Methodology Choices for Benchmarking Airports

Ian Kincaid
Director, Economic Analysis, Inter*VISTAS* Consulting Inc.

Dr. Michael W. Tretheway
Executive Vice President, Marketing and Chief Economist, Inter*VISTAS* Consulting Inc.[1,2]

Draft of 11 November 2009

1. Introduction

Benchmarking is a means by which to assess the performance of a firm, industry or sector by comparing its performance with other, similar firms, industries or sectors. A benefit of benchmarking is that it can be a fairly straightforward means to identify performance deficiencies or exceptional performance, without detailed and complex examination of processes. For example, it can be used to determine whether a firm has achieved high productivity levels by comparing its productivity with that of peer firms. Without this comparison it would be a difficult and complex process to determine whether the firm was in fact productive.

Benchmarking has become an increasingly popular tool used in the management, regulation and review of both private and public organisations. Like many other industries within the transportation sector, the airport industry, its stakeholder and researchers have used benchmarking in a number of different ways to assess and improve its performance. Examples includes annual studies by the Transport Research Laboratory (TRL)[3] and the Air Transport Research Society (ATRS)[4], benchmarking considered for possible use in the regulation of Aer Rianta,[5] and benchmarking studies commissioned by the management of individual airports.[6]

In this paper we examine the purposes for which benchmarking has been used (and purposes for which it could be used) and the impact this has on benchmarking approach used. We also consider many of the issues associated with performance benchmarking. To illustrate this discussion, our paper draws on examples from the airport industry, from other transportation modes, and from other sectors of the economy.

---

[1] Dr. Tretheway is also Adjunct Professor in the Sauder School of Business at the University of British Columbia.

[2] Correspondence should be directed to Mike.Tretheway@InterVISTAS.com, and Ian.Kincaid@InterVISTAS.com. Kincaid is located at InterVISTAS' Bath UK office. Tretheway is at the Vancouver Canada office.

[3] *Airport Performance Indicators* and *Review of Airport Charges*, Department of Air Transport, Transport Research Laboratory, Wokingham. This report is now published by Jacobs Consultancy.

[4] *Global Airport Performance Benchmarking Report*, Air Transport Research Society.

[5] The Irish Aviation Regulator considered, although ultimately did not adopt, the use of benchmarking as a factor in regulating airport terminal charges. See, for example, Commission Paper CP5/2001, section 3.4.2.3. Also see Irish Commission for Aviation Regulation (2001). *Report on Reasons for the Determination*.

[6] As an example, InterVISTAS Consulting undertakes benchmarking for Canada's Level II airports, with the benchmarking studies commissioned by the airports.

2. The Link Between the Use and Format of Benchmarking

This section provides an overview of the ways in which benchmarking has been used, how this affects the choice of performance measures, the level of aggregation applied and the selection of comparators or peer firms. Benchmarking has been used in a number ways:

- Assess managerial or firm performance. Benchmarking has been used as an internal management tool to assess an organisation's performance and to instigate change. Benchmarking has also been used by governments and other organisations to assess the performance of a company contracted to provide a particular service.
- Collaborative benchmarking. While similar to the item above, this approach tends to be a more collaborative process between airport operators or countries to assess their relative performance and identify areas of improvement.
- Price regulation. A notable example of this is the past use of benchmarking in the regulation of Aer Rianta in Ireland.
- National policy. Benchmarking has been used to inform policy decisions and to subsequently assess the impact of changes in policy.
- Supply chain or value chain efficiency. Here, benchmarking would not be applied solely to a firm and its peers. Instead, benchmarking would be applied to the entire supply or value chain. For example, in the case of airports, it may be that a particular hub airport is inefficient when measured against its peers, perhaps due to a higher level of infrastructure investment, but enables much greater efficiency in other supply/value chain members. A congested airport will often be assessed as efficient among its peers (less investment for a given number of movements) but at the cost of increasing airline operating costs and reducing airline capital productivity. Value chain benchmarking would attempt to sort out the contribution of airport investment on value chain efficiency.

Each of these approaches are described in more detail below along with their implications in terms of methodology and effectiveness.

2.1 Assess managerial or firm performance

Benchmarking has become an established management tool, used to identify best practices and to challenge established thinking and open organisations to new methods and ideas. Early benchmarking focussed on manufacturing processes. (Xerox Corporation is widely credited with initiating benchmarking in 1970s, which it used to improve its manufacturing and distribution processes by benchmarking against its major, generally Japanese, competitors.[7]) Benchmarking has since been used to in sales, marketing, pricing, product development, customer satisfaction, and in the public sector and non-profit organisations.

There are also instances of benchmarking being used to assess the performance of a company contracted to provide a particular service, albeit outside of the airport industry. For example, the provincial government of Ontario is using benchmarking to assess the performance of private companies contracted

[7] Another famous example of a benchmarking study is the MIT study of the automobile manufacturing (*The Machine That Changed the World*) which analysed in-depth automobile manufacturing processes at plants across the world and lead ultimately to significant changes in production processes in North America and elsewhere, by enabling managers to recognise what they had to do to achieve world class standards.

to operate correctional facilities, comparing their performance with similar facilities operated within the public sector.[8] If the private operator fails to perform at a level equal to or higher than the public sector equivalents, the company can be subject to financial penalties. To date, however, the Ontario government has yet to publish the results of this benchmarking analysis, so it is unclear how the benchmarking has progressed and what impact it has had.

As many airports have transformed from public utilities to commercial enterprises, there has been an increased interest in utilising benchmarking to improve performance. Benchmarking has been used by airport managers to examine a number of different aspects of the airport business:

- Pricing: benchmarking landing fees and other charges.
- Service quality: customer satisfaction levels, average queue times, incidences of delays.
- Cost: unit cost, such as operating or total cost per Work Load Unit (WLU).
- Productivity or efficiency: Total Factor Productivity (TFP), Variable Factor Productivity (VFP), or single factor productivity measures (e.g. WLU per labour hour). Arguably, TFP (or similar measures such as VFP) should become a key indicator by which airport performance is measured, as it provides a fairly comprehensive picture of the overall efficiency of the airport.

The comparator airports will generally be made of natural competitors to the airport or "best in class" airports whose performance the airport management may wish to emulate. For example, Heathrow Airport may wish to benchmark itself against Frankfurt, Paris, Amsterdam (competitors for European hub traffic) or Hong Kong, Singapore, Dubai (best in class). An airport catering to largely Low Cost Carrier airlines may wish to benchmark itself against airports serving similar airlines, rather than a major international hub.

