

**Making Economic Time-Series
Available to Users of Micro-
Computers in Ireland**

John Fitz Gerald

Special Article

in

QUARTERLY
ECONOMIC
COMMENTARY

October 1987

T. J. BAKER
S. SCOTT
R. A. STOREY





MAKING ECONOMIC TIME-SERIES AVAILABLE TO USERS OF MICRO-COMPUTERS IN IRELAND

by
John D. FitzGerald*

November 1987

1. Introduction

The major raw material used by economists who engage in applied research or analysis is the vast range of economic time-series which are published by many different bodies in all developed countries. The very range of sources and definitions gives rise to many problems before any consideration is given to matters of economic theory. "Datagrubbing" has traditionally taken a major part of the time devoted to any individual research project. With many economists engaging in overlapping areas of research and analysis there has been considerable duplication of effort in the past in developing suitable sets of data. Even if the producers of the raw material, economic time-series, do not see it as their duty to produce consistent data in suitable machine readable formats (can be read by a computer directly without retyping) covering a reasonable span of years, there is clearly an advantage to economists in co-operating in this onerous task.

This article examines the range of economic time-series which are available in computer databases or databanks in Ireland and considers how best these data can be made available to users of micro-computers. Section 2 of the paper sets out the background to the development of these databases and Section 3 describes their current scope and contents. Section 4 discusses the future development of databases. Section 5 examines some technical considerations on how best to access these data and Section 6 presents proposals as to how these data, currently only available on one mainframe computer, could best be made available to users of micro-computers in Ireland.

2. Background

The single biggest impetus to the creation of computer databases in Ireland came from the growing demands of macro-economic modellers for data in the late 1970s. The Central Bank and the Department of Finance led the way and by 1980 they had developed a very substantial database of annual time-series for Ireland. Because of the economies of scale in such work the task of developing and maintaining the database was undertaken on a co-operative basis by these two institutions. In the more recent past the ESRI has taken over from the Central Bank in helping to maintain this database, generally referred to as the Department of Finance/ESRI Databank. Partly to facilitate this work

*John D. FitzGerald is a Senior Research Officer with the Economic and Social Research Institute.

An earlier version of this paper was presented at the Irish Economic Association Conference at Kilkenny, May 1987.

the Central Bank, the Department of Finance and the ESRI agreed in the mid-1970s to share a common computer and common software. The notation and structure of the databank itself grew with the needs of the modellers using the data.

Because this database developed at a time when communications between computers were relatively underdeveloped, it was necessary to centralise it in one computer centre (the Civil Service Computer Centre, CCS) which could provide simultaneous interactive (instantaneous) access to all the users. In the late 1970s, this database was made available to all users of the CCS computer. This included the bulk of economists and analysts working in the State sector: the ESRI, the Department of Agriculture, the Department of the Environment and the Central Bank. However, this body of data was not readily accessible to researchers elsewhere, in particular those in the Universities and the private sector. The Central Bank, from the mid-1970s, has produced a folder of Irish economic statistics which contains quarterly and monthly data on a seasonally adjusted basis. The Department of Finance has produced a number of issues of a printed version of a subset of the data available in its databank (see Hurley and McQuaid, 1986; and FitzGerald, Keegan, McQuaid, and Murphy, 1983). However, this was of limited use to researchers and other analysts whose primary requirement was for data in a machine readable format. The Department of Finance and the ESRI have attempted to overcome this problem by making available, on an *ad hoc* basis, tapes of part of the database to researchers in the universities. However, problems of compatibility between different computers and computer programmes made this difficult. It also proved expensive because the access procedure was handled on a once-off basis with each request being separately processed.

The Department of Finance database itself has expanded steadily, with the ESRI and the Central Bank contributing additional data to it. The impetus for the development of this database has come from the research needs of these organisations. The latest version has undergone a major overhaul in the ESRI to meet the needs of the Medium Term Model.

