On the Accuracy of the SARD Approach Across Country Borders

Danish title: Præcisionsgraden af SARD-metoden på tværs af landegrænser

by
Emil Bryan Henningsen
(93108)

A thesis presented for the degree of
Master of Science in Finance and Accounting
(cand.merc.FIR)

Copenhagen Business School
Supervisor: Michael E. Jacobsen
September 16, 2019

Number of pages (characters): 78 (176.955)

Resumé

Formålet med denne afhandling er analysere præcisionsgraden af udvælgelsen af sammenlignelige selskaber til multipel værdiansættelse igennem 'Sum of Absolute Ranked Differences' metoden (SARD) udført på tværs af landegrænser. Den overordnede præmis for SARD er, at perfekte substitutter (peers) skal udvælges i forhold til sammenlignelighed i rentabilitet, risiko og vækst. Indtil nu er nøjagtigheden af SARD-metoden udelukkende blevet testet på en stikprøve indeholdende amerikanske selskaber. Dette muliggør videre undersøgelse til vurdering af præcisionsgraden af SARD-metoden globalt.

Min analyse er baseret på en international stikprøve indeholdende selskabsdata fra 27 OECD-lande henover perioden 2000-2019. Jeg undersøger den empiriske præcisionsgrad af SARD-metoden sammenlignet med peer udvælgelse baseret på brancheklassifikation. Jeg undersøger ligeledes hvorvidt en kombination af SARD og brancheklassifikation øger præcisionen i forhold til SARD udført på tværs af brancher. Endelig undersøger jeg hvordan nøjagtigheden af lande-specifikke SARD-estimater påvirkes igennem valget af stikprøve.

Mine resultater indikerer, at peer udvælgelsen igennem SARD er mere præcis end brancheklassifikation på globalt plan. Kombinationen af SARD og brancheklassifikation øger præcisionsgraden sammenlignet med SARD udført på tværs af brancher. Den landespecifikke analyse viser, at større internationale stikprøver er underordnede i forhold til præcisionen af estimater. I flere lande er udvælgelsen af peers mest nøjagtig når stikprøven begrænses til hjemlandet alene. Præferencen for udvælgelse i hjemlandet mindskes generelt igennem benyttelsen af den branche-baserede SARD-metode.

Præcisionsgraden varierer landene imellem. Mulige forklaringer er, at SARD påvirkes af branche-specifikke karakteristika og størrelsen på selskaberne. Den geografiske placering synes også at påvirke præcisionsgraden. Den branche-baserede SARD-metode nævnes som et muligt 'værn' imod lande-specifikke faktorer når peers udvælges på tværs af landegrænser. Nøjagtigheden af den globale SARD-udvælgelse synes altså påvirket af andre faktorer end ligheder i rentabilitet, risiko og vækst. Analytikere kan derfor træffe sub-optimale valg for globale selskaber hvis disse ikke adresseres forud for en eventuel implementering af SARD-metoden.

Contents

Chapter 1

Introduction

Prior literature on financial valuation generally acknowledge the multiple approach is the most popular technique to assess the value of stocks among analysts and portfolio managers (Demirakos, Strong, & Walker, 2004; Pinto, Henry, Robinson, & Stowe, 2015). In multiples-based valuations the relative price of a given target firm is given by the price of the comparable firms that share similar characteristics. If this relative pricing is reliable and agreed upon by market participants, then the value of firms should be identical to the multiples of comparable firms in regard to these characteristics. Thus, the multiple approach relies on the fundamental assumption that perfect substitutes should sell for the same price. This basic economic concept yields several practical issues for implementation. For instance, which scale factor should be applied (i.e. what multiples should be used), which selection variables capture these firm characteristics and ultimately how should comparable firms be identified? Prior research unanimously emphasizes the identification of these comparable firms as a crucial step in a successful multiple valuation (Plenborg & Pimentel, 2016). Three schools of thought on peer group selection exist. The first school argues peer group selection should be based on industry classification. Argument is that firms' operating in the same industries display economic similarities because of shared markets, identical product offerings and similar needs for production inputs. Intuitively, this should lead to some correlation in profitability, growth and risk for industry firms. Alford (1992) and Cheng and McNamara (2000) both found that peer selection based on a combination of industry classification and proxies for earnings and risk lead to more accurate estimates of valuation multiples.

