THE DEMOGRAPHY OF THE BUSINESS SECTOR AND THE USE OF STATISTICS DERIVED FROM BUSINESS REGISTER AND INSOLVENCY RECORDS

Marta F Haworth, Central Statistical Office, John R Kerr,
Department of Trade and Industry, United Kingdom
Small Firms Statistics Unit, DTI, St Mary's House, Sheffield S1 4PQ, UK.

KEY WORDS: Administrative records, business starts, failures, lifespan

Introduction

Statistics on the demography of business populations are of considerable interest to policy makers, the research community, trade bodies and firms themselves. Statisticians are required to produce timely and comprehensive data on business populations and analyse trends not just in broad aggregates but also for sub-populations relating to particular size groups, regions, industries and types of firms.

In the context of industrial and employment policy, there is particular interest in the formation of new businesses, in job generation and the role of small firms in this process, in the lifespan and growth of businesses and in business failures. Data on spatial and industry variation in job creation are relevant to regional industrial policy. There has long been an interest in role of small firms in the economy, including their contributions to growth and the development of a competitive economy (1). Trends in numbers of firms and in the numbers of business failures in particular have been attracting considerable public interest and have increasingly been regarded as indicators of the health of the economy.

Users increasingly demand better statistics, and yet many of the areas of interest are particularly difficult to survey. Elsewhere, for example in some service sectors, the available statistics are patchy. Additionally, deregulation and the requirement to reduce statistical demands on firms have resulted in resistance to extensions to statistical surveys. This is of particular importance with respect to small firms where compliance costs are often relatively the greatest.

This paper outlines the UK approach to these conflicting pressures for more and better statistics on one hand and less data collection on the other and presents statistics derived from administrative records on trends in the small firms population, analyses of business lifespan, and business failures. It also discusses how information from administrative records is combined with existing survey data to compile analyses of business population aggregates. Current UK experience with the use of administrative records is set against plans for future developments and the benefits which will result from register integration.

The Administrative Data Sources

The main administrative sources used for business demography statistics in the UK are:

taxation records from the administration of value added tax (VAT) which form the basis of the current main statistical business register (2).

company registration data obtained from the agency which administers incorporation of companies;

insolvency data from courts dealing with company winding up cases and published details of individual bankruptcies;

other sources including non-official data, for example, on business accounts held with major banks.

The Advantages and Limitations of Administrative Data

Although in combination the above sources contain quite a lot of detailed information, they share the usual problems and limitations associated with administrative sources (Annex A):

i) their coverage is affected by taxation and legislative changes or changes in administration, leading to discontinuities in time series (for example when there are VAT threshold changes);

ii) indicative information which is not critical for the primary use can be of variable or even inadequate quality -for example, turnover and classification on VAT records; classification of insolvency records;

iii) significant and variable processing lags can affect comparability of data over time, often reflecting factors such as administrative workloads or court

sessions, rather than changes in the statistical characteristics of the target populations;

iv) their coverage does not exactly match the business population studied - for example, many company incorporations relate to companies who are not actively trading.

v) statisticians have little control over them.

vi) where coverage is not total, it is often not representative of the population as a whole;

vii) units from different administrative sources may not be directly comparable.

Despite their limitations, these sources have some important advantages:

i) the statistical information is available at relatively little additional cost and without additional loading on business;

ii) the information is available for all businesses covered by the relevant legislation and non-responses or sampling errors do not present problems.

iii) the timeliness is often better than can be achieved in complex survey work.

Analyses Derived from Administrative Sources

Since 1980, data from the VAT register have been used to monitor annual changes in the size of the UK business population and the trends in registrations and deregistrations (start-ups and closures) which are causing these changes. Separate analyses have also been carried out for industry sectors and geographical areas.

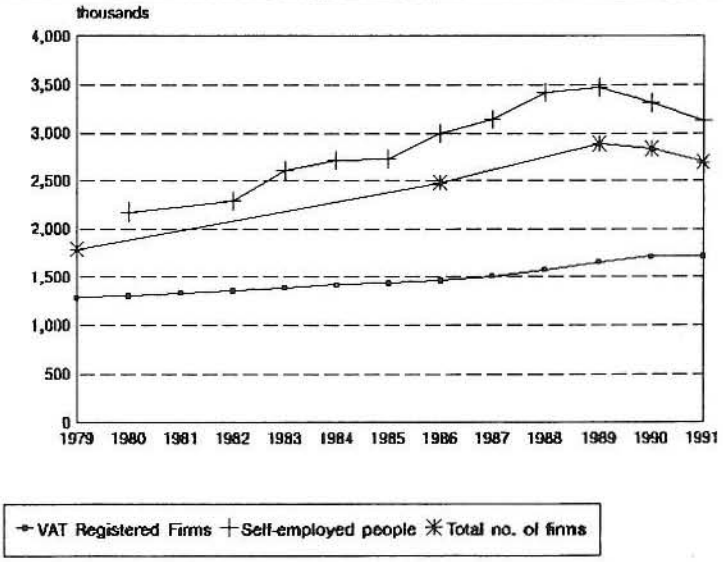
The data have also been used to study survival rates for businesses and the stages in their lifetime when they at greatest risk of failure. In addition the VAT register provides the basis for regular estimates of the size distribution of all firms.

Companies' House data provide a monthly count of the number of new incorporations and the stock of registered companies.

Data from banks is used to give quarterly estimates of new business start-ups, with limited industry and geographical breakdown.

Insolvency records have been used as the main indicator of business failures.

Chart 1
VAT registered businesses, self-employment and numbers of firms



source: Graham Bannock & Partners, CSO Bulletin 5/93, Labour Force Survey

Combining Data from Different Sources

In some cases data from administrative sources and from statistical sources are successfully combined; for example VAT-based information and data from the UK Labour Force Survey (LFS) (Annex B) are combined to compile estimates of the total number of firms (Chart 1) and also of the distribution of employment and turnover by size of firm (3).

The following paragraphs describe in more detail the use of the above sources for constructing statistics of:

i) size analyses of overall business population;

ii) lifespan analyses;

iii) business failures.

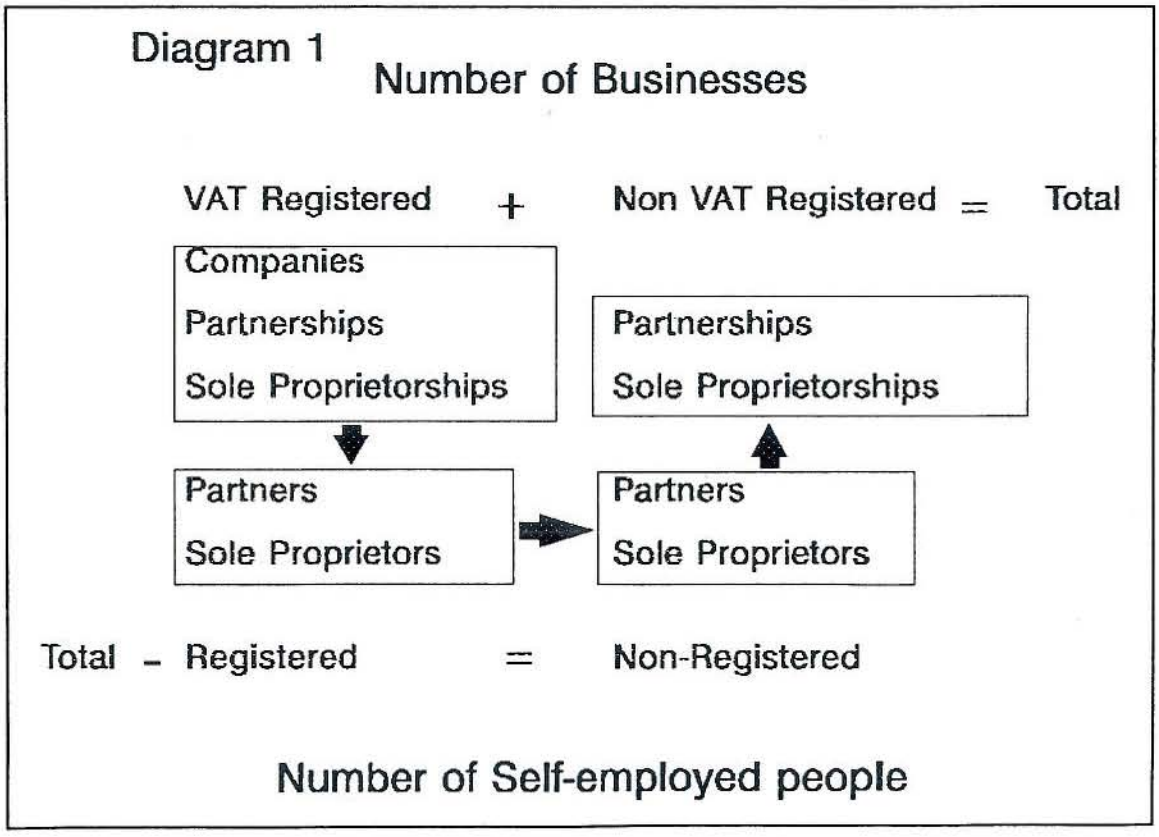
Analyses of UK Business Population by Turnover and Size

Estimates of the number of small firms and changes in the relative contribution to the UK economy by firms of different sizes form important inputs to policy discussions.

Although there is a variety of sources of data on the size distribution of firms, there is no single comprehensive source which can supply this information for all industry sectors. Estimates of the number of firms of different sizes are obtained by combining data from several sources at an aggregate level. The VAT register provides the basis for these estimates.

The starting point for the calculation is an analysis of the number of firms registered for VAT classified by legal form (sole proprietors, partnerships etc.), industry sector and turnover size band. The main exclusion from the VAT register is firms whose turnover is below the threshold for VAT (£37,600 in 1993/94) and who choose not to register. In order to take account of these firms, use is made of self-employment figures produced by the Employment Department from the Labour Force Survey (LFS).

The first step in using the self-employment data is to calculate the number of self-employed people who are registered for VAT. Subtracting this from the total number of self-employed gives the number of self-employed who are not registered for VAT, from which the number of non-registered businesses can be derived (See Diagram 1 below).



The number of self-employed people registered for VAT in each industry sector is obtained by multiplying the number of registered partnerships by an estimate of the number of partners per partnership and adding the result to the number of sole proprietorships. The number of partners per partnerships is derived from a further administrative source: tax records of the Inland Revenue.

The combined VAT and self-employment data provide a breakdown of the **total** number of firms by industry sector and turnover size band. The next stage is to use a transformation matrix to convert the breakdown by turnover size into a distribution of firms by employment size band. The transformation matrix, which gives the allocation of firms in each turnover band across employment size bands, is derived from survey data.