A further milestone occurred just over a year ago when the Irish CSO made available on the CCS computer a database, named EOLAS, containing some of their most frequently used published data. This new database is being constantly expanded by the CSO. Access to this new database was given to all users of the CCS computer on a pilot basis. More recently, the CSO have made available a sample diskette containing key series from this database and some of these data are even available on one Dublin computer bulletin board (a computer which can be contacted by other suitably equipped computers using the dial-up telephone system). Together the Department of Finance/ESRI databank and the CSO databank (EOLAS) cover the bulk of the data for Ireland which are used regularly in economic research and analysis.

The interest of those engaged in economic research in Ireland is not confined to data on the Irish economy. Those involved in forecasting obviously have a need for current data on developments in the outside world. Those working in many other areas have a need for time-series data for other countries as an input into work on the Irish economy or as an essential

ingredient of work on other economies. Work on the provision of such databases for other countries has proceeded much further than in Ireland. However, these databases, such as those maintained by the OECD (Organisation for Economic Cooperation and Development), the EUROSTAT (Statistical Office of the European Communities), and the UK CSO, are still not very accessible to Irish economists and, possibly more important, direct access may be very expensive. With this need in mind the Department of Finance first mounted five years ago a substantial body of the data produced by the OECD. The range of data was further expanded last year when the organisation of this database was revised (Costigan, FitzGerald, McQuaid and Redmond, 1986). It currently contains approximately 40,000 series which are available interactively and a further 100,000 which are available on tape. These data cover the bulk of the OECD's statistical publications.

3. Description of Databases Available

The databases currently available fall into three separate categories:

- (i) Data for Ireland maintained in the CSO Databank.
- (ii) Data for Ireland maintained in the Department of Finance/ESRI Databank.
- (iii) Data available from the OECD. Many of these series are made available on the CCS computer and additional data are available on tape.

These three databases are considered in turn.

(i) *The CSO Databank (EOLAS)* for Ireland at present contains approximately 90 data files, which in turn contain about 8,000 time-series and 3,500 cross-sectional variables. It covers the main areas of statistics published by the CSO. The areas that are at least partially covered at the moment are:

- Census of Population
- Labour Force Survey
- Labour Costs Survey
- Household Budget Survey
- Unemployment
- Quarterly Industrial Employment, Earnings and Hours Worked
- Monthly Industrial Production and Turnover
- Census of Industrial Production
- Quarterly Consumer Prices
- Monthly Wholesale Prices
- Monthly Agricultural Prices
- Monthly Retail Sales
- Building and Construction
- Monthly Trade, Transport and Tourism
- National Accounts
- Balance of Payments
- Quarterly Births, Marriages and Deaths
- Life Expectancy Tables
- Census of Agriculture
- Agricultural Data

This database is fully documented and the documentation is accessible while using the computer (on-line). It is updated on a regular basis; the majority of the short-term series are updated within 1 hour of release. The CSO will shortly market a set of 8 diskettes (floppy disks) containing a fairly comprehensive range of time-series for a number of these subject areas.

In using these data, experience suggests that the CSO has maintained its normal high standard of accuracy in transferring the data from its internal files to the computer database. However, these data suffer from all the problems which affect the CSO data published in more traditional format. There are discontinuities in some of the series. Most of these are documented in a computer file. Certain key series cover relatively short periods and in certain key areas there has been little attempt to link them to provide continuous time series over a reasonable span of years. For example, the National Accounts data begin in 1970 though data are available in different printed formats going back to 1947. Similarly the monthly trade data begin in 1972 and revisions which have been made to the annual totals have not been included in the monthly figures. As a result the monthly series do not sum to the correct annual figures. There are also discontinuities in these trade data due to problems with changes in classification. All of these problems are not related to the way EOLAS is maintained. They are carried through from the traditional format CSO publications.