The second school argues that pricing should uniformly rely on a peer group based on similar valuation fundamentals; that is similarities in profitability, risk and growth. Damodaran (2016) notes the following on the optimal peer group selection:

> A comparable firm is one with cash flows, growth potential, and risk similar to the firm being valued. It would be ideal if you could value a firm by looking at how an exactly identical firm - in terms of risk, growth and cash flows - is priced. Nowhere in

> this definition is there a component that relates to the industry or sector to which a firm belongs. Thus, a telecommunications firm can be compared to a software firm, if the two are identical in terms of cash flows, growth and risk

While explicitly making the case for the fundamental approach in peer selection, he does not pursue a more in-depth definition any further. The findings of Bhojraj and Lee (2002) and Dittmann and Weiner (2005) comply with the fundamental school of thought supporting that a peer selection based on similarities in economic characteristics is superior to a selection based on industry affiliation. Most recently, a third school has emerged, arguing the optimal peer group can be attained by comparing search traffic on websites. This generally implies investors' collective perceptions online is relevant when assessing the natural peers of a firm. Lee, Ma, and Wang (2015) support this claim by showing that two firms that are frequently co-searched by online users are fundamentally or economically similar (Plenborg & Pimentel, 2016). This peer selection technique is significantly more accurate than a peer selection based solely on industry affiliation.

Ultimately, financial literature has yet to provide on one straight answer on the preferable method to select the optimal set of peers in multiple valuations. However, some indications point in the favor of a collective peer selection combining both the industry- and fundamental ideas of thought. Lee et al. (2015) argue the current service and knowledge-based economy challenges the more traditional ideas of what constitute an industry and furthermore affects the accounting variables normally used to asses firm fundamentals of profitability and risk. Thus, new methods should be developed, and alternative data resources should be applied in future peer selection models (Plenborg & Pimentel, 2016).

Knudsen, Kold, and Plenborg (2017) propose the 'Sum of Absolute Rank Differences' (SARD) approach as a new method to identify the natural peers in multiple valuations. The model builds on intuition originating from clustering in data science and allows for an infinite number of selection variables. The authors select five variables that approximate valuation fundamentals in regard to profitability, risk and growth and identify firm-peers based on similarities in these across industries. Thus, the SARD approach initially agrees with the fundamental school of thought on peer selection as put forward in the quote of Damodaran. Knudsen et al. (2017) add to prior literature by showing the SARD approach yields considerably more accurate valuation estimates than when peers are identified through industry classification alone. In addition to the fundamental idea of SARD, the authors also propose the idea of a SARD selection performed within industry groups. Ultimately, this bridges the first- and second school of thought on peer selection offering a new promising identification method that is able to cluster similar firms based on well-defined variables including both industry- and fundamental characteristics of firms. Several research opportunities exist in regard to the overall assessment on the performance of the SARD approach. Most importantly, the model has only been tested on data consisting of US firms. Argument can be put forward, that it contradicts the overall idea of the fundamental approach when the potential peer pool is initially limited to firms within one country only. This should especially be the case for firms operating in smaller global markets, where a limited of number observations do not necessarily include the true 'natural' peers. Assuming the SARD selection variables are in fact the true proxies of the fundamentals driving market prices, the accuracy of SARD estimates should increase as the peer pool broadens. On the contrary, it can be argued, that differences across countries outweigh the benefits of the broader peer pool. Local regulation, national economic cycles, exchange rate risk and institutional background are just some examples of cross-border differences potentially influencing stock prices in effective global markets. If these effects are not captured by the selection variables of SARD this could lead to biased peer groups and wrongful valuation estimates. Furthermore, inconsistencies in the accounting practices applied in different countries would also result in a biased comparison of fundamentals.

