JUAN MANUEL GARCÍA LARA, BEATRIZ GARCÍA OSMA AND
BELÉN GILL DE ALBORNOZ NOGUER

Effects of Database Choice on International Accounting Research

Data availability is one of the traditional obstacles confronting researchers carrying out international empirical studies in accounting. In recent years several databases have claimed to offer comprehensive coverage of accounting and financial data of firms worldwide. We analyse whether the choice of database has an effect on the results of empirical studies. We find that the results of a simple empirical adaptation of the Ohlson (1995) model for fourteen member states of the European Union change significantly depending on the database chosen (Datastream, Global Vantage, Company Analysis, Worldscope, Thomson Financial, Financials and BvD Osiris). These differences are mainly attributable to differences in the samples across databases. When we match observations across all databases the differences persist but are much less pronounced. Our main conclusion is that database choice matters, as it leads to different results when the same research design is used.

Key words: Database comparability; International empirical research.

International accounting research has increased substantially in the last decade. This growth in research interest has coincided with the appearance of several databases that, since the early 1990s, offer comprehensive coverage of accounting and financial market data for companies worldwide. Normally, it is the previous literature, and particularly the benchmark papers, that set the standard as to which database to use in empirical research, thus facilitating comparisons across studies. Such is the case of U.S.- or U.K.-based studies, where researchers have almost exclusively retrieved data from Compustat/CRSP and Datastream, respectively.

Juan Manuel García Lara (jmglara@emp.uc3m.es) is a Lecturer in Accounting and Finance at Universidad Carlos III de Madrid, Beatriz García Osma (beatriz.garcia@uam.es) is a Lecturer in Accounting and Finance at Universidad Autónoma de Madrid and a PhD student at Lancaster University, and Belén Gill de Albornoz Noguer (noguer@cofin.uji.es) is a Senior Lecturer in Accounting and Finance at Universitat Jaume I de Castellón.
The authors appreciate the comments and insights of the discussant, Stewart Jones, as well as the helpful suggestions of Paulo Alves, Miguel Arce, Willem Buijink, Michael Bradbury, Stuart McLeay, Ken Peasnell, Maarten Pronk, José Luis Ucieda, and participants at the Abacus Forum. The paper has benefited from presentations at the IV Workshop on Empirical Accounting Research (Sevilla), the 2005 annual congress of the European Accounting Association and the Accounting Research Workshop at the University of Cádiz. We acknowledge financial contributions from the Spanish Ministry of Education (SEC2002 04608 C02 01/2 and SEJ2005 08644 C02 01), from AECA through the Carlos Cubillo Chair in Accounting and Auditing, and from the Universidad Autonoma de Madrid (Papiro). We also thank Shaling Li for her excellent research assistance.

However, database choice is not a trivial issue in international accounting research, as the lack of tradition and the existence of a wide choice of databases, have prevented the literature from leaning clearly toward any of the currently available sources of data. Differences across databases, if they exist, may lead researchers to conclusions that are difficult to extrapolate, potentially having important implications for the future development of accounting and corporate finance research.

Currently, the best known providers of accounting and other financial data for firms worldwide are Datastream, Global Vantage (or Compustat Global), Company Analysis, Worldscope, Thomson Financial, Extel Financials, BvD Osiris, BvD BankScope and BvD Amadeus. These databases specialize in financial statement data, although they sometimes provide additional data on other closely related items, such as market prices, analysts' forecasts, or corporate governance indicators. However, this is not always the case and costly annual subscriptions to general accounting databases often have to be complemented by purchasing other expensive data sources, such as CRSP and Compustat for U.S. data, I/B/E/S or JCF[1] for analysts' forecasts, local versions of Amadeus like FAME (U.K. and Ireland) or SABI (Spain and Portugal) for small national firms, local databases such as AspectHuntley for Australian data,[2] or databases that provide corporate governance related information such as Manifest for U.K. companies. Facing this broad choice of databases, some institutions opt to subscribe to database platforms like Thomson One Banker or Wharton Research Data Services (WRDS) to access several databases through the same interface.[3] Hence, database choice has not only empirical but also economic consequences.

Prior research has shown a limited amount of concern on how differences across databases could affect the results of international empirical accounting studies. For instance, Ball *et al.* (2000), using Global Vantage, imply that the differences between their results and those of Pope and Walker (1999), using Datastream, could be due to using different databases as sources for U.K. data. There is also some limited research on the existence of differences in certain items across databases for a single country, but no previous study has directly addressed the issue of how database choice may affect the results of international accounting research. This article provides the first evidence on the effects of database choice on the results obtained by such studies.

[1] The JCF Corp. offers several products through which different types of analysts' forecasts data can be retrieved. It is a growing competitor for I/B/E/S.

[2] AspectHuntley was created in 2003 as a result of the merger between the two leading providers of Australian financial data: Aspect Financial Pty Ltd and Huntley's Financial Services Pty Ltd. However, the use of this database is increasingly being supplanted by the use of database platforms, such as Thomson One Banker. We thank the discussant for providing this information.

[3] It should be taken into account that the use of several databases for research has high costs for the researcher, mainly related to (a) learning how to use the various databases, as they tend to differ quite considerably, and (b) matching data from different databases into a single dataset, which can be an onerous task.

Carrying out a simple market-based study in accounting using an international sample covering fourteen member states of the European Union, we show that database choice significantly affects the results of the model. Samples change substantially depending on the database, leading to significantly different results using identical research design. Specifically, we download all available data for fourteen EU member states for the period 1990–99, from seven different databases.[4] These are Datastream, Global Vantage, Company Analysis, Worldscope, Thomson Financial, Extel Financials and BvD Osiris. Using these data, we run a simple adaptation of the Ohlson (1995) model that regresses book value of shareholders' equity and earnings on the market value of the company, and find significant differences in the regression coefficients and measures of fit depending on the database chosen. When we match observations across databases, the differences almost disappear. This points to a size-related effect (the databases that offer fewer firm-year observations cover just the largest firms, while coverage increases as the average size of firms decreases), which would explain the differences.

In summary, our results contribute to international empirical accounting research by providing evidence on the effects of database choice. The findings are relevant to accounting researchers, as they raise concerns about the international comparability of the results of much accounting research, and their generalizability. More pragmatically, the results may also be of interest to academic institutions when deciding on the acquisition of international accounting data.

ACCOUNTING AND FINANCE DATABASES

Normally, research institutions invest in databases only when there is sufficient assurance of usage. Meek and Thomas (2004) point out that many such institutions in the U.S. are willing to incur the costs of databases for U.S. data (Compustat and CRSP) because many staff members and students use them. Oftentimes, researchers and PhD students are conditioned by the databases chosen by their institutions, driving even to a certain extent the type of research undertaken. When the host institution subscribes to several databases, the choice of database is commonly made on the grounds of (a) data availability, and (b) normal research practice.

Data availability refers to the number of observations covered by a database, both in terms of time-series and cross-section coverage (i.e., number of years and number of firms), although it may also refer to number of items available in the database. Data availability is an issue commonly cited in international accounting research. For example, Basu's (1999, pp. 90–1) discussion of Pope and Walker (1999) states that 'Pope and Walker collect data from 1976–96 for U.S. firms from Compustat PC Plus and for the U.K. firms from Datastream. However, they have fewer firms in each annual cross-section for both countries than researchers using

[4] From the previously mentioned databases, we exclude (a) BankScope, because it is a database specialized in financial institutions data, and (b) Amadeus, because, although it includes both pub lic and private companies, it concentrates mainly on private firms. Our empirical analysis focuses on publicly listed non financial firms and Osiris is the product offered by Bureau Van Dijk for public companies.

other databases such as Global Vantage.' Also Land and Lang (2002, footnote 4) argue that they use Global Vantage instead of Datastream or Worldscope because 'data availability, especially for early years, is best with Global Vantage'.

By normal research practice we refer to certain research contexts where one database has become mainstream and is widely used and accepted as the source of accounting data. For example, this is the case in the U.K., where most researchers have traditionally used Datastream, setting it as a standard and promoting its use even when other data sources are available, to facilitate the replication and comparison of results. However, there are currently no well-established databases for international accounting research, and particularly, for European-based accounting research. A review of the main international accounting research articles published in both U.S. and international journals shows that, quite often, researchers choose to combine several databases, and that although they appear to favour Global Vantage (out of the 32 papers identified, 16 use this database), several studies choose Worldscope (8), Datastream (5) or Extel Financials / Company Analysis (4), among other less well-know options. Table 1 offers a summary of the databases used to retrieve the accounting data underlying some key articles published in the main journals in this area. It can be observed readily that, in agreement with our previously explained 'normal research practice' argument, U.K.-based accounting research prefers Datastream to other data sources. Also, U.S.-based researchers appear to prefer using either Global Vantage or Worldscope, suggesting that, at least amongst U.S. academics, a certain tradition is already developing for international accounting research.

Below is a brief review of the databases currently available for research purposes depending on the existence or not of an established research practice and database usage standardization, paying particular attention to the case of international accounting research.

The Case of U.K. and U.S.: The Existence of Well-Established Research Databases

The main source of U.S. (and Canadian) accounting data is Compustat, a database maintained by Standard & Poor's Investment Services, Inc., a division of McGraw-Hill, Inc. Compustat has become the standardized database for accounting research, to the point that it is difficult to find a published empirical article that does not use it in the analysis of U.S. firms. The same could be said of CRSP, the database on share prices and other market information maintained by the Center for Research in Security Prices at the University of Chicago, and I/B/E/S or First Call, the most popular databases that offer analysts' forecasts and related data.[5]

[5] I/B/E/S International Inc. created their Academic Research Program over thirty years ago to provide both summary and individual analyst forecasts of company earnings, cash flows and other important financial items, as well as buy sell hold recommendations. I/B/E/S covers both U.S. and international quoted companies. In 2000, when its previous owner, the Primark Corporation, was acquired by Thomson Corporation, I/B/E/S was integrated with Thomson Financial / First Call. The First Call Historical Database, or FCHD, has been built from First Call's Real Time Earnings Estimates (RTEE) service and contains earnings information dating as far back as 1990, on 8,500 U.S. securities, 1,000 Canadian securities and 180 ADRs.