While most airport benchmarking has focussed on overall airport performance, it is possible for airports to benchmark specific services. For example, an airport could benchmark the unit cost or productivity of its ground handling service or cargo facilities against those provided at other airports. Airports can also benchmark against non-airport businesses. An obvious example would be retail, where an airport could benchmark the price and service quality of its retail facilities against equivalents on High-Street.

As with all forms of benchmarking, the development of meaningful performance indicators that reflect managerial performance is critical. For example, the unit operating cost of airports can differ depending on the degree of outsourcing undertaken by the airport. Some airports operate a number of activities themselves, such as ground handling, car parking and retailing, which many other airports outsource. Thus, one airport's operating costs may be higher simply because it engages in more services than other airports, and so may not be a suitable reflection of managerial performance.

Approaches have been developed to overcome these kinds of problems. For example, TRL "normalizes" the data so that performance indicators are based on a uniform set of activities. Another approach is to use statistical or econometric analysis to adjust the performance indicators for outside factors, an approach known as residual benchmarking, as is done for the ATRS reports. There are significant issues associated with both of these approaches, as discussed in Section 3.

[8] The performance indicators include incidences of violent disturbances, incidences of escape, incidences of unnatural death, as well as more mundane measures related to cost efficiency and productivity.

2.2 Collaborative benchmarking

Benchmarking started as a process conducted by individual firms to improve their processes and competitive position. Since then, benchmarking has also been being carried out in a collaborative manner by groups of firms or organizations.

Benchmarking customer satisfaction

An example of collaborative benchmarking is the former IATA Global Airport Monitor, now rebranded as AETRA and conducted in partnership with ACI,[9] which benchmarks (passenger) customer satisfaction on a quarterly basis. This benchmarking study originated with stakeholders (airlines), rather than with the airports themselves, although the initiative is now one supervised by the entities being benchmarked. Airports participate in the study on a voluntary basis; currently around 66 airports participate. In addition to assessing relative performance, the results of this benchmarking are also used for promotional purposes: *Number 1 Airport in the World / North America / Europe*, etc.

Similarly, Inter*VISTAS* Consulting Inc. conducts annual customer satisfaction benchmarking for a number of Canadian Airports Council small airports. The study benchmarks overall satisfaction with the airports as well as satisfaction with specific services – baggage delivery, ground transportation, retail, etc. The study also benchmarks other metrics such as average passenger spend rates and queue times.

Collaborative Benchmarking of European Air Navigation

Another example of collaborative benchmarking is that conducted by the Performance Review Commission (PRC) of EUROCONTROL, the European Organisation for Safety of Air Navigation.[10] Since 1997, the PRC has used air navigation data from all the member states to produce annual reports benchmarking the following performance indicators related to:

- Capacity and Delays: e.g., Delay Minutes per Flight, total cost of delays, average delay duration;
- Cost Effectiveness: e.g., Cost per Km Flown Enroute, Cost per Composite Flight-hour.[11], Flight-hours per ATCO hour (Air Traffic Control Officer).[12]
- Safety: the PRC has also attempted to benchmark safety but with limited success as there has been no agreement among member states on a set of key indicators for safety

The PRC has struggled to make progress in using the data to determine the reasons for performance differences between member states. Many of the reports produced include "health warnings" regarding

---

[9] According to ACI and IATA, AETRA is taken from the Latin word aethra, meaning the upper air or clear sky and is not an acronym.

[10] EUROCONTROL is a civil and military organisation established in 1963 to facilitate a safe, seamless pan-European Air Traffic Management (ATM) system. While the initial focus of the organisation was on safety and operations, its remit has expanded over time to include capacity management and development, operating costs, and fees and charges. EUROCONTROL is not an EU institution, but includes nearly all the EU members, as well as countries outside of the EU such as Switzerland, Turkey and Norway.

[11] *Enroute* refers to the high altitude, cruising part of the flight as opposed to the decent/landing and take-off/climb parts of the flight. The *composite flight hour* is a weighted average of the time spent enroute and the time spent in decent/landing and take-off/climb.

[12] The PRC has examined the possibility of developing a TFP indicator, but concluded that insufficient data was available to develop such an indicator.

comparison between performance of member states, as many external factors can impact on performance (weather, traffic mix, economies of scale, etc). This has severely limited the stated aims of the benchmarking exercise, which are to identify areas of best practice to be emulated and provide guidance on economic regulation of air navigation. The PRC has carried out some econometric analysis to adjust for these external factors (i.e., forms of residual benchmarking) but, to date, the analysis has been largely experimental and the PRC has stated that more rigorous analysis needs to be undertaken before any conclusions can be drawn.

2.3 Price regulation

Performance benchmarking is of interest to regulators as it potentially enables them to overcome some of the problems associated with information asymmetry. The regulator may have difficulty in obtaining adequate information on the regulated firm's operations and costs in order to determine the most economically efficient price cap. One approach by which to overcome this information asymmetry is to compare the regulated firm against other similar firms. In doing so, the regulator no longer needs detailed cost information to determine whether a company is efficient; instead, the regulator simply benchmarks the firm's unit cost or TFP against similar peer companies.

Regulation of Aer Rianta

A prime example of the use of benchmarking to determine a price cap is that by the Irish regulator, the Commission for Aviation Regulation.[13] The Commission was established in 2001 to regulate airport charges and aviation terminal service charges at three airports owned by Aer Rianta – Dublin, Cork and Shannon.[14] The price regulation used the common RPI –X formulation to determine the price cap applied to the airport. The Commission investigated using benchmarking to the guide the determination of the X value. The benchmarking study was commissioned in 2001, and compared the operational efficiency of the three Irish airports with comparator airports elsewhere in the world.

A total of 11 performance indicators were estimated related to cost efficiency (e.g., operating cost per WLU), revenue effectiveness (operating revenue per WLU) and service efficiency (WLU per employee). However, the primary indicator used by the Commission in formulating the price cap was *operating cost per WLU*.[15] The three Irish airports were each compared against a small selection of "peer" airports that handled similar volumes of passengers and that had similar operating requirements. In addition, Dublin Airport and the Aer Rianta group as a whole were benchmarked against a "best in class" group of European airports generally larger in scope than Dublin and recognised as innovators in their field. For example, the "best in class" comparator group for Dublin Airport included Brussels, Copenhagen and Stansted.

The benchmarking analysis found that while Dublin's operating cost per WLU was in line with its peer group, it was considerably higher than its "best in class" comparators. The Commissioned determined that

[13] Benchmarking has also been used in the price regulation of non-transportation sectors. For example, benchmarking has been used by regulators of the electricity sector in the UK, Ontario (Canada), Chile, Sweden and elsewhere. Regulators of the rail, water and telecoms industry in the UK have also used benchmarking to inform their economic regulation.