Second, Knudsen et al. (2017) include data restrictions in order to assess the validity of SARD compared to an industry benchmark. To attain this, firms are eliminated if the size of industry groups are insufficient to perform the benchmark. Thus, the performance of SARD has yet to be assessed on data containing firms with a limited number of observations within an industry. This relates to the conclusion that a combination of SARD and industry might offer a preferred alternative to the regular SARD-models. To my knowledge, no direct tests comparing the two versions of the SARD approach have been performed up until now.

The need for further assessment on the overall performance of the SARD approach serves as the starting point in this thesis. If cross-country differences influence global market pricing, it would contradict the fundamental idea underlying SARD. Since a predominant part of all peer selection literature have been derived for US data only it is important to asses robustness across borders. Dittmann and Weiner (2005) address the potential influence from cross-country differences when identifying firm peers' in-between countries. They argue that one single peer selection method does not necessarily work best for all countries in a broad global sample. Variation in preferred selection variables as well as the countries to be compared with seems to arise. In order to validate the robustness of SARD, these implications need to be examined. Building on the findings of Knudsen et al. (2017), I seek to assess whether the effectiveness of SARD stands in a broad global setting when the potential peers are found across national markets with varying accounting standards and institutional backgrounds. Furthermore, the aim is to address the eventual limitations of SARD in markets smaller than the US.

The thesis is organized as follows. The following section presents my overall research question as well as the hypotheses to be assessed throughout the thesis. I further describe the scientific method underlying the research and present the delimitations of the thesis. **Chapter 2** describes the theoretical foundation underlying the fundamental- and industry-based school of thought. **Chapter 3** provides a literature review of the central paper on the SARD approach as well as the central empirical studies on peer selection. In **Chapter 4**, I describe my research design and methodological approach. **Chapter 5** presents the empirical results. This chapter further includes robustness checks of empirical findings. In **Chapter 6**, my findings are related to the ones found in existing literature. I interpret the results obtained and highlight the practical relevance of my work. **Chapter 7** concludes the thesis.

1.1 Research question

As a natural continuation of the previous section, the main purpose of this thesis is to provide an answer to the following:

- *How well does the SARD approach perform identifying comparable peer firms to attain a precise estimate of the EV/EBIT multiple in a broad global setting consisting of different national markets?*

In order to answer this general research question, I divide the research question into hypotheses based on the considerations presented above. The first part of the analysis investigates the accuracy of SARD when the potential peer pool includes a broad sample of global firms from all around the world. A predominant part of prior studies on various peer selection models are centered on the accuracy when evaluating US firms only. In order to examine the robustness of SARD across national markets, I set three distinct peer pools of different sizes: EU firms only[1], global firms excluding US firms only and all firms. Following Knudsen et al. (2017), I benchmark the accuracy of SARD with the performance of a selection based solely on industry classification. Thus, the structure of the first part of the analysis relies on the following hypothesis:

- *Hypothesis 1 (H1): The selection of comparable firms based on the SARD approach yields higher accuracy when estimating EV/EBIT than the selection based on industry classification.*

H1 addresses whether a peer group based on the SARD approach is more accurate vis-a-vís a peer group based on industry alone. Furthermore, I test for the accuracy in-between each individual SARD-model examined. This allows me to assess whether the inclusion of individual SARD selection variables adds incremental informative value to the overall model in the large samples analyzed. I also examine how well an industry-based SARD

[1] i.e. European Union firms

approach performs compared to regular SARD-models. I develop and test an industry level-up SARD-model allowing a limited number of observations within an industry. This selection model is constructed through the ideas put forward in Knudsen et al. (2017) and Alford (1992). Expectation is that the combination of SARD and industry is more accurate than the regular SARD-models:

