SFB 649 Discussion Paper 2005-064

Worldscope meets Compustat: A Comparison of Financial Databases

Niels Ulbricht*
Christian Weiner*

* School of Business and Economics,
Humboldt-Universität zu Berlin,
Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Worldscope meets Compustat: A Comparison of Financial Databases[*]

Niels Ulbricht[†]
Christian Weiner[‡]

December 14, 2005

Abstract

With this study we are the first to systematically compare today's two major counterparts as a source of accounting and financial data for researchers: Compustat North America by Standard & Poor's and Worldscope by Thomson Financial. This investigation is conducted for U.S. and partly Canadian data over an extensive period from 1985 to 2003. We examine more than 650 data items available in both databases and address the question of whether or not the decision for one or the other source may have an impact on the outcome of research projects. It is probably commonly assumed that this impact is minor, but it also leaves room to question certain results. We show that the use of both databases should lead to comparable results, but also find that if, e.g. a size bias, is not treated with care the quality of results may differ considerable. Furthermore after 1998 the number of firms covered by Worldscope exceeds the one covered by Compustat by about one fourth.

---

[*] All remaining errors are our own. We would like to thank Ernst Maug and Ingolf Dittmann for helpful discussions and comments. We gratefully acknowledge support by the Rudolf von Bennigsen-Foerder foundation and the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

[†] Humboldt-Universität zu Berlin, School of Business and Economics, Spandauer Str. 1, 10178 Berlin, Germany; e-mail: ulbricht@wiwi.hu-berlin.de

[‡] Humboldt-Universität zu Berlin, School of Business and Economics, Spandauer Str. 1, 10178 Berlin, Germany; e-mail: weiner@wiwi.hu-berlin.de

1 Introduction

In recent years empirical questions and analyses are getting more and more attention within the fields of financial and accounting research. More than 70% of papers in leading financial and accounting journals are build on or backed up by empirical research. A significant number of these papers require accounting data to perform investigations. These underlying data deserve considerable attention, as the validity and power of the results rely on a well prepared dataset. In this study we investigate whether the choice of the data source may has a considerable impact on the outcome of an empirical research project. Specifically, we compare two competing data sources for financial and accounting data that are commonly used among researchers: Compustat North America by Standard & Poor's and Worldscope by Thomson Financial.

According to Standard & Poor's Compustat North America is the most complete database of U.S. and Canadian accounting data with 10,000 actively traded U.S. firms, 10,900 inactive firms as well as 1,100 Canadian firms. The history covers 20 years. According to Thomson Financial Worldscope provides 19,000 U.S. and international firms with a history beginning in 1980. About 5,500 firms are located in the U.S. and Canada. Beside these two sources there are several other data providers for accounting, financial and market data, e.g. Value Line Incorporated. The ValueLine database contains fundamental data (both current and historical) on more than 7,500 publicly traded North American, European, and Asian firms. It includes hundreds of items on each firm, with balance sheet and income statement data. Datastream also by Thomson Financial contains accounting and especially market data for the U.S. and numerous other countries, although regarding accounting data the number of available firms and the overall coverage is below that of Worldscope. Reuters Fundamentals from Reuters has a content of over 26,000 firms in more than 70 countries. 9,000 firms are from the U.S. with detailed balance sheet and income statement data. CRPS (Center for Research in Security Prices) provides market data for U.S. firms[1].

---

[1] Information are gathered for Compustat from www.standardandpoors.com, for Worldscope and Datastream from www.thomsonfinancial.com, for Value Line from www.valueline.com, for Reuters Fundamentals from www.reuters.com and for CRSP from gsbwww.uchicago.edu/research/crsp/.

In empirical research the database is one major source for questionable results due to an underlying selection distortion. This leads to the motivation for this study. The problem of whether or not results might have been influenced by the choice of using Compustat or Worldscope, has not yet been addressed. In order to examine this question we will describe both databases and point out considerable differences that we are able to identify. This is driven by the interest to determine the scope of an advantage to choose one or the other data source. Because of financial and time constraints research projects generally do not have the resources to use and match both data sources as their empirical basis.

How is it possible that the derived datasets differ when either Compustat or Worldscope is used? First, Standard & Poor's and Thomson Financial do not only rely on the documents disclosed by companies, like annual or quarterly reports, they also use contact to firm insiders, e.g. investor relation teams to infer more detailed information. Second, Thomson Financial and also Standard & Poor's established its own standard as to how certain accounting items are reported in their system in order to ensure comparability among data for different companies. Third, there are differences as to which and how many firms are included in either database. Although the discrepancy in data coverage has decreased considerably as both data providers broadened their firm base. For analyses of longer time series this aspect is still important. The reason for the difference in data coverage can most probably be explained by the origin of both data providers. In contrast to Standard & Poor's, Worldscope was originally developed by fund managers who wanted to systematically store accounting information of potential investments. This for instance may explain why Worldscope especially in early years suffers from a size effect. More important, more interesting and better visible firms, i.e. large firms, were added to the database first. While Worldscope now seems to be committed to add historic data to their database, the effect is still noticeable.

This study is limited to United States and partly Canadian data from 1985 to 2003. The reasons for these limitations are: first, we have only access to Compustat North America and second, multiple country aspects would let the scope of necessary analysis explode. The uniqueness of this study comes to a large extend from a profound understanding and knowledge of both databases. Most researchers focus on Compustat while Worldscope is typically covered by research projects that focus on international companies.

To name only the main difficulties: the codes to access data items are different, some data items may be stored in arrays and cannot be identified directly, firms and securities are distinguished via different methods, both databases use their own jargon or terminology in handbooks and access software. We will show that in most cases it would indeed not be worth the effort to work with both data sources as only minor changes of the results can be expected. Yet, there are conditions in which a researcher should a priori decide for one or the other data source.

The results of the study can be summarized in three major points. First, we show that for time periods after 1997 Worldscope covers considerably more firms than Compustat. This makes it the first choice for studies concerned with recent periods. On the other side Compustat shows a clear advantage in the number of firms covered before 1997. Therefore projects relying on long time series should use Compustat. Second, the overall distribution of variables over all years is significantly different between the two databases, while the distribution is similar within one year. Third, valuation of firms with a multiple approach leads to better predictors for Worldscope for all years between 1994 and 2002.

The rest of this study is organized as follows. Section 2 will introduce the relevant literature. Since this topic has not yet been addressed by other studies we give an overview of studies that examine different financial and accounting databases. In the third section of this paper we give detailed descriptive statistics for both datasets. This analysis contains the data coverage also with respect to certain quality requirements. Furthermore, we present detailed statistical descriptions for the most relevant accounting data items. This section also evaluates the usability of each dataset for time series purposes. Next, we investigate the data quality of accounting information in both databases. Finally, this section concludes with a statement as to the frequency of data errors, e.g. typos. The fourth section compares data of both datasets on firm levels which one would suspect to be equal. The fifth section presents results of a typical research question in finance and accounting. We value firms based on multiples of both datasets and compare the outcomes. Finally, the sixth section concludes.

2 Related Literature

Despite the widespread use of financial databases and the high relevance of accurate and reliable data in financial research, there exist only few studies that examine or compare data as well as databases. No paper covers Worldscope and Compustat together but there are some papers that compare Compustat with Value Line and CRSP, respectively. The reason is that Compustat or CRSP are used in about 95% of the studies that require accounting data, while only about 5% of the papers use Worldscope[2]. If studies are based only on U.S. firms then almost nobody uses Worldscope.

Papers examining accounting or financial data sources usually choose a very limited perspective. Some papers concentrate on reviewing the number of observations covered. While others point out one or two specific phenomena or errors of a specific data source. We aim to take a look from a broader perspective, by also including statistics that describe and compare the useability of data items with regard to standard empirical applications in finance, e.g. multiple valuation, DCF valuation or time series analysis.

One paper that compares two accounting databases is Kern and Morris (1994). They only focus on two variables. They present differences and similarities of the Compustat and ValueLine databases based on total assets and sales information. The mean differences of these two variables increase significantly from 1971 to 1990. The error tolerance used for this calculation is $10,000. This could lead to distorted results because the effect is larger for small than for large firms. They also present differences in effective tax rates to show variation.

---

[2] This is based on the analysis of 5 journals from 1995 to 2004.

Bennin (1980) compares CRSP and Compustat. He shows that Compustat contains price information that are as reliable as CRSP prices. This paper is from 1980, and we would expect that there are significant changes in data collection and processing, so that these results have little implications for today's research.

While there are few papers that compare databases, some papers observe that data are distorted or biased. One paper that links data requirements to a real research question is from Villalonga (2004a). She compares Compustat segment data with the Business Information Tracking Series (BITS) from the U.S. Bureau of the Census and shows that different data sources and a different level of detail have a large impact. A similar result comes from Schoar (2002). She shows that Compustat data produce different results in favor of a diversification discount.

