Information Paper

Construction of the Expanded Analytical Business Longitudinal Database, 2001-02 to 2012-13

Australia

2015

8171.0

ABS Catalogue No. 8171.0

INQUIRIES

For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070.

CONTENTS

BACKGROUND TO THE EXPANDED ANALYTICAL BUSINESS LONGITUDINAL DATABASE

PURPOSE OF THIS PAPER

The paper provides a summary of the methodology used to construct the initial version of the Expanded Analytical Business Longitudinal Database (EABLD). The paper also includes information about the sources of the data used, a description of potential uses and current access arrangements.

WHAT IS THE EXPANDED ANALYTICAL BUSINESS LONGITUDINAL DATABASE

The Australia Bureau of Statistics (ABS) is developing a new statistical asset, the Expanded Analytical Business Longitudinal Database, which integrates financial and business characteristics data for all active businesses in the Australian economy from 2001-02 to 2012-13. Using the ABS Business Register (ABSBR) as the integrating spine, the initial version of the EABLD links firm level administrative data from the Australian Taxation Office (ATO) with directly collected ABS survey data. The EABLD has been developed in partnership with the Department of Industry and Science (DIS) who also provided financial support.

The EABLD builds on the ABS' panel-based Business Longitudinal Database (BLD) which was created in 2005. The BLD contains longitudinal data from small and medium sized businesses and is designed to allow analysis of micro drivers of business performance over time. Successive panels (or waves) are followed for five years, with collected characteristics data linked to a small number of financial data items sourced from Business Activity Statement (BAS) data provided to the ABS by the ATO. Each panel includes approximately 2,000 businesses. The BLD is made available to researchers via a Confidentiialised Unit Record File (CURF) released in the Remote Access Data Laboratory.

Since the first BLD CURF was made available in 2009, there has been increasing demand for data which can accommodate more sophisticated analysis from researchers interested in the relationships between productivity, innovation and competition. This type of analysis requires additional financial and business characteristics information than available on the BLD. It also requires the full population of businesses including the large and complex firms, rather than only the sample of small and medium sized firms which are available in the BLD.

The EABLD has been created retrospectively, that is, data for the periods 2001-02 through to 2012-13 have been linked together at the one time, rather than progressively over the ten years as each year's information becomes available. This meant that detailed and complete information about business entries, exits and restructures which have occurred in the past were not available, so various methods were developed to ensure that the EABLD has a longitudinal perspective which best reflects that dynamic nature of the business population over time.

WHY CONSTRUCT AN EABLD

The EABLD has been created to demonstrate that integrating administrative and directly collected firm level data can provide a solid evidence base for productivity analysis, policy development and evaluation.

There is a wealth of information about businesses held across all levels of government in the form of administrative data. The ABS itself holds a substantial amount of firm level data sourced from business collections and administrative data provided by other government organisations. Until now much of these data have not been systematically integrated or linked, particularly not on a longitudinal basis. Combining high value data sources which have previously been used in isolation provides the opportunity to better support policy development through research and discussion. The ABS has recognised the increasing demand by government agencies and the analytical community for access to such integrated business micro level data, and with the assistance of the DIS, has created the EABLD. Research conducted using the EABLD will result in a greater amount of information available for use in policy formation. It will also enable the assessment of program outcomes using more advanced quantitative statistical techniques. The results from such program evaluation can in turn feed back into future policy design to achieve improved program outcomes.

The EABLD builds on the range of statistical data integration projects, particularly those focused on enhancing the suite of micro data and longitudinal initiatives, such as the Census Longitudinal Database and the ABS Migrants Census Data Enhancement Project.