- *Hypothesis 2 (H2): The industry level-up SARD approach yields more precise valuation estimates of EV/EBIT than the regular SARD approach.*

The assessment of **H1** and **H2** comprises the first part of my analysis. Utilizing the insights achieved from the initial hypotheses, the cross-country differences in estimation accuracy for 16 OECD countries is examined. Following the fundamental idea underlying SARD, expectation is that the probability of locating the 'true' natural peers is larger in a broader peer pool. Thus, the estimation precision should exhibit a positive correlation with the number of firms in the pool of potential peers. This should especially be the case for firms in smaller local markets containing a limited number of potential peers. Ultimately, I test whether:

- *Hypothesis 3 (H3): SARD estimates are more precise when the pool of potential peers is as large as possible (i.e. all firms in the sample) compared to estimations originating from smaller peer pools (i.e. when selection is limited to home country firms only).*

H3 is investigated by evaluating the country errors at various peer pools. In order to detect any potential impact from cross-country differences, these pools are set to either the home-, regional-, global- (excl. USA) or all firms level. An acceptance of **H3** indicates that the fundamental school of thought is robust across country borders when using the SARD approach. Rejection implies that SARD does not fully capture eventual cross-country differences and pre-defined peer pools should then be assessed individually for the different local markets examined. This would generally be in line with the findings of Dittmann and Weiner (2005), suggesting that selection models should be customized differently for firms in different countries. **H3** applies for both regular- and industry-based SARD-models. Generally, expectation is that the benefits of larger peer pools are greater for the industry-based models as compared to the regular models, since some national markets only include a limited amount of industry peers. SARD-models will be the consistent firm selection method applied in this thesis. Therefore, the performance of all alternative peer selection methods is excluded from the research design. SARD-model refers solely to the model as used in Knudsen et al. (2017).

1.2 The scientific method

I calculate the valuation error of estimates relative to the realized EV/EBIT-multiple for all firm-valuations on all SARD-models examined. Thus, as a prerequisite, the research design assumes realized market prices reflect the true firm values at a specific point in time, and the accuracy should be perceived as the ability to predict the observable market price. This is in line with a positivistic interpretation that market prices reflect aggregate supply and demand (i.e. markets are efficient). This thesis argues that comparable firms should ideally be found through a model based on the fundamental school of thought. However, it is a knowledge that models only partially comprise all information available. As a result, the driving factors for firm-peer identification can only be incompletely understood when assuming efficient markets. Thus, the research is characterized as neo-positivistic with a critical realistic ontology and a modified objective epistemology (Guba, 1990; Heldbjerg, 1997). The ontology acknowledges that one reality exists but can never fully be apprehended. The modified objective epistemology approximates objectivity since the assessment of the problem area will always be 'blurred' by own predispositions (Guba, 1990, p. 23). In contrast to sole positivism verifying a realistic phenomenon, neo-positivism seeks to explain and discover phenomena in a complex system (Heldbjerg, 1997).

The modified objective setting is accomplished through factual and realistic quantitative data (market- and accounting data) collected in an objective way. Complete objectivity cannot be achieved as some *manipulation* has been performed in regard to the handling of missing data and restricted samples. In all cases, the methodology strives to be 'as neutral as possible' which is in line with the modified objectivity (Guba, 1990, p. 21). I apply statistical measures to address the stated hypotheses which adds a positivistic dimension the research design.