The turnover/employment data are from a survey of nearly 3000 businesses. This is large enough to allow turnover/employment matrices to be calculated for three broad industry sectors: production and construction, distribution and other service industries. The Annual Census of Production and the Times 1000 are also used to estimate average employment for the largest firms.

Finally, checks are made of the accuracy of the resulting distribution. The most obvious check is to calculate the implied number of people employed in each industry sector and compare these totals with published official estimates. In sectors where the employment totals differ, adjustments are made to the transformation matrix for that sector.

At the end of 1991 there were an estimated 2.7 million businesses in the UK. This represents an increase of 900,000 businesses since 1979. More than 99 per cent of firms employed fewer than 100 employees and these accounted for nearly a half of all employment and over a quarter of turnover outside central and local government.

Business Lifespan Analyses

The number of VAT registered business has increased by over 400,000, or 33 per cent, since 1980. This net increase is the result of 2.4 million new registrations and nearly 2 million deregistrations. With such a changing population it is of great interest to examine how long firms, on average, continue to trade and whether there are periods in their lifetimes when firms are more vulnerable.

Data from the VAT register can provide evidence on both of the above questions. The VAT records include the date of registration and, where appropriate, the date of deregistration. Thus for each cohort of registrations it is possible to identify the numbers (and proportions) surviving for a given period of time. Amongst those businesses which had deregistered, the time elapsed between the dates of registration and deregistration provides a measure of their lifespan.

Businesses may have been trading for some time before they register for VAT and some of them will continue to trade after deregistration.

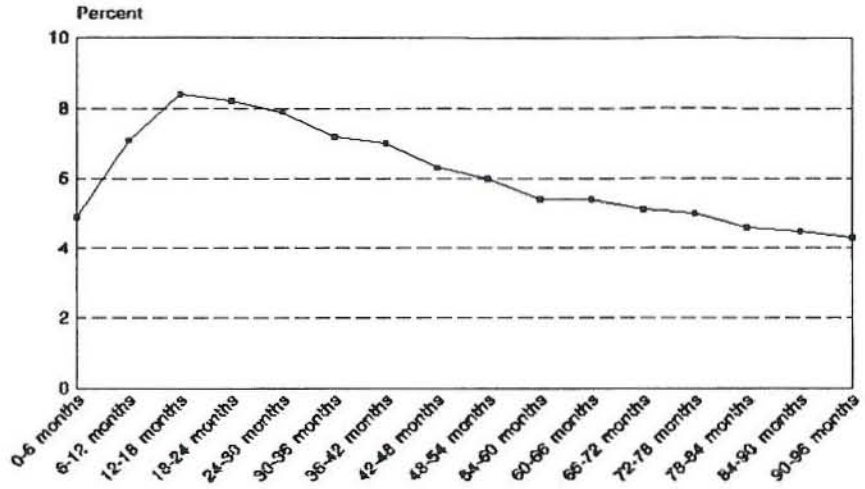
Consequently, the above estimates of survival rates and average lifespan provide a lower bound to the 'true' values for these parameters. The VAT database includes some information about the reason for deregistration and this can be used to give a broad indication of the extent to which the lifespan from registration is underestimated.

The vulnerability of businesses at different stages in their lifetime can be examined by calculating deregistration rates conditional on the business already having survived for a given period of time. Average conditional survival rates for the 1980s are shown in Chart 2. Businesses were at their most vulnerable between 12 and 30 months after registration. Thereafter, the probability of deregistration declined with the age of the business.

Business Failures

There is no precise or universally accepted definition of a "business failure". A whole spectrum of possible interpretations is possible, ranging from compulsory business closures due to insolvency to voluntary closures reflecting poor profitability, takeovers or businesses failing to meet their goals, or even closures due to retirement. There are many types of business closures which do not necessarily mean business failures - for example, many removals of records from the company register reflect restructuring of businesses or voluntary closures of non-trading concerns; many deregistrations from the VAT register arise because of changes in VAT threshold, rather than closure of the business.

Chart 2
Average deregistration rates



Number of deregistrations as percentage of number surviving at start of period
source: British Business, April 1987

The published statistics: Insolvencies

The official UK statistics published by the Association of British Chambers of Commerce (ABCC) on behalf of the DTI count company liquidations due to insolvency and bankruptcies of individuals, over 70% of which relate to individuals engaged in business. Insolvency, which occurs when a person or company is unable to pay debts on the due date, is included in the statistics only when it has been determined by courts or established by individual agreements.

The published official insolvency statistics give separate series for companies and individuals, with seasonally adjusted series to indicate the trend (4). Figures for Scotland are given separately from those for England & Wales. Breakdowns by industry are given for companies and for individuals. There are no satisfactory statistics of insolvencies by size of undertaking or by region because it would be impracticable and very costly to assemble the information. Although it can be argued that business insolvencies should be weighed rather than counted, in order to gauge the effect on the economy, there is no satisfactory measure of size or the value of a business at the time of insolvency - a business tends to decline over a period and there are problems of deciding at what time the size should be assessed, what measure should be used, and how to get the information retrospectively. A rise in numbers of insolvencies therefore indicates that business failure are becoming more painful, but it is not directly an indication of the trend in business failure or of the impact on the economy.

Other Official Statistics Relating to Business Failure

The Companies House Executive Agency (CHEA) administers the legal framework within which companies operate and the means by which those companies are formally registered, monitored and dissolved. The numbers of dissolutions and registrations can be compared and used to estimate the number of business starts and stops but they are not representative of businesses as a whole. The number of businesses being incorporated is dependent on the perceived benefits of incorporation which may change with the rates and thresholds of corporation tax and personal income tax. There are two registers at CHEA: the active register of companies regularly sending in their accounts and the dissolution register of companies that have failed to send in their accounts.

The CHEA publishes data on incorporated companies in its annual report (5). The number of company deregistrations includes not only liquidations due to insolvency but also liquidations for any other reason, as well as companies removed after persistently failing to provide accounts. The DTI uses the number

of companies on the active register as an indication of the total stock of companies.

The VAT register is also used as a measure of business starts and stops. It includes corporate businesses, partnerships and sole traders. Not all active businesses are registered for VAT: the coverage below the VAT turnover threshold is uncertain. Hence the number of registrations is not the true number of business start-ups. Nor are all VAT deregistrations equivalent to closures. Deregistrations include businesses which become exempt because of a change in the VAT threshold or a reduction in turnover and businesses which change their legal status, as well as voluntary closures for any reason. The Department of Trade and Industry publishes analyses (6) of the annual numbers of VAT registrations and deregistrations.

The Lord Chancellor's Department produces court-by-court statistics of winding-up petitions. Customs and Excise produce insolvency statistics on the number of winding up petitions that they have initiated. This covers a small proportion of the total number of insolvencies.

Other business failure statistics

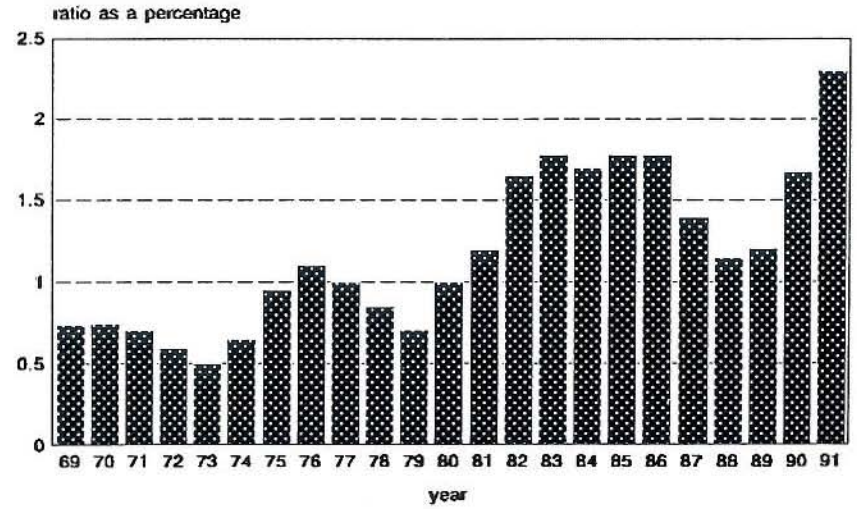
Several firms produce insolvency data: Dun and Bradstreet (D&B), KPMG, Touche Ross, Cork Gully and Grant Thornton. They produce their figures from the notices published in the London Gazette. Dun & Bradstreet publish quarterly press notices, giving the number of business insolvencies for the year so far for Great Britain, together with a regional breakdown.KPMG,Grant Thornton and Touche Ross publish statistics of receiverships but not all on the same basis, and there are no regular publication arrangements.

Trends in business failures

In the past two years, the number of insolvencies has risen sharply to record levels, both for incorporated companies and for individuals. Changes in the legal and administrative framework have introduced discontinuities (in particular, the Insolvency Act 1986 revised insolvency procedures from 1987 onward). The number of VAT deregistrations and Barclays' estimate of business closures also show increases over this period but at a much lower rate than implied by the insolvency figures.

During the 1980s the key feature is the rapid rise in the number of incorporated companies: the active company register has increased by about 40% to roughly a million. With the growth in the number of businesses, the number of company liquidations still represents just over 2% of the total number of active registered companies (Chart 3).

Chart 3
Ratio of company insolvencies to active registered companies in England & wales 1969-91



source: Companies House and DTI

Much of the increase in the proportion of companies becoming insolvent in the last decade has been attributed to the relative inexperience of those set up during the latter half of the 1980s, since new firms have been shown to be most vulnerable in the early years of their trading (Chart 2). Increased personal debt levels (associated with the growth in consumer credit) and corporate debt levels, increases in unemployment levels and the cumulative effects of bad debts due to other business failures have also been put forward as causes of the rapid increase in business failures. It has been suggested that changes in legislation may have had some effect.

Past attempts to establish relationships between insolvency levels and other economic variables have been of questionable value in providing reliable forecasts of insolvency levels. Work done in the 1970's on comparisons (7) with other economic indicators compared company liquidations with gross trading profit data (seasonally adjusted) and with GDP and FOP data. This confirmed an association between a fall in the numbers of insolvencies and growth in economic activity, but not a very close one. A limited degree of association between falls in the number of insolvencies and increases in company profits have also been found. This is not surprising: gross trading profits relate to all companies and the relatively few large companies account for the bulk of the total. On the other hand, most companies that become insolvent are very small and it is not obvious that their prosperity should be closely related to company profits as a whole. Recent exploratory work examined possible models involving (respectively) lagged company births, changes in GDP or growth in company debt to predict company liquidations. Changes in the number of insolvencies reflect changes in the level of economic activity but the relationship is not particularly close. The only conclusion which can be drawn is that the relatively high level of business failures is likely to persist in the aftermath of the current recession for the next 2 years or even longer.