3 Database Structure and Descriptives

3.1 Database Structure

We have developed two datasets of an equal data structure for the analysis, one with Worldscope data and one with Compustat data. Both datasets include only information on firm levels. Since Worldscope and Compustat also contain information on security levels it was the first step to clearly distinguish between these two data classes. In Worldscope all firms related entries are identified by the variable Perm ID (06105), where the last digit is zero in case of an entry on firm level or it is any other number in case of a security level entry. In Compustat entries are identified by their CUSIP number which serves as key for both security and firm level data. CUSIP represents the national identification number for firms from the United States and Canada. Furthermore, there is a "Global Venture Key" (GVKEY) which serves as a key for firm entries only.

Both databases distinguish between data items that have time series characteristics, i.e. change over time (total assets, net income, number of shares, monthly closing price), and items that relate to the current state and are assumed to stay relatively constant over time (name, address, country). The first category of variables is gathered with a certain frequency, i.e annual, monthly, weekly or daily. This type of information is hereafter referred to as time series variables. The second category of variables, hereafter referred to as static variables, are called current items in Worldscope and scalar items in Compustat. In total Worldscope covers 1,566 variables with information on either firm or security level that are either static or time series related. Compustat has a total of 1,307 variables[3], including I/B/E/S and ACE (Analysts' Consensus Estimates) data which was not available to me, since it requires additional licensing. This number also includes so called concepts, which are not actual data variables but stored formulas to derive information based on actual data, e.g. average five-year sales growth.

Table 1 gives an overview of available variables in Worldscope and Compustat grouped by category. For this study we regard only static and annual time series variables for firms domiciled in the U.S. or Canada, i.e. firms that are only traded on a stock exchange in theses countries. ADRs are disregarded. The Worldscope dataset is as of October 18th, 2004 and the Compustat data as of March 29th 2005. The time range of the study is limited by the earliest year with Compustat data available, which is 1985 and by the last year that is fully included in the Worldscope source, which is 2003. The Worldscope dataset contains information of 15,998 firms over the range from 1985 to 2003 and the Compustat dataset contains a total of 20,630 firms. Over the same range we have 146,154 firm-years in Worldscope and 170,607 observations in the Compustat dataset. In Table 2 the number of firm observations is expressed by year.

In the Worldscope database we can identify 151,173 firm-years with a single identifier and a valid year variable for the United States. For Canada we have 15,755 observations. From Compustat we get 167,179 firm-years with a CUSIP and a year variable for the Unites States and 17,159 for Canada. Table 2 displays the number of firms per year for both databases for the United States and Canada as well as the difference between the two databases.

Worldscope starts with a set of 2,945 firms in 1985 and increases the number of

---

[3] This number was derived from files in the /CSPRMPTS/NEW installation directory.

Table 1: General information

This table displays most important information about the Woldscope and Compustat datasets used for this analysis. Development information are based on the Worldscope Datatype Definition Guide and the Standard & Poor's Research Insight North America Data Guide, respectively. Additional information are based on our own research.

description	Worldscope	Compustat
# variables - total	1,566	1,307
# variables - timeseries	1,284	1,004
# variables - static	282	303
history, general	1980 - 2005	1985 - 2005
history, segment data	1980 - 2005	1998 - 2005
country coverage	56 countries	U.S., Canada

The Compustat time series figure includes a total of 213 I/B/E/S and ACE data items which need a special subscription.

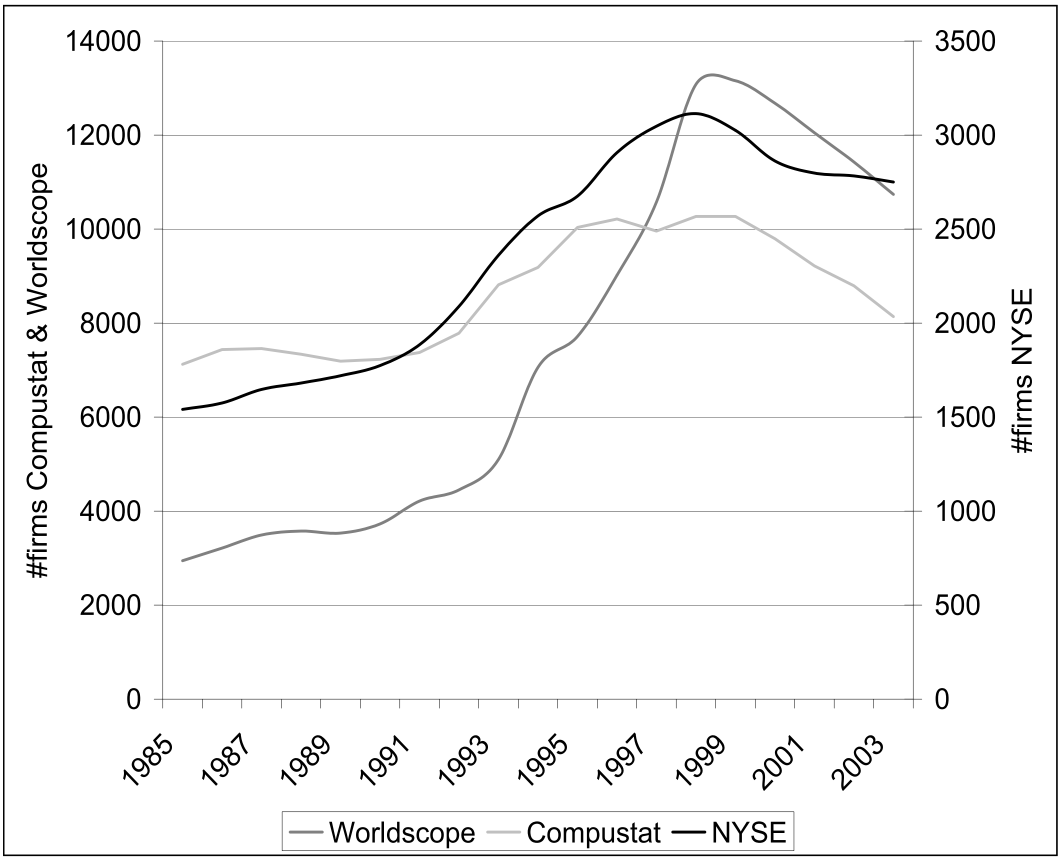
firms to a maximum of 13,156 firms in 1999 which is about 4 times higher. Between 2000 and 2003 the quantity declines to 10,738 firms. The development of the number of firms in Worldscope between 1985 and 2003 is coherent with market developments. Figure 1 displays the distribution of firms in Worldscope, Compustat and the number of firms that are listed at the New York Stock Exchange. We take this information as a proxy for newly listed firms in North America and Canada. The left y-axis shows the number of firms in both databases the right y-axis the number of listed firms at the NYSE. Especially between 1992 and 1999 a lot of new firms have been listed at the stock exchange. Between 2000 and 2003 Worldscope as well as NYSE cover significantly fewer firms. For Canada the distribution is not displayed but it looks similar. In 2004 we have not all data available , so the data give only an indication how fast new data updates are available. Compustat shows a different distribution. It starts with a high number of 7,124 firms and rise to 10,269 in 1998 which is an

Table 2: Availability of data

This table displays the number of observations available in Woldscope and Compustat that we can identify by Perm ID and year and by CUSIP and year, respectively. Data are separated by United States and Canada and year. The last two columns show the difference between the number of Compustat and Worldscope observations.

year	Worldscope		Compustat		Difference	
	U.S.	Canada	U.S.	Canada	U.S.	Canada
1985	2945	408	7124	375	4179	-33
1986	3213	447	7436	432	4223	-15
1987	3491	474	7458	495	3967	21
1988	3575	476	7337	511	3762	35
1989	3532	476	7190	517	3658	41
1990	3727	482	7228	519	3501	37
1991	4213	489	7374	530	3161	41
1992	4452	499	7786	539	3334	40
1993	5100	529	8815	692	3715	163
1994	7060	581	9185	812	2125	231
1995	7718	638	10036	918	2318	280
1996	9020	676	10214	952	1194	276
1997	10575	706	9958	993	-617	287
1998	13080	1145	10269	1184	-2811	39
1999	13156	1300	10268	1361	-2888	61
2000	12677	1343	9794	1420	-2883	77
2001	12047	1337	9215	1421	-2832	84
2002	11427	1308	8794	1457	-2633	149
2003	10738	1254	8137	1410	-2601	156
2004	9427	1187	3561	621	-5866	-566
avg per year	7559	788	8359	858	800	70

Figure 1: Number of firms



increase of about 50%. This pattern differs from the Worldscope database. In the following years the number of firms declines. For Canadian firms the number increases by about 400% from the minimum to the maximum. The difference between the two databases starts with 4,179 additional firms for Compustat. Then we see a sharp reduction, which leads to a balanced result in 1996 and 1997. After this, the number of firms in Worldscope passes the number in Compustat. In 2003 the difference is 2,601. For Canada the number of firms is always higher except in 1985, 1986 and 2004. The main findings are that new listings at the stock exchange can explain a large fraction of the number of firms over time. The correlation coefficients are higher than 90%.

