



BANCA D'ITALIA
EUROSISTEMA

BIRD
Bank of Italy: Remote
access to micro data

Instructions for use
Version 3.6

November 2020

Contents

| | | |
|-------|--|---|
| 1. | Introduction | 3 |
| 2. | Use of the system | 3 |
| 3. | User authorization..... | 4 |
| 4. | Statistical packages supported..... | 4 |
| 5. | Access to the system via e-mail | 5 |
| 6. | Review and system replies via e-mail | 5 |
| 7. | Limitations of the system | 6 |
| 8. | Confidentiality | 6 |
| 9. | General suggestions for use | 7 |
| 10. | Reporting research work carried out | 7 |
| 11. | Assistance support | 7 |
| 12. | Appendix 1: indications for the Stata language | 8 |
| 12.1. | Version supported and system defaults | 8 |
| 12.2. | End-of-line indicator | 8 |
| 12.3. | Access to the datasets | 8 |
| 12.4. | Suggestions for use | 8 |
| 13. | Appendix 2 : Indications for package R..... | 9 |
| 13.1. | Version supported and system defaults | 9 |
| 13.2. | Access to the datasets | 9 |
| 13.3. | Suggestions for use | 9 |

Use of the "BIRD" remote processing system

1. Introduction

The Bank of Italy's BIRD system allows researchers and economists to process data collected through its Survey of Industrial and Service Firms while maintaining the confidentiality of the individual data.

Users carry out their statistical analyses without having direct access to the micro data; they send an e-mail containing a program written in one of the prescribed languages and the system sends back an e-mail with the results of the calculations.

Use of the BIRD system is subject to the Bank of Italy accepting a request for the issue of a user license.

Authors who refer in their publications to results obtained with the BIRD system are required to cite the source (Bank of Italy and the name of the survey); however, they may not involve the Bank of Italy in any way since they are entirely and exclusively responsible for the use of the data.

2. Use of the system

Use of the system involves the following stages:

- a. users first register by filling in a form in which they provide ID data, accepting the conditions of use;
- b. once they have been authorized, users can carry out calculations by sending an e-mail to the system (with a heading that allows the sender to be identified) and a program to be run written in one of the prescribed languages;
- c. the credentials are recognized and the system carries out a series of formal checks on the commands used. If the check is successful, the system runs the program and, after an additional series of checks on the content, the results are sent to the authorized e-mail address.

For a description of the methodology and the information needed for the development of programs, see the documentation provided on the website of the BIRD system and, for what is not covered there, the annual supplements to the Statistical Bulletin devoted to the Survey of Industrial and Service Firms and to the Business Outlook Survey of Industrial and Service Firms.

3. User authorization

Access to the system is obtained by means of an authentication procedure based on the assignment to each user of a set of parameters that must be submitted together with the program for each calculation.

Interested users apply for authorization using the form available on the BIRD website. Every part of the form must be filled in and the application must be signed and accompanied by a copy of a valid ID document of the applicant.

These documents can be sent **by e-mail**, to the following address:

bird_assist@bancaditalia.it.

E-mail applications should contain PDF attachments of a **legible** scan of the paper-based documents requested.

The granting of authorization is notified with an e-mail message sent to the address given in the application form and containing the following parameters:

- **username** (chosen by the user) Please select an alphanumeric sequence (no other characters or punctuation marks).
- **password** (chosen by the user)
- **project** (a conventional name assigned by the system administrator)

The authorized e-mail address can be changed by sending a special e-mail to **bird_assist@bancaditalia.it** (see the section on "Assistance and support"). It is also possible to request the authorization of additional e-mail addresses, as long as they are related to the same applicant.

The authorized user profile is strictly individual and must not be shared, even temporarily. Users are authorized for the period of one year, after which time the authorization will no longer be valid; an extension can be requested via e-mail to **bird_assist@bancaditalia.it** (see the section on "Assistance and support").

4. Statistical packages supported

The supported processing languages are Stata and R. The choice of language is made through the statement ***package=** in the headings. For example, to submit a program in Stata, it will be necessary for the headings to include the statement

***package = stata**

to submit a program in R, it will be necessary for the headings to include the statement

***package = R**

Any other word that appears in the statement ***package =** generates an error message and the program will not be submitted to the system.

More details on the platforms available to the system can be found in the Appendices to this document.

