CER-ETH – Center of Economic Research at ETH Zurich

On the Construction of Common Size, Value and Momentum Factors in International Stock Markets: A Guide with Applications

Peter S. Schmidt, Urs von Arx, Andreas Schrimpf, Alexander F. Wagner and Andreas Ziegler

Working Paper 11/141
February 2011

Economics Working Paper Series

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

On the Construction of Common Size, Value and Momentum Factors in International Stock Markets: A Guide with Applications

Peter S. Schmidt
University of Zurich

Urs von Arx
University of Zurich, ETH Zurich

Andreas Schrimpf
Aarhus University

Alexander F. Wagner[†]
University of Zurich

Andreas Ziegler
University of Zurich, ETH Zurich

This version: 4.1.2011

Abstract

Demand is growing for a better understanding of how assets are priced in countries outside of the U.S. While financial data are available for many firms world-wide, it is important to have a reliable and replicable method of constructing high-quality systematic risk factors from these data. This paper first documents that appropriately screened data from Thomson Reuters Datastream and Thomson Reuters Worldscope can be used to replicate closely not only U.S. market returns and the corresponding momentum risk factor (as existing work has suggested), but also the widely-used U.S. size and value risk factors. We then build novel pan-European and country-specific momentum, size, and value risk factors. By comparing our pan-European market returns and risk factors with their counterparts in the U.S., we find that they are astonishingly highly correlated. The factors we compute are made available to other researchers.

JEL classification: C89, G12, G15

Keywords: Risk factors; value; size; momentum; international equity markets; asset pricing anomalies

[*] The authors appreciate financial support from the Swiss National Science Foundation, the Commission for Technology and Innovation of the Swiss Confederation (CTI), the Swiss Finance Institute, and the NCCR FINRISK (project B1: Corporate Finance, Market Structure and the Theory of the Firm) as well as financial and infrastructural support from the Zurich Cantonal Bank (ZKB). We are grateful to Harald Lohre for useful comments and suggestions at the beginning of our project. Helpful research assistance of Katja Bächtold, Carolin Hecht, and Veronika Sharonova is gratefully acknowledged.

† Corresponding author. Swiss Finance Institute – University of Zurich, Department of Banking and Finance, Plattenstrasse 14, CH-8032 Zurich, Switzerland, e-mail: alexander.wagner@bf.uzh.ch

1. Introduction

Many path-breaking results in empirical finance have been established for U.S. data by the investigation of the well-known Center for Research in Security Prices (CRSP) and COMPUSTAT dataset. Very prominently, the empirical failure of the one-factor model based on the Capital Asset Pricing Model (CAPM) has been documented using these data. For example, Fama and French (1993) show that their three-factor model – consisting of the market, value, and size risk factors – explains the cross-section of stock returns better than the one-factor model. Although there is an ongoing discussion of what the economic mechanism is by which passive investing in value firms and those with a relatively small market capitalization earns high expected returns, it has become common to control for these three factors in a wide range of applications. Moreover, Jegadeesh and Titman (1993) show for the U.S. that stocks having performed well in the past twelve months perform significantly better in the next 3-12 months than stocks which have performed poorly in the past twelve months. The high returns to momentum strategies have been difficult to rationalize by standard models, including the Fama-French three-factor model. In applications, researchers frequently include a momentum factor when modeling expected returns.

Researchers and practitioners alike are increasingly eager to determine the existence or non-existence of these anomalies in markets outside of the U.S. as well. Sometimes, a specific market per se is interesting; moreover, some factors may be more important in some countries than in others due to specific characteristics of individual markets. In addition to allowing the study of anomalies in different contexts (thus providing tests for theories that have been developed to explain anomalies in the U.S.), international analyses can address the argument that anomalies observed in the U.S. market may simply be a manifestation of survivorship or data-snooping biases (Kothari et al., 1995; Lo and MacKinlay, 1990; MacKinlay, 1995). Moreover, to implement standard applications in empirical finance such as long-run event studies or portfolio analyses also in non-U.S. markets, the analyst requires reliable predictions for expected returns based on an asset pricing model. In sum, there is a considerable need in the research community for high-quality data and reliable risk factors in international markets.

Fortunately, significant progress has been made in recent years on these fronts.[1] Perhaps the main challenge so far, however, is that the data employed have come from different data sources. Some studies use proprietary, country-specific datasets which are in general inaccessible to other researchers while other studies compile datasets from various sources.[2] Besides occasional questions regarding the exact procedures used in the data quality assurance, the most important practical drawback is the lack of availability of the constructed risk factors for other researchers. We believe that this lack of public availability has constituted a severe obstacle for researchers in international asset pricing or corporate finance.

In this paper, we show how an internally consistent, replicable financial dataset for the U.S. and European countries can be constructed and used to produce the well-known risk factors according to Carhart (1997), including the market, value, size, and momentum risk factors. We use Thomson Reuters Datastream and Thomson Reuters Worldscope data. It is well-known that data from Thomson Reuters Datastream can be prone to errors. For example, Ince and Porter (2006) show that the momentum effect is not detectable by using these raw data for the U.S. To circumvent these problems, Ince and Porter (2006) suggest some corrections that allow them to

---

[1] Several studies have employed international data to study empirical asset pricing models. For example Bauer et al. (2010), Fama and French (1998), Griffin (2002), Heston et al. (1999), Hou et al. (2006), Leippold and Lohre (2009), and Rouwenhorst (1998) use international datasets. Other studies use datasets from specific countries. Examples of this type of studies include: Ammann and Steiner (2008) (Switzerland), Artmann et al. (2010) (Germany), Dimson et al. (2003), Gregory et al. (2009), Nagel (2001) (all three U.K.), Schrimpf et al. (2007) (Germany), Vaihekoski (2004) (Finland) and Ziegler et al. (2007) (Germany).

[2] Griffin (2002), for example, uses data from the Pacific-Basin Capital Markets database (Japan), Thomson Reuters Datastream (U.K. and Canada) and CRSP/COMPUSTAT (U.S.). Vaihekoski (2004) uses a dataset from the Department of Finance and Statistics, Swedish School of Economics and Business Administration. Schrimpf et al. (2007) and Ziegler et al. (2007) use a database maintained at Humboldt University, Berlin, Germany.

obtain similar results for momentum in the Thomson Reuters Datastream dataset. In this paper, we build upon their screens and further expand them. We then go beyond price-based risk factors and consider the book-to-market equity ratio (BE/ME) and size factors. For this purpose, we use the Thomson Reuters Worldscope dataset.

In a first step, we compare our Thomson Reuters Datastream and Thomson Reuters Worldscope market returns and risk factors for the U.S. with important benchmarks: The market returns and momentum (WML), size (SMB), and value (HML) risk factors obtained from CRSP/COMPUSTAT data, as available on the website of Kenneth French, from here on referred to as the FF data (according to Fama and French, 1993).[3] We find that our market returns and risk factors replicate the FF counterparts remarkably well. The reliability of our thoroughly screened Thomson Reuters Datastream and Thomson Reuters Worldscope dataset is strengthened by additional analyses for stock portfolios which are separately sorted on size, BE/ME, and momentum as well as jointly sorted on size-BE/ME and size-momentum.

In a second step, we analyze pan-European market returns and pan-European risk factors on the basis of our thoroughly adjusted financial data. All European OECD countries are considered in this analysis. These Thomson Reuters Datastream and Thomson Reuters Worldscope risk factors cannot be compared with corresponding benchmarks as no publicly available pan-European risk factors exist so far. However, we show that for single European markets our market returns correlate strongly with corresponding well-known representative market indexes.

Finally, we compare our pan-European market returns and risk factors from the Thomson Reuters Datastream and Thomson Reuters Worldscope dataset with the U.S. market returns and risk factors from both FF and Thomson Reuters Datastream and Thomson Reuters Worldscope. Our results show that the market returns as well as the HML and WML factors in both regions

[3] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

are highly correlated with each other. Only the correlation between the U.S. and the European SMB factor is less strong. These results are both novel and surprising. Existing work for single markets (e.g. U.K. and the U.S.), such as Fama and French (1998) and Griffin (2002), has documented significantly lower correlations than those we uncover for the two regions.

We are, of course, not the first to calculate European risk factors from Thomson Reuters Datastream and Thomson Reuters Worldscope data. For example, An (2010), Ang et al. (2008), Angelidis and Tessaromatis (2008), Hou et al. (2006), Leippold and Lohre (2009, 2010), Liew and Vassalou (2000), and Moerman (2005) compute risk factors for single European countries from these databases. Our contribution relative to this existing literature is twofold. First; we do not only focus on specific applications, but also on data quality and replicability. Besides describing in detail all the steps in our screening procedure, we document that with our thoroughly screened data several benchmarks of U.S. and European market returns and risk factors can be matched. Second, our novel pan-European and country-specific risk factors will be made available to other researchers to facilitate research in international asset pricing.

The paper proceeds as follows. Section 2 deals with the data preparation. Section 3 explains the general construction of the risk factors, discussing in particular the choice of appropriate breakpoints for portfolio formation. Section 4 presents the market returns and risk factors for the U.S. market. Section 5 turns to the corresponding results for the pan-European and individual European markets. Section 6 compares the market returns and risk factors between the U.S. and pan-European markets in detail. Section 7 concludes.

2. Data preparation

In this section we describe the different steps of data preparation that are necessary to achieve an adequate data quality for the construction of risk factors and portfolios.[4]

Like Ince and Porter (2006, p. 465), we use Thomson Reuters Datastream constituent lists to construct our dataset. Besides research lists, we also use dead lists, Thomson Reuters Worldscope lists and for certain countries specific lists provided by Thomson Reuters Datastream and Thomson Reuters Worldscope. The Thomson Reuters Worldscope dataset is in principle available from 1980 onwards, but, as noted by the data provider, "statistically significant company and data item representation is best represented from January 1985 forward" (Thomson Financial, 2007, p. 4). Thus, we use data from 1985 onwards.[5] We use the "dead lists" of companies that cease to exist (due to mergers, bankruptcy or other reasons) to control for survivorship bias and Thomson Reuters Worldscope lists and sometimes additional lists to get a population as large as possible.[6] The lists are provided in Appendix A1 (U.S.) and A.2 (Europe).

On the basis of this initial sample (47,747 unique U.S. firms and 43,005 unique European firms), we first sort out firms which are obviously not a member of our population of interest. To do this we use firm characteristics which are assumed to be constant over time, thus employing "static screens." Specifically, our first screening procedure is to keep major listings (MAJOR="Y"), stocks located in the domestic market (e.g. GEOGN="UNITED STATES", for the U.S. and likewise for European countries) and firms of the equity type (TYPE="EQ"). There

4 This section is quite detailed and intended as a guide for researchers working with the Thomson Reuters Datastream and Thomson Reuters Worldscope database. Thus, a reader more interested in risk factor construction may directly jump to Section 3.

5 Ulbricht and Weiner (2005, p. 12-16, fig. 2-4) find a difference in the firm size structures between the Thomson Reuters Worldscope and COMPUSTAT databases which "diminishes over the years and is virtually not noticeable after 2002". Since Thomson Reuters Worldscope was "originally developed by fund managers", "more interesting and better visible firms, i.e. large firms, were added to the database first" (Ulbricht and Weiner, 2005, p. 3).

6 Nonetheless, it is very likely that not all dead stocks are captured by the dead lists (Ince and Porter, 2006, p. 470, note that firms like Atlantic Richfield Co., GTE Corp. and Honeywell are not included in the dead stock lists), and not all remaining firms are captured by the other lists available on Thomson Reuters Datastream and Thomson Reuters Worldscope.

are different reasons why firms are excluded by the static screens: either the firms are not major listings (e.g. preferred shares), foreign stocks, additional listings (e.g. closed-end-funds, REITs, ADRs, etc.) or there are simply no data available.