TABLE 1

DATABASE CHOICE FOR KEY RESEARCH PAPERS ON INTERNATIONAL ACCOUNTING

Paper	Journal	Countries under study	Database chosen
Aboody *et al.* (1999)	*JAE*	U.K.	Datastream International
Ali and Hwang (2001)	*JAR*	16 non U.S.	Global Vantage
Ashbaugh and Pincus (2001)	*JAR*	13	Worldscope
Baginski *et al.* (2002)	*TAR*	U.S., Canada	Compustat
Ball *et al.* (2000)	*JAE*	Australia, Canada, U.K., France, Germany, Japan, U.S.	Global Vantage Industrial/Commercial File
Ball *et al.* (2003)	*JAE*	Hong Kong, Malaysia, Singapore, Thailand	Global Vantage Industrial/Commercial File
Barth and Clinch (1996)	*CAR*	U.S., U.K., Australia, Canada	Compustat's Global Vantage
Basu *et al.* (1998)	*JBFA*	10	Global Vantage
Bhagat and Welch (1995)	*JAE*	U.S., Canada, U.K., Japan, Germany, France, Netherlands	Compustat's Global Vantage
Bhattacharya *et al.* (2002)	*TAR*	34	Worldscope
Black *et al.* (1998)	*JBFA*	U.K., Australia, New Zealand	Global Vantage
Bushman and Piotroski(2006)	*JAE*	38	Global Vantage
DeFond and Hung (2004)	*JAR*	33	Worldscope
Fan and Wong (2002)	*JAE*	Hong Kong, Korea, Thailand, Taipei, Singapore, Malaysia, Indonesia	PACAP database
García Lara *et al.* (2005)	*JBFA*	U.K., Germany, France	Datastream
Giner and Rees (2001)	*JBFA*	U.K., Germany, France	Extel Financial Company Analysis
Gordon and Joos (2004)	*TAR*	U.K.	Datastream International
Guenther and Young (2000)	*JAE*	France, Germany, U.K., U.S., Japan	Global Vantage
Haw *et al.* (2004)	*JAR*	22	Worldscope
Hope (2003)	*JAR*	22	Datastream, Global Vantage, Compustat, Global Access
Hung (2000)	*JAE*	21	Global Vantage Industrial/ Commercial Files
Joos and Lang (1994)	*JAR*	Germany, France, U.K.	Global Vantage Industrial/ Commercial File
Kallapur and Kwan (2004)	*TAR*	U.K.	Extel Company Analysis, Global Vantage
Khunara and Raman (2004)	*TAR*	U.S., Australia, Canada, U.K.	Global Vantage Industrial/ Commercial File
Land and Lang (2002)	*TAR*	Australia, Canada, Germany, France, U.K., Japan, U.S.	Global Vantage
Lang *et al.* (2004)	*JAR*	27	Worldscope
LaPorta *et al.* (1997)	*JF*	49	Worldscope, Extel
LaPorta *et al.* (1999)	*JF*	27	Worldscope
Leuz *et al.* (2003)	*JFE*	31	Worldscope
Monsen and Wallace (1995)	*CAR*	Denmark, Sweden, Finland, Norway, Iceland	Oslo Stock Exchange (manual)
Pope and Walker (1999)	*JAR*	U.K., U.S.	Compustat PC Plus, Datastream
Raonic *et al.* (2004)	*JBFA*	13 (all European)	Datastream
Seetharaman *et al.* (2002)	*JAE*	U.K., U.S.	Financial Times Extel Company Analysis, and Disclosure's Global Access
Young and Guenther (2003)	*JAR*	23	Compustat's Global Vantage

Note: The following abbreviations have been used: *CAR*, *Contemporary Accounting Research*; *JAE*, *Journal of Accounting and Economics*; *JAR*, *Journal of Accounting Research*; *JBFA*, *Journal of Business Finance and Accounting*; *JF*, *Journal of Finance*; *JFE*, *Journal of Financial Economics*; *TAR*, *The Accounting Review*.

In the U.K., Datastream has been the main source for accounting and financial information for many years. The number of items available, that is, the level of disaggregation of accounting data for U.K. firms in Datastream is larger than in other databases, and besides, it offers not only accounting data for the U.K., but also for firms from both developed and emerging economies, as well as information on daily share prices,[6] market indexes, macroeconomic indicators, bonds, foreign exchange rates, interest rates, commodities and derivatives. However, as a result of the acquisition of the Primark Corporation by Thomson Financial in 2000, Datastream stopped updating its accounting data in 2004 and subsequently substituted it with data from Worldscope. Therefore, our results are of special interest to U.K.-focused researchers, as they will not be able to use Datastream accounting data from now on. Clearly, this raises the question of which database will become mainstream for accounting research in the U.K., and whether comparisons with prior research will be difficult to undertake, particularly if significant differences exist between Datastream and other databases, especially Worldscope, given that it is the direct substitute for Datastream accounting data in many U.K. academic institutions.

In countries such as the U.K. and the U.S., researchers enjoy wider data availability and, although research practice has imposed one database as the standard, errors and differences across databases still exist and their potential consequences on empirical research have been pointed out in the literature. Early studies in this area were mainly concerned with the existence of errors in the databases (e.g., Rosenberg and Houglet, 1974; San Miguel, 1977; Bennin, 1980). More recently, Guenther and Rosman (1994) show that differences in industry codification between Compustat and CRSP can change the results and conclusions of certain studies, and they demonstrate that the results in Freeman and Tse (1992) change when they use CRSP's instead of Compustat's industry classifications. Closely linked to this result, a recent paper by Krishnan and Press (2003) shows that using NAISC (North America Industry Classification System) instead of SIC (Standard Industrial Classification) for categorization by industry can also significantly affect the results of studies that use industry classification as the basis of their analysis. Looking at specific items, Kern and Morris (1994) analyse the potential impact on empirical research of the differences between Compustat and Value Line databases, and show that for the same firms, both the figure of total assets and that of total sales can differ across databases. Similarly, Courtenay and Keller (1994) argue that errors in CRSP can affect studies that use trading volume data, and Elton *et al.*'s (2001) comparison of the CRSP mutual fund database with Morningstar shows large differences in both the monthly fund returns and the measures of risk-adjusted returns contained in both databases. Finally, Abarbanell and Lehavy (2000) show that the choice between I/B/E/S, First Call, Zacks and Compustat will likely have effects on the results of studies in the areas of earnings

[6] Market data coverage in the other databases considered is more limited, as they mainly offer fiscal year end prices.

management, earnings response coefficients and the value relevance of accounting information.

We expect that these errors and differences across databases found exclusively in the U.S. will also exist in the databases analysed in this paper, and that they may be even more marked when not just one but several countries are taken into account.

The Case of International Accounting Research: The Database Choice Dilemma

The standardization of database usage for market-based research in the U.S. and the U.K. is hardly the case for studies using European or international data. For international researchers interested in retrieving accounting and market data for listed firms worldwide there are several available databases, among which the best known are: Datastream, Global Vantage, Worldscope, Company Analysis, Thomson Financial, Extel Financials and BvD Osiris. The choice between these databases has usually been made on the grounds of: (a) database availability at the host institution, (b) availability of the required items for the particular analysis, (c) normal research practice and (d) time-series data availability and firm coverage. We have previously referred to most of these elements. Data constraints should not be underestimated, but it is likely to be the budgetary constraint that dominates the academic institution's decision to subscribe to a limited number of databases, particularly to databases that provide international data. Often, a single database is acquired, thus driving researchers into its use. If more than one database with similar data is available, then it is normal research practice that likely directs the choice, as we have argued previously. Below, we briefly describe the main data sources available for international accounting research.

Global Vantage is the name given to the international files of Compustat, which is commercialized by Standard & Poor's. Information from Global Vantage can be retrieved using Research Insight[7] or through database platforms such as WRDS. Similar to Compustat, an important feature of Global Vantage is that data are collected according to standardized definitions, researched and written by Standard & Poor's Compustat Services, Inc., which are in line with the regulations and standards of the Financial Accounting Standards Board, the Securities and Exchange Commission and U.S. Generally Accepted Accounting Principles and Procedures.[8] The database consists of four files: (a) Industrial/Commercial File, containing over 200 Income Statement, Balance Sheet, Flow of Funds and supplemental data items for firms in industry and commerce; (b) Financial

[7] Research Insight is a program designed by Standard and Poor's to retrieve data from Compustat and Global Vantage.

[8] Therefore, data are not presented as reported when known differences in definitions exist. They are made compatible on an international basis by adjusting for differences in accounting principles among different countries. To disentangle such adjusting processes is not an easy task for the researcher and it is beyond the scope of this paper.

Services File, containing approximately 500 Income Statement, Balance Sheet, Flow of Funds and supplemental data items for bank, insurance, broker/dealer, real estate, and other financial services industries[9]; (c) Issue File, for market-related data items, including monthly prices, dividends, shares traded, issued capital, and earnings per share; and (d) Currency File, containing month-end and average translation rate items and cross-rate tables for designated currencies. The Industrial/Commercial and Financial Files contain data relating to more than 13,000 international publicly traded companies in over eighty countries. Coverage extends to twelve years of annual data. Market information covers more than ninety local market indexes and over 110 currencies. For key items, updates of the database can be obtained daily over the internet or monthly via http delivery.

Worldscope, Thomson Financial, Extel Financials and Company Analysis are all products now controlled, like Datastream, by the financial division of the Thomson Corporation.

