Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC-ECCBSO-CBRT Conference on *"Uses of Central Balance Sheet Data Offices' information"*

Co-organised by the IFC, the European Committee of Central Balance Sheet Data Offices (ECCBSO) and the Central Bank of the Republic of Turkey (CBRT)

Özdere-İzmir, Turkey, 26 September 2016

Matching firm-level data sources at the Statistics Department of Banco de Portugal[1]

Paula Casimiro, Ana Bárbara Pinto and Tiago Pinho Pereira,
Bank of Portugal

[1] This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Matching firm-level data sources at the Statistics Department of Banco de Portugal

Paula Casimiro

Banco de Portugal and Chair of the BACH Working Group, ECCBSO

Ana Bárbara Pinto

Banco de Portugal and ERICA Working Group, ECCBSO

Tiago Pinho Pereira

Banco de Portugal

Abstract

Matching data from the Central Balance Sheet Database (CBSD) with other firm-level data sources for quality control (QC) purposes has been a common practice at the Statistics Department of Banco de Portugal. Data from annual and quarterly surveys of non-financial corporations (NFC) available in CBSD were matched with internal and external firm-level data sources. As internal data sources we have used bank loans granted by resident financial institutions from Central Credit Register (CCR), securities issues from the Securities Statistics Integrated System (SSIS), Monetary and Financial Institutions (MFIs) Interest Rates (MIR), and bank loans granted by non-resident financial institutions and group companies, exports and imports, and trade credits from Transactions and Positions with Non-Residents (*COPE)*, database. As external data sources we have used exports and imports and information related with business demography from Tax Authority and number of employees and wages paid from Ministry of Social Security. Despite some methodological issues that avoid a full comparison between the different sources of information, all sources of information benefit from the cross checking of firm-level data sources. We concluded that matching data from firm-level data sources is of utmost importance to assure the accuracy and reach a high level of quality of the NFC information, which allows Banco de Portugal to publish useful information for firms' decision making such as the Enterprise and Sector Tables and the Central Balance Sheet Studies.

Keywords: firm-level databases, non-financial corporations, data matching

JEL classification: C81

Contents

1. Introduction

Matching data from the Central Balance Sheet Database (CBSD) with other firm-level data sources for quality control (QC) purposes is a common practice at the CBSD of Banco de Portugal.

Every year, data from both the annual and the quarterly survey of non-financial corporations (NFC) are matched with the Central Credit Register (CCR), the Securities Statistics Integrated System (SSIS), the data from Transactions and Positions with Non-Residents (*COPE*, *Comunicação de Operações e Posições com o Exterior*, in the Portuguese acronym) - and the Monetary and Financial Institutions (MFIs) Interest Rates (MIR) in order to assure the accuracy of the information reported by the NFC.

The CCR database contains information about all the loans above 50 Euros granted by resident financial institutions, while the SSIS database contains detailed data on issues and portfolios on a "security-by-security" and "entity-by-entity" basis. The COPE database contains information of flows and positions reported by resident legal entities with yearly transactions with the rest of the world above 100.000 Euros.

During the QC process of the annual and quarterly surveys of NFC, loans reported by firms in their balance-sheets are compared with the information available at the CCR, the SSIS and the COPE databases, while exports and imports of goods and services are compared with the COPE database. This comparison has been very useful for CBSD, although there are some methodological issues that do not allow for a complete matching of the data, especially in the case of the COPE database.

Regarding loans, benefits from the CCR and the SSIS database are twofold: on one hand, they allow the distinction between bank loans and bonds when firms do not specify the sources of their funding; on the other hand, in the case of bank loans, since the CCR only contains loans from resident financial institutions, it is possible to obtain, by a residual approach, the amount of loans granted by non-resident financial institutions to Portuguese NFC. The amount of loans granted by non-resident financial institutions can also be obtained directly from the COPE database. Usually, data on loans granted by non-resident financial institutions obtained by the residual approach matches the data obtained from the COPE database. It is also possible to obtain from the COPE database the intra-group loans from non-resident firms.

With respect to exports and imports of goods and services, the existence of data from COPE for a company that does not report exports and imports of goods and services in the CBSD surveys possibly allows to fill a gap in the CBSD database. However, there are several explanations for the absence of a complete matching between COPE and CBSD database, such as the existence of trade credits, business group relationships or cash pooling.

Monetary and Financial Institutions' (MFIs) statistics have detailed data on new, renegotiated and outstanding loans granted by monetary financial institutions on a "loan-by-loan" and "entity-by-entity" basis. Hence, it is possible to match this information with the implicit interest rates on the accounting information sent by NFC to the CBSD.

Data from Tax Authority includes information on intra-European Union (EU) and extra-EU exports and imports of goods and services, total sales and value added taxes (VAT), R&D tax incentives (deductions to R&D expenditures), business register for VAT purposes, income paid to or received from non-resident entities and interest paid or received by natural resident people.

Finally, data from the Ministry of Social Security contains the number of employees and the wages paid, by firm, on an annual basis.

Throughout the paper, we provide an integrated time series analysis of CBSD, CCR, SSIS, MIR and COPE databases from 2011 to 2015, as well as additional comments on matching databases.

2. Firm-level data sources

2.1. Internal data sources

Statistics based on the CBSD, CCR, SSIS, MIR and COPE databases are regularly published on the Statistical Bulletin and on the BPstat | Statistics Online, the interactive dissemination database available at Banco de Portugal website. In this section, we provide a brief description of each one.

- Central Balance Sheet Database (CBSD)

CBSD exists since 1983, based on accounting data of individual firms. From 2006 onwards, annual CBSD data has improved considerably and has been based on obligatory financial statements, which allowed the monitoring of almost all Portuguese NFC (about 370.000), instead of only a sample of them.

The major goal of the CBSD is to contribute to a better understanding of the operating and financial performance of NFC. CBSD data are useful to produce statistics about NFC, to derive the NFC sector for National Accounts, to estimate several items for Balance of Payments (BoP), to update business registers, and to produce sectoral benchmarks, namely Sector Tables and Enterprise and Sector Tables (Brites, 2013).

Yearly data of the CBSD database is obtained from *Informação Empresarial Simplificada* (IES). IES is a mandatory annual report through which NFC submit their annual accounts (balance-sheet, income statement, statement of changes in equity, cash flow statement and the annex to the financial statements) simultaneously to the Tax Authority, Ministry of Justice, Banco de Portugal and Statistics Portugal.

IES is reported within six and a half months of the economic year end, which, for most enterprises resident in Portugal, corresponds to 15 July of the year following the reference year.

Data reported by enterprises through IES is subject to QC by Banco de Portugal mainly to ensure that the accounting information for the economic year is coherent and complete and that the main aggregates are consistent throughout the years.

QC comprises the matching of data reported through IES with other internal data sources of Banco de Portugal, such as CCR, SSIS, MIR and COPE, as well as with external data sources, such as Tax Authority and Social Security.

- Central Credit Register (CCR)

Following Casimiro (2013), the Portuguese CCR database was launched in 1978, first including only the credit liabilities of NFC and, from 1993 onwards, also the credit liabilities of households.

Reporting institutions to the Portuguese CCR are banks, savings banks and mutual agricultural credit banks (MFIs), other non-monetary financial institutions and public agencies that grant credit, and NFC buying loans from the resident financial sector.

The main purpose of the CCR is to contribute for the financial stability by helping financial institutions in assessing the credit risk of their current or new credit clients, since they can access CCR data. Insurance companies undertaking credit and bond insurance can access CCR data, although they do not report it.

Data reported to Portuguese CCR include the borrowers ID (for residents, the tax identification number is used), the credit drawn (amounts outstanding at the end of the month), credit undrawn (irrevocable credit commitments), personal guarantees (potential credit liability), type or purpose of the loan, collateral (type and value), periodic repayments (for some types of loans granted to private individuals), original and residual maturities, credit defaults and write-offs, and specific flags for Banco de Portugal internal use of the data (e.g. securitized loans and loans used as collateral in Eurosystem financing operations).

- Securities Statistics Integrated System (SSIS)

The SSIS of Banco de Portugal was established in 1999. It was created to store, manage and explore data on securities issues and portfolios on a "security-by-security" and "investor-by-investor" basis, excluding investors in the households sector, whose data are aggregated by the investor's country. This database comprises securities other than shares and shares and other equity. The assembled data include, on a monthly basis, stocks and transactions, with the ISIN code being used for the identification of the securities (Dias, 2013).

Regarding issues, the SSIS collects data on securities issued by resident entities in Portugal, irrespectively of the fact that those issuances take place in the Portuguese market or in external markets. A multiplicity of sources are used such as the Lisbon Stock Exchange, the Portuguese Securities Market Commission, the Portuguese Treasury and Debt Management Agency and commercial databases.

In the case of portfolios, comprehensive information on holdings of domestic and foreign securities by resident investors and holdings of domestic securities by non-resident investors is collected. Data are reported mainly by custodians (e.g. banks, dealers and brokers). Direct reporting by resident investors with relevant portfolios deposited abroad is also applicable.

The leading aim of SSIS is the production of statistics on issues and portfolios of securities, the design of "from-whom-to-whom" tables crossing issuers and holders, and the supply of input data for MFIs, BoP and National Accounts statistics.

- Monetary and Financial Institutions (MFIs) Interest Rates (MIR)

Besides micro data regarding the end-of-month balance-sheet of MFIs (mainly deposits received and loans granted), these institutions also communicate individual information concerning banking interest rates on new and renegotiated loans to NFC.

According to Santos (2013), Banco de Portugal created this new requirement in June 2012 with the aim of obtaining representative data on new loan operations, in a context of financial stability assessment. This new requirement only applies to MFIs granting at least 50 Million Euros per month in new loans to NFC. Furthermore, solely euro denominated operations and loans to euro area resident entities are taken into account.