After these static screens, 30,400 firms remain for the U.S. and 21,435 for Europe. For these firms, we then extract time series data from the database. The time series draws are separated into yearly data (Thomson Reuters Worldscope) and monthly data (Thomson Reuters Datastream). To break down the yearly information into a monthly frequency, we use the Thomson Reuters Worldscope fiscal year end information (Thomson Reuters Worldscope item 05350).[7]

For the correction of the monthly data we apply dynamic screens suggested by Ince and Porter (2006) as well as additional filters. Table 1 summarizes the employed screening procedures.

[Table 1 here]

Table 2 and Table 3 list the number of firms for different stages of the data preparation process as well as the actual employed number of firms in case of the value weighted factors for the U.S. and for Europe. From the 30,400 (21,435) firms that remain after the static screens, 15,241 (12,218) fulfil the minimum requirements of having at least one point in time with jointly a non-missing dscd code (DSCD), common equity (WC03501), fiscal year end (WC05350), number of shares (NOSH) and price (P). Of these, 14,203 (11,086) U.S. (European) firms pass the time

[7] Occasionally, the fiscal year entry (such as "12/1999") is missing, but at least one item of the actual Thomson Reuters Worldscope company-specific data is known. In such cases, to avoid losing these datapoints, we fill in the fiscal year information if the fiscal year information of either the preceding (e.g. "12/1998") or succeeding year (e.g. "12/2000") in the data is contained in the data. If fiscal year ends from the year before and after the missing fiscal year end information are known, but from a different month, we use the latest month (e.g. if the preceding fiscal year end is "12/1998" and the succeding fiscal year end is "09/2000" then we use "12/1999" as the fiscal year end for 1999).

series screens described in Table 1. In the end we use 13,343 (11,054) U.S. (European) firms to construct the value weighted market factor, 11,114 (9,462) firms to construct the SMB and HML factors and 11,654 (10,035) firms to construct the WML factor. All numbers are for unique firms over the whole time span.

The U.S. sample (with respect to the SMB and HML factors) starts with a little less than 2,000 firms in the early eighties, rises to a maximum of about 5,500 in the year 2000 and falls from then on steadily until about a bit more than 4,000 firms in 2008. The European sample (with respect to the SMB and HML factors) starts with less than 1,000 firms in 1987[8], rises to more than 4,000 firms in 1998 and then stays between 4,000 and 5,000 firms until 2008. The detailed listing of the evolution of the number of firms can also be seen in Table 2 and Table 3.

[Tables 2 and 3 here]

Some further issues cannot be fixed by the suggestions of Ince and Porter (2006), but are important for the present application. Most important, the exchange affiliation is only recorded for the current point in time. We choose to use all stocks which are available on Thomson Reuters Datastream and Thomson Reuters Worldscope, which means that there are not only NYSE-, AMEX- or NASDAQ-listed stocks in the U.S. sample. We note that this implies that our U.S. sample is drawn from a different population than the sample population described by Fama and French (1993). The alternative, using only firms listed on the NYSE, AMEX, or NASDAQ at the end of the sample period, would result in a sample suffering from survivorship bias.

There are two additional issues for European stocks, which either are not relevant or of

[8] Since most exchange rate series available on Thomson Reuters Datastream start in 1987 (or later), we do not calculate joint European SMB, HML and WML factors before 1987, because we cannot calculate returns denominated in one currency and also cannot express market capitalizations for value weighting in one joint currency.

minor relevance for U.S. stocks. First, the adoption of the Euro in January 2002 implies that there exist two currencies in all countries that switched to the Euro. Data of companies which are traded after January 2002 are all dominated in Euros, whereas data of companies which are delisted before January 2002 are denominated in the old currency of the respective country. This can easily be fixed. We use the fixed euro conversion rate and express all cash values (like size) in euro values.[9]

Second, for some European countries dividend data are obviously erroneous. We observe that for some companies dividends are of a magnitude of about ten times the actual price series, which means that screening procedures like S06 or S07 (see Table 1) result in unusually high returns of several hundered percents whenever dividend payments are distributed. A casual inspectation shows that sometimes dividend payments made later are a fraction of the unusually high dividends, which leads us to the conjecture that a decimal or other error occurred. In order to correct this issue, we apply the following procedure (see Table 1, screen S05): Whenever a dividend payment is observed that is greater than 50% of the adjusted price, we divide the Thomson Reuters Datastream dividend by a certain value.[10] We apply this screen also to the U.S. dataset, although this issue is not of practical relevance there.

---

[9] Note that this procedure leaves the returns unaffected. Since value weighted market returns are generated by weighting with lagged size, this transformation may have a noticeable effect on value weighted market returns (and other return series which use value weighting, such as the risk factors) if a significant number of companies exit the sample before the euro changeover. This effect will be stronger the closer the relation between average returns and size is.

[10] The problem of the unusually high dividends is especially severe for the following countries: Belgium, Greece, Italy, Luxembourg, Portugal, Slovakia, Spain and Turkey. It turns out that dividing by 10, 100 or 1000 works well. In the case of Greece, Iceland, Italy and Turkey whenever a dividend payment is observed that is greater than 50% of the adjusted price, we divide dividends by 1000, in the case of Belgium, Czech Republic, Greece, Ireland, the Netherlands, Portugal, Slovakia, Spain, the U.K. and the U.S. we divide dividends by 100, in the case of Luxembourg we divide dividends by 30 and in the case of Austria, Denmark, Finland, France, Germany, Norway, Poland, Sweden and Switzerland we divide dividends by 10.

3. Risk factors

This section describes the constuction of risk factors as proposed by Fama and French (1993) and Carhart (1997) (Section 3.1) as well as the calculation of breakpoints for the allocation of stocks to portfolios employed in this paper (Section 3.2).

3.1. Construction

Fama and French (1993) introduced risk factors based on individual stock characteristics. To gain market-wide factors from individual firm characteristics Fama and French (1993) sorted stocks on these characteristics and used the difference in portfolio returns between high rated and low rated stocks according to these characteristics. In particular, they proposed one factor based on the difference in portfolio returns between stocks with a small market capitalization and stocks with a big market capitalization (small-minus-big – SMB) and one factor based on the difference between stocks with a high book-to-market equity ratio and a low book-to-market equity ratio (high-minus-low – HML). This empirical model has become standard in the empirical asset pricing literature. Following the recipe of Fama and French (1993) other factors based on individual stock characteristics have been proposed in the literature, most notably the momentum factor proposed by Carhart (1997), which is based on the observation by Jegadeesh and Titman (1993) that stocks with a high past performance (winners) outperforms stocks with a low past performance (losers) in the next 3-12 months. This factor is based on the difference between winner and loser portfolios and is often referred to as WML (winners-minus-losers). We follow this method to construct the factors SMB, HML and WML. Our Thomson Reuters Datastream and Thomson Reuters Worldscope dataset of monthly observations begins in

December 1985 and ends in December 2008.[11] The return calculation is based on closing prices of the last trading day of each month. If a stock is not traded on the last trading day, the last valid trading price is used. The Thomson Reuters Datastream total return indices which we use for return calculation include dividends and account for stock splits.

We calculate book equity as Thomson Reuters Worldscope common equity (WC03501) plus deferred taxes (WC03263), if available. For all sorts we use only stocks with available book equity which is greater than zero. Size is either the Thomson Reuters Datastream market value (MV) or the product of the Thomson Reuters Datastream unadjusted price (UP) with the Thomson Reuters Datastream number of shares (NOSH). BE/ME for the sorting month June is calculated as book equity divided by size of the preceding December. We sort all stocks each June, beginning in 1984. To be included in the June sort of year τ a stock must have a positive book value and size available in December of the previous year τ-*1*. Furthermore, to calculate value weighted returns, a stock needs to have available size from the preceding month, a valid return, positive book value, as well as price and number of shares.

In order to construct the SMB and HML factors, all remaining stocks are sorted each December into three BE/ME groups (breakpoints are discussed in Section 3.2). Furthermore, we sort these stocks each June into two size groups. From the intersection of the two size groups, small (S) and big (B), and the three BE/ME groups, low (L), medium (M) and high (H), we form six portfolios, which are held for one year.[12] The six portfolios contain small size and low BE/ME stocks (S/L), small size and medium BE/ME stocks (S/M), small size and high BE/ME stocks (S/H), big size and low BE/ME stocks (B/L), big size and medium BE/ME stocks (B/M),

[11] Note that we therefore begin with the porfolio formation in June 1986 and with the calculation of return series in July 1986.

[12] When a stock is no longer available in our dataset we invest the share of this stock into the other stocks in the respective portfolio group according to the employed weighting scheme.

as well as big size and high BE/ME stocks (B/H). Panel A of Table 4 illustrates the sorting procedure.

[Table 4 here]

From the monthly value weighted returns of these six portfolios we construct the factors SMB and HML for month t as follows:

$$\mathrm{SMB}_t = \frac{\mathrm{r}_t^{\mathrm{S/L}} + \mathrm{r}_t^{\mathrm{S/M}} + \mathrm{r}_t^{\mathrm{S/H}}}{3} - \frac{\mathrm{r}_t^{\mathrm{B/L}} + \mathrm{r}_t^{\mathrm{B/M}} + \mathrm{r}_t^{\mathrm{B/H}}}{3}, \tag{1}$$

$$\mathrm{HML}_t = \frac{\mathrm{r}_t^{\mathrm{S/H}} + \mathrm{r}_t^{\mathrm{B/H}}}{2} - \frac{\mathrm{r}_t^{\mathrm{S/L}} + \mathrm{r}_t^{\mathrm{B/L}}}{2}. \tag{2}$$

$\mathrm{r}_t^{\mathrm{X/Y}}$ denotes the returns of a portfolio stocks belonging to size class X (either S or B) and BE/ME class Y (either H, M or L) in month t based on the portfolio formation in last June.

In order to construct the momentum factor, we first define our momentum measure. For each portfolio-formation month t-1 we calculate for each stock the mean return from month t-12 to month t-2 and use this mean return to compile three momentum groups. This sorting takes place every month. We also construct two size groups each month. To be included in the sort, the stock return has to be available in every month from t-12 to t-2 and size must be available in month t-1. From the intersection of the two size groups, i.e. small (S) and big (B), and the three momentum groups losers (L), medium (M) and winners (W), we form six portfolios. The six portfolios contain small size and loser momentum stocks (S/L), small size and medium momentum stocks (S/M), small size and winner momentum stocks (S/W), big size and loser momentum stocks (B/L), big size and medium momentum stocks (B/M), as well as big size and winner momentum stocks (B/W). The sorting procedure is also illustrated in panel B of Table 4.

We construct the factor WML for month *t* as the difference of the mean returns of the two winner portfolios minus the mean returns of the two loser portfolios:

$$WML_t = \frac{r_t^{S/W} + r_t^{B/W}}{2} - \frac{r_t^{S/L} + r_t^{B/L}}{2}. \qquad (3)$$

$r_t^{X/Z}$ denotes the returns of a portfolio stocks belonging to size class X (either S or B) and momentum class Z (either W, M or L) in month *t* based on the portfolio formation in month *t*-1.

