

The use of security-by-security databases for portfolio investment statistics

João Cadete de Matos¹, Paula Casimiro² and Maria do Carmo Aguiar³

1. Introduction

Statistical data on securities, periodically released by the Bank of Portugal, are compiled on the basis of the Securities Statistics Integrated System (SIET – *Sistema Integrado de Estatísticas de Títulos*). This system was developed by the Statistics Department of the Bank of Portugal, with the purpose of gathering in a single repository all the information deemed necessary to comply with reporting requirements on securities. SIET makes it possible to meet user needs, at both the national and international level. Quite ambitious in its aims, the system has been a challenge for data quality managers and a source of opportunities for data “explorers”.

One of the statistical outputs that benefited from the development of this system was the portfolio investment data collection system for balance of payments (BOP) and international investment position (IIP). Until 1999, the Portuguese portfolio investment (PI) statistics relied on an asymmetric system for assets and liabilities. Although in the latter case, the data collection system implemented since 1991 was already based on a security-by-security data model, on the assets side the inexistence of a unique and standardised identifier prevented the application of this method, and so data was collected from respondents aggregated by type of security, country of the issuer and currency of denomination. Another distinctive feature of both systems was the periodicity of the data collected. While monthly flows were available for both PI assets and liabilities, in the case of end-of-period positions the data collection ranged from monthly data for the liabilities side, to annual data for the assets side.

As already mentioned, the existing data collection system was implemented in the early 1990s, and by the end of the decade it was necessary to make some changes and adjustments, mostly due to the need to improve the periodicity – of the end-of-period statistics – and coverage – in terms of the variables collected – of the PI assets.

With developments in financial markets worldwide, securities statistics have increasingly gained importance. Therefore, subjects related to coverage, quality and harmonisation of securities statistics produced in the various countries are a growing concern at the international level and, in particular, within the scope of the European System of Central Banks (ESCB). In this context, integrated statistical systems enable a more efficient and harmonised production of statistical data. SIET, being an integrated system that includes data on issues and portfolios and covers all the economy’s institutional sectors, makes it possible to cope efficiently with most information requirements in the field of securities statistics.

This paper overviews SIET architecture: inputs, processing and enrichment modules, and outputs. It describes the Portuguese experience in compiling PI statistics, specifically the benefits of using a security-by-security database. A glance into the near future foresees the

¹ Director, Statistics Department, Bank of Portugal, joao.cadete.matos@bportugal.pt.

² Head, Balance of Payments Statistics, Bank of Portugal, paula.casimiro@bportugal.pt.

³ Head, Securities Statistics, Bank of Portugal, maria.carmo.aguiar@bportugal.pt.

integration of information from the Centralised Securities Database (CSDB), an ongoing project of the European Central Bank (ECB).

2. SIET features and main components

SIET is an information system that stores data on securities issues and portfolios on a security-by-security and investor-by-investor basis, except for investors in the households institutional sector, whose data are aggregated by the investor's country. This means that data considered relevant for statistical analysis are collected, validated and stored for each security, each issuer and each investor. The existence of a reference database with individual information on securities and issuers allows the collection of statistical information from reporting entities on a security-by-security basis. This approach implies lower reporting costs, given that there is no need for reporters to aggregate background information according to multiple criteria. Furthermore, it enables better information monitoring and a greater flexibility when exploring data and building statistical analysis.