[14] In 2004, the Irish government split Aer Rianta into three separate airport authorities. In addition, the legislation narrowed the role of the Commission to regulate airport charges at Dublin Airport only (Cork and Shannon are no longer subject to price regulation).

[15] As discussed in Annex A, if estimated correctly, this measure is generally equivalent to VFP.

Dublin Airport should be able to considerably reduce this efficiency gap and proposed setting an X that would achieve such as result. A similar approach was used to determine the X for Shannon and Aer Rianta as a whole. Cork was found to be more efficient than its peers and so had no specific price cap applied.

The approach used to benchmark the Aer Rianta airports came under criticism, particularly from Aer Rianta itself. The major complaint was that the benchmarking study did not compare like with like. As mentioned previously, the costs of the airports can differ depending on the degree of outsourcing undertaken by the airports. In addition, other external factors may have distorted the assessment of cost efficiency. For example, one airport may face higher labour costs as it operates in a higher labour cost country. Equally, service quality, traffic mix, traffic volume and traffic "peakiness" can also impact on costs. The prime indicator used, operating cost per WLU, was also criticised as a) it is a partial productivity measure that does not consider investment, and b) the use of WLU as a measure of output may be to narrow and not capture the full range and quality of outputs provided by the airports.

Also, while the Commission did review the price cap approximately once a year to assess compliance and to adjust the cap for inflation and other factors, no additional benchmarking has been undertaken since the initial study. It can be argued that, as a result, the Commission may not have a correct picture of the operational efficiency improvements achieved by the regulated airports. While an airport may indeed be achieving its efficiency targets, this may simply be because of technological and other factors which are benefiting the industry as a whole, rather than specific actions taken by the airport itself. Arguably, ongoing benchmarking would be required to truly assess the efficiency improvements achieved by the airport. To date though, the Commission has not indicated that it will conduct any further benchmarking analysis.

The use of benchmarking in the price regulation of airports remains controversial. As part of its quinquennial review of its regulated airports in 2000-02, the UK CAA explored the possibility of using benchmarking as part of the formulation of the price cap. After consulting with industry, and undertaking a "test" benchmarking review of Manchester Airport, the CAA concluded that issues associated with data quality and the methodologies for adjusting the data precluded the use of benchmarking in the near future. However, the CAA takes the view that benchmarking may be able to play a part in setting price caps in the future, once some of these issues have been resolved.

2.4 National policy

The use of benchmarking in a policy context has considerable attraction. In the same way that firms can learn from the activities of other firms, governments can enhance policy by examining the relative performance of other countries. Benchmarking can be used to determine what sectors of the economies are lagging behind international competitors and what sort of policy changes have been effective in improving sectoral performance.[16] For example, have countries that have privatised their airports experienced greater improvements in efficiency, investment, pricing, etc in that industry than countries that kept airports within the public sector? Benchmarking can also be used to assess the whether a policy change has been effective by comparing before and after performance with that of other countries.

The performance indicators depend on the exact nature of the policy question being addressed, but typically can include indicators related to investment, pricing, cost, productivity, service quality and user

[16] There have also been studies benchmarking the overall economic performance of countries, such as the World Economic Forum's *Global Competitiveness Report*. In 2005/06, Finland ranked number 1, followed by the USA, Sweden and Denmark.

take-up (e.g., traffic levels). Preferably, the performance indicators should be national averages or aggregates. However, in some sectors of the economy (including airports), there is a severe lack of reliable, aggregate data (this issue, as it relates to the airport industry, is discussed in more detail in Section 3). As a result, the comparison sometimes has been conducted at the firm level.

The choice of comparators also depends on the policy question. If the requirement is to determine where improvements can be made, the comparators may be countries that are viewed as "best in class", or which have enacted radical or interesting policy changes. If the benchmarking is to determine the effectiveness of a policy change, the comparators may be a control group of similar countries that have not enacted any policy change.[17]

Using Benchmarking to Inform Policy

An example of benchmarking being used to inform national policy is international benchmarking conducted by the Productivity Commission of Australia.[18] As part of a program of major reforms of the Australian economy in the 1990s (the National Competition Policy), the commonwealth government commissioned a international benchmarking review, which was published in 1995. This review covered a broad range of major infrastructure service industries – electricity, telecommunications, rail freight, road freight, ports (referred to as waterfront), aviation (airports, air navigation and airlines), gas supply and coastal shipping. The benchmarking focussed on a small number of core indicators:

- prices (revenue per unit of output, or a weighted average of prices);
- service quality (service interruptions, on-time delivery, etc., depending on the sector);
- labour productivity (a partial productivity measure);
- capital productivity (a partial productivity measure).

The analysis was carried out at the individual infrastructure provider level, rather than country-wide or state-wide averages. For example, the performance of Sydney Airport was compared against Melbourne, Cairns, Tokyo, Auckland, Hong Kong, London Gatwick, etc. The range of countries selected for comparison varied depending on the industry examined but included countries in Asia, North America and Europe.

The study highlighted the performance gap between the best Australia infrastructure providers and the best in the rest of the world, as well as the gap between best and worst Australian infrastructure providers. An example of the gap analysis, taken from the final study report, is provided in Figure 1. The chart indicates that, with the exception of waterfront coal, the best (i.e., lowest priced) infrastructure in Australia was found to be higher priced than the best foreign counterpart.[19] The findings of the report were deemed to support the argument for the implementation of the National Competition Policy (which was already underway, in any case).

The Productivity Commission's view is that the impact of the benchmarking on policy was largely indirect. Its chief value has been in informing the policy debate and creating greater awareness of Australia's