3.2. *Choice of breakpoints*

In each of the above sorts, we need to choose breakpoints to divide portfolios. This issue is most relevant for the size breakpoints and arises to a lesser extent for the BE/ME and momentum sorts. With respect to size in the U.S., Fama and French (1993, p. 8) calculate breakpoints from the NYSE sample only, but apply the breakpoints to the whole sample of NYSE, AMEX, and NASDAQ stocks.[13] Unfortunately, it is impossible to separate the NYSE stocks in our sample from other stocks (at least not over the whole time span). Therefore, we use an approximation by using breakpoints calculated from the whole sample, but aiming to mirror the Fama and French (1993) NYSE breakpoints. By considering the number of firms in each of the six size-BE/ME portfolios reported on Kenneth French's website, we can calculate the average of the empirical breakpoints which separates small and big stocks in those portfolios. Panel A of Table 5 shows the corresponding results. The mean (median) of this breakpoint is the 0.81 (0.81) quantile for the period from 07/1986 to 12/2008. Furthermore, the minimum of this breakpoint is the 0.77 quantile and the maximum is the 0.84 quantile, which suggests that this breakpoint is quite stable over time. Therefore, we use in our application the 0.80 quantile as a breakpoint for the

[13] NYSE breakpoints are also freqently used by other researchers. For example: Ang and Chen (2002, p. 455), Adrian and Franzoni (2009, p. 540), and Hodgson et al. (2002, p. 625) calculate breakpoints from all NYSE stocks and sort all stocks on NYSE, AMEX and NASDAQ into portfolio groups according to the NYSE breakpoints. Campbell (1996, p. 316-317), Chen et al. (1986, p. 394-395), Cochrane (1996, p. 587) and Ferson and Harvey (1991, p. 391) use size portfolios constructed from NYSE stocks.

separation of small and big stocks. The empirical mean (median) FF breakpoints for the BE/ME portfolios are the 0.36 (0.36) and 0.70 (0.70) quantiles. For the seperation among the the three BE/ME groups we use the 0.30 respectively the 0.70 quantiles. The breakpoints actually used are reported in the "actual" column of Table 5. We do not use mean or median empirical breakpoints since the breakpoints we actually employ are more common in similar applications and are roughly close to the mean or median empirical breakpoints. We apply this approximation procedure to all portfolios involving size. Panel B of Table 5 shows the breakpoints implied by the FF data for the size-momentum sort into six portfolios.

[Table 5 here]

4. Results for the U.S. market

This section compares U.S. market returns and risk factors from our dataset with the corresponding series from Kenneth French's website (Sections 4.1 and 4.2). In addition, we investigate the quality of our dataset by comparing portfolio groups single and double sorted on characteristics from Kenneth French's dataset with ours (Section 4.3).

4.1. Market returns

Table 6 shows averages (avg.) and standard deviations (σ) for value weighted and equal weighted U.S. market returns from the FF and Thomson Reuters Datastream datasets as well as correlations between both return series (ρ) over time.[14] The value weighted market returns are quite similar, with an average monthly return of 0.81% for the FF data and an average monthly

[14] Since Kenneth French does not provide EW market returns on his website, corresponding returns from the CRSP database are reported.

return of 0.82% for our Thomson Reuters Datastream data. The correlation coefficient between the FF and our Thomson Reuters Datastream value weighted returns is 0.95. The Thomson Reuters Datastream equal weighted market return of 1.34% is higher than the equal weighted market return from the FF dataset, which is 0.90%. Given that the FF dataset uses only NYSE, AMEX and NASDAQ stocks, whereas our Thomson Reuters Datastream dataset contains potentially all U.S. listed stocks available, it is not surprising that the equal weighted average return of the Thomson Reuters dataset is higher than the equal weighted average return of the FF dataset since small stocks, listed on regional exchanges, will be covered only in the former. Smaller stocks are generally considered to gain higher returns because of a premium for illiquidity or default risk.[15] However, importantly, the correlation between the two equal weighted market returns is 0.97.

[Table 6 here]

4.2. Risk factors

We now analyze the time series of the U.S. SMB, HML and WML factors. The corresponding results are also shown in Table 6. The average values for the SMB factors are rather low and amount to 0.04% per month (FF) and 0.06% (Thomson Reuters Datastream and Thomson Reuters Worldscope). The correlation coefficient between the two SMB factors based on the FF and our Thomson Reuters Datastream and Thomson Reuters Worldscope dataset is 0.93. The HML factors yield higher average values than the SMB factors and are very similar with 0.34% per month for the FF dataset and 0.30% per month for our Thomson Reuters Datastream and

[15] Leippold and Lohre (2009, Section 6.3) discuss the impact of illiquidity on momentum returns and conclude that "the least momentum profits occur for the most liquid stocks and that profitability is increasing with illiquidity." Vassalou and Xing (2004, p. 866) conclude that "Small firms earn higher returns than big firms, only if they also have high default risk."

Thomson Reuters Worldscope dataset. The correlation coefficient between the two HML factors is 0.87. The WML factors have the highest average values with 0.86% per month (FF) and 0.76% per month (Thomson Reuters Datastream and Thomson Reuters Worldscope). The correlation coefficient between both factors is 0.96. In sum, we are able to replicate very closely the properties of the benchmark risk factors, suggesting that the screens are effective in transforming the raw data into a data series suitable for further analysis.

4.3. Portfolios sorted on size, BE/ME, and momentum

To further evaluate the quality of our sample, we sort all sample stocks separately on the characteristics size, BE/ME and momentum. We compare the individual portfolios of each sort with portfolios provided by Kenneth French. We report means, standard deviations and correlation coefficients of the average monthly returns over time to compare the portfolios with each other.

First, we sort all stocks in our sample according to their size and group them into ten size groups according to the empirical breakpoints inferred from the FF data, as described in Section 3.2 (see also Table A.1). The results are shown in Table 7. The correlation coefficients, ranging between 0.93 and 0.96, show that the returns of our size portfolios behave very similarly to the returns of the FF size portfolios.

Note also that the average returns for the ten size groups are very similar for the FF and our Thomson Reuters Datastream and Thomson Reuters Worldscope sample. The only exception is the smallest size group, in which the average return in our sample exceeds the FF average return by 0.27 percentage points per month, suggesting the presence of a size effect in Thomson Reuters Datastream and Worldscope data. This finding is consistent with our conjecture about the small, non-NYSE/AMEX/NASDAQ stocks being responsible for the higher equal weighted Thomson Reuters Datastream returns.

Next, we consider the results for ten BE/ME groups. Here, we form portfolio groups by employing decile breakpoints (see Table A.2). The results are also shown in Table 7. The average returns for the ten FF BE/ME groups are approximately increasing in BE/ME. We observe the same behavior for our ten Thomson Reuters Datastream and Thomson Reuters Worldscope BE/ME groups. The correlations are somewhat smaller than in the case of the size groups, but still very high, ranging from 0.83 to 0.94.

Table 7 also shows the same figures for the ten momentum group groups, again by employing decile breakpoints (see Table A.3). The ten momentum groups of each sample show an almost monotonic behavior between momentum and average returns. The average returns of the Thomson Reuters Datastream groups are substantially higher than the average returns in the FF groups in case of the tenth and ninth group, suggesting that the small non-NYSE/AMEX/NASDAQ stocks are mostly contained in these groups. The correlations of the momentum groups range between 0.87 and 0.94.

[Table 7 here]

Next, we compare Thomson Reuters Datastream and Thomson Reuters Worldscope and FF portfolios sorted on two characteristics jointly. Overall, the twenty-five portfolios sorted on size-BE/ME and size-momentum calculated from Thomson Reuters Datastream and Thomson Reuters Worldscope data are quite similar to the corresponding portfolios provided by Kenneth French when evaluated in terms of return correlations. There are some notable differences in average returns, though.

Panel A of Table 8 shows the detailed results. For most of the size groups there seems to be a positive monotonic relation between BE/ME and average returns. However, for the BE/ME groups we observe a different behavior regarding size, depending on the specific group. For low BE/ME stocks, we detect an "inverted size effect" (Fama 1991, p. 1588), which means that big firms yield higher average returns than small firms. However, this effect is much more pronounced in the FF sample. Thus, the biggest difference in the average returns of the Thomson Reuters Datastream and Thomson Reuters Worldscope and FF size-BE/ME return series can be found in the small size/low BE/ME (S/L) group. This finding suggests that a significant portion of the non-NYSE/AMEX/NASDAQ stocks have a low BE/ME. For the second and third BE/ME group there seems to be no relation between size and average returns. In the fourth and the highest BE/ME group a size effect with high returns in the small size groups and low returns in the big size groups can be observed in both samples. The correlations of the 25 size-BE/ME Thomson Reuters Datastream and Thomson Reuters Worldscope portfolios with the 25 size-BE/ME FF portfolios range between 0.76 (B/H-portfolio) and 0.96 (S/H-portfolio).

[Table 8 here]

We report the results for 25 size-momentum portfolios in the panel B of Table 8. In case of the FF portfolios, we observe an "inverted size effect" in the loser and the second momentum group and a size effect in the third, fourth and winner groups. In contrast, for the Thomson Reuters Datastream and Thomson Reuters Worldscope size-momentum portfolios we observe a size effect in all momentum groups, except in the loser group. In each of the size groups we observe a momentum effect, which means that the average returns of the winner portfolio are always higher than the average returns of the looser portfolio. The correlations of the twenty-five size-momentum returns between the FF- and the Thomson Reuters Datastream sample range between 0.90 and 0.96.

5. Results for the European market

This section presents the results for the pan-European market returns and risk factors (Sections 5.1 and 5.2). We also compare our self-created local market indices with publicly available local market indices to evaluate data quality (Section 5.3).

5.1. Market returns

Table 9 shows averages (avg.) and standard deviations (σ) for value weighted and equal weighted pan-European market returns from the Thomson Reuters Datastream database. The value weighted market return is on average 0.81% per month. Furthermore, the equal weighted market return amounts to 0.88% per month on average.

[Table 9 here]

5.2. Risk factors

We next compile overall risk factors of all European OECD countries. The results are also shown in Table 9. The average value for the SMB factor is rather low and amounts to -0.02% per month. The HML factor yields a higher average value than the SMB factor with 0.48% per month. The WML factor has the highest average value with 1.11% per month.

5.3. Market returns for single European countries

In order to evaluate the quality of our sample we compare self-created market indices from different European countries with market indices available on Thomson Reuters Datastream. In Table 10 we present results for the market returns of twenty-three European countries. We report average percentage values of known local indices with a sufficiently long time series, as well as value weighted and equal weighted market returns calculated from Thomson Reuters Datastream firm-level data. Furthermore, we present correlation coefficients of the value weighted and equal weighted market returns with the respective index(es). Two time periods are examined: a long period (07/1989 – 08/2009) and a short period (07/1999 – 08/2009).[16]

[Table 10 here]

There are differences by construction between the publicly available local indexes, which we use for comparison and the self-compiled value weighted indexes. First, the local indexes are usually calculated with the free float market capitalization as index weights, whereas we use total market capitalization. Second, we use price and dividend data to compile the indices, whereas some local indexes employed for comparison incorporate only price information.[17] When possible, we use Thomson Reuters Datastream total return indices, which include dividend payments. However, these indices are not always available and therefore we use also pure price indices for

[16] To report results as uniform as possible for all European markets we use a different sample period as in the U.S. case. Although a few markets seem to have a broad coverage back to 1986, most markets seem to be covered much better a few years later. Therefore we chose 07/1989 as the start date for our investigation. Furthermore we use a later ending data than in the U.S. case (08/2009 instead of 12/2008).

[17] For example, the U.S. value weighted market returns on CRSP without dividends is on average 0.14 % (per month) lower than the CRSP value weighted market return with dividends for the period ranging from July 1986 to December 2008.

comparison purposes.[18] The third difference is that indexes like FTSE or MSCI do not include all stocks available because of the limited investability of small stocks. The remaining indices are either broad market indices (BAS (Belgium), ISEQ (Ireland), SPI (Switzerland), LSE (Luxembourg), WGI (Poland), ICEXALL (Iceland)); indices restricted to a certain number of firms (CAC40 (France), AEX (Netherlands)) or indices which cover a certain portion of the total market capitalization (BUX (Hungary), SAX (Slovakia)).

Panel A reports the results for all countries with available data for both periods. Panels B-G report results for countries for which we use different time periods, due to data availability restrictions.[19] We observe that for the twenty biggest European stock markets[20] the correlations with the local indices for the 07/1999 – 08/2009 period are at least 0.95 or even higher. For the 07/1989 – 08/2009 period, the thirteen biggest markets have at least a correlation of 0.97 with the respective local indices. Furthermore, it is a satisfying result that for the biggest stock markets the Thomson Reuters Datastream calculated indices are almost perfectly correlated with the comparison indices. The seven biggest stock markets (UK, France, Germany, Switzerland, Spain, Italy and the Netherlands) all have at least correlations of 0.98 (0.98) in the period from 07/1999 – 08/2009 (07/1989 – 08/2009) with the respective comparison indices. Correlation coefficients in all countries are higher than 0.93 (0.95) for the long (short) period except for Luxembourg, Poland (long period), Slovakia (data are only available for the short period) and

[18] The Swiss Performance index (SPI), the Warsaw General Index (WGI), The Share Index of the Budapest Stock Exchange (BUX) and the Slovak Share Index (SAX) include dividend payments by construction. Furthermore, we use total return indices for the following countries: Austria (short period), Denmark (short period), Finland (short period), France (both periods), Germany (short period), Ireland (both periods), Italy (short period), Netherlands (both periods), Norway (short period), Portugal (both periods), Spain (short period), Sweden (short period), Turkey (both periods), U.K. (both periods), Luxembourg (second period), Greece (both periods), Hungary (MSCI) and Czech Republic (both periods). All other indices are pure price indices.