DATA SOURCES INCLUDED IN THE EABLD

DATA SOURCED FROM THE ATO

ATO Administrative Data

ATO data included in the EABLD has been supplied to the ABS under the *Taxation Administration Act 1953* which requires that such data are only used for statistical purposes. Once the information is provided to the ABS, it also becomes subject to the *Census and Statistics Act 1905*. No individual information collected under the *Census and Statistics Act 1905* is provided back to the ATO or any other organisation for administrative or regulatory purposes. Any discussions within this paper regarding data limitations are in the context of using the data for statistical purposes, and not in the ability of the data to support the ATO's core operational requirements. For more information about the data sourced from the ATO, refer to the ATO website: <www.ato.gov.au>.

Data from three ATO administrative sources are included in the EABLD. The structures set up by businesses to report obligations to the ATO can change at any time as a result of mergers, acquisitions, business closures and commencements. This, and other issues particular to each data source, can impact on the ease of integration with ABS directly collected data. A list of data items can be provided upon request.

Business Activity Statement (BAS)

Business Activity Statements are submitted to the ATO by businesses to report their Goods and Services Tax (GST) obligations. The data items available include total sales, other GST-free sales, non-capital purchases, capital purchases, export sales and wages and salaries. Businesses can choose to report their BAS information to the ATO on either a monthly, quarterly or annual basis depending on the size and, to some degree, preferences of the businesses. However businesses can elect to apply cash or accrual accounting methods and so summing all periods of a financial year may be affected by seasonality or other factors.

Business Income Taxation (BIT)

Business Income Taxation forms are submitted to the ATO by businesses to report taxable income or loss. There are four main types of businesses that report annual income tax; these are partnerships and partners; companies; sole traders; and trusts and beneficiaries. Across the four different types of BIT forms the majority of items reported are similar, however, the level of detail required can be quite different. There can also be differences in the time periods covered. Further, if one business acquires another business, the acquired business may still be required to submit a taxation return for that part of the fiscal year when it operated independently.

Pay As You Go (PAYG)

Employing businesses are responsible for collecting the personal income tax obligations of their employees and providing employees with an annual payment summary at the end of each fiscal year. They then report this information to the ATO through a Pay As You Go statement. Within large, complex structures, an Ultimate Holding Company (UHC) can be responsible for reporting employment information for all members of the group. If the UHC acquires new firms or sells those that are within the current group this can impact on what is reported, and by which firm.

Collective reporting

The majority of businesses report their BAS, BIT and PAYG obligations for a single ABN. However, as part of the Australian Tax Office's goal to reduce reporting burden, a group of businesses may elect to collectively report their taxation obligations. The individual members of a reporting group do not report individual totals, as a nominated representative reports the group total obligations on their behalf. The collective reporting arrangements allowed can differ for the three types of taxation, which means the reporting groups created may not necessarily be the same across all three types. In addition, businesses can form, alter and disband collective reporting arrangements throughout the fiscal year. This can result in some transactions being recorded under one business and others under the nominated representative within the fiscal year.

ABS SURVEY DATA

The majority of ABS business surveys use the ABS Business Register as the source for their survey frame. This facilitates linking ABS survey data to the ATO data to form the EABLD. Currently, data from three different ABS collections have been included in the EABLD. A list of data items can be provided upon request.

Business Characteristics Survey

The Business Characteristics Survey (BCS) is conducted annually and collects detailed information relating to the use of information technology, innovation and a broad range of other non-financial characteristics (such as collaborative arrangements and competition). The BCS scope consists of all employing business entities (defined as those businesses which register for the ATO's Pay As You Go Withholding scheme) in the Australian economy except for those in:
- Standard Institutional Sector Classification of Australia (SISCA) 3000 (General Government)
- SISCA 6000 (Rest of the World)
- Australian and New Zealand Standard Industrial Classification 2006 (ANZSIC06) Division O (Public administration and safety)
- ANZSIC06 Division P (Education and training)
- ANZSIC06 Groups 624 (Financial asset investing) and 633 (Superannuation funds)
- ANZSIC06 Groups 954 (Religious services) and 955 (Civic, professional and other interest groups)
- ANZSIC06 Subdivision 96 (Private households employing staff)

The BCS utilises a stratified random sample methodology; the sample size may vary from year to year (for 2012-13, the total sample size was approximately 12,000 businesses). The reference period is the year ended 30 June.