Future Work

The development of the Inter-Departmental Business Register (IDBR) will improve the quality and range of administrative data that can be used to examine the demography of the UK business population.

The IDBR uses two administrative sources: records of traders registered for VAT and employers registered with the Inland Revenue for 'Pay As You Earn' (PAYE) tax purposes. The PAYE data includes information on the number of employees, though low paid workers are excluded.

Although the use of two sources creates problems of matching and consistency it has the benefit that it is much less likely to be affected by changes in either of the administrative sources.

To maintain consistency for statistical enquiries the administrative units (VAT traders, PAYE scheme units) in the IDBR will be linked to statistical units (enterprise, local unit). These links will enable more reliable estimates of births and deaths to be made by increasing the likelihood of detecting deregistrations and re-registrations which were for reasons related solely to the tax system.

Although very small businesses (those with a turnover below the VAT threshold and no employees liable for tax) will be excluded from the IDBR, the inclusion of employment data will improve the estimates of size distribution of businesses by reducing the dependency on data from a range of surveys.

Studies have shown that small firms make a disproportionately large contribution to net job generation. Hitherto research into job creation has had to make use of commercial databases (8). These suffer from the problems of administrative sources discussed earlier, but often have further deficiencies with respect to small firms. The coverage is generally poor and there may not be any system for regularly updating the information on the database. The IDBR will contain historical records which can be used to provide much more reliable estimates of patterns of growth of businesses. The wide coverage of the IDBR will also allow these analyses to be carried out with a more detailed industry classification.

References:

(1) Bolton Committee, 'Report of the Committee of Inquiry on Small Firms'; HMSO; 1971

(2) Perry J, 'The Development of a Business Survey Frame from Administrative Data'; ICES; 1993

(3) Companies House Executive Agency Annual Report; HMSO.

(4) 'Business Briefing'; The Association of British Chamber of Commerce.

(5) Daly M and McCann A, 'How many small firms?'; Employment Gazette; February 1992.

(6) 'VAT Registrations and Deregistrations in the UK (1980-1991)'; CSO Bulletin 5/93; 1993.

(7) 'Insolvency Statistics for England and Wales'; Economic Trends No 257; March 1975.

(8) Daly M, Campbell M, Robson R and Gallagher C, 'Job creation 1987-89: The contribution of small and large firms'; Employment Gazette; November 1991.

ANNEX A

UK Administrative Data Sources: their Coverage, Advantages and Limitations

1. **Business Register (VAT Registrations and Deregistrations)**

Coverage

Information on some 1.7 million legal units (companies, sole proprietors, partnerships etc) is collected in the course of administering the VAT system.

All business with turnover above a given threshold (currently £37,600) are required to register for VAT, other than those trading in exempt goods and services (mainly health and education services). Businesses with a turnover below the threshold may register on a voluntary basis.

Advantages

i) A balanced indicator of births and deaths of firms on a consistent basis.

ii) Offers scope for analysis by region and by company size.

iii) Allows the dynamics of the business sector to be studied.

iv) Information is available at little additional cost to government and without additional burden on businesses.

v) data are available for all registered businesses; this allows detailed analyses without problems associated with sampling errors.

Limitations

i) Figures only available since 1980.

ii) Coverage not complete and trends in the non-registered population may be different (Self-employment grew at a faster rate than VAT registered businesses in the 1980s).

iii) Registrations and deregistrations are not the same as births and deaths - there is a slight tendency for the registration series to lag trends in the number of births. However, the difference is not so great as to invalidate the analysis.

iv) Registrations and deregistrations can be affected by changes in the regulations governing VAT; allowances for the effects of such changes will be necessary and may also lead to discontinuities in the time series. The increase in the VAT threshold from £25,400 to £35,000 in 1991 led to an estimated 14,000 reduction in registrations and 41,000 increase in deregistrations.

2. Companies House

Coverage

All incorporated businesses in Great Britain are registered on the Companies House register.

Advantages

i) Complete coverage of all incorporated businesses.

ii) Timeliness.

Limitations

i) Not all new incorporations correspond to births - A business may have been trading for some time before incorporation.

ii) Removals from the register are not an accurate indicator of business failures. Many are only removed when it becomes apparent that they are no longer trading and removals are also dependent on workload priorities within the Agency.

iii) Narrow coverage - looks solely at companies and does not include individuals, i.e. sole traders, partnerships excluded.

3. Insolvencies

Coverage

Official insolvency figures are compiled from two main sources: The Insolvency Service Executive Agency and Companies House Executive Agency.

The Insolvency Service Executive Agency covers all individual Voluntary Arrangements, Deeds of Arrangement, Bankruptcy Orders and Compulsory liquidations which go through local courts.

Companies House Executive Agency covers all insolvency procedures undergone by companies.

Advantages

i) Clear legal framework;

ii) Series back to 1960;

iii) A clear definition; figures indicate the number of painful business failures.

Limitations

i) Consistency of the series over time is questionable because of legislative changes and changes in administrative procedures and the factors which affects the choice of individual businesses whether to seek corporate status.

ii) Comparable and meaningful figures of business start-ups not available.

iii) The insolvency statistics relate to numbers of businesses but do not measure the effect on economic activity or unemployment.

iv) Outside the corporate sector, it is not possible to distinguish accurately between business failures and personal bankruptcies.

v) Insolvencies are not representative of business failures in general - they make up only a small and extreme portion of the total.

4. Other Sources

Clearing Banks

Coverage

Two banks use information on numbers of new business accounts opened to estimate the number of new business start-ups. Both are based on estimates of the number of new non-personal accounts which correspond to business start-ups together with survey estimates of market share.

NatWest have provided 6 monthly estimates of start-ups during 1985-1992, and quarterly estimates since end 1992. Barclays have quarterly estimates going back to 1988 for start-ups and closures.

Advantages

i) Timeliness: estimates are usually published within about a month of the end of the relevant period.

Limitations

i) The market share estimates are subject to sampling errors, and to date no assessment has been made of the reliability of the overall estimates.

ii) The two banks' estimates do not always agree and there is no systematic difference in the two sets of estimates.

ANNEX B

LABOUR FORCE SURVEY

First conducted in 1973, the LFS was carried out every two years until 1983 and every year between 1984 and 1991. A quarterly survey was introduced in Spring 1992. The LFS provides estimates of numbers of self-

employed people, categorised into those with and without employees.

Estimates of entries into and exits from self-employment are available. With the annual survey these were based on questions asking about labour market status one year previously. The number of people who say they have moved in or out of self-employment seldom correspond with the overall net change in self-employment between successive surveys and adjustments are needed to the estimates of entries and exits to reconcile these two figures. As a result the estimates can only give a broad indication of the level of entries and exits.

The quarterly survey has an 80 percent sample overlap between successive quarters. This will allow for direct comparisons of labour market status at different points in time and more reliable estimates of entries and exits.

IMPACT OF STATISTICS CANADA ESTIMATES ON CANADA'S FISCAL EQUALIZATION SYSTEM

Krishna Sahay and Clancy Barrett
Krishna Sahay, 20th Floor, Section P, Tunney's Pasture
Ottawa, Ontario K1A OT6

Introduction

Federal states must deal with issues of regional disparity and the unequal fiscal capacities of sub-national governments, such as provinces or states. Since Confederation, in 1867, Canada's federal government has transferred funds to provincial governments in order to permit all provinces to provide comparable levels of public services to their residents at comparable levels of taxation. At present, the federal government administers three major programs designed to assist provinces in the provision of public services: Established Program Financing, the Canada Assistance Plan and the Equalization Program under the Fiscal Arrangements Act.

This paper focuses on the Equalization Program and on Statistics Canada's role in its administration. Under the Fiscal Arrangements Act, the Chief Statistician of Canada provides an annual certificate showing data used in the calculation of equalization payments. All statistical estimation procedures require the use of professional judgement and carry an irreducible amount of statistical estimation error. The primary purpose of this study is to identify those variables which have a major impact on provincial equalization entitlements and to examine whether statistically insignificant variations in estimates could have material revenue consequences.

The plan of the paper is as follows: Section 1 presents a brief overview of the equalization program, including the formulae used in the calculation of entitlement payments and the theoretical effect of changes in statistical variables. Section 2 describes the simulation experiments that were carried out; Section 3 describes the classification of variables by the level of impact on equalization payments and provides an evaluation of variables to identify those which both have a high impact on equalization and carry significant statistical risk. Section 4 provides a brief conclusion.[1]

1. The Equalization Program

1.1 Background

The equalization program is based on the concept of equalization of provincial "fiscal capacity". The fiscal capacity of a province is based on the availability of tax bases within the province and is, therefore, a measure of the ability to raise revenues, rather than a measure of actual revenues. The Equalization program involves the transfer of federal funds to those provinces whose fiscal capacity is below an agreed standard. Over time this standard has varied, and at different times, has been the equalization of revenue to: a national average; the richest province; the two richest provinces; and has evolved today to a five-province standard consisting of Quebec, Ontario, Manitoba, Saskatchewan and British Columbia.

Provincial governments have many sources of revenue and those that are subject to equalization have also varied over the years. At present the fiscal agreement considers 37 revenue sources, ranging from personal and business taxes to very specific mineral revenues.

For Fiscal Year 1988-89, the federal government transferred $7.3 billion to seven of Canada's ten provinces under the equalization program. In the receiving provinces, as a group, these transfers amounted to 15 per cent of the revenues from their own sources. The transfers were particularly important to the Atlantic Provinces; in Newfoundland and Prince Edward Island equalization payments approached 50 per cent of provincial revenues from their own sources.

[1] Due to space limitations only excerpts from the analytical tables and results are shown in this paper. For detailed results please contact the authors.

1.2 Calculation of Entitlements

A province's equalization entitlement is calculated on the basis of 37 revenue sources. Revenue sources include items such as personal and corporate income taxes, sales taxes and property taxes. Corresponding to each revenue source is a revenue tax base. The revenue tax base is an economic variable to which a tax rate is applied to yield revenue. For instance, personal and corporation income, retail sales and value of property are tax bases which determine the value of the revenue sources listed. The number of variables composing a revenue tax base varies significantly from base to base. A total of 210 variables are used in the calculation of the 37 tax bases, and of these Statistics Canada provides 158.[2] Equalization entitlements are calculated on the basis of the revenue tax bases and hence the influence of any particular variable on entitlements depends on its relative importance in the revenue base calculation.