5. Access to the system via e-mail

Users access the BIRD system by preparing a program written in one of the supported languages. The program must be contained in the text of an e-mail that satisfies the following requirements:

- it must be in **plain text, with no formatting (i.e. no colours, no bold characters, no Vcards, no requests for confirmation of reading, no annexes, no disclaimers, and no automatic signatures)**; in addition, researchers should not use characters such as apostrophes and special characters of some languages, such as ü, ß, æ, Å, ç, and ñ;
- it must **contain a heading in the first four lines**, as in the example given below for a user whose name is "johndoe", whose password is "secret", whose project is "invind" and who intends to submit a program written in Stata:

```
*user = johndoe  
*password = secret  
*project = invind  
*package = stata
```

As can be seen, each of the four lines of authorization is preceded by an asterisk. The four lines must follow each other and be separated by a carriage return.

Everything contained in the message after the line following four headings is sent to the processing environment and must therefore be written in the language chosen.

Upper and lower case letters and spaces do **not** count for the headings, although they may do for the language used (see the Appendices).

Users are recommended to check that there are no spaces after the end of each line of the program to be run. because if there are, there could be errors in the calculations.

The e-mail must be sent from the address that has been authorized and that is associated permanently with the authorization credentials supplied. **The system rejects e-mails when the credentials are correct but it comes from an e-mail address different from that which has been authorized.**

The e-mail address to which messages containing processing requests are to be sent is **bird@bancaditalia.it**.

The e-mail server does not process messages which come from persons who are not authorized or who do not comply with the rules referred to above or whose content does not consist of a program written in one of the supported languages.

6. Review and system replies via e-mail

The program is run by the system automatically, in the time required on the basis of the resources available and the demand of other users. At the end of the processing, the output is **reviewed**.

The persons responsible for the running of the system examines the output to check its conformity with the policies for confidentiality and use of the system. Subsequently, the output is sent to the user at the authorized e-mail address or withheld, in which case the user is sent an e-mail giving the reason.

The time needed to process the program depends on its complexity and the concurrent use of the resources, including those for the handling of the incoming and outgoing mail. The time needed for the review depends on the complexity of the processing and any confidentiality aspects that emerge. Normally replies arrive within a very short time, but users are requested to wait at least two working days before asking for information.

The system sends the results of calculations that do not violate the confidentiality of the elementary data automatically; however, any manual security reviews are carried out exclusively on working days from Monday to Friday. At any time the Bank may control processing manually. Any jobs that get blocked at the end of the review process during a weekend or a holiday period are examined as soon as possible.

7. Limitations of the system

It is not possible to submit several programs written in different languages in the same e-mail. By contrast, it is possible to submit several programs written in the same language in the same e-mail.

The Bank reserves the right to set limits in terms of machine time, dynamic memory, mass storage, number of statements of the program, and lines of output, on the basis of the resources available, in order to protect the stability and maximum usability of the system.

Very long outputs are always subjected to a security check; beyond a certain length, currently set equal to 5,000,000 bytes, they are automatically cancelled.

It is not possible to save permanent datasets. Temporary datasets can be saved in a specific way for each application package, as shown in the Appendices.

On grounds of confidentiality, as detailed in the following section, a dedicated interpreter excludes programs from the processing phase that contain certain forbidden commands and words.

Attention is also drawn to the characteristics and limitations of each of the supported platforms, as shown in the Appendices.

8. Confidentiality

In order to permit both the confidentiality of the individual data and sufficiently functional statistical and econometric processing, criteria have been established to ensure confidentiality of individual data.

Processing that could reveal data referring to individual firms is therefore forbidden. For example, the printing of single records or parts thereof is prohibited (e.g., the command "list" in Stata has been inhibited). The system is designed to

prevent the execution of some commands that could result in violation of the confidentiality of the firms in the sample.

The check on the confidentiality of individual data is also carried out for the logic of the program submitted. To this end, the Bank of Italy can verify the content of the calculations made and keep them for an unlimited time; it also reserves the right not to send all or part of the results of calculations in order to protect the confidentiality of firms' data.

The e-mail address **bird@bancaditalia.it** may not be used to send messages other than for the submittal of processing programs according to the standards and formats prescribed in this manual. The address is managed automatically and incoming messages are read and sorted automatically too.

Abuse of the remote processing system can lead to the suspension or revocation of the authorization, without prejudice to any legal consequences.

9. General suggestions for use

In order to optimize the use of resources and minimize the probability of a security check, with the consequent delay, or the cancellation of all the work submitted, it is recommended that users should:

- **not attempt to print elementary data;**
- **not create statistical tables with cells having a very small number of observations;**
- **not create frequency distributions of continuous variables;**
- **avoid creating excessively long outputs.**

10. Reporting research work carried out

In order to permit the collection of bibliographic material on their projects, authors are kindly required to send a copy of the research work carried out using the BIRD system to the following e-mail address: **bird_assist@bancaditalia.it**. Users of LaTeX, where available, are invited to attach a text file containing the bibtex headings of the work.