[19] For the sake of clarity we do not report more than one comparison index. The only exception is Hungary for which we report in the second period also results for the MSCI-Index, besides the BUX, for which we report results for both periods. Since the BUX is a blue chip index and covers only the largest companies traded on the Budapest Stock Exchange (which contains thirteen firms in May 2010), the MSCI index is in principle better suited than the BUX. However, in the first period this index is not completely available (in contrast to the BUX).

[20] Table A.6 in the Appendix lists all European OECD countries on their market capitalizations as by June 2008. All further remarks about aggregated market size of European countries refer to Table A.6.

Iceland (data are only available for the 01/2003 – 08/2009 period). Moreover, in three cases correlations in the long period are slightly higher than in the short period (no more than 0.01 difference) whereas correlations in the short period are for some countries (Luxembourg and Poland) considerably higher (more than 0.09 difference) than in the long period.

We suspect that the relatively low correlation of our indices with the comparison indices for Luxembourg, Slovakia and Iceland can be explained by the fact that companies which have an influence on the respective local market returns are nevertheless so small that they are not sufficiently covered by Thomson Reuters Datastream and Thomson Reuters Worldscope.[21]

In sum, we conclude that the European dataset compiled from company-level Thomson Reuters Datastream and Thomson Reuters Worldscope data yields, with some exceptions for tiny markets, quite reliable results after the correction of data errors as described in this paper.

6. Comparison between the U.S. and pan-European markets

The new European risk factors that we have obtained allow us to ask and answer a novel research question: How do the risk factors of the U.S. and the European economic regions compare with each other, and in particular, how highly correlated are they? While a complete answer of this question is outside the scope of this paper (which is focused more on data issues), some initial results can be obtained that motivate further research. For this analysis, we consider the period from 07/1989 to 12/2008, i.e., the overlapping time period. Table 11 compares the European risk factors with the corresponding Thomson Reuters Datastream and Thomson

[21] For example, a closer examination reveals that over 50% (in terms of the market capitalization) of the SAX index is not covered by Thomson Reuters Datastream data when we try to find the corresponding companies in April 2001 (according to Bratislava Stock Exchange, 2001) within our Thomson Reuters Datastream and Thomson Reuters Worldscope sample. Most companies are not covered by Thomson Reuters Worldscope, others are covered by Thomson Reuters Worldscope, but Thomson Reuters Datastream provides no market data or the stocks are excluded by one of our screens.

Reuters Worldscope and FF U.S. factors.

[Table 11 here]

Interestingly, the averages do not differ much. The average return of the U.S. market portfolio calculated with Thomson Reuters Datastream and Thomson Reuters Worldscope (FF) data amounts to 0.70% (0.71%) whereas the European market factor earns an average return of 0.81% per month. The difference between the two sets of HML and WML factors is of a similar magnitude. The U.S. SMB factor earns a (positive) average return of 0.11% (0.10%) and the European SMB factor earns a (negative) average return of -0.02% per month. The biggest difference between the European factors and their U.S. equivalents in terms of absolute values occurs for the returns of the HML and WML factors.

The correlations between Europe and the U.S. confirm the impressions of the average returns: The market, HML, and WML factors are rather highly correlated across regions with correlation coefficients of 0.80 (0.79), 0.57 (0.54) and 0.67 (0.65), respectively. By contrast, the SMB factor has the lowest correlation with coefficient of 0.21 (0.20).

To get further insights of how U.S. and European risk factors are related, we plot the cumulative returns of the factor portfolios for each market against each other. Besides the Thomson Reuters Datastream and Thomson Reuters Worldscope calculated factor portfolios for both regions we also plot the FF factor portfolios for the U.S. This is done in Figures 1-4.

[Figures 1 to 4 here]

Figure 1 shows the value weighted market factors. It can be observed that the two cumulative market portfolio returns generally tend to move into the same directions. From 02/91 to 02/05 the cumulative return earned by the U.S. market portfolio appears to be higher than the cumulative return earned by the European market portfolio. But from 03/05 to 04/09 the cumulative return earned by the European market portfolio was higher. This observation might be partly due to an appreciation of the Euro against the US$, since we denominate all returns in US$. The figures for the HML and WML factors (Figures 3 and 4) look similar. The two cumulative returns of the factor portfolios show in general the same trends. In case of the HML factor portfolio the cumulative return is, for both markets, virtually zero until 01/00 and then increases in both markets until 12/06. Note that in case of the HML and the WML factor the cumulative portfolio return on the European market is in general higher than on the U.S. market. The cumulative values of the SMB factor (see Figure 2) are near zero for both markets, but also show most of the time the same trending behaviour. For example, the spike on 03/00 is observed in both markets, although the spike is much more pronounced for the U.S. factor.

We believe these to be novel results that can stimulate further inquiry. Most studies on international asset pricing only report correlations between single countries, and not regions like the U.S. and Europe. For example, Fama and French (1998, Table 6) report a correlation of 0.51 between U.S. and U.K. market returns for the period from 1975 to 1995, and Griffin (2002, Table 1) reports a correlation of 0.68 for the period from 1981 to 1995. These numbers are rather small compared to the correlation of 0.80 reported in this study between the two regions. One interpretation of the present findings is that correlations on a higher aggregation level between U.S. and European stock markets are higher than those on the country-level. Besides the aggregation effect, another explanation might be that U.S. and European stock markets have become more integrated in more recent times, at least with respect to some determinants of stock returns. When studying market integration, existing work has mostly focused on correlations of market returns. The results here provide a first view on correlations of risk factors. Future research is needed to address this topic more fully. (Naranjo and Porter (2010) investigate comovement of momentum returns from 40 countries, finding that the correlations change across various sub-periods. Schrimpf et al. (2011) model time-varying co-dependence with a copula-based approach and consider value, size, and momentum returns for a smaller set of countries.)

7. Conclusion

A major obstacle for research in international asset pricing and corporate finance has been a lack of reliable and publicly available data on international risk factors and portfolios. With this paper, we aim to make a step towards overcoming this obstacle. Specifically, this paper provides a detailed analysis of how to construct high-quality, replicable portfolios and risk factors from Thomson Reuters Datastream and Thomson Reuters Worldscope data.

We first outline appropriate screens and data filters by which the quality and the reliability of the data can be raised significantly. This is demonstrated for the U.S., for which we show that the discussed data screening procedures lead to portfolios and risk factors based on Thomson Reuters Datastream and Thomson Reuters Worldscope data that have very similar properties as those obtained from CRSP and Compustat. Furthermore we expand the analysis to European stock markets, showing that the correlations of our self-compiled value weighted indices with well-known representative stock market indices are very high. Additionally, we calculate pan-European risk factors, including all European OECD countries. A first result obtained with these data is that the pan-European market returns and risk factors appear to be astonishingly highly correlated with their counterparts in the U.S., with the exception of the size factor. To facilitate research on international asset pricing, the European risk factors computed in this paper are made freely available to other scholars.

References

Adrian, T., Franzoni, F., 2009. Learning about Beta: Time-varying Factor Loadings, expected returns, and the conditional CAPM. Journal of Empirical Finance 16, 537–556.

Ammann, M., Steiner, M., 2008. Risk Factors for the Swiss Stock Market. Swiss Journal of Economics and Statistics 144 1–25.

An, J., 2010. Can the Long-Run Risks Explain the International Value Premium? Evidence Using Last Century Data. Working Paper, Cornell University.

Ang, A., Chen, J., 2002. Asymmetric Correlations of Equity Portfolios. Journal of Financial Economics 63, 443–494.

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2008. High Idiosyncratic Volatility and Low Returns: International and Further U.S. Evidence. Journal of Financial Economics 91, 1–23.

Angelidis, T., Tessaromatis, N., 2008. Does Idiosyncratic Risk Matter? Evidence from European Stock Markets. Applied Financial Economics 18, 125–137.

Artmann, S., Finter, P., Kempf, A., Koch S., Theissen, E., 2010. The Cross-Section of German Stock Returns: New Data and New Evidence. Working Paper, Centre for Financial Research.

Bauer, R., Cosemans, M., Schotman, P.C., 2010. Conditional Asset Pricing and Stock Market Anomalies in Europe. European Financial Management 16, 165-190.

Bratislava Stock Exchange, 2001. Monthly Statistics – July 2001.

Campbell, J.Y., 1996. Understanding Risk and Return. The Journal of Political Economy, 104, 298–345.

Carhart, M.M., 1997. On Persistence in Mutual Fund Performance. The Journal of Finance 52, 57–82.

Chen, N.-F., Roll, R. Ross, S.A., 1986. Economic Forces and the Stock Market. The Journal of Business 59, 383–403.

Cochrane, J.H., 1996. A Cross-Sectional Test of an Investment-Based Asset Pricing Model. The Journal of Political Economy 104, 572–621.

Dimson, E., Nagel, S., Quigley, G., 2003. Capturing the Value Premium in the United Kingdom. Financial Analysts Journal 59, 35–45.

Fama, E.F., 1991. Efficient Capital Markets: II. The Journal of Finance 46, 1575–1617.

Fama, E.F., French, K.R., 1993. Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics 33, 3–56.

Fama, E.F., French, K.R., 1998. Value versus Growth: The International Evidence. The Journal of Finance 53, 1975–1999.

Ferson, W.E., Harvey, C.R., 1991. The Variation of Economic Risk Premiums. The Journal of Political Economy 99, 385–415.

Gregory, A., Tharyan, R., Huang, A., 2009. The Fama-French and Momentum Portfolios and Factors in the UK. Working Paper, University of Exeter.

Griffin, J.M., 2002. Are the Fama and French Factors Global or Country Specific? The Review of Financial Studies 15, 783–803.

Heston, S.L., Rouwenhorst, K.G., Wessels, R.E., 1999. The Role of Beta and Size in the Cross-Section on European Stock Returns. European Financial Management 5, 9–27.

Hodgson, D.J., Linton, O., Vorkink, K., 2002. Testing the Capital Asset Pricing Model Efficiently under Elliptical Symmetry: A Semiparametric Approach. Journal of Applied Econometrics 17, 617–639.

Hou, K., Karolyi, G.A., Kho, B.C., 2006. What Factors Drive Global Stock Returns? Manuscript, Ohio State University.

Ince, O.S., Porter, R.B., 2006. Individual Equity Return Data from Thomson Datastream: Handle with Care! The Journal of Financial Research 29, 463–479.

Jegadeesh, N., Titman, S. , 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. The Journal of Finance 48, 65–91.

Kothari, S.P., Shanken, J., Sloan, R.G., 1995. Another Look at the Cross-Section of Expected Stock Returns. The Journal of Finance 50, 185–224.

Leippold, M., Lohre, H., 2009. International Price and Earnings Momentum. Working Paper, Imperial College - Business School.

Leippold, M., Lohre, H., 2010. Data Snooping and the Global Accrual Anomaly. Working Paper, University of Zurich.

Liew, J., Vassalou, M., 2000. Can Book-to-Market, Size and Momentum Be Risk Factors That Predict Economic Growth? Journal of Financial Economics 57, 221–245.

Lo, A.W., MacKinlay, A.C., 1990. Data-Snooping Biases in Tests of Financial Asset Pricing Models. The Review of Financial Studies 3, 431–467.

MacKinlay, A.C., 1995. Multifactor Models Do Not Explain Deviations from the CAPM. Journal of Financial Economics 38, 3–28.