The actual calculations, are carried out for each revenue source separately and involve four steps as shown in Figure 1 on following page.

As Figure 1 suggests, there are a number of factors which affect provincial entitlements:

- Entitlements depend on the size of provincial revenue bases which measure each province's ability to raise revenue, but are not directly related to whether or not a province chooses to exercise that ability to raise revenue.
- While a province's tax rates do not directly affect its entitlements under this formula there is an indirect effect due to the fact that the national average tax rate is simply a weighted average of the tax rates in all provinces. The weights are the provincial shares for each tax base. Changes in the weights of large provinces or in the tax rates of large provinces could have significant effects on the entitlements of all provinces.
- In addition to the effect on national averages, changes in the revenue tax base or tax rates in standard provinces have an impact on all provinces while non-standard provinces affect only their own entitlements.

2. The number of variables used in each year can be counted in various ways, depending on the treatment of things such as lagged values of variables. In the count used here each variable is counted only once, regardless of the number of lagged values that are also reported.

Figure 1

Calculation of Entitlements: Single Revenue Base

1. Calculate the implicit **national average tax rate** for that base as:

$$t_c = \frac{\text{Total Revenues from source for all provinces}}{\text{Total tax base for all provinces}}$$

2. Calculate the **hypothetical per capita** revenue from the tax base in the five standard provinces if these provinces had imposed a tax rate equal to the national average tax rate:

$$R_s = t_c * \frac{\text{(Total revenue tax base for five standard provinces)}}{\text{Population of five standard provinces}}$$

3. Calculate the **hypothetical per capita** revenue (R_i) from the tax base in the province "i", using the formula from step 2 and data for province "i"

4. Calculate the provincial entitlement for the province from this revenue source as:

Entitlement = Population of province * (R_s - R_i)

The entitlements from any particular revenue base could be positive or negative. The final provincial entitlement is the sum of entitlements from all of the revenue sources.

To analyze the effects of changing variables which constitute the different revenue tax bases the following section reports on simulation experiments in which different revenue tax bases are increased by 1 per cent. An increase in a revenue tax base requires changes in one or more variables used in the calculation of the base. In order to facilitate presentation the paper begins by analyzing changes in the revenue tax base, without identifying which variables are changed. Subsequently there is a brief discussion of the actual variables. Since the equalization formula is based on per capita results for each province the distribution of population between provinces is an exceptionally important variable. For this reason simulations of changes in population are presented separately in a later section.

1.3 Analysis of the Effect of Changing Revenue Base Variables

The equalization calculations shown in Figure 1 can be summarized by the following equation:

$$(1) \quad \frac{E_i}{P_i} = t_c\left(\frac{B_s}{P_s} - \frac{B_i}{P_i}\right)$$

where:

- E_i = entitlement to province i.
- P_i = population in province i.
- t_c = national average tax rate for a given revenue source calculated by dividing total revenue source (TR), by the total revenue tax base (B_c)
- B_s/P_s = per capita revenue tax base of the standard provinces
- B_i/P_i = per capita revenue tax base in province i.

To facilitate the analysis Equation 1 can be rearranged as follows:

$$(2) \quad E_i = \frac{B_s}{B_c} TR\left(\frac{P_i}{P_s} - \frac{B_i}{B_s}\right)$$

From equation (2), a province's equalization entitlements are determined by the difference in its population and revenue base shares relative to the standard provinces. Thus, a province with a large share of the total population and a small share of the revenue base would have a large equalization entitlement. If the population share and the revenue base shares are equal then no equalization entitlements will exist. If the shares are approximately equal then any entitlements will be small and even small percentage variations in the revenue base can cause large percentage changes in entitlements.

Table 1 shows the relationships between changes in a revenue base and entitlements from that base. The relationships are presented as elasticities, i.e. each entry in the figure shows the percentage change in entitlements from a 1 per cent change in a revenue base. Also, for reasons of simplicity only, the table presents the elasticity for the special case of a province whose tax rate is equal to the national average tax rate. More general results are included in a detailed analytical appendix which may be obtained from the authors on request. This simplifying assumption implies that the elasticities shown reflect the effect of changes in the revenue tax base only and not of tax rates.

As Table 1 indicates, the elasticity of entitlement ($\mathcal{E}_{Ei,Bi}$) with respect to a revenue base is a complex function of the ratios of population and revenue bases to the corresponding quantities in the standard provinces. However, the economic interpretation of the relationships is straightforward. In Table 1, B* represents the ratio of the revenue base in the "i^{th}" province to the revenue base in the five standard provinces. In the following, we refer to this ratio as the provincial share. Similarly, P* is the population share of the "i^{th}" province and [P*-B*] represents the difference between the province's population share and its share of the revenue base. Thus, the main factors governing changes in a province's entitlements are that province's revenue base share and population shares. Clearly, $[B^*/(P^*-B^*)]$ is not defined when $P^*=B^*$. In other words, when the province's population share and its share of the revenue base are equal its entitlement from the base is zero and therefore the elasticity of entitlements is undefined. When P^* and B^* are approximately equal the estimated elasticity is likely to be numerically very large, but has very limited economic significance since the dollar amounts involved are likely to be small.

Table 1

Elasticity of Entitlement with Respect to Revenue Base, $t_i = t_s$

	Own Province Impact	Other Provinces Impact
Standard Province	$-\frac{B^*}{P^*-B^*}+\frac{B^*}{P^*-B^*}\frac{P_i}{P_s}$	$\frac{B_i}{B_s}\cdot\frac{B^*}{P^*-B^*}$
Non-Standard Province	$-\frac{B^*}{P^*-B^*}$	0
$where\ B^* = \frac{B_i}{B_s},\ P^* = \frac{P_i}{P_s}$		

Table 2

Effect of 1 Per Cent Increase in Revenue Tax Base: Payroll Taxes - Base Effect Only
($'000)

	NFLD	PEI	NS	NB	QUE	ONT	MAN	SASK	ALTA	BC
NFLD	-317	0	0	0	143	266	21	16	0	69
PEI	0	-71	0	0	32	60	5	4	0	15
NS	0	0	-612	0	221	413	33	26	0	106
NB	0	0	0	-455	179	334	27	21	0	86
QUE	0	0	0	0	-3641	3108	249	192	0	800
ONT	0	0	0	0	2367	-5484	353	273	0	1137
MAN	0	0	0	0	272	507	-751	31	0	131
SASK	0	0	0	0	254	474	38	-583	0	122
ALTA	0	0	0	0	599	1118	89	69	-2151	288
BC	0	0	0	0	748	1395	112	86	0	-2190
Percentage Change in Total Entitlements From All Tax Bases										
NFLD	-0.038	0.000	0.000	0.000	0.017	0.032	0.003	0.002	0.000	0.008
PEI	0.000	-0.040	0.000	0.000	0.018	0.034	0.003	0.002	0.000	0.009
NS	0.000	0.000	-0.073	0.000	0.027	0.049	0.004	0.003	0.000	0.013
NB	0.000	0.000	0.000	-0.059	0.023	0.043	0.003	0.003	0.000	0.011
QUE	0.000	0.000	0.000	0.000	-0.107	0.092	0.007	0.006	0.000	0.024
ONT	0.000	0.000	0.000	0.000	0.056	-0.130	0.008	0.006	0.000	0.027
MAN	0.000	0.000	0.000	0.000	0.034	0.064	-0.095	0.004	0.000	0.016
SASK	0.000	0.000	0.000	0.000	0.056	0.104	0.008	-0.127	0.000	0.027
ALTA	0.000	0.000	0.000	0.000	0.016	0.030	0.002	0.002	-0.058	0.008
BC	0.000	0.000	0.000	0.000	0.095	0.177	0.014	0.011	0.000	-0.278

Simulation results illustrating the effect of changing a revenue base by 1 per cent are shown in Table 2. The results are presented in a matrix format where the diagonal elements represent the "own province" impact of an increase in the revenue base. Off-diagonal elements in each column represent the effect on other provinces. When the provincial tax rates are all equal to the national average [Table 2], a 1 per cent increase in the revenue base in Newfoundland resulted in a decline of $317,000 in entitlements to Newfoundland, but had no effect on other provinces since, as a non-standard province, Newfoundland has no direct effect on other provinces. In contrast, changes in the standard provinces have a significant influence on the outcomes for other provinces. As the table indicates, elasticities can vary significantly from province to province.

2.2 Analysis of Changes in Population

The distribution of population amongst provinces can affect each of the revenue tax base calculations of entitlement, and therefore can have potentially significant effects on a province's entitlements. This section examines the impact of population change on equalization entitlements.

We examine two different ways in which the population of a particular province can change:

- An autonomous increase in the population of a province, leading to a corresponding increase in the total population of Canada and;
- An increase in population of a province caused by shifts from other provinces, with the total population of Canada remaining unchanged.

2.2.1 Effects of an Increase in Aggregate Population

- A population increase in a standard province results in an increase in entitlements in that province and a decrease in entitlements in all other provinces.
- An increase in population in a non-standard province results in an increase in entitlements in it's own province and has no impact on the other province.
- On a per capita basis, it is not clear whether a standard province would have an increase or a decline in entitlements even though the total entitlements associated with the revenue tax base would increase. The key factor determining the effect on a per capita basis is, not surprisingly, how the provinces base share compares to its population share. The per capita entitlements of other provinces would, however decline.
- The per capita entitlements of a non-standard province would increase whenever its population increases.

2.2.2 Effects of a Population Shift

An alternative analysis of population examines the impact of a shift in population from one province to another.

A shift in population either from a standard or non-standard province will increase the entitlements of the receiving province, whether it is a standard or non-standard province. However, the magnitude of the impact will be influenced by whether the population receiving and sending provinces are standard provinces.

If the population shifts towards a standard province, the entitlements for all provinces, other than Ontario decline. Further, the total entitlements of all provinces combined is reduced. On the other hand, when population is shifted between standard provinces, such as Ontario and Quebec, the total entitlements remains unchanged and there is simply a shift in entitlements from one standard province (Quebec) to another (Ontario).

Simulation results from this analysis can be obtained from the authors.

2.2.3 Actual Changes in Population

Sections 2.2.1 and 2.2.2 discuss changes in the distribution of population in Canada which are caused either by an increase in total population or by a shift in population. Recent experience suggests that a combination of these two factors is at work in Canada. Immigrants, who are now a major component of any Canadian population increase, appear to be settling disproportionately to Ontario, B.C. and Alberta. In addition, there appear to be internal migration to these provinces. Based on the results shown, these population movements would lead to a shift in entitlements from Atlantic Canada, Saskatchewan and Manitoba and an increase in Ontario and B.C. in particular. Since these latter

provinces do not now receive equalization payments the net effect of recent population patterns is to reduce the total entitlements under the program.