Barron and Smithers design and produce the Thomson Extel Financials core database, which is available from Thomson via three alternative platforms: Company Analysis, Thomson One Banker or FTP feeds. Extel Financials is a database containing basic accounting data for 26,500 companies from over fifty-five different countries starting from 1985. This includes what they claim to be full quoted coverage for the U.K., France, Netherlands and Switzerland in Europe, and Singapore, Hong Kong, Malaysia, Thailand and Australia in Asia and the Pacific. In contrast with other databases, like Datastream, Worldscope or Global Vantage, the key feature that distinguishes Extel Financials is that the data are in a 'structured as reported' format, meaning that the data items are (a) the same as those reported in the annual report and accounts, and (b) structured in a way that strikes a balance between providing enough data to satisfy sophisticated customers and ensuring that it is still easy to use. There are potentially 1,450 data fields that can be retrieved.[10] The data are organized as a series of pyramids that show the relationships between the items.[11] The data are collected from native language reports,

[9] The number of items available for every given company in both the Industry/Commercial and Financial Files is smaller than that in Compustat for U.S. and Canadian firms, as the accounting items are provided on a more aggregated basis. Data are collected using consistent sets of items developed by examining financial statements from a variety of countries and identifying items that are widely reported by companies regardless of their geographic location, business activity and accounting practices.

[10] In reality, only a proportion of the 1,450 items will contain data for any one company. In the U.K., where accounting standards require a high level of detail in the annual accounts, a large industrial company will have data for around 450 of these fields.

[11] Each pyramid starts with a named grand total, such as 'total assets'. The next level breaks this total into its constituents; for example 'total assets' is broken into 'fixed assets' and 'current assets'. Subsequent levels break the components down into their components, and so on. Thus, at each level, the pyramid shows how the data items relate to those in the level below and the level above.

following a detailed process to allocate information to the correct fields.[12] The database is updated via weekly CD ROM or by FTP.

Company Analysis is a Windows-based product created in 1995 by Mike Barron and Jeremy Smithers that allows powerful manipulation of the Extel Financials database, which is claimed to be the origin of the data. This product is very attractive to the researcher as it has, from our point of view, the user-friendliest interface of all the databases under study, with Datastream likely placed at the opposite end of the scale. Data retrieval from Company Analysis is significantly less time-consuming than for any of the other databases under study.

Extel data were previously incorporated by The Financial Times in 'Financial Times Extel Company Analysis', and in Primark as 'Primark Extel Company Analysis', the underlying data being available only through Company Analysis. After Primark was taken over by the Thomson Corporation, the information from Extel was also offered through the database platform software Thomson One Banker. Therefore, data retrieved from Extel Financials using Thomson One Banker or directly using Company Analysis should come from the same source. However, the samples obtained differ substantially, and the data obtained from Extel Financials through Thomson One Banker appear to be closer to Worldscope than to the data retrieved from Extel using Company Analysis.

Worldscope claims to be the only source needed for comprehensive and detailed accounts (more than 1,500 different items) and market data for over 40,000 public companies (31,000 active and 9,000 inactive) in more than fifty developed and emerging markets, for up to twenty years of historical data, including annual, interim and preliminary data. Worldscope fundamentals, which are updated daily, are available on Thomson One Banker, Global Access, Datastream, FTP site (in ASCII format), customized data files and third party platforms. Worldscope data are collected from corporate documents such as annual reports and press releases, exchange and regulatory agency filings, and newswires. Similar to Global Vantage, the data are not presented as reported in companies' financial statements but standardized.

Information from Thomson Financial can be retrieved through Thomson One Banker.[13] Thomson Financial combines, in a somewhat obscure way, information from several databases (Compustat and Worldscope, among others).

Finally, BvD Osiris is owned by Bureau van Dijk Electronic Publishing, SA, and it contains a wide range of accounting and other items for firms from over 120

[12] To describe this process, the database providers claim that: 'data editors are typically graduates with a background in accountancy . . . taking about seven hours to enter the data for one company; much more for a very complex company. Also, after inputting, the data undergoes more than 150 validation routines to ensure that the editor has not made any mistakes. If this test is successfully passed, the data is [*sic*] handed onto a senior editor who re checks the entered data. In addition, a Quality Assurance team constantly rechecks a large proportion of the population, to ensure that any errors are identified.'

[13] Thomson One Banker allows direct data download from its website and also provides an add in feature that permits downloading of large datasets into Microsoft Excel.

countries. Osiris claims to cover all publicly listed companies worldwide, as well as other major non-listed firms that are primary subsidiaries of publicly listed firms, or in certain cases, when clients request information from a particular company.

Bureau van Dijk also provides databases that offer coverage of non-listed small and medium size firms. Amadeus is their Pan-European database containing information on firms from thirty-four European countries. They also provide more detailed information on a national basis through the national versions of Amadeus, like FAME for the U.K. or SABI for Spanish and Portuguese firms. Data from the Bureau van Dijk databases can be retrieved using their own interface (which is the same for each of their databases), with two similar versions available: on-line and DVD. Osiris data can also be retrieved through WRDS. Apart from accounting data, Osiris provides information regarding ownership, news, ratings, share prices, forecasts, etc. The core accounting data are provided by World'Vest Base, and regional data are retrieved from Korea Information Service, Teikoku Databank (Japan), Huaxia International Business Credit Consulting Company (China), Multex (U.S.) and Edgar Online (U.S.). Different data sources are also used for other purposes. Accounting data are available as reported and also in what they call a standardized format to allow cross-border comparisons. BvD Osiris covers 24,700 listed, 900 unlisted and 2,690 delisted companies. The database is updated weekly over the internet and ten times a year on the DVD version. Data availability starts from 1986 although, as is the case with the rest of databases, the number of items available at the beginning of the coverage period is very limited.

In addition to these databases that provide accounting and financial data, accounting researchers might also wish to refer to databases for international data on other items such as analysts' forecasts—for example from I/B/E/S or JCF—or corporate governance data from diverse national providers such as Manifest, or Perfect Information for U.K. data, or ISS for U.S. data. In fact, specialized data on items such as CEO turnover or pension plan details are only available in a limited number of databases, and commonly, researchers do not use the more general accounting databases described previously to retrieve such data.

Hence, there is a wide variety of databases offering what would appear to be the same product: comprehensive international accounting data for firms worldwide. We test whether the data offered by these databases are in fact comparable, and whether differences, if they exist, may have an impact on the results and findings of market-based international accounting research.

RESEARCH DESIGN

Our main objective is to test if database choice has an effect on the results of empirical accounting studies using international data. To that end, we replicate the behaviour of a researcher who, having only one database available, intends to run a regression of market value on earnings and book value. This admittedly simple regression is run because (a) this is a specification commonly used in

capital markets research in accounting, and (b) a finding of differences amongst databases using the most parsimonious model possible would strongly support the prediction that database choice has important implications for research. To test for differences among databases, we collected all data available for the period 1990–99 in seven databases (Datastream, Global Vantage, Worldscope, Company Analysis, Thomson Financial, Extel Financials and BvD Osiris), for all firms covered in fourteen European countries (all the EU member states prior to the May 2004 enlargement, except Luxemburg, i.e., Austria, Belgium, Denmark, France, Finland, Germany, Greece, Ireland, Italy, Portugal, Spain, Sweden, the Netherlands and the United Kingdom). Financial firms are excluded from the analysis. This process generates seven sets of data, one for each database.[14] The information retrieved is in local currency, i.e., euros for member states in the Euro-zone, pounds sterling for the U.K., Danish crowns for Denmark and Swedish crowns for Sweden.

From each of these seven different sets of data, we remove missing data for the four variables necessary to carry out the empirical analysis, namely (a) market share price, (b) book value of shareholders' equity, (c) bottom line net income and (d) number of shares. After excluding firms with negative book value of shareholders' equity, we then calculate earnings per share and shareholders' equity per share, and remove the first and last percentiles of market share price, book value of shareholders' equity per share, and bottom line net income per share. Once the sample is constructed for each dataset, the following valuation model is run:

$$P_t = \alpha + \beta BVPS_t + \gamma EPS_t + e_t, \tag{1}$$

where P is the market share price, $BVPS$ is the book value of shareholders' equity per share, EPS is bottom-line earnings per share and t is the time period indicator.

As previously argued, the rationale underlying the choice of model (1) to test for differences across databases is that it employs the three most frequently used variables in market-based accounting research: prices, book values and earnings. Furthermore, for any given firm with data available in every database, these three are the basic items for which all databases should have data. Were other less obvious data items used, such as environmental provisions or loans to directors, they might be available only in certain databases. This would complicate the inferences to be drawn about differences across databases, as the samples would be expected to differ anyway. If discrepancies are found among databases using a parsimonious model, then we can safely assume that the differences in the results would be even more significant if we were to use less common data items. A full comparison of the databases under study would require the downloading of all available data (all items, all years, all firms) so that conclusions could be drawn. Clearly, that is an extensive task that, although potentially interesting, is not the focus of our current analysis.

[14] Data were downloaded around January 2004 using the subscription packages licensed to Lancaster University, except data from BvD Osiris, which were downloaded at the University of Valencia, during a research visit by Dr García Lara.

TABLE 2

NUMBER OF FIRM YEAR OBSERVATIONS BY DATABASE AND COUNTRY, 1990 99

	Company Analysis	Datastream	Extel Financials	Global Vantage	BvD Osiris	Thomson Financial	Worldscope
Austria	320	248	424	263	117	431	431
Belgium	398	193	530	259	196	558	555
Denmark	541	389	634	387	279	918	918
Finland	268	367	304	298	135	562	561
France	2,057	1,483	2,668	1,491	944	3,368	3,366
Germany	1,674	3,170	2,059	1,538	599	2,967	2,965
Greece	94	277	92	61	191	661	360
Ireland	326	98	313	243	90	370	369
Italy	604	331	842	425	218	1,004	998
Portugal	72	359	94	136	70	392	390
Spain	454	343	681	455	189	813	810
Sweden	408	759	560	506	367	965	964
The Netherlands	762	527	977	703	319	1,078	1,077
United Kingdom	5,661	5,317	8,909	5,837	1,593	8,601	8,582

RESULTS

Descriptive Statistics

As can be seen in Table 2, the final sample size varies substantially across databases for all countries. If the sample changes across databases, this will obviously affect the results of any type of empirical study. Determining which sample gives a better picture of the whole population is beyond the scope of this article and likely not to be testable anyway.