Quality of the results arise through a systematic- and consistent scientific approach (Heldbjerg, 1997) with a high level of both validity and reliability. Internal validity is the compliance between theory and empirical data, whereas external validity refers to whether findings can be generalized to other contexts. Reliability refers to the credibility and trustworthiness of results implying future researchers should be able to replicate results if the analysis was repeated (Andersen, 2008). The thesis generally exhibits a high degree of internal validity. I ensure the compliance of the stated hypotheses and the research design by deriving the theoretical relation between accounting fundamentals and EV/EBIT. This establishes the theoretical link between the selection model being tested and the empirical data used for the analysis. Following the neo-positivistic approach, the sum of parts is not similar to each individual part (Heldbjerg, 1997). I accommodate this by conducting various sub-analysis and robustness checks to clarify the dynamics affecting the SARD approach in a broad global setting. All methodologically choices are critically based on existing literature. Thus, own predispositions are mitigated through the dependency on the 'critical tradition' within the scientific field. This increases the overall validity of the thesis.

Reliability is high. All selection models are mathematically formulated and can be accessed through the scripts enclosed (**Appendix B**). Furthermore, all data is publicly available. Thus, any future researcher can replicate and validate the results obtained. External validity is more uncertain because of data selection. This is especially the case when examining individual sub samples (e.g. country samples). Findings might not stand in alternative samples. In fact, it is explicitly pointed out how particular circumstances can lead to varying results. However, the neo-positivistic approach allows the dependency on the choice of sample which in turn should be perceived as a premise more than a methodological limitation. Knowledge is system based rather than universal (Heldbjerg, 1997) which is aligned with the intention of the study on the SARD approach in a broad global setting including different national markets.

1.3 Delimitations

The empirical analysis is restricted to the period 2000 to 2019. I only consider data from firms with headquarters in OECD countries. Focus is strictly on the identification of comparable firms. Thus, I do not address any other implementation issues concerning multiple valuation. For example, I restrict the analysis to include the EV/EBIT multiple only. I do not test alternative error measures or which averaging procedure that preferably should be used. All specific decisions related to these issues will be explicitly motivated in the methodology chapter. Any changes in these decisions are strictly due to robustness checks performed to validate the resistance of the empirical results obtained. I do not intend to change the original SARD approach as put forward by Knudsen et al. (2017). This applies for both the model development and selection variables included. Ideally, the SARD approach should be tested against any alternative selection model in a broad global setting. I limit my analysis to test: 1) whether the regular SARD approach offer a higher degree of valuation accuracy vis-a-vís estimates from peer groups that are based on industry classification and 2) whether the industry-based SARD approach yields more accurate estimates than the regular SARD approach. As such, I limit the thesis to focus on the fundamental- and industry-based schools of thought only. Thus, I delimit theory, literature and selection models from the ideas and suggestions arising from the third school of thought concerned with the selection of peer groups based on similarities in internet co-searches.

.

Chapter 2

Theoretical foundation

The purpose of this section is to establish the theoretical foundation underlying the fundamental- and industry-based schools of thought on peer group selection. Initially, it is shown how the value of the EV/EBIT multiple can be derived from a fundamental equity valuation model. Following this, the definition of industry used in this thesis is highlighted. I motivate the theoretical link between industry affiliation and inter-firm comparability in profitability, risk and growth.

2.1 EV/EBIT derived from the DCF model

The discounted cash flow model (DCF model) determines the value of a firm through the discounted value of the free cash flows (FCFF) which a firm is expected to realize in the future (Schreiner, 2009). As such, FCFF at time, τ, denotes the after-tax cash flow available to all investors in the firm, including both debt- and equity holders. FCFF is the share of the net operating profit after tax (NOPAT) not reinvested into the firm

$$NOPAT_\tau = EBIT_t \times (1 - t) \tag{2.1}$$

$$FCFF_\tau = NOPAT_\tau - \Delta\text{Invested capital}_\tau \tag{2.2}$$

Starting from equation **2.1**, FCFF is obtained through **2.2**, by deducting the cash withheld in the core operation of the firm denoted by ΔInvested capital$_\tau$. This includes adding back depreciation and amortization and deducting increases in net working capital and capital expenditures (CAPEX) (Schreiner, 2009). FCFF is used to pay both dividends and debt or retain cash for future activities. Thus, FCFF is not concerned with cash distribution to shareholders, but rather total cash generation to the enterprise (Gode & Ohlson, 2006, p. 3). Consequently, the DCF model estimates the enterprise value, consisting of the market value of equity and net interest-bearing liabilities (Petersen, Plenborg, & Kinserdal, 2017). Now assume a constant growth scenario in which firm capital structure, tax rate, reinvestment rate and cash flows are constant growing in perpetuity, T. Assume further an efficient capital market, defined as a market that fully reflects all information available to all market participants (Malkiel & Fama, 1970). In this setting the enterprise value of a firm, EV, at time τ, would be expressed by