Moerman, G.A., 2005. How Domestic is the Fama and French Three-Factor Model? An Application to the Euro Area. Working Paper, Erasmus Universiteit Rotterdam.

Nagel, S., 2001. Accounting Information Free of Selection Bias: A New UK Database 1953-1999. Working Paper, London Business School.

Naranjo, A., Porter, R.B., 2010. Risk factor and industry effects in the cross-country comovement of momentum returns. Journal of International Money and Finance 29, 275-299.

Rouwenhorst, K.G., 1998. International Momentum Strategies. The Journal of Finance 53, 267–284.

Schrimpf, A., Schröder, M., Stehle, R., 2007. Cross-sectional Tests of Conditional Asset Pricing Models: Evidence from the German Stock Market. European Financial Management 13, 880–907.

Schrimpf, A., Wagner, A., Ziegler, A., 2010. International Value and Momentum Investment Styles. Working Paper, Aarhus University.

Thomson Financial, 2007. Worldscope Database: Datatype Definitions Guide. Issue 6.

Ulbricht, N., Weiner, C., 2005. Worldscope Meets Compustat: a Comparison of Financial Databases. Working Paper, Humboldt University Berlin.

Vaihekoski, M., 2004. Portfolio Construction for Tests of Asset Pricing Models. Financial Markets, Institutions & Instruments 13, 1–39.

Vassalou, M., Xing, Y., 2004. Default Risk in Equity Returns. The Journal of Finance 59, 831–868.

Ziegler, A., Schröder, M., Schulz, A., Stehle, R., 2007. Multifaktormodelle zur Erklärung deutscher Aktienrenditen: Eine empirische Analyse. Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung 59, 355–389.

Table 1: Dynamic screens

Screen identifier	Short description	Items involved
S01	We delete all zero returns (with returns calculated from the total return index) from the end of the sample until the first non-zero return (cf. Ince and Porter, 2006, p. 465).	Total Return Index
S02	We delete all zero values (with returns calculated from the price index) from the end of the sample until the first non-zero value (cf. Ince and Porter, 2006, p. 465).	Price Index
S03	We delete all so-called "Penny-stocks" with prices less than one unit of the domestic currency (cf. Ince and Porter, 2006, p. 473).	Unadjusted Price
S04	We set all returns to missing for which the price is greater than 1,000,000 of the domestic currency.	Price Index
S05	We divide all dividends by a fixed value, which are greater than half the adjusted price (a detailed treatment on this issue is given in Section 2).	Price Index, Dividends
S06	If there are no observations in the total return index, then price and dividend (if available) information are used to compile returns, if at least price information is available.	Total Return Index, Price Index, Dividends
S07	We compare the Thomson Reuters Datastream total return index with the self-created total return index constructed from price and dividend (if available) data and use the self-created index if the difference between the total return index is greater than 0.5 in absolute terms (cf. Ince and Porter, 2006, p. 473).	Total Return Index, Price Index, Dividends
S08	We compare the Thomson Reuters Datastream market value with the self-created market value, calculated by multiplying the unadjusted price with the number of shares and set the market value to missing if the difference in terms of the self-created market value is greater than 0.5 in absolute terms.	Total Return Index, Price Index, Dividends
S09	We set all returns to missing, for which the return is greater than 890%.	Return
S10	We delete the returns for which R_t or R_{t-1} is greater than 300% and $(1+R_t)(1+R_{t-1})-1$ is less than 50% (cf. Ince and Porter, 2006, p. 473-474, fn. 4).	Total Return Index
S11	All stocks are excluded which are not listed on U.S. exchanges.	Exchange Mnemonic
S12	We search the Extended Name for suspicious word parts "pref", "prf", "%", "duplicate", "dupl" and set, if necessary, the returns to missing (cf. Ince and Porter, 2006, p. 471).	Extended Name

Table 2: Number of firms for the U.S. market

Year	List	Corrected	Market	SMB/HML	WML
1984	1,905	1,877	1,862	1,670	1,683
1985	1,992	1,956	1,940	1,840	1,848
1986	2,139	2,085	2,057	1,908	1,914
1987	2,269	2,217	2,192	1,984	1,996
1988	2,514	2,442	2,406	2,056	2,080
1989	2,625	2,542	2,491	2,248	2,264
1990	2,594	2,508	2,457	2,351	2,358
1991	2,660	2,537	2,479	2,329	2,325
1992	3,027	2,899	2,821	2,369	2,396
1993	3,200	3,078	2,987	2,693	2,714
1994	3,799	3,672	3,563	2,901	2,925
1995	4,800	4,615	4,502	3,317	3,430
1996	5,352	5,197	5,067	4,294	4,339
1997	6,070	5,853	5,703	4,806	4,840
1998	7,126	6,566	6,272	5,225	5,253
1999	8,153	7,087	6,612	5,378	5,410
2000	8,741	7,383	6,871	5,562	5,477
2001	8,398	6,245	5,969	5,277	5,094
2002	8,071	5,637	5,325	4,856	4,586
2003	7,632	5,599	5,306	4,765	4,487
2004	7,470	5,757	5,390	4,788	4,572
2005	7,494	5,729	5,381	4,772	4,531
2006	7,471	5,807	5,407	4,758	4,540
2007	7,375	5,621	5,256	4,684	4,512
2008	6,973	5,148	4,883	4,414	4,248
All	15,241	14,203	13,343	11,114	11,654

Note: We report the number of firms for different stages of the data preparation process. The numbers shown correspond to July for each year. The "List" column refers to the number of firms for the list which we use to draw time series data. This list is already corrected for static items, which are foreign stocks, non-equities, and additional listings as described in Section 2. In addition, we require the dscd code (DSCD), common equity (WC03501), fiscal year end (WC05350), number of shares (NOSH) and prices (P) to be non-missing, which are minimum requirements to be included in the portfolio sorts. The "corrected" column refers to the number of firms after the time series screens depicted in Table 1. The same minimum requirements as for the "list" stocks are also imposed. The "Market", "SMB/HML" and "WML" columns report the number of firms which are actually used to compile the Market Return, the SMB and HML factors and the WML factor. Note that these numbers refer to the value weighted case. In the equal weighted case the numbers would be slightly higher, since the size values of the preceding month are not needed. The "All" row reports the number of unique firms observed over the whole time span.

Table 3: Number of firms for the European market

Year	List	Corrected	Market	SMB/HML	WML
1984	713	672	609		
1985	852	812	741		
1986	983	935	920		
1987	1,270	1,219	1,196	895	860
1988	2,084	1,999	1,957	1,085	1,115
1989	2,819	2,720	2,664	1,800	1,822
1990	3,192	3,080	3,018	2,484	2,505
1991	3,451	3,315	3,230	2,811	2,800
1992	3,580	3,393	3,281	2,989	2,982
1993	3,668	3,469	3,343	3,017	3,011
1994	3,862	3,673	3,532	3,145	3,157
1995	4,060	3,768	3,641	3,320	3,336
1996	4,246	3,944	3,797	3,415	3,422
1997	5,179	4,792	4,602	3,552	3,594
1998	5,615	5,200	4,984	4,317	4,345
1999	5,972	5,485	5,288	4,554	4,588
2000	5,854	5,279	5,117	4,429	4,424
2001	6,142	5,472	5,350	4,511	4,482
2002	6,282	5,277	5,128	4,717	4,657
2003	6,023	4,965	4,798	4,453	4,317
2004	5,990	4,942	4,729	4,289	4,244
2005	6,189	5,076	4,870	4,269	4,210
2006	6,548	5,401	5,212	4,374	4,370
2007	6,755	5,605	5,436	4,717	4,570
2008	6,777	5,425	5,317	4,751	4,727
All	12,218	11,086	11,054	9,462	10,035

Note: We report the number of firms for different stages of the data preparation process. The numbers shown correspond to July for each year. The "List" column refers to the number of firms for the list which we use to draw time series data. This list is already corrected for static items, which are foreign stocks, non-equities, and additional listings as described in Section 2. In addition, we require the dscd code (DSCD), common equity (WC03501), fiscal year end (WC05350), number of shares (NOSH) and prices (P) to be non-missing, which are minimum requirements to be included in the portfolio sorts. The "corrected" column refers to the number of firms after the time series screens depicted in Table 1. The same minimum requirements as for the "list" stocks are also imposed. The "Market", "SMB/HML" and "WML" columns report the number of firms which are actually used to compile the Market Return, the SMB and HML factors and the WML factor. Note that these numbers refer to the value weighted case. In the equal weighted case the numbers would be slightly higher, since the size values of the preceding month are not needed. The "All" row reports the number of unique firms observed over the whole time span. Note that the "European" sample is composed of all European OECD countries.

Table 4: Portfolio sorts for factor construction

Panel A: Six size-BE/ME portfolios

		BE/ME		
		low	medium	high
size	small	S/L	S/M	S/H
	big	B/L	B/M	B/H

Panel B: Six size-momentum portfolios

		momentum		
		losers	medium	winners
size	small	S/L	S/M	S/W
	big	B/L	B/M	B/W

Note: This table illustrates the sorting procedure which is used to create six size-BE/ME and six size-momentum portfolios which are the building blocks of the SMB, HML and WML factors. Panel A: All stocks are divided into two size groups by their market value (small (S) and big (B)). Simultaneously all stocks are also divided into three BE/ME groups (low (L), medium (M) and high (H)). Panel B: All stocks are divided into two size groups by their market value (small (S) and big (B)). Simultaneously all stocks are also divided into three groups depending on the average returns of the last twelve month, by skipping the most recent one (losers (L), medium (M) and winners (W)). For a discussion of the breakpoints see Section 3.2 and Table 5.

Table 5: Breakpoints for double sorts

	Mean	Median	Minimum	Maximum	Actual
Panel A: Breakpoints for size and BE/ME					
$size_{BP1}$	0.81	0.81	0.77	0.84	0.80
BE/ME_{BP1}	0.36	0.36	0.29	0.42	0.30
BE/ME_{BP2}	0.70	0.70	0.62	0.76	0.70
Panel B: Breakpoints for size and momentum					
$size_{BP2}$	0.82	0.82	0.77	0.86	0.80
mom_{BP1}	0.39	0.40	0.22	0.55	0.30
mom_{BP2}	0.70	0.71	0.53	0.84	0.70

Note: We use the number of portfolio constituents provided by Kenneth French to calculate size and BE/ME breakpoints (Panel A) as well as size and momentum breakpoints (Panel B), which apply to the whole sample (not only to NYSE stocks). We do so to find breakpoints for our sample, which are close to the FF breakpoints. The table shows the size breakpoint ($size_{BP1}$) and the two BE/ME breakpoints (BE/ME_{BP1} and BE/ME_{BP2}) for the building blocks of the SMB and HML factors (Panel A) as well as the size breakpoint ($size_{BP2}$) and the two momentum breakpoints (mom_{BP1} and mom_{BP2}) for the building blocks of the WML factor (Panel B). We report mean, median, minimum and maximum of these empirical FF breakpoints. Furthermore, we report the breakpoints actually employed in this study (column "actual"). The time period ranges from 07/1986 to 12/2008.

Table 6: Market returns and risk factors for the U.S. market

	FF			TR			
	Avg.	σ	t	Avg.	σ	t	ρ
VW	0.81	4.54	2.98	0.82	4.47	3.05	0.95
EW	0.90	5.43	2.75	1.34	5.30	4.22	0.97
SMB	0.04	3.40	0.18	0.06	3.11	0.33	0.93
HML	0.34	3.12	1.81	0.30	3.03	1.63	0.87
WML	0.86	4.48	3.14	0.76	4.89	2.54	0.96

Note: This table reports descriptive statistics for the time series of monthly value weighted (VW) and equal weighted (EW) market returns as well as the returns of the SMB, HML and WML factors in %. We compare two different U.S. datasets with each other: The FF and Thomson Reuters Datastream and Thomson Reuters Worldscope (TR) as described in Section 2. We report the sample average (Avg.), the sample standard deviation (σ), the t-statistic (t) and the correlation coefficient between the two samples (ρ). The t-statistic tests the hypothesis whether the mean of the tested series is zero. A rejection indicates that the mean is different from zero. In samples with a size greater than 31 (as it is here the case), the distribution under the null is well approximated by a standard normal distribution. The time period ranges from 07/1986 to 12/2008. All returns are in percent per month and are denominated in US$.

