



Department of Econometrics and Business Statistics

<http://business.monash.edu/econometrics-and-business-statistics/research/publications>

**The Australian Macro Database:
An online resource for
macroeconomic research in Australia**

Timur Behlul, Anastasios Panagiotelis, George
Athanasopoulos, Rob J Hyndman and Farshid Vahid

December 2016

Working Paper 17/01

The Australian Macro Database: An online resource for macroeconomic research in Australia

Timur Behlul

Alphabeta Consulting

Level 7, 4 Martin Place, Sydney, New South Wales 2000, Australia

Email: timur.behlul@alphabeta.com

Anastasios Panagiotelis*

Department of Econometrics and Business Statistics,

Monash University, Caulfield Campus,

VIC 3145, Australia

Email: anastasios.panagiotelis@monash.edu

George Athanasopoulos

Department of Econometrics and Business Statistics,

Monash University, Caulfield Campus,

VIC 3145, Australia

Email: george.athanasopoulos@monash.edu

Rob J Hyndman

Department of Econometrics and Business Statistics,

Monash University, Clayton Campus,

VIC 3800, Australia.

Email: rob.hyndman@monash.edu

Farshid Vahid

Department of Econometrics and Business Statistics,

Monash University, Clayton Campus,

VIC 3800, Australia.

Email: farshid.vahid@monash.edu

* Correspondence to: Anastasios Panagiotelis, Department of Econometrics and Business Statistics, Monash University, PO Box 197, Caulfield East, VIC 3145, Australia.

Acknowledgment: We thank Professor Mardi Dungey for testing the database and providing valuable feedback early in its development. We are also grateful for the support of the Australian Research Council through the Discovery Project Grant DP1413220.

7 December 2016

JEL classification: C55, C82

The Australian Macro Database: An online resource for macroeconomic research in Australia

Abstract

A website that encourages and facilitates the use of quantitative, publicly available Australian macroeconomic data is introduced. The Australian Macro Database hosted at ausmacrodata.org provides a user friendly front end for searching among over 40000 economic variables, sourced from the Australian Bureau of Statistics and the Reserve Bank of Australia. The search box, tags and categories used to facilitate data retrieval, are described in detail. Known issues with the website and future plans are discussed in the conclusion.

Keywords: Economic time series, Macroeconomic data, Australian Bureau of Statistics, Reserve Bank of Australia, FRED.

1 Introduction

To encourage and facilitate quantitative research into the Australian economy, it is crucial that economists have easy access to current data. In the US, web based services such as FRED and Quandl compile data from different sources and provide an easy-to-use interface for economists to download data. In this paper, we introduce a new website, ausmacrodata.org, that provides this role for Australian data. This website was created as part of a Discovery Project ¹ funded by the Australian Research Council.

A large number of Australian economic time series are collected and made available online by the Australian Bureau of Statistics (ABS). Unfortunately, the ABS website does not always cater to the specific needs of economic researchers. First, the ABS is not only responsible for collecting economic data, but also data on health, education, the environment, and other important issues. Second, the ABS does not necessarily distinguish between data releases that contain quantitative data and data releases that contain text data such as explanatory notes. Third, even where quantitative data are available for download via a MS-Excel spreadsheet, often multiple series are included in a single file, sometimes across different sheets making it difficult to find a specific series without downloading and browsing through multiple large spreadsheets. Fourth, it is difficult to filter series by geographic location or by the frequency with which data are released. Fifth, publicly available economic data from other sources, in particular the Reserve Bank of Australia, are not made available via the ABS website.