These caveats should not detract from the major development which the CSO Databank represents. As time passes the problem of the short time horizon of many series will be reduced by the addition of new data. The availability of the databank will facilitate the CSO in its own work. However, unless the CSO has the resources and is willing to devote them to providing linked series in certain key areas, these data will continue to pose problems for users. While making the job of users easier, the ready access to data in machine readable form will not eliminate the need to examine these data carefully before using them.

(ii) The Department of Finance/ESRI Databank for Ireland grew out of the needs of macro-economic modellers. The range of data available reflects this interest. They are almost all annual time-series and are presented in a National Accounts framework. The complete contents of National Income and Expenditure Tables A.1 to A.22 and A.27 are available. These data have in all cases been carried back to at least 1960, and in many cases to 1958 or 1953. Where discontinuities exist they have been adjusted to produce consistent series by linking relevant sets of data. In the case of the disaggregated data on consumption, the data are maintained on both the old basis and the new basis used in the most recent publication. This ensures availability of consistent data. With the CSO's co-operation, this approach will be adopted in the case of any further changes in definition.

The databank also includes all the data for Ireland published, with a long lag, in the EEC National Accounts together with some additional data, especially for investment, which are published in UN and OECD National Accounts publications. (For example a breakdown of investment by the sector of the economy undertaking it.) These data are supplied to the relevant bodies by the CSO but not, as yet, published elsewhere in Ireland. They generally

become available in this database a number of months after the publication of National Income and Expenditure, a year to two years before they become available elsewhere. However, users should be aware that some of the more detailed disaggregations are considered by the CSO to be of a lesser reliability than the more aggregated data. The CSO has in fact reduced the level of disaggregation in the sectoral analyses of GDP pending development of more reliable sources.

In addition to the National Accounts data, there is a range of series covering certain categories of public expenditure, tax revenue and the national debt. All these series are designed to be consistent with the National Accounts and are not easily related to the data published in the Government's Estimates. The rates of indirect taxes and direct taxes are covered in very considerable detail. Other series which have been needed for economic research in the Department of Finance and the ESRI include a limited range of Labour Force data, obtained from the CSO, and some monetary data supplied by the Central Bank. However, the monetary data are particularly weak as a result of problems obtaining consistent data for long periods due to major definitional changes in the early 1980s.

Only a very limited range of data on a subannual basis is maintained in this databank. The range of cross-section data is also very restricted. The database includes the 1975 Input-Output Table for Ireland as published by the CSO, together with additional transformations undertaken by Murphy (1984) with the help of the CSO. Some income distribution data drawn from the Revenue Commissioners' Accounts are also included for isolated years.

Because of the very limited resources available, the data have not undergone as rigorous a checking process as have the CSO data. As a result the user should check the data before use and report any errors to those responsible. Experience indicates that the series derived directly from the National Accounts are extremely reliable. A range of different cross-checking procedures have been built into the generation process which ensures that the data conform exactly to the data published by the CSO. Errors occur most frequently in series which have been constructed from a range of sources. To facilitate users the procedure employed to generate the database is stored in the database and is available to all users.

The approach taken by those maintaining this large database is rather different from that of the CSO. The first priority is to maintain consistent series over a long period. In many cases this may involve merely rebasing series. However, the frequent minor changes in definition adopted by the CSO have necessitated much additional work. In all cases the CSO has provided extensive help in linking these series. In particular the National Accounts section of the CSO has provided much additional material to help in this exercise. However, this does not mean that the CSO would necessarily approve of the approach taken in each particular case.

This databank is continuously developing in line with the needs of researchers in the Department and the ESRI. By the end of the year a further body of data, compiled by M. Ross of the ESRI, on the total costs of public employment will be included in it. These series will begin in 1947/48 and will run up to 1985. It will also have benchmark figures for 1926/27 and 1938/39. As

a result, it will make available consistent and accurate time-series extending in many cases for 40 years. These data will cover wages, salaries, pensions, and other labour costs of Central Government and Local Authorities. Separate series will be available for each Department.

