



BANCO DE PORTUGAL
EUROSISTEMA

GUIDE FOR RESEARCHERS USING BANCO DE PORTUGAL MICRODATA RESEARCH LABORATORY (BPLIM) DATA

This version: September 26, 2018



1 General Remarks

The Banco de Portugal Microdata Research Laboratory (BPLIM) is part of the Economics and Research Department of Banco de Portugal (BdP) and was created with the objective of facilitating to researchers the access and use of micro datasets about the Portuguese economy. Primarily, these are datasets owned by BdP, but there are other datasets supplied by third parties. Data can only be accessed on BPLIM's dedicated servers. BPLIM has two servers. The **internal server** is primarily intended for use by BdP own researchers (internal researchers) while the **external server** has remote access capabilities and is intended for external researchers. Internal researchers are not bound by any restriction in terms of data access but restrictions apply to external researchers. Below we detail all relevant aspects involving data access at BPLIM.

2 Data Available for Researchers

2.1 What are the characteristics of BPLIM's datasets?

All BPLIM datasets are stripped of elements that allow for direct identification of companies or individuals. Whenever possible, the datasets contain unique unit identifiers common across datasets: examples of these are *tina* – the tax identification number anonymized for companies – and *bina* – the bank identification number anonymized. Datasets are available in Stata format and are stored in an efficient way that minimizes file size. Labels are applied to all variables and value labels to all categorical variables. Whenever possible labels can be displayed in Portuguese and English and datasets are accompanied by a manual and a set of auxiliary files that succinctly describe the data. BPLIM datasets may also have companion files which extend the capabilities of the original data as well as harmonized variables to guarantee comparability over time and across datasets. Datasets are updated regularly based on a data extraction (“data freeze”) at a specific point in time.

2.2 What data are available?

The complete list of datasets including a short description of each one is available in Appendix A.

3 Data Access

3.1 Who can gain access to the data?

Access is only allowed to researchers who intend to use the data for scientific purposes. Internal users (those with a contractual link to the BdP) have access to all datasets available at BPLIM. Access by external researchers is subject to several restrictions. Appendix A summarizes the access restrictions imposed to each dataset.

3.2 How can researchers request access to the data?

Access to BPLIM data implies the submission of a project. The proposal must: (1) contain a short description of the research project; (2) identify all researchers involved in the project and their affiliation; and (3) identify the datasets, timeframe, and variables required. All external researchers with access to the data are required to sign a



confidentiality agreement and send a copy of their vitas. A copy of these documents can be found in Appendix B. If the project consists of a master or doctoral dissertation then the supervisor has to be part of the project team and must also sign the confidentiality agreement. BPLIM staff will work with the researchers to identify the required datasets and, if needed, to build a customized dataset.

3.3 Approval process

Upon submission of all documentation the project will be evaluated to ensure that it addresses a legitimate research question. Compliance with the guidelines is also an important factor for recurring external researchers (those that have already participated in BPLIM projects). Once the project is approved data access will be granted.

3.4 How can the data be accessed?

When applying for the project the researcher must specify if he/she wants to access an account in the **internal** or the **external server**.

Accounts open at the **internal server** can only be accessed at the installations of BdP (on-site access) either at Lisbon or Porto. BdP internal researchers can login into BPLIM's internal server from their desktop using their network login credentials (this requires installation of specific software¹). Users not affiliated to BdP will be considered external researchers. They will be assigned a login and password for BPLIM's internal server and will be granted access to a terminal from which it is not technically possible to transfer, download, copy, paste or print any data. There is a limited number of terminals available so external researchers must book their visits well in advance.

If the account is on the **external server** then the data can be accessed remotely using a safe connection (we use the *nomachine* software). With this connection it is not possible to exchange files between the external server and the local computer (similar to the terminal access to the internal server). In special circumstances explained below the external researcher may be granted access to the data without having an account at the server (**surrogate access**). In this case the external researcher will not have any direct contact with the data and BPLIM staff will act as a proxy for the researcher regarding data access. This means that BPLIM staff will execute the scripts written by the external researcher and share the outputs with the researcher after validation.