3.2 Descriptive Statistics

This section will discuss statistical properties of both datasets in the following regards. First, we group firms by the level of detail for which data are available. Second, we present statistics for the most frequently analyzed accounting items. Third, we compare the data availability for time series analysis and comment on the frequency of typing errors.

Firms grouped by the level of detail

From a practical perspective the plain number of observations does not describe the usability of a dataset very well. It is also important that for a given firm and year sufficient data items are available to analyze or use an economic model. We examine the level of detail of the available information for two practical research scenarios. The first scenario captures the usability given that one intends to perform a firm valuation based on multiples. The second observes the data situation that one is interested in a discounted cash-flow (DCF) or residual income (RI) model. Since a multiple valuation is less demanding in terms of required accounting information, we will call all firms with sufficient data being of "basic quality" and all firms with sufficient accounting information for a DCF or RI model being of "high quality".

For the basic quality category we require the following items to be available, information in parentheses specifies the corresponding Compustat and Worldscope item: total assets (at, 02999), total liabilities (lt, 03351) sales (sale, 01001), net income (ni, 01551), EBIT (ebit, 18191), EBITDA (oibdp, 18198), SIC code (sich, 19506).

For the high quality category we require that each firm-year observation completely covers the additional accounting items: current assets (act, 02201), current liabilities (lct, 03101), net property plant and equipment (ppent, 02501), depreciation, depletion and amortization (dp, 01151), taxes (txt, 01451), dividends (dvc, 04551), pre-tax income (pi, 01401), long-term debt (dltt, 03251), minority interest on balance sheet (mib, 03426). The following table 3 presents the number of observations for the basic and high quality category for each year from 1985 to 2003. The total number of observations for each year is derived by requiring at least either sales or total assets to be available.

In total Compustat contains a considerably higher number of firms up to 1997, while from 1998 Worldscope's coverage is on average about 25 percent broader than that of Compustat. The data quality seems to be an advantage of Compustat. On average 77 percent of Compustat observations have sufficient information for the basic category, while only 54 percent of the Worldscope observations fulfill the same requirement. For the high quality category the picture looks about the same, 64 percent of Compustat and only 42 percent of Worldscope observations enter each category. Although after 1998 the absolute number of observations is always higher in the Worldscope database. This should make Worldscope to be the first choice for studies that focus on this more recent period.

Statistics for key accounting items

Table 4 displays the general dimension of variables in each database. To do so, we select 14 representative variables from the balance sheet, income and cash flow statement that are commonly used for research purposes. The definitions for each variable based on Worldscope's and Compustat's data guides can be found in Appendix A. In this table we show statistics for all firms and information available from 1985 to 2004. The first line of every data item refers to Worldscope data the second to Compustat data.

For net sales (Worldscope code: 01001, Compustat code: sale) we identify 135,697 and 165,997 firm-year information, respectively. The lowest sales values are about -$70,660,000 for Worldscope and -$20,370,000 for Compustat. We assume that these negative values are based on returns and discounts. The 25th-percentile is close to zero, which gives an indication of the number of small firms in either database. The mean of net sales differs significantly between the two databases. It can also be seen that there is a large difference between mean and also median values, which indicates a different firm size structure in both datasets. The standard deviation supports these findings. The difference between the two databases is significant. Figure 2 to figure 4 support this finding. We plot the distribution of the natural logarithm of

Table 3: Level of detail of accounting information

This table displays the number of available firm-year observations. Basic quality refers to the requirements for a standard multiples valuation, while high quality refers to the level of detail of accounting information to conduct a standard discounted cash-flow or residual income analysis. The last column presents the total number of observations in the dataset. The last row shows averages.

year	Compustat Quality					Worldscope Quality				
	basic		high		total	basic		high		total
1985	1	(0%)	0	(0%)	7348	3	(0%)	3	(0%)	3353
1986	8	(0%)	4	(0%)	7690	11	(0%)	9	(0%)	3660
1987	7146	(92%)	5967	(77%)	7761	91	(2%)	78	(2%)	3935
1988	6982	(91%)	5730	(75%)	7636	419	(10%)	254	(6%)	4020
1989	6858	(92%)	5596	(75%)	7471	560	(14%)	338	(9%)	3971
1990	6859	(92%)	5629	(75%)	7485	1430	(35%)	1097	(27%)	4112
1991	6961	(91%)	5740	(75%)	7618	3196	(69%)	2520	(55%)	4603
1992	7319	(92%)	6059	(76%)	7967	3581	(74%)	2849	(59%)	4846
1993	7773	(85%)	6397	(70%)	9094	4214	(77%)	3094	(56%)	5497
1994	8210	(86%)	6789	(71%)	9509	5383	(79%)	4127	(60%)	6850
1995	9021	(87%)	7535	(73%)	10355	5974	(79%)	4592	(61%)	7520
1996	9130	(87%)	7693	(73%)	10489	6819	(78%)	5201	(59%)	8759
1997	8862	(86%)	7482	(73%)	10275	7724	(75%)	5900	(57%)	10276
1998	9038	(85%)	7673	(72%)	10660	9924	(75%)	7929	(60%)	13179
1999	9018	(84%)	7666	(71%)	10794	9967	(74%)	7941	(59%)	13392
2000	8466	(81%)	7180	(69%)	10395	9235	(71%)	7300	(56%)	12952
2001	7786	(79%)	6668	(68%)	9811	8812	(71%)	7065	(57%)	12371
2002	7262	(77%)	6211	(66%)	9438	8428	(72%)	6750	(57%)	11768
2003	6647	(75%)	5737	(65%)	8811	7664	(69%)	6126	(55%)	11090
avg	7018	(77%)	5882	(64%)	8979	4918	(54%)	3851	(42%)	7692

Table 4: Coverage of main variables

This table displays the descriptive statistics of several important variables. The first line refers to Worldscope data the second line to Compustat data. Values are in 10,000,000. We cover all available values from 1985 to 2004. Sales is net sales, cogs is cost of goods sold, dda is depreciation and amortization, oi is operating income, ni is net income, cash is cash and equivalents, inv is inventories, ppeg is gross property, plant and equipment, ppen is net property, plant and equipment, ta is total assets, wc is working capital, debt is total debt, eqty is common equity, capex is capital expenditure. N refers to the number of firms, p25 and p75 are quartiles, std is the standard deviation. Significance of means is based on the parametric t-test. Significance of medians is based the non-parametric Wilcoxon rank sum test.

var	n	min	P25	mean	median	P75	max	std
sales	135697	-70.66	2.20	142.88	11.93	58.87	25632.90	637.68
	165996	-20.37	1.18	129.83^{c}	7.12^{c}	43.04	26398.90	651.62
cogs	103696	-48.38	1.17	99.41	7.83	39.77	19489.50	488.60
	165928	-36.66	0.68	88.94^{b}	4.15^{c}	26.88	20340.30	492.33
dda	112541	-32.00	0.11	9.27	0.58	3.15	2995.85	52.35
	159279	-0.79	0.04	7.48^{c}	0.25^{c}	1.72	3230.12	47.28
oi	134392	-1340.11	-0.01	13.78	0.79	5.20	3061.18	74.05
	163309	-0.27	0.00	20.13^{c}	0.30^{c}	3.19	37299.84	326.50
ni	134465	-5612.19	-0.10	6.13	0.36	2.58	4485.13	52.69
	146382	-4.05	0.00	17.95^{c}	0.00^{c}	1.22	19932.50	191.88
cash	113966	-0.92	0.17	16.67	1.03	4.87	17815.75	152.45
	166521	-0.59	0.10	30.92^{c}	0.67^{c}	3.79	43465.51	388.64
inv	108941	-58.59	0.01	15.33	0.91	6.25	4914.90	72.00
	103846	-0.57	0.00	0.74^{c}	0.01^{c}	0.12	284.33	5.81
ppeg	104931	0.00	0.88	132.78	5.76	38.18	32008.27	662.67
	165780	-920.74	-0.05	15.56^{c}	0.44^{c}	4.05	5854.00	100.35
ppen	130047	-188.98	0.37	65.53	2.45	18.12	12806.34	316.45
	165904	-8572.66	-0.15	9.69^{c}	0.21^{c}	2.57	4201.70	76.92
ta	134789	0.00	4.10	365.27	19.29	94.14	126403.20	2602.35
	167179	0.00	1.60	322.44^{b}	10.33^{c}	64.79	148410.10	2622.55
wc	104824	-8351.60	0.10	11.48	2.14	9.05	4499.90	113.21
	165511	-4267.50	0.00	2.19^{c}	2.73^{c}	3.80	10365.00	40.02
debt	134237	0.00	0.16	124.65	2.65	21.81	807021.41	3114.05
	165851	0.00	0.09	99.53^{c}	1.35^{c}	16.03	96173.20	1091.08
eqty	134541	-122459.72	1.34	62.36	6.07	26.98	22425.68	451.06
	166667	-2229.50	0.44	55.85^{c}	3.34^{c}	18.57	22423.43	299.86
capex	133157	-23.68	0.06	10.77	0.47	3.16	3317.67	60.99
	150562	-9.90	0.03	10.30	0.27^{c}	2.21	6502.80	67.31