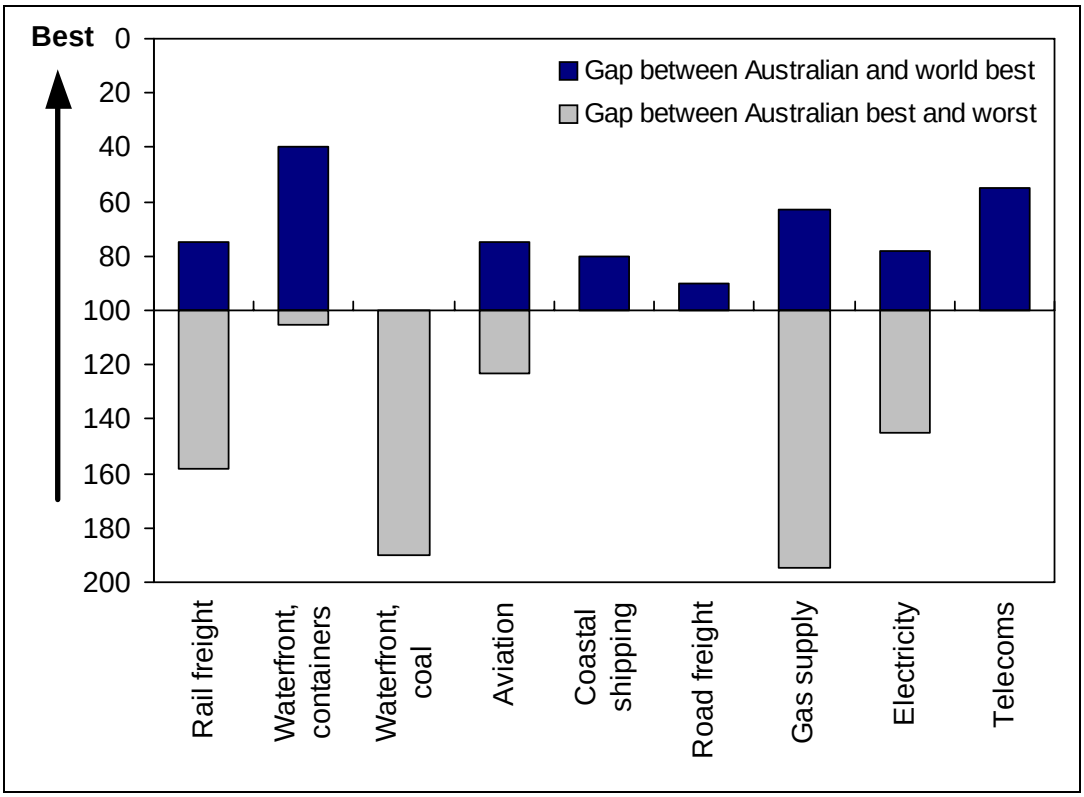
[17] In this case, it is also useful to have data collected over a considerable period of time, to allow before and after comparisons.

[18] Some of this benchmarking was conducted by predecessors to the Productivity Commission, such as the Bureau of Industry Economics.

[19] There also existed a considerable performance range within Australia itself. For example, the worst (i.e., most expensive) gas supply provider charged nearly double the best (i.e., cheapest provider).

relative economic performance.[20] One reason for this is that benchmarking rarely produces clear-cut evidence on the comparative performance of Australian infrastructure. The Productivity Commission views that it is generally hard to determine whether differences in performance are due to internal practices or external factors outside the control of the industry.

Figure 1: Sample Presentation of Benchmarking Gap Analysis
Price Performance Gaps – Australia and Best Observed
(Index relative to Australian best = 100)



Source: Reproduced from *International Benchmarking Overview 1995*, Bureau of Industry Economics, Report 95/20, November 1995.

Assessing the Impact of Government Policy

Two policy changes that have come under considerable scrutiny over the years has been the deregulation of the rail sector in Canada and the U.S. and the deregulation U.S. airline industry. Benchmarking has occasional been used in this analysis as it provides a means, albeit imperfect, to control for other factors not related to policy (e.g., fuel prices, recession, technological change).

In Canada, the rail industry was partially deregulated in 1967, in an attempt by the government to revive a heavily loss-making industry and enable it to compete with other modes, primarily trucking.[21] Conveniently

[20] One exception to this was the container ports, which underwent radical government reforms after the benchmarking showed that they were significantly more expensive, less productive and poorer quality than international counterparts. Follow-up benchmarking in 2003 found that the ports had significantly improved their performance relative to international comparators.

[21] Further deregulation occurred in 1987 and 1996.

for researchers and policy-makers, the rail industry in the U.S. did not undergo a similar deregulation process until the late 1970s. The rail industries in both countries are broadly similar, using the same gauge, the same type of trains and are primarily involved in the transportation of goods, rather than passengers (one key difference is that, in terms of track and volumes carried, the U.S. rail industry is several times larger than its Canadian counterpart). This situation provided a prime opportunity to assess the impact of policy by comparing the performance of the industries in the two countries. A number of studies compared the productivity (TFP) of the railways in the two countries, and found that the Canadian railways saw very distinct improvements in productivity following the 1967 reforms, relative to the U.S. railways.[22] Furthermore, the Canadian railways were found to have achieved a higher *level* of productivity despite the natural conditions favouring the U.S. railways (economies of scale, milder weather, greater population density). The results of these studies had considerable impact on the debate on rail policy in both countries, leading the way for further reforms and deregulation. The analysis benefited from the availability of reliable, detailed and largely compatible data on the rail sector in both countries.

There have also been studies that have benchmarked productivity and other measures of the U.S. airline industry to assess the impact of deregulation in 1978.[23] One challenge with this approach has been the quality of data available on non-U.S. airlines. While the U.S. government maintains exceptional detailed and comprehensive data on the U.S. airline industry, data from many other parts of the world is far more limited and often incompatible. Nevertheless, its has generally be found that there was some improvement in productivity at U.S. airlines following deregulation, relative to international airlines.

2.5 Supply chain or value chain efficiency

Airports are only one part of the commercial aviation supply chain or value chain. They account for only 4-8% of total cost of commercial airline services.[24] Just as air carriers may be willing to trade off higher capital costs to reduce fuel, labour or other costs,[25] air carriers may also trade off the "airport factor-of-production" for efficiencies in other factors.

---

[22] For example, see D.W. Caves, L.R. Christensen, J.A. Swanson and M.W. Tretheway (1982), "Economic Performance of U.S. and Canadian Railroads: The significance of ownership and the regulatory environment", in W.T. Stanbury and F. Thompson, eds., Managing Public Enterprise, Praeger. Also see K.D. Freeman, T.H. Oum, M.W. Tretheway and W.G. Waters II (1987), The Growth and Performance of the Canadian Transcontinental Railways: 1956-1981, Centre for Transport Studies, University of British Columbia.

[23] See D.W. Caves, L.R. Christensen, M.W. Tretheway and R. J. Windle (1987), "An Assessment of the Efficiency Effects of U.S. Airline Deregulation via an International Comparison", in E. E. Baily, ed., Public Regulation: New Perspectives on Institutions and Policies, Cambridge: The MIT Press, 1987, pp. 285-320.

[24] The range is due to a) different levels of airport charges in different jurisdictions (e.g., U.S. airport costs are subsidized and thus are quite low relative to other jurisdictions), and b) the wide fluctuations in the cost of fuel which affects the share of all non-fuel costs in the total airline cost structure.