3.5 Which are the restrictions imposed on access to BPLIM's datasets?

Internal researchers are able to access all datasets available at BPLIM provided the access is through the internal server but they will have the same access conditions as external researchers if accessing data through the external server.

¹ BdP researchers should contact the information systems department (DSI) through the 20 500 extension number and request installation of the Linux emulation software.



The access policy for an external researcher is more restrictive. External researchers are granted access to all datasets that can be placed in the external server (these are the datasets classified as “remote” in the column designated “external research access” for the tables in Appendix A). However, if working in a joint project with internal researchers they may be allowed access to other datasets. These datasets are identified with the “co-authorship” classification and have to be placed in the internal server.

Datasets with the classification of “no” in the column “external research access” cannot be accessed by external researchers under any circumstance – data access in joint projects is by internal researchers only. In very particular situations, in projects that do not have internal researchers, it may be possible for an external researcher to gain surrogate access to a dataset. These datasets have the “surrogate” classification (see subsection 3.7).

Datasets made available to external researchers are subject to a confidentiality classification as follows: **low**, **medium**, or **high**. If the variables in a dataset are classified as **low** then the data is only anonymized. However, if a dataset is classified as **medium** then the dataset will contain variables with “perturbed” values, not the original values. Datasets classified with a **high** level of confidentiality contain variables with “perturbed” values and are restricted to a subsample of the observations. The external researcher will always have the possibility of running scripts on the original datasets.

3.6 How are projects that involve co-authorship between BdP internal researchers and external researchers handled?

In joint projects where the external researcher does not have access to the data the internal researcher has the responsibility to ensure that the information shared with his/her external co-authors respects the confidentiality requirements associated with the data.

Projects that require data access by BdP internal researchers and external researchers are designated “mixed projects”. If the mixed project is implemented in the external server then the distinction is irrelevant and all researchers are treated as external researchers. However, if the project is in the internal server BPLIM will open an account for the BdP internal researcher(s) and place there all the data requested. External co-authors can be added as users if the datasets made available in that account have a confidentiality classification of “low”. If the account contains datasets with a higher classification level, then the internal user may request the creation of a second “parallel” account. This second account will have both the internal and external researchers as users but the data available will respect the requirements imposed on access for external researchers. It will be the responsibility of the internal researcher to run all scripts on the original data placed on the first account that only he/she has access to. It will also be his/her responsibility to make sure that external researchers do not have any type of access to confidential data; in particular, external co-authors cannot access the project account with the original data, the desktop computer of the internal co-author, or any logs that might contain confidential data. In a mixed project all interaction with BPLIM has to be done via the BdP internal researcher.



3.7 How can an external researcher gain surrogate access to a dataset?

External researchers not involved in joint projects with BdP internal researchers can still gain access to datasets considered to be more sensitive. Appendix A identifies the files that can be accessed this way. To gain surrogate access the external researcher needs to submit a detailed project to BPLIM. The project will be evaluated according to the relevance of the topic for the research agenda of BdP. Only projects deemed relevant will be granted surrogate access. If BPLIM decides to support the project then the external researcher will be assigned a data expert at the microdata laboratory (BPLIM) who will interact with him/her and will help prepare and run the scripts on the original data. All outputs shared with the external researcher are subject to the usual disclosure restrictions. These access conditions for external researchers are not waived in any circumstance.

3.8 How can external researchers work with perturbed data?

A perturbed dataset contains “perturbed” variables which are variables where the original values have been replaced by different values. A perturbed file is easily identified because the file name will contain “..._P...”. Perturbed variables are also clearly identified because their name ends in “_p”.

Datasets with perturbed variables are made available to facilitate the elaboration of the scripts to manipulate/analyze the data. However, results of analysis performed on datasets with perturbed variables are not valid for research purposes. External researchers can always run the scripts on the original data. If the project involves a co-authorship with internal researchers at the bank (“joint project”) then that responsibility falls upon the internal researcher since he/she has access to the original data. If the project does not involve internal researchers then BPLIM staff will run the scripts on the original data. This is a process that takes time as we first need to replicate the results with the perturbed data before moving on to the original data. To speed up the process researchers working with perturbed data are encouraged to discuss the elaboration of the scripts with BPLIM staff.

3.9 Transfer of external files

BdP internal researchers working in the internal server can freely copy files to and from their account in the internal server and thus can merge BPLIM datasets with other data. BPLIM will assist with the anonymization of external datasets to ensure compatibility among datasets. Note that BPLIM identifiers (eg: *tina* and *bina*) are specific to a project and may not be valid if files are exchanged between accounts. If the account is shared with external researchers (“mixed project”) then the internal researcher has to respect the guidelines that apply to external researchers. At the request of the internal researcher BPLIM can anonymize and/or perturb the datasets intended to be used in mixed projects. External researchers may also request that external files be placed on their accounts. These requests need to contain an explanation of the external file as well as a justification for the need. BPLIM staff will assist with the merging of the external datasets with BPLIM datasets. External datasets are typically aggregated but it might be possible to add external datasets with finer granularity. These situations will be analyzed case by case and discussed with BPLIM staff.



4 Data Analysis and Statistical Software

4.1 Tools

Once the account for the project is created the researcher(s) will have access to a Linux environment that has Stata and R available. A number of packages that extend the capabilities of the base distributions for Stata and R have been pre-installed. BPLIM staff will assist if the researcher needs to add additional packages.

5 Output Extraction

5.1 Can the researcher transfer any files from the server?

External researchers are never allowed to transfer files to/from BPLIM's servers. The same policy applies to BdP internal researchers working on the external server.

However BdP internal researchers are free to transfer files to and from their accounts in the BPLIM internal server. When an account in the internal server is created for an internal researcher then the project area from the internal server is mapped to the researcher's personal desktop. This means that internal researchers can easily transfer files to and from BPLIM's internal server including data and output logs. In case the internal researcher is working with external co-authors it will be his/her responsibility to ensure that all files shared with the co-authors respect the guidelines imposed for the release of information.

5.2 What restrictions apply to the release of output logs?

As a general principle an output log should never contain information that reveals individual records of the dataset, and thus it should contain only aggregate-level information. This means that listings of individual records, tables with cells obtained with three or less observations, statistical measures with standard errors of zero, etc., are not allowed. Plain text files (including Latex and comma or tab separated values) are the preferable format but other formats may be acceptable provided that data content can be easily checked (e.g. small Excel files). Graphical outputs should be produced in either ".eps" or ".png" format. In mixed projects the responsibility of checking adherence to these principles falls upon the internal researcher. If requested by the internal researcher, BPLIM will assist with that process. In projects involving only external researchers, it will be the responsibility of BPLIM staff to verify conformability of all outputs. Output verification depends on staff availability and may take longer in periods when the workload is higher. BPLIM staff will only answer requests for output extraction sent by email and will take as long as needed to ensure that all confidentiality requirements are safeguarded.



6 Data Archive and Storage

6.1 Archival and replicability of work.

Users are responsible for keeping a well-documented copy of all code files (syntax files) used to prepare their analysis. Once a project ends BPLIM will keep only a copy of the syntax files and all datasets that were placed in the “initial_dataset” folder. These are typically BPLIM created datasets. All other files will be deleted unless the researcher explicitly requests that certain external file(s) (see subsection 3.9) be archived. Only in exceptional and well justified circumstances will BPLIM agree to archive data files created by the researchers. It is the responsibility of the researcher to ensure that all materials for proper replication of the results are archived in the project. Archives are kept for ten years since the closing of the project.

7 Publication of Research Papers

7.1 Citations and research outputs

All BPLIM datasets should be cited according to the information provided in the data manual. If available, the DOI should be referenced. When the topic analysed by the external researcher(s) bears special relationship to BdP’s statutory tasks the researcher is encouraged to discuss their results with BdP staff. In this case, BdP may ask to see a copy of the work prior to any public release and may provide suggestions regarding the research project. Moreover, in situations where data access is granted based on the relevance of the topic (“surrogate access”) researchers are not only encouraged to discuss their results with BdP staff but must also seek BdP approval prior to any public release of their results. As soon as available, researchers are required to send to BPLIM a copy of all research outputs (working paper, conference proceedings, paper, thesis, etc.) related to the project.

8 Contacting BPLIM

8.1 How can I contact BPLIM?

The preferred way to contact BPLIM is through email: BPLIM@bportugal.pt. For projects already ongoing the subject line should always include the project reference (eg: SEPE_XXX_XX_JDOE). If necessary you can contact us at:

Address: Banco de Portugal
Microdata Research Laboratory
Rua do Almada, 71
4050-036 Porto
Portugal
Ph: +351 222 077 247.



APPENDIX A

Datasets Available at BPLIM

1. Central Balance Sheet Database

The Central Balance Sheet Database is constructed and made available by Banco de Portugal and provides economic and financial information on non-financial corporations operating in Portugal. It contains annual data from 2006 onwards and is mostly based on information reported through *Informação Empresarial Simplificada* (IES, Simplified Corporate Information). Until 2009 the data were reported according to the *Plano Oficial de Contabilidade* (POC) – The National Plan of Accounts. In 2010, a new accounting system, the *Sistema de Normalização Contabilística* (SNC) – The Accounting Normalization System was put in place. The main goal of SNC is to make the reporting of accounting information compliant with International Accounting Standards (IAS). Thus, all data reported after 2009 follow the SNC. This is a structural change that is reflected in the way the dataset is organized and poses some challenges in the comparability of the variables over time. BPLIM provides two different sets of files: the annual firm-level set and the firm harmonized panel. The first data set contains the data as reported while the harmonized panel contains variables that were made consistent over time. For more information see the data manuals.

The **Datasets** available are:

Name	Data Range	Confidentiality Classification	External Researcher Access
Firm Level			
Yearly Data:			
<i>Apr2017 Extraction</i>	2006-2015	Low ⁽¹⁾	Remote
<i>Nov2017 Extraction</i>	2006-2016	Low ⁽¹⁾	Remote
<i>Jun2018 Extraction</i>	2006-2016	Low ⁽¹⁾	Remote
Panel Data Harmonized:			
<i>Apr2017 Extraction</i>	2006-2015	Low	Remote
<i>Nov2017 Extraction</i>	2006-2016	Low	Remote
<i>Jun2018 Extraction</i>	2006-2016	Low	Remote

(1) There are two variables that if requested by the external researcher will be made available in the perturbed format.



2. Credit Register Data

The Central Credit Responsibility (CCR) database reports credit supply by all credit-granting institutions in Portugal. Data is collected monthly with the objective of supporting participants in the risk assessment of credit concession. Banco de Portugal collects information at the credit exposure level but produces several datasets with different levels of granularity: exposure, firm-bank-level and firm-level data. The data sets can be aggregated to quarterly or annual values. For the firm-bank level data BPLIM produces a harmonized panel which can be aggregated to the firm level. All datasets are updated annually. For more information see the data manuals.

Name	Data Range	Confidentiality Classification	External Researcher Access
<u>Firm Level</u>			
Monthly/Yearly Data:			
<i>Apr2017 Extraction</i>	1999-2016	Medium	Remote
<i>Apr2018 Extraction</i>	1999-2017	Medium	Remote
Panel Data Harmonized:			
<i>Apr2018 Extraction</i>	1999-2017	Medium	Remote
<u>Firm-Bank Level</u>			
Monthly/Yearly Data			
<i>Apr2017 Extraction</i>	1980-2016	Medium	co-authorship / Surrogate
<i>Apr2018 Extraction</i>	1980-2017	Medium	co-authorship / Surrogate
Panel Data Harmonized:			
<i>Apr2018 Extraction</i>	1999-2017	Medium	co-authorship / Surrogate
<u>Exposure Level:</u>			
Monthly/Yearly Data			
<i>Apr2017 Extraction</i>	1980-2016	Medium	co-authorship / Surrogate
<i>Apr2018 Extraction</i>	1980-2017	Medium	co-authorship / Surrogate



3. Bank Balance Sheet

The Monetary Financial Institutions (MFIs) Balance Sheet Database reports detailed information on the assets and liabilities of the MFIs operating in Portugal. This data set can be of interest for both research and the monitoring and evaluation of emerging risks in the Portuguese financial system. The dataset is updated annually.