c, b, a indicates significance at 1%, 5%, 10% level net sales for both databases for a period from 1985 to 2003 and for the years 1985 and 2002 separately. The first plot is based on 135,697 Worldscope sales values and 165,996 Compustat sales values. As pointed out before both plots show that the Worldscope database has a considerable size effect in comparison to the Compustat database. This effect diminishes over the years and is virtually not noticeable after 2002 (plots between 1986 and 2001 are not shown here, but do support this statement). Cost of goods sold, total assets and capital expenditure also show similar distributions but with significant differences in means and medians. Depreciation, depletion and amortization and common equity show a strong difference in the minimum values while the other figures are similar. Operating income and net income show high differences for minimum and maximum values. The standard deviation is also larger for Compustat. Cash and equivalents are similar for most figures except the maximum value and the standard deviation. Inventories and property, plant and equipment are completely different. Working capital is similar for medians but different in the tails of the distribution. Debt is similar for means and medians but different for maximum values. In general, the number of firm-years differs between 30,000 and 60,000, which can be seen as one reason for different values.

In table 4 we could show that the overall data coverage is different between the two databases. Therefore, we now compare the selected variables for each year separately. We present two representative years in table 5. We suppress detailed descriptive statistics and display only means, medians and standard deviations as well as the number of observations.

In 1997 we find a similar amount of information for net sales, operating income, net property, plant and equipment, total assets, total debt, common equity and capital expenditure. The other variables deviate by about 10% to 25%. In 2002 cost of goods sold, cash and equivalents, gross property, plant and equipment, working capital can be seen more often in Worldscope, while the other variables have more values in Compustat. The difference between the number of observations is between 5% and 20%. Compared to the overall distribution the values per year are more similar. In 1997 net sales, cost of goods sold, total assets, total debt and capital expenditure

Figure 2: Distribution of log sales - 1985 to 2003

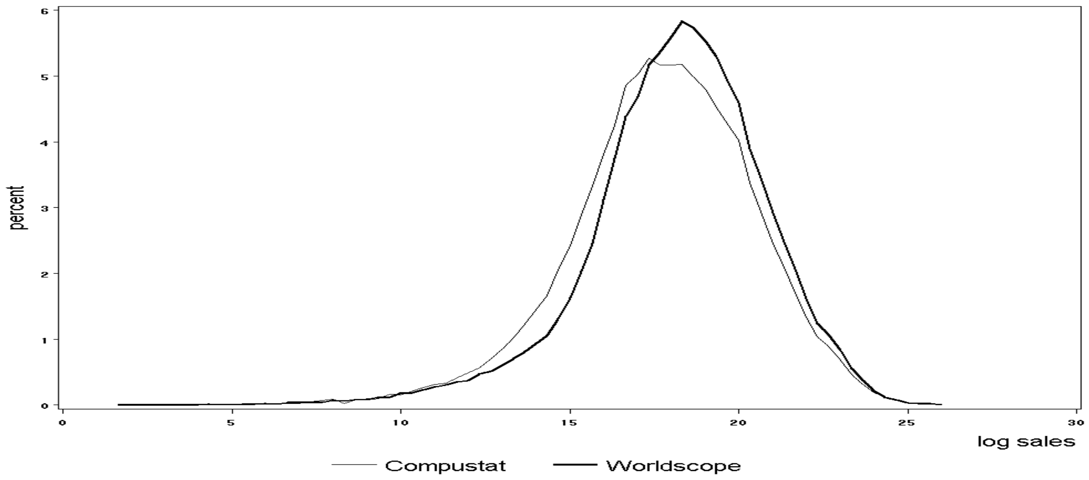


Figure 3: Distribution of log sales - 1985

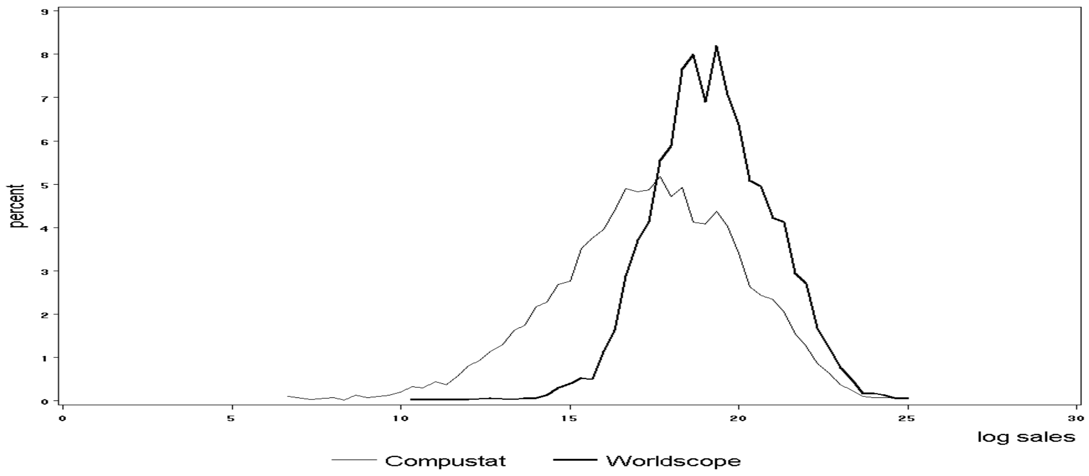


Figure 4: Distribution of log sales - 2002

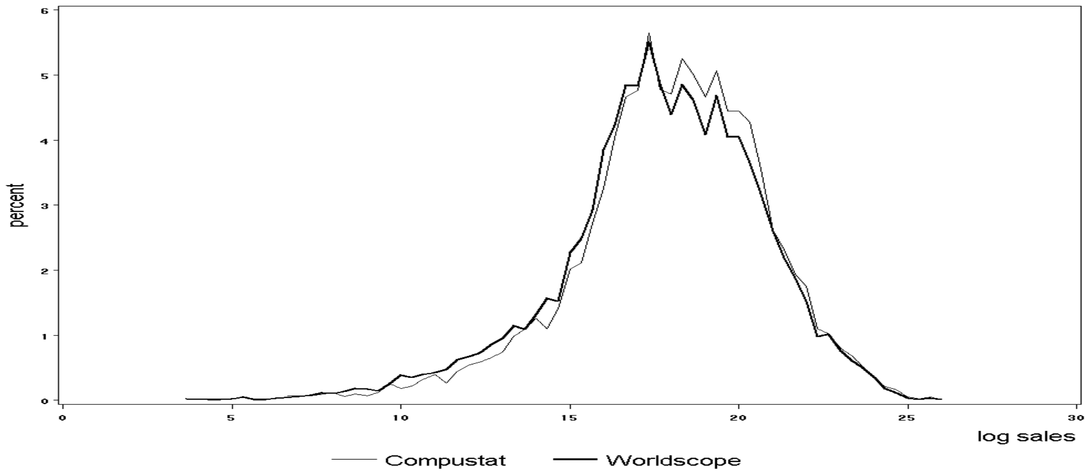


Table 5: Yearly coverage of main variables

This table compares the selected variables for two representative years. The first line refers to Worldscope data the second line to Compustat data. Values are in 10,000,000. We cover all available values from 1997 and 2002. Sales is net sales, cogs is cost of goods sold, dda is depreciation and amortization, oi is operating income, ni is net income, cash is cash and equivalents, inv is inventories, ppeg is gross property, plant and equipment, ppen is net property, plant and equipment, ta is total assets, wc is working capital, debt is total debt, eqty is common equity, capex is capital expenditure. N refers to the number of firms, p25 and p75 are quartiles, std is the standard deviation. Significance of means is based on the parametric t-test. Significance of medians is based the non-parametric Wilcoxon rank sum test.

