



To whom it may concern:

I would like to bring to your attention a new industry dataset that Statistics Canada is placing in Canadian universities under the Data Liberation Initiative (DLI).

The data consist of the underlying industry statistics that are used by Statistics Canada's productivity program to produce labour and total factor productivity summary statistics.

The Productivity Accounts develop and maintain a large database in support of the productivity program—what some refer to as the KLEMS (Capital, Labour, Energy, Materials and Services) database. KLEMS integrates time series data on gross output, materials inputs, service inputs, energy purchases, labour, investment and capital. Each of these data series is calculated in both nominal dollars and real (chained volume) dollars. Price indices are collected for each of these series. Finally, KLEMS classifies these series using three different levels of aggregation—corresponding to the S, M, and L levels used in Statistics Canada's input-output accounts. The period covered by the database extends from 1961 to the current reference year (2003 at present). We have recently generated a new database in the North American Classification System (NAICS) to complement the earlier database that used the Standard Industrial Classification system (SIC).

Statistics Canada's Productivity Accounts build first off an integrated set of production accounts—that generate GDP from final demand, from income data and from industry level data—and produce a set of integrated, coherent accounts. The Productivity Group takes this integrated set of accounts and produces a set of estimates of labour services and capital services that are coherent with the output estimates. For example, on the labour side, the Productivity Group chooses amongst various source data (there are multiple sources, i.e. household versus employer surveys, each giving different estimates of labour inputs), ensures the boundaries of the labour sources agree with the boundaries of the industry output data, and produces a set of labour inputs. In the case of capital services, the Productivity Accounts start with investment data taken from a survey of investment, reconcile and modify them to accord with National Accounts boundaries, and then estimate capital services making use of rates of return that are derived from the National Accounts estimates of profits or surplus taken from the input/output tables.

.../2

In addition to the data that we have distributed to the local DLI representative on a CD-ROM, we have included a manual on statistical methodology used for construction of the TFP estimates and the underlying series.

We have also included a paper that discusses how an analyst might use the database in such a way as to take into account the fact that it contains ‘aberrant’ observations.

This is done because, in constructing the KLEMS database, the Productivity Accounts integrate data from different sources within Statistics Canada that are not always perfectly comparable with one another. This process serves to improve data accuracy or suitability by contributing to the production of time series that are consistent over time. By their nature, the survey systems that provide data to the SNA are often not ‘time-series’ consistent. Industry and product classification systems have changed over time. Surveys (such as the Annual Survey of Manufactures) have changed their coverage. Surveys have been restratified. Each of these changes may improve survey estimates at a given point in time—but serve to render analysis over time less coherent. While rough corrections are often provided by survey programs to account for the impact of changes in coverage or classification, the survey programs rarely provide all of the changes that are required to provide time-series coherence. The Productivity Accounts assemble the data and modify the data so as to improve its overall coherence.

But this process is not perfect. And unusual observations may not all consist of errors. Therefore, the database needs to be investigated carefully before it is used for particular purposes. The paper that accompanies the KLEMS database is there to remind users of the desirability of asking whether the data meet their purpose—not to suggest a particular solution for every project.

I should also stress that we have provided the data at a more detailed industry level than Statistics Canada publishes—with the warning that not all series are equal in quality. Statistics Canada works at more detailed levels than are generally published for a number of reasons—not the least of which is the need to see where improvements can best be made. We provide the more detailed level of data here because they may prove useful for some projects. But we caution the user not to treat every industry or every series as being equal in quality. More detail on the quality of each series can be found on Statistics Canada’s Website

http://www.statcan.ca/cgi-bin/imdb/p2SV.pl?Function=getDocumentation&AC_Id=29216&AC_Version=1&ul=ul&lang=en&db=IMDB&dbg=f&adm=8&dis=2.

Finally, I should note that additional data is available for some purposes. Information on labour and asset types can be divided more finely than is done in the public database. We do not include finer details in the dataset that we are releasing because it is often not possible to do so without violating the confidentiality provisions of the *Statistics Act*. But we do have a doctoral and post-doctoral program that permit access to these data by academics working here in Ottawa. If there are projects that require more detailed data, I encourage you to contact me at the email address below.

I trust the data that we are depositing in the DLI will be useful to your faculty and students.

Yours sincerely,

A handwritten signature in black ink, reading "J. R. Baldwin". The signature is written in a cursive style with a large, sweeping initial "J".

John R. Baldwin
Director
Micro-economic Analysis Division
Statistics Canada
K1A 0T6
john.baldwin@statcan.ca
613-951-8588