[25] An air carrier that invests in new aircraft which are more fuel efficient would be judged to have suffered a decline in capital productivity, even though total factor productivity may have been improved. This can work in the other direction as well. An air carrier may have 'best in class' variable factor productivity by having overinvested in aircraft and other airline capital, but be below its peers in total factor productivity. Variable factor productivity measures are dangerous if their evaluation ignores the level of capital. Neoclassical economics allows estimation of variable cost functions (equivalently production functions for variable factors or variable factor productivity functions) ***but only if the level of capital is included as a regressor***. A variable cost function or a variable factor productivity function which omits the level of capital stock, is not consistent with the economic theory of production. Measuring airport variable factor productivity is not a "solution" to the lack of a measure of airport capital. Economic theory requires a measure of airport capital, whether estimating a total cost function or a variable

This is not a trivial matter. The operational performance and congestion of an airport has dramatic implications for airline productivity. The capacity and efficiency of an airport has a direct and profound impact on airline operating costs and capital productivity. A congested airport raises costs and requires longer aircraft block times.[26] The latter means that a given aircraft will only be able to operate a reduced number of flights per day. While this may seem insignificant for a single flight, the requirement for an additional 15 minutes for each flight at a carrier's hub will mean a) an average loss of roughly one hour productive service each day for each aircraft, b) a decrease in airline capital productivity of roughly 10%, and c) the need for a fleet that is perhaps 10% larger than would be the case if the airport were uncongested. Congested airports also mean a greater number of travelers will miss connections. This raises airline costs due to additional ground and flight staff time and interrupted trip expenses.[27]

What is often overlooked is that airport investments often have dramatically lowered *airline* operating and capital costs. However, these cost reductions appear on airline accounting ledgers as a) lower labour costs, and b) higher aircraft productivity resulting in lower aircraft ownership costs. As an example, if an airport undertakes a capital investment program that reduces average block times by 15 minutes,[28] then a hub carrier operating 1000 flights per day through the hub will generate operating cost savings of roughly $350 million *per annum*. The higher aircraft productivity will save over 90,000 block hours annually, equivalent to the annual flying time of roughly 40 aircraft. At an average aircraft price tag of $100 million, this is a $4 billion reduction in airline balance sheet needs.

Benchmarking only an airport relative to its peers may find that the airports which make the greatest contribution to the commercial aviation value chain are assessed to be the least efficient and highest cost airports. Such benchmarking would be a misleading indicator of airport performance. Benchmarking studies which only measure airport "outputs" as the number of aircraft movements, number of passengers or a similar measure of workload units are especially vulnerable to misleading findings on airport performance relative to peers. Given the importance of airport congestion as a driver of airline labour, capital and other expenses, the benchmarking measures of airport output must also include a measure of congestion or delay.

Including a congestion or delay measure for airport benchmarking is a challenge, of course, as airports typically do not include such measures in their annual reports. Many airports do provide such measures in their master plans and forecast documents, but these are more difficult for researchers to obtain. A number of jurisdictions collect and publish data for on-time flight performance.[29] The raw data underlying these measures can be processed to construct measures of delay at airports. Alternatively, airlines or airports can be surveyed to seek quantitative or qualitative measures of delay. An example might be a survey

---

cost function, or estimating a total factor productivity regression vs. a variable factor productivity regression. The lack of a measure of capital does not justify estimation of relationships which are inconsistent with economic theory.

[26] Block times are the time from when an aircraft departs the gate at the origin airport to the time when the aircraft arrives at the destination airport gate. Block times include not only the normal flight time between origin and destination, but also the additional time aloft due to queues on approach (e.g., circling or reduced en route flight speeds) and ground taxi/delay times.

[27] Meal vouchers, hotel accommodations, etc. for interrupted travelers.

[28] This might be due to additional runway capacity, more efficient de-icing operations, greater number of gates reducing wait times for gates and ramp loading/unloading of passengers, etc.

[29] E.g., the U.S. and Australia.

question which asked whether the number of flights which experience congestion or weather delay fall into one of five performance bands.[30,31]

A survey to carriers using the airport is an alternative means to assess the impact of the airport on the commercial aviation value chain. Such a survey should be targeted at airline dispatchers, not airport stations managers (who are often focused on customers service and real estate issues).

The key point is that airport benchmarking which ignores the impact of airport investment and operational efficiency, can produce a misleading assessment of the airport's contribution to economic efficiency. Of particular importance is that airport benchmarks which are based only on measures of airport variable inputs (with no control for airport capital) are biased and should be discouraged.

2.6 Summary of the Uses of Benchmarking

Table 1 provides a summary of the broad uses for benchmarking and the type of measures and analysis undertaken in each case.

[30] While weather is beyond the control of most airport management, airports can and do invest in landing systems (e.g., category IIIa) which greatly reduce the number of flights which are delayed due to low visibility weather, in cross wind runways which enable operations when the prevailing wind is not prevailing, etc.

[31] As an example a question could be "Indicate the number of days when aircraft are delayed an average 5 or more minutes due to congestion – runway or de-icing: a) less than 5 days per year, b) 5-10 days per year, … Indicate the number of days when aircraft are delayed an average of 15 minutes or more …"

Table 1: Use of Benchmarking

Purpose	Types of Measure	Level of Aggregation	Comparators
Assess performance	▪ Price ▪ Customer satisfaction ▪ Service quality ▪ Unit cost ▪ Efficiency (TFP, VFP, etc)	Airport or individual services	▪ Best in class ▪ Natural competitors
Collaborative benchmarking	▪ Price ▪ Customer satisfaction ▪ Service quality ▪ Unit cost ▪ Efficiency (TFP, VFP, etc.)	Airport or individual services	▪ Other group members
Price regulation	▪ Efficiency (TFP, VFP, etc.)	Airport	▪ Best in class or similar peer airports
Assess policy	▪ Price ▪ Service quality ▪ Unit cost ▪ Efficiency (TFP, VFP, etc.) ▪ Investment ▪ Throughput or take-up	National or airport	To inform policy: ▪ Best in class ▪ Competitor countries ▪ Countries that have enact major policy reform To assess policy outcomes: ▪ Control group of counties that have not enacted policy change
Measure Contribution to Value Chain Efficiency	▪ Same as above, but must also include measure of congestion or delay	Airport	▪ Other group members

3. Issues in Benchmarking Airports

This section provides a more in-depth, technical discussion of some of the issues touched on in the previous section:

- availability and quality of data;
- adjusting the data to provide meaningful comparisons;
- use of residual benchmarking, with reference to the ATRS study.

3.1 Availability and Quality of Data

One of the main challenges in benchmarking airport performance is obtaining workable data. The availability and quality of data on airport activities and finance varies considerably around the world, depending on the ownership structure of the airports, accounting practices in each country, and the data collected by government statistical agencies.

While most private and not-for-profit airports publish detailed financial accounts, the financial accounts of some publicly owned airports are not readily available in any great detail. Differing accounting practices can also create data inconsistencies. For example, some airports treat interest expenses as an operating cost, while others treat it as a non-operating expense. U.S. airports treat grants and monies received from the Passenger Facility Charge (PFC) as offsets to capital expenditures rather than revenues;[32] however, in Canada, the Airport Improvement Fee (AIF, similar to the PFC) is generally treated as a revenue source. The BAA depreciates runways over period of up to 100 years, while many other airport operators use shorter periods of between 20 and 40 years. Airports are also subject to differing tax regimes. To some extent, the data can be adjusted in order to apply a consistent accounting standard, but this can be difficult task requiring very detailed data.

The issues of data quality and consistency has made the development of more complex metrics such as TFP even more challenging. While a number of techniques exist for estimating such measures (Tornqvist index, DEA), all require detailed and accurate data on the airport inputs and outputs. However, it is possible to provide a rough approximation of TFP using real unit cost, provided that the deflation of unit cost is based input price inflation rather consumer price inflation. The general equivalence of TFP and real unit cost is discussed in Annex A.[33] While this approach is not without its challenges, it can provide a reasonable approximation where more complex approaches cannot be used.

[32] This treatment of PFC revenue results in highly distorted balance sheet measures of airport capital for U.S. airports. An alternative is to construct a perpetual inventory measure of airport capital stock using airport investment (including contributions from PFC revenues). For a description see D.W. Caves, L.R. Christensen, J.A. Swanson and M.W. Tretheway (1981), "Economic Performance of U.S. and Canadian Railroads: The significance of ownership and the regulatory environment," in F.W. Thomson and W.T. Stanbury eds, *Managing Public Enterprise*, New York, Praeger.

[33] Note that there is no equivalence of a VFP regression without capital stock as an argument and a variable cost function. The latter must have the level of capital stock as an argument, while VFP is devoid of any measure of capital. VFP regressions with a measure of capital stock can have an equivalence to a variable cost function, although with a highly restrictive specification.

In general, there is a dearth of aggregate data collected by government or industry associations by airport operations and finance. Generally data on passenger, cargo and aircraft traffic does exist,[34] but more detailed information on airport operations and finances is not available in national statistics. This is not the case in other transportation sectors. The U.S. government collects very detailed data on the U.S. airline industry including traffic volumes, financial information, fuel consumption, employment, delays, lost luggage, passenger complaints, ticket price data.[35] Likewise, detailed information is available on the North American rail industry through government statistics and data collected by the Association of American Railroads and the Railway Association of Canada. Having this type and quality of data on the airport industry would highly valuable in benchmarking national policy on airports as well as for other uses of benchmarking.[36]

3.2 Adjusting the data to provide meaningful comparisons

As already mentioned in Section 2, a difficulty using the results of benchmarking to improve performance or develop policy is understanding the true causes for observed differences in the performance metrics. Do the differences in the performance indicators reflect managerial performance (policy performance), are they the result of data inconsistencies, or are they the result of external factors that management (or policy makers) cannot affect?

For example, benchmarking landing fees may appear to be a straightforward exercise. However, airports around the world structure their fees and charges in quite different ways. Some airports charge the airlines using just a weight-based landing fee which covers all services at the airport; other airports have a plethora of additional charges including passenger terminal fees, gate utilisation fees, fees for use of the FID and announcement systems, and other charges for specific airport services. In addition, many airports levy fees charged directly to the passengers, such as the AIF or PFC. A more meaningful comparison for an airport manager would be the total fees and charged levied on the carrier and its passengers for a flight by a typical A320, B767, A340, etc. with a load factor of X%.[37] Not only can airport management use this information to assess its airport's price competitiveness, but it can also be used to address misconceptions held by airlines and the public regarding the airport's pricing levels.[38]

Even with this adjustment there are many other issues that can affect the comparison of pricing levels. Some airports may have higher fees than others simply because they provide more services. While some airport's fees cover ground handling, fire and emergency service and cargo handling, other airports do not provide these services (these service are instead provided by third party vendors), and so have lower fees. This issue also affects benchmarking of other performance measures such as cost and productivity. However, if this measure (or any similar measure) is to be used to assess management performance, consideration needs to be given to factors outside of management control that may affect the comparison. There are a range of factors that can impact the comparison of airport performance:

---

[34] Even here there are gaps. For example, cargo data collected by the federal government in Canada is notoriously unreliable, and seriously understates cargo volumes.

[35] This data can be found at www.bts.gov (Bureau of Transportation Statistics).

[36] In our experience, there is also very limited data on the marine sector (ports, seaways, shipping).

[37] This type of approach is used by the TRL in its annual review of airport charges.

[38] For example, while one airport may appear have exceptional high landing fees, when other charges are also considered, it may be in line with or lower cost than its competitor airports.

- Degree of contracting out by the airport. Where certain services are contracted out to third parties, who charge the airlines directly, this lowers the operating costs of the airport. Also, the extent to which the airport develops non-aeronautical activities can impact costs, productivity, revenues, etc. Airport operators have at least partial control over some of these factors. However, depending on the nature of the benchmarking exercise it may be useful to control for these factors (e.g., assessing aeronautical fees).
- Weather. Airports in cold climates require snow clearing and de-icing equipment increasing their cost base (unless, of course, these services are contracted out). Likewise, airports subject to inclement weather (hurricanes, high winds) may have higher costs or reduced output. Potentially, airports in hot areas could have reduced output levels as aircraft sometimes have to operate with lower take-off weights (hot air is less dense).
- Government subsidy or assistance. Airports operating within the public sector may benefit from government subsidy for operating or capital costs. Also, these may have access to lower cost financing as the loans or bonds are backed by the government. Public sector airports (or indeed some private airports) may receive certain services from the government at no cost or at a lower cost than those airports who provide the services themselves (or who contract them from the private sector). For example, fire and emergency services and security.
- Traffic mix. International passengers tend to require more infrastructure and space than domestic passengers (e.g., customs and immigration, higher baggage loads). They also tend to generate higher revenues to the airport, through higher fees and charges to the airline and higher retail spend rates. Also, the proportion of connecting passengers, versus O/D passengers, can also have cost and revenue implications. The amount of cargo handled by the airport also impacts its cost and revenue structure.
- Capacity constraints. The capacity of the airport may be constrained by factors outside of management control. For example, the airport may be subject to night curfews, noise quotas or slot constraints. Likewise, the airport's ability to manage airport capacity may also be limited by government regulation. Some airports may be able to apply peak period pricing while others are forced to apply undifferentiated fees, impacting on productivity and revenues.
- Cost of Living. Some airports face higher labour costs as they operate in higher labour cost countries. The cost of labour for Heathrow Airport is likely to be several multiples higher than that of New Delhi Airport.
- Service Quality. Productivity gains or cost reductions may be achieved by lowering the level of service at the airport. A comparison based purely on financial measures may miss this aspect of performance.
- Economies of Scale. Airport productivity can potential be enhanced by economies of scale. Airports with higher traffic volumes have higher productivity levels than smaller airports simply because it benefits from economies of scale. Arguably, this is a factor outside of management control.[39]
- Congestion. Unless service quality is controlled for, airports which are congested may appear to have very high productivity (e.g., movements per runway), when in fact they reduce the economic efficiency of commercial aviation.