Economic Activity Survey

The Economic Activity Survey (EAS) is conducted annually and collects financial information for all businesses operating in the Australian economy except for those in:
- SISCA 3000 General Government (apart from those in ANZSIC06 Subdivision 28 Water Supply, sewerage and drainage services which are included)
- ANZSIC Subdivision 62 (Finance)
- ANZSIC Subdivision 63 (Insurance and superannuation funds)
- ANZSIC Subdivision 75 (Public Administration)
- ANZSIC Subdivision 76 (Defence)
- ANZSIC Subdivision 96 (Private households employing staff)

The EAS utilises a stratified random sample methodology; the sample size may vary from year to year (for 2012-13, the sample size was approximately 22,000 businesses). The reference period is the year ended 30 June.

Survey of Research and Experimental Development by Businesses (BERD)

This survey collects information on Research and Experimental Development (R&D) undertaken by Australian businesses. R&D is defined in accordance with the OECD standard as 'creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications'.

The survey scope is all Australian businesses and the private non-profit institutions mainly serving them. The frame is constructed based on a list of businesses provided to the ABS by the Department of Industry and Science and is sourced from their R&D Tax Incentive Scheme.

Up to and including 2010-11 the BERD was conducted annually as a census of all in-scope businesses. From 2011-12, this was changed to a biennial stratified random sample of approximately 4,000 businesses. The survey frame is constructed using a list of businesses which have registered for R&D tax incentives which is linked to the ABSBR. The reference period was unchanged and remains the year ended 30 June.

INTEGRATION METHODOLOGY

The integrating spine for the EABLD is the ABS Business Register (ABSBR). The Australian Business Number (ABN) is used as the primary linking variable in the construction of the EABLD.

As the EABLD is being created retrospectively, access to a snapshot of the ABSBR as it was at the end of June in each of the EABLD reference periods is not available, a proxy has been used. This proxy has been sourced from the comprehensive common frame which is used to conduct ABS business surveys. The common frame created in June of each of the EABLD reference periods has been used.

ABS BUSINESS REGISTER

The ABS uses an economic statistics model to describe the characteristics of businesses and the structural relationships between related businesses. Most businesses and organisations in Australia are required to obtain an ABN. They are then included on the whole-of-government register of businesses, the Australian Business Register (ABR), which is maintained by the ATO. The ABS allocates businesses from the ABR to one of two sub-populations:

Non-profiled population

The majority of businesses on the ABR have simple structures and the unit registered for an ABN will satisfy ABS statistical reporting requirements. For these businesses, the ABS aligns its statistical units structure with the ABN. That is, the ABN unit is used as the statistical unit for all integrated ABS economic collections. These businesses form the non-profiled population.

Profiled population

For those businesses where the ABN is not considered suitable for ABS statistical requirements, the ABS maintains its own units structure through direct contact with the business. This population consists typically of large, diverse and complex structured businesses, and constitute the profiled population. For businesses in the profiled population, statistical units comprise the Enterprise Group (EG) and the Type of Activity Unit (TAU). The range of activities carried out across the EG can be very diverse. The TAU is set up to represent a grouping of one or more businesses within the EG that cover all the operations within an industry sub-division and for which a basic set of financial, production and employment data can be reported.

Together these two sub-populations (the non-profiled and profiled populations) make up the ABS Business Register.

POINT-IN-TIME INTEGRATION

The reported ATO data, which is based on taxation reporting structures, has been adapted to fit the economic units model structure used on the ABSBR. A process has been developed to transform the reported ATO data to facilitate its use for statistical purposes. Initially, within the ATO collective reporting arrangements, data reported by the nominated representative is apportioned across all members of the group, so that every business has taxation data representative of activity (rather than representative of a structure created for reporting to the ATO).

