



Michael Anyadike-Danes

Contact: m.anyadike-danes@aston.ac.uk

Aston University matching of BSD and FAME data

Background

The data reported in these tables have been compiled by merging firm-level records from two different databases: the Business Structures Database (BSD) and FAME. The strength of the BSD is its near universal coverage of all firms in the UK, however it only provides information on a limited range of firm characteristics (employment, turnover, location, structure and industry sector). The strengths and weaknesses of FAME are complementary to those of the BSD, its coverage is considerably more limited, especially of smaller firms, but it reports a very wide range of characteristics, mainly those which appear in a firm's annual accounts. What we have done here is to build on the strengths of both these by using firm-level identifiers to match records from each dataset and then merge them.

The following sections describe the two source datasets in more detail and provide some information on the merged BSD-FAME dataset.

1. BSD database

The BSD is a series of annual 'snapshots' of the Inter-Departmental Business Register, an administrative database which captures information from VAT returns and employer Pay As You Earn (PAYE) tax and social security records as well as ONS business surveys. The unit of analysis is an "employer enterprise" -- a business with at least one employee (since an employee can work for more than one firm summing over firms produces an estimate of jobs rather than employment) -- which we refer to as an enterprise. Using the unique firm identifier, firm-level records from the successive annual snapshots have been linked together. The status (as a live employer enterprise) is inferred from data on PAYE tax and social security contributions from their employee(s). The 'birth' of a firm is dated by the first appearance of non-zero employment and its 'death' is treated symmetrically and dated by the disappearance of the last employee. The data do not distinguish between *de novo* births and those which results from the break-up of an existing firm, similarly the data do not distinguish between the closure of a firm and its disappearance due to Enterprise data

includes birth; death; employee numbers for each year; industrial classification; and for in most cases, annual turnover.

2. FAME database

FAME contains information on companies registered at Companies House in the UK. It covers company financials, in detailed format, with up to 10 years of history, detailed corporate structures and the corporate family, shareholders and subsidiaries. It includes 7 million companies in the UK, 200,000 companies in a summary format. Details of 700,000 companies which are active but not required to file accounts, or have yet to file accounts, and 4 million companies which are no longer active. The data on these firms are collected from various sources most notably the national official bodies in charge of collecting company accounts data. The data are then compiled and organised by Bureau van Dijk (BvD) in a consistent format following strict guidelines.

A standard company report includes a balance sheet, profit and loss account, turnover, employees and industry codes. Moreover, the data includes detailed ownership and subsidiary information which is based on millions of links between firms and their shareholders and subsidiaries worldwide. A link establishes an ownership relationship between a firm, its shareholder and its subsidiaries and data lists the direct and indirect shareholders and subsidiaries of a given company with their percentage ownership. From this information, it is possible to distinguish between foreign-owned and domestically-owned firms where the nationality of a firm is determined by the ultimate parent's country of ownership. The data also lists the subsidiaries of a given company together with their percentage of ownership and it can be determined whether a firm is a multinational (if it has one or more foreign subsidiaries) or a purely domestic firm. FAME data relies on company reports to Companies House. This means that the data is often up to two years old. It also means that because only large companies are obliged to report employment, turnover and assets that medium and smaller companies don't share this data. FAME data is unverified and also includes worldwide figures for companies.

3. BSD-FAME matching

Data from FAME have been matched to firm-level records in the BSD using unique firm-level identifiers. The details of the match are set out in a summary table (see Table A3.1 in 'MSB Aston growth & business characteristics'). Out of the 9,377 MSBs on the BSD in 2010, 7,185 are on FAME, a match rate of about three quarters. However this match rate is not a constant, indeed FAME coverage has been systematically improving over time: coverage in the early years (2001 to 2004) is closer to one half than to three quarters. Of course this means it is difficult to draw definitive conclusions about the time series properties

of data from BSD-FAME unless there is reason believe that the characteristic of interest is not likely to have been affected by this 'selection effect'.

It is worth noting, though, that the match rate for MSBs is very much better than the match rate for non-MSBs. Again, looking at 2010 we find of 1.5 million non-MSBs on the BSD less than 200 thousand can found on FAME (a match rate of 12%, i.e. about one sixth of the MSB match rate). The picture becomes a little clearer if we look at the rather more fine-grained data (see Table A3.2 in 'MSB Aston growth & business characteristics'). It emerges, as might have been anticipated from what we already know about the FAME dataset, that the match rate is very much better at the 'large' end of the firm size distribution (for larger than MSB firms) than it is at the 'small' end (for smaller than MSB firms). In 2010 the larger than MSB match rate was above 80% (561 FAME out of BSD 666), whilst the smaller than MSB match rate was just over 10% (187 thousand FAME out of 1.5 million BSD). This positive association between size and the match rate holds across the MSB category too. As we can see from the middle of the table, the match rate for the larger MSBs (£250m to £500m) is about 80%, whilst for the smallest MSBs (£25m to £50m) it is just 75%, it varies across the MSB size range but by just five percentage points. The clear implication is that conclusions about differences between MSBs and larger firms, and comparing MSBs of different sizes, might be drawn quite reliably, but selection bias may make conclusions about the differences between MSBs and smaller firms rather less robust.