	1997				2002			
var	n	mean	median	std	n	mean	median	std
sales	9923	127.06	8.13^{b}	615.13	8761	192.96^{a}	10.27^{c}	896.50
	9730	136.59	8.72	617.09	9240	168.43	7.32	787.56
cogs	9906	85.95	4.70	459.54	8760	130.51^{b}	5.57^{c}	674.87
	7419	96.19	6.43	473.14	7395	110.21	4.61	575.81
dda	9615	7.12^{b}	0.28	42.84	8470	12.36	0.46	70.34
	8019	8.66	0.45	48.37	7839	11.95	0.49	67.25
oi	9819	20.23^{a}	0.30^{a}	314.99	8697	34.50^{c}	0.26^{c}	524.67
	9654	14.02	0.70	67.46	9217	15.82	0.23	93.37
ni	8627	12.71^{c}	0.02	84.45	8470	41.79^{c}	0.24^{c}	326.78
	9697	7.22	0.32	37.37	9207	2.54	0.04	94.11
cash	9935	29.00^{b}	0.90^{c}	339.94	8792	59.61^{c}	1.59^{c}	662.03
	8002	18.16	1.06	237.05	7927	24.27	0.95	202.34
inv	5968	0.69^{c}	0.02^{c}	5.38	5245	0.96^{c}	0.01^{c}	7.03
	7447	15.02	0.59	69.21	7725	15.15	0.20	75.33
ppeg	9900	16.54^{c}	0.67^{c}	95.95	8757	23.43^{c}	0.51^{c}	151.49
	7088	132.64	4.46	650.57	7319	160.89	3.55	827.91
ppen	9905	11.31^{c}	0.37^{c}	63.19	8762	8.64^{c}	0.14^{c}	131.35
	9022	61.44	1.70	297.58	9044	78.64	1.39	376.76
ta	9958	303.46	12.23^{c}	2208.37	8794	618.87	20.51^{c}	4580.20
	9629	355.22	16.58	2364.87	9252	521.15	16.91	3863.57
wc	9882	2.04^{c}	2.95^{c}	10.61	8739	1.39^{c}	1.96^{c}	29.08
	7048	12.34	2.17	94.32	7358	11.48	0.81	124.20
debt	9891	92.40	1.48^{c}	896.98	8764	201.79	2.28^{c}	2045.86
	9536	111.20	2.00	986.31	9225	170.30	1.47	1693.92
eqty	9925	52.18^{b}	4.25^{c}	217.24	8769	88.98	5.15^{b}	451.86
	9629	59.57	5.04	233.72	9220	81.58	4.48	408.02
capex	8904	11.21	0.35^{b}	72.56	7829	14.54^{b}	0.34^{c}	84.63
	9525	10.80	0.36	63.91	9174	11.77	0.22	69.62

c, b, a indicates significance at 1%, 5%, 10% level show no significant difference in means. Cost of goods sold, depreciation, depletion and amortization and net income are not different in medians. The other values are significantly different but this could come from a different dataset size. Only operating income, inventories, gross and net property, plant and equipment as well as working capital show high differences in standard deviations. Especially property, plant and equipment are formed by a combination of several other variables, which could lead to these spreads between Worldscope and Compustat values. In 2002 the dispersion increases, while the number of comparable datasets deviates on average by only 10%. For all variables one can see an increase in standard deviations compared to 1997. This automatically leads to higher differences. On the other hand the number of firms stays almost constant. For depreciation, depletion and amortization, total assets, total debt and common equity the means are not significantly different. For the other variables we see highly significant differences. Summarized, the differences between the two databases are significant but given the fact that the number of firms is not equal we see similar distributions in general.

Usability for times series analysis

Many economic models are built on the analysis of time series data. In order to compare the usability of the Worldscope and Compustat database for this purpose we investigate for a set of key variables the number of firms with a complete history dating back to a certain year.

This information is presented in table 6 to table 8. The results can be read as follows: if one is interested in a time series of EBIT data starting in 2003 with a length of five years, then one can work with 7,035 firms if one relies on Compustat data and with 6,897 firms if one works with data based on Worldscope. These values can be found in the row of the year 1999.

Considering times series of up to seven years Worldscope shows slightly more available firms although not for all variables but for the important ones, like sales, total assets, etc. In contrast, Compustat is clearly better for time series analysis of eight or more years.

Table 6: Usability for time series analysis I

This table displays the number of firms with complete time series beginning in 2003 and dating back to the specified year for the following variables: current assets (CA), depreciation depletion and amortization (DDA), dividends (Div), EBIT (EBIT), and EBITDA (EBITDA). cs refers to Compustat data and ws to Worldscope data.

year	CA		DDA		Div		EBIT		EBITDA	
	cs	ws	cs	ws	cs	ws	cs	ws	cs	ws
2003	7192	6795	8433	7273	8592	8472	8737	8221	8500	7913
2002	7059	6755	8250	7200	8457	8411	8549	8057	8315	7701
2001	6581	6655	7688	7079	7981	8288	7995	7797	7750	7422
2000	6227	6411	7233	6844	7534	8033	7540	7360	7287	6984
1999	5791	6029	6723	6482	7026	7591	7035	6897	6775	6526
1998	5256	5403	6046	5849	6330	6890	6339	6155	6093	5807
1997	4663	3769	5354	4262	5622	5195	5624	4561	5394	4259
1996	4448	3139	5102	3476	5360	4268	5368	3775	5144	3500
1995	4131	2703	4756	2974	5004	3699	5013	3261	4796	3032
1994	3544	2385	4115	2635	4344	3288	4343	2889	4155	2682
1993	3203	1729	3703	1951	3932	2535	3934	2254	3735	2062
1992	2920	1589	3149	1713	3373	2087	3376	1871	3204	1717
1991	2667	1479	2885	1581	3093	1946	3093	1755	2936	1605
1990	2469	1312	2687	1382	2885	1642	2889	1506	2738	1377
1989	2306	1249	2517	1316	2708	1572	2709	1439	2567	1314
1988	2178	1207	2387	1268	2565	1513	2564	1381	2431	1258
1987	2089	1091	2290	1135	2458	1320	2454	1212	2326	1110
1986	1970	985	2146	1008	2303	1169	2294	1070	2179	976
1985	1807	939	1978	964	2116	1106	2113	1013	2011	928

Table 7: Usability for time series analysis II

This table displays the number of firms with complete time series beginning in 2003 and dating back to the specified year for the following variables: long-term debt (LTD), minority interest on (balance sheet) (MIB), net income (NI), property plant and equipment net (PPE), and pretax income (PI). cs refers to Compustat data and ws to Worldscope data.

year	LTD		MIB		NI		PPE		PI	
	cs	ws	cs	ws	cs	ws	cs	ws	cs	ws
2003	8773	8537	8298	8477	8746	8651	8555	8409	8747	8625
2002	8619	8492	8084	8405	8560	8613	8404	8334	8561	8593
2001	8058	8368	7478	8242	8006	8515	7855	8161	8007	8494
2000	7612	8076	6985	7886	7548	8285	7419	7792	7551	8262
1999	7106	7665	6421	7446	7044	7906	6909	7295	7046	7877
1998	6406	6940	5692	6709	6348	7160	6216	6553	6350	7138
1997	5678	5197	4968	4957	5635	5387	5497	4822	5635	5376
1996	5419	4368	4678	4173	5374	4465	5243	4062	5375	4455
1995	5059	3744	4292	3615	5022	3781	4891	3521	5023	3777
1994	4393	3327	3635	3205	4356	3365	4244	3111	4357	3358
1993	3986	2579	3224	2463	3947	2609	3854	2382	3945	2604
1992	3404	2123	2682	2017	3410	2151	3315	2003	3386	2149
1991	3136	1984	2416	1879	3129	2006	3048	1867	3106	2003
1990	2921	1682	2203	1584	2920	1692	2841	1590	2899	1694
1989	2748	1613	2022	1506	2740	1620	2665	1517	2721	1622
1988	2615	1547	1896	1442	2594	1563	2528	1458	2577	1564
1987	2499	1346	1767	1250	2467	1353	2421	1268	2465	1361
1986	2349	1195	1641	1078	2307	1195	2276	1129	2305	1203
1985	2145	1129	1476	998	2122	1131	2091	1070	2122	1139

Table 8: Usability for time series analysis III

This table displays the number of firms with complete time series beginning in 2003 and dating back to the specified year for the following variables: sales (sales), total assets (TA), and total liabilities (TL). cs refers to Compustat data and ws to Worldscope data.

year	sales		assets		TL	
	cs	ws	cs	ws	cs	ws
2003	8747	8624	8782	8633	8773	8549
2002	8562	8580	8635	8596	8619	8516
2001	8013	8471	8079	8485	8058	8398
2000	7563	8235	7637	8221	7607	8095
1999	7057	7854	7133	7814	7100	7665
1998	6359	7103	6437	7065	6401	6911
1997	5644	5366	5713	5330	5675	5124
1996	5386	4448	5455	4457	5418	4321
1995	5035	3770	5098	3788	5059	3755
1994	4367	3356	4425	3368	4394	3339
1993	3957	2605	4021	2612	3990	2587
1992	3398	2152	3464	2148	3431	2130
1991	3119	2007	3189	2005	3162	1987
1990	2914	1700	2974	1699	2948	1681
1989	2736	1636	2796	1628	2773	1601
1988	2591	1577	2662	1563	2641	1540
1987	2483	1384	2525	1360	2505	1340
1986	2322	1231	2375	1203	2355	1185
1985	2136	1160	2171	1140	2152	1121

4 Database Matching

The maximal number of data items that contain information for one single firm in the Worldscope dataset is 618, for Compustat it is 925. For some variables historical information is not available. We call these static variables, while all others are time series variables.