$$EV_\tau = \sum_{\tau=1}^{\infty} FCFF_\tau(1+g_\tau)^T(1+WACC_\tau)^{-T} = \frac{FCFF_{\tau+1}}{WACC_\tau - g_\tau} \tag{2.3}$$

Where g_τ denotes the constant growth rate of FCFF. $WACC_\tau$ represents the weighted average cost of capital, consisting of both equity and net interest-bearing debt. The $WACC_\tau$ is calculated taking into account the relative weights of each component of the capital structure consisting of either the market value of equity (MVE) or the market value of net-interest bearing debt financing (NIBD). Thus, the WACC is expressed by the required return on equity, R_E, the cost of debt financing, R_D, as well as the corporate tax rate, t

$$WACC_\tau = \frac{NIBD_\tau}{NIBD_\tau + MVE_\tau} \times R_{D,\tau} \times (1-t) + \frac{MVE_\tau}{NIBD_\tau + MVE_\tau} \times R_{E,\tau} \tag{2.4}$$

The WACC is generally perceived as the appropriate discount rate when estimating the value of a firm through future cash flows in DCF.

The fundamental value drivers underlying the EV/EBIT multiple can be derived through equation **2.2** and **2.3** with the addition of two new terms. First, denoting the rate of reinvestment in firm core operations by θ, seen as the proportion of NOPAT 'ploughed back' into the firm at time τ

$$\theta_\tau = \frac{\Delta\text{Invested capital}_\tau}{NOPAT_\tau} \tag{2.5}$$

Second, introducing the return on invested capital (ROIC) as the rate of return for each dollar invested in core operations only from period $\tau - 1$ to τ (Schreiner, 2009, p. 34)

$$ROIC_\tau = \frac{NOPAT_\tau}{\text{Invested capital}_{\tau-1}} \tag{2.6}$$

In the constant growth setting with a fixed reinvestment rate, a logical relation exists between ROIC and θ and the growth rate: $g_\tau = ROIC_\tau \times \theta_\tau$ and $\theta_\tau = \frac{g_\tau}{ROIC_\tau}$. Recalling that FCFF can be explained as NOPAT less reinvestment in core operations, this term can now be expressed through NOPAT and θ. Recall further, that NOPAT is EBIT less corporate taxes. Thus, equation **2.3** can then be restated

$$\begin{aligned} FCFF_\tau &= NOPAT_\tau \times (1-\theta_\tau) \\ &= EBIT_\tau \times (1-t_\tau) \times (1-\frac{g_\tau}{ROIC_\tau}) \end{aligned} \tag{2.7}$$

$$EV_\tau = \frac{EBIT_{\tau+1} \times (1-t_\tau) \times (1-\frac{g_\tau}{ROIC_\tau})}{WACC_\tau - g_\tau} \tag{2.8}$$

Ultimately, an expression for the EV/EBIT multiple can be obtained by dividing $EBIT_\tau$ on both sides. This yields the following relation[1]

$$\frac{EV_\tau}{EBIT_\tau} = \frac{(1+g_\tau) \times (1-t_\tau) \times (1-\frac{g_\tau}{ROIC_\tau})}{WACC_\tau - g_\tau} \tag{2.9}$$