Table 7: One way sorts on size, BE/ME and momentum for the U.S. market

	FF		TR			FF		TR			FF		TR		
	Avg.	σ	Avg.	σ	ρ	Avg.	σ	Avg.	σ	ρ	Avg.	σ	Avg.	σ	ρ
	size					BE/ME					momentum				
Group 1	0.79	6.12	1.06	5.56	0.96	0.71	5.27	0.67	5.44	0.94	-0.26	8.28	-0.29	9.62	0.94
Group 2	0.79	6.37	0.78	5.97	0.94	0.79	4.81	0.70	4.68	0.92	0.50	6.32	0.50	6.99	0.92
Group 3	0.85	5.93	0.81	5.77	0.94	0.90	4.72	0.81	4.54	0.90	0.59	5.31	0.40	5.92	0.91
Group 4	0.78	5.75	0.80	5.53	0.94	0.87	4.72	0.80	4.42	0.90	0.74	4.66	0.50	5.08	0.90
Group 5	0.85	5.63	0.76	5.31	0.93	0.83	4.50	1.01	4.47	0.90	0.65	4.31	0.62	4.48	0.87
Group 6	0.84	5.17	0.77	5.06	0.94	0.77	4.49	0.94	4.29	0.86	0.65	4.32	0.69	4.28	0.93
Group 7	0.91	5.08	0.87	4.96	0.94	0.94	4.35	0.84	4.69	0.85	0.80	4.27	0.81	4.28	0.89
Group 8	0.85	5.15	0.81	5.19	0.94	0.85	4.30	0.78	5.05	0.86	1.02	4.22	0.87	4.33	0.93
Group 9	0.86	4.71	0.83	4.83	0.95	0.98	4.51	1.21	4.92	0.83	0.87	4.69	0.98	4.81	0.90
Group 10	0.76	4.45	0.76	4.40	0.94	0.99	5.26	1.03	6.02	0.86	1.31	6.41	1.44	6.98	0.93

Note: We report descriptive statistics for the time series of ten size, BE/ME and momentum groups. We compare two different U.S. datasets with each other: The dataset provided by Kenneth French (FF) and the dataset compiled from Thomson Reuters Datastream and Thomson Reuters Worldscope data (TR) as described in Section 2. We report the sample average (Avg.), the sample standard deviation (σ) and the correlation coefficient between the two samples (ρ). The time period ranges from 07/1986 to 12/2008. All returns are in percent per month and are denominated in US$.

Table 8: Two way sorts on size-BE/ME and size-momentum for the U.S. market

	FF					TR									
	Average					Average					ρ				
Panel A: size-BE/ME portfolios															
	L	2	3	4	H	L	2	3	4	H	L	2	3	4	H
S	0.06	0.89	0.96	1.20	1.15	0.61	0.84	0.98	1.16	1.26	0.92	0.93	0.93	0.93	0.96
2	0.51	0.80	1.06	1.05	1.03	0.55	0.84	0.97	1.03	1.07	0.93	0.91	0.90	0.90	0.89
3	0.59	0.84	0.96	0.97	1.23	0.56	0.86	0.93	0.89	1.09	0.93	0.92	0.88	0.89	0.83
4	0.85	0.88	0.82	1.06	0.93	0.79	0.85	0.81	0.91	1.27	0.93	0.92	0.92	0.89	0.87
B	0.80	0.90	0.77	0.80	0.89	0.75	0.81	1.02	0.66	1.07	0.95	0.92	0.90	0.85	0.76
Panel B: size-momentum portfolios															
	L	2	3	4	W	L	2	3	4	W	L	2	3	4	W
S	-0.22	0.58	0.91	1.16	1.56	0.28	0.62	0.95	1.11	1.62	0.96	0.94	0.94	0.95	0.96
2	0.10	0.65	0.93	1.09	1.36	0.32	0.63	0.81	0.98	1.40	0.93	0.90	0.92	0.91	0.95
3	0.38	0.61	0.80	0.89	1.30	0.37	0.66	0.82	0.98	1.22	0.93	0.91	0.93	0.91	0.94
4	0.23	0.79	0.85	0.97	1.19	0.43	0.63	0.75	0.94	1.13	0.95	0.92	0.93	0.92	0.92
B	0.36	0.70	0.62	0.91	1.06	0.29	0.49	0.62	0.82	1.11	0.92	0.91	0.93	0.92	0.93

Note: We report descriptive statistics for the time series of twenty-five size-BE/ME (Panel A) and size-momentum portfolios (Panel B). We compare two different U.S. datasets with each other: The dataset provided by Kenneth French (FF) and the dataset compiled from Thomson Reuters Datastream and Thomson Reuters Worldscope data (TR) as described in Section 2. We report the sample average (Average) and the correlation coefficient between the two samples (ρ). The time period ranges from 07/1986 to 12/2008. All returns are in percent per month and are denominated in US$.

Table 9: Market returns and risk factors for the European market

	Avg.	σ	t
VW	0. 81	4.98	2.50
EW	0.88	4.37	3.08
SMB	-0.02	2.34	-0.15
HML	0.48	1.88	3.90
WML	1.11	3.40	4.98

Note: This table reports descriptive statistics for the time series of monthly value weighted (VW) and equal weighted (EW) market returns as well as the returns of the SMB, HML and WML factors in %. The sample includes firms from all European OECD countries. The data are from Thomson Reuters Datastream as described in Section 2. We report the sample average (Avg.), the sample standard deviation (σ) and the t-statistic (t). The t-statistic tests the hypothesis whether the mean of the tested series is zero. A rejection indicates that the mean is different from zero. In samples with a size greater than 31 (as it is here the case), the distribution under the null is well approximated by a standard normal distribution. The time period ranges from 07/1989 to 12/2008. All returns are in percent per month. All returns are in percent per month and are denominated in US$.

Table 10: Comparison with European Indexes

Panel A:	07/1989 - 08/2009					07/1999 - 08/2009				
	Avg.			ρ		Avg.			ρ	
	Com.	VW	EW	VW	EW	Com.	VW	EW	VW	EW
Austria (FTSE)	0.57	0.75	0.71	0.97	0.85	0.77	0.77	0.80	0.98	0.86
Belgium (BAS)	0.38	0.59	0.83	0.97	0.86	-0.04	0.05	0.51	0.96	0.86
Denmark (FTSE)	0.68	0.76	0.84	0.96	0.70	0.76	0.78	0.83	0.98	0.76
Finland (FTSE)	1.05	1.01	1.21	0.97	0.70	0.50	0.64	1.11	0.99	0.68
France (CAC40)	0.71	0.77	1.05	0.99	0.77	0.25	0.38	1.17	0.99	0.81
Germany (FTSE)	0.58	0.71	0.81	0.98	0.78	0.38	0.44	0.87	0.99	0.79
Ireland (ISEQ)	0.65	0.84	1.40	0.98	0.85	-0.02	0.18	1.29	0.97	0.83
Italy (FTSE)	0.44	0.75	0.71	0.99	0.90	0.08	0.20	0.45	0.99	0.90
Netherlands (AEX)	0.77	0.78	0.92	0.98	0.84	-0.03	0.13	0.68	0.99	0.85
Norway (FTSE)	0.69	0.94	1.22	0.97	0.81	0.98	0.99	1.07	0.98	0.83
Portugal (MSCI)	0.65	0.76	1.27	0.94	0.81	0.11	0.31	1.30	0.96	0.76
Spain (FTSE)	0.73	0.96	0.98	0.99	0.84	0.54	0.46	0.80	0.98	0.78
Sweden (FTSE)	0.96	1.08	1.27	0.97	0.80	0.69	0.69	1.19	0.99	0.82
Switzerland (SPI)	0.80	0.88	0.78	0.99	0.81	0.23	0.25	0.68	1.00	0.81
Turkey (MSCI)	5.14	5.47	6.35	0.93	0.90	2.94	2.47	3.74	0.95	0.93
United Kingdom (FTSE)	0.73	0.73	0.64	1.00	0.73	0.22	0.26	0.49	1.00	0.73
Panel B:	01/1992 - 06/1999									
Luxembourg (MSCI/LSE)	1.04	1.65	1.78	0.61	0.61	0.39	0.29	0.64	0.83	0.77
Panel C:	03/1992 - 08/2009									
Greece (MSCI)	1.00	1.03	1.80	0.93	0.71	0.08	0.02	0.64	0.95	0.73

Table 10 (continued): Comparison with European Indexes

Panel D:	02/1993 - 08/2009					07/1999 - 08/2009				
	Avg.			ρ		Avg.			ρ	
	Com.	VW	EW	VW	EW	Com.	VW	EW	VW	EW
Poland (WGI)	2.31	2.77	3.82	0.86	0.76	1.14	0.84	1.90	0.94	0.84
Hungary (BUX)	2.05	1.85	2.58	0.97	0.79	1.14	0.86	1.97	0.98	0.56
Hungary (MSCI)						0.98	0.86	1.97	0.99	0.57
Panel E:	08/1996 - 08/2009									
Czech Republic (FTSE)	1.18	1.21	1.14	0.95	0.67	1.60	1.47	1.65	0.95	0.59
Panel F:										
Slovakia (SAX)						1.32	2.15	2.51	0.67	0.69
Panel G:										
Iceland (ICEXALL)						-0.44	0.75	0.38	0.57	0.72

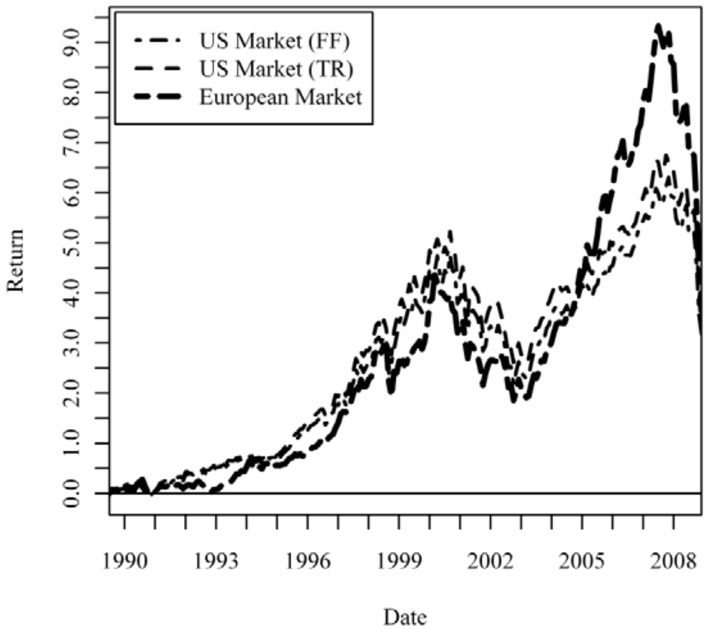
Note: In this table we report basic descriptive statistics of Thomson Reuters Datastream calculated value weighted (VW) and equal weighted (EW) market returns and compare these indexes with publicly available indexes (denoted as Com.). For most countries we report two different sample periods: a long one, typically ranging from 07/1989 - 08/2009 and a short one, typically ranging from 07/1999 - 08/2009, exceptions are indicated. We use the following country-specific indexes for comparison: Brussels All Share (BAS, Belgium), CAC40 (France), FTSE (Austria, Denmark, Finland, Germany, Italy, Norway, Sweden, United Kingdom, Czech Republic), Ireland SE Overall (ISEQ, Ireland), AEX (the Netherlands), MSCI (Portugal, Luxembourg, Greece, Hungary), Madrid SE General (IGBM, Spain), Swiss Performance Index (SPI, Switzerland), Istanbul Stock Exchange National-100 (ISEN100, Turkey), Luxembourg SE General (LSE, Luxembourg), The Share Index of the Budapest Stock Exchange (BUX, Hungary), Warsaw General Index (WGI, Poland), Slovak Share Index (SAX, Slovakia), OMX Iceland All Share (ICEXALL, Iceland). We report the sample average (Avg.) and the correlation coefficient between the two samples (ρ). Averages are reported in percent per month.