Reported data includes the date of the operation, maturity of the loan, initial period of interest rate fixation, amount, annualized interest rate, the existence or not of collateral, the nature of the loan (new or renegotiated), borrower ID and residence.

Information on the interest rates of outstanding loans is also available.

- Transactions and Positions with Non-Residents (COPE)

According to Marques (2011), the collection and compilation of BoP data was set in 1993, based on monthly reports by resident banks, which communicated and classified transactions with non-residents on their own behalf and on behalf of their customers. Also, the report of transactions with non-residents settled without the intermediation of the resident banking system was mandatory and it was done by direct reporting to Banco de Portugal.

From 2013 onwards, the system of communication to Banco de Portugal changed, giving rise to direct reporting by economic agents on monthly transactions and positions with non-resident counterparts (so-called *COPE*, *Comunicação de Operações e Posições com o Exterior*, in the Portuguese acronym). Entities with transactions with non-residents above 100.000 Euros per year started to report and classify their transactions and positions with non-residents directly to Banco de Portugal, even if they have the intermediation of the resident banking system. Reports by resident banks without the classification of transactions are now only used to validate information submitted by entities.

Reported data is very granular and includes exports and imports of goods and services (including travel and tourism), rights and operations over tangible and intangible assets, unilateral transfers, real estate investment, shares, units of participation and other equity securities, debt securities, performing and non-performing loans, trade credits, bank deposits, margin accounts, financial derivatives and employee stock options and transfers between accounts breakdown by nature (asset or liability), maturity (short or long term), direct investment relationship (no relationship, voting rights lower than 10% or voting rights greater or equal than 10%), and transaction type (capital or income).

2.2. External data sources

- Tax Authority

Banco de Portugal has been receiving firm-level data from the Tax Authority since 2014, in the sequence of an information exchange agreement. Data is sent four-times a year and it is available from 2006 onwards.

This database includes the monthly amount of extra-EU exports and imports of goods and services, the quarterly amount of intra-EU exports and imports of goods and services, the fields of the VAT return (e.g. amount of sales, intra-EU imports of goods and services and other operations that originate VAT), the annual amount of tax incentives for R&D, the register of active companies for VAT purposes, with the date of beginning and end of activity, income paid to non-resident entities, interest on savings paid to resident natural persons, and income obtained from non-resident entities.

- Social Security

Data from Social Security was available until 2013 and comprised the annual number of employees and wages paid by firm.

3. Results

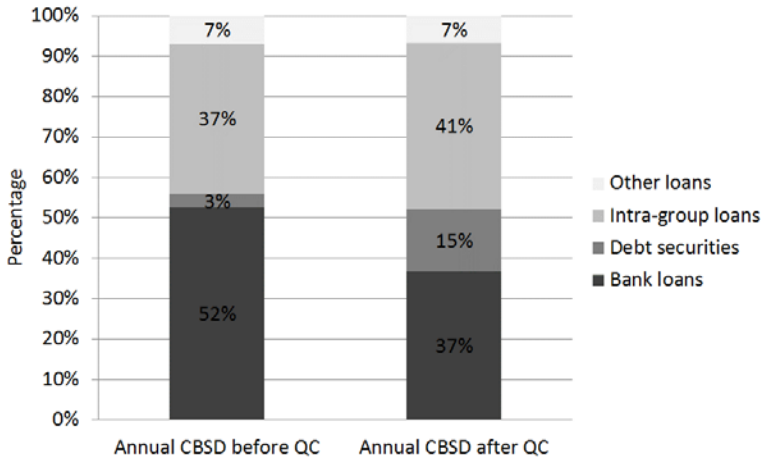
In this section, we provide some examples of data matching and how it improves the quality of the databases. The QC process of the CBSD occurs every year and quarter according with the periodicity of the data sources. Annual QC is done after the submission of the IES and it involves not only human resources from Banco de Portugal but also a group of undergraduate students that manually validate the information sent by a sample of NFC[1].

3.1. Borrowings structure and their sources

Chart 1 below shows the total borrowings structure of the Portuguese NFC in 2015, breakdown by sources. For total borrowings we mean the sum of bank loans, debt securities issued, loans from group companies and other loans. During the QC process, data on bank loans, debt securities and intra-group loans is matched with the CCR, the SSIS and the COPE databases, which explains the differences between the initial and the present situations. The CCR database provides information on loans granted by resident financial institutions, while SSIS database provides data on debt securities issues and COPE database on non-resident banks and intra-group loans.

[1] This sample is generated from the universe of more than 370.000 companies, according to some criteria. Usually, the final sample of validated companies represents more than 1% of the universe and more than 50% of the turnover of Portuguese NFC.

Chart 1: Borrowings structure of NFC (2015)

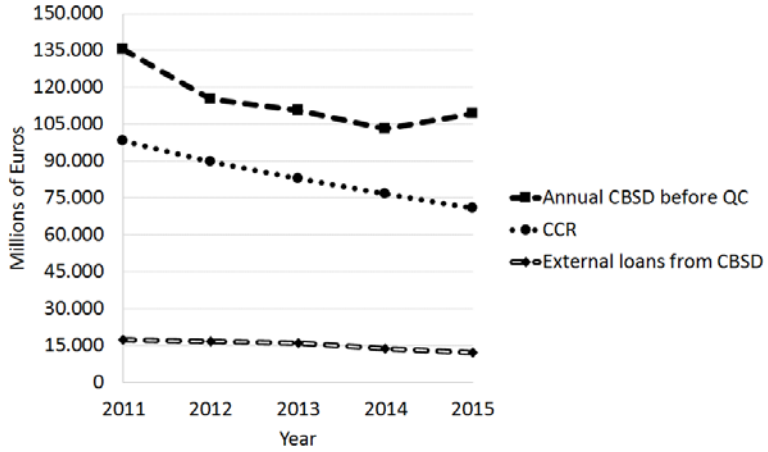


3.2. CBSD vs. CCR database

Chart 2 shows the amount of loans granted by financial institutions in the CBSD before QC and in the CCR database, as well as loans granted by non-resident financial institutions[2] from the CBSD, which are not available at the CCR database. External loans usually account for 15% of the loans granted by financial institutions and for 5% of total borrowings.

As it can be observed, loans granted by financial institutions in the CBSD before QC are higher than in the CCR database. This happens because of the existence of external loans and because NFC usually do not detail their sources of financing and include all of them in a single item, which is loans from financial institutions. External loans can be obtained through the comparison with the COPE database or the NFC' annual report, if available. Otherwise, NFC can also be contacted to clarify the data.

Chart 2: Loans granted by financial institutions before QC



[2] Also denoted as "external loans" throughout this paper.

Chart 3 shows the situation after QC. As it can be seen, the amount of loans granted by financial institutions at the CBSD moves closer to the amount in the CCR database and, if we deduct to this amount the value of external loans we can observe an almost perfect matching between databases.

Chart 3: Loans granted by financial institutions after QC

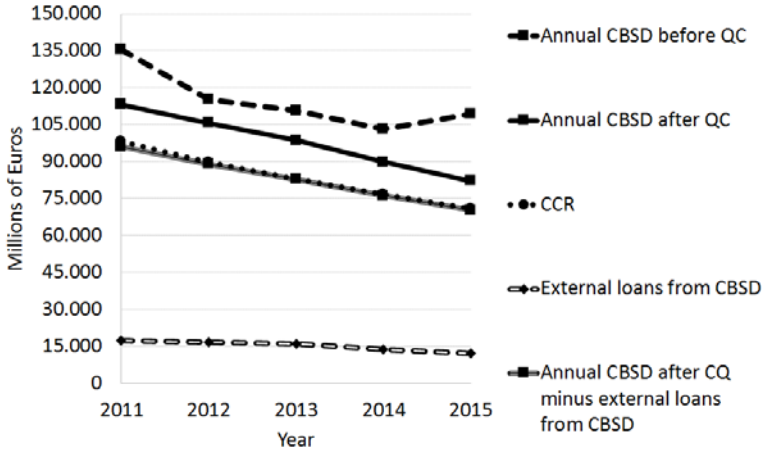


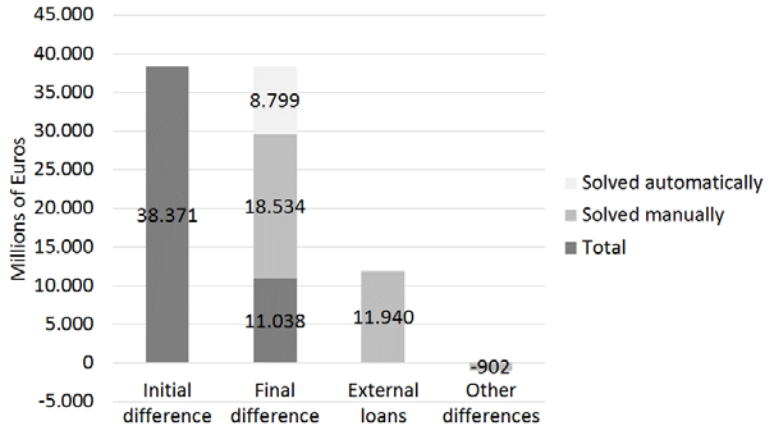
Chart 4 shows initial and final differences between the CBSD and the CCR database before and after QC.

In a first stage, NFC with large differences between the CBSD and the CCR databases and without external loans in the previous years in the CBSD[3], and for which the amount of loans outstanding in the CCR database is less or equal than the borrowings reported through IES are treated automatically, with the amount of loans outstanding in the CCR database being incorporated in the CBSD. According to Chart 4, this automatic procedure solved about 23% (corresponding to 8.799 Millions of Euros) of the initial difference between the databases.

