A security-by-security approach to deriving investment fund security transactions data from stock data

Frank Mayerlen[1]

Introduction

The concept of security-by-security (s-b-s) reporting represents perhaps the most promising approach to reducing the statistical burden for reporters while simultaneously improving the quality and detail of statistical data. For the compiler of statistics, the processing of s-b-s data requires a substantial change in the production environment, entailing a move away from the simple "adding up" of aggregated information, towards the processing of micro data in relational databases. A further, and perhaps even more important change, from the compilation perspective, is the requirement for securities reference information, in addition to the information provided by reporting agents.[2] The need for appropriate reference data is often seen as an element of s-b-s approaches that could entail risks, particularly in cases where reporting agents are likely to hold "exotic securities" that may not be (fully) covered by the reference information.

This paper proposes a statistical solution for the compilation of investment fund balance sheet and transaction data using an s-b-s approach. In particular, it addresses the problem of incomplete or missing reference information, proposing built-in fallback solutions that would be automatically phased out as reference data improve, to allow for the robust production of reliable statistics at a stage when the reference information may (still) have some gaps.

The paper describes a test application for the compilation of investment fund statistics following an s-b-s approach. Owing to space limitations, some details are not covered, in particular as regards conceptual design and implementation. The most relevant of these issues have been mentioned in the footnotes.

Practical problems with security-by-security approaches

From a conceptual point of view, all data reported using an s-b-s approach will be matched with (enriched by) reference information from a securities reference database, to create a comprehensive basis for further processing or consistent aggregation required in a variety of contexts. In this connection, gaps or errors in reference data are a major practical concern for compilers of statistics, since these data are necessary for classifying statistical information by instrument, currency and issuer's residency and sector. Another practical problem arises when the reporting agent cannot provide a valid identifier for some securities and instead must report aggregate information. This means that the s-b-s compilation

[1] Principal Economist-Statistician, European Central Bank, Directorate General Statistics, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany, E-mail: frank.mayerlen@ecb.int

[2] The European System of Central Banks (ESCB) will be sharing the Centralised Securities Database (CSDB) for securities reference information purposes. This will ensure consistency in the production of Eurosystem statistics.

environment must also be able to cope with aggregated data and make the related revaluation adjustments.

A forward-looking approach to establishing an efficient, robust and sound production system

How "precise" does an s-b-s system for producing macroeconomic statistics need to be?

In developing an s-b-s production system for macroeconomic statistics, user needs – including data accuracy requirements – should be a major driving force. One should usually refrain from requiring overall "book-keeping accuracy", even if theoretically feasible and appealing to some users or producers of statistics. It may be impossible to achieve such a high level of accuracy, given that the reported information and the reference data come from two independent data sources. Moreover, the cost involved in attaining (close to) 100% accuracy and coverage for the reference data would be disproportionate, and rarely justifiable in terms of the increased usability of the statistical data. With a more efficient (but slightly less precise) statistical approach, the data compiled on an s-b-s basis will still be considerably more precise and detailed than almost any statistics that are reported solely as aggregates.

Defining the data to be required from reporting agents

S-b-s reporting allows the use of statistical approaches to data requirements that are much more in line with reporting agents' internal business systems than are the requirements for aggregated reporting. Consequently, the reporting burden can be minimised if this information by itself is sufficient for statistical purposes. On the other hand, s-b-s approaches often provide some additional data at almost no cost when they are available in the business systems.

Therefore, in determining the data requirements for the forthcoming new statistics on euro area investment fund assets and liabilities, it has been agreed that, subject to the views of the compiling institution, reporting agents shall not be required to report transaction data for securities, in the event that they implement an s-b-s reporting framework for reporting stocks. The rationale for this is that often transaction data are not readily available in business systems, and are more efficiently derived by the producers of statistics. At the same time, the reporters agreed in this case to provide price data for each securities position. This is readily available to the funds and greatly facilitates the production of the statistical data. The agreed data requirement is expected to facilitate the production of high quality statistics and is considered to offer a good cost benefit compromise for both reporting agents and compilers of statistics.[3]

Designing the s-b-s production environment

All s-b-s production environments need to make use of both the reported information and the securities reference information. Given that one reference database may be used by several production environments, it is clearly more efficient to invest in a sound procedure for setting

[3] The data requirement may also cover the reporting of deposit (and loan) positions on a "currency-by-currency" basis, but without any revaluation adjustments. Based on these data, accurate transactions can be derived separately for each currency.

default values in the reference database than fixing problems in each individual production environment. However, for guaranteed robust functioning, each production environment should incorporate fallback solutions to remedy any gaps or errors in the reference data that have not been addressed previously. The production environment must, in any event, include a default procedure to provide for those cases in which securities identifiers cannot be matched with reference data, owing to reporting errors or to gaps or errors in the reference data. For the sake of overall consistency among different statistics, it is also essential to ensure that the production system's estimation methods will "switch off" in cases where the reference data are complete.