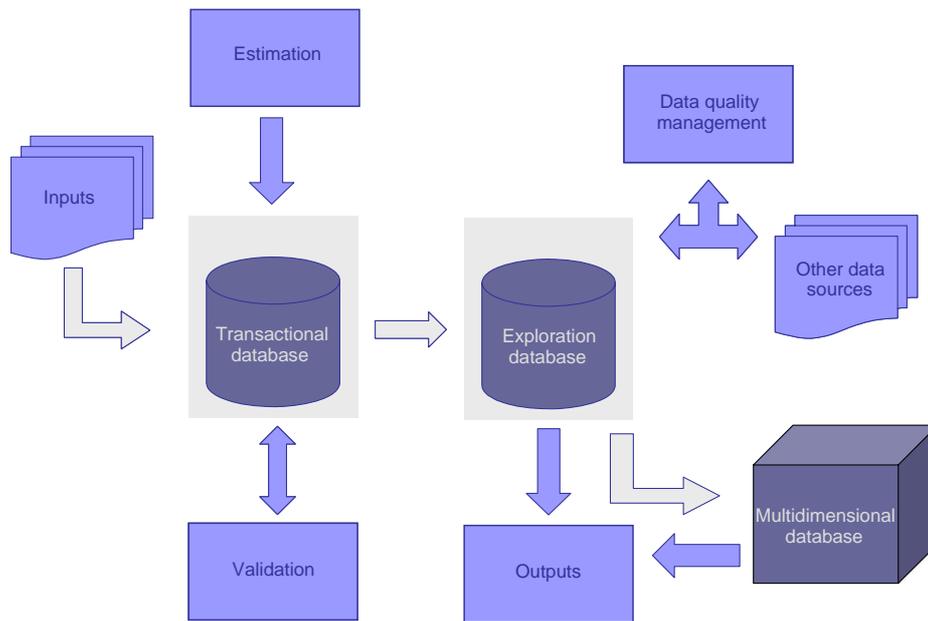
SIET comprises two information segments: one on securities issues and the other on securities holders. In the segment on securities issues, information is collected on securities issued by resident entities in Portugal, issues taking place in either the Portuguese market or external markets. Data on issues are collected from several sources, such as the Institute of Registries and Notaries, the Ministry of Justice, the Securities Market Commission, and the Portuguese Treasury and Government Debt Management Agency, among others. In accordance with the provisions of Article 13 of its Organic Law, Banco de Portugal shall ensure the production of securities statistics covering issues by Portuguese residents and it may require of any public or private body the direct supply of information deemed necessary for the compilation of these statistics.

At the securities portfolios level, detailed information is collected on investments by residents in domestic and foreign securities, as well as on the portfolios of non-resident investors in domestic securities. Data are reported according to a Bank of Portugal Instruction on "Securities Statistics – Transactions and Positions". Additionally, information on the features of foreign securities held by resident investors is obtained from commercial databases.

SIET replaced a relatively large set of procedures for the collection of data on securities, while extending the coverage of reporting to all investing sectors and enabling a better quality control. As regards data on issues, the procedures for researching and collecting information underwent some evolution, and reporting on securities' holders was totally changed. In fact, the report on securities portfolios had previously covered only information on external transactions and on operations conducted by monetary financial institutions (MFIs). Today all sectors are covered.

Figure 1 illustrates SIET architecture. The system relies on two relational databases and one analytical database. Collected data are stored and validated (first level of quality control) in the "transactional database". Data estimation of missing information is also done in this database. Validated and "enriched" data are copied daily to the "exploration database". A second level of quality control is performed on aggregate data, by carrying out consistency tests and comparisons with other information sources. Statistical outputs are produced from the "exploration database" as well as from the "multidimensional database". This analytical database was developed recently and is a quite powerful tool since it enables user-friendly multidimensional analysis of the information.

Figure 1
SIET architecture



SIET stores information on the type of instrument, the institutional sector and the residency of the issuer/investor, prices (quotations), transactions and positions associated with securities issues (issues, redemptions and outstanding amounts) and transactions and positions associated with securities portfolios (purchases, sales, stocks). Classification of securities and entities follows the European System of National and Regional Accounts (ESA/95). Securities are preferably identified through the ISIN code (International Securities Identification Number), and resident issuers/investors through the NPC (Portuguese acronym for the fiscal number of collective entities). Standard unique identification codes are fundamental for sharing and integrating information from different sources. The ISIN code accomplishes this purpose for securities. Unfortunately, for entities (such as issuers or investors) there is no such code at an international level.

SIET was intended to be a system that would make it possible to address all the Bank of Portugal's needs at the level of statistical information on securities. The development and implementation of the system was quite challenging and ambitious, but the outcome was very positive.

While preparing the project, reporters were contacted and the new reporting scheme was discussed. One may think that asking the reporters to send individual information on transactions and positions would be burdensome. However, most reporters were also developing their own information systems, and sending individual information was easier and less expensive than aggregating it according to several different statistical criteria. Compilers at the Statistics Department needed to handle much more data. Gradually, methods were developed for rapidly identifying possible errors or discrepancies. As in all new systems, there was a learning curve.

SIET is an open system in the sense that new components may be developed and integrated. The component related to the estimation of missing information is currently being enhanced and, in the near future, information on foreign securities will be obtained from the CSDB (see section 6. The near future), benefiting from the increased quality of this information.

3. The Portuguese experience in compiling PI statistics

There are different possibilities available when designing a data collection system for portfolio investment statistics, in the context of BOP and IIP. The possible systems may differ according to the targeted reporting agents, periodicity and level of aggregation, and the corresponding results vary in terms of implementation and running costs, data availability and quality, and response burden.

3.1. Selection of a data collection model for PI statistics

A data collection model for PI statistics can be defined as a combination of several features: the level of detail of the information collected (aggregated or on a security-by security basis); the type of information collected (both stocks and flows or collecting one and deriving the other); the collection method used (census or a sample survey); and the reporting channel (indirect reporting – settlement or custodian based – or end investor direct reporting).

Aggregate reporting has the advantage for the compiler of reduced costs of implementation and maintenance, resulting in a relatively small amount of data to keep. However, it also holds the risk of potential miscalculation or the use of non-generalised aggregation procedures by the different reporting entities, and it carries greater difficulties in cross-checking the data and in reconciling flows and stocks. Another non-negligible aspect is the greater risk of misclassification or double-counting between portfolio investment and direct investment, since this distinction will have to be implemented by each individual respondent and it may provide limited information about the weight of a given investor in a company's equity capital. For respondents, aggregate reporting usually also means a greater reporting burden in terms of details and breakdowns to be reported, the need to keep and maintain (in every respondent's system) a security-by-security database from which to derive such breakdowns, and the need to make adjustments in the reporting systems every time new or additional output requirements emerge.

On the other hand, security-by-security reporting ensures accuracy and consistency of the data, although it implies a shift of costs and work from the respondent to the compiler in terms of aggregation procedure and maintenance of an individual securities database. The implementation of new requirements becomes more flexible and, in some cases, may not even imply the need to introduce changes in the respondents' reporting systems. This type of system can be used to derive flows from high-frequency stock data, reducing the reporting burden for reporting agents and allowing for quality checks at a very detailed level. The reporting burden will also be reduced since the amount of detail (in terms of breakdowns) to be reported by respondents decreases. Finally, security-by-security reporting is useful for the calculation of interest on an accruals basis, and it may support synergies with other statistics, such as money and banking statistics and national financial accounts. The main disadvantages of security-by-security data collection models are the considerable costs of setting up and maintaining them, and their dependence on the availability of unique securities identifiers.

According to the targeted respondents, three major reporting channels can be distinguished. The first option is indirect settlement-based reporting by domestic banks for their own transactions and transactions on behalf of their clients. This alternative has the advantages of keeping the size of the reporting population relatively small while providing high-frequency, timely data. It is easily adaptable to security-by-security reporting and carries minor problems concerning double-counting between portfolio and direct investment. The main problems come from the widespread use of netting and clearing techniques, which prevent the collection of gross investment and disinvestment, and the need for complementary reporting (eg for settlements through accounts with foreign banks). In addition, pure stock statistics have to be collected separately, through one of the other possible channels.

A second option is direct reporting by resident issuers and end investors, which can ensure full reconciliation between flows and stocks and the collection of related income on an accrual basis. The distinction between direct and portfolio investment does not constitute a problem either. The major drawback of this alternative is the potentially large size of the reporting population, namely in the case of households. Also in the case of some specific sectors it may be difficult to receive timely and high-frequency data. The implementation of security-by-security reporting may be more difficult for sectors unfamiliar with this way of reporting and storing of information and, finally, statistical principles and methodology can differ from accounting principles used by a great number of respondents.

The third option is the indirect reporting by custodians or other financial intermediaries involved in securities transactions and holdings. This reporting channel has the same advantages of the first alternative (timely and high-frequency data, relatively small reporting population, easy to adapt to security-by-security reporting, allowing for micro-checks of the data) and at the same time it permits a full reconciliation between stocks and flows. However, it will require some complementary information collected directly from the end investors in the case of securities held in custody abroad. Additional challenges will be the exclusion of repo-type transactions/positions and direct investment holdings.

The selection of a direct or indirect reporting scheme depends, of course, on the national specificities, like the size of the targeted population or the reporting practice. Direct reporting is more suitable for banks' own holdings but indirect reporting may be the only practical approach for households. For other sectors, the most suitable reporting channel depends on several factors, such as the average size of companies. Indirect reporting has advantages in terms of timeliness, efficiency and adaptability to a security-by-security system. However, there may be difficulties in collecting specific data such as repo transactions, or in distinguishing between portfolio and direct investment, and it will have to be supplemented with direct reporting in some cases, taking special care to avoid gaps (lack of coverage) or overlaps (double-counting).

3.2. The Portuguese model

The need to change the portfolio investment data collection system for BOP and IIP purposes led to the deep consideration of several dimensions of the problem, including the selection of a more appropriate level of detail and reporting channel, as described above. The experience gathered from the simultaneous existence of a security-by-security (sec-by-sec) reporting system (for PI liabilities) and an aggregated one (for PI assets) facilitated the choice of a data model of the sec-by-sec type. Some benefits of a sec-by-sec system as compared to an aggregated approach were evident at that time. On the compilers' side, the quality of the final statistics and the data control checks are facilitated and enhanced if items of data are collected on an individual basis. On the respondents' side, the need to aggregate the data means that each one of them will have to keep a database of individual securities and run aggregation procedures, increasing by the number of respondents the workload needed to produce these statistics. The contacts held with respondents confirmed that they preferred a security-by-security solution, and by that time the widespread use of the ISIN code in the financial markets overcame the practical difficulties of implementing such a system in the case of resident's investment in foreign securities.

As to the selection of the respondents, the existing system was based on indirect reporting by resident custodians, complemented by direct reporting from end investors holding securities in custody abroad; it appeared that this would continue to be an appropriate solution, especially in terms of reliable and timely data.

Portugal participates in the Coordinated Portfolio Investment Survey (CPIS) of the International Monetary Fund (IMF) since its first edition in 1997, and also plans to participate in the forthcoming Coordinated Direct Investment Survey (CDIS). These initiatives are

important in promoting the availability and consistency of bilateral data and enhancing the quality of the information available to meet users' demand. However, their objectives depend to a great extent on the number of countries participation and, above all, on the common methodologies and definitions used. For both questionnaires, a sec-by-sec system is the only option that guarantees that securities are correctly classified and that the risk of double-counting between portfolio and direct investment is minimised.

4. Consistency across statistics

As mentioned previously, one of the major drawbacks of a sec-by-sec reporting system is the significant costs involved in its development and further maintenance. The solution envisaged for reducing these costs was to adopt the system also for other statistics compiled in Banco de Portugal Statistics Department. In this context, SIET was implemented, as described above. This option carried additional benefits: respondents do not have to extract from their systems only the information that is relevant for BOP and IIP (ie residents' investment in foreign securities or non-residents' investment in national securities), but provide data on all investors' transactions/holdings in all securities; some data needed for financial accounts (residents' investment in national securities) was not previously available; and an integrated collection system finally produces more consistent statistics.

The data collected through this system are, therefore, very rich in terms of the information they may provide to compilers and users of PI statistics. Not only are traditional variables such as (detailed) type of security, (detailed) institutional sector of the resident investor/issuer, and a full geographical breakdown (on the assets side) available, but other possible details may be provided for analysis, such as the currency of denomination or the institutional/economic sector of the non-resident issuer. Changes in the underlying methodology may also be introduced with minor efforts, since the data collection system was designed in a broader manner comparing to the required output.

SIET's information is used as input for the compilation of a wide set of statistics produced at the Bank of Portugal (in addition to the compilation of portfolio investment figures and related income), which are also disseminated at the national and international level:

- In the field of monetary and financial statistics, SIET enables the validation and detailing of the information reported in MFI and non-monetary financial institutions' balance sheets, as regards their securities issues and own portfolios. Additionally, investment funds statistics are also produced using SIET.
- SIET information is used for the compilation of public finance statistics, in order to obtain the issues of the different general government subsectors and to determine the share of these securities that is taken by general government entities (for data consolidation purposes).
- Values reported to the central balance sheet data office on the activity of non-financial corporations are also compared with SIET data, for the purpose of quality control, of the information collected by both systems, regarding securities issued and purchased by these entities.
- Finally, at the level of national financial accounts, SIET data are used to compile assets and liabilities items of the various institutional sectors of the economy, in debt securities and equities, broken down by type of security and maturity.

SIET promotes consistency across statistics produced by the Bank of Portugal. In fact, securities issues statistics are an output of SIET. In addition to the component of analysis and release at the national level, these statistics are reported to international organisations, namely the ECB and the Bank for International Settlements (BIS). Furthermore, SIET

provides information on securities to several surveys and studies promoted internationally, which are a very important tool for financial and economic analysis. They include the CPIS, which is conducted by the IMF on an annual basis, the ECB annual statistical survey on market structures, the BIS international banking statistics, and the Eurostat government debt questionnaire.

Research on capital market structures and financial stability uses both macro and micro data on securities issuers and holders. SIET has been a source of information to the annual Financial Stability Report of the Bank of Portugal (since 2004), as well as to the Financial Soundness Indicators – an initiative of IMF to identify the strengths and weaknesses of the financial sector, with 62 participating countries (end-2005 reference data), and to the Financial Sector Assessment Programme – conducted by the IMF in Portugal in 2006.