The results of a 1 per cent population increase in each of the provinces, are shown in Table 3.

Table 3, indicates that when a province's population increases its own entitlements increase by a large amount. For instance, a 1 per cent increase in the population of Quebec would lead to an increase in its entitlements in 1988-89 of $187 million or 5.5 per cent of its total entitlement. At the same time there would be declines in total entitlements for all other provinces ranging from $1.7 million in P.E.I. to $122 million in Ontario. The table also confirms that when the population of a non-standard province increases there is no effect on the other provinces for a population increase in a non-standard province.

As noted earlier, any net shift in entitlements to Ontario (or indeed B.C. or Alberta) leads to a reduction in actual equalization payments since they do not currently receive equalization payments.

Table 3

Effect of 1 Per Cent Increase in Population
($'000)

	NFLD	PEI	NS	NB	QUE	ONT	MAN	SASK	ALTA	BC
NFLD	23459	0	0	0	-7343	-10414	-1202	-1124	0	-3301
PEI	0	5300	0	0	-1659	-2353	-272	-254	0	-746
NS	0	0	36372	0	-11385	-16147	-1863	-1742	0	-5118
NB	0	0	0	29459	-9221	-13078	-1509	-1411	0	-4145
QUE	0	0	0	0	187297	-121588	-14032	-13118	0	-38538
ONT	0	0	0	0	-121747	214558	-19927	-18630	0	-54731
MAN	0	0	0	0	-13995	-19849	42397	-2142	0	-6291
SASK	0	0	0	0	-13083	-18556	-2141	39777	0	-5882
ALTA	0	0	0	0	-30836	-43735	-5047	-4719	98515	-13862
BC	0	0	0	0	-38472	-54565	-6297	-5887	0	105442

Percentage Change in Total Entitlements

NFLD	2.797	0.000	0.000	0.000	-0.875	-1.242	-0.143	-0.134	0.000	-0.394
PEI	0.000	2.993	0.000	0.000	-0.937	-1.329	-0.153	-0.143	0.000	-0.421
NS	0.000	0.000	4.356	0.000	-1.363	-1.934	-0.223	-0.209	0.000	-0.613
NB	0.000	0.000	0.000	3.820	-1.196	-1.696	-0.196	-0.183	0.000	-0.538
QUE	0.000	0.000	0.000	0.000	5.521	-3.584	-0.414	-0.387	0.000	-1.136
ONT	0.000	0.000	0.000	0.000	-2.877	5.070	-0.471	-0.440	0.000	-1.293
MAN	0.000	0.000	0.000	0.000	-1.760	-2.497	5.333	-0.269	0.000	-0.791
SASK	0.000	0.000	0.000	0.000	-2.860	-4.057	-0.468	8.695	0.000	-1.286
ALTA	0.000	0.000	0.000	0.000	-0.827	-1.172	-0.135	-0.126	2.641	-0.372
BC	0.000	0.000	0.000	0.000	-4.890	-6.935	-0.800	-0.748	0.000	13.401

3. Classification of Variables by Impact on Entitlements

This section reports the results of simulation experiments to classify variables by their impact on entitlements. Detailed simulation results for each variable (including population variables) used in the calculation of entitlements are provided in appendix tables obtainable from the authors upon request.

Each variable affects entitlements through its impact on the revenue tax base or revenue source on a per capita basis. Revenue tax bases or revenue sources can be made up of a large or small number of variables and therefore a particular variable may have a small impact on entitlements even when revenue tax base or revenue source of which it is a member has an important impact on entitlements. At the risk of some repetition, the following sections present analyses of individual variables as well as revenue tax bases.

3.1 Criteria for Degree of Impact from a 1 Per Cent Change in a Variable.

To assess the impact a variable has on the total entitlements of a province a ranking of the impact is required. Five categories of impact: (1) Very High, (2) High, (3) Medium, (4) Low and (5) Negligible are given in Table 4.

Table 4
Criteria for Ranking of Impact

Very High Impact:	Greater than 1% Change in Total Entitlements
High Impact:	Between .1% and 1% Change in Total Entitlements
Medium Impact:	Between .01% and .1% Change in Total Entitlements
Low Impact:	Between.001% and .01% Change in Total Entitlements
Negligible Impact:	Less than .001% Change in Total Entitlements

3.2 Revenue Tax Base

This section classifies entire revenue tax bases by impact. The listing ranks each revenue base in descending order of revenues subject to equalization.

Revenue bases with a "Very High Impact" or "High Impact" on entitlements account for over 90 per cent of total revenue sources subject to equalization:

- Personal Income Tax Revenue Bases, General and Miscellaneous Sales Tax Revenue Bases, and Provincial-Local Propery Tax Revenue Bases have a very high impact on total entitlements and are also the top three ranked revenue sources subject to equalization.

- Eleven revenue tax bases have a high impact on total entitlements. Of these, eight are included in the next ten ranked revenue sources. The revenue bases ranked as high impact correspond to revenue sources that account for over 28 per cent of the total revenue source subject to equalization.

3.3 Input Variable Simulation

The results for revenue bases shown in Table 5 can be broken down further by analyzing the individual variables which are used in the calculation of tax bases. The impact of a 1 per cent increase for the input variables prepared by Statistics Canada are recorded in detailed appendices which are available on request.

The ranking of the input variables is summarized in Table 5. Four variables have a "Very High" impact on equalization payments. Of theses, three are calculated by the Department of Finance and Revenue Canada and include: Basic Federal Tax, simulated tax payments and simulated provincial distribution of income taxes. The fourth "Very High" impact available is Population and is provided by Statistics Canada.

Table 5
Summary Impact of Input Variables

Ranking	Statistics Canada	Non-Statistics Canada	Total
Very High	1	3	4
High	22	5	27
Medium	49	16	65
Low	59	7	66
Negligible	27	21	48
Total	158	52	210

Of the 27 high impact variables 22 are supplied by Statistics Canada of these:

- Population is the single most important variable as an input variable, affecting the calculations of entitlements from all revenue bases;
- Eight of the high impact input variables are used in the calculation of the "General and Miscellaneous Sales Taxes" revenue tax base.
- Five Statistics Canada variables that have a high impact on total entitlements are used in the calculation of the "Provincial-Local Property Tax Revenues" revenue tax base.
- The remaining nine variables are used in the calculation of entitlements from various other revenue tax bases such as non-commercial vehicle licences, alcoholic beverages, gasoline taxes, etc.

3.4 High Impact Variables and Estimation Errors

Twenty-two revenue tax base input variables and population which have a "high" or "very high" impact on equalization payments are supplied by Statistics Canada. From the point of view of this program changes in the scope, concepts, methodology or operational procedures which result in changes in the estimates of these variables would have a potentially material impact on provincial entitlements. In addition to any methodological changes, large statistical estimation errors for any of these variables would be a cause for concern since this would imply the confidence interval around any point estimate could be large enough to affect payments to the provinces.

It would be useful to use the estimated coefficients of variation for high impact variables to further classify them into "high variability" variables. However, a number of variables used are obtained from administrative data and others are not survey based. Hence, for most of the series used in this analysis it has not been possible to obtain these estimates. Instead, for each high impact variable we examined the record of revisions to obtain a "proxy" for estimation error. A number of variables are not subject to revision and, therefore, we are unable to estimate their variability in this way.

Clearly, this is at best an imperfect way of measuring the variability of data. For instance, amongst the variables with high variability is the national accounts estimate of wages and salaries. In the equalization calculations "final" national accounts estimates for wages and salaries are used. These estimates are based on actual taxation data and, therefore, are likely to be firmly based. However, initial estimates of wages and salaries are derived from preliminary indicators which are subject to revision. Because of this difference in estimation methodology between first and final estimates our "proxy" for statistical variability is likely to overstate the statistical risk associated with using this series in the equalization program. Nevertheless, using "proxy" estimates of variability, nine of the high impact variables are identified as having a high variability; four variables from General and Miscellaneous Sales Taxes, four variables from Provincial-Local Property Tax Revenues and the wages and salaries variable from Payroll Taxes.

Table 6 on the last page provides a province-by-province summary of the impact of a 1 per cent increase in the nine high impact/high variability variables.

3.5 Population Simulation

The estimates of population by province have the largest impact of any single Statistics Canada variable. This general issue has been discussed at length in Section 2.2. Additional tables, showing how changes in population affect entitlements from individual revenue bases are available on request.

4. Conclusions

The foregoing analysis of variables that affect equalization payments to provinces suggests that there are only nine variables produced by Statistics Canada which have a high impact on payments and may also have a high variability.

The main conclusion from the analysis is that although there are a number of variables which could **potentially** have a significant impact on equalization payments the actual risks are probably quite small. Nine variables have a sufficiently high variability to potentially have material impact on equalization payments. These are: Wages and Salaries from the National Income and Expenditure Accounts; four variables which affect General and Miscellaneous Sales Taxes - Service Establishment Sales, Investment in Machinery and Equipment (including repairs), Capital and Repair in the Primary Sector, and the Cost of Construction; four affecting the calculation of Provincial and Local Tax Revenues -- Personal Disposable Income, Net Provincial Income at Factor Cost, Residential Capital Stock and Capital Stock-Commercial.

The variables are identified as **potentially** having a significant impact but this does not necessarily imply an actual impact. In particular, the capital stock variables and the national income variables are identified as being both highly variable and having high potential impact because these variables are revised several times and undergo significant changes in successive revisions. However, the estimates used in the actual calculation of equalization payments are the final estimates and these data are not subject to revision.

Population estimates are not included in the list above because they have not been identified as having a high variance. However, population estimates are the most significant of the data prepared by Statistics Canada used in the Equalization Program. Any changes in the distribution of population amongst provinces, whether due to statistical error or to definition and concept changes could have significant impacts on payments under the program.

Immigration has always been a major source of Canada's population growth and, in recent years, immigrants have been settling disproportionately in the provinces of Ontario, British Columbia and Alberta. In addition, there appears to be internal migration to these provinces. Based on the results shown, these population movements would lead to a shift in entitlements from Atlantic Canada, Saskatchewan and Manitoba and an increase in Ontario and B.C. in particular. Since these latter provinces do not now receive equalization payments the net effect of recent population patterns is to reduce the total entitlements under the program.