The derivation of the EV/EBIT multiple shows the intrinsic value drivers through the DCF model, which is a central fundamental equity valuation model in finance literature (Petersen et al., 2017). Equation **2.9** shows how EV/EBIT can be seen as a function of profitability (ROIC), risk (WACC), growth (g) and the corporate tax-rate (t) through DCF. By formulation, the value of the multiple will, all else being equal, increase with growth (g) and decrease with higher levels of cost of capital (i.e. higher $WACC$). Considering the assumption, that g is a product of ROIC and the reinvestment rate, θ, this is equivalent to stating that EV/EBIT, all things being equal, will increase as ROIC (i.e. profitability) increases. In addition to this, the corporate tax rate affects the value negatively. In sum, this is identical to the well-known expression of *Economic Value Added* (EVA) denoting firm excess return by $EVA_\tau = (ROIC_{after\ tax,\tau} - WACC_\tau) \times \text{Invested capital}_\tau$ (Petersen et al., 2017). The estimated market value will be above the book value of invested capital when the present value of expected EVAs is positive and below book value when the present value of expected EVAs is negative. Thus, classic valuation theory underlie the fundamental school of thought on peer selection: if the market value of equity and debt approximates the present value of expected cash flows based on profitability, risk and growth, these variables should explain a significant portion of the variation in EV/EBIT multiples across firms (Bhojraj & Lee, 2002). Peer groups should therefore be selected by approximating these variables the best way possible.

This thesis is limited to the analysis of the EV/EBIT multiple only. However, before moving on it is important to leave some theoretical comments on the value drivers of equity-based multiples such as price-earnings (P/E) and price-to-book (P/B). Both multiples are highly used in practice as well as in the prior literature on identification of comparable firms. Theoretical derivations of these are beyond the scope of this thesis. However, since P/E is included in a later robustness check, it is beneficial for the reader to know

[1] The derivation of the EV/EBIT multiple follows different passages in Schreiner (2009) and Petersen et al. (2017).

that equity-based multiples can also be shown as direct functions of profitability, risk and growth. Both P/B and P/E depend on the return on equity (ROE), risk referred to as the required rate of return, r_e, and the future constant growth rate (g). A fundamental difference between equity- and EV-based multiples is that EV-based multiples value the free cash flow available to the total firm (debt and equity), whereas the equity-based multiples only are concerned with the cash distribution to equity holders.

2.2 Industry-peer comparability

2.2.1 Definition of an industry

An industry is defined as a group of firms that offer a product or service or class of products or services that are close substitutes for one another (i.e. high and positive cross elasticity of demand) (Kotler, Keller, Brady, Goodman, & Hansen, 2016). The nature of this definition is widely used in both finance, management and economics. From an empirical perspective, various industry classification systems deliver a systematic mapping of firms' industry affiliation[2]. Nightingale (1978) criticizes the empirical use of these international identification systems because the theoretical boundaries of an industry is not strictly defined. For instance, geographies and varying activities among firms are two examples where a pre-defined industry class can be perceived as highly ambiguous. While acknowledging the potential pitfalls in the theoretical industry definition, it is beyond the scope of this thesis to discuss these issues any further.

In this thesis, firms' industry affiliation will be defined through the Global Industry Classification Standard (GICS). According to the CICS guidebook, the sources of information used to classify firms' industry affiliation include both quantitative- and qualitative data from annual reports, financial statements and investment research reports (Standard & Poor's, 2018). As most industry classification systems, GICS is constructed as a hierarchical system including four levels of mapping: sub-industry, industry, industry group and sector. All firms classified are assigned an 8-digit GICS code. This 8-digit code imply the highest level of detail within the classification system which is the sub-industry of a given firm. Removing the last two digits gives the 6-digit GICS code representing the industry. Industry group and sector follow as the 4- and 2-digit code respectively. Thus, all firms belong to a designated 'industry' at all four levels in the GICS system and are linked together in regard to the GICS classification levels. As of 2018, the GICS system includes 11 sectors, 24 industry groups, 69 industries and 158 sub-industries[3]. