, and expected in the future to attain, higher productivity growth, ii) faced, and is expected to continue facing, lower input price inflation than competitive industries in the economy. This guideline is appropriate during the specified period of price cap regulation when all of the regulated firm's services are subject to price cap regulation.

Given the prominent position of productivity growth within price cap regulation, in section 3 we provide a discussion of the major issues pertaining to the characterization of productivity growth targets. In order for price cap regulation to emulate competitive market forces, the productivity target must be achievable by the average provider within the industry during the price cap period. The target must also reflect long-term productivity trends as opposed to short-term fluctuations and one-time anomalies. Lastly, the productivity target must be invariant to manipulation by regulated firms, and regulators.

The basic guideline is appropriate when all of the regulated firm's services are subject to price cap regulation. In section 4 we describe a method pivoting off of the basic guideline that shows how the X factor should be modified when a regulated firm produces both "capped services" (those subject to price cap regulation) and "uncapped services" (those not subject to price cap regulation). The X factor should be modified when the prices of uncapped services are growing at a rate that differs from the rate dictated by the basic (all-service) X factor. Failure to do so generally imposes too stringent an offset on the regulated firm.

2. The Basic Approach

Price cap regulation is intended to replicate the discipline of competitive market forces. Competitive forces compel firms to realize productivity gains and to pass these gains on to their customers in the form of lower prices, after accounting for increases in input prices. Therefore, if all industries in an economy were competitive, output prices in the economy would grow at a rate equal to the difference between the growth rate of input prices and the rate of productivity growth.

If a regulated industry operated typically like any industry in a competitive economy, the discipline of competitive forces could be replicated by limiting the rate of growth of regulated prices to the economy-wide rate of price inflation. This restriction would require the regulated industry to realize the same productivity gains that are realized in other industries of the economy, and to pass these gains on to customers, after adjusting for the typical rate of input price inflation. Therefore, the X factor is zero when the regulated industry is capable of achieving exactly the same productivity growth rate and faces exactly the same rate of input price inflation as other industries of the competitive economy.

More generally, a positive X factor reflects at least one of the following two conditions:

(1) The regulated industry is capable of increasing its productivity more rapidly than are other industries of the economy,

(2) The prices of inputs employed in the regulated industry grow less rapidly than do the input prices faced by other industries of the economy.

In symbols, the average growth rate of prices in a price cap regime, PG, contains two elements, an inflation rate (IR) and an offset (or X factor) to the rate of inflation:

$$PG = IR - X.$$

$$X = \{[WG(Economy) - WG(Industry)] + [TFPG(Industry) - TFPG(Economy)]\}$$

where; WG is the input price growth rate, and TFPG is the TFP growth rate. Under price cap regulation, if the regulated industry is able to achieve more rapid productivity growth, or to face lower input price inflation than other industries, then the regulated industry should be required to pass the associated benefits on to customers in the form of lower prices.[1]

Generally, the rate of inflation is taken to be an average growth rate in output prices for the economy. One example is the Gross Domestic Product Price Index (GDPPI).[2] Typically regulators calculate the historically justified X factor and then set the price cap X factor equal to the historical rate unless elements can be identified which will cause future values to depart systematically from historical values. If future X factors are expected to differ from past ones then adjustment to the historical values must be made. These identifiable factors are customarily most important when price caps are being first introduced. Efficiency incentives created by the switch to price caps could result in higher future rates of productivity growth than were historically observed, although there is little evidence to substantiate this claim.

Any productivity gains associated with the switch to price caps must be temporary and dissipate when price caps have been in effect for some time. It is also possible for future rates of productivity growth to diverge from past rates, even if price caps already exist for some time. Indeed, productivity growth rates can decline. For example, one of the main sources of measured productivity growth arises from output growth due to scale economies. Since many regulated industries exhibit economies of scale, if future output growth decreases from the historical rate due to competitive pressures, then future productivity growth rates would be lower than those historically observed.

3. Productivity Target

Under price cap regulation a firm has the potential to improve its financial position by becoming more productive through the introduction of new processes, products, and management practices. The extent to which an incumbent carrier is able to improve its financial viability depends, in part, on the X factor set by the regulator. Therefore, the selection of an appropriate productivity target is critically important to the success of price cap regulation. The main issues surrounding the selection criteria of the productivity target are reviewed in this section.

3.1 Industry TFP Target

One of the key requirements for proper price cap regulation is to base the offset on an industry-wide productivity index, instead of the performance of a particular regulated firm. The X factor should reflect, in part, the differential between the regulated industry's productivity growth, and that of the overall economy. If price cap regulation is to emulate competitive markets the regulated firms should be rewarded for superior productivity performance.

Superior performance must be defined in terms of outperforming rivals and not oneself. This feature improves the firm's incentive to become relatively more productive, and captures the essence of a competitive situation.[3] Since a firm can increase its profit, if it can achieve greater productivity gains than those reflected by the industry productivity target, it has an incentive to be more productive than the industry as a whole. Firms that fall short of the industry productivity target see their profit erode, and thereby have an incentive to improve and exceed the industry productivity norm. Thus it is the nexus between industry productivity gains and firm profit potential that drives the regulated firm to allocate scarce resources in a manner that facilitates its ability to surpass industry-wide productivity growth.

3.2 Long Term TFP Target

Price cap regulation encompasses a target productivity growth rate for the regulated industry. However, productivity growth rates fluctuate yearly. The proper implementation of price caps (currently) configured necessarily requires the ability to distinguish the long-term trend in TFP from short-term fluctuations. Use of the secular productivity trend causes the average price of the regulated firm to adjust to a long-term productivity potential, thereby contributing to the stability of the regulatory plan.

Secular productivity trends mitigate the impact of one-time events on annual TFP growth. Not all productivity gains achieved in one period are sustainable into future periods. For example, labor cost reductions in one year by the regulated industry, (cost reductions by one firm should not have any influence on the productivity target, since it is an industry-wide target) may lead to temporary productivity gains. However, these new gains may not be sustainable for more than a couple of years. Conversely, a short-term drop in revenue, for example as a result of a downturn in aggregate demand in the economy, may result in a temporary output reduction that is not matched by corresponding input cost reductions. This would not result in a reduction of the long-term productivity growth trend.

Cyclical fluctuations in productivity also arise because some factors of production, such as capital and skilled labor, are fixed in the short-run. For example, when economic activity contracts, firms do not reduce its capital to the same degree that output declines because capital adjustment is very costly. Moreover, using long term productivity growth rates in X factor calculations accommodates the lumpiness of investment. Large capital projects, embodying technological advances, may be required over some time period, followed by a period of relatively low investment. These lumpy, and discrete capital additions initially lead to higher costs and thereby lower productivity growth. However, once the new capital is deployed, productivity growth increases. Short-term productivity fluctuations are exacerbated in capital-intensive industries, such as telecommunications, resulting from timing mismatches of costs and revenues. Significant expenditures may be required in the early stages of a capital project, but the benefits of that investment may not be realized for a number of years in the future. As a result industry TFP growth may be initially understated, while overstated in later periods.

3.3 Immutable TFP Target

An important feature of price cap regulation arises from the commitment that regulators will not view increased firm profitability as a source of regulatory failure. Indeed, the potential for increased profit is a fundamental premise underlying price cap regulation. Incentives are diluted if the productivity target is altered by the recent past performance of the regulated firm. If the potential exists within the regulatory framework for increasing the TFP target in response to large productivity gains in prior periods, then the incentive to innovate in the future by investing in cost reducing production methods, and developing new goods, and services will be substantially reduced or eliminated.

Review and revision of the X factor by the regulator within the price cap period is inappropriate under price cap regulation. If the productivity target is reviewed within the price cap period then the review creates an erratic and unpredictable target, undermining the incentive for productivity improvement under price cap regulation. Blunting of productivity-improving incentives also operates if price cap performance results are monitored over a relatively short period of time, with the possibility of recalculating the productivity target. In this situation, as the case of frequent productivity reviews, the potential for recapture of past productivity gains will severely dull efficiency incentives.

A firm must have a reasonable degree of certainty that it will retain the benefits of increased productivity beyond the industry norm that results from deployment of new technology, new product development, and restructuring or improvements in operations. In the absence of reasonable certainty, the firm will not have sufficient incentives to undertake such activities. To ensure that the regulatory framework provides proper efficiency incentives, price cap performance review periods should be sufficiently long so that the threat of recapture of past productivity gains is minimized.

4. Limited Regulatory Span

The discussion of the basic framework for price cap regulation in section 2 assumed that all of the regulated firm's services are subject to price cap regulation. Price cap regulation is generally applied to only a subset of the services supplied by the regulated firm. For example, in telecommunications basic local services are typically regulated while long distance services are often unregulated. There would not be any need to distinguish between regulated and non-regulated services in price cap plans if all prices grew at the same rate. Additionally, the basic characterization of the X factor provided above could be implemented without modification if productivity and input price and quantity data that pertained exclusively to regulated operations were available. However, joint products and common costs generally make it impossible to derive productivity growth rates and input price growth rates separately for "capped services" (those subject to price cap regulation) and for "uncapped services" (those not subject to price cap regulation). Consequently, the guideline described above must be modified to define an appropriate X factor for capped services.

In this section we describe a method pivoting off of the basic guideline that shows how the X factor must be modified when a regulated firm produces both capped services, and uncapped services. The X factor for capped services must be decreased when the prices of non-capped services are falling more rapidly (or growing more slowly) than prices dictated by the basic (all-service) X factor. Failure to so so imposes to stringent an offset on the regulated firm.[4]

Calculation of the offset begins with the fact that the average growth rate of prices (PG) for a regulated firm are separable into growth rates for prices subject to price caps (PGC) and growth rates for prices that are uncapped (PGU). The weighted sum of output price growth rates is shown as:

$$PG = \alpha PGC + (1-\alpha)PGU = IR - X$$

$$X = \{[WG(Economy) - WG(Industry)]$$
$$+ [TFPG(Industry) - TFPG(Economy)]\}$$

where α is the revenue share of capped services. Isolating the average growth rate of capped service prices leads to:

$$PGC = IR - X^*$$

$$X^* = X + Xa$$

$$Xa = [(1-\alpha)/\alpha]\{PGU - WG(Industry) + TFPG(Industry)\}$$

The rationale that underlies this adjustment is straightforward. Price cap regulation is designed to compel the firm to pass on anticipated (industry-wide) productivity gains to customers in the form of lower prices, after correcting for increases in input prices. If the prices of non-capped services are falling more rapidly than they would be if they reflected only anticipated productivity gains and input price increases, then the firm is actually passing on to customers of non-capped services more benefits than price cap regulation for the firm's entire operations would dictate. Under these circumstances, the prices of capped services must decline by less than the rate implied by price cap regulation based on the firm's entire operations. This reduction implies a reduced X factor relative to the basic all-service offset. Similarly, the appropriate X factor for capped services must be increased when the prices of non-capped services are falling less rapidly than prices dictated by the basic (all-service) X factor.

In general, the X factor must be based on the services subject to price caps. Prospective differential growth rates between capped and non-capped prices necessitate an offset adjustment for the basic X factor. Failure to do so significantly biases the offset imposed on the regulated firm.

The magnitude of the appropriate adjustment to the X factor can be substantial. To illustrate this fact, consider the following example drawn from the current Canadian telecommunications price cap plan.

Example:

TFP Growth(Industry)	4.2%
Less TFP Growth(Economy)	1.0%
Plus Input Price Growth(Economy)	3.0%
Less Input Price Growth(Industry)	2.7%
Equals	
X Factor	3.5%
Plus (uncapped revenue, 60%/ capped revenue, 40%) Multiplied by	
Uncapped Price Growth	-3.8%
Less Input Price Growth(Industry)	2.7%
Plus TFP Growth(Industry)	4.2%
Equals Uncapped Adjustment	
Xa Factor	-3.45%

X* (Capped)= X(Overall) + Xa = 3.5% - 3.45% = 0.5%

5. Conclusion

This article provides three main messages. First, there is a simple guideline to the selection of an appropriate X factor in price cap regulation plans. The guideline states that the X factor reflects the extent to which the regulated industry has achieved more rapid productivity growth and faced lower input price inflation than other sectors of the economy. Second, in order for price cap regulation to emulate competitive market forces, the productivity target must not be firm specific, but relate to the industry. The target must also reflect long-term productivity trends as opposed to short-term fluctuations, and the productivity target must be invariant to strategic manipulation by regulated firms, and the regulator. The third message is that the offset guideline must be modified to account for limited spans of regulatory control. The details of the appropriate adjustments are intuitive and their magnitudes are readily calculated. Failure to make the adjustments result in X factors that deviate significantly from their most appropriate levels.

Notes

* The unabridged version of this article can be found online at www.csls.ca. Email: jeff.bernstein@carleton.ca.

1 The offset is often referred to as the productivity offset. However, this terminology is incorrect since the differential in input price growth rates between the regulated firm and the economy also forms part of the offset.

2 In a strict sense the regulated industry's services should be netted out of the economy-wide measures. However, in practice, the subtraction from economy-wide measures is not undertaken since a regulated industry's services form a small part of the overall economy.

3 Theoretically it may be appropriate to apply an industry-wide productivity target to a given firm that excludes the firm's own productivity performance. Therefore, the firm's productivity would have no effect on the industry productivity measure included in its price cap formula. However, this would require a separate productivity target for each firm operating under price caps. This may not be a practical option.

4 Similarly, the X factor for capped services must be increased when the prices of non-capped services are falling less rapidly than prices dictated by the basic (all-service) X factor.

References

Bernstein, J.I., Sappington, D. (1999). Setting the X Factor in Price-Cap Regulation Plans. *Journal of Regulatory Economics*, 16 (1), 5-25.

What Does Total Factor Productivity Measure?

Richard G. Lipsey
Simon Fraser University
Kenneth Carlaw*
University of Canterbury

This paper is intended more to ask questions than to assert answers. Most strong assertions that we do make should be taken only as hypotheses. We hope in presenting the paper to learn enough to force us to change our mind on many issues. We do not, however, believe that the confusion around the interpretation of TFP numbers is unique to us. The quotations given below show the different interpretations that various eminent economists put on TFP measures.

(1) "Economists tend to think of productivity as measuring the current state of technology used in producing the goods and services of an economy (or industry or firm), and want to interpret the changes in such a measure as reflecting 'technological change', shifts in the production possibilities frontier. For this purpose, it is usual to focus on one or another version of 'multi-factor productivity'." (Griliches: 1010)

(2) "When economists speak of productivity growth, this is essentially what they mean — the growth rate is the economy's ability to produce output from a given stock of inputs. In ordinary parlance, one can think of productivity growth as 'working smarter' rather than 'working harder'. ... [T]otal factor productivity of an economy only increases if people 'work smarter' and learn to obtain more output from a given supply of inputs. Improvements in technology — the invention of the internal combustion engine, the introduction of electricity, of semiconductors — clearly increase total factor productivity." (Law: 6&7)

(3) "The defining characteristic of [total factor] productivity as a source of economic growth is that the incomes generated by higher productivity are external to the economic activities that generate growth. These benefits "spill over" to income recipients not involved in these activities, severing the connection between the creation of growth and the incomes that result." (Jorgenson, 1995 pp. xvii.)

(4) "The central organising concept...[is] the division of observed growth in output per worker into two independent and additive elements: capital-labour substitution, reflected in movements around the production function; and increased efficiencies of resource use, as reflected by shifts in this function. To maintain additivity, ...the analysis...could not be applied cumulatively without introducing an interaction term between capital substitution and increased efficiency. ...[T]he residual productivity debate never did attempt to answer the question, of what is the residual composed?

This remains the dominant question." (Metcalfe: 620)

(5) "...over the long run...the indices [of TFP] do in fact reveal the increased productivity associated with technological possibilities, either in the form of technical progress or through a better use of all available technologies." (Statistics Canada, March 1996: 119)

(6) "Technological progress *or* the growth of total factor productivity is estimated as a residual from *the* production function.... Total factor productivity is thus the best expression of the efficiency of economic production and the prospects for longer term increases in output." (Statistics Canada 13-568: 50-51, italics added)

(7) "[T]otal factor productivity (TFP) measures the output produced by given amounts of labour and capital together. High TFP indicates a high level of technology and means that both capital and labour can earn large rates of return while cost of production remains low." (Dollar and Wolff: 3).

(8) "The pioneers of this subject were quite clear that this finding of large residuals was an embarrassment, at best 'a measure of our ignorance'." (Griliches, 1994: 1)

(9) "A growth-accounting exercise [conducted by Alwyn Young.] produces the startling result that Singapore showed no technical progress at all." (Krugman: 55) "Singapore will only be able to sustain further growth by reorienting its policies from factor accumulation toward the considerably more subtle issue of technological change." (Young: 50)

Quote (1) says that economists think of changes in TFP as measuring technological change. Quote (2) says that TFP measures all improvements in technology, including such things as the introduction of electricity and the motor car. In direct contrast, quote (3) tells us that TFP measures only externalities and other free gifts associated with economic growth. Quote (4) tells us the TFP measures are valid only for measurements taken over relatively short periods of time. In contrast, quote (5) tells us TFP measures the effects of technological change and increases efficiency over long periods of time. Quote (6) goes even further and says that TFP measures the *prospects* for longer term increases in output. Quote (7) says the height of TFP measures the height of the technology in use and the magnitude of the returns to capital. In direct contrast to all the others, quote (8) cautions us that TFP is a measure of our ignorance, a measure of what we do not know. The quotes in (9) assume that low TFP measures for Singapore indicate that during the period when its per capita income rose from third-world levels to those of industrialised countries, it underwent no technological change and was thus on an unsustainable path such as that followed by the USSR.

It is something of an understatement to say that all of these statements about TFP cannot be correct. We are interested in long run technological change. We believe that much of the world's population has been living through a series of massive technological shocks that are associated with what is called information and communications technologies (ICTs). Some economists doubt this judgement because, among other things, there was no evidence of big technological shocks in the measurements of total factor productivity (TFP) over the 1990s. If anything, judging from the TFP figures, technological change seems to have slowed over the last two decades of the 20th century.

In this paper, we discuss this objection. To do this, we discuss what TFP does and does not measure, relating productivity measures to details of what is known about technological change as a result of the writings of such students of technology as Nathan Rosenberg, Christopher Freeman and Paul David. We conclude that standard TFP measures that purport to measure technological change do not do so;

nor do they measure the importance of externalities and other "free gifts" in the growth process.

Indeed, we argue that looking for the importance of new technological knowledge in measures of the type of externalities studied by Arrow in his classic 1962 article, or in observed discrepancies between private and social rates of return to R&D, is to look in the wrong place. The importance of technological change for economic growth is in the technological complementarities that it creates, not in the externalities. These complementarities go largely unmeasured by TFP calculations. Understanding these issues is a key to understanding the place of technological change in the growth process and is also important in avoiding some common misinterpretations of empirical measures of productivity change. We go on to argue that whatever may or may not be measured by TFP, it cannot be understood as measuring technological change or technological dynamism.

The Technological Background

We conceive of society as starting with two basic factors, labour, L, and the endowments provided by nature, R.[1] Society then produces output, Y, using these endowments along with two main bodies of created assets, physical capital, K, and human capital, H. The measurement of output poses massive problems, which we assume to be solved as we want to focus on the input side.[2]

We use a wider definition of technology than is usual. For us, technology is the idea set of how to create economic value, while the economic structure includes the embodiment of these ideas in such things as machines, plant layout, firm organization and location, infrastructure and financial institutions. Our definitions of technology and structure are described in full in the Appendix of the unabridged version of this paper.

Technological Change

The overall technology systems of all growing economies evolve along paths that include both small incremental improvements and occasional jumps. To distinguish these, investigators often define two categories. An innovation is *incremental* if it is an improvement to an existing technology. An innovation is *radical* if it could not have evolved through incremental improvements in the technology that it displaces-e.g., artificial fabrics could not have evolved by incremental improvements out of the natural fabrics that they displaced in many uses.

An extreme form of radical innovation is called a general purpose technology (GPT). GPTs share some important common characteristics: they begin as fairly crude technologies with a limited number of uses; they evolve into much more complex technologies with dramatic increases in their efficiency, in the range of their use across the economy and in the range of economic outputs that they help to produce. As they diffuse through the economy, they are improved in efficiency and expanded in their range of use.

As mature technologies, they are widely used for a number of different purposes, and they have many *technological complementarities* in the sense of co-operating with, and sometimes requiring amendments to, many other technologies, as well as creating myriad possibilities for the invention of new technologies.[3] The steam engine, the dynamo and the internal combustion engine are examples of major GPTs in the field of energy generation. An important thing about a GPT is that it creates a new research program for invention and innovation. As the research program proceeds, new opportunities open up exponentially. Then as the GPT matures, the number of new opportunities created per unit of time may fall steadily, causing the returns to further investment in invention and innovation to fall steadily. Note this is not the diminishing returns as the capital stock grows with technology constant.

Instead, it is diminishing returns caused by a reduction in the new rate of creation of new investment opportunities as a GPT matures.

Most technology is embodied in new capital equipment whose accumulation is measured as gross investment. So technological change and investment are interrelated, the latter being the vehicle by which the former enters the production process. Anything that slows the rate of embodiment through investment, such as unnecessarily high interest rates, will slow the rate of growth, just as any slowdown in the development of new technology will do so in the long term. So, from the fact that new investment and growth in employment can statistically "account for" most economic growth, it does not follow that these are the main causes of growth. Both technological change and investment are needed.[4]

Holding technology constant

Since we are considering attempts to measure technological change, the concept of constant technology is critical. To hold technology constant conceptually we need to do the following: hold all product, process, and organizational technologies constant at what was known at some base period; accumulate more physical capital that embodies the technologies then in use, or others that were known but not in use; accumulate more human capital in the form of more education in what was known at that time. Then calculate the increase in output. This is a measure of what could have been achieved without any alteration in technological knowledge. Now calculate the actual increase in output. The difference is "due to" or "enabled by" technological change in the sense that it could not have happened without such change.

Measured over the long run, say a century, the difference due to technological change would be massive. Here are just a few illustrative examples of what the constant-technology experiment would reveal.

- Feeding 6 billion people with the agricultural technologies of 1900 would have been literally impossible.[5] Sooner or later Malthusian checks would have become a reality as ever expanding populations would have encountered increasing food shortages. (Among other things this shows that population and the labour force cannot be taken as independent of technology.)
- Pollution would have become a massive problem. By our standards, 1900 technologies were heavily polluting and to increase production sufficiently to employ all the new capital would have led to major increases in pollution.
- Resource exhaustion would have been serious. Most new technologies are absolutely saving in resources.[6] Thus to produce today's manufacturing and service output with 1900 technologies would have required vastly more resources than are currently used.

In another demonstration of these points, the calculations of the Club of Rome in the 1970s showed the folly of believing that production could long be increased at current rates with no change in technology. The Club's predictions of doom were falsified by continued technological advance, which invalidated their calculations on resource exhaustion and unsustainable pollution. But these mistaken predictions do show for how few decades current world growth rates could be sustained in a world of static technology.[7]

Continuity and discontinuity in technological change

Aggregate growth models use an aggregate production function of the form:

$Y= F(K, H, L)$. Such models display a continuously declining marginal product of physical and human capital that is only increased by continual gradual increases in the productivity constant A. But one of the most important propositions that matter later in this paper is that technological change is not like that. It does not cause continuous variations in the marginal productivity of capital as more capital is accumulated and as technological change slowly and continuously alters the relation between inputs and outputs. Instead major technological advances of the sort now called new GPTs cause discontinuities in the opportunities for new investment and radically alter the relation between inputs and outputs at the microeconomic level. As Lipsey and Bekar put it some time ago in another context:

> "[As] any one technology evolves over time ...its productivity [may] approach an upper limit beyond which further improvements are difficult if not impossible. [However] when there is a shift from one technology to another and then to a third, there is no reason to expect any particular relation between the increments of output that arise in moving between technologies... Consider, for example, successive energy technologies... The factories of the First Industrial Revolution were powered by water power (K_1). Then, in the early nineteenth century, the transition was made to steam (K_2). Then in the period from 1890 to 1930 electricity replaced steam... (K_3). At some future time, nuclear fission and/or fusion may replace fossil fuels as the main generators of power for electricity (K_4). Still later, electricity itself will give way to some hard-to-imagine, new energy source... (K_5).
>
> Now consider the increment in total constant-dollar value of production $[Y(K_{n-1}) - Y(K_n)]$ that is due to each shift from one technology to the next technology. No matter how the impact is measured, there is nothing in physics or economics to suggest that the increments in going from one technology to the next have to be ordered in the following way:
>
> $[Y(K_2)-Y(K_1)] > [Y(K_3)-Y(K_2)]>[Y(K_4)-Y(K_3))]>[Y(K_5)-Y(K_4)]$
>
> Indeed, there are no currently known general principles to suggest any particular relation, and we can see may different ones in history. Historical experience suggests that, as time passes, one technological change may bring massive gains, to be followed by another that brings smaller gains, to be followed by a shift that brings larger gains — and that this is true on each of the many fields in which technology is used."

Interpreting the potential of each GPT in terms of the research program for invention and innovation that it creates, Lipsey and Bekar are saying that the richness of successive programs associated with successive GPTs may stand in any relation to each other. A GPT will only be introduced if it is more efficient than the technology it replaces. But GPTs that meet that criteria, some will have very much richer potential for dependant innovation than others. Some will therefore lead to more rapid and more fundamental changes in technology and productivity than others.

Is there a productivity paradox?

Models of both exogenous and endogenous growth that use an aggregate production function do not explicitly model technology. They conceal technology and technological change in the black box of the aggregate production function that transforms inputs into total output

(GDP). Technology is itself unobservable in these models, while changes in technology are observed only as a residual when changes in measured outputs are not fully matched by changes in measured inputs. Changes in the TFP residual are assumed to measure changes in the production function and hence to be related to changes in technology. The so-called productivity paradox relates to the observation at the micro economic level of major technological changes associated with the ICT revolution combined with the observation of low rates of productivity growth at the macro level as measured by TFP (and other productivity measures).

We do not accept the existence of a productivity paradox (assuming productivity can be properly measured). We argue two key points briefly in the text and in more detail in the appendix, which compares the model we use for viewing long term growth and technological change with standard macro models.

First, new fundamental technologies, even where they do greatly raise productivity and living standards over the long term, typically have long gestation periods. They often need decades to develop their full potential.[8] Also, the whole structure of the economywhat we call the facilitating structure as outlined in the appendixoften needs to be altered. As Paul David (1991) has emphasized, this requires much time, so that these technologies may show only small macro benefits in the short term of several decades, and but big payoffs over the very long term of half a century or more.

Second, and much more fundamentally, we argue, in contrast to the macro models mentioned above, that there is no necessary relation between technological changes and productivity changes, however the latter are measured. For example, from 1770-1820 the whole of British society and the whole basis of the economy was transformed by the First Industrial Revolution that took work out of the homes and put it into proto-factories — sheds containing hand-powered machines — and actual factories containing water-powered machinery. Although this was a fundamental social and technological transformation that contained the seeds of most of what followed later in the 19th century, changes in productivity and real wages were small or non-existent (Crafts 1985).[9] In contrast, changes associated with the second phase of the Industrial Revolution when steam power was brought into the factories and combined with improved versions of automated textile machinery caused productivity and real wages to rise substantially over the period 1820-1870.

The conclusion to be drawn from these and other similar historical events is that the only thing we can say about a new technology that replaces an existing one is that it must be expected to bring some gain (or else it would not be adopted). But the margin of gain measured in terms of such things as reduced cost and/or increased revenues may be small or large or anything in between. In other words, *productivity changes and technological changes do not stand in any invariant quantitative relation to each other*: *nor are measured productivity changes any necessary indication of the depth of some technological transformation that the society is undergoing.*

Natural Resources Made Explicit

The problems with total factor productivity as a measure of technological change can be illustrated with a preliminary contrast between two positions. (The problems are studied in detail in the unabridged version of this paper.)

Contrasting predictions

The first is the prediction following from the standard neoclassical aggregate production function that measured capital and labour could have been increased at a constant rate from 1900 to 2000 with constant technology and no change in living standards. The second is our argument in a previous section that such an event would have had catastrophic effects on living standards.

To reconcile these conflicting positions we need to recognize that for the neoclassical prediction to hold the capital that would need to grow would include such resource inputs as agricultural land, mineral resources, "waste disposal" ecosystems, fresh water resources, and a host of other things that the standard measurements of capital ignore. Following Solow (1957), however, economists typically define the stock of physical capital used in their growth models to include the stock of natural resources, land, minerals, forests etc. Yet everything that is then assumed about capital is appropriate to physical and human capital and takes no account of the specific problems of natural resources. For example, although the stocks of plant and equipment can be increased more or less without limit, the stocks of arable land and mineral resources are constrained within fairly tight limits.

Since we see society as starting with people and what nature provides, and since the evolution of stocks of capital often obey laws that differ from those that govern the evolution of the stocks of natural resources, we find it conceptually useful in our theorizing to separate the stock of natural resources from the stock of created physical capital. Natural resources are used up in the process of production and some, such as petroleum, cannot be replaced while others, such as trees and the fertility of land, typically (although not invariably) can be. The problems associated with renewable resources are not unlike those associated with physical capital. Most renewable resources such as the air and water that remove pollution and nurture fishing stocks, are naturally renewable up to some maximum rate of exploitation, after which help is needed up to some higher rate of exploitation, above which the stocks may deteriorate even when helped.

To account for these stocks, we need to deduct the natural resources used in current production, and treat as gross capital investment the amounts spent in maintaining existing resources (e.g., the productivity of land) restoring those used up (e.g., reforestation) and discovering new supplies (such as mineral exploration).

Importantly, technological advance alters the economic value of existing natural resources. Some values may be lowered, such as when the invention of the electric motor and the internal combustion engine lowered the value of coal reserves, while other values may be greatly enhanced (sometimes starting from a base of zero value), such as when the introduction of the gasoline engine and the automobile greatly increased the value of petroleum reserves.

These changes in resource values are sometimes consciously created by technological change, as when methods of using low grade iron ore and tailings from previous operations, were invented. At other times, they are the unconscious result of technological advances pursued for other purposes, as when the internal combustion engine, which started out as a stationary engine driven by coal gas, finally settled on petroleum products as its most efficient fuel.

The absence of explicit resource inputs from the neo-classical growth model, poses no problem as far as income goes because all of the value of consumed resources must show up as income for the labour and capital involved in extracting and processing them. But we know that resources are important inputs into any real production function. Rosenberg (1994) has shown, for example, that the so-called American System of Manufacturers, which was the basis of America's overhauling Europe as the prime industrial country, was based on processes that made lavish use of natural resources while economizing on scarce labour. (Capital was also lavishly used at first but, later, the innovations became capital saving as well as labour saving.)

Modeling resources

To illustrate some of the problems associated with the omission of natural resources let the underlying production function be:

(1) $Y = AK^{\alpha}L^{\beta}R^{\delta} \quad \alpha+\beta+\delta = 1$

where K is produced capital and *R* is natural resources, agricultural land, forests, minerals, air, water, etc.

Now let *K* and *L* increase at a constant rate *u*. If nothing else happens, there will be diminishing returns as more labour and capital are applied to a given resource base. Output will be increasing at the rate $(\alpha+\beta)u$ and so per capita real income will be shrinking. ($Y_t = Y_o e^{(\alpha+\beta)t}$, which, since $\alpha+\beta < 1$, is a declining series.)

Now reallocate some of the capital formation to creating technological change in the resource industries. The new technologies are resource saving (as the evidence cited earlier in this paper shows that they are). Assume that as a result of the R&D, resources are also growing at the rate *u*, measured in *efficiency units*. (This is exactly the same as Harrod neutral technological change, only the input whose efficiency is growing is resources rather than labour.) Total income, capital and labour are now all growing at the rate *u* while per capita income is no longer falling. However, if we measure *R* in physical units, *R* will be constant, while *A* will be growing at the rate δu.

If we now take the data generated by (1) and fit a constant returns to scale production function without *R* in it, we will get a perfect fit with the equation :

(2) $Y = BK^{\varepsilon}L^{1-\varepsilon}$

where $B = AR^{\delta}$ and ε will exceed α by the share of capital in the costs of producing resources and $1-\varepsilon$ will exceed β by the share of wages in resources costs. Thus we can fully explain the growth process that is actually being driven by "resource enhancing technological change" in terms of a constant returns production function containing only two inputs, *K* and *L*, and with an unchanged productivity parameter. So measured TFP will be zero and all the increase in output will be ascribed to increases in measured labor and measured capital.

We conclude that because resources are not specifically modelled in the neoclassical growth model, much of the substantial amount of technical change that goes into increasing the productivity of given natural resources will show up as increases in capital, labour (and depending on accounting procedures, possibly also as R&D) and, not as shifts in the production function.

Notice that this conflicts with Griliches' treatment of unrecorded inputs in his discussion of errors in measuring TFP. He argues, not implausibly, that an increase in any unmeasured input would raise output without raising measured costs and so would add to measured TFP. *The present discussion shows that technological change that increases the efficiency of an unmeasured input may show up as an increase in measured costs and hence leave TFP unchanged. This is one of the many reasons why changes in TFP do not measure technological change.*[10]

Conclusions

Here are some of the conclusions that the present study has reached and concerns that it has raised about TFP.

- TFP cannot simultaneously measure all technological change and just the free gifts from externalities and scale effects.
- All improvements in technology, such as the internal combustion engine, do not "clearly raise TFP".
- Increases in output that would not have occurred without technological change (i.e., for which technological change is a necessary condition) do not necessarily cause TFP to increase.
- TFP does not measure "prospects for longer term increases in output" since, among other reasons, new GPTs tend to be associated with up-front costs and downstream benefits.
- There is reason to suspect that TFP does not adequately reflect the increase in a firm's capital value created by R&D activities that are realised through sale rather than exploitation by the developing firm. Yet these are often technological advances created by the use of valuable resources.
- TFP does not adequately capture the effects of those technological changes that operate by lowering the cost of small firms and then allowing large subsequent increases in sales and outputs.
- TFP does not adequately measure the massive amount of technological change that gets embodied in physical capital where the change tends to be recorded as an increases in capital rather than a change in productivity.
- When full equilibrium does not pertain, as in the midst of any lagged adjustment process, the marginal equivalencies needed for successful aggregation do not obtain and there is every likelihood that increases in productivity of labour and capital will be recorded as increases in the quantities of labour and capital inputs.
- New technologies often lead to large up front costs of R&D and learning by doing and using that are incurred in the expectation of future benefits that will be missed when current outputs are related to current costs. The amount of this activity may vary with the life cycle of GPTs and other major technologies and so, thus, may measured TFP.
- Neither TFP nor externalities measure the technological complementarities by which an innovation in one sector confers benefit on other sectorsbenefit for which those in other sectors would be willing to pay but do not have to do so.
- Low TFP numbers for the Asian Tigers do not mean they are in the same boat as was communist Russia; they are quite compatible with successful technology enhancing policies and technological transformation of a country through domestically generated or imported capital.
- TFP is as much a measure of our ignorance as it is a measure of anything positive.

It seems to us that, whatever TFP does measureand there is cause for concern as to how to answer that question — it emphatically does not measure technological change. In the long term, we are interested in increases in output per unit of labour, resources (and waiting in the Austrian sense of the term). While people are of course free to measure anything that seems interesting to them, the degree of confusion surrounding TFP, particularly the assumption that low TFP numbers imply a low degree of technological dynamism, would seem to us to justify dropping the measure completely from all discussions of long term economic growth. Even if that does not happen, as we are sure it will not, every TFP measure should carry the caveat: *changes in TFP do not in any way measure technological change.*

Notes

* This article is drawn from a much longer paper of the same title that can be found at www/sfu.ca/~rlipsey. References omitted from these excerpts are given in the full version. Note that the excerpts give the first part of the paper in its entirety and then jump to the conclusions most of which depend on arguments not given in these excerpts. Email: Richard G. Lipsey: rlipsey@sfu.ca; Ken Carlaw: k.carlaw@econ.canterbury.ac.nz

1 What is provided by nature is exogenous but similar resources may also be created by human effort as when a forest is replanted or a like restocked.

2 As we later argue, the importance of technological change is in preventing the decline of the marginal productivity of capital that would inevitably occur with constant technology. A similar force operates on the output side. If the technology of consumers goods and services were to have been held constant at those existing at some earlier point in time, say 1900, diminishing utility of income would be a reality, as consumers wondered what to do with a third and fourth horse and buggy and train trip to the nearby seaside. Technological changes in consumers goods constantly present consumers with new consumption possibilities and removes at least the inevitability of rapidly declining marginal utility of income as income rises over time, (which does not prevent marginal utility of income from declining at a point in timewhen, of necessity, technology is constant).

3 For a detailed consideration of these characteristics and a development of the definition that follows in the text see Lipsey, Bekar and Carlaw, Ch. 2 in Helpman (1998). The entire book is of value in studying GPTs.

4 Nonetheless consider the choice between two polar cases, one could either live in a society in which technology advanced but was only embodied through "replacement investment" since net investment (and hence measured capital accumulation) was zero, or in a society in which nothing was known that was not known in 1900 and more and more investment had been made in 1900-style productive facilities to produce 1900-style goods and services. We wager that most people would prefer the former alternative. As Solow long ago observed: "One could imagine this [growth] process taking place without net capital formation as old-fashioned capital goods are replaced by the latest models, so that the capital-labour ratio need not change systematically." (Solow, 1957:316)

5 Of course, population is endogenous and it is not clear how much population would have increased if food producing technologies had remained frozen at their 1900 levels. However, Western practices of cleanliness had already lowered death rates in the West and had led to large increases in life expectancy with a resulting population boom, and these practices were already being extended to the less developed countries. Thus some large population expansion would certainly have occurred.

6 "Total materials requirements per unit of (constant) GDP have declined between 1.3% per year in Germany, 2% per year in Japan, and 2% per year in the Netherlands" (Grubler: 240).

7 A further problem arises in altering the capital labour ratio with fixed technology. In so far as the population increase and new capital is merely replicating existing productive facilities staffed by new workers, production can be expanded at more or less constant returns to scale. But this process employs more persons while leaving constant their productivity and hence their real wages and living standards. Raising living standards with static technology requires increasing the capital labour ratio. Although new technologies often do this, existing technologies, especially in manufacturing, typically have little scope for varying the capital labour ratio, factor proportions being more or less built into them. It would be impossible, for example, to take a plant designed in 1900 to produce steam engines and increase the amount of capital per worker by 500%. There was room for some substitution of capital for labour within the confines of existing technology of steam engine construction, but not much. So, if the economy were to grow by increasing the capital per head from its level in 1900 to its level in 2000 without altering technology, it would become increasingly difficult to find places in which the extra capital could be profitably employed. Much of it would end up in non-manufacturing activities, while manufacturing areas, which were, in actuality, some of the main sources of rising living standards throughout the 20th century, would be carried on in unchanged ways with little increase in capital per head.

8 For a more detailed description of the evolution of a GPT see Lipsey, Bekar & Carlaw, Chapter 2 in Helpman (1998).

9 For another interesting piece of evidence, the old technology of hand loom weavers persisted side by side with the new technology of automated weaving for more than 50 years. The number of weavers and their wages fell steadily but the process took a very long time. This could not have happened if the new technology had been radically more efficient than the old, illustrating that there is no necessary likes between big technological change and big productivity gains.

10 This is a caution for macro modellers and those who calculate TFP from aggregate data. However, industry studies that include resources as inputs of downstream industries and as outputs of basic industries may catch these productivity gains by showing a ceteris paribus decline in resource inputs. (Although we would have to know much more than we now do to say whether or not this gain would be washed out when aggregating over all the sectors.)

Labour or Total Factor Productivity: Do We Need to Choose?

Timothy C. Sargent and Edgard R. Rodriguez[1]
Economic Studies and Policy Analysis Division
Department of Finance, Canada

I. Introduction

Which productivity measure is the 'best' has recently been the subject of some debate in academic and policy circles. On the one hand there are those who argue that total factor productivity (TFP)[2] is the appropriate measure of productivity growth, and that labour productivity is a much cruder measure.[3] On the other hand, there are those who argue that TFP depends too much on arbitrary assumptions, and that labour productivity is more closely related to current living standards, which is what society ultimately cares about. In this note, we shall argue that both measures have their place, and that neither tells the whole story.

II. The Basic Neoclassical Model of Productivity

In the basic neoclassical model, output is assumed to depend on the quantity of labour employed, the level of the capital stock, and the level of TFP, a parameter that governs the relation between the inputs, capital and labour, and output. TFP is commonly identified with the level of technology but it actually incorporates a wide variety of factors, such as the internal organization of firms and the level of worker effort. Growth in TFP is usually calculated as a residual, by subtracting the relative contributions of the growth in labour input and capital input from output growth.

When using this framework, one important consideration to keep in mind is that in the neoclassical model growth in the capital stock is not an exogenous determinant of growth in the long run. Rather, it is an endogenous variable that depends on TFP growth. In the long run, if TFP were to stop growing, capital intensity would also stop growing, and there would be no further labour productivity growth and ultimately no further improvements in standards of living.

Thus by adopting the neoclassical model of TFP growth, one is more than just calculating the relative importance of capital in explaining labour productivity growth, one is adopting a *theory* of growth, and one in which causation runs *from* TFP growth *to* capital stock growth.[4] This matters for which measure of productivity growth is most useful, because it means that, in the long run, capital stock growth will tend to follow TFP growth, and not the other way around.

III. Capital-Embodied Technical Change

Models of capital-embodied technical change, such as that of Solow (1960), postulate that much of productivity growth can be attributed to increases in the *quality* of capital goods. In these models technological change is said to be *embodied* in new capital goods.[5]

Incorporating capital-embodied technical change into the neoclassical model provides an additional source of growth. In the long run, labour productivity growth depends on both disembodied technological change and capital-embodied technical change. As with the basic neoclassical model, the rate of capital accumulation should not be thought of as an exogenous determinant of productivity growth. Rather, it is determined by the rate of technological change, both embodied and disembodied. This is because there are still diminishing returns to capital. Even though the quality of new capital goods is increasing, this is a purely exogenous process, and is not affected by an increase in investment.

How should we interpret standard measures of TFP when there is capital-embodied technical change? The answer depends crucially on the extent to which capital stock measures are adjusted by statistical agencies for quality changes. In the case where there is no adjustment made for quality changes, measured TFP growth will pick up both the disembodied portion of technological change and the embodied portion. Thus the conventional growth accounting procedure gives the relevant measure of productivity growth for understanding long-run changes in labour productivity growth. What it does not do is give the portion of long-run growth that is due to capital-embodied technical change.[6]

IV. Productivity in New Growth Models

While the neoclassical model assumes TFP to be essentially determined outside the model, 'new' growth theory, or endogenous growth theory, attempts to explain the evolution of TFP. All new growth theories stress the importance of ideas as the underlying 'engine' of growth. However, they differ in what is posited to be the primary determinant of ideas: some authors stress investment in human capital, others stress investment in new capital goods, or spending on research and development.

What does new growth theory imply for growth accounting? If we assume that TFP depends on investment in human capital, or on that part of spending on R&D not measured as investment, then the standard growth accounting framework largely carries through.[7]

The situation is different if TFP growth depends on capital accumulation, as in Romer (1987), where greater capital accumulation leads to a greater variety of capital goods and thus greater capital quality. If we apply the standard growth accounting procedure in the context of Romer's growth model, we find that TFP depends on both disembodied technological change and the rate of increase of the variety of capital goods. Because the latter is endogenously determined by the resources society devotes to saving, this means that TFP now follows capital accumulation, rather than being a cause of it. As a result, TFP is no longer a better guide to long term trends in productivity than labour productivity.

V. Conclusions

What then do we believe to be the most appropriate measure of productivity growth, TFP or labour productivity? Our view is that the choice should depend on several factors, including the time period of interest, the quality and comparability of the capital stock data, and the growth model assumed.

The Time Period of Interest

In most models of growth, diminishing returns to capital accumulation implies that capital intensity cannot grow faster in the long run than (adjusted) TFP growth. Any tendency for labour productivity to grow faster than TFP is merely a temporary phenomenon. Nevertheless, this adjustment to the long run could well be prolonged, and it is possible for capital intensity to grow more quickly than adjusted TFP for a considerable period of time. So if the intent is to examine trends in the economy over a period of less than a decade or so, then labour productivity is a better guide. If the intent is to examine long run trends in the economy, then TFP may well be a better choice than labour productivity.

The Quality and Comparability of the Capital Stock Data

If there are important biases in the estimates of the capital stock used to construct measures of TFP growth, then it will clearly be better to rely on measures of labour productivity, which are estimated directly using readily available estimates of value added and labour inputs. This is especially relevant when making cross-country comparisons, because the procedures used by different statistical agencies to deal with, for example, depreciation and aggregation, may be quite different.[8]

What the Underlying Model of Productivity Growth Is Assumed to Be

Most attempts to measure TFP involve an implicit assumption about the underlying model of growth. This means that simply estimating TFP growth is not enough: one needs to know how to interpret the resulting time series:

- In the *neoclassical model of growth*, when no adjustments are made for changing capital quality, measured TFP growth reflects both disembodied and embodied technological change, and remains the best guide to long-run trends in productivity. When adjustments are made to the capital stock, measured TFP only reflects disembodied technological change, and additional data are required to construct a measure of embodied technological change.
- In *new growth models in which physical capital accumulation is the engine of growth*, capital intensity drives TFP growth, not the other way around. Therefore trends in capital accumulation are more relevant for examining the growth process than are trends in TFP growth.

Given that both the short run and the long run matter to policy makers, and that there is considerable debate over how to measure the capital stock, and over which is the 'right' model of economic growth, it seems reasonable to argue that both labour productivity and TFP are useful indicators of growth trends in the economy, but that neither should be relied upon exclusively.

Notes

1 The authors would like to thank Bing-Sun Wong, Jeremy Rudin and Benoit Robidoux for helpful conversations. A longer version of this paper is available at www.csls.ca under the International Productivity Monitor. The views expressed in this paper are our own and should not be attributed to the Department of Finance. Email: Sargent.TimothyC@fin.gc.ca

2 Statistics Canada refers to TFP as 'Multifactor Productivity', or MFP.

3 See for example May (2000).

4 Although the neoclassical model is at the heart of how economists usually interpret measures of TFP, it is possible to give an index number interpretation to TFP statistics that does not presuppose a particular economic model. See Hulten (2000) for more details.

5 Models of capital-embodied technical change are often called 'vintage' models, because capital goods of more recent vintages are more productive-of greater quality-than older capital goods.

6 When prices of capital goods are adjusted for quality change, the situation becomes more complicated. Statistical agencies such as Statistics Canada and the Bureau of Labor Statistics now adjust downwards the prices of computers and certain related capital goods to reflect quality change. This means that the real capital stock of these machines is now measured in efficiency units, to give a measure of the effective capital stock. In this case standard growth accounting procedures will only reveal the disembodied part of technological change. To calculate the part of technological change that is embodied in capital goods, one must use the difference between the non hedonically-adjusted price index and the hedonically-adjusted price, as in Greenwood et al. (1997). This will give an empirical estimate of the rate of capital-embodied technological change to calculate the long-run growth path of the economy. Note though that a complication arises if hedonic pricing methods used to calculate not only the effective capital stock, but also to the real value of output. Greenwood et al. (1997) argue that the appropriate way to measure real output is to use the price of consumption goods, not investment goods, to deflate the value of real investment. Failure to do so biases upward real output and therefore labour productivity and disembodied TFP, and in consequence reduces the relative importance of capital-embodied technical change.

7 Because there are still diminishing returns to physical capital, the (physical) capital-labour ratio cannot grow faster than adjusted TFP in the long run, and so TFP growth is still a good guide to long-term growth trends. The pattern of causation still runs from TFP to physical capital: the only change is that now TFP is determined by some other factor, such as investment in human capital or R&D spending.

8 See Coulombe (2000) for more on this topic in the context of U.S.-Canada comparisons. There is also the question of how to aggregate different components of the capital stock. While statistical agencies typically use historical cost to weight different kinds of capital goods, Jorgensen and Griliches (1967) argue that the more appropriate technique from a conceptual standpoint is to weight assets by an estimate of their marginal products. In this way measures of the capital stock can capture increases in the quality of capital resulting from substitution towards capital goods with higher marginal products. Failure to do so will bias downwards estimates of the capital stock and so bias upwards measures of TFP growth. The other reason why official estimates of capital may depart from the theoretically ideal measure is that not all capital goods are used at full capacity at all points during the business cycle. Just as the expression for labour input should only reflect the number of persons employed, so too should the measure of the capital stock include only capital employed. Although in the long run changes in capacity utilisation should even out, in the short term, failure to adjust for changes capacity utilisation will tend to impart a pro-cyclical bias to measured TFP.

References

Coulombe, Serge, 'Three Suggestions to Improve Multi-Factor Productivity Measurement in Canadian Manufacturing', paper presented to the *Centre for the Study of Living Standards* Conference Canada-US Manufacturing Productivity Gap. Ottawa, (January 2000), http://www.csls.ca/jan/Coulombe.pdf .

Greenwood, Jeremy, Zvi Hercowitz and Per Krussel, 'Long-Run Implications of Investment-Specific Technological Change,' *American Economic Review* 87, no.2 (June 1997), pp.342-362.

Hulten, Charles, 'Total Factor Productivity: A Short Biography', NBER Working Paper no.7471 (January 2000).

Jorgensen, Dale W. and Zvi Griliches, 'The Explanation of Productivity Change,' *Review of Economic Studies* 34, no.3 (July 1967), pp.349-383.

May, Doug, comments at the CSLS Conference on the Canada-U.S. Manufacturing Productivity Gap, (January 2000), http://www.csls.ca/jan/May.pdf .

Romer, Paul M., 'Growth Based on Increasing Returns due to Specialization,' *American Economic Review* 77, no.2 (May 1987), pp.56-62.

Solow, Robert M., 'Investment and Technical Progress,' in *Mathematical Methods in the Social Sciences* (Kenneth J. Arrow, Samuel Karlin and Patrick Suppes eds), Stanford University Press (1960).

The Challenge of Total Factor Productivity Measurement

Erwin Diewert *
Department of Economics
University of British Columbia
Vancouver, Canada

In order to measure industry total factor productivity accurately, we require reliable information not only on the outputs produced and the labour input utilized by the industry but we also require accurate information on *eight* additional classes of input used by the industry, namely intermediate inputs; reproducible capital inputs; inventories; land; resources; working capital, money, and other financial instruments; knowledge capital; and infrastructure capital. This article reviews the issues in developing reliable estimates for output, labour input, and the eight additional classes of input for total factor productivity measurement by industry.

The Definition of Total Factor Productivity

The *total factor productivity* of a firm, industry or group of industries is defined as the *real output* produced by the firm or industry over a period of time divided by the *real input* used by the same set of production units over the same time period. However, it turns out to be difficult to provide a meaningful definition of real output or real input due to the heterogeneity of outputs produced and inputs utilized by a typical production unit.

On the other hand, it is possible to provide meaningful definitions of *output growth* and *input growth* between any two time periods using index number theory. Essentially, any sensible *quantity index* aggregates up a weighted average of the rates of growth of each of its components over the two periods in question, which provides a meaningful aggregate growth rate. The two periods are generally taken to be consecutive periods (the chain principle is used in this case) or the current period and a base period (the fixed base principle is used in this case).

Thus total factor productivity growth of a production unit over *two* time periods can be meaningfully defined as an output quantity index divided by an input quantity index where the quantity indexes utilize the output and input price and quantity data that pertain to the production unit for the two periods. Thus if outputs grow faster than inputs, we say that there has been a total factor productivity improvement. Over long periods of time, advanced economies have achieved rates of total factor productivity growth in the range of about 0.5 to 1.5 percentage points per year; i.e., aggregate output has grown about 0.5 to 1.5% faster than aggregate input. Clearly, TFP growth is an important determinant of improvements in living standards. Note that simply measuring TFP growth does not tell us anything about what causes this growth. However, in order to have explanations for TFP growth, it is first necessary to measure it accurately.

First, we note that there can be problems in comparing the TFP growth for an industry which has a large proportion of intermediate inputs relative to its gross output compared to an industry that uses very little intermediate input. To make the TFP growth rates for the two industries comparable, it is necessary to treat intermediate inputs as negative outputs and aggregate them up with the gross outputs of the production unit under consideration. Then TFP growth is defined as an index number aggregate of gross outputs and (negative) intermediate inputs divided by an index number aggregate of primary inputs.

A second technical problem associated with the measurement of TFP growth is that it is difficult to figure out what is the "correct" way of aggregating heterogeneous labour inputs. One immediately thinks of classifying workers according to their "occupations" and then the relevant price and quantity variables to enter into the index number formula are the hours worked by each occupational type along with the corresponding average (or marginal) wage rates. However, it proves to be extremely difficult to define homogeneous occupational classes over even moderate periods of time. Productivity researchers eventually decided to disaggregate hours worked by the demographic characteristics of the worker such as age, sex race, years of schooling and so on.

Why is it so Difficult to Measure the Total Factor Productivity of an Industry?

In order to measure the TFP growth of a firm or an aggregate of firms, it is necessary to have accurate price and quantity information on all of the outputs produced by the set of production units for the two time periods under consideration as well as accurate price and quantity information on all of the inputs utilized.

Gross Outputs

In order to measure the productivity of a firm, industry or economy, we need information on the outputs produced by the production unit for each time period in the sample along with the average price received by the production unit in each period for each of the outputs. In practice, period by period information on revenues received by the industry for a list of output categories is required along with either an output index or a price index for each output. In principle, the revenues received should not include any commodity taxes imposed on the industry's outputs, since producers in the industry do not receive these tax revenues. The above sentences sound very straightforward but many firms produce thousands of commodities so the aggregation difficulties are formidable. Moreover, many outputs in service sector industries are difficult to measure conceptually: think of the proliferation of telephone service plans and the difficulties involved in measuring insurance, gambling, banking and options trading.

Intermediate Inputs

Again, in principle, we require information on all the intermediate inputs utilised by the production unit for each time period in the sample along with the average price paid for each of the inputs. In practice, period by period information on costs paid by the industry for a list of intermediate input categories is required along with either an intermediate input quantity index or a price index for each category. In principle, the intermediate input costs paid should include any commodity taxes imposed on the intermediate inputs, since these tax costs are actually paid by producers in the industry.

The major classes of intermediate inputs at the industry level are materials, business services, and leased capital. The current input-output framework deals reasonably well in theory with the flows of materials but not with intersectoral flows of contracted labour services or rented capital equipment. The input-output system was designed long ago when the leasing of capital was not common and when firms had their own in house business services providers.

This lack of information means the current input-output accounts will have to be greatly expanded to construct reliable estimates of real value added by industry. At present, there are no surveys (to our knowledge) on the interindustry flows of business services or for the interindustry flows of leased capital. Another problem is that using present national accounts conventions, leased capital resides in the sector of ownership, which is generally the Finance sector. This leads to a large overstatement of the capital input into Finance and a corresponding underestimate of capital services into the sectors actually using the leased capital.

Labour Inputs

Using the number of employees as a measure of labour input into an industry will not usually be a very accurate measure of labour input due to the long term decline in average hours worked per full time worker and the recent increase in the use of part time workers. However, even total hours worked in an industry is not a satisfactory measure of labour input if the industry employs a mix of skilled and unskilled workers. Hours of work contributed by highly skilled workers generally contribute more to production than hours contributed by very unskilled workers. Hence, it is best to decompose aggregate labour compensation into its aggregate price and quantity components using index number theory. The practical problem faced by statistical agencies is: how should the various categories of labour be defined.

Another important problem associated with measuring real labour input is finding an appropriate allocation of the operating surplus of proprietors and the self employed into labour and capital components.

Reproducible Capital Inputs

When a firm purchases a durable capital input, it is not appropriate to allocate the entire purchase price as a cost to the initial period when the asset was purchased. It is necessary to distribute this initial purchase cost across the useful life of the asset. National income accountants recognize this and use depreciation accounts to do this distribution of the initial cost over the life of the asset. However, national income accountants are reluctant to recognize the interest tied up in the purchase of the asset as a true economic cost. Rather, they tend to regard interest as a transfer payment. Thus the user cost of an asset (which recognizes the opportunity cost of capital as a valid economic cost) is not regarded as a valid approach to valuing the services provided by a durable capital input by many national income accountants.

The treatment of capital gains on assets is even more controversial than the national accounts treatment of interest. In the national accounts, capital gains are not accepted as an intertemporal benefit of production but if resources are transferred from a period where they are less valuable to a period where they are more highly valued, then a gain has occurred; i.e., capital gains are productive according to this view.

However, the treatment of interest and capital gains poses practical problems for statistical agencies. For example, which interest rate should be used?

The distinction between depreciation (a decline in value of the asset over the accounting period) and deterioration (a decline in the physical efficiency of the asset over the accounting period) is now well understood but has still received little recognition in the latest version of the SNA.

A further complication is that our empirical information on the actual efficiency decline of assets is weak. We do not have good information on the useful lives of assets. The UK statistician assumes machinery and equipment in manufacturing lasts on average 26 years while the Japanese statistician assumes machinery and equipment in manufacturing lasts on average 11 years.

A final set of problems associated with the construction of user costs is the treatment of business income taxes: should we assume firms are clever and can work out their rather complex tax-adjusted user costs of capital or should we go to the accounting literature and allocate capital taxes in the rather unsophisticated ways that are suggested there?

Inventories

Because interest is not a cost of production in the national accounts and the depreciation rate for inventories is close to zero, most productivity studies neglect the user cost of inventories. This leads to misleading productivity statistics for industries where inventories are large relative to output, such as retailing and wholesaling. In particular, rates of return that are computed neglecting inventories will be too high since the opportunity cost of capital that is tied up in holding the beginning of the period stocks of inventories is neglected.

The problems involved in accounting for inventories are complicated by the way accountants and the tax authorities treat inventories. These accounting treatments of inventories are problematic in periods of high or moderate inflation.

Land

The current SNA has no role for land as a factor of production, perhaps because it is thought that the quantity of land in use remains roughly constant across time and hence it can be treated as a fixed, unchanging factor in the analysis of production. However, the quantity of land in use by any particular firm or industry does change over time. Moreover, the price of land can change dramatically over time and thus the user cost of land will also change over time and this changing user cost will, in general, affect correctly measured productivity.

Land ties up capital just like inventories (both are zero depreciation assets). Hence, when computing ex post rates of return earned by a production unit, it is important to account for the opportunity cost of capital tied up in land. Neglect of this factor can lead to biased rates of return on financial capital employed. Thus, industry rates of return and TFP estimates will not be accurate for sectors like agriculture which are land intensive.