Table 11: European and U.S. risk factors

	Europe			U.S. (TR)				U.S. (FF)			
	Avg.	σ	t	Avg.	σ	t	ρ	Avg.	σ	t	ρ
Mkt	0.81	4.98	2.50	0.70	4.24	2.53	0.80	0.71	4.33	2.50	0.79
SMB	-0.02	2.34	-0.15	0.11	3.24	0.52	0.21	0.10	3.52	0.44	0.20
HML	0.48	1.88	3.90	0.28	3.15	1.38	0.57	0.31	3.23	1.48	0.54
WML	1.11	3.40	4.98	0.93	5.07	2.80	0.67	0.98	4.66	3.23	0.65

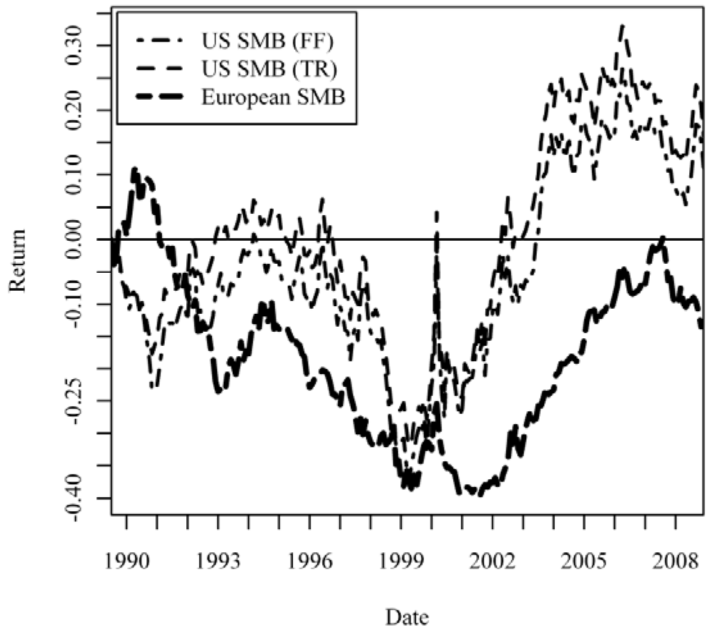
Note: We report descriptive statistics for the time series of the market portfolio (value weighted market return) as well as the SMB, HML and WML factors. We compare the factors compiled with U.S. data from Thomson Reuters Datastream and Thomson Reuters Worldscope and from the data library of Kenneth French (FF) with the factors compiled with European data. The data are from Thomson Reuters Datastream and Thomson Reuters Worldscope data (TR) as described in Section 2 and from Kenneth French´s Webpage (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). We report the sample average (Avg.), the sample standard deviation (σ), t-statistics (t) and the correlation coefficient between the two samples (ρ). The t-statistic tests the hypothesis whether the mean of the tested series is zero. A rejection indicates that the mean is different from zero. In samples with a size greater than 31 (as it is here the case), the distribution under the null is well approximated by a standard normal distribution. The time period ranges from 07/1989 to 12/2008. The European sample consists of all European OECD countries. All returns are in percent per month and are denominated in US$.

Figure 1: Cumulative returns on the market portfolio



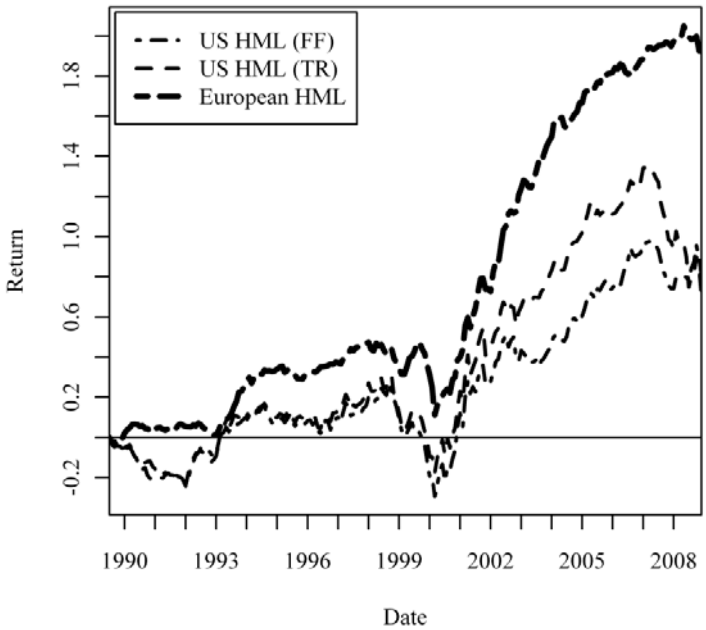
Note: This Figure shows cumulative returns on the market portfolio (value weighted market return). We report the graphs for the market portfolio calculated from Thomson Reuters Datastream data (US Market (TR)), for the U.S. market from Kenneth French's Website (US Market (FF)) and for the market portfolio calculated from a sample of all European OECD countries (European Market).

Figure 2: Cumulative returns on the SMB portfolio



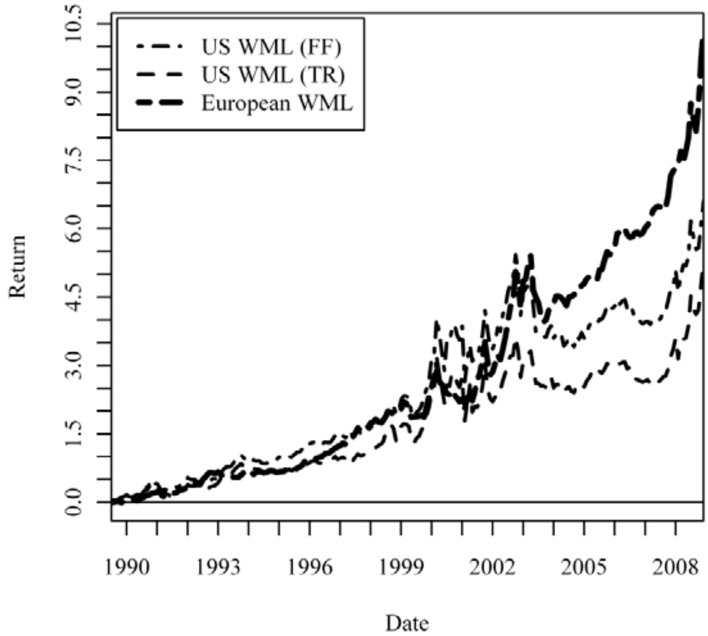
Note: This Figure shows cumulative returns on the SMB portfolio. We report the graphs for the SMB portfolio calculated from Thomson Reuters Datastream data (US SMB (TR)), for the U.S. SMB portfolio from Kenneth French's Website (US SMB (FF)) and for the SMB portfolio calculated from a sample of all European OECD countries (European SMB).

Figure 3: Cumulative returns on the HML portfolio



Note: This Figure shows cumulative returns on the HML portfolio. We report the graphs for the HML portfolio calculated from Thomson Reuters Datastream and Thomson Reuters Worldscope data (US HML (TR)), for the U.S. HML portfolio from Kenneth French's Website (US HML (FF)) and for the HML portfolio calculated from a sample of all European OECD countries (European HML).

Figure 4: Cumulative returns on the WML portfolio



Note: This Figure shows cumulative returns on the WML portfolio. We report the graphs for the WML portfolio calculated from Thomson Reuters Datastream data (US WML (TR)), for the U.S. WML portfolio from Kenneth French's Website (US WML (FF)) and for the WML portfolio calculated from a sample of all European OECD countries (European WML).

A. Supplementary Appendix

A.1. Constituent lists for the U.S. sample

We collect data from the following list types: research lists (FUSAA, FUSAB, FUSAC, FUSAD, FUSAE, FUSAF, FUSAG)[22], dead lists (DEADUS1, DEADUS2, DEADUS3, DEADUS4, DEADUS5, DEADUS6) and Thomson Reuters Worldscope lists (WSUS1, WSUS2, WSUS3, WSUS4, WSUS5, WSUS6, WSUS7, WSUS8, WSUS9, WSUS10, WSUS11, WSUS12, WSUS13, WSUS14, WSUS15, WSUS16, WSUS17, WSUS18).[23]

A.2. Constituent lists for the European sample

We collect data from the following lists: WSCOPEOE, ALLAS, DEADOE (Austria); WSCOPEBG, FBDO, DEADBG (Belgium); WSCOPEDK, FDEN, DEADDK (Denmark); WSCOPEFN, FFIN, DEADFN (Finland); WSCOPEFR, FFRA, ALLFF, DEADFR (France); WSCOPEBD, FGER1, FGER2, DEADBD1, DEADBD2 (Germany), WSCOPEIR, FIRL, DEADIR (Ireland); WSCOPEIT, FITA, DEADIT (Italy); WSCOPENL, FHOL, ALLFL, DEADNL (Netherlands); WSCOPENW, FNOR, DEADNW (Norway); WSCOPEPT, FPOM, FPOR, FPSM, DEADPT (Portugal); WSCOPEES, FSPN, DEADES (Spain); WSCOPESD, FSWD, DEADSW (Sweden); WSCOPESW, FSWS, DEADSW (Switzerland); WSCOPETK, FTURK, DEADTK (Turkey); WSCOPEUK, FBRIT, DEADUK (U.K.); WSCOPELX, FLUX, DEADLX (Luxembourg); WSCOPEGR, FGREE, FGRPM, FGRMM, FNEXA, DEADGR (Greece); WSCOPEHN, FHUN, DEADHU (Hungary); WSCOPEPO, FPOL, DEADPO

[22] Note that the lists FUSAA-FUSAG contain the same information as the FAMERA-FAMERZ lists, employed by Ince and Porter (2006, p. 465). However, FUSAA-FUSAG comprise only seven instead of twenty-six lists.

[23] The lists “FUSAA, FUSAB, …, FUSAG”, “DEADUS1, DEADUS2, …, DEADUS6” and “WSUS1, WSUS2, …, WSUS18” are special constituent list of all available firms availiable provided by Thomson Reuters Datastream and Thomson Reuters Worldscope.

(Poland); WSCOPECZ, FCZECH, FCZECHUP, DEADCZ (Czech Republic); FSLOVAK, FSLOVALL, DEADSLO (Slovakia); WSCOPEIC, FICE, DEADIC (Iceland).

These lists are basically selected from three categories: Worldscope lists, research lists and dead lists. Worldscope list begin with “WSCOPE” and end with a two-letter country code. Worldscope lists exist for all countries employed in this study, except Slovakia. Research lists aim to cover all equities listed in a specific country. Datastream provides two kind of those lists. The first kind begins with “ALL” and ends with a two-letter country code. The second kind begins with “F” and ends with a three-to-five letter country code. For all countries at least one of these lists is provided by Thomson Reuters Datastream and Thomson Reuters Worldscope Dead lists are used to keep the sample free of a survivorship bias, since the other lists typically contain only active stocks. Dead list begin with “DEAD” and end with a two-letter country code. Dead list exist for all countries employed in this study.

Besides these three list types we use additional lists for some countries. These lists are either main market lists (Portugal, Greece), second market lists (Portugal), or new market lists (NEXA - Greece). In addition we use the FCZECHUP list in case of the Czech Republic.