Name	Data Range ²	Confidentiality Classification	External Researcher Access
Monthly Data:			
<i>Apr2017 Extraction</i>	1997-2016	Medium	Remote
<i>Jun2018 Extraction</i>	1997-2017	Medium	Remote

² The data is available from September 1997.



4. Quadros de Pessoal

Quadros de Pessoal (QP) is available through a Protocol celebrated between Banco de Portugal and the Ministry of Labour and Social Solidarity and may only be used for research purposes by internal researchers of the Bank and their co-authors. It covers all public and private firms with at least one employee. QP contains yearly data on firms, establishments, and their workers with detailed information on worker characteristics. For more information see the data manual.

Name	Data Range	Confidentiality Classification	External Researcher Access
<u>Firm Level</u>			
Yearly Data	1982-2013	Low	no
<u>Establishment Level</u>			
Yearly Data	1982-2013	Low	no
<u>Worker Level:</u>			
Yearly Data	1982-2013	Low	no



APPENDIX B

Declaration on Confidentiality and Use of Data

Full name: _____

Institution: _____

Professional address: _____

Personal address: _____

E-mail: _____

Phone: _____

Title and brief description of research project with mention to the Banco de Portugal data to be used:

Purpose of the research project: _____

Termination date of the project: _____



The Researcher undersigned declares, for whatever purposes appropriate, that his use of the information made available for research complies with and respects the following rules:

- 1) The Researcher acknowledges that the data made available:
 - a) Are confidential and belong to Banco de Portugal;
 - b) Must not be used by unauthorized persons, which includes third parties other than the undersigned Researcher to whom Banco de Portugal supplies, or explicitly allows access to, the data.

- 2) The Researcher commits to treating the data with all appropriate confidentiality safeguards, and agrees not to divulge the data to unauthorized persons. This implies that:
 - a) The Researcher cannot duplicate, transfer, divulge, publish or reveal the data to third parties;
 - b) In no circumstances can the data be transferred or made accessible, under any form, from data platforms not belonging to Banco de Portugal;
 - c) The Researcher has only access to anonymised data and shall not make any attempt to identify individuals, firms, financial institutions or other entities described individually by the data.

- 3) The Researcher shall only use the data for scientific purposes; thus, the Researcher cannot divulge the data under their original form or in a way that allows their apprehension by third parties, nor can he use them for other purposes, namely commercial.

- 4) The Researcher must mention the source of the data in any publication stemming from the research project, including, in particular, tables and charts compiled on the basis of the data.

- 5) The researcher accepts that:
 - a) The research project shall be developed in close cooperation with Banco de Portugal staff;
 - b) When the topic analysed by the Researcher bears special relationship to Banco de Portugal's statutory tasks, Banco de Portugal may ask for a copy of the work carried out using the data and make suggestions regarding the research.

- 6) The Researcher understands that any contravention of the abovementioned rules may:
 - a) Be reported by Banco de Portugal to the competent authorities;
 - b) Be subject to civil or criminal prosecution under the applicable law; and
 - c) Result in cessation of access to the data.



- 7) The Researcher assumes full and exclusive responsibility for the use of the data.
- 8) The Researcher will cease to have access to the data as soon as the objective of the research is fulfilled or the termination date of the project, as agreed by the Researcher and Banco de Portugal, is attained. However, this deadline can be extended, in particular to allow for the replication of the results of the research.
- 9) In addition to the dispositions of item 6), the competent court to legally settle any dispute arising from this declaration of confidentiality and use of data is the District Court of Lisbon (or, in Portuguese, *Tribunal da Comarca de Lisboa*), with exclusion of any other jurisdiction.

I declare that the information provided by me is truthful and that I shall observe the rules for the use of the data stated above, to which I fully subscribe.

Place and date

Signature