In a second stage, NFC with large differences regarding CCR database which are not solved automatically are distributed for manual QC. Manual matching solved around 48% (corresponding to 18.534 Millions of Euros) of the initial difference between the databases. At the moment, the difference between the two databases remains at 11.038 Millions of Euros (29% of the initial difference), which corresponds approximately to the amount of external loans (11.940 Millions of Euros). As pointed out before, these loans are not generally covered by the CCR and are manually inserted according to firms' annual reports, direct contact or matching with COPE database.

[3] NFC with external loans in the previous years are distributed for manual QC.

Chart 4: Initial and final differences between CBSD and CCR databases (2015)



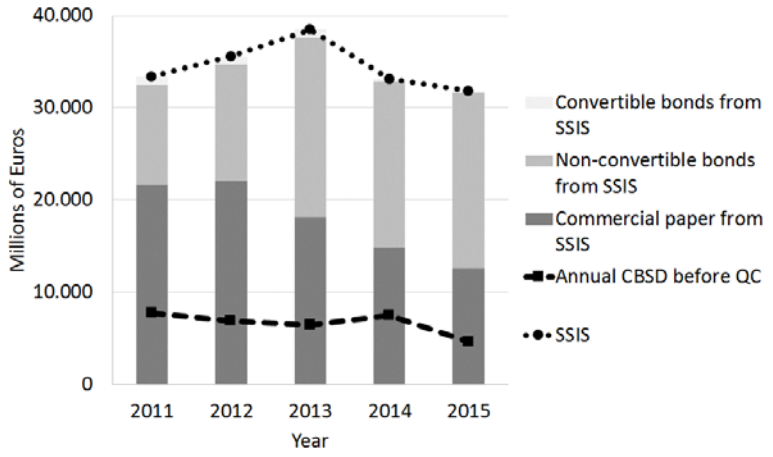
Besides external loans, there is a fraction of the final difference that is explained by specific circumstances such as time lags between the CBSD and the CCR databases, bankruptcy[4] or lawsuits against banks (Other differences, which represent 2% of the initial difference in absolute value). Also, there will always be a small difference between databases, even after QC, because only firms with material differences between CBSD and CCR database are distributed for manual QC.

3.3. CBSD vs. SSIS database

During the QC process, matching with SSIS database is also done. Chart 5 shows the comparison between the outstanding amount of debt securities in the CBSD and in SSIS database. The amount from SSIS is generally much greater given that NFC incorrectly recognize the majority of their funding as loans from banks. On one hand, NFC usually do not detail their sources of financing and include all of them in a single item, which is loans from financial institutions. On the other hand, there are cases in which firms contact a bank to contract a loan, the bank agrees, and then securitizes the loan due to tax advantages. This loan will be consider by the firm as bank loan but in fact in a debt security.

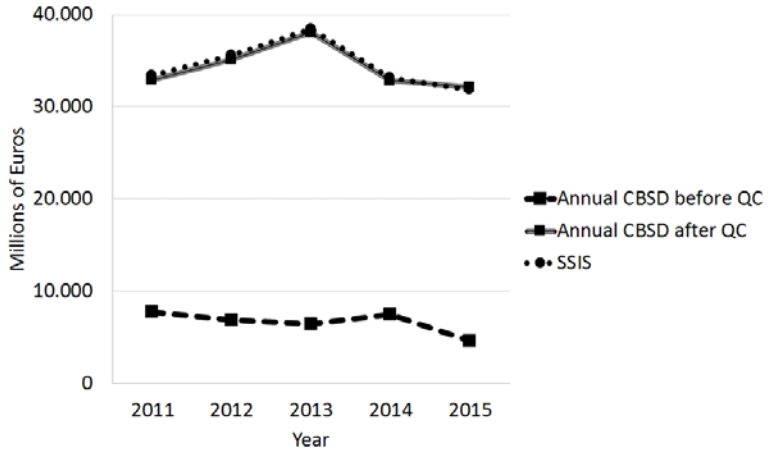
[4] Bankrupt firms usually submit their IES with many figures equal to zero, namely loans granted by financial institutions, while in the CCR database these loans remain. This happens because banks that report to the CCR continue to recognize these loans in their balance-sheets.

Chart 5: Financing through debt securities before QC



After QC, the amount of outstanding debt securities at the CBSD nearly overlaps the amount at the SSIS database (Chart 6). Here, the QC is also organized at two stages. First, for NFC with large differences between the CBSD and the SSIS database and for which the amount of outstanding debt securities in the SSIS database is less or equal than the borrowings reported through IES, automatic matching with the SSIS database is done. Then, the remaining situations are distributed for manual validation.

Chart 6: Financing through debt securities before QC



It should be stressed that automatic matching of CCR and SSIS database is done simultaneously to prevent unbalanced balance-sheets. Every year and quarter there is automatic incorporation in the CBSD of data from other sources namely CCR and SSIS databases, as well as bank and group loans from COPE database. As a result of this incorporation, the new total borrowings of CBSD could go above or go below the original ones, generating unbalanced balance sheets.

If, after matching the data from other data sources with the CBSD, the new total borrowings of CBSD go above the original ones and this excess is due to the original Other loans item of total borrowings, then the Other loans are adjusted to make the new total borrowings equal to the initial ones. If adjusting the initial Other loans item of total borrowings does not solve the unbalance, empirical evidence tells us that firms recognize borrowings in other liabilities' items which not total borrowings. Hence, the automatic procedure will take values from these other liabilities' items to

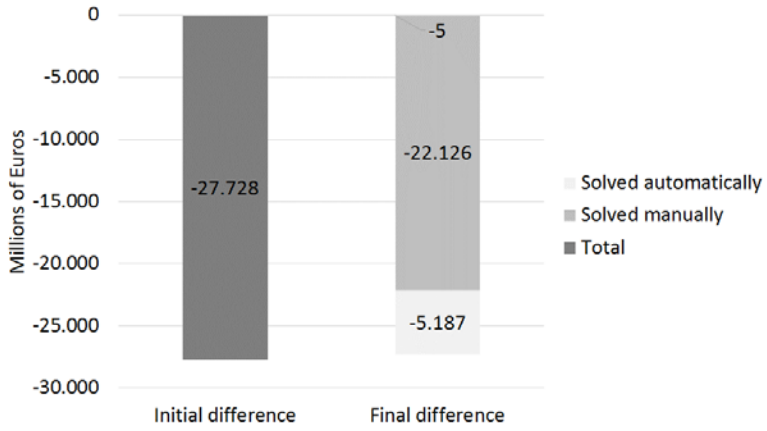
total borrowings in order to match the data available from the other sources and will solve the unbalance.

In the cases that unbalance could not be solved acting this way, data from other sources is only automatically matched until the amount that prevents unbalanced balance-sheets and that respects the maximum difference allowed between CBSD and the other sources. The firms which are not solved automatically are distributed for manual validation.

If the new total borrowings go below the original ones, original total borrowings are kept equal, with the bank loans being matched with the CCR and the COPE databases, intra-group loans with the COPE database, and the loans through debt securities matched with the SSIS database. The excess in the original total borrowings is distributed for intra-group loans and other borrowings, according to the borrowings structure of the previous year.

Chart 7 shows initial and final differences between the CBSD and the SSIS database before and after QC. From the initial difference of -27.728 Millions of Euros (SSIS greater than CBSD), almost 80% (corresponding to 22.126 Millions of Euros) were manually inserted and 19% (corresponding to 5.187 Millions of Euros) were automatically inserted into CBSD after consultation of the SSIS database. In the end of the QC, the two databases are almost fully matched, with the amount of outstanding debt securities in CBSD being slightly greater than in SSIS database (which means a final difference of -5 Millions of Euros, 0,02% of the initial difference in absolute value).

Chart 7: Initial and final differences between CBSD and SSIS databases (2015)



As in the case of comparisons with the CCR database, a complete matching is not possible mainly because some firms report liquid (deducted of fees) or mark-to-market values to the CBSD database while in the SSIS database figures appear at their gross or nominal value.

On the other way around, SSIS database also benefits from inputs of the CBSD. For example, in the case of debt issuance by companies belonging to the same business group, the consultation of the annual reports of companies during the QC of the CBSD annual data allows the identification of the correct issuer, which sometimes is incorrectly identified at the SSIS database. If non-resident companies are, indeed, those which issue the securities, the issue should not be considered at the SSIS, since they are non-resident. However, during the QC process of the CBSD, if

the consolidated annual report indicates that securities are, actually, issued by a domestic firm of the group, contacts between the two areas are made and SSIS database is updated if necessary.

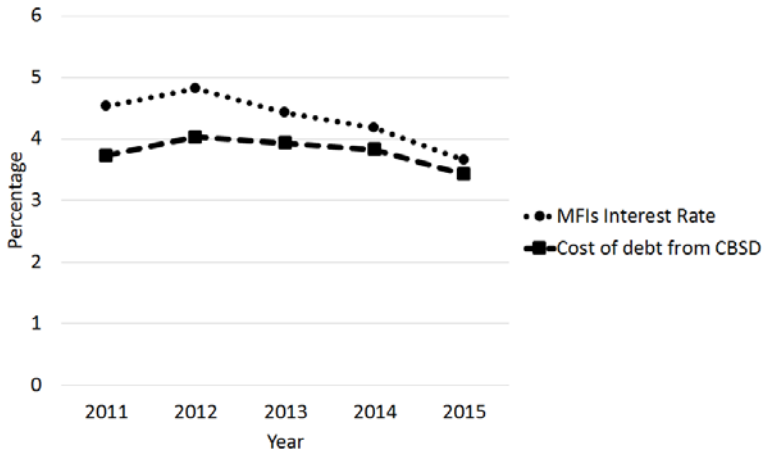
3.4. CBSD vs. MIR database

MIR database is used in the QC process of CBSD to detect eventual cases of a wrong report by firms. Chart 8 illustrates the interest rate of outstanding loans granted by MFIs to NFC, as well as the cost of debt from CBSD, defined as the interest paid divided by total borrowings.

Although this cost of debt contains other sources of financing besides loans from banks (and eventually with lower interest rates) it is a proxy for the interest rates that are actually paid by firms and, thus, it can be compared with MFIs' interest rates.

As it can be observed, in recent years, the overall cost of debt of NFC is not too far from the interest rates that are actually paid by NFC to MFIs.

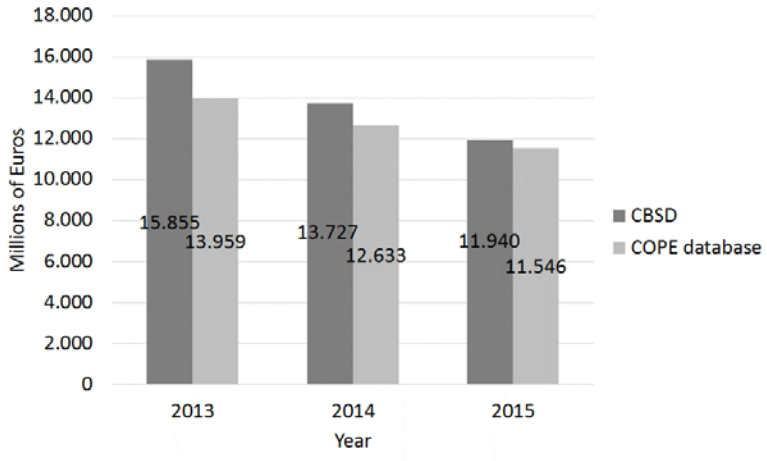
Chart 8: Cost of debt of NFC



3.5. CBSD vs. COPE database

Chart 9 shows the amount of external loans in the CBSD and COPE database. As it can be seen, external loans in the two databases are almost completely matched, which is not a surprise given that, on the one hand, COPE database is one of the sources used to fill the gaps regarding external loans in the CBSD, and, on the other hand, information from the CBSD is also used by BoP for QC purposes.

Chart 9: External loans in the CBSD and COPE database



On the other hand, comparisons between the amount of exports and imports of goods and services from the CBSD and the COPE and the Tax Authority databases are also made during the QC of the CBSD annual data. Contrarily to what happens with the CCR and SSIS databases, an almost complete matching is not possible because there are several methodological differences between the databases. However, information is used in the CBSD whenever it is needed to fill in missing values or confirm non-expected values according to the historical data for a given firm.

First, not all of the amounts recognized in the income statement correspond to effective financial flows. There are fractions of exports and imports that are not immediately paid. CBSD works according to an accounting perspective, while COPE database only recognizes exports or imports when there are financial flows. Hence, one of the sources of differences between the two databases is the existence of trade credits. Indeed, if we subtract the amount of trade credits to the exports and imports of the CBSD, there is an approximation to the COPE figures (Charts 10 and 11).

Besides trade credits, causes for an incomplete matching between CBSD and COPE database include the existence of transactions with resident branches of non-resident firms and with non-resident branches of resident firms[5], intra-group cash pooling, misclassification of transactions, time lags between the two databases, same operations reported in different companies of the same business group in each of the two systems, and the utilization of non-resident bank accounts owned by resident firms.

Regarding Tax Authority data, information sent to Banco de Portugal is divided by intra-EU and extra-EU trade. However, it was detected for some companies a duplication of values, given that transactions for which the goods are sent to an extra-EU location, but the counterpart is an intra-EU company were considered in both extra-EU and intra-EU systems by the Tax Authority. Also, non-resident branches of resident firms are treated by the Tax Authority as non-resident entities, while in accounting they are included in the report of resident firms. Anyway, this information is used in certain situations as a reference for CBSD QC.

[5] Non-resident branches of resident firms are considered by COPE database as non-resident entities, while in IES an accounting perspective prevails and they are considered as part of the resident firm.

Chart 10: Exports of goods and services from CBSD and COPE database

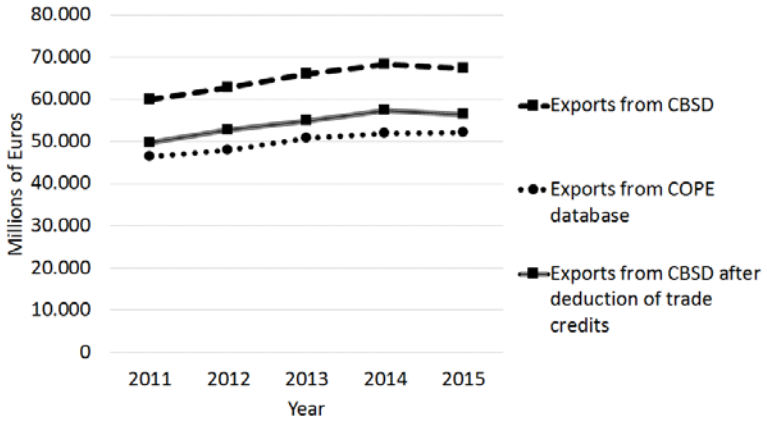
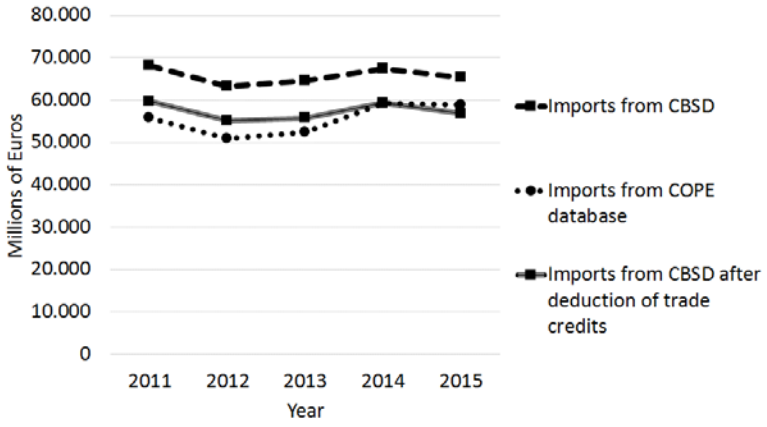


Chart 11: Imports of goods and services from CBSD and COPE database



3.6. CBSD vs. Social Security data

Social Security data used at the CBSD were a result of a pre-processed query to the original Social Security data. From this query, only a file with the total number of employees and their wages by firm and year was made available to the CBSD. As a result, some differences arouse, especially regarding the number of employees, given that, through IES, firms report their average number of employees during the year and not the total one, as in the pre-processed file from the Social Security data.

Consequently, the number of employees from the Social Security data is greater than the CBSD. However, the trend is very similar as it can be seen in Chart 12. Wages from the CBSD and the Social Security are presented in Chart 13.

Chart 12: Number of employees from the CBSD and the Social Security databases

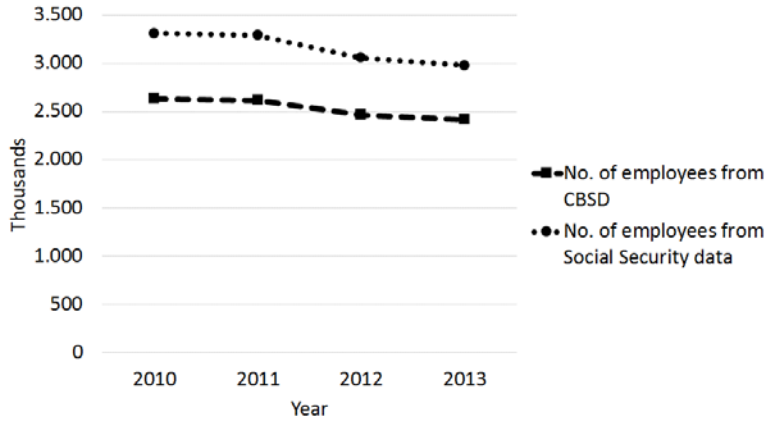
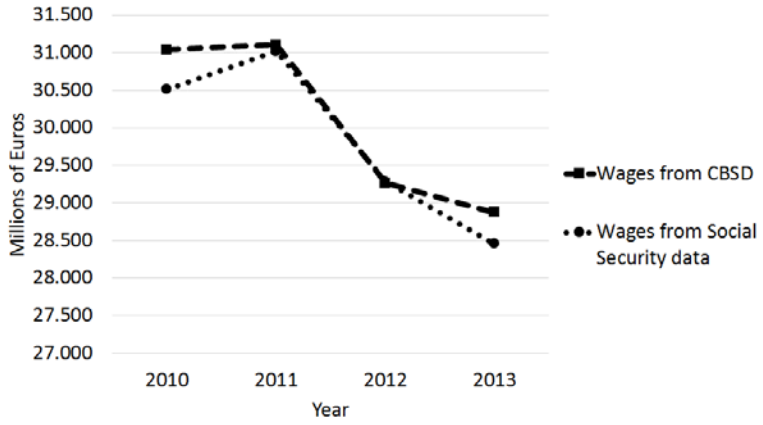


Chart 13: Wages from the CBSD and the Social Security databases



4. Conclusions

Matching firm-level databases is essential to ensure the quality of statistics. This paper presents the specific case of the CBSD of Banco de Portugal, which benefits from the existence of both internal and external databases that allow filling the gaps of the information submitted by NFC.

The Statistics Department of Banco de Portugal manages several databases, namely the CBSD, the CCR, the SSIS, the MIR and the COPE databases. Besides internal databases, the CBSD also has or had access to external data sources such as the exports and imports from the Tax Authority and the number of employees from Ministry of Social Security. Every year, during the QC process of data sent by NFC, the Central Balance-Sheet Office uses information of these sources to improve the quality of its data.

It is important to stress that not only the Central Balance-Sheet Office, but also all the other divisions benefit from the integration and interchangeability of the databases managed by the Statistics Department of Banco de Portugal. Frequently, inputs from one division are used for other divisions to improve their data, by matching or by validation of their own reports.

Data from the CCR and the COPE databases on loans granted by financial institutions and group companies, and from the SSIS on debt securities issued by NFC are automatically and manually matched with the CBSD, overcoming misclassifications in the reported data that otherwise would only be solved by direct contact to firms, which would be a very slow process.

Sometimes, it is not possible to fully match the databases. If, in the case of CCR, nearly all the amount that is not matched derives from external loans, which are usually available from the COPE database, in the case of exports and imports the sources of differences between CBSD and COPE database is broader. Trade credits are the main justification, but there are many others such as transactions with resident branches of non-resident firms or between non-resident accounts, intra-group cash pooling and equal operations reported in different companies of the same business group.

It was also illustrated that independently of the database used, the trends are fairly the same, which is important in the sense that, even in the absence of some source, the other sources available allow the characterization of a given phenomenon.

To sum up, manage and match several databases contributes to improve the quality of statistics. In the particular case of the CBSD, it was demonstrated that matching databases allows filling the gaps of NFC reports, although there are some methodological differences, whose knowledge also contributes to better understand the boundaries of each data source.

REFERENCES

Brites, M. (2013), "Efficient ways of dealing with accounting data from enterprises", Workshop on Integrated Management of Micro-databases, Porto, June 2013

Casimiro, P. (2013), "A flexible approach to credit statistics: The use of the Portuguese Central Credit Register for statistical compilation", Workshop on Integrated Management of Micro-databases, Porto, June 2013

Dias, L. T. (2013), "Micro-databases on securities statistics as a key issue for satisfying user needs", Workshop on Integrated Management of Micro-databases, Porto, June 2013

Marques, C. (2011), "Guidelines for a flexible and resilient statistical system: the architecture of the new Portuguese b.o.p/i.i.p system", 58th World Statistics Congress of the International Statistical Institute, Dublin, August 2011

Santos, C. (2013), Bank interest rates on new loans to non-financial corporations - one first look at a new set of micro data, Financial Stability Report of Banco de Portugal, May 2013.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC-ECCBSO-CBRT Conference on *"Uses of Central Balance Sheet Data Offices' information"*

Co-organised by the IFC, the European Committee of Central Balance Sheet Data Offices (ECCBSO) and the Central Bank of the Republic of Turkey (CBRT)

Özdere-İzmir, Turkey, 26 September 2016

Matching firm-level data sources at the Statistics Department of Banco de Portugal[1]

Paula Casimiro, Ana Bárbara Pinto and Tiago Pinho Pereira,
Bank of Portugal

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Matching firm-level data sources at the Statistics Department of Banco de Portugal

Ana Bárbara Pinto • Banco de Portugal and member of the ERICA Working Group, ECCBSO

BANCO DE PORTUGAL
EUROSISTEMA

IFC / ECCBSO / CBRT Conference on "Uses of Central Balance Sheet Data Offices' information"

Özdere-İzmir, September 26th, 2016

Outline

1. **Firm-level data sources at the Statistics Department of Banco de Portugal**

   a. **Internal data sources**
      i. The Central Balance-Sheet Database (CBSD)
      ii. The Central Credit Register (CCR)
      iii. The Securities Statistics Integrated System (SSIS)
      iv. Monetary and Financial Institutions (MFIs) Interest Rates (MIR)
      v. Transactions and Positions with Non-Residents (COPE)

   b. **External data sources**
      i. Tax Authority
      ii. Social Security

2. **Some results from matching firm-level data sources**

3. **Conclusions**

Internal data sources

Central Balance-Sheet Database (CBSD)

- Created in 1983
- Contains **individual and consolidated (IFRS and National GAAP) accounting data on non-financial corporations** (NFC)
- Useful for the production of **statistics about NFC and sectoral benchmarks**, the derivation of **NFC sector for National Accounts**, the **estimation of several items for BoP**, for **updating business registers** and **risk assessment**

Central Credit Register (CCR)

- Launched in 1978
- Contains information about all the **loans above €50 granted by resident financial institutions** (e.g. borrowers and lenders ID, amount, guarantees, maturity)
- Contributes for the financial stability by **helping financial institutions in assessing the credit risk** of their current or new credit clients

Internal data sources

Securities Statistics Integrated System (SSIS)

- Established in 1999
- Contains detailed data on **issues** and **portfolios** on a **"security-by-security"** and **"entity-by-entity"** basis
- Allows the production of **statistics on issues and portfolios of securities**, the design of "**from-whom-to-whom" tables crossing issuers and holders**, and the supply of **input data for MFIs, BoP and National Accounts** statistics.

MFIs Interest Rates (MIR)

- New requirement created by Banco de Portugal in June 2012, with the aim of obtaining **representative data on new loan operations**, in a context of financial stability assessment
- Applies to MFIs granting at least 50 million euros per month in new loans to NFC resident in the euro area
- Reported data includes the **maturity** of the loan**, initial period of interest rate fixation**, **amount**, **annualized interest rate**, **borrower ID** and **residence**

Internal data sources

Transactions and Positions with Non-Residents (COPE)

- Set in 1993, based on monthly reports by resident banks and directly from some large entities
- Since 2013, all entities with yearly transactions with the rest of the world above €100.000 started to report and classify their **transactions and positions with non-residents** directly to Banco de Portugal
- Reported data includes **exports and imports** of goods and services, **loans**, **trade credits**, and several other operations and its breakdown by nature, maturity, direct investment relationship and transaction type

External data sources

Tax Authority

- Received at Banco de Portugal since 2014
- Contains **data from 2006 onwards**
- Includes, among others, **extra-EU and intra-EU exports and imports** of goods and services, the fields of the VAT return, the annual **amount of tax incentives for R&D**, and the **register of active companies** for VAT purposes, with the date of beginning and end of activity

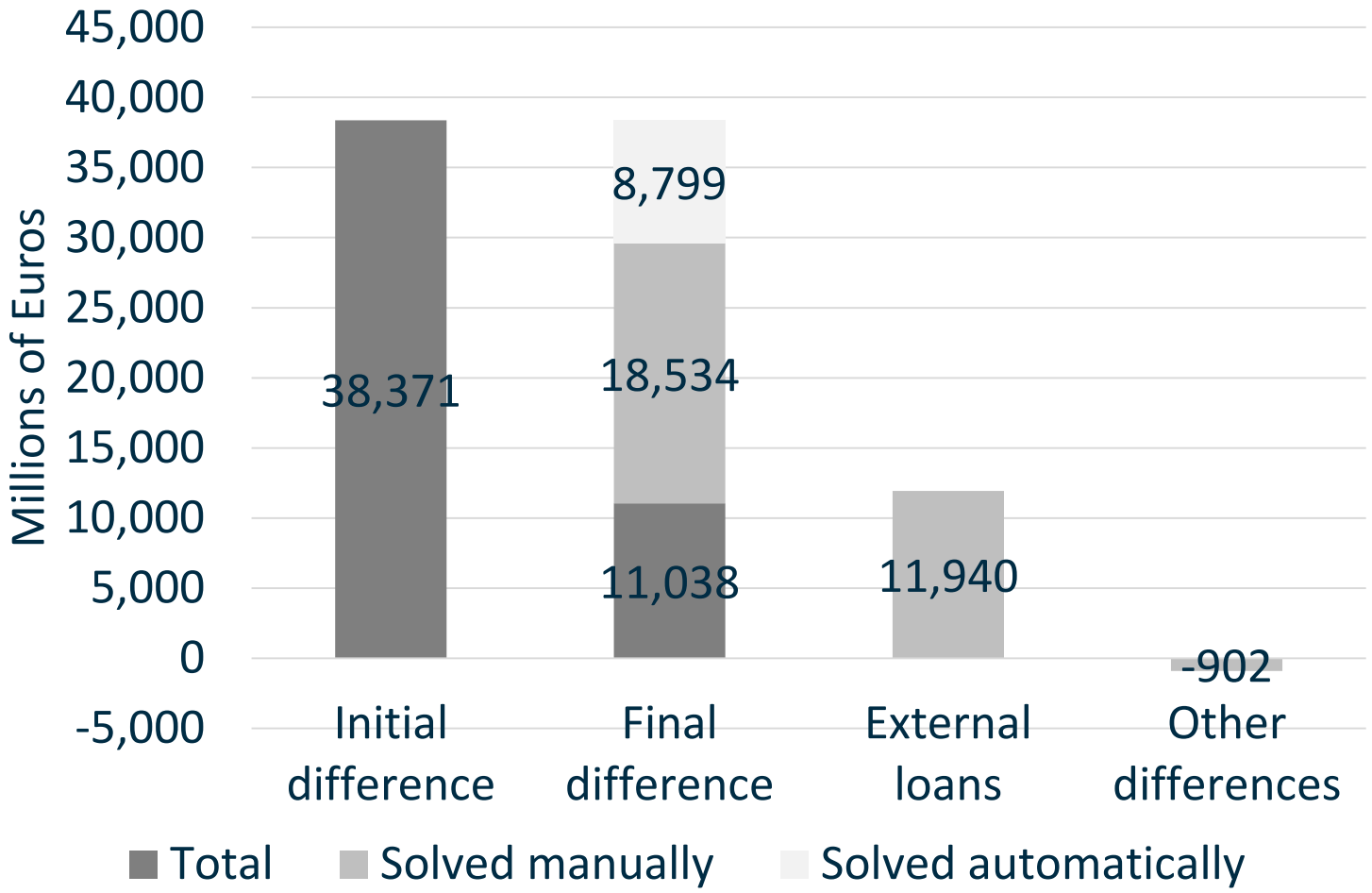
Social Security

- Received at Banco de Portugal until 2013
- Contains data on the **annual number of employees** (paid and unpaid), **number of hours worked**, and the **amount of wages paid** by firm

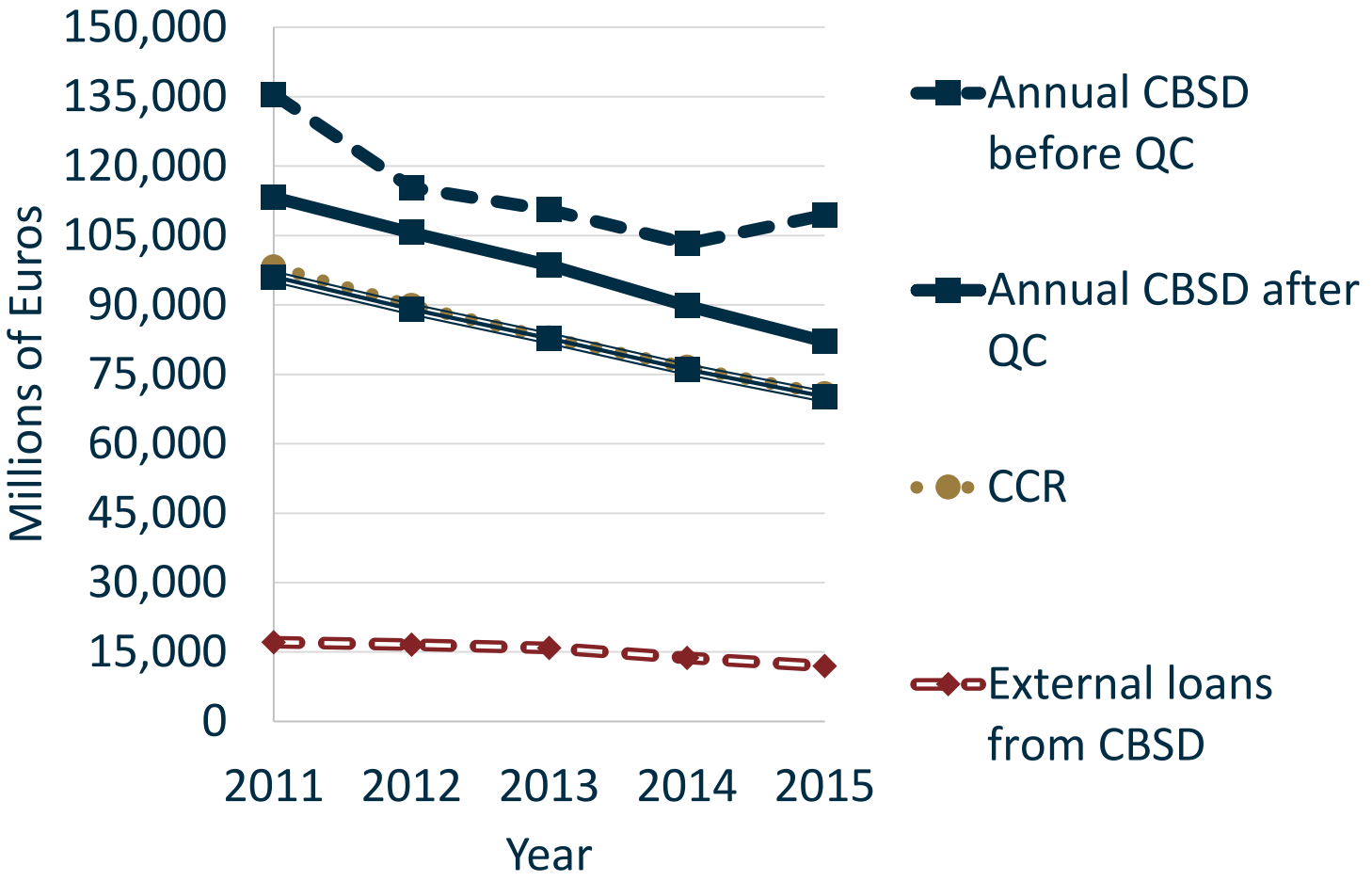
- **Every year, data submitted by NFC to the Central Balance-Sheet Data Office of Banco de Portugal is subject to a quality of control (QC) process**

- **In this process, data from the firm-level data sources previously identified are matched with the CBSD and contribute to fill the gaps of CBSD**

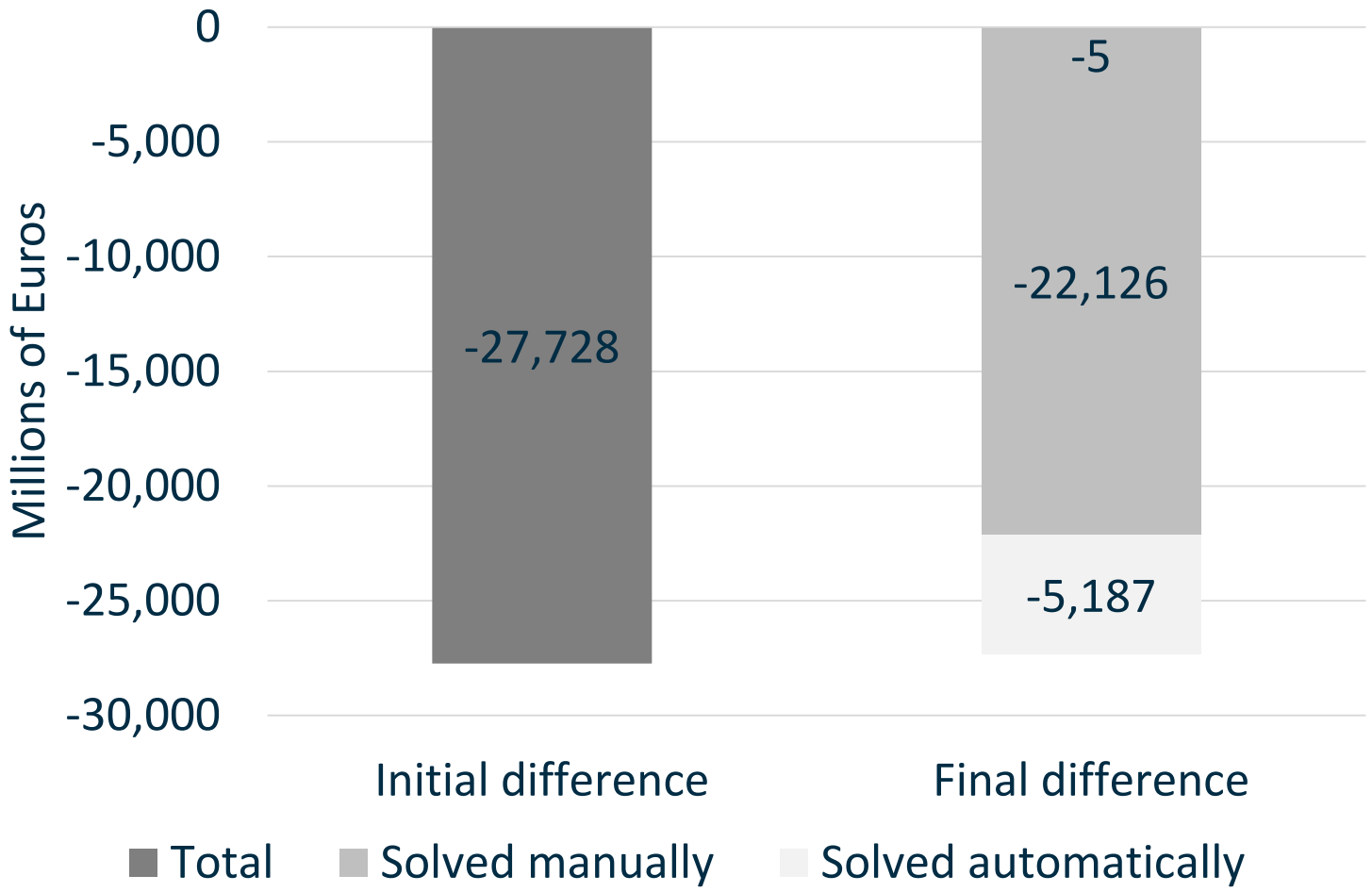
Results: CBSD vs. CCR database (2015)