Interestingly, the database that offers by far the fewest number of observations is BvD Osiris. A detailed analysis of the data shows that this database only offers valid information to estimate model (1) from 1996 onwards, and hence almost half of our sample period is not covered. Although accounting data are available in BvD Osiris from 1986 onwards, market capitalization (item 35907) appears to be available only from 1996. Given this data limitation, we report all results for BvD Osiris, but we avoid any reference to them in the subsequent discussion of our results.

Analysing the largest sample, the U.K., we observe that the sample ranges from 8,909 firm-year observations using Extel Financials, to 5,317 in Datastream. Thus, the sample in Extel Financials is 67 per cent larger than that in Datastream. This difference in the sample across databases is similar for the other countries under study. For example, in Germany the sample size ranges from 3,170 in Datastream

TABLE 3

MARKET CAPITALISATION BY DATABASE AND COUNTRY

	Company Analysis		Datastream		Extel Financials		Global Vantage		BvD Osiris		Thomson Financial		Worldscope	
	Median	*SD*	Median	*SD*	Median	*SD*	Median	*SD*	Median	*SD*	Median	*SD*	Median	*SD*
Austria	67	565	108	531	79	491	139	637	110	626	98	528	98	528
Belgium	107	1,929	535	2,504	74	1,472	281	2,492	165	2,584	112	1,437	113	1,441
Denmark	691	10,280	1,006	10,269	680	8,873	1,403	9,909	813	13,258	425	7,633	425	7,633
Finland	355	4,045	197	2,796	308	12,511	384	4,791	346	19,441	100	9,233	100	9,241
France	136	5,726	276	6,643	102	4,386	223	6,163	137	8,214	80	4,027	80	4,027
Germany	185	4,882	72	3,447	182	4,352	236	8,660	181	12,399	117	3,812	117	3,813
Greece	321	11,397	48	406	334	2,026	228	6,345	230	1,499	53	656	35	159
Ireland	53	691	349	1,263	78	540	119	874	151	2,750	76	750	76	750
Italy	151	4,401	342	5,256	145	3,978	258	5,971	281	6,324	140	4,192	141	4,204
Portugal	208	1,954	57	2,494	291	2,366	83	1,412	216	1,727	41	1,198	42	1,201
Spain	229	5,236	471	6,004	219	4,326	314	3,351	534	8,743	183	4,017	187	4,024
Sweden	3,127	60,162	443	8,633	2,316	23,406	2,992	57,870	1,038	63,016	1,095	18,595	1,101	18,604
The Netherlands	171	10,272	353	11,059	112	3,555	216	9,563	262	12,837	119	8,360	119	8,364
United Kingdom	51	4,457	56	3,423	39	4,617	112	3,694	92	10,099	46	3,155	46	3,158

Note: Market capitalization is given in millions of monetary units. Monetary units are expressed in pounds sterling for the United Kingdom, in Danish crowns for Denmark, Swedish crowns for Sweden, and in euros for all other countries.

to 1,538 in Global Vantage; and in France, from 3,368 in Thomson Financial to 1,483 in Datastream. Except for Germany and the U.K., Thomson Financial always offers the widest firm-year coverage for all countries. The data coverage offered by Worldscope is virtually identical to that offered by Thomson Financial, except in the case of Greece. There is no clear picture as to which database offers the lowest firm-year availability as it varies substantially across countries. Thus, for academic research subscribers, Datastream offers the most restricted data availability for eight countries (Austria, Belgium, France, Ireland, Italy, Spain, the Netherlands and the U.K.), Global Vantage for three countries (Denmark, Germany and Greece), and Company Analysis for three countries (Finland, Portugal and Sweden). Clearly, as indicated above, such differences in the sample sizes must have an important impact on the results of our empirical analysis. If the samples are so different, it seems inevitable that the results will also vary, unless every sample is built to be a perfect representation of the whole population, which is unlikely.

One of the consequences of the coverage differences among the databases under study is the problem of introducing size bias into empirical research. It is expected that all databases will include the larger firms, while they will probably differ in their coverage of the smaller firms. Table 3 provides an insight into possible size bias. For each country, this table gives the median market capitalization of companies by database. Given the sample sizes set out previously in Table 2, we can now see that for eight out of the fourteen countries (Denmark, Finland, France, Germany, Italy, Portugal, Spain and the U.K.) the database with the largest number of observations is also the database with the smallest median firm size. Also, the largest median firm size in nine countries (Belgium, France, Ireland, Italy, Spain, the Netherlands, Finland, Sweden and Germany) corresponds to the database with the smallest data coverage. Therefore, those databases that cover fewer firms seem to focus on the largest ones, as could be expected. This size bias likely also introduces survivorship bias into the sample, as larger firms tend to survive longer than their smaller counterparts. Figure 1 represents graphically the relationship between firm size and coverage by the databases under study for the largest of the samples, that is, the U.K. We observe that the median market value decreases monotonically as firm coverage increases. Therefore, size bias appears to be an issue for databases covering the smallest number of firms, like Global Vantage, Osiris and Datastream. Researchers should consider the consequences of this potential bias in their research designs.

Table 4 shows the descriptive statistics of the per share variables used to estimate regression (1), which vary noticeably across databases. This is not surprising, considering the pronounced differences in the samples across databases.

Regression Results

For brevity, regression results using the seven databases are only provided for the two largest economies in the EU, those of the U.K. and Germany. These results are reported in Panels A of Tables 5 and 6. To give a straightforward indication of whether the differences between databases are significant, Panel B of each table contains a matrix indicating whether the differences in the coefficients between

TABLE 4

DESCRIPTIVE STATISTICS

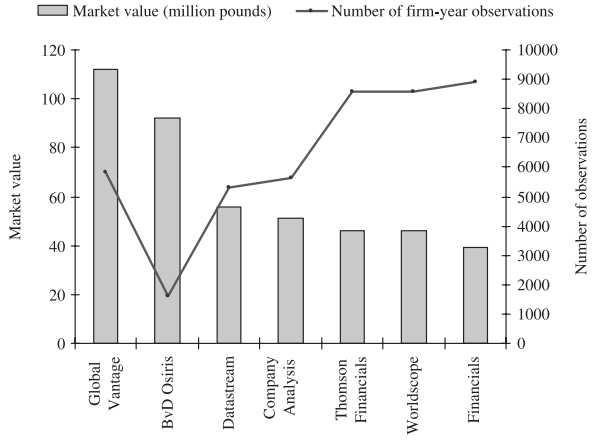
	Company Analysis		Datastream		Extel Financials		Global Vantage		BvD Osiris		Thomson Financial		Worldscope	
	Median	*SD*	Median	*SD*	Median	*SD*	Median	*SD*	Median	*SD*	Median	*SD*	Median	*SD*
Austria														
BVPS	36.8	77.1	42.9	132.4	25.7	49.5	39.4	74.7	37.0	87.2	20.9	38.9	20.9	38.9
Share price	47.8	86.5	48.8	80.6	38.3	52.9	53.5	89.0	34.5	37.2	36.3	60.3	36.3	60.3
EPS	2.6	10.2	3.7	12.1	1.5	7.7	3.3	11.1	2.7	8.0	1.8	6.8	1.8	6.8
Belgium														
BVPS	60.2	236.0	58.5	175.4	36.5	110.5	74.7	1,464.9	23.8	289.8	32.4	80.9	32.2	79.8
Share price	94.1	337.7	124.9	277.2	57.8	133.4	99.1	297.3	48.5	171.4	54.8	112.0	54.4	108.4
EPS	5.5	30.2	5.5	18.2	2.9	13.4	7.2	57.4	2.6	31.4	2.3	9.9	2.3	9.7
Denmark														
BVPS	333.0	1,184.8	317.7	2,158.42	190.2	288.5	324.1	807.8	250.2	257.5	171.9	234.3	171.9	234.3
Share price	460.0	9,559.1	430.0	11,857.4	238.9	2,663.2	390.0	758.4	258.0	500.5	240.0	1,055.4	240.0	1,055.4
EPS	28.2	135.5	29.1	234.4	14.5	41.7	29.4	97.5	21.8	29.5	13.2	36.2	13.2	36.2
Finland														
BVPS	12.8	13.8	13.0	24.4	6.9	5.7	12.6	22.7	9.2	7.1	4.8	6.2	4.8	6.2
Share price	18.7	21.0	13.8	20.0	9.1	10.7	15.8	21.3	10.5	9.2	7.5	11.3	7.6	11.3
EPS	1.3	2.2	1.4	3.0	0.6	1.3	1.5	3.1	1.1	1.2	0.5	1.3	0.5	1.3
France														
BVPS	35.2	76.6	37.8	69.7	24.8	42.5	36.3	167.3	18.1	45.2	23.8	43.1	23.8	43.1
Share price	60.2	124.9	70.0	120.4	38.1	50.1	61.5	103.3	32.0	41.2	35.8	53.2	35.8	53.2
EPS	3.3	9.2	3.8	9.0	2.0	4.7	3.2	21.7	1.6	3.8	1.9	5.2	1.9	5.2
Germany														
BVPS	72.4	79.8	72.8	102.1	15.8	40.6	64.0	62.6	13.7	53.2	12.7	39.7	12.7	39.7
Share price	138.0	146.7	145.7	278.5	28.6	107.6	133.7	162.3	23.8	127.9	28.0	130.2	28.0	130.3
EPS	5.0	13.6	3.9	14.1	1.1	7.0	4.4	11.8	1.3	5.8	0.9	7.6	0.9	7.6
Greece														
BVPS	4.2	44.7	3.7	6.7	2.6	2.6	4.2	5.9	1.5	3.1	1.3	1.8	1.2	1.8
Share price	16.0	739.0	7.6	11.8	9.3	11.2	12.2	16.4	5.6	11.7	3.0	8.3	2.1	3.2