The new website ausmacrodata.org addresses these issues, providing a user friendly front end for economic researchers. The original motivation behind the project was to develop a large macroeconomic database for the estimation of coincident indices and forecasting in the context of the dynamic factor model literature (see Forni, Hallin, Lippi, & Reichlin, 2000; Stock & Watson, 2002). However, the website could be of value to macroeconomists, labour economists, or indeed any economic researcher that uses Australian quantitative data. In Section 2, we outline the series available on ausmacrodata.org. In Section 3 we discuss how best to navigate the site, which will depend on the needs of the user. In particular we discuss the concepts of categories and tags. In Section 4, we discuss how data can be viewed and downloaded. We conclude in Section 5 and discuss future plans for ausmacrodata.org.

¹DP140103220: “Macroeconomic forecasting in a ‘Big Data’ world”.

Catalogue Number	Description
3101.0	Demographic Statistics
5206.0	National Accounts: National Income, Expenditure and Product
5220.0	National Accounts: State Accounts
5232.0	National Accounts: Finance and Wealth
5302.0	Balance of Payments and International Investment Position
5368.0	International Trade in Goods and Services
5609.0	Housing Finance
5625.0	Private New Capital Expenditure and Expected Expenditure
5676.0	Business Indicators
6202.0	Labour Force
6302.0	Average Weekly Earnings
6345.0	Wage Price Index
6354.0	Job Vacancies
6401.0	Consumer Price Index
6416.0	Residential Property Price Indexes: Eight Capital Cities
6427.0	Producer Price Indexes
6457.0	International Trade Price Indexes
8501.0	Retail Trade
8731.0	Building Approvals
8750.0	Dwelling Unit Commencements
8755.0	Construction Work Done
9314.0	Motor Vehicle Registrations

Table 1: Summary of ABS catalogues that contain data provided by the Australian Macro Database.

2 Data

The data available at the Australian Macro Database come from two major sources. The majority of the data are released by the Australian Bureau of Statistics (ABS). To explain the data that are available via ausmacrodata.org, it is necessary to briefly explain how data are organised at the ABS. All ABS data are organised into *catalogues*, each identified by a unique *catalogue number*. Only a subset of these catalogues, summarised in Table 1, are available via ausmacrodata.org. The ABS catalogues that are excluded contain data that are not macroeconomic in nature, and are instead related to other domains such as health or the environment. Other ABS catalogues are excluded since they contain explanatory notes on the data collection process rather than quantitative data.

Each catalogue is organised into *tables* where each table is essentially a spreadsheet containing multiple *series*. In total, data on 39,394 variables, sourced from the ABS, are made available through the Australian Macro Database.

In addition to ABS data, additional series are obtained from the Reserve Bank of Australia (RBA). A list of these series can be found at <http://www.rba.gov.au/statistics/tables/>. Generally

these involve banking, money, credit, lending, interest rates, as well as inflationary expectations. Some series published by the RBA come from other sources, for example series on business and consumer sentiment are measured by NAB and MI-Westpac respectively, while other series made available by the RBA are themselves sourced from the ABS. In total, 910 variables are sourced from the RBA, bringing the total number of variables available via ausmacrodata.org to 40,304 variables.

The Australian Macro Database checks for newly updated data daily at 3am Australian Eastern Standard Time. A current schedule of ABS updates is also provided in the Frequently Asked Questions on ausmacrodata.org/FAQ.

3 Navigating the Site

Links to download some of the most commonly-used series are provided on the homepage of ausmacrodata.org, which is shown in Figure 1. These series are real GDP growth, real GDP, the current account, the unemployment rate, inflation rate, change in house prices, the target cash rate, the Australian–US dollar exchange rate and retail sales. For each of these series, the latest value, the units of measurement, the dates for which historical data are available, and the frequency of the series, are provided on the home page, giving an overall snapshot of some key macroeconomic variables.

For other series, the best way to navigate the site will depend on the exact series one is searching for. For many commonly used series it is best to browse through the *categories* tab. For more obscure series, it is possible to browse through tags which can be filtered by their own search function. There is also a generic search function at the top right hand side of the page. We now explain categories, tags and the search box in more detail.