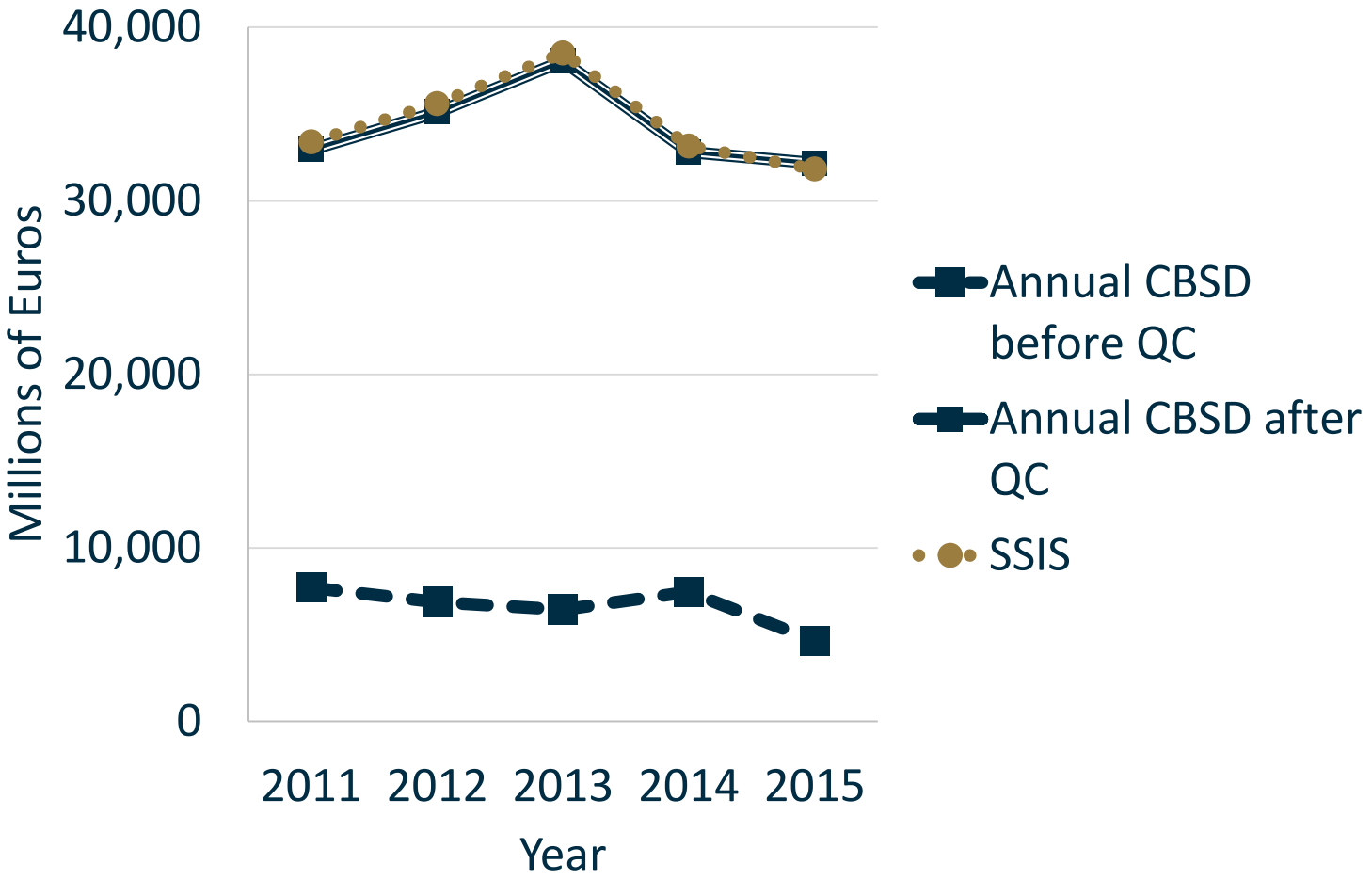
Results: CBSD vs. CCR database



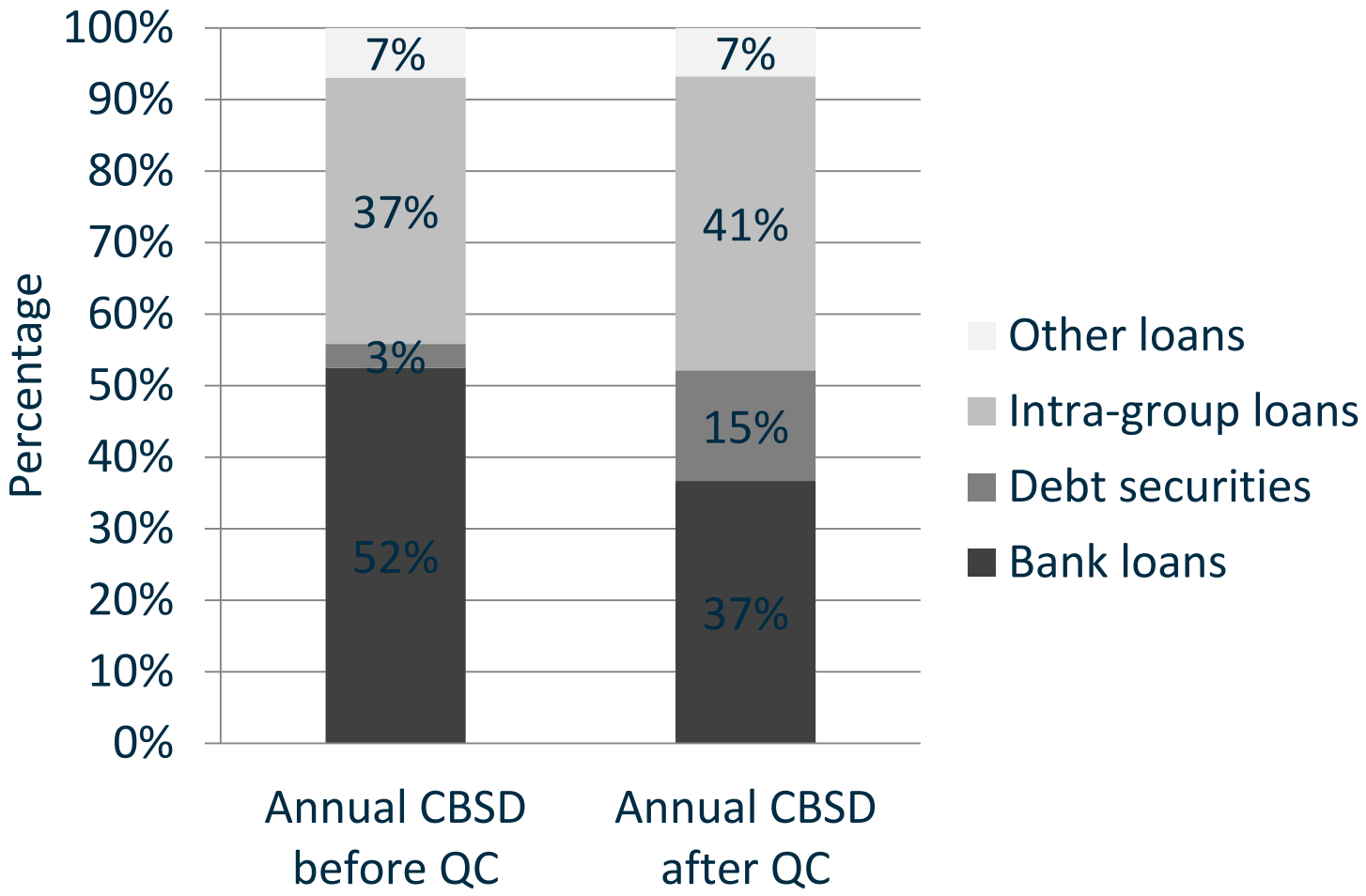
Results: CBSD vs. SSIS database (2015)