BVPS	0.7	1.4	1.7	1.6	0.8	1.1	1.1	1.4	0.8	1.4	0.7	1.2	0.7	1.2
Share price	1.2	5.3	2.5	3.9	1.3	2.0	2.1	3.1	1.7	5.1	1.2	2.6	1.2	2.6
EPS	0.1	0.2	0.2	0.3	0.1	0.2	0.2	0.3	0.2	0.3	0.1	0.2	0.1	0.2
Italy														
BVPS	1.6	3.8	1.9	6.0	1.8	3.9	1.9	2.6	2.0	11.0	1.6	3.3	1.6	3.3
Share price	1.8	3.6	2.7	4.4	2.0	3.2	2.7	3.6	3.9	8.3	2.1	3.2	2.1	3.2
EPS	0.1	0.4	0.2	0.5	0.1	0.3	0.1	0.3	0.2	0.6	0.1	0.3	0.1	0.3
Portugal														
BVPS	9.2	3.8	10.6	217.6	2.1	3.9	10.6	5.7	2.4	2.6	6.6	7.2	6.6	7.2
Share price	16.3	13.0	11.1	221.1	5.0	6.8	12.5	12.4	3.4	5.7	5.2	6.7	5.2	6.7
EPS	0.9	0.8	0.6	35.6	0.2	0.4	0.6	1.1	0.2	0.3	0.2	1.1	0.2	1.1
Spain														
BVPS	10.2	10.9	9.1	28.4	5.8	10.3	10.5	12.1	4.6	5.6	5.3	9.6	5.2	9.6
Share price	15.1	20.6	15.0	31.4	7.4	13.0	15.3	20.6	9.6	10.1	7.2	11.1	7.2	11.1
EPS	1.0	1.7	0.9	4.6	0.5	1.4	0.9	1.7	0.5	0.8	0.4	1.2	0.4	1.2
Sweden														
BVPS	66.2	54.8	59.4	545.8	45.9	32.5	76.5	262.1	39.5	89.1	35.6	29.6	35.6	29.7
Share price	122.8	103.4	85.0	94.5	74.6	58.7	108.0	102.3	72.1	88.7	60.0	50.3	60.0	50.3
EPS	8.3	11.3	6.2	33.0	4.9	6.9	9.0	28.5	4.2	11.3	3.7	6.3	3.7	6.3
The Netherlands														
BVPS	13.0	31.8	13.2	15.3	9.4	83.4	12.4	19.0	6.5	9.7	8.8	15.2	8.8	15.2
Share price	26.1	43.0	29.0	27.1	16.5	67.8	26.4	28.5	16.7	13.9	16.0	20.6	16.0	20.6
EPS	2.0	4.5	2.2	2.6	1.3	4.4	1.9	2.9	1.1	1.5	1.2	2.2	1.2	2.2
United Kingdom														
BVPS	0.7	1.3	0.8	1.4	0.6	1.2	0.9	1.2	0.9	1.8	0.7	1.2	0.7	1.2
Share price	1.5	2.4	1.6	2.3	1.2	1.9	1.9	2.4	1.5	2.5	1.3	1.9	1.3	1.9
EPS	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.3	0.1	0.2	0.1	0.2

Notes: BVPS book value of shareholders' equity per share, excluding preferred shares; Share price market share price at the balance sheet date, *P*; EPS bottom line earnings per share. See exact variable definitions in Appendix A. All per share values are given in units of sterling pounds for the U.K., Swedish crowns for Sweden, Danish crowns for Denmark and euros for all other countries.

FIGURE 1

MEDIAN MARKET VALUE AND FIRM COVERAGE OF EACH DATABASE SAMPLE (U.K. ONLY)



pairs of databases are significant or not. The analysis of differences for the other countries is provided in Appendix A.

Panel A of Table 5 shows the results for the U.K. It can be observed readily that the coefficients and measures of fit change substantially across databases. The intercept ranges from 0.78 in Thomson Financial to 1.15 in Global Vantage; the coefficient of the book value of shareholders' equity (β) from 0.28 in Datastream to 0.69 in Worldscope; the earnings multiple (γ) from 3.82 in Extel Financials to 6.88 in Global Vantage; and, finally, the adjusted coefficient of determination from 0.43 in Global Vantage to 0.53 in Company Analysis. As observed in Panel A of Table 6, differences are even more pronounced for Germany, with the intercept ranging from 5.33 in Thomson Financial to 124.10 in Datastream, the coefficient of the book value of equity from 0.77 in Datastream to 2.45 in Thomson Financial, the earnings multiple from 1.44 in Worldscope to 3.29 in Datastream, and the adjusted R-squared from 0.15 in Datastream to 0.61 in Thomson Financial and Worldscope. Panels B of Tables 5 and 6 show the significance of these differences: we only fail to find significant differences between Extel Financials and Worldscope in the case of the U.K. and between Worldscope and Thomson Financial in the case of Germany. As can be observed in the results reported in Appendix A, for the remaining countries, only in the case of Ireland does the choice of database seem to be less of an issue as just Company Analysis offers significantly different results.

It should be noted that all intercepts and slope coefficients in Tables 5 and 6 have the expected signs, are statistically significant and take reasonable values.

TABLE 5

UNITED KINGDOM

Panel A: Regression results

	α *t* stat	β *t* stat	γ *t* stat	Adj. R^2
Company Analysis	0.86	0.50	6.72	0.53
	28.87	*14.80*	*25.95*	
Datastream	1.04	0.28	6.80	0.48
	31.11	*7.41*	*25.34*	
Extel Financials	0.83	0.66	3.82	0.49
	40.55	*24.49*	*20.58*	
Global Vantage	1.15	0.37	6.88	0.43
	32.21	*9.92*	*23.74*	
BvD Osiris	1.12	0.54	4.67	0.39
	13.82	*5.42*	*7.69*	
Thomson Financial	0.78	0.68	4.33	0.52
	39.17	*23.52*	*19.73*	
Worldscope	0.79	0.69	4.18	0.51
	39.95	*24.12*	*19.14*	

Panel B: Significance of difference in coefficients α, β and γ across databases

	CA	DS	EF	GV	OS	TF
DS	α***,β***					
EF	β***,γ***	α***,β***,γ***				
GV	α***,β**	α**,β*	α***,β***,γ***			
OS	α***,γ***	β**,γ***	α***	γ***		
TF	α**,β***,γ***	α***,β***,γ***	γ*	α***,β***,γ***	α***	
WS	α*,β***,γ***	α***,β***,γ***		α***,β***,γ***	α***	

$$Model: P_t \quad \alpha + \beta BVPS_t + \gamma EPS_t + e_t,$$

where P is the market share price, $BVPS$ is the book value of shareholders' equity per share, EPS is bottom line earnings per share and t is the time period indicator.

All t statistics are heteroskedasticity consistent. We estimate the statistical significance of the differences between coefficients in regressions using different databases as $\theta_1 \ \theta_2$ divided by $\sqrt{\sigma_1^2 + \sigma_2^2}$, where θ_i is the estimated coefficient and σ_i the standard error for variable i. ***, ** and * indicate that the difference is significant at a 1%, 5% and 10% level. The following abbreviations have been used: CA Company Analysis; DS Datastream; EF Extel Financials; GV Global Vantage; OS BvD Osiris; TF Thomson Financial; WS Worldscope.

This implies that none of the databases under analysis provides implausible data for the U.K. or Germany. This is also true for the Netherlands. However, this is not always the case, particularly for the smaller countries in the sample. For example, for Austria, γ is not significant using Datastream, Global Vantage,

TABLE 6

GERMANY

Panel A: Regression results

	α *t* stat	β *t* stat	γ *t* stat	Adj. R^2
Company Analysis	75.57	0.91	2.13	0.40
	16.20	*13.67*	*7.10*	
Datastream	124.10	0.77	3.29	0.15
	19.75	*11.47*	*7.08*	
Extel Financials	9.42	1.84	1.61	0.57
	5.38	*23.68*	*2.82*	
Global Vantage	43.90	1.60	1.82	0.49
	9.91	*22.92*	*4.35*	
BvD Osiris	15.79	0.87	2.65	0.54
	7.06	*7.33*	*3.37*	
Thomson Financial	5.33	2.45	1.49	0.61
	3.69	*34.83*	*3.59*	
Worldscope	5.96	2.44	1.44	0.61
	4.22	*34.27*	*3.46*	

Panel B: Significance of difference in coefficients α, β and γ across databases

	CA	DS	EF	GV	OS	TF
DS	α***,γ**					
EF	α***,β***	α***,β***,γ**				
GV	α***,β***	α***,β***,γ**	α***,β**			
OS	α***	α***	α**,β***	α***,β***		
TF	α***,β***	α***,β***,γ***	α*,β***	α***,β***	α***,β***	
WS	α***,β***	α***,β***,γ***	β***	α***,β***	α***,β***	

$$Model: P_t \quad \alpha + \beta BVPS_t + \gamma EPS_t + e_t,$$

where P is the market share price, $BVPS$ is the book value of shareholders' equity per share, EPS is bottom line earnings per share and t is the time period indicator.

All t statistics are heteroskedasticity consistent. We estimate the statistical significance of the differences between coefficients in regressions using different databases as $\theta_1 \quad \theta_2$ divided by $\sqrt{\sigma_1^2 + \sigma_2^2}$, where θ_i is the estimated coefficient and σ_i the standard error for variable i. ***, ** and * indicate that the difference is significant at a 1%, 5% and 10% level. The following abbreviations have been used: CA Company Analysis; DS Datastream; EF Extel Financials; GV Global Vantage; OS BvD Osiris; TF Thomson Financial; WS Worldscope.

Thomson Financial or Worldscope, and in the case of Datastream, it takes a negative sign. For the other countries, the results are less striking; however, the usage of Datastream, Company Analysis, and Global Vantage appears more problematic.

Specifically, the estimation based on Datastream results in an insignificant β for Denmark, Greece, Italy, Spain and Sweden, taking a negative value for Greece; while γ is negatively significant for Portugal. In the case of Company Analysis, β is not significant for Denmark, negative for Portugal and negatively significant for Belgium; while γ is not significant for Ireland and France; additionally, Company Analysis data result in negative intercepts for Ireland and Greece. Finally, using Global Vantage results in non-significant β coefficients for Finland, France and Portugal, being negative in the case of Portugal. On the other hand, Global Vantage is the only database that generates a positive intercept value for Denmark. While there are no clear patterns, the number of observations available is obviously a key factor in the analysis.

Consistent with the descriptive statistics, the regression results show that some of the databases provide very similar data. Specifically, Thomson Financial and Worldscope produce significantly different coefficients only for Greece, indicating the samples provided by these two databases are fairly consistent. Also, Extel Financials and Worldscope generate similar coefficients in eight out of the fourteen countries. Overall, the results observed for these fourteen member states of the EU confirm our prediction that the use of different databases to analyse the same country with the same research design yields significantly different coefficients.

Further Analysis: Matching Samples

The results detailed in the previous section are explained by differences in the samples across databases. However, it is unclear whether these results are (a) driven exclusively by heterogeneous database firm coverage (i.e., the samples contain different firms), or (b) partially attributable to differences in the figures available for the same firms in different databases, which could be caused by errors, by the adjustments or standardization made to as-reported data or by a different variable definition among databases.[15]

We analyse the common observations in all databases to find out to what extent, if at all, the databases offer different observed values for the same firms. This test sets up a rather unrealistic scenario because in practice researchers do not normally have access to all the databases, and once a database is chosen, no adjustments are made to the sample on the basis of observations matching those in other databases, regardless of whether other databases are available or not. This is why, for our main analysis, we work with all available observations, in an attempt to replicate as closely as possible the decisions made by a researcher who has chosen a particular database. The study by Land and Lang (2002) is one example of this pattern of behaviour. They claim to use Global Vantage, instead of Datastream or Worldscope, because of wider data availability. However, once they have chosen Global Vantage, they simply use all observations available in that database, not constraining their sample to common observations across the three databases. However, even though our objective is to replicate what the researcher would do and then determine whether the results would change using

[15] As can be seen in Appendix B, variable definition is probably not an issue in our analysis.

a different database, constraining the samples to common observations across all databases will provide additional evidence on the observed differences and whether they can be attributable, at least partially, to differences in the figures available for the same firms in alternative databases.

Observations are matched across the different databases using the SEDOL code. BvD Osiris is excluded in the matching, as it would constrain the sample to the period 1995–99, and would mean losing six years of observations. Table 7 shows the median values of shareholders' equity, market capitalization and net income for the firms in common, both in terms of the figures appearing in financial statements and on a per share basis. Table 7 also shows the number of common observations. Overall, we have approximately 4,000 common observations, of which approximately a half are for U.K. firms. After matching the samples, we find that for some countries it would not be feasible to perform any additional analysis, as the resulting sample is very small. This is the case for Ireland, Greece and Portugal.

The descriptive statistics in Table 7 show that the observed values (both in terms of the figures appearing in financial statements and on a per share basis) are similar across databases, and appear much closer than for the unmatched sample. Analysing these descriptive statistics, we can identify two clear groups or clusters with extremely similar numbers. In fact, there appear to be two main sources of data. On the one hand, Extel Financials, Thomson Financial and Worldscope offer very similar values for the three variables analysed, both in terms of the financial statement amounts and on a per share basis. The other three databases, Company Analysis, Datastream and Global Vantage, also offer similar values, and with regard to the financial statement variables there is in fact little difference across the two groups of databases. However, some marked differences remain with regard to the per share data, which is likely to be attributable to differences in the number of shares reported as common shares outstanding at the balance sheet date across the different databases.

Clearly, these descriptive statistics indicate that the use of a matched sample will probably reduce the differences between estimated regression coefficients. Indeed, when we replicate the regression analysis using the common sample across the six databases, only for Germany do the differences, albeit reduced, remain substantial and significant. For all other countries, the differences between coefficients are much less pronounced.[16] In fact, for two countries, Italy and Spain, there is not a single significant difference in the coefficients across the databases. Some differences still remain in the other countries, although much less pronounced than with the unmatched samples.

There is still considerable accounting method diversity across the EU member states under analysis. Some of the databases attempt to create comparable numbers between firms and across countries by standardizing accounting values, while others prefer to present data as reported. The databases that standardize values do not disclose information on how this procedure is carried through,

[16] These results are available from the authors upon request.

TABLE 7

MEDIAN VALUES OF SELECTED VARIABLES FOR THE MATCHED SAMPLE

	Company Analysis		Datastream		Extel Financials		Global Vantage		Thomson Financial		Worldscope		N. obs
	Entity	Per share	Entity	Per share	Entity	Per share	Entity	Per share	Entity	Per share	Entity	Per share	
Austria													
Shareholders' equity	352.0	38.6	337.2	44.6	352.0	32.2	351.1	38.6	352.0	32.1	352.0	32.1	
Market capitalization	440.0	53.8	440.0	53.8	440.0	39.7	450.8	53.5	440.4	39.7	440.4	39.7	77
Net income	30.4	3.3	19.5	3.7	30.4	2.1	24.0	3.0	23.2	2.2	23.2	2.2	
Belgium													
Shareholders' equity	430.9	63.6	432.2	64.3	436.3	22.4	423.8	67.0	429.7	22.4	429.7	22.4	
Market capitalization	937.6	164.7	842.2	165.1	968.0	46.2	901.0	160.2	927.1	43.0	927.1	43.0	44
Net income	52.6	9.4	52.0	9.2	54.6	2.3	51.7	9.7	52.6	2.2	52.6	2.2	
Denmark													
Shareholders' equity	1,126.0	277.6	1,104.7	315.8	1,126.0	191.5	1,104.7	315.8	1,104.7	180.0	1,104.7	180.0	
Market capitalization	1,557.9	455.0	1,484.6	455.0	1,640.4	268.0	1,502.6	380.0	1,740.8	260.0	1,740.8	260.0	145
Net income	111.6	27.9	118.9	31.3	109.5	16.2	115.9	31.3	111.6	15.9	111.6	15.9	
Finland													
Shareholders' equity	375.2	10.7	383.3	11.7	375.2	7.6	380.1	10.8	364.3	7.6	364.3	7.6	
Market capitalization	535.8	11.9	513.6	11.9	535.8	9.8	549.7	11.1	575.0	9.8	575.0	9.8	83
Net income	33.5	1.0	33.5	1.1	33.5	0.8	36.6	1.0	35.4	0.8	35.4	0.8	
France													
Shareholders' equity	123.8	37.1	121.2	35.6	123.8	22.9	121.9	36.7	122.0	23.1	122.0	23.1	
Market capitalization	194.0	59.5	188.8	60.0	191.1	38.1	194.9	61.2	192.7	41.3	192.7	41.3	442
Net income	10.0	3.3	9.5	3.3	9.5	2.1	10.0	3.3	10.2	2.1	10.2	2.1	
Germany													
Shareholders' equity	129.8	68.9	131.8	71.4	129.0	13.3	130.3	67.5	129.8	12.1	129.8	12.1	
Market capitalization	275.7	136.5	253.6	131.4	257.5	26.3	265.9	136.6	273.0	25.6	273.0	25.6	619
Net income	12.6	4.7	12.5	4.8	12.2	1.2	12.3	4.8	12.8	1.1	12.8	1.1	
Greece													
Shareholders' equity	191.0	9.6	201.1	10.5	191.0	4.0	205.4	7.8	195.0	3.8	100.7	2.3	
Market capitalization	453.0	17.3	469.8	17.9	453.0	7.3	492.8	19.1	453.4	9.8	242.7	5.1	11
Net income	25.4	0.5	25.0	0.6	25.4	0.4	27.9	0.6	24.1	0.4	15.4	0.3	

(continued)

	Company Analysis		Datastream		Extel Financials		Global Vantage		Thomson Financial		Worldscope		N. obs
	Entity	Per share	Entity	Per share	Entity	Per share	Entity	Per share	Entity	Per share	Entity	Per share	
Ireland													
Shareholders' equity	241.5	0.8	141.2	0.8	139.5	0.8	93.6	0.8	139.0	0.8	139.0	0.8	
Market capitalization	315.8	1.8	192.7	1.8	277.2	1.8	231.6	2.1	277.2	1.6	277.2	1.7	3
Net income	18.6	0.2	22.3	0.3	22.3	0.2	21.2	0.3	22.3	0.2	22.3	0.2	
Italy													
Shareholders' equity	241.5	2.6	247.2	2.6	241.5	2.7	239.5	2.8	240.0	2.6	240.0	2.6	
Market capitalization	315.8	4.9	285.4	4.9	315.8	4.1	326.0	5.6	382.6	4.4	382.6	4.4	74
Net income	18.6	0.2	17.5	0.2	18.6	0.2	18.7	0.2	18.6	0.1	18.6	0.1	
Portugal													
Shareholders' equity	171.0	10.0	171.0	10.5	171.0	2.2	172.4	10.0	171.0	2.2	171.0	2.2	
Market capitalization	265.9	22.0	268.0	22.0	265.9	8.2	268.0	22.4	256.6	8.1	256.6	8.1	15
Net income	10.9	1.4	10.9	1.2	10.9	0.4	11.1	1.5	10.9	0.3	10.9	0.3	
Spain													
Shareholders' equity	197.8	9.0	202.5	9.3	197.8	4.4	199.4	9.5	197.8	4.3	197.8	4.3	
Market capitalization	517.9	15.3	517.9	15.3	517.9	7.2	522.1	15.7	517.9	7.2	517.9	7.2	137
Net income	28.2	0.9	28.2	0.9	28.2	0.4	28.5	0.9	28.2	0.4	28.2	0.4	
Sweden													
Shareholders' equity	701.3	63.9	701.3	72.4	701.3	49.3	701.3	72.4	701.3	46.4	701.3	46.4	
Market capitalization	1,654.3	117.5	1,653.6	117.5	1,654.3	96.0	1,653.6	107.1	1,757.9	86.6	1,757.9	86.6	83
Net income	103.7	8.0	103.7	9.1	103.7	6.8	103.7	8.9	103.7	6.5	103.7	6.5	
The Netherlands													
Shareholders' equity	144.5	12.5	145.3	13.4	144.5	8.9	144.8	12.5	144.4	9.1	144.4	9.1	
Market capitalization	299.1	25.4	298.6	25.4	299.1	20.3	300.6	25.5	297.5	20.3	297.5	20.3	212
Net income	24.3	2.2	24.3	2.1	24.3	1.6	24.1	2.1	23.8	1.5	23.8	1.5	
United Kingdom													
Shareholders' equity	44.0	0.9	45.0	0.9	45.4	0.7	43.6	0.9	42.6	0.7	42.6	0.7	
Market capitalization	100.9	2.1	99.6	2.0	103.7	1.7	101.8	2.1	101.3	1.6	101.3	1.6	2,039
Net income	5.9	0.1	5.8	0.1	5.9	0.1	5.9	0.1	5.8	0.1	5.8	0.1	