(iii) The Database of OECD data maintained by the Department of Finance is the single largest economic database in the country. The vast bulk of the data published by the OECD in traditional format is supplied to the Department of Finance on tape. These data normally arrive in the Department substantially before the date of publication of the document in book form. In the case of the more frequently used publications their complete contents are made available interactively in the database. In total these amount to over 40,000 annual, semi-annual, quarterly, and monthly series. They are updated regularly by the OECD. As the vast range of data is quite intimidating to first time users the best approach is to consult the traditional format publications to familiarise oneself with what is available. The range of data currently available interactively in CCS is as follows:

OECD Economic Outlook: Most of the data in this publication, together with historical series and forecasts up to one year ahead for a wide range of variables are available. Many of these data are not actually published in the Outlook. They include National Accounts aggregates for each OECD member country and groups of members, together with data on trade, employment, exchange rates, and interest rates.

OECD Main Economic Indicators: This publication covers a wide range of short-term indicators for member states. The data are predominantly monthly in periodicity. They include data on business surveys and cyclical indicators not published in the traditional format.

OECD Indicators of Industrial Activity: These data include indices of industrial output, output prices, and orders, broken down by detailed sector for each member state.

OECD Labour Force Data: The contents of the annual and quarterly publications covering population, labour force, employment, and unemployment.

OECD Capital Stock Series: A limited range of series is available for 12 countries, excluding Ireland.

The contents of the OECD National Accounts Volumes and Monthly Trade Series are available on tape. Because of the vast range of data on these tapes and the number of tapes involved, they have not been extensively used to date.

Full details of all of these sets of data are given in the appendices of the ESRI Technical Paper *Database Access Using TROLL on the CCS Computer* by Costigan, FitzGerald, McQuaid and Redmond, 1986.

4. Future Development of Databases

The range of data in computer databases available to economists using the CCS computer has expanded rapidly in recent years. However, there are a number of areas where economists would like to see further developments. The most obvious gap is the absence of a significant body of monetary data. The Central Bank collects a considerable volume of such data which it publishes in its monthly and quarterly publications. However, there is no sign of its making these data available to any outside user in a machine readable format.

In addition these data suffer from serious problems due to changes in definitions in the early 1980s. When the Central Bank eventually comes to make this database available it is to be hoped that it will make an effort to provide linked series going back before 1980. The Central Bank also currently receives the IMF Financial Statistics on tape each month. Until now, problems with software have prevented wider access to these very useful data.

The Department of the Environment is currently developing a sophisticated database of series on the building sector using TROLL on the CCS computer. If at some stage in the future it wished to make some of these data available to the wider public, the technical problems would clearly be the same as for the other databases discussed above.

Finally, the Department of Energy has a major EEC contract to develop a database of energy data. However, to date, this project has proceeded rather slowly. Its final form has not been decided. Whatever form it eventually takes, it is important that it takes into account the needs of potential users elsewhere in the public sector and in the wider academic community.

The CSO obviously has a key role in the provision of computer databases for economic analysts. Until recently it was rather slow to develop its own database and it has limited experience of the needs of users. Its approach has, so far, not been very satisfactory from the point of view of researchers due to the relatively low weight placed by the CSO in the past on the provision of consistent series over long time periods in certain key areas (e.g. the National Accounts). For other analysts who use shorter data series this may not pose too serious a problem. Its position and role may preclude the CSO from eventually superseding the work of the Department of Finance and the ESRI in developing and maintaining such a database of series for use in macro-economic research. However, the advances which it has made in the recent past, in spite of constrained resources, are very much to be welcomed, and it must be encouraged to play a central role in setting standards for the development of databases by other public bodies.

5. Technical Access Problems for Users of Micro-Computers

This section considers the format in which the data may be distributed to users who do not have access to the CCS computer. (The method of accessing these different databases on the CCS computer is dealt with in detail in the Technical Paper referred to above.)