Against a background of turmoil in international financial markets, as the one recently observed, the availability of detailed information on the portfolio assets of financial institutions has proved to be essential in the reassessment of risk exposure.

5. Opportunities and challenges

From our experience, the development of an integrated system for securities statistics has provided major improvements and opportunities in this field:

- Information for statistical purposes is classified by statistical experts and follows a common methodological framework.
- Calculations are performed locally according to internally defined algorithms.
- Valuation adjustments follow uniform criteria.
- Consistency between transactions and positions are monitored in detail.
- Outstanding amounts issued and held are compared at the security level.
- Outputs are compiled according to multiple criteria without having to ask the reporters to do so.

There are still some challenges regarding the compilation of securities statistics. In particular, in the case of the PI liabilities, the direct reporting option is not usually available, and indirect reporting through resident custodians is limited to the extent that non-resident investors use the resident financial system. Additionally, the two options are unable to provide a geographical breakdown of liabilities by creditor country. The Portuguese case follows a mixed approach, ie PI liabilities are calculated based on the net balance of all cross-border custody holdings between issuers, central securities depositories (CSDs), resident custodians and resident end investors. The potential risk of misclassification or double-counting with direct investment is taken care of by relating the PI data with the direct investment surveys.

The geographical allocation of PI liabilities' end-of-period positions and related income by creditor country is the main limitation for the time being. It is an important issue for concern since it also limits the compilation and dissemination of meaningful bilateral IIP statistics. In this context, an exploratory analysis of CPIS data is being undertaken. This intends to derive measures for overcoming some of its limitations, such as the existence of non-published confidential data, the geographical allocation of securities held as foreign reserve assets (collected through another, confidential, survey – SEFER) and the holdings of countries not reporting to the CPIS.

In the case of PI assets, the indirect reporting system via custodians may not be able to capture all the relevant data, even when complemented by direct reporting of securities held

in custody abroad, especially in the case of households. Although this is not considered to be a significant problem for the time being, since it is expectable that Portuguese households use the resident banking system for their investment decisions, it may become increasingly important in the future, namely in a context of a more integrated financial system at the level of the European Union (EU) and, more specifically, of the euro area. In this context, a third-party reporting schema could be further analysed and developed at the EU/euro area level.

6. The near future

The compilation of statistical information on securities entails a number of difficulties, both in terms of classification and valuation and at the level of the holders' identification. The problems are not related to the lack of information sources, given that there are several commercial databases that provide information on individual securities and several ESCB central banks maintain their own databases. However, in some cases, there are gaps and, in other cases, information is not consistent between different sources. These were the main reasons for the development, at the ECB, of a reference securities database with information on a security-by-security basis – the CSDB.

The purpose of the CSDB is to set up a database with complete, consistent, validated and updated information on all securities relevant to the ESCB's statistical objectives. The existence of a single database should promote consistent results and efficient data collection and compilation. This database uses information from commercial databases and other sources, which include the National Central Banks (NCBs) that maintain security-by-security databases. Data quality management will benefit from the cooperation between the different CSDB participants.

The Bank of Portugal has actively collaborated in this project since its inception. Information on Portuguese securities, extracted from SIET, is being sent to the ECB on a monthly basis. Also, monthly extracts of the CSDB are being used for data quality checking.

From the statistical viewpoint, the CSDB serves two purposes: to supply information for the compilation of aggregates for the euro area, such as securities issues statistics, and to supply reference information on securities and issuers, so as to cope with the collection of statistical information on a security-by-security basis, and enabling the production of improved aggregate statistics.

The development of the CSDB is being carried out gradually: in phase 1, completed in May 2005, the system was implemented at the ECB; in phase 2, currently ongoing, mechanisms will be implemented for online access and application-to-application communication for NCBs.

Currently, several countries are already collecting information on portfolio investment on a security-by-security basis for the production of balance of payments and international investment position statistics. In the near future, all euro area members will follow this approach, using the CSDB for the classification of information on securities. In this sense, the CSDB will be a major contribution to a more efficient production of harmonised statistics.

We conclude by re-emphasising that recent changes in financial markets, due to globalisation and innovation, brought new challenges and demands to statistics. Simultaneously, technological evolution continues to provide opportunities to develop increasingly integrated systems, based on item-by-item data. These developments, together with an increasing collaboration between institutions at national and international levels, are definitely contributing to the production of more accurate, reliable, and comparable statistics.