The Case of the Missing Data:  
Implications for Productivity Measurement

Harry Postner  
(formerly Economic Council of Canada)

Session 4B “Service Sector Productivity Measurement II”

April 12  8:30 - 10:00 AM
THE CASE OF THE MISSING DATA:
IMPLICATIONS FOR PRODUCTIVITY MEASUREMENT

by

HARRY H. POSTNER
(Formerly Economic Council of Canada)

Paper prepared for
Conference on Services Sector Productivity and the Productivity Paradox
Organized by the Centre for the Study of Living Standards
Ottawa, Ontario, April 11-12, 1997
1. INTRODUCTION

This paper is directly oriented to the two main concerns of our Conference: (1) problems of measuring Service sector real output and productivity growth and (2) examination of the “productivity paradox” — why new information technology has evidently not raised officially measured service industries’ productivity. The particular orientation of the paper, however, has certain outstanding features that should be born in mind.

First, the analysis is primarily conceptual, but is nevertheless concerned with basic statistical problems of coverage and methodology. We try to raise issues that other papers do not tend to handle. There is a distinct forward-looking bias rather than a purely historical analysis. And the reader will find a lot of discussion about design and retrieval of computer software — what else is the essence of the “information revolution”?

To attract a wide audience the exposition is kept brief and essentially nontechnical, though some prior acquaintance with the issues is assumed. It will be clear, therefore, that some of the problems do require a more detailed treatment. More details are sometimes available in this writer’s previous papers and are sometimes in a current or planned research stage. In effect, the paper reflects the writer’s work on problems of: (1) statistical units, (2) industrial classification systems, (3) accounting for costs of economic change, (4) the goods/services/information trichotomy, and (5) new forms of economic organization.

In addition, there is a distinct Canadian content. Anyone searching for Joseph Schumpeter or Paul Romer will not find them here. But the reader will find, and for very good reasons, Harold Innis (Canada’s greatest economist) and Marshall McLuhan (“the medium is the message”).

There are two pervasive Systems that influence the coverage and viewpoint of this paper. First we have a dominating position attained by the 1993 System of National Accounts [1993] with or without some amendments proposed by Carter & Postner [1996] to reflect the growing information economy. Second, we have the dominating position that will be obtained by the new North American Industry Classification System [1996] with or without some minor amendments also proposed by Carter & Postner [1996]. In any event it will be very helpful if the readers of this paper have some familiarity with the 1993 SNA and the 1996 NAICS. Nevertheless the highlights of the two Systems are briefly outlined in the course of the paper — at least those highlights that are most relevant to our concerns.

Section 6 below attempts to put it all together. There is a Summary showing some possible resolutions of the “productivity paradox”. It should be noted, of course, that these possible resolutions are purely hypothetical in nature because of the basic conceptual orientation of the paper. The Summary, however, could still provide guidelines for future empirical investigation — when all the evidence is available.
Finally, the Conclusion Section 7 points to new challenges for the national statistical agency. The old rules of the statistical game are no longer working. The time has arrived to begin re-thinking the dimensional relationships between producers and users of economic statistics. Indeed, the critical role of relationships also turns up in Section 4 on new network forms of economic organization.

As one perceptive Canadian social thinker has put it: “People don’t just want more information; people want relationships”.

2. PROBLEMS OF IDENTIFYING A SERVICE SECTOR: THE RISE OF AN INFORMATION SECTOR

Until recently there were no apparent problems of identifying what is meant by the Service sector vis-à-vis the Goods producing sector. True, there were some borderline cases but these did not really matter. It was important to identify the Service sector because the behaviour of measured productivity growth (and levels) evidently differs in the Service sector as compared to the Goods sector (each sector, say, taken as a whole).

However, once we begin to examine the information revolution and its impacts, the rules of the game change. There is no longer a simple Goods/Services dichotomy. There is now and indeed there has been for many years a Goods/Services/Information trichotomy that must be covered and taken into account. This is, of course, a basic motivation underlying the 1996 NAICS.