Technically, the production environment for investment fund statistics must be able not only to process s-b-s information, but also to deal with data referring to other balance sheet categories reported on an aggregated basis, and with aggregated (fallback) reporting of securities positions. Note that because of limited space, the processing of such aggregated data is not addressed in detail here.

The entire compilation procedure may be organised in a layered approach, in which stocks are compiled first, and then the transaction data based on this information are derived.

Calculating stock data

Where data on prices and outstanding amounts are concerned, the production of statistics on stocks of securities by the described s-b-s approach relies entirely on the s-b-s data reported by the investment funds. These data are reported without any further information by instrument category, currency, or issuer's sector or residency, which reporting agents generally consider difficult to provide. These data are instead sourced from the reference database, independent of the actual reporting. Other (non-securities) balance sheet components, eg deposits (and loans), must be reported with the required breakdowns.

Remedying gaps in the securities reference data

Given that the funds report prices as well as outstanding amounts,[4] gaps in reference data affect only the statistical breakdowns (categorisation) of the balance sheet components, not the total value of a securities position. Securities reference data cover (1) instrument category; (2) currency in which issue is denominated; (3) issuer's sector; and (4) issuer's residency. When reference data are "incomplete", either the entire set of relevant securities reference information is missing (ie the security is not covered in the reference database or cannot be matched) or data on one or more individual variables are missing. The proposed estimation method explained below is designed to fill these gaps by deriving *missing* information from the most similar *complete* records, thereby exploiting the information available to the extent possible.

Given the four reference variables for stock statistics, there are fifteen possible combinations of gaps in the reference data, including the case where there are no reference data at all.

[4] In practice, the funds have to report two of the following three variables for each security: (1) price; (2) number of securities or outstanding amount; and (3) total value of the position.

Table 1

Different error types in the reference information

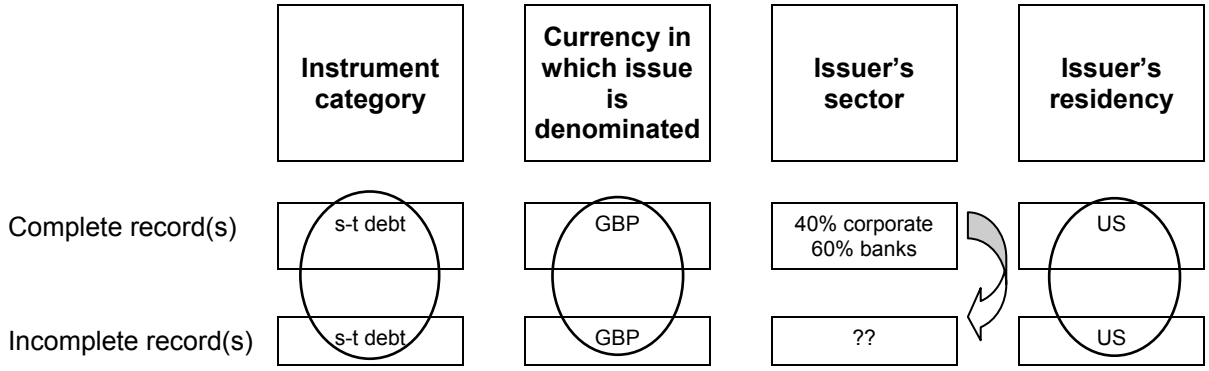
[X marks missing data]

Error type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Instrument category	X				X	X	X	X	X	X	X				
Issuing currency	X	X			X	X					X	X		X	X
Issuer sector	X	X	X		X			X		X			X	X	
Issuer residency	X	X	X	X					X	X	X	X			

After classifying the gaps by "error type", the estimation procedure searches for complete securities records with similar reference information for all fields that are filled in, including the complete and incomplete record. Data are then estimated only for the field(s) with gaps, on the basis of the complete record(s). The advantage of this method is that the estimation, for example, of the issuer's sector for a short-term (s-t) debt security issued by a US resident does not affect data on securities issued by euro area residents. The estimation is carried out separately for assets and liabilities.