Diagram 1: Collective reporting of tax data distributed across member ABNs
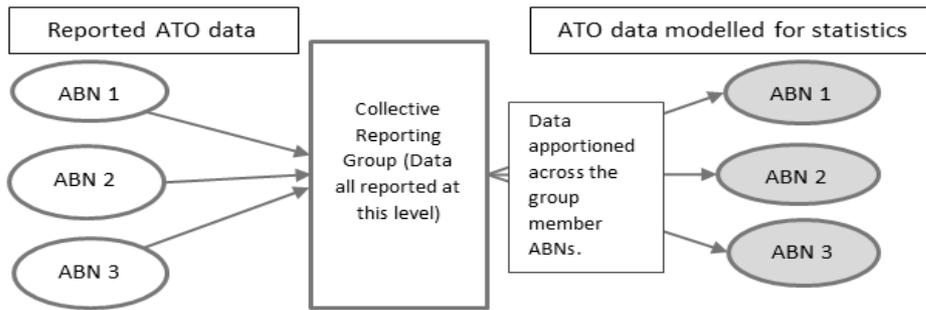
After this step, the ATO datasets containing the BAS, BIT and PAYG data are integrated with the ABSBR.

For units in the ABS non-profiled population, the ABN used in the ATO reported data is the equivalent unit to that used on the ABSBR (as there is a one-to-one relationship between them). Integrating the ATO financial information with the ABSBR information is straight-forward for these units. The ABS data collected from the Business Characteristics Survey, Economic Activity Survey and Survey of Research and Experimental Development by Businesses is able to be directly linked to the relevant taxation data.

Diagram 2: ATO data at the ABN level integrated with ABS Non-profiled Population
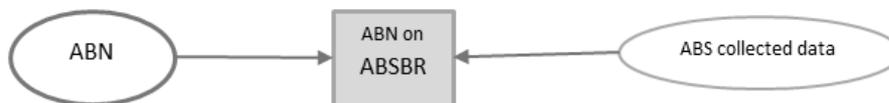
For units in the ABS profiled population, ATO reported data from all ABNs under a given EG is aggregated to this EG level, and then apportioned across the ABSBR TAUs within the EG based on employment information. This produces a dataset which contains modelled ATO data for units based on the reporting structures on the ABSBR. Again, ABS collected data is then directly linked to this taxation data.

Diagram 3: ATO data at the ABN level distributed to ABS Profiled Population
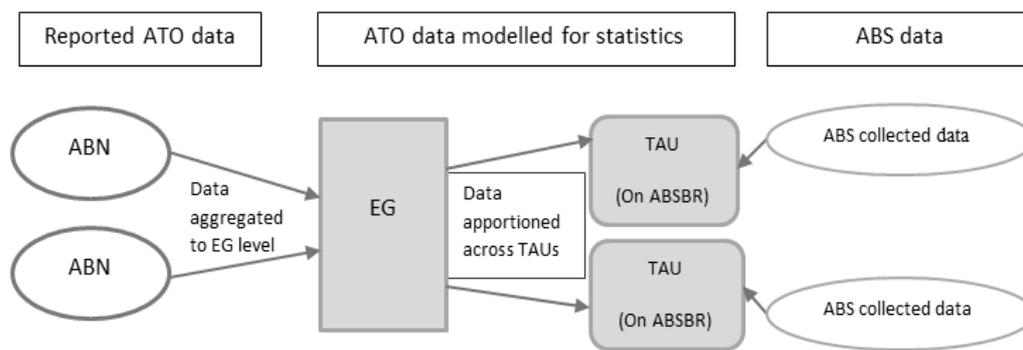
This process has been undertaken for each of the financial periods included (2001-02 to 2012-13) in the EABLD, based on the reporting structures current in each of the ATO and ABSBR datasets at the end of the relevant period.