3.1 Categories

To ensure that the site is easy to use, we have developed a hierarchy of categories to facilitate navigation. Hovering the mouse pointer over a category will show a drop-down menu with further subcategories. This is displayed in Figure 2. We now describe the top level of our hierarchy of categories.

National Accounts, Flow of Funds and International Trade: This category contains series related to the components of national accounts: household and government consumption, public and private investment, imports and exports at both a national and state level.

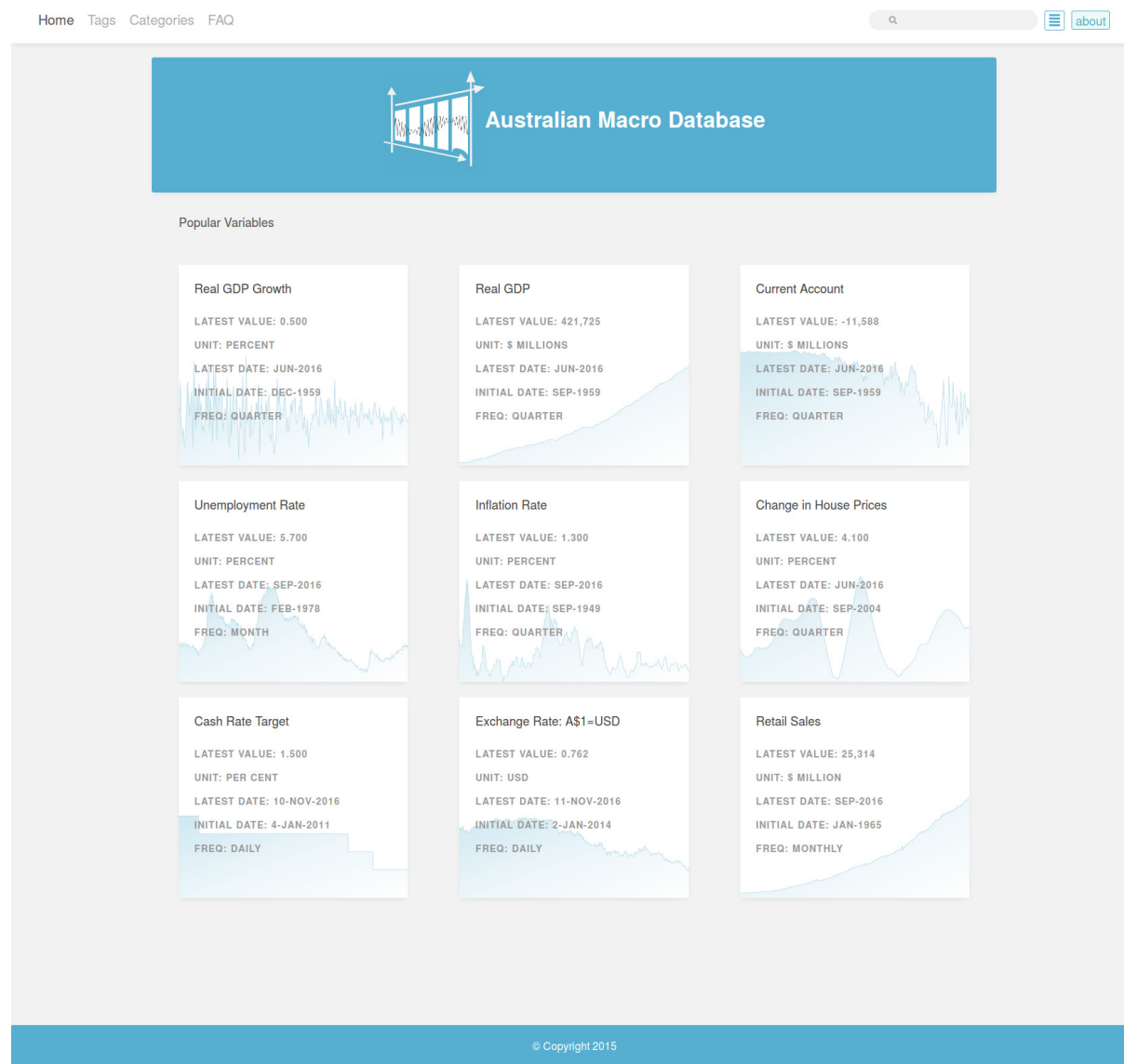


Figure 1: Homepage of ausmacrodata.org showing information about nine major economic series.