[39] It can be argued that management can pursue strategies to increase traffic volumes so that the airport can benefit from economies of scale. However, there are also historic, political and geographic factors that can affect traffic volumes.

One approach to account for some of these issues is to normalise the data, so that the benchmarking is based upon a uniform set of activities. This is the approached used by TRL in its annual *Airport Performance Indicators* report. The data is adjusted to reflect the costs, revenues and employment associated with a core set of activities. These activities include: provision of runways, taxiways and aprons; provision and operation of the terminals; provision of retail and food & beverage within the terminal. Examples of non-core activities include baggage handling, car parking, air traffic control, other non-aeronautical activities such as airport hotels and leasing of airport land. The data for each airport is adjusted to strip out costs, revenues, employment associated with non-core activities.[40]

This approach requires detailed data to carry out these adjustments as well as certain amount of judgement to determine costs and revenues in a hypothetical situation. Furthermore, the approach is not very effective in adjusting for factors such as traffic mix or economies of scale. To some extent, these factors can be adjusted for by benchmarking airports of a similar size and/or traffic mix. This, though, limits the pool of peer airports that can be compared, and may exclude airports of interest (e.g., best in class). Another approach used to address these issues is residual benchmarking, which is discussed in the next section.

It is worth noting that adjusting for these factors may not always be necessary or appropriate. For example, exploiting economies of scale may well be a policy objective. Differences in financing costs between public and privates are relevant to an analysis of the benefits of airport privatisation. The degree of outsourcing or non-aeronautical revenues development may well be relevant to an assessment of managerial performance.

3.3 Use of residual benchmarking, with reference to the ATRS study

Residual benchmarking involves undertaking econometric or statistical analysis of performance indicators to determine the extent to which certain factors explain the differences between various airports. Generally, the raw or gross performance measure is regressed against a number of variables related to factors that may be of interest (e.g., traffic volumes for economies of scale, traffic mix, measures of service quality). The approach has two benefits:

- It enables greater understanding of what factors may be responsible for changes in performance.
- It allows performance measures to be adjusted for "external" factors not relevant to the study, which may distort the findings, leaving a residual measure.

This approach is most commonly applied to TFP productivity measures. The annual benchmarking report by the Air Transport Research Society (ATRS) is an example of this approach.[41] Since 2002, this annual study has reported cost efficiency and productivity measures for around 100 airports located in North America, Europe and the Asia Pacific. The report includes estimates of residual TFP (in the 2002 and 2003

[40] For example, for airports that provide ground handling and parking services, the costs, revenues and employment associated with those activities are stripped out the airport data. The revenues of these airports are further adjusted to reflect the fact that if they did not provide these services, they would contract to a third party provider who would pay the airport s concession of percentage of profits. Some adjust are made for differences in corporate income taxes, but not all differences in accounting practices have been adjusted for, due to the complexity involved.

[41] *Global Airport Performance Benchmarking Report*, Air Transport Research Society.

reports) and residual VFP (in the 2003 and 2004 reports).[42] The logarithmic regression model used to estimated residual VFP in 2004 relates the raw measure to factors such as:[43]

- Passenger traffic volumes (economies of scale);
- Percentage of international traffic (traffic mix);
- Percentage of revenue from non-aviation activities (impact of non-aeronautical activities on productivity);
- Percentage of air cargo in total traffic (impact of cargo activities);
- Capacity constraints (impact of capacity constraints on productivity).

However, there are many serious problems with the ATRS residual productivity regression analysis. First, many of the results from the residual TFP/VFP analysis conducted by the ATRS do not appear to be robust and bring in to question the validity of the findings. For example, in each of the three reports produced, the impact of airport size has varied dramatically as seen below.

ATRS estimate of impact of airport size on productivity

Year	Elasticity of Productivity with respect to airport size	Meaning
2002	0.35	an airport of double the size will have 35% higher TFP
2003	0.15	an airport of double the size will have 15% higher TFP
	0.16	an airport of double the size will have 16% higher *VFP*
2004	-0.18[44]	an airport of double the size will have 18% lower *VFP*

Source: ATRS reports for 2002, 2003, 2004

According to the ATRS results, the impact of economies of scale were reduced by half in a single year, and in 2004 it reversed sign – larger airports were now less efficient than smaller airports. Newark Airport, which was the 4th most productive North American airport in 2002 on ATRS' residual productivity measure became the second worst performing airport in 2004. The explanation given for the negative estimate in 2004 is that many of the airports (particularly in North America and Europe) are large airports that have exhausted their economies of scale. This is a curious explanation given that just a year previously this did not appear to be the case, even though the sample of airport is largely the same each year. A more likely explanation is that there are problems with the model specification or the data that require further examination. Other coefficient estimates also exhibit considerable (unexplained) variation. For example, the estimated coefficient on cargo traffic is positive in 2003 (airports with a higher proportion of cargo traffic

[42] The ATRS provide no explanation for decision to drop measurement of TFP in the 2004 report.

[43] Based on the "pooled" model using data from all airports. Separate models were estimated for North American, European and Asia Pacific airports with differing sets of explanatory variables.

[44] Result from the "pooled" model. The separate models estimated for North American, European and Asia Pacific airports produced airport size elasticities of –0.298, -0.647 and +0.06 respectively.

are more productive), but negative in 2004 (airports with a higher proportion of cargo traffic are less productive).