LONGITUDINAL INTEGRATION

The primary purpose of the ABSBR is to facilitate the production of information about active businesses in the Australian economy for a given time period and it is used as the source of almost all business survey frames. Hence, unit structures on the ABSBR are altered as often as necessary to reflect the changes that occur within the business population. Making such changes is vital to ensure ABS outputs are fit-for purpose. However, to then identify the reasons why businesses enter and exit the ABSBR across different time periods can be difficult. As the EABLD is being created retrospectively, detailed and complete information about business entries, exits and restructures which have occurred in the past was not available, so various methods were developed to ensure that the EABLD has a longitudinal perspective which best reflects that dynamic nature of the business population over time.

It is important to note that a business entry or exit does not necessarily equate to a business start-up or closure. There may be a number of reasons why a business entry or exit event has occurred, including events relating to selling a business and changes in a business' structure due to merger or takeover activities. Further, businesses may transition into and out of the profiled population based on the guidelines used by the ABS, such as a medium sized enterprise growing in size and complexity which is then moved from the non-profiled to the profiled population.

CONTRIBUTION TO ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT PROJECTS

The creation of the EABLD has enabled Australia to take part in two Organisation for Economic Co-operation and Development (OECD) distributed microdata projects looking at the Dynamics of Employment (DynEmp) and Micro Drivers of Productivity (MultiProd), which previously had not been possible.

DynEmp

Employment growth is at the heart of the policy debate. Cross-country evidence on the role of small and young firms in contributing to employment creation is still scant. DynEmp is a project led by the OECD aimed at providing new empirical evidence on the role of creative destruction, start-ups and young firms to support the design of better policies for employment and productivity growth based on confidential firm level data from national business registers.

DynEmp is a distributed microdata project currently involving approximately 20 countries. The main findings across all participating countries so far is that young, small businesses account for a disproportionate amount of jobs and that the start-up rate of new businesses has been falling over the past decade. The ABS has run the DynEmp routine on the EABLD, and Australian results broadly conform to the patterns evident in other countries. For example, young firms contribute disproportionately to job creation and the start-up rate of new businesses fell over the period 2001-02 to 2012-13. The OECD will include Australian DynEmp data in their next release planned for later in the year.

In respect of Australia's contribution to the OECD's DynEmp project, broad level aggregates from the PAYG data were compared to closely equivalent series such as annual wages and salaries from the national accounts, labour force measures of employment and hours worked, and average weekly earnings. Overall these coherence checks confirmed that the methodology used to integrate the various datasets resulted in data that are consistent with other ABS data series. For example, when comparing appropriately scoped annual totals from PAYG with average weekly earnings the industry averages were quite similar.

Multiprod

Employment patterns and productivity growth play a central role in shaping the welfare of societies and the competitiveness of countries. 'MultiProd' is an OECD project to study productivity patterns, investigating the extent to which different policy frameworks can shape firm productivity and examining the way resources are allocated to more productive firms through a process known as allocative efficiency.

MultiProd is proceeding in the same manner as the DynEmp project in that an OECD-supplied statistical routine is provided to participating countries and the results are collected and analysed by the OECD. The main goal of MultiProd is to determine the drivers of productivity (both labour and multi-factor) and to assess whether there are similar patterns across the participating countries. The ABS intends to produce results for MultiProd in the 2015-16 financial year.

ACCESSING THE EABLD

Access to the EABLD is currently available in the following ways:

- The ABS undertakes work on behalf of the client. This involves contacting the ABS to determine the research and analysis to be undertaken. This is a fee for service arrangement where varying levels of complexity of statistical analysis can be undertaken. Under this arrangement confidentialised summary information and results are returned to the client.

- The ABS runs statistical analysis programs provided by the client. This is a process where the client provides statistical code to be run against the EABLD. This is a fee for service arrangement and confidentialised information is returned to the client. The client may use the services of a third party, such as an academic institution, with skills in statistical analysis to create and supply the code to the ABS.