It also contains series related to assets and liabilities of households, financial and non-financial corporations and governments, and the balance of payments and international finance.

Labour Statistics: The category contains series pertaining to the labour market, including employment, weekly earnings and job vacancies.

Industry: This category contains data on variables including capital expenditure, sales and inventories measured at an industry level. Some of the specific industries for which variables are available are building and construction, housing, finance, retail and automobiles.

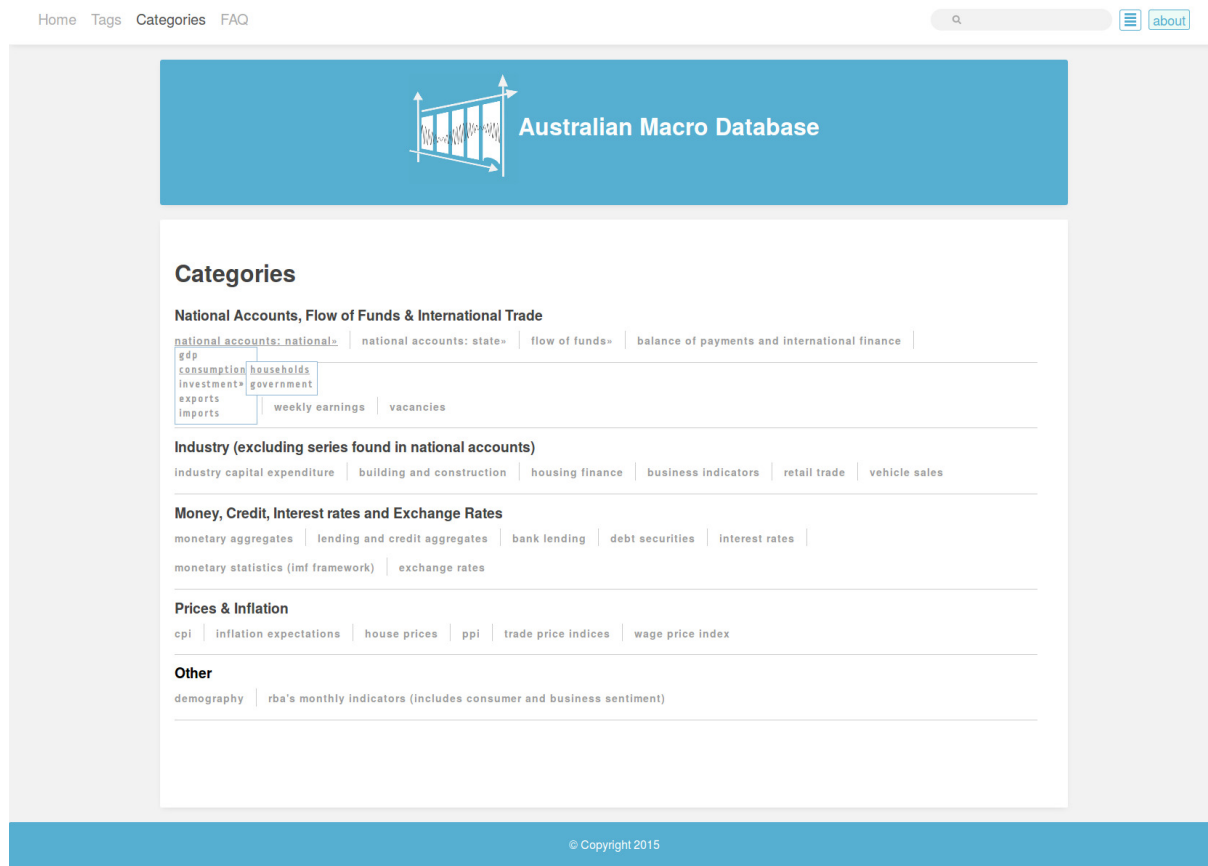


Figure 2: Categories of ausmacrodata.org including drop down menus that become visible when the mouse pointer is hovered over them.