11. Assistance support

Assistance concerns use of the system and the archives and preferably is provided by e-mail, using the **bird_assist@bancaditalia.it** address. The object of the request should be indicated in the object line. **The address cannot be used to submit calculations.**

12. Appendix 1: indications for the Stata language

12.1. Version supported and system defaults

Version supported: 10.0 for Windows. The system contains all the .ado files normally present in that version of the Stata package. Commands which are automatically placed at the top of each calculation and which cannot be personalized:

```
set more off , permanently;  
set maxiter 300 , permanently;  
set mem 300m;
```

12.2. End-of-line indicator

In Stata programs to be submitted to the BIRD system it is necessary to use the semi-colon (;) to indicate the end of a line. For this purpose it is necessary to put the command "#delimit;" before each calculation. The simple carriage return can cause errors in the recognition of commands. It should also be remembered that each line cannot contain more than 72 characters.

12.3. Access to the datasets

The dataset of the **Survey of Industrial and Service Firms** has the symbolic name **\$db_001** and that of the Survey of Construction Firms has the name **\$db_cos**. The dataset of the Special Survey on the impact of the Coronavirus is accessible under the symbolic name **\$db_iseco**.

The dataset of the **Business Outlook Survey of Industrial and Service Firms** has the symbolic name **\$db_sondstor** for the time stacked database, whereas data collected in the individual waves are contained in the datasets **\$db_sondxxxx**, where xxxx is the reference year of the wave. The dataset of the Business Outlook Survey of Construction Firms has the symbolic name **\$db_sondcost**.

For an optimal allocation of resources it is required that only the variables and the records of interest are loaded with the dataset. Please use a syntax as in the example that follows:

```
use <list of variables of interest>  
if <conditions on variables>  
using $db_001;
```

For real-life examples, please see the dedicated documentation.

12.4. Suggestions for use

- Do **not** use commands capable of violating the confidentiality of individual data (e.g. **list**) or the integrity of the system. Do not attempt to use system commands such as **exit** or the prefix **!**. Programs containing forbidden commands or options are automatically cancelled. The repeated use of forbidden commands or options may lead to revocation of the authorization, without prejudice to any legal consequences.
- Stata distinguishes between upper and lower case letters. It is recommended that users use lower case letters unless expressly requested otherwise.

- Some Stata commands, such as **probit** and **logit**, may produce very long outputs. **Calculations whose output is too long may be subjected to a security check or cancelled.** For some commands, the **“, nolog”** option can shorten the output. In general, appropriate use of the **“noisily”** and **“quietly”** commands can shorten the total length of the output.
- In order to save intermediate datasets, it is necessary to use the symbolic reference **\$mydata**, followed by the username, as in the example given below, where the user “maria” saves the temporary file **myfile.dta**:

```
save $mydata\maria\myfile, replace;
```

Intermediate datasets are cancelled at the end of each job.

13. Appendix 2 : Indications for package R

13.1. Version supported and system defaults

Supported version: 3.3.1 for Windows.

At present, it is not possible to memorize intermediate datasets or other files, such as graphics or PDF files in language R.

13.2. Access to the datasets

The files containing the datasets of the **Survey of Industrial and Service Firms** and of the **Survey of Construction Firms** are named **indann.rdata** and **indann_cos.rdata** respectively. To access the datasets, write:

```
invind <- get(load("indann.rdata"))  
cos <- get(load("indann_cos.rdata"))
```

The dataset of the **Special Survey on the impact of the Coronavirus can be loaded with the following statement:**

```
iseco <- get(load("iseco.rdata"))
```

The dataset of the **Business Outlook Survey of Industrial and Service Firms** can be loaded with the following statement:

```
sstor <- get(load("sondstor.rdata"))
```

whereas the datasets containing data from the individual waves of the survey are loaded using the following:

```
sdXXXX <- get(load("sondXXXX.rdata"))
```

(XXXX is the reference year of the wave).

The dataset of the **Business Outlook Survey of Construction Firms** can be loaded with the following statement:

```
sondcost <- get(load("sondcost.rdata"))
```

13.3. Suggestions for use

- **Do not** use commands capable of violating the confidentiality of individual data or the integrity of the system. Jobs containing forbidden commands or options are automatically cancelled. The repeated use of forbidden commands or options may lead to revocation of the authorization, without prejudice to any legal consequences.
- **Calculations whose output is too long may be subjected to a security check or cancelled.** Users should choose the option for each command that ensures the shortest length compatible with the level of detail required.
- It should be remembered that a job may be cancelled if it involves an excessive use of resources in terms of machine time, dynamic memory or mass storage. To keep the memory space allocated to a minimum it is suggested to work on data frames containing only the variables of interest.