The matching between two databases is one of the critical problems in empirical research. The only clear reference between the two databases is the CUSIP number of each company. Because this is the main identifier of the Compustat database it is always available for Compustat. In Worldscope the CUSIP is a common static variable. It is available for 98.3% of the observations. Table 9 presents the number of possible matches between the two databases and the matching quality for selected variables. We only use the CUSIP as well as the name to identify possible matches. Other identifier like SEDOL or ISIN on the other hand show a poor performance. The first three columns show the number of firms that can be identified in Worldscope but have no counterpart in Compustat, the number of firms that are available in both databases and can be matched as well as the number of firms that are only available in Compustat.

In 1986 the total number of observations is 7,903. 466 (5.90%) are available in Worldscope but have no counterpart in Compustat or it is not possible to identify the counterpart. 2,241 (28.36%) companies are available in both databases and have one CUSIP identifier or the same name. 5,195 observations are available in Compustat but not in Worldscope or it is not possible to identify the counterpart. Between 1986 and 1993 Compustat contains many more firms than Worldscope, which leads to a matching quantity below 40%. From 1994 to 2003 the number of firms in both databases is similar. The number of successful matches increases from 50% to 65% in 2003. The second result that is shown in the table is the matching success of important accounting variables. All information are based on successfully matched observations. We report the number of firms that have the same value for the variable considered. Due to rounding problems we allow an absolute deviation between two values of 0.5%. In 1986 we detect 1,959 firms that have the same values for total assets. This is the only variable where the number of matches decreases significantly between 1986 and 2003. The variables capital expenditure and net sales have a constant value of about 80%. Cash and equivalents has a constant level of 85%. Common equity declines from 81.08% in 1986 to 71.00% in 1999 and then increase to 78.77%. Net income has a constant low level between 55% and 65%. Overall one can see that the matching quality is high for the important variables.

5 Multiple Valuation

This paper not only presents descriptive information about the two databases. We will also document the performance for one typical research question in finance and accounting. The approach is based on a multiple valuation procedure. The result is a comparison between the estimated enterprise value and the observed enterprise value. We use two independent datasets from each database with the same restrictions. We do not require that a company has to be available in both datasets.

This method calculates the enterprise value to earnings multiple for firm k as followed:

$$\frac{EV_k}{E_k} = \frac{\text{market capitalization}_k + \text{total debt}_k - \text{cash \& short-term investments}_k}{\text{EBIT}_k - \text{non-operating interest income}_k}. \tag{1}$$

Market capitalization is calculated by the number of shares outstanding multiplied by the unadjusted share price. EBIT refers to earnings before interest and taxes. We subtract cash and short term investments as well as the corresponding revenue figure to get a multiple that represents the operating activities of a firm.

The assumption is that there exists a local linear relationship between the value of the firm and the multiple that can be written as:

$$EV_{k,t} = \beta_{k,t} E_{k,t} + \epsilon_{k,t}. \tag{2}$$

Here $EV_{k,t}$ is the enterprise value of the firm k at time t, E refers to earnings of firm k, ß is the estimated multiple of peer group G for firm k, where each comparable

Table 9: Data matching

This table displays the yearly matching quality between Worldscope and Compustat. The first three columns show the matching quality between the two databases. In 1986 466 firms are available in Worldscope but not in Compustat or the identification is not possible, 5,195 are available in Compustat but not in Worldscope or cannot be identified, 2,241 are available in both databases. Based on the firms that could be matched we describe the matching quality for important variables. Columns ta, ce, ca, eq, ni, sa count the number of equal values for assets, capital expenditure, cash, common equity, net income and net sales. The second line of each year shows the percentage matching quality. We allow an absolute deviation of 0.5%.

	Database Matching			Variable Matching					
year	ws	match	cs	ta	ce	ca	eq	ni	sa
1986	466	2241	5195	1959	1761	1922	1817	1291	1802
	5.90	28.36	65.74	87.42	78.58	85.77	81.08	57.61	80.41
1987	473	2433	5025	2131	1950	2111	1983	1350	1959
	5.96	30.68	63.36	87.59	80.15	86.77	81.50	55.49	80.52
1988	466	2630	4707	2308	2135	2249	2165	1458	2106
	5.97	33.70	60.32	87.76	81.18	85.51	82.32	55.44	80.08
1989	446	2618	4572	2297	2162	2236	2141	1555	2069
	5.84	34.28	59.87	87.74	82.58	85.41	81.78	59.40	79.03
1990	429	2645	4583	2313	2206	2255	2167	1591	2121
	5.60	34.54	59.85	87.45	83.40	85.26	81.93	60.15	80.19
1991	673	2947	4427	2552	2483	2537	2398	1746	2375
	8.36	36.62	55.01	86.60	84.26	86.09	81.37	59.25	80.59
1992	682	3118	4668	2554	2626	2680	2506	1581	2504
	8.05	36.82	55.13	81.91	84.22	85.95	80.37	50.71	80.31
1993	685	3726	5089	2879	3193	3222	2976	1823	3020
	7.21	39.22	53.57	77.27	85.70	86.47	79.87	48.93	81.05
1994	1167	5166	4019	3749	4226	4374	4022	3099	4014
	11.27	49.90	38.82	72.57	81.80	84.67	77.86	59.99	77.70
1995	1250	5719	4317	4119	4738	4865	4429	3717	4484
	11.08	50.67	38.25	72.02	82.85	85.07	77.44	64.99	78.41
1996	1636	6565	3649	4630	5434	5609	5090	4294	5155
	13.81	55.40	30.79	70.53	82.77	85.44	77.53	65.41	78.52
1997	2301	7193	2765	5062	6011	6158	5529	4554	5685
	18.77	58.68	22.55	70.37	83.57	85.61	76.87	63.31	79.04
1998	2760	8714	1555	5847	7164	7164	6205	5369	6908
	21.18	66.88	11.93	67.10	82.21	82.21	71.21	61.61	79.27
1999	2690	8654	1614	5559	7130	7013	6144	5275	6885
	20.76	66.78	12.46	64.24	82.39	81.04	71.00	60.95	79.56
2000	2543	8176	1618	5454	6723	6892	6008	4951	6500
	20.61	66.27	13.12	66.71	82.23	84.30	73.48	60.56	79.50
2001	2260	7680	1535	5013	6313	6484	5627	4677	6149
	19.69	66.93	13.38	65.27	82.20	84.43	73.27	60.90	80.07
2002	1944	7088	1706	4711	5757	6012	5299	4146	5555
	18.10	66.01	15.89	66.46	81.22	84.82	74.76	58.49	78.37
2003	1686	6439	1698	4424	5283	5526	5072	4104	5102
	17.16	65.55	17.29	68.71	82.05	85.82	78.77	63.74	79.24

firm can be identified by the index j. Then ß at time t is defined by the enterprise value to earnings multiple

$$\beta_{k,t} = \text{median}_{j \in G} \left(\frac{EV_{j,t}}{E_{j,t}} \right). \tag{3}$$

Here ϵ_k is the absolute percentage estimation error. For firm k at time t it is defined through

$$\epsilon_{k,t} = \left| \frac{EV_{k,t} - \beta_{k,t} * E_{k,t}}{EV_{k,t}} \right|. \tag{4}$$

We have several limitations that we apply to both datasets. The raw dataset contains all firms from the United States and Canada that are available between 1994 and 2002. All required accounting data, market values and industry membership data are available for year t. The fiscal year end is the calendar year. The variables cash and short-term investments, non-operating interest income as well as total debt are set to zero if data are not available. The market price is chosen from the last trading day of April in year t+1. The firm has only one type of stocks. Total assets have to be higher than 1,000,000. EBIT has to be positive and higher than 0.1. An SIC-code has to be available over time, if not, a static SIC-code is used. After elimination through the restriction 26,205 firm-year observations for Worldscope and 20,567 firm-year observations for Compustat remain in the sample. Table 10 shows the descriptive statistics for the sample.

The enterprise value is larger for Compustat, while total assets are larger for the Worldscope database. Sales, total debt and EBIT are almost equal for mean, median and standard deviation for both databases. The enterprise value to EBIT ratio shows a much higher mean for Worldscope, while the median is similar but higher for Compustat.

The results of the approach are displayed in table 11. We present the mean, median and the standard deviation of valuation errors for each year and database and compare the results.