Money, Credit, Interest Rates and Exchange Rates: This category includes data on monetary variables including different measures of the money supply, credit and lending aggregate and also interest rates and exchange rates.

Prices and Inflation: This category contains the components of the consumer price index at national and city level as well as house prices. Data on both actual inflation and inflationary expectations as well as the trade price index and wage price index are also available here.

Other: This category includes two subcategories. The first is demography and includes births, deaths and marriages, and domestic and international population flows. The second includes surveys of business and consumer sentiment measured by the RBA.

3.2 Tags

Tags are keywords that allow researchers to isolate groups of series in which they may be interested. For example, a researcher interested in finding data on insurance for the state of Victoria measured at a quarterly frequency can use the tags ‘insurance’, ‘victoria’ and ‘quarterly’, which will filter series that have all three of these tags, reducing the number of series under consideration from over 40000 to just six series. A screenshot that shows this example is provided in Figure 3.

Due to the large number of variables, it was necessary to establish tags in an automated fashion. As stated earlier, data from the ABS come from broadly defined *catalogues*; for example, Catalogue number 5206.0 has the title **Australian National Accounts: National Income, Expenditure and Product**. Each catalogue includes multiple *tables*; for example, Table 1 of Catalogue 5206.0 has the title **Key National Accounts Aggregates**. Each table includes multiple *series*; for example, one series in Table 1 of Catalogue 5206.0 is **Gross domestic product: Chain volume measures percentage changes**, which is real GDP growth. The list of tags is constructed from the words used in the titles of the relevant series, table and catalogue. In the example here, all words in bold will be tags for real GDP growth, excluding so-called “stop” words such as ‘and’. Words in the title of the table and catalogue are included as tags since, for many series, the series title on its own is rather uninformative and can be a single word such as ‘manufacturing’.

In addition to the automated selection of tags, some manual adjustments were also made. Tags were created that include multiple words, for example ‘new south wales’ is treated as a single tag. Also, some tags were added manually; for example, all series with the ‘gross domestic product’ tag were given an additional ‘gdp’ tag, and all variables that were chain volume measures were given the tag ‘real’.

Since this leads to over 1000 tags, we have taken several steps to facilitate the use of tags. First, the tags that appear in the most number of series appear on the first page of results. Second, one can search for tags using a separate search box. Third, tags themselves are organised into four sections.

Concepts: This includes tags that are broad economic concepts (e.g., ‘prices’, ‘expenditure’), different measurement concepts (e.g., ‘gross’ and ‘net’), as well as miscellaneous tags that are attached to many series (e.g., ‘non’ and ‘during’).

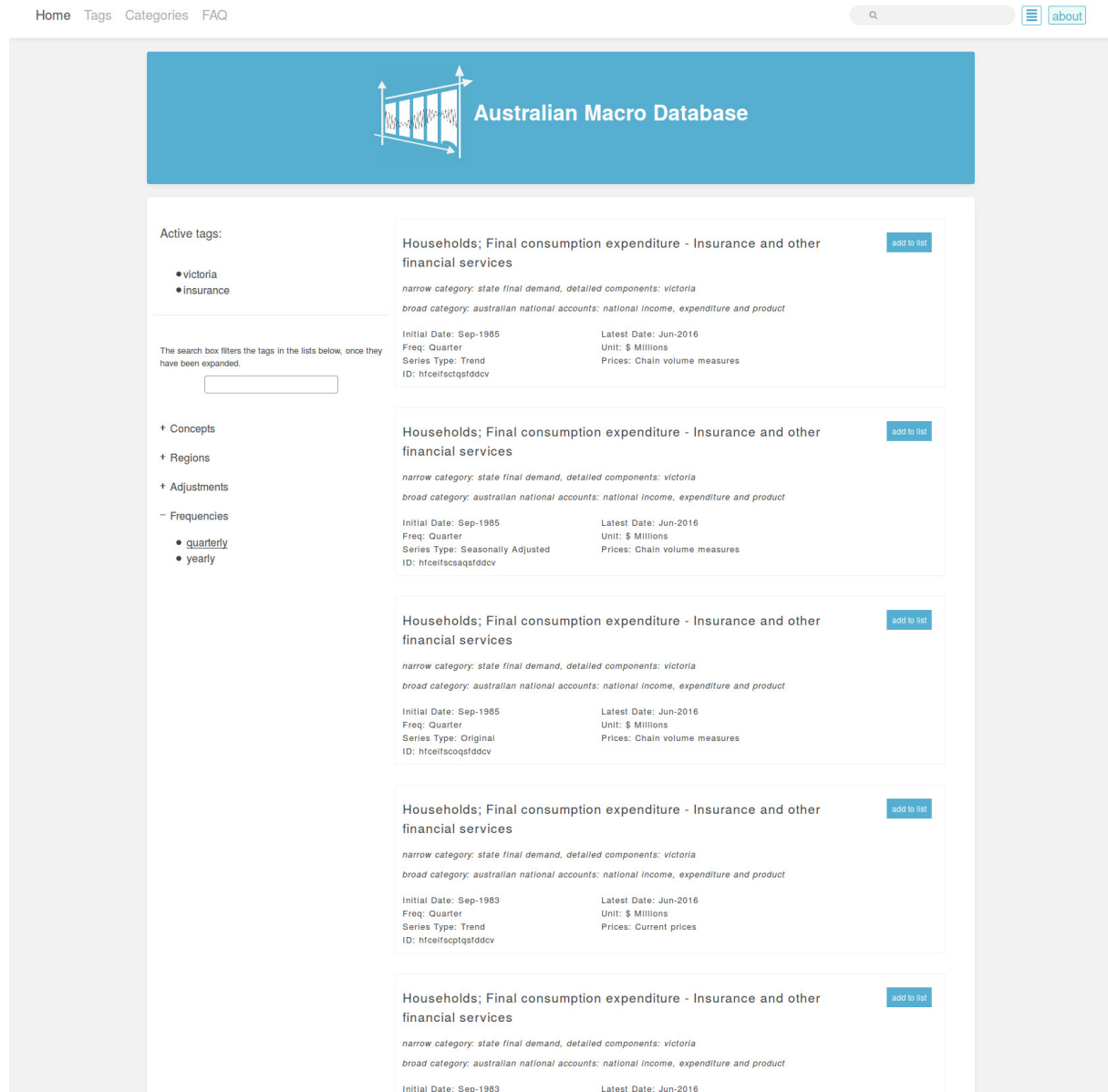


Figure 3: An example of using tags in austmacrodata.org. The tags ‘insurance’ and ‘victoria’ have already been chosen and are ‘active’ tags that filter out the series. By expanding frequencies, one can choose to add either the ‘quarterly’ or ‘yearly’ tag to filter the series even further. The screenshot has been cropped.

Regions: These tags relate to geographical location and include state and territory names (e.g., ‘nsw’, ‘victoria’), the names of state capitals (e.g., ‘sydney’, ‘melbourne’), as well as the tag ‘australia’ which relates to aggregate data.

Adjustments: These tags refer to ways that the data may or may not have been adjusted, and includes the tags ‘original’, ‘real’, ‘nominal’, ‘seasonally adjusted’ and ‘trend’.

Frequencies: These tags refer to the frequency with which data is collected, namely ‘quarterly’, ‘yearly’, ‘monthly’ and ‘daily’.

3.3 Search Box

Rather than browsing through categories or tags, it is possible to use a generic search box in the top right corner of the page. This search function is also centred around the concept of tags but in a way that is more hidden from the user’s perspective.

Each word entered into the search box is treated as a separate string and matched to a string in the list of tags. There are some exceptions to this, for example if ‘new south wales’ is entered into the search box, it will be treated as a single string and matched to the ‘new south wales’ tag. Stop words are also removed.

Each search string is matched to a list of tags using the SQL percent wildcard at the end of the string. This allows the end of the search term to differ slightly from the tag. The series with tags that match the search terms are returned in order of a relevance score. The relevance score will be highest when the search terms match tags that appear in the series name, lower when the search terms match tags that appears in the table name, and lowest when the search terms match tags that appears in the catalogue name.

4 Display and Download

Once a series is selected, several pieces of information are displayed. An example of this screen is provided in Figure 4. In the centre of the page is a time series plot of the data with a zoom function so that the user can focus on a specific range of dates. Along the left hand side of the screen are the date of the most recent update, a description of the series, a link to the source, and further information such as definitions or breaks in the series. By clicking on the ‘view dataset’ tab, the raw numbers of the series can be viewed.

The data can be download in either .csv or .txt format. An additional useful function is to ‘add to list’. This is inspired by the ‘shopping cart’ found on a number of popular retail websites. This function allows users to continue browsing and once they have found all the series required, all series in the list can be downloaded in both .csv and .txt format in a single zip file.