and thus, some differences across databases could be due to this standardization process. However, we do not observe any clear pattern in terms of differences between those databases that claim to standardize the data and those that claim to present data as reported and, overall, the results suggest that the differences introduced by the adjustments made to the as-reported data in some of the databases are not sufficiently important to inhibit comparability across databases, at least as far as shareholders' equity, earnings and market value are concerned. Finally, regarding the reliability of the data provided, the results suggest that, for the items under consideration, errors are not a significant problem across the databases examined. This conclusion, however, might not apply to other less commonly used items.

SUMMARY AND CONCLUSIONS

We address the question whether database choice affects the results and findings of international empirical research on accounting. We analyse the results of running a simple empirical adaptation of the Ohlson (1995) model for fourteen EU member states, obtaining the data from seven widely used databases: Datastream, Global Vantage, Company Analysis, Worldscope, Thomson Financial, Extel Financials and BvD Osiris.

The results show that parameters and measures of fit from a simple empirical model regressing book value of shareholders' equity and earnings on market value of the firm differ significantly across databases. There are two main conclusions. First, differences between databases exist and lead to differences in the results of even a very simple empirical study using key accounting variables. The results suggest that these differences are mainly attributable to heterogeneous firm-coverage across databases. When focusing on the common observations across all databases, the differences disappear almost completely. Second, our results are shown to be relevant both for researchers using empirical data in a European context, and also for academic institutions with an interest in investing in new data sources. Differences observed across databases also pose questions regarding the generalizability, reliability and comparability of the results of empirical studies using European accounting data. Which database offers more generalizable data is beyond the scope of the present study, but our results suggest that this could be an interesting issue to pursue further, because the differences we find are mainly attributable to variation in firm coverage across databases, and one might think that researchers should use the data set that gives a more accurate picture of the whole population. Regarding reliability, although our results suggest that differences in the observed values of common items across databases are minimal, further research on data reliability could also be of interest to accounting academics.

REFERENCES

Abarbanell, J., and R. Lehavy, 'Differences in Commercial Database Reported Earnings: Implications for Inferences in Research on Analyst Forecast Rationality, Earnings Management, and the

Information Content and Value Relevance of Earnings', Working Paper, University of North Carolina and University of California, Berkeley, 2000.

Aboody, D., M. E. Barth and R. Kasznik, 'Revaluations of Assets and Future Firm Performance: Evidence From the U.K.', *Journal of Accounting and Economics*, January 1999.

Ali, A., and L. Hwang, 'Country Specific Factors Related to Financial Reporting and the Value Relev ance of Accounting Data', *Journal of Accounting Research*, Spring 2000.

Ashbaugh, H., and M. Pincus, 'Domestic Accounting Standards, International Accounting Standards, and the Predictability of Earnings', *Journal of Accounting Research*, December 2001.

Baginski, S. P., J. M. Hassel and M. D. Kimbrough, 'The Effect of Legal Environment on Voluntary Disclosure: Evidence From Management Earnings Forecasts Issued in U.S. and Canadian Markets', *The Accounting Review*, January 2002.

Ball, R., S. P. Kothari and A. Robin, 'The Effect of International Institutional Factors on Properties of Accounting Earnings', *Journal of Accounting and Economics*, February 2000.

Ball, R., A. Robin and S. Wu, 'Incentives Versus Standards: Properties of Accounting Income in four East Asian Countries', *Journal of Accounting and Economics*, December 2003.

Barth, M. E., and G. Clinch, 'International Accounting Differences and Their Relation to Share Prices: Evidence From U.K., Australian, and Canadian Firms', *Contemporary Accounting Research*, Spring 1996.

Basu, S., 'Discussion of "International Differences in the Timeliness, Conservatism, and Classification of Earnings" ', *Journal of Accounting Research*, Supplement, 1999.

Basu, S., L. Hwang and C. Jan, 'International Variation in Accounting Measurement Rules and Analysts' Earnings Forecast Errors', *Journal of Business Finance and Accounting*, November 1998.

Bennin, R., 'Error Rates in CRSP and COMPUSTAT: A Second Look', *Journal of Finance*, December 1980.

Bhagat, S., and I. Welch, 'Corporate Research and Development Investments: International Compar isons', *Journal of Accounting and Economics*, March May 1995.

Bhattacharya, U., H. Daouk, and M. Welker, 'The World Price of Earnings Opacity', *The Accounting Review*, July 2003.

Black, E. L., K. F. Sellers and T. S. Manly, 'Earnings Management Using Asset Sales: An Inter national Study of Countries Allowing Noncurrent Asset Revaluation', *Journal of Business Finance and Accounting*, November 1998.

Bushman, R. M., and J. D. Piotroski, 'Financial Reporting Incentives for Conservative Accounting: The Influence of Legal and Political Institutions', *Journal of Accounting and Economics*, September 2006.

Courtenay, S. M., and S. B. Keller, 'Errors in Databases Revisited: An Examination of the CRSP Shares Outstanding Data', *The Accounting Review*, January 1994.

DeFond, M. L., and M. Hung, 'Investor Protection and Corporate Governance: Evidence From Worldwide CEO Turnover', *Journal of Accounting Research*, May 2004.

Elton, E. J., M. J. Gruber and C. R. Blake, 'A First Look at the Accuracy of the CRSP Mutual Fund Database and a Comparison of the CRSP and Morningstar Mutual Fund Databases', *Journal of Finance*, December 2001.

Fan, J. P. H., and T. J. Wong, 'Corporate Ownership Structure and the Informativeness of Accounting Earnings in East Asia', *Journal of Accounting and Economics*, August 2002.

Freeman, R., and S. Tse, 'An Earnings Prediction Approach to Examining Intercompany Information Transfers', *Journal of Accounting and Economics*, December 1992.

García Lara, J. M., B. García Osma and A. Mora, 'The Effect of Earnings Management on the Asym metric Timeliness of Earnings', *Journal of Business Finance and Accounting*, April/May 2005.

Giner, B., and W. Rees, 'On the Asymmetric Recognition of Good and Bad News in France, Germany and the United Kingdom', *Journal of Business Finance and Accounting*, November/December 2001.

Gordon, E. A., and P. R. Joos, 'Unrecognized Deferred Taxes: Evidence From the U.K.', *The Accounting Review*, January 2004.

Guenther, D. A., and A. J. Rosman, 'Differences Between COMPUSTAT and CRSP SIC Codes and Related Effects on Research', *Journal of Accounting and Economics*, July 1994.

Guenther, D. A., and D. Young, 'The Association Between Financial Accounting Measures and Real Economic Activity: A Multinational Study', *Journal of Accounting and Economics*, February 2000.

Haw, I., B. Hu, L. Hwang and W. Wu, 'Ultimate Ownership, Income Management, and Legal and Extra Legal Institutions', *Journal of Accounting Research*, May 2004.

Hope, O., 'Disclosure Practices, Enforcement of Accounting Standards, and Analysts' Forecast Accuracy: An International Study', *Journal of Accounting Research*, May 2003.

Hung, M., 'Accounting Standards and Value Relevance of Financial Statements: An International Analysis', *Journal of Accounting and Economics*, December 2000.

Joos, P., and M. Lang, 'The Effects of Accounting Diversity: Evidence From the European Union', *Journal of Accounting Research*, Supplement, 1994.

Kallapur, S., and S. Y. S. Kwan, 'The Value Relevance and Reliability of Brand Assets Recognized by U.K. Firms', *The Accounting Review*, January 2004.

Kern, B. B., and M. H. Morris, 'Differences in the COMPUSTAT and Expanded Value Line Databases and the Potential Impact on Empirical Research', *The Accounting Review*, January 1994.

Khurana, I. K., and K. K. Raman, 'Litigation Risk and the Financial Reporting Credibility of Big 4 Versus Non Big 4 Audits: Evidence From Anglo American Countries', *The Accounting Review*, April 2004.

Krishnan, J., and E. Press, 'The North America Industry Classification System and its Implication for Accounting Research', *Contemporary Accounting Research*, Winter 2003.

Land, J., and M. H. Lang, 'Empirical Evidence on the Evolution of International Earnings', *The Accounting Review*, October 2002.

Lang, M., K. V. Lins and D. P. Miller, 'Concentrated Control, Analyst Following, and Valuation: Do Analysts Matter Most When Investors are Protected Less?', *Journal of Accounting Research*, June 2004.

LaPorta, R., F. Lopez de Silanes and A. Shleifer, 'Corporate Ownership Around the World', *The Journal of Finance*, April 1999.

LaPorta, R., F. Lopez de Silanes, A. Shleifer and R. W. Wishny, 'Legal Determinants of External Finance', *Journal of Finance*, July 1997.