It turns out that there are significant differences between the two databases. In 1995 and 1996 the number of firms is similar. In the following years the number

Table 10: Descriptive statistics

This table displays the descriptive statistics of the datasets for the valuation procedure. Data are collected for a period from 1994 to 2002. Enterprise value (ev) is the sum of market capitalization and total debt. EBIT is earnings before interest and tax. Values are in millions.

variable	Worldscope			Compustat		
	mean	median	std	mean	median	std
ev	2910.86	171.66	17173.43	3455.21	204.00	21336.95
total assets	4437.36	195.60	31360.38	4239.91	158.34	32477.98
sales	1596.47	99.27	7203.12	1547.62	92.24	7464.46
total debt	1388.83	22.71	12917.01	1311.47	19.51	13658.49
EBIT	185.02	7.06	1131.96	191.41	6.44	1223.06
ev/EBIT	21.99	8.08	4180.06	11.64	10.17	1656.26

of firms in Worldscope increases while in Compustat it decreases. The mean and the standard deviation show that there are some extreme values in the Worldscope database. This is coherent with the descriptive statistics, where the average enterprise value to EBIT ratio is much higher for Worldscope than for Compustat. Therefore, we concentrate on median values. One can see a similar pattern over time but the median values in Worldscope are always below Compustat values. The difference decrease from 12% in 1995 to 3% in 2002. In 1999 both databases show the highest valuation error with 45% and 51%, respectively. In the following years errors decrease and become more similar. The Wilcoxon rank sum test indicates that the difference of medians is significant in each year.

6 Conclusion

This paper documents that there are some similarities but also significant differences between the Compustat and Worldscope databases. The overall conclusion is that

Table 11: Valuation results

This table displays mean, median and standard deviation (std) of valuation errors based on Woldscope and Compustat datasets from 1992 to 2002. N is the number of observations. The valuation error is defined as the percentage deviation between the estimated and the observed enterprise value. Estimated values are based on the median enterprise value to EBIT multiple from an industry peer group. Diff shows the difference of medians. Significance in medians is based on the non-parametric Wilcoxon rank sum test.

year	Worldscope				Compustat				diff
	n	mean	std	median	n	mean	std	median	
1994	1803	0.41	0.60	0.29	2425	0.59	1.18	0.41	0.11^c
1995	2513	0.51	1.42	0.32	2594	0.68	2.11	0.44	0.12^c
1996	2807	1.02	18.16	0.36	2714	0.74	5.21	0.43	0.07^c
1997	3128	0.83	9.65	0.42	2569	0.71	2.09	0.46	0.04^c
1998	3391	0.99	12.49	0.44	2331	0.74	1.98	0.48	0.04^b
1999	3331	1.02	9.81	0.45	2244	0.84	2.54	0.51	0.06^c
2000	3328	0.65	1.53	0.39	2092	0.70	1.80	0.43	0.04^c
2001	3095	0.99	17.72	0.38	1773	0.73	3.06	0.41	0.03^c
2002	2809	0.97	9.62	0.37	1738	0.69	5.14	0.40	0.03^c

c, b, a indicates significance at 1%, 5%, 10% level

for U.S. firms Worldscope is as competitive as Compustat even though nobody uses Worldscope for research with U.S. data. We find no statistical or methodological reason why Worldscope should not be used for research. The number of firms and the coverage of variables is similar. Applied to multiple valuation the results are different. One can see a significantly lower valuation error for Worldscope independently of the number of firms in the underlying dataset.

References

- Alford, Andrew W. (1992): The effect of the set of comparable firms on the accuracy of the price-earnings valuation method, Journal of Accounting Research, 30, pp. 94-108.

- Bennin, Robert (1980): Error rates in CRSP and Compustat: a second look, Journal of Finance, 35, pp. 1267-1271.

- Bhojraj, Sanjeev, and Charles M. C. Lee (2001): Who is my peer? A valuation-based approach to the selection of comparable firms, Journal of Accounting Research, 40, pp. 407-439.

- Bhojraj, Sanjeev, and Charles M. C. Lee / Derek K. Oler (2003). "What's my line? A comparison of industry classification schemes for capital market research." Journal of Accounting Research, 41, pp. 745-774.

- Clarke, Richard N. (1989): SICs as Delineators of Economic Markets, Journal of Business, 62, pp. 17-31.

- Fama, Eugene F., and Kenneth. R. French (1997): Industry costs of equity, Journal of Financial Economics, 43, pp. 153-193.

- Fan, Joseph P. H., and Larry H. P. Lang (2000): The measurement of relatedness: an application to corporate diversification, Journal of Business, 73, pp. 629-660.

- Guenther, David A., and Andrew J. Rosman (1994): Differences between Compustat and CRSP SIC codes and related effects on research, Journal of Accounting and Economics, 18, pp. 115-128.

- Kern, B. Beth, and Michael H. Morris (1994): Differences in the Compustat and expanded Value Line databases and the potential impact on empirical research, Accounting Review, 69, pp. 274-284.

- Ramnath, Sundaresh, Steven Rock, and Philip Shane (2001): Value Line and I/B/E/S earnings forecasts, working paper, Georgetown University.

- Schoar, Antoinette S. (2002): Effects of corporate diversification on productivity, Journal of Finance, 57, pp. 2379-2403.
- Villalonga, Belén (2004): Diversification discount or premium? new evidence from BITS establishment-level data, Journal of Finance, 59, pp. 479-506.

Appendix A: Data Definition

variable	Worldscope	Compustat
capex	Capital expenditure represents the funds used to acquire fixed assets other than those associated with acquisitions. (*04601*)	Capital expenditure represents cash outflow or the funds used for additions to the firm's property, plant and equipment. (*capx*)
cash	Cash and short-term investments represent the sum of cash and short-term investments. (*02001*)	Cash and equivalents represent cash and all securities readily transferable to cash. (*che*)
cogs	Cost of goods sold represents specific or direct manufacturing cost of material and labor entering in the production of finished goods. (*01051*)	Cost of goods sold represents all costs directly allocated by the company to production, such as material, labor and overhead. (*cogs*)
co. eqt.	Common equity represents common shareholders' investment in a company. (*03501*)	Common equity represents the common shareholders' interest in the company. (*ceq*)
depr.	Depreciation represents the process of allocating the cost of a depreciable asset , depletion refers to cost allocation for natural resources and amortization relates to cost allocation for intangible assets. (*01151*)	Depreciation and amortization represent non-cash charges for obsolescence of and wear and tear on property, allocation of the current portion of capitalized expenditures, and depletion charges. (*dp*)
inventories	Inventories represent tangible items or merchandise net of advances and obsolescence acquired for either resale directly or included in the production of finished goods manufactured for sale in the normal course of operation. (*02101*)	Inventories represent merchandise bought for resale and materials and supplies purchased for use in production of revenue. (*invt*)
net income	Net income represents the net income the company uses to calculate earnings per share. (*01751*)	Net income represents income or loss by a company after expenses and losses have been subtracted from revenues and gains for the fiscal period including extraordinary items and discontinued operations. (*ni*)

net sales	Net sales represents gross sales or other operating revenues less discounts, returns and allowances. (*01001*)	Net sales represents gross sales (the amount of actual billings to customers for regular sales completed during the period) reduced by cash discounts, trade discounts, and returned sales and allowances for which credit is given to customers. (*sale*)
op. income	Operating income represents the difference between sales and total operating expenses. (*01250*)	Operating income represents the operating income of a company after deducting expenses for cost of goods sold, selling, general, and administrative expenses, and depreciation. (*oiadp*)
pp&e gross	Gross property, plant and equipment represents tangible assets with an expected useful life of over one year which are expected to be used to produce goods for sale or for distribution of services. (*02301*)	Gross property, plant and equipment represents the cost of tangible fixed property used in the production of revenue. (*ppegt*)
pp&e net	Net property, plant and equipment represents gross property, plant and equipment less accumulated reserves for depreciation, depletion and amortization. (*02501*)	Net property, plant and equipment represents the cost, less accumulated depreciation, of tangible fixed property used in the production of revenue. (*ppent*)
total assets	Total assets represent the sum of total current assets, long-term receivables, investment in unconsolidated subsidiaries, other investments, net property, plant and equipment and other assets (*02999*)	Total assets represent current assets plus net property, plant, and equipment plus other non-current assets. (*at*)
total debt	Total debt represents all interest bearing and capitalized lease obligations. (*03255*)	Total debt represents debt obligations due more than one year from the company's balance sheet date, plus debt in current liabilities. (*dt*)
working cap.	Working capital represents the difference between current assets and current liabilities. (*03151*)	Working capital represents the difference between total current assets minus total current liabilities as reported on a company's balance sheet. (*wcap*)

Variable identification is in parentheses. Data are based on the Worldscope Datatype Definition Guide and the Standard & Poor's Research Insight North America Data Guide.