A.3. Additional Tables for empirical FF breakpoints

Table A.1: Breakpoints for the ten size portfolios by FF

	Mean	Median	Minimum	Maximum	Actual
$size_{BP1}$	0.50	0.49	0.39	0.59	0.45
$size_{BP2}$	0.63	0.63	0.53	0.70	0.60
$size_{BP3}$	0.71	0.72	0.62	0.76	0.70
$size_{BP4}$	0.77	0.78	0.70	0.82	0.75
$size_{BP5}$	0.83	0.83	0.77	0.86	0.80
$size_{BP6}$	0.87	0.87	0.82	0.90	0.85
$size_{BP7}$	0.90	0.91	0.87	0.93	0.90
$size_{BP8}$	0.94	0.94	0.92	0.95	0.93
$size_{BP9}$	0.97	0.97	0.96	0.98	0.96

Note: We use the number of portfolio constituents provided by Kenneth French to calculate size breakpoints, which apply to the whole sample (not only to NYSE stocks). We do so to find breakpoints for our sample, which are close to the FF breakpoints. The table shows the nine size breakpoints ($size_{BP1}$, … , $size_{BP9}$). We report mean, median, minimum and maximum of these empirical FF breakpoints. Furthermore we report the breakpoints actually employed in this study (column "actual"). The time period ranges from 07/1986 to 12/2008.

Table A.2: Breakpoints for the ten BE/ME portfolios by FF

	Mean	Median	Minimum	Maximum	Actual
BE/ME_{BP1}	0.16	0.17	0.10	0.22	0.10
BE/ME_{BP2}	0.27	0.27	0.20	0.33	0.20
BE/ME_{BP3}	0.36	0.36	0.29	0.42	0.30
BE/ME_{BP4}	0.45	0.44	0.36	0.51	0.40
BE/ME_{BP5}	0.53	0.53	0.44	0.60	0.50
BE/ME_{BP6}	0.62	0.61	0.52	0.69	0.60
BE/ME_{BP7}	0.70	0.70	0.62	0.76	0.70
BE/ME_{BP8}	0.79	0.79	0.75	0.84	0.80
BE/ME_{BP9}	0.89	0.88	0.84	0.92	0.90

Note: We use the number of portfolio constituents provided by Kenneth French to calculate BE/ME breakpoints, which apply to the whole sample (not only to NYSE stocks). We do so to find breakpoints for our sample, which are close to the FF breakpoints. The table shows the nine BE/ME breakpoints (BE/ME_{BP1}, … , BE/ME_{BP9}). We report mean, median, minimum and maximum of these empirical FF breakpoints. Furthermore we report the breakpoints actually employed in this study (column "actual"). The time period ranges from 07/1986 to 12/2008.

Table A.3: Breakpoints for the ten momentum portfolios by FF

	Mean	Median	Minimum	Maximum	Actual
mom_{BP1}	0.19	0.19	0.07	0.32	0.10
mom_{BP2}	0.30	0.31	0.14	0.45	0.20
mom_{BP3}	0.40	0.40	0.22	0.55	0.30
mom_{BP4}	0.48	0.48	0.30	0.64	0.40
mom_{BP5}	0.55	0.56	0.38	0.71	0.50
mom_{BP6}	0.63	0.63	0.46	0.77	0.60
mom_{BP7}	0.70	0.71	0.54	0.84	0.70
mom_{BP8}	0.77	0.78	0.62	0.89	0.80
mom_{BP9}	0.86	0.87	0.73	0.94	0.90

Note: We use the number of portfolio constituents provided by Kenneth French to calculate momentum breakpoints, which apply to the whole sample (not only to NYSE stocks). We do so to find breakpoints for our sample, which are close to the FF breakpoints. The table shows the nine momentum breakpoints (mom_{BP1}, … , mom_{BP9}). We report mean, median, minimum and maximum of these empirical FF breakpoints. Furthermore we report the breakpoints actually employed in this study (column "actual"). The time period ranges from 07/1986 to 12/2008.

Table A.4: Breakpoints for the six size and BE/ME portfolios of FF

	Mean	Median	Minimum	Maximum	Actual
$size_{BP1}$	0.60	0.60	0.54	0.65	0.60
$size_{BP2}$	0.75	0.76	0.70	0.79	0.70
$size_{BP3}$	0.85	0.86	0.82	0.88	0.80
$size_{BP4}$	0.93	0.93	0.92	0.94	0.90
BE/ME_{BP1}	0.27	0.27	0.20	0.33	0.20
BE/ME_{BP2}	0.45	0.44	0.36	0.51	0.40
BE/ME_{BP3}	0.62	0.61	0.52	0.69	0.60
BE/ME_{BP4}	0.79	0.79	0.75	0.84	0.80

Note: We use the number of portfolio constituents provided by Kenneth French to calculate size and BE/ME breakpoints, which apply to the whole sample (not only to NYSE stocks). We do so to find breakpoints for our sample, which are close to the FF breakpoints. The table shows four size breakpoints ($size_{BP1}$, …, $size_{BP4}$) as well as four BE/ME breakpoints (BE/ME_{BP1}, …, BE/ME_{BP4}). We report mean, median, minimum and maximum of these empirical FF breakpoints. Furthermore we report the breakpoints actually employed in this study (column "actual"). The time period ranges from 07/1986 to 12/2008.

Table A.5: Breakpoints for the six size and momentum portfolios of FF

	Mean	Median	Minimum	Maximum	Actual
$size_{BP1}$	0.62	0.62	0.54	0.68	0.60
$size_{BP2}$	0.77	0.77	0.71	0.81	0.70
$size_{BP3}$	0.86	0.86	0.82	0.89	0.80
$size_{BP4}$	0.94	0.94	0.92	0.95	0.90
mom_{BP1}	0.30	0.31	0.14	0.45	0.20
mom_{BP2}	0.48	0.48	0.30	0.64	0.40
mom_{BP3}	0.63	0.63	0.46	0.77	0.60
mom_{BP4}	0.77	0.78	0.62	0.89	0.80

Note: We use the number of portfolio constituents provided by Kenneth French to calculate size and momentum breakpoints, which apply to the whole sample (not only to NYSE stocks). We do so to find breakpoints for our sample, which are close to the FF breakpoints. The table shows four size breakpoints ($size_{BP1}$, …, $size_{BP4}$) as well as four momentum breakpoints (mom_{BP1}, …, mom_{BP4}). We report mean, median, minimum and maximum of these empirical FF breakpoints. Furthermore we report the breakpoints actually employed in this study (column "actual"). The time period ranges from 07/1986 to 12/2008.

A.4. Market capitalization of European OECD Countries as by June 2008

Table A.6: Countries ranked on market capitalization

Rank	Country	Market Cap.
1	UK	3,293,137.90
2	France	2,411,038.70
3	Germany	1,559,243.60
4	Switzerland	1,233,377.30
5	Spain	1,040,680.50
6	Italy	919,115.17
7	Netherlands	801,532.25
8	Sweden	455,478.12
9	Norway	399,487.29
10	Belgium	339,405.15
11	Finland	299,218.37
12	Denmark	242,390.28
13	Austria	241,672.29
14	Greece	232,578.61
15	Turkey	202,218.14
16	Poland	189,458.54
17	Portugal	122,562.82
18	Ireland	98,269.30
19	Czech Republic	81,379.51
20	Hungary	43,757.04
21	Luxembourg	40,036.11
22	Iceland	23,147.83
23	Slovakia	3,428.81

Note: The table shows all European OECD countries ranked by their total market capitalization (Market Cap.) in million US$ in June 2008. The data are from Thomson Reuters Datastream.

Working Papers of the Center of Economic Research at ETH Zurich

(PDF-files of the Working Papers can be downloaded at www.cer.ethz.ch/research).

10/140 L. Leinert
How do unanticipated discoveries of oil fields affect the oil price?

10/139 H. Gersbach, M. T. Schneider and O. Schneller
Basic Research, Openness, and Convergence

10/138 L. Bretschger and V. Kappel
Market concentration and the likelihood of financial crises

10/137 M. T. Schneider and R. Winkler
Growth and Welfare under Endogenous Lifetime

10/136 V. Hahn
Sequential Aggregation of Verifiable Information

10/135 A. Bommier, M.-L. Leroux and J.-M. Lozachmeur
On the Public Economics of Annuities with Differential Mortality

10/134 A. Bommier, A. Chassagnon and F. Le Grand
Comparative Risk Aversion: A Formal Approach with Applications to Saving Behaviors

10/133 A. Bommier and B. Villeneuve
Risk Aversion and the Value of Risk to Life

10/132 L. Bretschger and S. Valente
Endogenous Growth, Asymmetric Trade and Resource Taxation

10/131 H. Gersbach and N. Surulescu
Default Risk in Stochastic Volatility Models

10/130 F. Schwark
Economics of Endogenous Technical Change in CGE Models - The Role of Gains from Specialization

10/129 L. Bretschger, R. Ramer and F. Schwark
Long-Run Effects of Post-Kyoto Policies: Applying a Fully Dynamic CGE model with Heterogeneous Capital

10/128 M. T. Schneider, C. Traeger and R. Winkler
Trading Off Generations: Infinitely-Lived Agent Versus OLG

10/127 V. Kappel
The Effects of Financial Development on Income Inequality and Poverty

10/126 M. T. Schneider
The Larger the Better? The Role of Interest-Group Size in Legislative Lobbying

10/125 A. Ziegler
Individual Characteristics and Stated Preferences for Alternative Energy Sources and Propulsion Technologies in Vehicles: A Discrete Choice Analysis

10/124 P. F. Peretto and S. Valente
Resource Wealth, Innovation and Growth in the Global Economy

09/123 H. Gersbach and M. T. Schneider
Tax Contracts and Elections

09/122 V. Hahn
Why the Publication of Socially Harmful Information May Be Socially Desirable

09/121 A. Ziegler
Is it Beneficial to be Included in a Sustainability Stock Index? A Panel Data Study for European Firms

09/120 K. Pittel and L. Bretschger
The Implications of Heterogeneous Resource Intensities on Technical Change and Growth

09/119 E. J. Balistreri, R. H. Hillberry and T. F. Rutherford
Trade and Welfare: Does Industrial Organization Matter?

09/118 H. Gersbach, G. Sorger and C. Amon
Hierarchical Growth: Basic and Applied Research

09/117 C. N. Brunnschweiler
Finance for Renewable Energy: An Empirical Analysis of Developing and Transition Economies

09/116 S. Valente
Optimal Policy and Non-Scale Growth with R&D Externalities

09/115 T. Fahrenberger
Short-term Deviations from Simple Majority Voting

09/114 M. Müller
Vote-Share Contracts and Learning-by-Doing

09/113 C. Palmer, M. Ohndorf and I. A. MacKenzie
Life's a Breach! Ensuring 'Permanence' in Forest Carbon Sinks under Incomplete Contract Enforcement

09/112 N. Hanley and I. A. MacKenzie
The Effects of Rent Seeking over Tradable Pollution Permits

09/111 I. A. MacKenzie
Controlling Externalities in the Presence of Rent Seeking

09/110 H. Gersbach and H. Haller
Club Theory and Household Formation

09/109 H. Gersbach, V. Hahn and S. Imhof
Constitutional Design: Separation of Financing and Project Decision

09/108 C. N. Brunnschweiler
Oil and Growth in Transition Countries

09/107 H. Gersbach and V. Hahn
Banking-on-the-Average Rules

09/106 K. Pittel and D.T.G. Rübbelke
Decision Processes of a Suicide Bomber – Integrating Economics and Psychology

08/105 A. Ziegler, T. Busch and V.H. Hoffmann
Corporate Responses to Climate Change and Financial Performance: The Impact of Climate Policy

09/104 S. Valente
Endogenous Growth, Backstop Technology Adoption and Optimal Jumps

09/103 K. Pittel and D. Rübbelke
Characteristics of Terrorism

09/102 J. Daubanes
Taxation of Oil Products and GDP Dynamics of Oil-rich Countries

09/101 S. Valente
Accumulation Regimes in Dynastic Economies with Resource Dependence and Habit Formation

08/100 A. Ziegler
Disentangling Specific Subsets of Innovations: A Micro-Econometric Analysis of their Determinants

08/99 M. Bambi and A. Saïdi
Increasing Returns to Scale and Welfare: Ranking the Multiple Deterministic Equilibria

08/98 M. Bambi
Unifying time-to-build theory

08/97 H. Gersbach and R. Winkler
International Emission Permit Markets with Refunding