Leuz, C., D. Nanda and P. D. Wysocki, 'Earnings Management and Investor Protection: An International Comparison', *Journal of Financial Economics*, September 2003.

Meek, G. K., and W. B. Thomas, 'A Review of Markets Based International Accounting Research', *Journal of International Accounting Research*, Vol. 3, No. 1, 2004.

Monsen, W., and W. A. Wallace, 'Evolving Financial Reporting Practices: A Comparative Study of the Nordic Countries' Harmonization Efforts', *Contemporary Accounting Research*, Spring 1995.

Ohlson, J. A., 'Earnings, Book Values, and Dividends in Equity Valuation', *Contemporary Accounting Research*, Spring 1995.

Pope, P. F., and M. Walker, 'International Differences in the Timeliness, Conservatism, and Classification of Earnings', *Journal of Accounting Research*, Supplement, 1999.

Raonic, I., S. McLeay and I. Asimakopoulos, 'The Timeliness of Income Recognition by European Companies: An Analysis of Institutional and Market Complexity', *Journal of Business Finance and Accounting*, January/March 2004.

Rosenberg, B., and M. Houglet, 'Error Rates in CRSP and COMPUSTAT Data Bases and Their Implications', *Journal of Finance*, September 1974.

San Miguel, J. G., 'The Reliability of R&D Data in COMPUSTAT and 10 K Reports', *The Accounting Review*, July 1977.

Seetharaman, A., F. A. Gul, and S. G. Lynn, 'Litigation Risk and Audit Fees: Evidence From U.K. Firms Cross Listed on U.S. Markets', *Journal of Accounting and Economics*, February 2002.

Young, D., and D. A. Guenther, 'Financial Reporting and International Capital Mobility', *Journal of Accounting Research*, June 2003.

APPENDIX A

SIGNIFICANCE OF THE DIFFERENCE IN COEFFICIENT ESTIMATES α, β AND γ ACROSS DATABASES FOR OTHER COUNTRIES INCLUDED IN THE SAMPLE

	CA	DS	EF	GV	OS	TF
Austria						
DS	α***,β***,γ*					
EF		α***				
GV	β**	α***,β***	β***			
OS	α**,β**	α***	α**	β***		
TF	α***,β***	α***,β****	α***,β***	α**	α*,β***	
WS	α***,β***	α***,β***	α***,β***	α**	α*,β***	
Belgium						
DS						
EF	α***					
GV	α***,β***	α***,β***	α***,β***,γ*			
OS	β***	β**	α***,β***	α***		
TF	α***		β*	α***,β***,γ*	α***,β***	
WS	α***			α***,β***,γ*	α***,β***	
Denmark						
DS						
EF		α**				
GV	α***	α***	α***,β***,γ***			
OS	α**,γ*	α***	α***,β***,γ***	α***,β**,γ**		
TF	α***	α***	α***,β**,γ**	α***,β***	α**,γ**	
WS	α***	α***	α***,β**,γ**	α***,β***	α**,γ**	
Finland						
DS	α*,β***,γ*					
EF	α**	α***,β**				
GV	α***,β***	α**	α***,β***			
OS	β***	α**	α**,β***	α***		
TF	α***	α***,β***		α***,β***	α***,β***	
WS	α***	α***,β***		α***,β***	α***,β***	
France						
DS						
EF	α***,γ***	α***,γ***				
GV	α***,β***,γ***	α***,β***,γ***	α***,β***,γ**			
OS	α***,β***	α***,β***	β***,γ***	α***,β***,γ***		
TF	α***,γ***	α***,γ***	α*,β*	α***,β***,γ*	β***,γ***	
WS	α***,γ***	α***,γ**	α*,β*	α***,β***	β***,γ***	
Greece						
DS	α**,β**,γ*					
EF	α**,β**	α**,β***				
GV	α**,β*,γ*	α**,β***				
OS	α**,β*	α***,β***	α**,β*			
TF	α**,β**	α***,β***,γ*	α**		α**	
WS	α**,β**	α***,β***	α***	α**,β***	α***	α***,β**,γ*
Ireland						
DS	α*					
EF	α**					
GV	α***,β*		α*			

APPENDIX A

(CONTINUED)

	CA	DS	EF	GV	OS	TF
OS						
TF	α^{***}					
WS	α^{***}					
Italy						
DS	α^{***}					
EF	β^{*}	α^{***},β^{***}				
GV	β^{***}	α^{***},β^{***}	β^{***}			
OS	α^{*}		α^{***}	α^{***},β^{**}		
TF	β^{***},γ^{*}	α^{***},β^{***}			α^{***}	
WS	β^{***}	α^{***},β^{***}			α^{***}	
Portugal						
DS	$\alpha^{***},\beta^{***},\gamma^{***}$					
EF	α^{**},β^{*}	β^{***},γ^{***}				
GV		$\alpha^{***},\beta^{***},\gamma^{***}$	α^{***},β^{***}			
OS	α^{***}	β^{**},γ^{**}	α^{**}	α^{***}		
TF	$\alpha^{***},\gamma^{***}$	β^{***},γ^{***}	γ^{***}	$\alpha^{***},\beta^{***},\gamma^{***}$	α^{**},γ^{*}	
WS	$\alpha^{***},\beta^{*},\gamma^{***}$	β^{***},γ^{***}	γ^{***}	$\alpha^{***},\beta^{***},\gamma^{***}$	α^{**},γ^{*}	
Spain						
DS	α^{***},γ^{**}					
EF	$\alpha^{**},\beta^{**},\gamma^{***}$	α^{***},β^{*}				
GV		α^{***}	α^{***},γ^{**}			
OS	γ^{*}	α^{***}	α^{**}			
TF	$\alpha^{**},\beta^{*},\gamma^{***}$	α^{***}		$\alpha^{***},\gamma^{***}$	α^{*}	
WS	$\alpha^{**},\beta^{**},\gamma^{***}$	α^{***},β^{*}		$\alpha^{***},\gamma^{***}$	α^{**}	
Sweden						
DS	$\alpha^{***},\beta^{***},\gamma^{***}$					
EF	α^{***},β^{**}	$\alpha^{***},\beta^{***},\gamma^{***}$				
GV	$\alpha^{***},\beta^{***},\gamma^{***}$	β^{**}	$\alpha^{***},\beta^{***},\gamma^{***}$			
TF	$\alpha^{***},\beta^{***},\gamma^{**}$	$\alpha^{***},\beta^{***},\gamma^{***}$	γ^{*}	$\alpha^{***},\beta^{***},\gamma^{***}$	α^{***},β^{***}	
WS	$\alpha^{***},\beta^{***},\gamma^{**}$	$\alpha^{***},\beta^{***},\gamma^{***}$		$\alpha^{***},\beta^{***},\gamma^{***}$	α^{***},β^{***}	
The Netherlands						
DS	$\alpha^{*},\beta^{***},\gamma^{***}$					
EF	α^{***}	α^{***},β^{**}				
GV	$\alpha^{**},\beta^{***},\gamma^{**}$		α^{***}			
OS	β^{***}	γ^{*}	α^{***},β^{**}	α^{*},β^{*}		
TF	α^{***}	$\alpha^{***},\beta^{***},\gamma^{***}$		α^{***},β^{*}	α^{***},β^{***}	
WS	α^{***}	$\alpha^{***},\beta^{***},\gamma^{***}$		α^{***},β^{*}	α^{***},β^{***}	

$$Model\text{: } P_t \quad \alpha + \beta BVPS_t + \gamma EPS_t + e_t,$$

where P is the market share price, $BVPS$ is the book value of shareholders' equity per share, EPS is bottom line earnings per share and t is the time period indicator.

All t statistics are heteroskedasticity consistent. We estimate the statistical significance of the differences between coefficients in regressions using different databases as $\theta_1 - \theta_2$ divided by $\sqrt{\sigma_1^2 + \sigma_2^2}$, where θ_i is the estimated coefficient and σ_i the standard error for variable i. ***, ** and * indicate that the difference is significant at a 1%, 5% and 10% level. The following abbreviations have been used: CA Company Analysis; DS Datastream; EF Extel Financials; GV Global Vantage; OS BvD Osiris; TF Thomson Financial; WS Worldscope.

APPENDIX B

VARIABLE DEFINITION BY DATABASE

	Company Analysis	Datastream	Extel Financials	Global Vantage	BvD Osiris	Thomson Financial	Worldscope
BVPS (book value per share)	[{eq} {eq.s.ps}]/ {mkt.nsh}	DS305/ DSIC	(EX.ShareholdersEquity EX.ShareholdersEquityPreferShare)/EX.Common SharesOutstanding	(G193–G268)/ CSHO	Shareholders Funds (35902)/No of outstanding shares (28)	TF.TotalCommonEquity/ TF.CommonShares Outstanding	WS.TotalCommonEquity (03501)/WS.Common SharesOutstanding (05301)
P (share price)	{mkt.shpc}	DSMV/ DSIC	EX.PriceClose	Prccm	Mark Cap (35907)/ No of outstanding shares (28)	TF.YrEndMarketCap/ TF.CommonShares Outstanding	WS.PriceClose (05001*)
EPS (earnings per share)	{ni}/ {mkt.nsh}	(DS625 + DS193)/ DSIC	EX.NetIncome/EX.Com monSharesOutstanding	(G378 + G381)/ CSHO	Net Income (35904)/ No of outstanding shares (28)	TF.NetIncome/ TF.CommonShares Outstanding	WS.NetIncome (01751)/ WS.CommonShares Outstanding (05301)

Note: We also define EPS for Thomson Financial as (TF.NetIncome + TF.ExtraItemsAndGnLsSaleofAsset)/TF.CommonSharesOutstanding, and for Worldscope as (WS.NetIncome + WS.ExtraItemsAndGnLsSaleofAsset)/WS.CommonSharesOutstanding. The reason for using these additional definitions for EPS in these two databases is that it is not clear in the database manual definitions whether or not TF.NetIncome and WS.NetIncome include extraordinary items. Regardless of the definition used, the results do not change qualitatively.