Resources

Examples of resource inputs include depletion of fishing stocks, forests, mines and oil wells and improvement of air, land or water environmental quality (these are resource "outputs" if improvements have taken place and are resource "inputs" if degradation has occurred).

The correct prices for resource depletion inputs are the gross rents (including resource taxes) that these factors of production earn. Resource rents are usually not linked up with the depletion of resource stocks in the national accounts although some countries, including the U.S. and Canada, are developing statistics for forest, mining and oil depletion.

The pricing of environmental inputs or outputs is much more difficult. From the viewpoint of traditional productivity analysis based on shifts in the production function, the 'correct' environmental quality prices are marginal rates of transformation while, from a consumer welfare point of view, the 'correct' prices are marginal rates of substitution.

The above seven major classes of inputs and outputs represent a minimal classification scheme for organizing information to measure TFP at the sectoral level. Unfortunately, no country has yet been able to provide satisfactory price and quantity information on all seven of these classes. To fill in the data gaps, it would be necessary for governments to expand the budget of the relevant statistical agencies considerably. This is one area of government expenditure that cannot be readily filled by the private sector.

There are also additional types of capital that should be distinguished in a more complete classification of commodity flows and stocks. In the following subsections, we will comment on some of the measurement problems associated with these more esoteric kinds of capital.

Working Capital, Money and other Financial Instruments

Firms hold money and other forms of working capital so since there is an opportunity cost associated with holding stocks of these assets over an accounting period, these assets must provide useful services in the production process. In theory, the demand for working capital and other financial assets could be modeled in the same way that the demand for physical inventories is modeled. However, the firm's demand for money is complicated by the fact that the need for money is somewhat dependent on the price level (and changes in the price level). It turns out that both in the consumer and producer theory contexts, it is not a trivial matter to derive the "right" price deflator for monetary balances.

Increasingly, nonfinancial firms hold an array of "regular" financial instruments such as stocks, bonds, insurance policies and mortgages but also of "esoteric" financial instruments such as futures contracts, currency and commodity options and other contracts that manage risks. Obviously, the demand for these commodities that involve risk in an essential way is not easy to model. Although there is a huge theoretical literature on this topic, no clear direction seems to have been provided to statistical agencies on how to calculate appropriate prices and quantities for these risky financial instruments.

Knowledge Capital

In view of the recent stock market boom involving firms that provide knowledge intensive or high tech products, it is important to be able to define a firm's stock of knowledge capital. However, it is difficult to define what we mean by *knowledge capital* and the related concept of *innovation*. We attempt to define these concepts in the context of production theory.

We think in terms of a local market area. In this area, there is a list of establishments or production units. Each establishment produces outputs and uses inputs during each period that it exists. *Establishment knowledge* at a given time is the set of input and output combinations that a local establishment could produce during that given time period t. It is the economist's period t production function or period t production possibilities set. *Establishment innovation* is the set of *new* input-output combinations that an establishment in the local market area could produce in the current period compared to the previous period; i.e., it is the growth in establishment knowledge or the increase in the size of the current period production possibilities set compared to the previous period's set. Since the statistical agency cannot know exactly what a given establishment's production possibilities is at any moment in time, it will be difficult to distinguish between *substitution* of one input for another within a given production possibilities set versus an *expansion* of the production possibilities set; i.e., it will be difficult to distinguish between substitution along a production function versus a shift in the production function.

How can we measure knowledge capital? Given the way we have defined knowledge (as time dependent, firm specific production possibility sets), it is extremely difficult to measure knowledge and changes in knowledge (innovation). Some of the possible input-output combinations that a production unit can produce are imbedded in its capital equipment and the accompanying manuals. Other possible combinations of inputs and outputs might be imbedded in its patents or the unpublished notes of the scientists that developed the patents. Yet other combinations might be imbedded in the brains of its workers. However, there are certain stocks that we can measure that will probably be positively correlated with the size of local knowledge stocks. A *science and technology statistical system* should concentrate on collecting information on these knowledge related stocks, such as stocks of patents, research and development expenditures; education and training undertaken in the firm.

Infrastructure Capital

Examples of infrastructure capital inputs are: roads; airports; harbors; water supply; electricity supply; sewage disposal; garbage disposal; telephone, and cable TV and internet hookup.

Many of the above stocks will appear in the list of reproducible capital stocks if privately owned. However, it still may be useful to distinguish the various types of infrastructure capital from ordinary structures. Publicly owned roads present special problems: they provide valuable services to business users but their price to the users is zero. Here is another example (in addition to the example of environmental prices) of demand prices being quite different from supply prices.

Before moving on to other productivity related topics, we sum up the above material on measuring inputs and outputs of a production unit. We note that most total factor productivity studies use only the information associated with output and inputs of intermediate inputs, labour, and reproducible capital. Typically, labor productivity studies use only information on output and labour input while many total factor productivity studies use only information from categories output, labour input, and reproducible capital. I believe that these productivity studies are of very limited use. A more meaningful productivity study would use information on all categories and use at least the first six categories. However, the valuation problems for the remaining categories are formidable, both from the practical and conceptual points of view.

In the following section, we note that there are some additional measurement problems that arise at the sectoral level that are due to the impossibility of calculating accurate input-output coefficients for real commodity flows across industries that add up properly.

On the Difficulty of Obtaining Accurate Real Input-Output Coefficients

All of the productivity comparisons for manufacturing industries between Canada and the U.S. rely on the information on gross output and intermediate input flows that can be obtained from the country current and constant dollar input-output tables. But there are some *additional* measurement difficulties that are associated with the use of the constant dollar (or real) industry input-output tables.

- The *same* commodity price deflator is generally used to deflate the appropriate commodity value flows for *each and every industry*. This procedure is only correct if *each* industry produces precisely the same mix of micro commodities within each of the 1000 broad commodity classes in the commodity classification and micro commodity prices are constant across industries.
- Even worse, the *same* commodity price index that is used to deflate outputs across industries is also used to deflate intermediate inputs across industries.

In my opinion, the above problems make the use of constant dollar input-output tables as a source of data for industry productivity studies a very risky undertaking. These data are bound to be filled with measurement errors. Thus there is an urgent need for statistical agencies to take a new look at the existing input-output methodology.

In addition to the above *conceptual* problems with the constant dollar input-output accounts, there are some other *practical* problems associated with the current statistical system:

- *The present input-output commodity classification has remained frozen in time since the 1950's.*
- Not only is the classification system outdated, but even using the existing classification system, *the information on interindustry value flows is very incomplete*.

The reader will now understand why all intercountry comparisons of productivity growth at the industry level should be taken with a large grain of salt. The information base upon which these comparisons are based is far from being adequate. Statistical agencies, the government and ultimately the public will have to allocate additional resources so that the limitations of the currently available data from the system of input-output accounts can be remedied.

National Productivity Measurement versus Industry Productivity Measurement

Obviously, it would be useful if we could obtain accurate information on the growth of total factor productivity at the industry level because then we could determine more precisely where the growth (or lack of growth) is originating. However, as we have seen in the previous section, the current input-output statistics that are available from national statistical agencies in all countries are far from being accurate.

But the situation is not nearly as bleak when we attempt to measure TFP growth at the national level. This is due to the fact that generally speaking, deliveries to final demand made by the aggregate production sector are in fact accurately measured and moreover, there are reasonably accurate price indexes that are constructed for the various components of final demand. Moreover, at the level of the entire market economy, intermediate inputs collapse down to just imports plus purchases of government and other nonmarket inputs. This simplification of the hugely complex web of interindustry transactions of goods and services explains why it is much easier to measure productivity at the national level than at the industry level. Also, when we measure the input of primary factors of production at the national level, we do not have to worry about errors that might have been made in classifying these inputs into an industry. Similarly, we do not have to keep track of changes in the classification of firms to industries and of sales of used assets from one industry to another. Thus measurement of total factor productivity at the national level is likely to be much more accurate than the measurement of total factor productivity at the industry level.

Conclusion

The current system of industry statistics that is used by every advanced country today has not kept up with the evolution of the world economy from primary and manufacturing production to the production of services. As a result, intercountry comparisons of total factor productivity growth at the industry level are not likely to be very accurate. Ultimately, the public will have to support additional resources being allocated to statistical agencies so that this neglect of services measurement can be addressed.

Notes

* The unabridged version of this article with full references is found at www.csls.ca under the *International Productivity Monitor*. Email: diewert@econ.ubc.ca