SFB 649 Discussion Paper Series

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Nonparametric Risk Management with Generalized Hyperbolic Distributions" by Ying Chen, Wolfgang Härdle and Seok-Oh Jeong, January 2005.

002 "Selecting Comparables for the Valuation of the European Firms" by Ingolf Dittmann and Christian Weiner, February 2005.

003 "Competitive Risk Sharing Contracts with One-sided Commitment" by Dirk Krueger and Harald Uhlig, February 2005.

004 "Value-at-Risk Calculations with Time Varying Copulae" by Enzo Giacomini and Wolfgang Härdle, February 2005.

005 "An Optimal Stopping Problem in a Diffusion-type Model with Delay" by Pavel V. Gapeev and Markus Reiß, February 2005.

006 "Conditional and Dynamic Convex Risk Measures" by Kai Detlefsen and Giacomo Scandolo, February 2005.

007 "Implied Trinomial Trees" by Pavel Čížek and Karel Komorád, February 2005.

008 "Stable Distributions" by Szymon Borak, Wolfgang Härdle and Rafal Weron, February 2005.

009 "Predicting Bankruptcy with Support Vector Machines" by Wolfgang Härdle, Rouslan A. Moro and Dorothea Schäfer, February 2005.

010 "Working with the XQC" by Wolfgang Härdle and Heiko Lehmann, February 2005.

011 "FFT Based Option Pricing" by Szymon Borak, Kai Detlefsen and Wolfgang Härdle, February 2005.

012 "Common Functional Implied Volatility Analysis" by Michal Benko and Wolfgang Härdle, February 2005.

013 "Nonparametric Productivity Analysis" by Wolfgang Härdle and Seok-Oh Jeong, March 2005.

014 "Are Eastern European Countries Catching Up? Time Series Evidence for Czech Republic, Hungary, and Poland" by Ralf Brüggemann and Carsten Trenkler, March 2005.

015 "Robust Estimation of Dimension Reduction Space" by Pavel Čížek and Wolfgang Härdle, March 2005.

016 "Common Functional Component Modelling" by Alois Kneip and Michal Benko, March 2005.

017 "A Two State Model for Noise-induced Resonance in Bistable Systems with Delay" by Markus Fischer and Peter Imkeller, March 2005.

018 "Yxilon – a Modular Open-source Statistical Programming Language" by Sigbert Klinke, Uwe Ziegenhagen and Yuval Guri, March 2005.

019 "Arbitrage-free Smoothing of the Implied Volatility Surface" by Matthias R. Fengler, March 2005.

020 "A Dynamic Semiparametric Factor Model for Implied Volatility String Dynamics" by Matthias R. Fengler, Wolfgang Härdle and Enno Mammen, March 2005.

021 "Dynamics of State Price Densities" by Wolfgang Härdle and Zdeněk Hlávka, March 2005.

022 "DSFM fitting of Implied Volatility Surfaces" by Szymon Borak, Matthias R. Fengler and Wolfgang Härdle, March 2005.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

HUMBOLDT-UNIVERSITÄT · ZU BERLIN ·

023 "Towards a Monthly Business Cycle Chronology for the Euro Area" by Emanuel Mönch and Harald Uhlig, April 2005.
024 "Modeling the FIBOR/EURIBOR Swap Term Structure: An Empirical Approach" by Oliver Blaskowitz, Helmut Herwartz and Gonzalo de Cadenas Santiago, April 2005.
025 "Duality Theory for Optimal Investments under Model Uncertainty" by Alexander Schied and Ching-Tang Wu, April 2005.
026 "Projection Pursuit For Exploratory Supervised Classification" by Eun-Kyung Lee, Dianne Cook, Sigbert Klinke and Thomas Lumley, May 2005.
027 "Money Demand and Macroeconomic Stability Revisited" by Andreas Schabert and Christian Stoltenberg, May 2005.
028 "A Market Basket Analysis Conducted with a Multivariate Logit Model" by Yasemin Boztuğ and Lutz Hildebrandt, May 2005.
029 "Utility Duality under Additional Information: Conditional Measures versus Filtration Enlargements" by Stefan Ankirchner, May 2005.
030 "The Shannon Information of Filtrations and the Additional Logarithmic Utility of Insiders" by Stefan Ankirchner, Steffen Dereich and Peter Imkeller, May 2005.
031 "Does Temporary Agency Work Provide a Stepping Stone to Regular Employment?" by Michael Kvasnicka, May 2005.
032 "Working Time as an Investment? – The Effects of Unpaid Overtime on Wages, Promotions and Layoffs" by Silke Anger, June 2005.
033 "Notes on an Endogenous Growth Model with two Capital Stocks II: The Stochastic Case" by Dirk Bethmann, June 2005.
034 "Skill Mismatch in Equilibrium Unemployment" by Ronald Bachmann, June 2005.
035 "Uncovered Interest Rate Parity and the Expectations Hypothesis of the Term Structure: Empirical Results for the U.S. and Europe" by Ralf Brüggemann and Helmut Lütkepohl, April 2005.
036 "Getting Used to Risks: Reference Dependence and Risk Inclusion" by Astrid Matthey, May 2005.
037 "New Evidence on the Puzzles. Results from Agnostic Identification on Monetary Policy and Exchange Rates." by Almuth Scholl and Harald Uhlig, July 2005.
038 "Discretisation of Stochastic Control Problems for Continuous Time Dynamics with Delay" by Markus Fischer and Markus Reiss, August 2005.
039 "What are the Effects of Fiscal Policy Shocks?" by Andrew Mountford and Harald Uhlig, July 2005.
040 "Optimal Sticky Prices under Rational Inattention" by Bartosz Maćkowiak and Mirko Wiederholt, July 2005.
041 "Fixed-Prize Tournaments versus First-Price Auctions in Innovation Contests" by Anja Schöttner, August 2005.
042 "Bank finance versus bond finance: what explains the differences between US and Europe?" by Fiorella De Fiore and Harald Uhlig, August 2005.
043 "On Local Times of Ranked Continuous Semimartingales; Application to Portfolio Generating Functions" by Raouf Ghomrasni, June 2005.
044 "A Software Framework for Data Based Analysis" by Markus Krätzig, August 2005.
045 "Labour Market Dynamics in Germany: Hirings, Separations, and Job-to-Job Transitions over the Business Cycle" by Ronald Bachmann, September 2005.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

046 "Paternal Uncertainty and the Economics of Mating, Marriage, and Parental Investment in Children" by Dirk Bethmann and Michael Kvasnicka, September 2005.

047 "Estimation and Testing for Varying Coeffcients in Additive Models with Marginal Integration " by Lijian Yang, Byeong U. Park, Lan Xue and Wolfgang Härdle, September 2005.

048 "Zeitarbeit in Deutschland: Trends und Perspektiven" by Michael C. Burda and Michael Kvasnicka, September 2005.

049 "Courtesy and Idleness: Gender Differences in Team Work and Team Competition" by Radosveta Ivanova-Stenzel and Dorothea Kübler, September 2005.

050 "Do Factor Shares Reflect Technology?" by Benjamin Bental and Domonique Demougin, September 2005.

051 "Optimal investments for risk- and ambiguity-averse preferences: A duality approach" by Alexander Schied, September 2005.

052 "Relational Contracts and Job Design" by Anja Schöttner, September 2005.

053 "Explicit characterization of the super-replication strategy in financial markets with partial transaction costs" by Imen Bentahar and Bruno Bouchard, October 2005.

054 "Aid Effectiveness and Limited Enforceable Conditionality" by Almuth Scholl, August 2005.

055 "Limited Enforceable International Loans, International Risk Sharing and Trade" by Almuth Scholl, August 2005.

056 "Stock Markets and Business Cycle Comovement in Germany before World War I: Evidence from Spectral Analysis" by Albrecht Ritschl and Martin Uebele, November 2005.

057 "An empirical test of theories of price valuation using a semiparametric approach, reference prices, and accounting for heterogeneity" by Yasemin Boztuğ and Lutz Hildebrandt, November 2005.

058 "Integrable e-lements for Statistics Education" by Wolfgang Härdle, Sigbert Klinke and Uwe Ziegenhagen, December 2005.

059 "What does the Bank of Japan do to East Asia?" by Bartosz Maćkowiak, December 2005.

060 "Portfolio Value at Risk Based On Independent Components Analysis" by Ying Chen, Wolfgang Härdle and Vladimir Spokoiny, December 2005.

061 "How much of the Macroeconomic Variation in Eastern Europe is Attributable to External Shocks?" by Bartosz Maćkowiak, December 2005.

062 "The Impact of Industry Classification Schemes on Financial Research" by Christian Weiner, December 2005.

063 "The Conglomerate Discount in Germany and the Relationship to Corporate Governance" by Christian Weiner, December 2005.

064 "Worldscope meets Compustat: A Comparison of Financial Databases" by Niels Ulbricht and Christian Weiner, December 2005.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".