Figure 1

Matching of complete and incomplete records for the estimation of reference data



Where there is no matching complete security record in which all complete fields coincide, the incomplete record is downgraded to a record in which all fields are missing (error type 1 in Table 1), and is distributed in a second round of processing, following the overall structure of all complete securities records (again, separately for the asset and liability sides).[5]

The estimation procedure guarantees that the statistical categorisation is applied to all balance sheet categories where this is required. The entire balance sheet can now be aggregated for reporting.

[5] A solution with a stepwise downgrade is also conceivable. The downgrading procedure would in theory also (as a fallback) allow securities positions to be reported in the aggregate without any detailed instrument classification, provided that the amounts are small in relative terms. The statistical classification would then be estimated by the procedure described.

Derivation of transactions and investment income

Conceptual approach

The stock data as presented above form the basis for the derivation of transaction and investment income data for securities following an s-b-s approach (ie separately for each security). Transactions that occurred during a period between t_0 and t_1 will be derived according to the formulae shown below and using the following variables: *EQ* (number of shares); p_{EQ} (price per share in euros); *DT* (debt securities, nominal amount); p_{DT} (debt securities price in euros); *DIV* (dividend per share in euros); *DAYS* (number of days between t_0 and t_1); *ACF* (daily accrual factor for debt).

Securities transactions (equity):[6]
$$TR_{EQt0t1} = \left(EQ_{t1} - EQ_{t0}\right) * \frac{\left(p_{EQt1} + p_{EQt2}\right)}{2} \quad (1)$$

Securities transactions (debt):
$$TR_{DTt0t1} = \left(DT_{t1} - DT_{t0}\right) * \frac{\left(p_{DTt1} + p_{DTt2}\right)}{2} \quad (2)$$

Dividend income (equity):[7]
$$IN_{EQt0t1} = \frac{\left(EQ_{t1} + EQ_{t2}\right)}{2} * DIV \quad (3)$$

Interest income (debt):
$$IN_{DTt0t1} = \frac{\left(DT_{t1} + DT_{t2}\right)}{2} * DAYS * ACF \quad (4)$$

Practical problems

Formulae (1) and (2) above require price information for t_0 as well as t_1. Since price data are reported together with the corresponding stocks, they will not be available from reporting agents where positions are zero at either t_0 or t_1. In such cases, the price information is sourced from the reference database, which also provides the necessary income information (dividends and accrual factors).

In practice, the reference database may lack the income information, and possibly also the price data. In this case, using only the price at t_1 or t_0 to estimate the transactions may not be sufficiently accurate, especially during periods with strong moves in share prices. Moreover, an approach is needed to derive transactions for those positions where the reference information for the stock data has been estimated, ie where no price information can possibly be available on an s-b-s basis, as there is no link to an individual security. The proposal in both cases is to "downgrade" these records to aggregate information (AGG) and to apply price indices derived from those s-b-s records where full price data are available.[8]

Deriving transactions using dedicated price indices

The basic idea behind this approach is analogous to the solution for stock data described above, in which missing information is also derived from the most similar *complete* records.

---

[6] If applicable, the number of shares is automatically adjusted for stock splits that occurred between t0 and t1. The relevant corporate action information should be available from the reference database.

[7] If applicable, dividends are automatically adjusted for stock splits that occurred between t0 and t1. The relevant corporate action information should be available from the reference database.

[8] A similar approach may be considered at some stage for income data, where average accrual factors or dividends may be applied to those instruments or positions for which no income information is available.

For transactions, volume-weighted Laspeyres[9] price indices are derived from the records with complete reference data for each equity or debt category (X) of the securities item. These price indices (PR_{Xt0t1}) are then applied so as to price-adjust the aggregate balance sheet positions at t_0 when complete reference information is unavailable (AGG). As a last step, transactions are calculated as the difference in (adjusted) aggregate positions.

Adjustment of positions at t_0: $$AGG_{Xt0\_adj} = AGG_{Xt0} * PR_{Xt0t1} \quad (5)$$

Calculation of transactions: $$TR_{t0t1} = AGG_{Xt1} - AGG_{Xt0\_adj} \quad (6)$$

Aggregation and reporting

Both the directly derived transactions and income data, and the transactions estimated using the indices, are stored in a single table, together with the statistical reference information by sector, residency, currency and instrument category. Based on this information, different (but fully consistent) aggregates can be constructed for the investment fund statistics and, in principle, for other purposes as well, such as external statistics.

[9] From a conceptual point of view, a Fisher price index may be more consistent with the method for deriving transactions according to formula (1).