The ABS is currently reviewing micro data access arrangements for clients and other options for access to the EABLD may become available as the review progresses.

POTENTIAL FUTURE DEVELOPMENTS

As noted at the beginning of this paper, the creation of an enduring firm level statistical asset will aid in building the capacity to undertake firm level analysis of micro-economic drivers of performance, competitiveness, productivity. It will improve the evidence base for policy development and evaluation leading to more targeted expenditure of government funds. A large array of information about business is captured as part of government administration and by integrating these data, there would be an increase in capacity to evaluate the effectiveness of government initiatives in a strategic cross-portfolio context. The construction of an EABLD for the periods 2001-02 to 2012-13 represents the foundation of a dynamic and enduring firm level statistical asset, able to be expanded and enhanced through the inclusion of additional datasets. Integration into the EABLD of Personal Income Taxation (PIT) data from the ATO would represent a significant step towards the creation of a fully Longitudinal, Linked Employer/Employee Database.

The method used to construct the EABLD supports the integration of ABS business survey data and other administrative data. Examples of other data that could be integrated are patents, trademarks and designs data from IP Australia; insolvency data from the Australian Financial Security Authority; or Department of Industry and Science program participation data. Enhanced access for analysts will also be a core goal of future development.

As the above are medium to long term objectives, in the short term the ABS plans to progress the following areas:

REFINED INTEGRATION METHODOLOGY

During construction of the EABLD a number of assumptions were made when apportioning data to address collective reporting arrangements and the differences between ATO and ABS reporting structures. Further investigation into improved apportioning methods will continue. This will be achieved through gaining a better understanding of tax reporting arrangements, via interactions with the ATO, and investigation into potential data sources available to assist with apportioning. In addition, the ABS will continue to keep abreast of developments within the field. For example, recent research on German data suggests that relabelling business identifiers introduces bias when examining the drivers of employment growth among young and established businesses (Hethy-Maier and Schmieder, 2013).

INTERNATIONAL COLLABORATION

During the course of constructing the EABLD, the ABS investigated other comparable datasets produced internationally, such as those by Statistics New Zealand and Statistics Canada. The key lesson learnt was that each country faces its own challenges due to their statistical units model and how this aligns with administrative data. For example, Statistics Canada identifies false entry and exit by tracking labour movements and abstracts from merger and acquisitions by projecting the current business structure backwards. In contrast, researchers interested in organic employment growth wish to distinguish employment growth due to acquisition from that arising within the business itself, which is possible in countries such as Sweden (Davidsson and Delmar, 2006).

The ABS will continue to collaborate internationally to share findings and strengthen international partnerships to maintain current understanding of global developments. This will assist with the improvement of Australian versions of various integrated micro level datasets including the future development of firm level statistical assets including redevelopment of the ABS Business Register.

REFERENCES

Davidsson, Per and Frederic Delmar "High Growth Firms and Their Contribution to Employment: The Case of Sweden 1987-1996" in Davidsson, Per; Frederic Delmar and Johan Wiklund (2006) *Entrepreneurship and the Growth of Firms*, Edward Elgar Publishing, Northhampton, Massachusetts.

Hethy-Maier, Tanja and Johannes F. Schieder (2013) "Does the Use of Worker Flows Improve the Analysis of Establishment Turnover? Evidence from German Administrative Data" National Bureau of Economic Research (NBER) Working Paper No. 19730.

FOR MORE INFORMATION . . .

www.abs.gov.au the ABS website is the best place for data from our publications and information about the ABS.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FAX 1300 135 211

EMAIL client.services@abs.gov.au

PHONE 1300 135 070

FREE ACCESS TO STATISTICS

All ABS statistics can be downloaded free of charge from the ABS web site.

WEB ADDRESS www.abs.gov.au