Clearly these results are not robust and it is troubling the ATRS reports do not appear to address this issue. This is compounded by the fact that the ATRS will not provide the raw TFP/VFP data even to those who purchase the reports, nor provide any details on the estimation of the raw TFP/VFP measures.[45]

A second serious problem with the ATRS residual analysis is that the VFP regressions are inconsistent with economic theory. As described in Section 2, neoclassical economics requires that a variable cost function have the level of capital included. If a measure of capital is excluded from the variable cost function or its estimating equation, the specification is incorrect and the results are biased and potentially meaningless. The same applies to VFP regression. A VFP regression which fails to include a measure of capital is mis-specified, inconsistent with neoclassical economics and likely biased. A justification offered by the ATRS group that produces the report is that a measure of capital is not available. Two comments should be made. First, the available data from airports would allow construction of a measure of airport capital using a perpetual inventory methodology, although it would require some work to gather the data and construct it. Second, the unavailability of data is not an excuse to estimate a relationship inconsistent with economic theory. ATRS engages in considerable efforts to attract media attention to its results and in those comments makes statements regarding which airports are the most efficient. These statements, based on a methodology inconsistent with economic principles, are potentially misleading, and seriously degrades the policy dialogue on airports.

In general, residual benchmarking measures, while useful for some aspects of assessing the performance of managers of individual airports, have serious limitations which must be considered. Developing these measures requires econometric and/or statistical analysis. Such analysis can be easily biased by incorrectly specifying the factors that influence the benchmarking measure, by using one 'functional form' rather than another, by using an incorrect statistical technique, or by using a distorted sample of firms/countries for the analysis, etc.

If the statistical analysis underlying the computation of residual benchmarking measures is not robust, then the residual performance measures can be quite different depending on the econometric model used, the sample selected, etc. This would appear to be the case with residual TFP/VFP measures estimated by the ATRS. As demonstrated above, the results of the ATRS residual benchmarking analysis are specious and should be treated with high caution or disregarded. It is strongly urged that the ATRS attempt to refine and improve the analysis before producing more residual VFP results, or discontinue this aspect of the study for the time being. As well, ATRS should make its data available for peer review.

4. Conclusions

This paper has discussed the many ways in which benchmarking has been used in the airport industry and in other parts of the economy. This included assessment of managerial performance, collaborative benchmarking, price regulation, and informing and assessing national policy. The format of benchmarking differed in each case to match the requirements of the study.

[45] The raw TFP/VFP measures were estimated using a parametric method - the ATRS estimates a production function and infers the raw TFP/VFP score from the production function. Unfortunately, the ATRS does not report any of the parameters of its production function, only the resultant TFP/VFP estimates.

The paper also examined some of the key issue associated with benchmarking, primarily difficulties in providing meaningful comparisons and the approaches developed to improve the benchmarking comparisons such as residual benchmarking.

4.1 Is benchmarking with limitations better than no benchmarking?

While benchmarking has many issues and limitations associated with it, it is certainly not without value. It is a useful tool to identify deficiencies and excellence in performance. It can spur competitive forces and shake up conventional thinking (referred to as *paradigm blindness* by some benchmarking practitioners).

The answer to whether benchmarking with limitations is better than no benchmarking also depends on how the benchmarking is used. If benchmarking is being used to calculate the price cap applied to an airport, then the limitations of benchmarking can have major implications. Even minor errors in the benchmarking analysis that feeds into the price cap could result in a price cap costing the airport millions of dollars in unnecessarily forgone revenues (or conversely, allowing the airport operator to collect excessive revenues). On the other hand, if the benchmarking is being used to assess the impact of policy reforms then a broad, but imprecise measure may be sufficient. The benchmarking in this context would be able to identify the directionality and broad scale of impact, even if the level of precision is limited.

Benchmarking can serve as an effective decision-aid tool, but decision makers must be aware of the limitation of the analysis, and the analysis itself must demonstrate sufficient robustness.

Annex A: The General Equivalence of Real Unit Cost and TFP

There is a general equivalence of TFP and inflation adjusted unit cost. TFP is the ratio of aggregate output to aggregate input:

$$\text{TFP} = \frac{\text{Aggregate output quantity index}}{\text{Aggregate input quantity index}}$$

Unit cost is the ratio of total cost to aggregate output:

$$\text{Nominal unit cost} = \frac{\text{Nominal total cost}}{\text{Aggregate output quantity index}}.$$

Unit cost is the sum of the costs of individual factors of production (labour, capital, energy and materials). From the quantities of the individual factors of production, the aggregate input index is constructed. Dividing total cost by the aggregate output index produces an input price index (specifically, the 'dual' input price index). Another way of expressing this is:

$$\text{Nominal total cost} = \text{Aggregate input quantity index} * \text{Input price index}$$

Combining the last two equations yields:

$$\text{Nominal unit cost} = \frac{\text{Aggregate input quantity index}}{\text{Aggregate output quantity index}} * \text{Input price index}$$

or:

$$\text{Nominal unit cost} = (1 / \text{TFP}) * \text{Input price index}$$

or:

$$\text{TFP} = \text{Nominal unit cost} / \text{Input price index}$$

$$= \text{Real unit cost.}$$

That is, a TFP index is equal to an inflation adjusted unit cost index.[46]

[46] There is also an equivalence of TFP regressions and cost function regressions. This is proved in K.D. Freeman, T.H. Oum, M.W. Tretheway and W.G. Waters II (1987), *The Growth and Performance of the Canadian Transcontinental Railways: 1956-81,* see p. 203ff.

While there is conceptual equivalence of direct measures of TFP and real unit costs, there are some subtleties to the equivalency of the two. If these subtleties are not treated correctly, a computation that seems to be real unit cost may be a biased measure of TFP.

First, to get real unit cost, nominal unit cost is not divided by a consumer price index, but rather by an index of input prices. A few industries publish an input price index: for example the U.S. rail industry publishes the Rail Cost Adjustment Factor which is an input price index. The U.S. airline industry (not the airport industry) publishes an index of increases in the prices of airline inputs. However, most industries and nations do not publish industry specific indices of input prices.

In the absence of an industry specific input price index, a producer's price index may be used as a proxy – at the very least it is superior to using a CPI. Another proxy would be an index of wages, as typically it moves closer to producer price indices than to consumer price indices.

Second, to use real unit cost as a TFP equivalency it must be measured based on total cost, specifically an economist's definition of total cost which includes the annual return on equity capital. Operating cost is not a sufficient measure of total cost for the equivalency.[47] Total accounting cost would be better than operating costs, although it should be augmented by a normal return on equity capital.

Third, in some sectors, data is so limited that neither total cost or operating cost data is available. This is especially the case when infrastructure services are provided by government departments, rather than by a corporation with its own accounting books. Highway and marine/air navigation services are often provided via government departments, with their costs imbedded in overall departmental budgets.

[47] Note that real unit operating cost, if it excludes any capital costs, is roughly equivalent to Variable Factor Productivity (VFP).