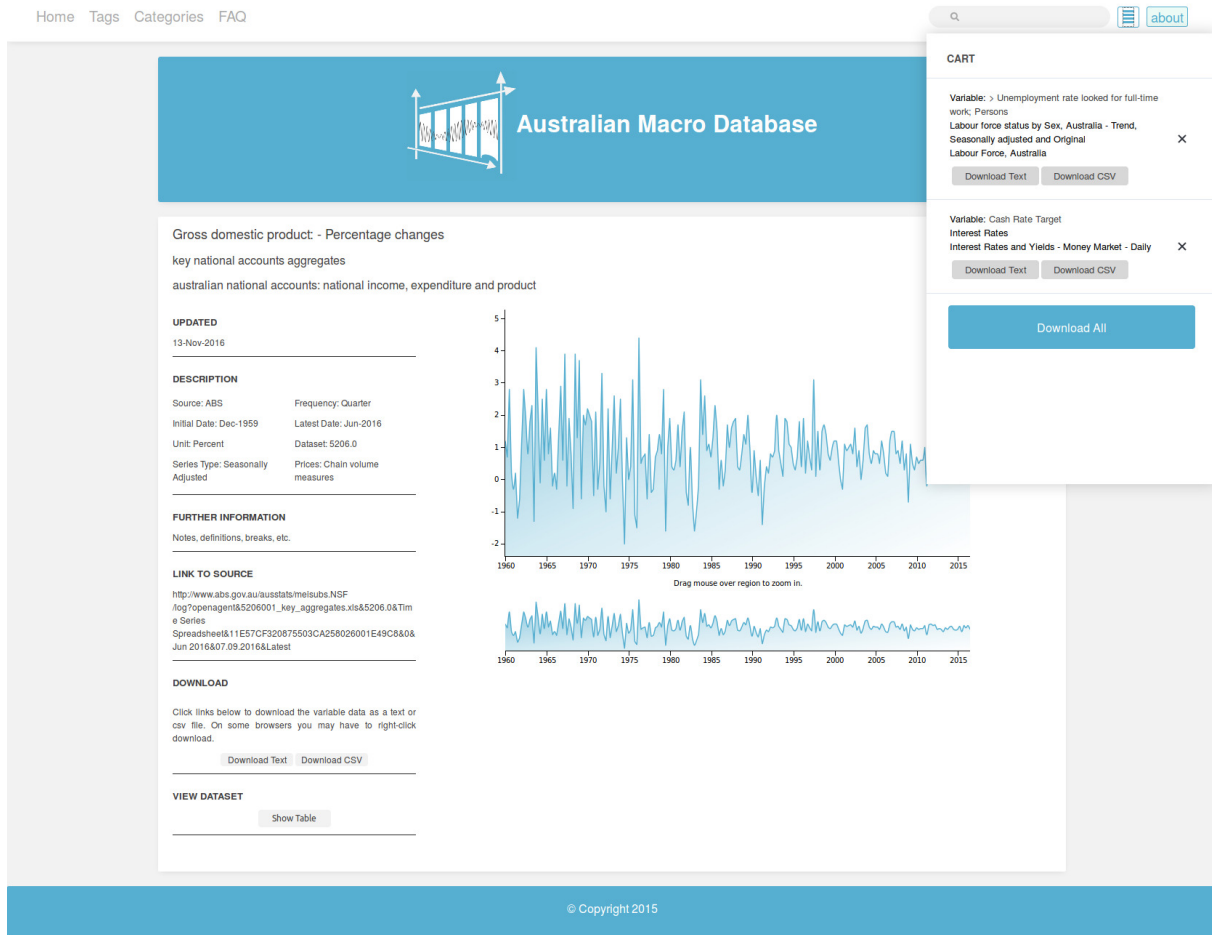


Figure 4: Screenshot of the download screen for seasonally adjusted percentage changes of GDP. The 'shopping cart' of series is displayed at the top right.

5 Known Issues and Future Plans

Unlike FRED and Quandl, there is currently no large administrative team supporting the website. Instead ausmacrodata.org has been set up to be self-updating in a largely automatic manner. This limitation has led to some known issues that users should be aware of. For example, a search for 'gdp' will return 15 results with the title 'Gross domestic product - Percentage changes', some of which are adjusted for trend or seasonality, some of which are real or nominal, and some of which are only available at yearly frequency. Although it would be preferable for the variable to appear once with active tags appearing just below, the resource constraints of the project made this an infeasible task for a website containing over 40,000 individual variables.

In spite of these issues, we believe that the ausmacrodata.org resource will be a useful tool for research into the Australian economy for academics and practitioners alike. This adds to other publicly available data sources put together by Australian research teams such as the

Australian Real-Time Macroeconomic Database fbe.unimelb.edu.au/economics/macrocentre/artmdatabase, UQ International Comparisons Data Base uqicd.economics.uq.edu.au/, and the data archives of the National Centre for Econometric Research (NCER) www.ncer.edu.au/. The Australian Macro Database is an ongoing project and we encourage feedback by sending an email to ausmacrodata@gmail.com. In the future we intend to provide links to download data compiled for specific research papers. Currently, there is a link to the dataset of 151 variables used in the forecasting study of Jiang, Athanasopoulos, Hyndman, Panagiotelis, and Vahid (2017). A point of difference from existing repositories such as the NCER is that ausmacrodata.org will provide a link to a continuously updated version of the data in addition to a link providing the exact data that was used in the paper. Also, researchers who develop APIs for reading data from ausmacrodata.org into their software are encouraged to contact us and we will make these APIs available through the website.

References

- Forni, M, Hallin, M, Lippi, M, & Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *The Review of Economics and Statistics*, **82**(4), 540–554
- Jiang, B, Athanasopoulos, G, Hyndman, RJ, Panagiotelis, A, & Vahid, F. (2017). *Macroeconomic forecastion for Australia using a large number of predictors* (Working paper No. 17/02). Monash University.
- Stock, JH & Watson, MW. (2002). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, **97**(460), 1167–1179