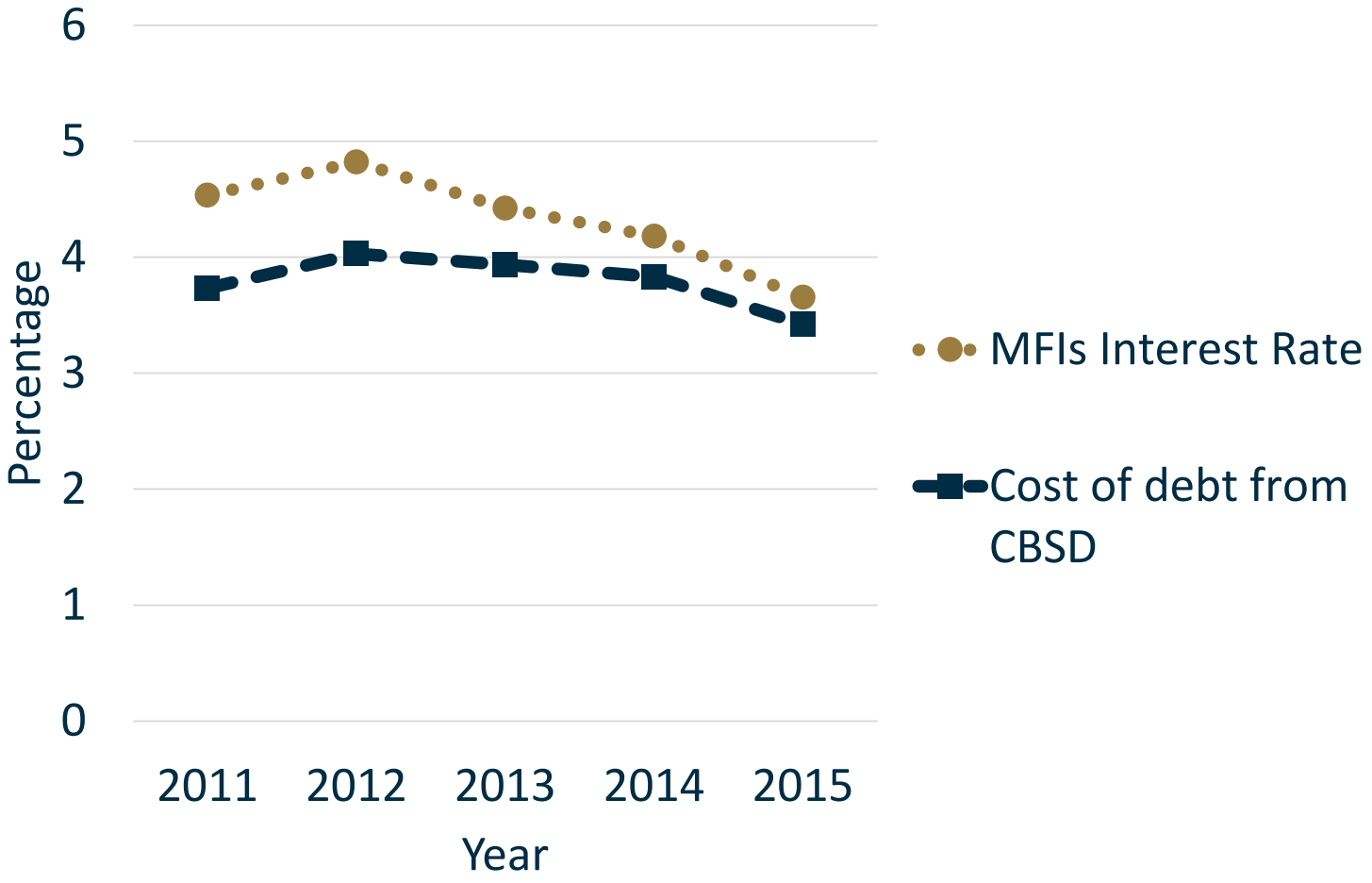
Results: CBSD vs. SSIS database



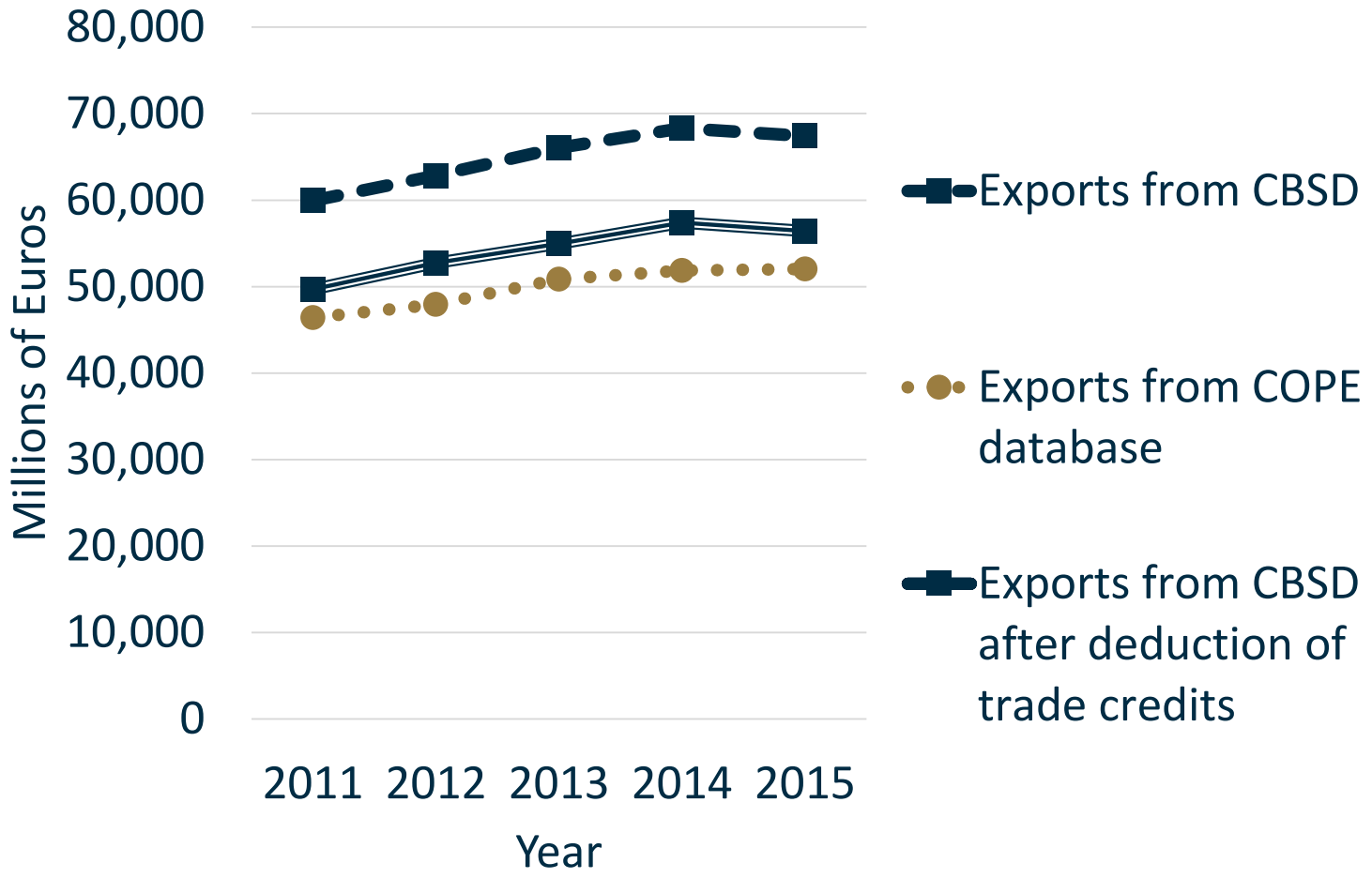
Borrowings structure of NFC (2015)



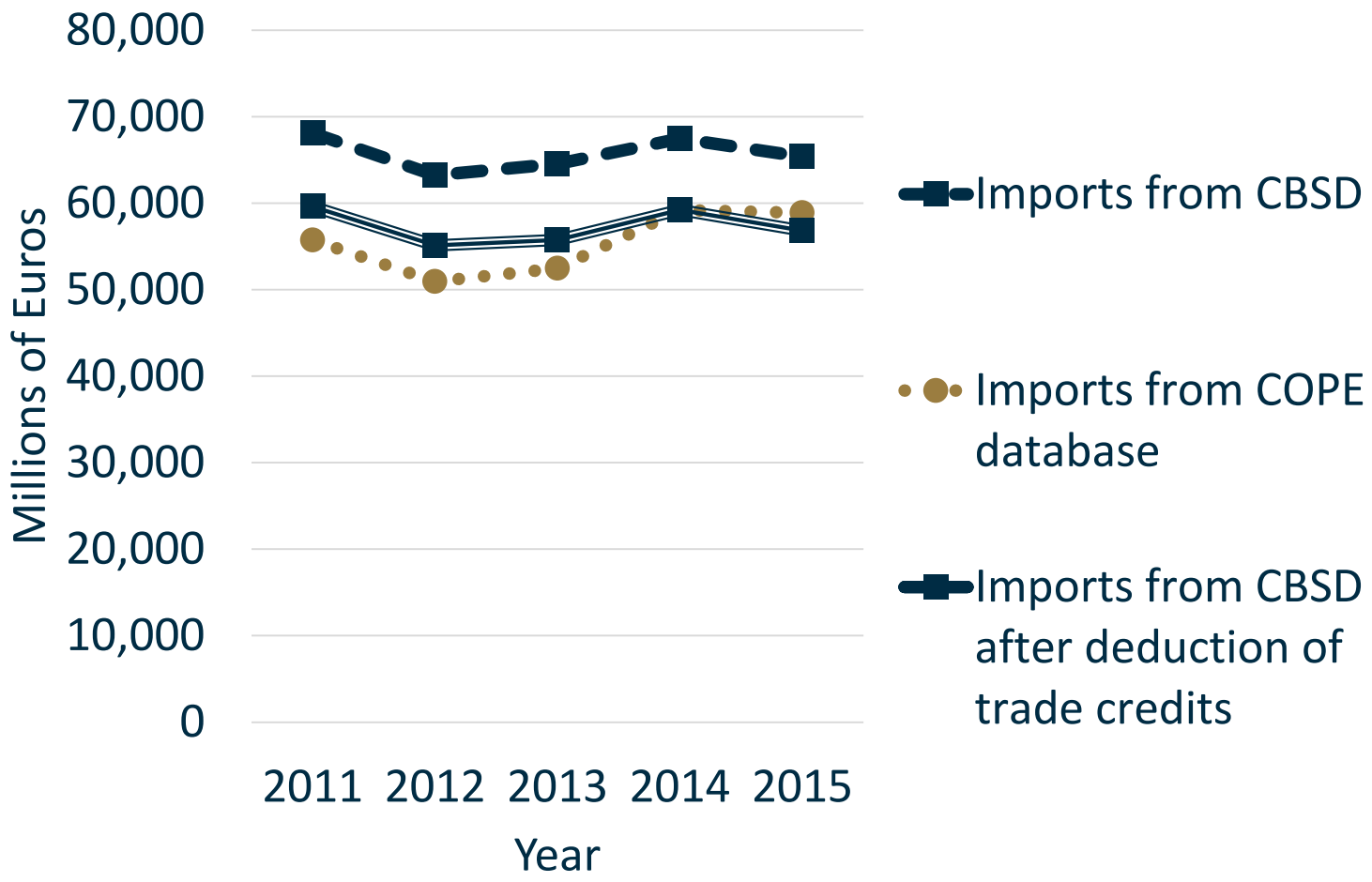
Results: CBSD vs. MIR database



Results: CBSD vs. COPE database



Results: CBSD vs. COPE database



Conclusions

1. **Matching firm-level databases is of crucial importance for the quality of statistics**

2. **Banco de Portugal manages a wide range of internal databases with the information of ones being used as an input for others**

3. **A complete matching is not always possible due to methodological differences between databases**

4. **After the quality control, even in the case of incomplete matching, all the databases show the same trend**

Thank you for your attention!

apinto@bportugal.pt