Table 6
Effect of High Impact/High Variability Variables

Own Province Effect of High Impact/High Variability - Variables Produced by Statistics Canada 1988-89
(Per Cent of Initial Entitlement)

	NFLD	PEI	NS	NB	QUE	ONT	MAN	SASK	ALTA	BC
Investment in Capital and Repair of Mach./Equip	0.09	0.07	0.16	0.14	0.22	0.28	0.19	0.39	0.17	0.67
Capital and Repair, Mach/Equip Primary Sector	0.02	0.02	0.02	0.02	0.01	0.02	0.04	0.15	0.05	0.08
Cost of Construction	0.06	0.06	0.09	0.08	0.14	0.15	0.12	0.26	0.11	0.36
Service Establishment Sales	0.03	0.04	0.05	0.04	0.06	0.07	0.09	0.15	0.05	0.25
Wages and Salaries ex. Supplementary Labour Income	0.04	0.04	0.07	0.06	0.09	0.11	0.09	0.13	0.06	0.28
Personal Disposable Income	0.03	0.02	0.05	0.03	0.15	0.15	0.10	0.10	0.05	0.39
Net Provincial Income at Factor Cost	0.03	0.04	0.07	0.05	0.11	0.10	0.09	0.15	0.06	0.27
Residential Capital Stock	0.15	0.18	0.26	0.21	0.36	0.44	0.36	0.63	0.22	1.14
Capital Stock - Commercial	0.13	0.11	0.24	0.19	0.34	0.35	0.32	0.50	0.24	1.00

Own Province Effect of High Impact/High Variability - Variables Produced by Statistics Canada 1988-89
($'000)

	NFLD	PEI	NS	NB	QUE	ONT	MAN	SASK	ALTA	BC
Investment in Capital and Repair of Mach/Equip	764	121	1,326	1,114	7,316	11,810	1,541	1,788	6,184	5,201
Capital and Repair, Mach/Equip Primary Sector	161	32	135	135	422	719	289	673	1,898	636
Cost of Construction	462	112	773	581	4,617	6,457	957	1,170	4,050	2,849
Service Establishment Sales	270	71	449	346	2,076	3,118	701	698	2,014	2,000
Wages and Salaries ex. Supplementary Labour Income	371	69	585	449	3,165	4,841	732	588	2,131	2,186
Personal Disposable Income	215	39	441	263	4,959	6,126	786	447	2,004	3,101
Net Provincial Income at Factor Cost	281	65	543	416	3,601	4,376	743	680	2,292	2,086
Residential Capital Stock	1,274	310	2,198	1,653	12,148	18,753	2,833	2,872	8,182	8,951
Capital Stock - Commercial	1,071	195	1,973	1,495	11,462	14,728	2,558	2,266	9,033	7,863

STATISTICS OF INCOME DIVISION'S USES OF ADMINISTRATIVE BUSINESS TAX RECORDS: AN OVERVIEW

Jeri Mulrow, Susan Hinkins, Jonathan Shook, IRS
Jeri Mulrow, IRS, R:S, P.O. Box 2608, Washington, DC 20013-2608

INTRODUCTION

The Statistics of Income (SOI) Division is a statistical group within the Internal Revenue Service (IRS). The SOI Division takes annual samples of tax returns, cleans and checks the data, and provides population and subpopulation estimates of tax, income, and other financial items of interest to economists and policy makers. In this paper, we discuss only one of the SOI samples, the annual sample of corporate tax returns.

Before describing some specific aspects of the SOI environment, an overall picture of our current state can be described using the perspective of a dynamic system. The dynamic system model described in Kauffman's The Origins of Order is used to describe biological processes: the evolution of species, the development of proteins, etc. But it can also be loosely adapted as an analogy for the development or evolution of a sample design and information system.

Greatly simplified, this model begins with a space of possible characteristics. In our example, this would include all the sample design options, data collection methodologies, and estimation techniques that we can choose from. For each point in this space, a measure of its "fitness" is defined. In the case of the sample design and estimation, "fitness" could be a general measure of "usability", encompassing all the customers' needs: accessibility, efficiency, accuracy, cost, etc. This results in a "fitness landscape" over the space of possible options. This is usually a "rugged fitness landscape" as it has more than one local maximum; i.e. there is more than one set of viable combinations.

Naturally such a system will migrate towards the peak, the best choice. If it were a static world, the best design and product would be achieved and no further adjustments would be necessary.

But it is not a static world; the landscape changes over time. Properties of the population change; the users' needs changes; budgets change. The computer capabilities and relative costs have certainly improved over the last 20 years and are still changing. There are advances in statistical techniques, in many cases directly related to the increased capacity for fast, inexpensive computing. Therefore the relative fitness of a design will change over time, and it is necessary to constantly evaluate and make adjustments in order to stay at the top of the "fitness landscape". Usually these are incremental changes: increasing or decreasing sample sizes, changing strata definitions, modifying estimation techniques, etc.

Eventually it may happen that the rugged landscape changes so much that incremental changes are not enough. A "long jump across a rugged fitness landscape" is required; one needs to move to an entirely different set of options in order to make needed improvements in the "fitness" of the final product.

The SOI Division could be described as being in the process of making such a long jump, making more than incremental changes, in order to meet the changing needs of our users and taking advantage of changes in technology.

Since the landscape is constantly changing, the jump may not land at the top of the higher fitness peak. But it is necessary to make the jump in order to move in the direction of a significantly better product. Many people at SOI are involved in different aspects of this major shift. But the creative vision and the direction for this long jump are due to the director, Fritz Scheuren.

In this paper, we give some historical background about the SOI environment, describe briefly the current corporate program, and give some examples of the direction we hope to move with a long jump.

1. INFORMATION FROM ADMINISTRATIVE DATA

The Statistics of Income (SOI) program of the IRS came into being soon after the adoption in 1913 of the Sixteenth Amendment to the Constitution and the subsequent enactment of the first modern income tax law, the Revenue

Act of 1916. In that Act, Congress specifically called for the annual publication of statistics by stating that:

> The preparation and publication of statistics reasonably available with respect to the operation of the income-tax law and containing classifications of taxpayers and of income, the amounts allowed as deductions and exemptions, and any other facts deemed pertinent and valuable shall be made annually by the Commissioner of Internal Revenue, with the approval of the Secretary of the Treasury.

The wording contained in the 1916 Act has been repeated, with practically no change, in each major rewrite of the Internal Revenue Code since that time. Consequently, the Statistics of Income Division has published annual tabulations and reports on the operation of the income tax laws for every year since 1916 (Scheuren & Petska, 1992).

The following sections describe the type of data available from corporate tax returns and some considerations for making this information more useable.

IRS Administrative Data

The SOI Division has access to two basic sets of administrative data within the IRS. The first and underlying basis of the IRS administrative data is a vast volume of documents supplied by taxpayers and IRS agents regarding tax liabilities and payments. In 1916, this was a large, but manageable, data set to work with directly. The early SOI reports were population summaries. Now the corporate population is over four million. The majority of this information is still being supplied on paper documents, which fill warehouses. Fortunately the use of electronic data transfer is increasing.

For the administration of tax collecting, the IRS also maintains more accessible, computerized data bases. A nationwide master file system, which has an account for every active taxpayer required to file U.S. income taxes, including a limited amount of information from every tax return filed with the IRS. The data primarily relate to tax consequences: taxes paid and taxes due, amendments to the tax return, refund requests, and information about audits. However, there is also some demographic data and descriptive information such as a change in filing status, which may include a merger or acquisition, a filing of bankruptcy or re-consolidation, address changes, and some information on parent companies and their subsidiary filers. Some of the information is more easily accessible than others, and the SOI Division has not had access to or used much of the administrative data that is available. Just recently, efforts are being made to collect some of this additional information on the master file system based on requests from the Treasury Department.

The master file keeps accounts open for five years after the last activity. Therefore, there are varying amounts of information available for different taxpayers and companies. Companies who have had a lot of activity over the years may have accounts that have records dating back to the inception of the master file system. Others may only have the last five years worth of data or less depending on the date of incorporation. Thus, some longitudinality exists, but the structure is not complete or easily defined.

The master file is maintained basically for keeping track of taxes owed and taxes paid. It is not designed for providing the details, available from the tax return, that are of interest to economists and policy makers. For example, it does not contain balance sheet detail, all the tax credit items, or schedule components. Also, although the information on the master file is generally reliable, certain variables are less reliable than others. Data directly related to tax consequence are extremely reliable, such as, taxes paid. On the other hand, data not related to tax consequences may not be quite as good. Industry code is an example. It is a self reported item which is on the master file but is not checked for accuracy or consistency from year to year.

The primary users of IRS tax data within the Federal government are the Office of Tax Analysis in the Department of the Treasury, the Bureau of Economic Analysis in the Department of the Commerce, and the Joint Committees on Taxation in the U.S. Congress. The master file system, although it has information on every taxpayer, does not have enough detail for the extensive economic modeling performed by these primary users. In order to provide the detail needed for their work, the SOI Division collects additional information from a sample of approximately 85,000 tax returns annually.

Over the nearly 80 years of SOI data collection, the population has changed

dramatically, and with it, the amount and types of data being collected. The following sections briefly describe two basic aspects of the sample data:

- What information is collected from each corporate return, and
- How many and what type of corporations are included.

Qualities of the Corporate Tax Return

In the early years, only a modest amount of economic and tax data was available on SOI's files. For example, the 1917 corporation file consisted of approximately 30 data items covering the population of corporation tax return filers. At that time, the collected data were predominantly used by Treasury Department officials to estimate revenue and to conduct research on tax policy, which are still two of the primary uses of the data.

By contrast, the modern day SOI corporation program collects a wealth of financial, economic, and tax data. For example, the 1991 corporate program collects approximately 1800 data items from a sample of eight of the 1120 Corporate income tax return types and 16 of the supporting forms and schedules. In addition, the SOI foreign studies program collects and disseminates data from three more specialized forms attached to the corporate return having to deal with foreign owned corporations.

Depending on the specific 1120 corporate form type, the corporate tax return generally includes detailed schedules for the income statement, balance sheet, tax computation, cost of goods, dividends, reconciliation of book income to tax return income, and an analysis of retained earnings. Specialized corporation returns such as the Form 1120L, Life Insurance, and 1120-PC, Property and Casualty, also include business specific schedules. For example, the Form 1120L provides schedules for such items as policyholder dividends, increase/decrease in reserves, and policy interest. Similarly, the Form 1120-PC provides schedules to compute premiums earned and losses incurred.

Beyond the data captured from the internal schedules of a Form 1120, many of the attached forms, used to calculate bottom line totals, are also collected. For example, most of the data from the Schedule D, Capital Gains and Losses, used to calculate the amount of capital gains reported on the income statement, are collected. Similarly, most of the detailed data available on the Form 4626, Alternative Minimum Tax--Corporations, the Form 4562, Depreciation and Amortization, and the Form 3800, General Business Credit are collected by SOI. See Department of the Treasury, Publication 16 (Pub. 16) for a complete listing of the forms, schedules, and number of data items abstracted for the 1991 program.

Much of this detailed corporate tax data, whether viewed in aggregate or in microdata form, is absolutely essential to assess the efficacy of current tax law and estimate the effects of proposed changes in the law. Additionally, economic and tax items such as corporate profits, total tax liability after credits, cash distributions, total depreciation, interest received, and interest paid are used extensively by the Commerce Departments's Bureau of Economic Analysis in its National Income and Product Accounts.

Beyond economic and tax data, a limited amount of organizational information is available for collection. For example, information on consolidated/non-consolidated, initial or final filing, 50 percent or more foreign ownership, and number of shareholders are routinely collected. However, parent/subsidiary relationships and merger/reorganization activities are currently not well captured by SOI although this information is of interest. Tax law requires the parent corporation to complete Form 851, Affiliations Schedule, identifying its affiliated corporations included in the consolidated return. However, large consolidated returns frequently include hundreds of subsidiaries, making the data capture particularly onerous. Consequently, Form 851 data are not currently collected by SOI. Reorganization, merger and acquisition data are frequently provided on attachments describing the taxpayer actions and the entities involved. As mentioned, SOI does capture the taxpayer reported indication as to initial, final, or merging status of the filer. However, data are not currently collected from taxpayer provided attachments as to the entities involved.

One of the most critical data items in the SOI corporate file is the industry code assigned to a given corporation return. The industrial classification used by SOI conforms to the Enterprise Standard Industrial Classification (ESIC), which classifies companies, rather than individual establishments. It follows closely the detailed Standard Industrial Manual (SIC) which is designed as a means of classifying

establishments. Both, the ESIC and SIC are publications issued by the Office of Management and Budget. Some departures from the ESIC system were made by SOI for financial industries in order to reflect particular provisions of the Internal Revenue Code. See Department of the Treasury's Publication 647 for an example of the relationships between the SOI industry code and the SIC and ESIC codes.

Although SOI's classification is designed to apply to the company rather than the establishment, its application to corporation income tax return statistics has limitations. A return is classified by industry based on the activity accounting for the largest percentage of total receipts. This means, large corporations with diversified activities are included in only one industry, even though many of their business operations are unrelated to that industry. Consequently, statistics for an industry may be understated by amounts reported by corporations whose principal activity lies elsewhere, or overstated by amounts reported by corporations who have substantial business operations in other industries.

A final consideration concerning corporate tax return data is its comparability (or lack thereof) to publicly available sources. In general, the largest caveat must be that accounting requirements for tax purposes frequently differ significantly from Generally Accepted Accounting Principles (GAAP) employed for financial statement purposes. Differences in statutory and book accounting for items like depreciation, installment sales, and amortization of goodwill can lead to significant differences in reported profits. Additionally, for tax purposes, companies may file a consolidated return but in filing their financial statements, they may report their results separately. These data differences make any effort to replace missing tax data with publicly available financial data, a somewhat dubious procedure. The SOI Division makes no attempt to reconcile these differences and collects only tax data.

Data Collection: From Census to Sample

Along with the increased complexity of the data, in terms of the number of different forms and the amount of information per return, the total number of returns has also increased dramatically since 1916. Up through the early 1950's the SOI statistics were based on a census of all returns filed with IRS. Since then the statistics have been based on samples of returns. The change from a census to a sample might be considered a "long jump" rather than an incremental change.

The sample size has stayed relatively constant or decreased in size since 1951. The overall sampling rate has decreased from nearly 40% in 1951 to the current level of just over 2%. The sample size is approximately 85,000 returns out of a population of nearly 4 million active corporate returns.

The dramatic decrease in the relative size of the sample over time has caused the structure of the sample design to become more complex and to be increasingly dominated by the larger corporation tax returns. The population of corporation returns is highly skewed, with a relatively few large corporations accounting for a large percentage of the total money amounts. For example, in 1990 the largest 0.5% of the corporations contained 89% of Interest Income, 82% of Long Term Capital Gains, and 98% of the Foreign Tax Credit. Obviously these largest returns must be in the sample in order to make reasonable estimates of populations totals. Of the 85,000 sampled 1990 corporations, approximately 20,000 were these largest corporations which are selected with certainty.

An essentially fixed sample size, an increasing number of returns in the population, and the need to sample all "large" returns, result in dramatic reductions in the sampling rates for the smaller size classes. If SOI's users were only interested in estimates of annual, total amounts, this might not be a concern. But the users have other needs, including interests in subpopulations of smaller corporations. These needs must currently be met with the remaining 65,000 in the sample. Therefore, the sample design has become increasingly more complex over time. The sample design is currently a highly stratified probability design with 48 strata defined by tax form type and size of corporation. A description of the 1990 sample of corporate returns can be found in Publication 16.

Returns are initially selected for the SOI sample from the IRS master file system. They are obtained for statistical processing, where the data is put into an electronic format, cleaned, and checked for inconsistencies. In recent years, SOI has made tremendous production gains in this area due to a modernization of the computer systems used to collect the data. Instead of weeks to obtain a clean record it now may take

less than an hour depending on the size of the return.

Due to substantial penalties for misreporting, the detailed income and expenditure data on tax returns are generally regarded as more reliable than similar survey data. Even so, SOI goes to great lengths to protect against non-sampling errors, such as those due to taxpayer or data entry errors. Extensive on-line tests for consistency and reliability are made based on the structure of the tax law and the improbability of various data combinations. The SOI Division also has a rigorous quality review program.

Missing data are not much of a problem. Typically, less than one percent of the data are missing or inconsistent. Missing items can sometimes be obtained through telephone or written follow-ups, but more often imputation procedures are employed. For item imputation, prior year information is used, and variations on hot deck and nearest neighbor methodologies have been used with current data.

Longitudinal Data

Because of the interest in longitudinal studies and in improving estimates of change for various economic variables, SOI has designed the sampling selection process so that there is an overlap of sampled returns from one year to the next. To do this, SOI began using the Taxpayer Identification Number (TIN), an individual's Social Security Number (SSN) or a corporation's Employer Identification Number (EIN), as a basis for sample selection in 1968. This procedure allows for overlap of companies in year-to-year samples while retaining randomness within a given year. It is not a panel per se since SOI does not guarantee that any particular company will be in the file over several years. However, longitudinal data are available for many of the corporations in the sample, particularly the larger ones. The overlap in samples decreases over time, and the smaller a corporation is, the less likely SOI is to have information for it several years in a row.

The procedure using the Taxpayer's Identification Number has changed over the years. In the corporate program, from 1968 to 1978, random digits were selected in specific positions of the EIN. The corporate sample was actually a cluster sample under this type of sampling procedure. However, there was a concern that the assignment of EINs to corporations might cause an appreciable intra-cluster correlation, since EIN's have some built in structure to them.

Beginning in tax year 1978, SOI decided to use a transformation of the EIN rather than the EIN itself. This method was first proposed and studied at the Bureau of the Census by B.J. Tepping. The general formula for computing the transform is: $Y = c * X \pmod p$ where Y is the transformed number and equals the remainder when $c * X$ is divided by p; X is the EIN; p is a large prime number; and c is a constant which is relatively prime to the number of subsets the population is partitioned into.

This transformation accomplishes two important purposes: (1) the transformed number is pseudo-random, and (2) the transform, corresponding to a given EIN, is always the same. The companies in the sample from a given stratum will be a random sample. If p and c remain constant over the years then (1) the sample is self-adjusting for births and deaths and (2) there will be a large overlap in the sampled returns from year to year. For example, beginning with the 65,469 unique, non-1120S returns in the 1987 sample, 36,002 are also present in the following three years' samples. To properly use these longitudinal data, however, one must take into account the sample design and the selection mechanism. Basically, the larger corporations and corporations that increase in size over time will be over-represented, and corporations that decrease will be under-represented.

This sample selection mechanism also provides a subset representing a simple random sample from the 1987 population, that should be selected in every following sample. This subset is defined by taking corporations with EIN's that would be selected at the lowest sampling rate. It can be used to estimate 'birth and death' rates for some classes of corporations. However, it is a small sample containing, 6,165 corporations with only 3,623 corporations with data for all 4 years.

2. LOOKING TO THE FUTURE

Historically, SOI, like many other government statistical agencies, has maintained a strong descriptive or enumerative focus rather than an analytic or inferential one (Norwood, 1989.)

Changes in the "fitness landscape" have resulted in the need for SOI to consider major changes in the process and the product. More

information and more accessible data are required in the following ways:

- More timely data
- Simpler data sets
- More available micro-data (masked to maintain confidentiality)
- More information

In the following sections, we present some examples of major limitations of the data base. We outline the accessibility of the data and we include our current and future plans to provide better products for our users, by expanding our statistical, graphical, and processing tools.

Timeliness

Because the SOI corporate data are used to analyze the effects of current tax policy, to estimate effects of proposed policy changes, and to measure and analyze the U.S. economy, timeliness is important for the efficacy of these data. Unfortunately, timeliness is a problem with the administrative data based on tax returns.

For example, the 1991 corporate tax returns will generally be filed between June 1991 and March 1993. Most of the largest or most complex returns will be filed after September of 1992. Sample selection continues through June of 1993. After sample selection, there is still the process of retrieving selected tax returns, abstracting the data, checking, and cleaning the data file. In rare cases, the returns may not be available to SOI at all, because another branch of the IRS is using it or the corporation has been granted a special exception by the IRS due to special circumstances.

The data base for the 1991 corporate activity would typically not be complete until October, 1993 and the tabulations would not be published until December, 1993 or January, 1994. The data collection process is being changed to dramatically shorten the time to enter and check the data, and provide more information with fewer errors. This should move the production of the final data base and the final estimates up by four to six months over the existing system.

But this is not enough. The users need estimates on demand, or a continuum of data over time. This is our ultimate goal, which requires major changes in our sampling perspective and in estimation techniques. In particular, increased use of model-based estimates will be required. As a starting point, an incremental change was made beginning with the 1990 program; SOI is providing advance data estimates and an advance data file before the sample is complete. For the 1991 data, advance data were available by May 10, 1993.

Because the "late" returns are not like the "early" returns, the properties of the late returns need to be modeled. The first models used have been simple ratio adjustments, based on prior year results. But in the future we are considering estimating the propensity to be in the "early" sample versus "late", to be used to weight the advance sample.

Since the distribution of many of the economic and tax variables is extremely skewed, a very, very small number of the corporations account for a large percentage of the total amount, making these entities extremely influential to the data base and the tabulations. If even a few of these largest corporations are missing, the resulting tabulations, and possibly conclusions, may change dramatically. Some of these critical corporations will not be in an advance sample. Since these largest corporations are so influential and, at least for certain variables, so unstable from year to year, modeling these records for the advance data is not a reasonable option (Hinkins & Mulrow, 1992). Therefore, a small survey has been added to the administrative data base. For these critical corporations, if at the time of the advance data, these tax returns have not been filed or have not been available for statistical processing , a short questionnaire is sent directly to the corporation requesting information on approximately 20 tax items. In this way, at least some of the current information for these corporations is obtained and used in the file.

The move from providing only one final product to producing a continuum of data files and estimates, with associated measures of reliability or confidence, is a major component of our "long jump". It requires more than incremental changes; it requires a shift in perspective and in the tools needed.

Simpler Data Sets

The fact that the sample design is fairly complex is both a strength and a weakness of the data. There are many advantages to the complex sample design. It generally accommodates the many needs of the primary users within a world of fixed and sometimes shrinking resources. It allows our primary users to make more accurate estimates and

projections.

To correctly use the microdata resulting from such a complex sample requires a detailed understanding of the design, how it changes over time, and the mechanism for sample selection. For a less sophisticated or more casual user, i.e. one who is not intimately involved with the sample design and the data structure over the years, it is very difficult to correctly take advantage of this information. It is even likely that such a user will draw incorrect conclusions from the data because properties of the sample design were ignored. For example, as mentioned earlier, if the sample overlap from year to year is used to estimate change, the estimates will be biased unless the correct adjustments are made for the sample design and selection mechanism.

We cannot simplify the sample design, but we are considering methods of providing users with alternative data files that could be tailored for their specific needs and be equivalent to a simple random sample. This would make the information in the corporate sample more accessible.

Disclosure and Public Access

Finding ways to obtain wider public access, while protecting taxpayer's confidential information, is considered extremely important to SOI. Several outside users have expressed extreme interest in tax microdata for running economic models. However, tax return data are protected by law from public scrutiny, and strict procedures govern the handling of returns and computer tape files containing such information. Even after specific identifiers, such as name, address, and EIN are removed, the remaining tax return data may still be confidential. It may be possible to identify firms by linking taxpayer information with publicly available data; access to microdata such as Standard and Poor's COMPUSTAT and Dun and Bradstreet data compounds the problem of releasing business tax data (Spruill, 1984). It is difficult to address disclosure concerns while preserving data utility (Greenberg, 1990).

Currently, SOI has a public use file for individual tax data, but nothing comparable for the business tax data. However, a "blurred" microdata file for specific use by the Texas Legislative Budget Board has been developed by KPMG Peat Marwick in cooperation with SOI. The microdata file is actually a composite of information from 11 different data sources including the SOI corporate sample. Although many disclosure concerns have been addressed in the blurring procedure, SOI is still investigating the data further. The question of data utility still remains. Univariate statistics have been preserved throughout the process, but no analysis of the underlying correlation structure has been undertaken. Thus, the blurring procedure is a good start towards producing a corporate microdata file, but still has a ways to go. The SOI Division will use this as a foundation for future projects and research in the area of developing a business public use file.

More Information

For SOI, the focus has been to strictly gather information available on the tax return only. Not much emphasis has been put on trying to obtain information from other sources or to expand the data beyond the initial narrow bounds. For example, the SOI data are collected from pre-audited returns and no attempt is made to update the data with any further information the IRS may receive after the initial filing of the return. The data are just a snapshot in time. For some users, this is a limitation of the data.

Another limitation for our users is that the population unit of interest does not always correspond to the corporate unit represented by the tax return. Statistics Canada's work on a Business Register (Armstrong,1990) shows both the great usefulness and importance of documenting the relationships between corporate units, and also the difficulty of keeping track of these relationships.

This is a problem inherent in the tax return data available. If a tax return corresponds to more than one unit of interest, it is not necessarily possible to extract the information corresponding to the parts. It would be useful to at least know something about these relationships, and how they change over time. That is, the user would like to know that the 1989 tax return corresponds to corporate unit X and that the 1990 tax return corresponds to corporate units X and Y.

The plans for improving information of this type involve several processes. The corner stone of this process is being able to explicitly define the corporate units of interest, and the rules for defining relationships between them. Once we have at least a simple set of definitions, the more straight forward part is to see what information is available for tracking these relationships.

Finally, a system must be devised for using the available data to automatically track and make this information accessible.

We don't have the cornerstone yet, but we are working on it. The definition of the corporate units of interest (establishments, companies, corporations, etc.) depends on our users. The priority with which we work on this also depends to some extent on our users. We are currently discussing with economists just what kinds of definitions and relationships are of interest.

A Changing IRS Environment

Some fairly dramatic changes in the tax collection environment appear to be coming. Not surprisingly, some will aid us and some will make our task more difficult.

For example, there is a move to an increased use of electronic filing. This has begun in the Individual tax filing (1040's) and will expand to the corporate area sometime in the future. The advantages of electronic filing, both to the taxpayer and to IRS, are many and obvious. For SOI's collection of the administrative data, the advantages could be great. There should be more data available, sooner, with fewer errors. The information can be available to more than one IRS user, so there should no longer be a problem of SOI not being able to get the information because another part of IRS has the return.

But this system cannot eliminate all errors, and with electronic filing the taxpayer's original data are no longer available as a source to fall back on. If inconsistent information is found on a schedule it may be because the taxpayer made an error or because there was an error in transmission. Currently we can go back to the original tax return to check the information. In the future, this may no longer be an option. Also, it will require more care and control so that data are not "electronically lost".

There may also be a movement afoot to reduce the amount of data that the taxpayer must supply. Instead of supplying all the information to substantiate the tax calculation, the taxpayer may only be required to initially send in the major items of concern. The taxpayer could be required to send in additional information after the initial information was reviewed. Obviously, this would seriously effect the type of administrative data available and necessitate SOI using a very different approach for collecting a sample of complete administrative data.

3. SUMMARY AND ACKNOWLEDGEMENTS

There have been significant changes in our "landscape", both in terms of our users' needs and in terms of available technology, making a "long jump" necessary. A common difficulty in such a situation is focusing all of the separate parts into a complete picture; seeing the need to make a major shift. The SOI Division is starting to see this need and lay the ground work for such a change.

Credit and praise should be given to those inside the division who have recognized the need and are willing to meet the challenge of making a "long jump". Thanks should be given to those outside the division, our primary users, for supporting and encouraging our efforts in these times of change.

References

Ahmed, Y., Scheuren, F. (1991). The U.S. Statistics of Income Program: Issues and Challenges, *Bulletin of the International Statistical Institute*, No. 48.

Armstrong, G. (1990). An Overview of the Generation of Statistical Entities. Working paper, Statistics Canada, Ottawa, Canada.

Dalenius, T. (1988). Controlling Invasion of Privacy in Surveys. Stockholm, Sweden: Statistics Sweden.

Fellegi, I. (1987). Some Mathematical Statistical Research Problems in Statistics Canada. A paper presented at the Annual Meeting of the Statistics Society of Canada. Quebec, Canada.

Greenberg, B. (1990) Disclosure Avoidance Research at the Census Bureau, *Paper presented at the 1990 Annual Research Conference*, Arlington, VA.

Harte, J.M. (1986). Some Mathematical and Statistical Aspects of the Transformed Taxpayer Identification Number: A Sample Selection Tool Used at IRS, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 603-608.

Hinkins, S., Scheuren, F. (1986). Hot Deck Imputation Procedures Applied to a Double Sampling Design, *Survey Methodology*, 12, 181-

196.

Hinkins, S., Scheuren, F. (1989). Evaluating Sample Design Modifications: Balancing Multiple Objectives, *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 654-658.

Hinkins, S., Mulrow, J., Collins, R. (1990). Design and Use of an Imbedded Panel in the SOI Corporate Sample, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 760-765.

Hinkins, S., Mulrow, J. (1992). Preliminary Estimates from the 1990 SOI Corporate Sample, *Statistics of Income Turning Administrative Systems into Information Systems*, 115-118.

Jones, H., McMahon, P., (1984). Sampling Corporation Income Tax Returns for Statistics of Income, 1951 to Present, *Proceedings of the American Statistical Association, Section on Survey Research Methods*,437-442.

Kauffman, S.A. (1993). *The Origins of Order*. Oxford University Press.

Mulrow, J. (1990). Description of the Sample and Limitations of the Data, *Statistics of Income 1987, Corporation Income Tax Returns*,Publ. 16, 9-15.

Mulrow, J., Woodburn, L. (1990). An Investigation of Stratification Errors, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 754-759.

Norwood, J.L. (1989). The Influence of Statistics on Public Policy. *Proceedings of the Symposium on Statistics in Science, Industry and Public Policy*, 30-42.

Petska, T., Scheuren, F., Wilson, B. (1992). The Statistics Corner: The Statistics of Income Program of the Internal Revenue Service, *Business Economics*, 56-60.

Petska, T., Scheuren, F. (1992). The Statistics Corner: Information From Tax Returns: The Statistics of Income Program of the Internal Revenue Service: Part 2, *Business Economics*, 66-69.

Scheuren, F., Petska, T. (1993). Turning Administrative Systems into Information Systems, *Journal of Official Statistics*, Vol. 8, No. 4, 109-120.

Spruill, N. (1983). The Confidentiality and Analytic Usefulness of Masked Business Microdata, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 602-607.

Sunter, A.B. (1987). Implicit Longitudinal Sampling from Administrative Files: A Useful Technique. *Journal of Official Statistics*, 2, 2.

Tepping, B.J. (1969). Memo to Mr. J.F. Daly, dated January 15, 1969, U.S. Department of Commerce, Bureau of the Census.

Triplett, J. (1991). The Federal Statistics System's Response to Emerging Data Needs. *Journal of Economic and Social Measurement*, 17, 3-4, 155-177.

Westat, Inc. (1974). Results of a Study to Improve Sampling Efficiency of Statistics of Corporation Income. Bethesda, MD (unpublished technical report).

U.S. Department of Treasury (1992). *Statistics of Income - 1990 Corporation Income Tax Returns*, Publication 16, Internal Revenue Service.

U.S. Department of Treasury (1991). *A General Description of the Corporation Source Book*, Publication 647, Internal Revenue Service.