We should be clear that the new Information industries sector (called the Information & Cultural industries sector in Canada) is not just a split off from the original Services sector. The Information sector includes inter alia: (1) newspaper, periodical and book publishing — formerly classified to manufacturing, (2) sound recording production and music publishing — also formerly included in manufacturing, and (3) radio, television and cable programming and broadcasting — usually assigned to the Goods sector. Most important, the pervasive telecommunications industry, in all its aspects, is removed from Goods and is now completely classified to the Information sector. But this is not all!

The Information sector of the NAICS creates distinct 4-digit industries for both database publishing (producers and distributors) and software publishing (design and support) formerly hidden somewhere in computer services. There are in addition two important information industries at the 3-digit level: (1) information providers per se that include libraries, museums and archives (formerly part of education services), news syndicates (formerly a n.e.c. business service), and a long string of on-line information vendors that are either entirely new (e.g., Internet access providers) or were formerly buried somewhere in n.e.c. configurations; and (2) an equally long string of data and transaction processing operators — some newly identified (data entry & optical scanning) and some broken out of computer services n.e.c. for the first time.
But what about CD-ROM publishers? At the moment the NAICS does not provide a single class for CD-ROM — it depends on the content. These publishers, of growing importance as purchasers of copyright, are most likely to be found in either: (1) database publishing (databases), (2) software publishing (software & multimedia products), or (3) integrated record production (sound recordings).

The rationale for creating a new Information (and Cultural) industries sector is explained in NAICS [1996], now available on the Web. The System sector is further elaborated in Carter & Postner [1996] with suggested additions involving, e.g., custom software and training packages. The wide range of industry commodities embodied in the new sector share some distinct economic properties, including their inherent intellectual and creative processes often protected by copyright law, not typical of the “remaining” Goods sector and Service sector industries. There are also some peculiar economic problems such as: (1) who owns uncopyrighted information?, and (2) who pays for use of public information? — that are outside our present concerns.

It might be noted that the NAICS does not create an Internet “industry” or a World Wide Web “industry”. Such industries can indeed be broken out of the NAICS somewhat analogous to breaking out and exposing a so-called Tourist industry from the old SIC. Anyone interested in trying their luck at this task is, however, advised to first consult Postner [1994] — the theory and practice of satellite classification systems. The creation of satellite systems of classification is a non-trivial exercise and requires prior knowledge of satellite accounting systems from the SNA [1993].

What can we conclude from all this that is most relevant to the main concerns of the paper?

First, I think we should be very modest about claims made on the basis of historical estimates of productivity growth stemming from obsolete industrial classification systems. The presently available results do not provide an acceptably sharp distinction between “Goods” and “Services”. Information industries are not distinguished at all and many important information commodity items are simply missing. Second, it is hoped that the NAICS will be made fully operational as soon as possible and that historical data can be re-classified to permit effective time series analysis. Third, I do not claim that an application of NAICS will somehow cause the celebrated productivity growth slowdown to “disappear”. But it would permit a deeper analysis, particularly when we come to the essential task of examining the “productivity paradox”.

3. PROBLEMS OF MEASURING INFORMATION AND SERVICE SECTORS PRODUCTIVITY

It seems clear from the preceding analysis that we cannot avoid discussing real output and productivity for both a Service sector and an Information sector. Moreover, the Information sector is a prime producer of other industries’ intermediate inputs as well as fixed capital inputs often in intangible form. It is therefore easy to see that productivity estimates for any sector can be sensitive
to the accounting and measurement of the newly identified Information sector. The situation, in fact, is compounded by the occurrence of technology changing “too fast” (Carter [1986]).

This leads us to a selective discussion of issues that occur in both NAICS [1996], particularly Agreements numbered 13 & 18, and the SNA [1993], particularly as elaborated by Carter & Postner [1996] and further clarified by Hill [1997]. One way or another, we cannot avoid examining the measurement issues of “accounting for economic change” (Postner [1995]). However, because many (not all) of these problems have been discussed elsewhere, though not with reference to official measures of real output and productivity growth, the exposition is kept brief. Readers are referred to the original sources for further details.

First it should be clear that the reproduction and mass duplication of magnetic and optical media — CD-ROMS, software, audio, video and multimedia products — are classified to the Manufacturing sector. But the original publishing i.e., design and documentation and on-line retrieval, of software are placed in the new Information sector. This distinction is the result of direct consultation between the people behind NAICS and business sources within the relevant industries. It is motivated to facilitate the agencies’ production of statistical data — an effort that is difficult to argue with. But in order to understand statistical data we must take the trouble to learn its origins. Locating all software in a single class, as in the presently obsolete SIC, leads to misleading results that do not reflect industrial organization practices with respect to software access and creation.

These considerations might partly explain why software developments, the very essence of the new information technology, do not appear to be impacting productivity growth estimates. The present obsolete industrial classifications are not sufficiently sensitive to picking up the key roles of software as both an intermediate input and an intangible produced fixed asset — depending on the period of use and the stage of creation of the software per se. The situation is made more complex when it comes to accounting for the exercise of copyright (sale and use) of original software code. Some of these considerations also arise with respect to CD-ROM and databases. Indeed, presently available estimates of total factor productivity growth do not even have a place for intangible fixed assets — whether the estimates are on a “gross” or a “net” basis. It is true that incorporation of such assets in productivity estimates would require expected life assumptions and obsolescence patterns — but this is not an impossible task. In any event, we do have life expectancies for copyright when exercised.

Before continuing it should be observed that the above analysis is essentially based on a synthesis of the SNA [1993] and the NAICS [1996].

When we consider the Information sector, in a growth-accounting context, it is necessary to deal with real outputs and not just nominal outputs. The general issue of deflating nominal values is discussed by other papers; so is the issue of quality-adjusted real outputs. But it is also possible to obtain real output measures directly by careful analysis of the tasks embodied in a production process. Such a measure is available for the writing of original software source code — even taking account of the many different (over 400) computer languages.

This is accomplished, in effect, by ignoring what the source code says and concentrating on
what the code does. It distinguishes and measures: (1) inputs — the number of different commands
the software can accept; (2) outputs — how many types of information it can generate; (3) inquiries
— how many different sorts of questions a user can ask; (4) files — how many it can handle
simultaneously; and (5) interfaces — the number of potential links with other software. The five
categories are combined by appropriately-defined weights (see The Economist [1993]). The
combination then yields a measure of “function-points” real output, imputed to software engineering
projects, that can be aligned with projects’ total nominal costs and revenues. One key lesson here:
ignore the impossible and concentrate on the feasible. The method covers both systems and
management software and works for both cross-section comparisons and time-series analysis.

Finally, some writers have already conjectured that excessive technological change may not
be a “good thing” or at best a “mixed blessing”. The problem again is to operationalize the
conjecture with appropriate distinctions. A suggested package and associated classification of
excessive costs of economic and technological change would be: (1) start-up costs, (2) change-over
& re-tooling costs, (3) re-organization and re-structuring costs, and (4) close-down costs (see Postner
[1995]). These costs should be considered as “excessive” over and above some norm. The “norm”,
however, is not always obvious, but neither is a “slowdown”.

The costs of economic change ceteris paribus would typically manifest themselves by
depressing measured productivity growth rates over the medium term — whether the costs are
correctly capitalized and amortized or incorrectly expensed. One of the cost components, close-
down costs, is often lost or misplaced in official data collections and, indeed, may not be capable of
industrial allocation until all final historical revisions have been performed. It might be noted that
close-down costs should include the capital writeoff costs of unforeseen obsolescence as well as
environmental clean-up costs and labour buyout adjustments. A special case of close-down costs
is downsizing costs.

An earlier paper, Postner [1994a], spelled out these and other considerations in further detail.
An important component of start-up costs and change-over costs would be computer network
systems (re)design and related custom software and (re)training packages. There will be significant
periods when “the LAN is down” and staff are on “another” computer language training course.
These latter costs are difficult to identify outside the NAICS [1996]. But there also exist software
reuse repositories that are not yet captured by any classification systems.

What can we conclude from all this? It seems clear that we need implementation of both
NAICS [1996] and SNA [1993] as soon as possible. Once again, presently available productivity
results based on the old SIC and the old SNA, are not capable of fully reflecting the new information
technology — particularly on the software side. It is, however, not easy to say what a new set of
historical results would really look like even with fully implemented Systems. Indeed, there are
considerations that go beyond both NAICS and SNA. One of these is correctly accounting for the
costs of technological change — outlined above. Two other considerations and their possible
repercussions are briefly outlined in the following two sections. We would then be prepared to “put
it all together”.

5
4. NETWORK FORMS OF ECONOMIC ORGANIZATION
WHAT ARE THE ISSUES?

The official data used to measure industry and sectoral real output and productivity growth ultimately stem from a complex statistical infrastructure erected by the national statistical agency. To focus on essentials, we will only consider the methodological sources underlying production statistics — sometimes called principal industrial statistics. In fact the availability and collectability of such statistics also underlies any SIC, including the NAICS [1996]. The discussion in the previous sections concerning new features of the NAICS pre-supposes that principal industrial statistics are collectible, at least down to the newly classified 4-digit industrial level.

Briefly, a statistical infrastructure is composed of: (1) central registers (and profiles) of economic units, (2) a collection of survey interfaces, and (3) a list of standard economic events accounting for the creation, dissolution and updating of economic units. The standard economic statistical unit basic to production statistics is the so-called “establishment”. Most statisticians recognize that the “establishment” should be interpreted in a flexible manner; in fact, we prefer to simply use the term “business unit” (see also Ryten [1993]).

The theory and practice underlying the business unit ultimately pre-supposes either one of two forms of economic organization: (1) markets, characterized by legal contracts and market prices; and (2) hierarchies, characterized by established routines and administrative fiat. It is, however, very difficult to apply the business unit to a third form of economic organization: (3) networks, often characterized by informal, consensual and reputational relationships. (The “art” of consensing was first investigated by Postner [1976].)

It is true that statisticians have come to include “joint ventures” and “strategic alliances” in their infrastructural arsenal embodying the business unit. This is a step in the right direction. But it still pre-supposes that the unit has a formal going-concern existence with a contact location and statistics that are accountable and reportable in the usual manner. When we come to economic activities supposedly found in the new Information sector, however, the old rules of the statistical game no longer apply. The Information sector (as well as Services) is the prime locus of network forms of economic organization.

Briefly, there are at least three principal problems: (1) information and cultural products as economic commodities often are not capable of absorbing the standard rules of business accounting or simple arithmetic; (2) digitalized economic transactions performed over electronic media leave no paper trail or even electronic trail for statistical reporting (no “bit” tax is on the horizon); and (3) a significant element of economic activity occurring within network forms of organization is purely voluntary or informal or relational — outside the present limited scope of production statistics. In addition there is the well-known problem: (4) economic activities occurring in cyberspace have no address, no physical location and may stem from virtual transient-concerns that can “add up” in importance. (The updating impulses of standard events do not work.) So all the precious accounting and standard precepts fundamental to statistical collection are violated!
Before continuing, it must be noted that the “network” is not a statistical unit of any kind. It is, however, an increasingly important form of economic organization and indeed appears in different (hybrid) forms, not of all which require the electronic medium (Bureau of Industry Economics [1992] provides many examples and a guide to the literature).

The question then remains: If we are to have an Information sector and a Services sector in the NAICS that are reasonably all-inclusive, what new kind of statistical unit should we be looking for in order to keep track of a rapidly changing economy? At the moment, it appears to this writer that a “business-as-usual” approach to the new Information sector (and other sectors too) would run the risk of missing significant collections of economic activity — a lot of which should be included in production activity — and therefore undermining our estimates of real output and productivity growth. (Some earlier ideas are outlined in Carter & Postner [1996].)

But we need a lot more than a new statistical unit. (A new unit is easy to “name” but that is all.) We need: (1) new dimensions of survey interfaces, and (2) a new list of nonstandard (imputed) economic events. This writer’s feeling is that we should spend more time examining the potential for new kinds of community-based surveys and media-sourced monitoring, as well as transversally-integrated business surveys. There is the risk of course of double-counting, but there are worse crimes in life than double-counting, namely “no-counting”! This writer believes it is possible to construct a new statistical unit, fortified by new survey interfaces and nonstandard events, that would be relevant to fully exposing and accounting for the new network economy.

At the moment the NAICS [1996] is reasonably up-to-date in terms of industrial classification. But it is reasonably out-of-date in the broader terms of statistical appropriation. It is not advisable to simply isolate NAICS, or any other classification system, from matters of economic organization during periods of rapid economic change.

5. THE INNIS-McLUHAN CONNECTION

Harold Innis (1894-1952), Canada’s greatest economist, spent the last few years of his productive life studying the “bias of communication”. His ideas were taken up and extended by his great disciple, Marshall McLuhan (1912-1980). McLuhan was not an economist by training; he was educated in English literature. But after the death of Harold Innis, McLuhan shook the media world of communications with his creative ideas embodied in books, lectures, radio, TV and newspapers. McLuhan lived long enough to witness the computer hardware/software revolution and, therefore, he is worth quoting at length [1980]. Further analysis of the Innis-McLuhan connection can be found in Menzies [1996] and a forthcoming paper by this writer.

Consider then the following quotation from McLuhan’s last piece of writing:
In the ‘80s, there will be a general awareness that the technology game is out of control, and that perhaps man was not intended to live at the speed of light...

The ‘80s ... will witness a dramatic increase in the conservative backlash against runaway technology and change. Excessive speed of change isolates already-fragmented individuals and the accelerated process of adaptation takes too much vitality out of communities. By sheer attrition the social group is reduced to the condition of an anemic individual without the energy to adapt to the demands of survival...

Everybody will come to feel that his job, his family, his pension and his very identity are threatened by every kind of change... When people feel a threat to their identity, when they sense a danger to their self-image, they become very anxious and even violent...

The frontier of the ‘80s are much more inward, numerous and elusive than in the old hardware days. It might even be said that at the speed of light man has neither goals, objectives nor private identity. He is an item in the data bank — software only, easily forgotten — and deeply resentful.

There is not much economics or statistics in this quotation; more a matter of physiology and psychology. It seems clear, however, that what McLuhan says about the ‘80s could also apply to the ‘90s or beyond. If McLuhan were alive today, he would probably be writing for Wired. He would probably be particularly impressed by the Net and the Web and their, as yet, unrealized potential for “humanizing” the information technology revolution.

For our purposes we can ask the following. Suppose Marshall McLuhan was correct in his concern for runaway technology and change. Suppose the impacts and backlash described in the quotation are a faithful representation of the ‘80s and the ‘90s. Precisely how, then, could these impacts show up in official statistics of economic growth and productivity change during these time periods? Are the official economic statistics sensitive to the matters raised by McLuhan? Or are the official economic statistics somehow immune to these matters — calling for more general social indicators of well-being? It should be clear that unless we have answers to these questions, there is not much we could say about the productivity paradox in all its aspects.

It is also possible, of course, that McLuhan was simply wrong and that there is no substance in these matters. But substance or not, we could agree that economic and social statistics sensitive, in one way or another, to the concerns of McLuhan (and other similar thinkers) should be made available by national statistical agencies.
6. PUTTING IT ALL TOGETHER

This section contains two Summary Highlights that confront the two major concerns of the Conference. Most of the points have already been raised in the course of the paper; here they are simply put together.

A. Do official data produced by statistical agencies for service industries truly capture changes in real output and productivity?

Summary of Empirical Considerations

(1) Services are not correctly identified in present industrial classifications that are at least 15 years out of date.

(2) We need to break out Information and Cultural industries from traditional Goods and Services and further extend the Information (and Cultural) sector of the NAICS [1996].

(3) Productivity data do not correctly account for intangible fixed capital assets along the lines advocated by SNA [1993] and clarified by Hill [1997].

(4) Productivity data do not correctly reflect: up-front costs, re-organization & down-sizing costs, and close-down costs.

(5) The different phases of software formation and retrieval — as a manufactured good, as an information commodity and as a service — are mismeasured or unidentified in official data. The relative efficiencies of software codes are not taken into account.

(6) Statistical units underlying the proposed NAICS [1996] for the Information (and Cultural) industries and Services do not fully capture network forms of economic organization and their production activities. Coverage cannot be guaranteed unless NAICS is complemented by appropriate satellite classification systems.

Conclusion: Official data do not reflect underlying reality. But we could still go ahead to the second question.

B. The Productivity Paradox: Why is the information technology revolution having no apparent payoff in terms of productivity growth?
Summary of Possible Resolutions

(1) With full implementation of SNA [1993] and NAICS [1996] the productivity paradox would disappear. (This resolution is not likely.)

(2) Technology is changing “too fast” — reflected in rising up-front costs and rising close-down costs including bankruptcy — even though the costs are officially mis-allocated.

(3) The “era” of intellectual property either: (a) raises the costs of protecting ownership of intellectual creations — litigation and registry delays, or (b) raises the costs of not protecting ownership — pressure to be first at “any cost”, built-in obsolescence, and excessive secrecy.

(4) The information revolution largely resides within network forms of economic organization that are presently either: (a) obscured, or (b) simply missing — in the official data.

(5) The McLuhan issues significantly depress officially measured productivity growth data — when all is said and done. (This resolution is not likely.)

(6) The productivity payoff from the information revolution is selective and has not yet matured to have a general economic impact similar to other revolutions. We are still waiting for user-friendly and intelligent electronic agents that are sufficiently powerful to act in our individual economic interests.

Conclusion: The “year 2000” problem in software code probably implies no short-term relief of the productivity paradox. But in the meantime note that legal shenanigans are a common element of some of the possible resolutions. This is why it is so important to cover network forms of economic organization — where legal shenanigans tend to be minimized and social capital (trust) tends to be maximized.
7. NEW CHALLENGES FOR NATIONAL STATISTICAL AGENCIES

This Conference is a good opportunity to pose some new challenges for national statistical agencies. In this concluding section we go a little beyond the strict terms of reference stated in the preceding section. Yet the challenge also follow naturally from the preceding considerations. The general tone is now decidedly forward-looking with an appeal to a larger audience. Perhaps some of the following points may provide the themes of the next CSLS Conference!

It seems best to merely list 10 sets of new (and not-so-new) challenges, each with a brief description. We might note that the “challenges” are not necessarily posed to statistical agencies alone; they also apply mutatis mutandis to the applied economics profession.

Five General Challenges

(1) Can statistical agencies adopt a more “open door” policy with respect to the detailed methodologies and short-comings of their officially produced data? Where are the data “good enough” for sophisticated analysis? Where are the data “bad enough” not to warrant further analysis?

(2) What do statistical agencies regard as their major collection and methodological problems at this time? How can the profession of applied economists (outside the agencies) help out in these respects? Is there a role for new “relationships” between statistical producers and users of economic data?

(3) How can statistical agencies “educate” the users of their economic data with respect to the complex statistical infrastructure underlying the data? Are professional users of economic data really interested in learning about statistical infrastructure?

(4) Can statistical agencies provide their constituency with acceptable guidelines for modifying and adapting their data to fit the needs of particular economic research projects? For example, the NAICS [1996] may not satisfy all users; how can users perform industrial reclassifications in a systematic manner?

(5) What could be done by statistical agencies to set up “regional” field offices at headquarters of major multinational enterprises? How are statistical agencies “coping” without field offices on orbiting commercial satellites? Could we learn more about international “joint ventures” between different national statistical agencies?
Five Specific Challenges

(6) Is it realistic to think of “quality-adjustments” (both up and down) for literally thousands of different commodity items that enter the official data for real output and productivity growth? Is it realistic to think of “early” price adjustments for the literally hundreds of new commodities (and processes) that enter the market each year, most of which disappear within the year? Could we establish some key priorities, e.g., the relative efficiencies of original software codes (both as inputs and outputs)?

(7) Where are national statistical agencies per se classified in the NAICS [1996]? I would recommend their classification somewhere in the Information (and Cultural) industries sector. If statistical agencies do not produce information, then who does? The very essence of information is to be self-reflexive.

(8) How should we economically classify the proceedings of this Conference? Can we consider the proceedings and papers of the Conference as a bona fide production activity that should be included (or imputed) somewhere in the NAICS? In my view, this Conference is a production activity occurring as part of a network form of economic organization. Fortunately, no attempt is being made to impose ownership on intellectual property, but we do have reputational concerns.

(9) Can statistical agencies begin tilting away from the strict economic considerations of economic output and productivity growth, to more general social indicators of well-being? This would permit sensitivity to the concerns of Marshall McLuhan and other Canadian scholars and social activists.

(10) What plans have statistical agencies made to account for changes in the “genetic material” of the economic system itself? An historical example would be: the change from pure barter to the creation of a medium of exchange. A future (forthcoming) example would be: the change from our present economic system to one based mainly on electronic agents and markets.
REFERENCES