At present it is relatively easy to transfer either a limited subset or all of the data in the database maintained by the Department of Finance and the ESRI to other users who use the TROLL computer package. However, the only other computer centre where TROLL is currently available is UCC. As an alternative, the TROLL computer package has been altered to facilitate the transfer of data to other programmes on an IBM mainframe (See Costigan *et al.* 1986). However, this procedure requires each user to have access to an industry standard tape drive and, even with such access, there may be problems in reading the data on non-IBM machines. The format in which the data are copied to tape will generally require further processing depending on the package into which they are to be read. This places substantial extra burdens on potential users.

These methods of sharing data with a wider public are being rendered increasingly obsolete with the development of micro-computers. The bulk of econometrics packages currently in use in Ireland will run on such micros and even such large mainframe packages as TROLL can be run on specially modified micros. In the longer term, the development of communications will allow direct transfer of such data over telephone lines using such facilities as EIRPAC. At present the problems in arranging such communications between widely differing types of computers are rather complex. Until such facilities are in widespread use there is a need to make the data from the different databases available on floppy disks which can be read by the family of IBM-compatible micro-computers. However, even given this latter constraint, there is a wide range of different formats which could be used. While there are some signs that a range of standard formats for data transfer between different programmes is developing in the area of spreadsheets, these may not be suitable for all potential users.

In the end the standard adopted by the CSO will probably set the pace for Irish users. The set of diskettes which, as indicated above, the CSO intends to market in the near future will be written in DIF format which is suitable for many spreadsheets. However, many of the other analysis packages used by economists cannot read data written in such a form directly. However, the CSO has invited comments from potential users on this matter and, depending on the reaction, it may be prepared to change or add to the range of formats in the future. The approach adopted by the OECD is to supply a special programme on the diskette which will translate the data into a range of different formats. In the longer term this may be the best solution to the problem for Irish suppliers of data.

While this issue may seem rather technical to many readers, if a suitable range of formats is not decided on now it will cause serious problems for all potential users. Much time could be spent translating data into a range of different formats. Most users of econometrics software do not enjoy delving into hexadecimal code and writing programmes to transform data!

6. Dissemination of the Databases to a Wider Audience

The CSO's decision to market its data in diskette form must be viewed with great interest. This service will, hopefully, expand over time and will be of major benefit to all those who currently use CSO data in a printed form.

The other two databases described above are currently maintained by the Department of Finance and the ESRI. The ESRI is currently considering how best to make available a subset of these data on floppy disk to outside users. With this in mind a questionnaire is included with the *Quarterly Economic Commentary* which potential users are invited to fill in. The scope and range of services to be provided will be determined by the potential demand. Obviously such a service will have to at least cover its costs through user charges. If there is sufficient potential demand it is likely that the service could come in two forms: either a wide range of data from the databases with limited additional analysis, or else a tailor-made package covering a user's specific interests with some additional analysis (e.g. seasonal adjustment, moving averages, etc.). The frequency of revision of the data would be determined by the user's needs.

With the advent of the CSO's diskette service and the potential development of the Department of Finance/ESRI Database the range of services available to Irish economic analysts will be considerably increased. It is to be hoped that other public bodies will follow suit and that retyping of data by a large number of users will become a thing of the past.

References

Costigan, J., J. FitzGerald, A. McQuaid and A. Redmond, 1986. *Database Access Using TROLL on the CCS Computer*, ESRI Technical Series, Number 1.

FitzGerald, J., O. Keegan, A. McQuaid and A. Murphy, 1983, *Department of Finance Data Bank of Economic Time Series*, Department of Finance Research Paper, 2/83.

Hurley, D.G. and A. McQuaid, 1986. *Department of Finance Databank of Economic Time Series. Part 1: National Accounts Data (adjusted)*, Department of Finance Research Paper, 1/86.

Murphy, A., 1984. *Analysis of the 1975 Input-Output Tables*, Department of Finance Research Paper, 1984.

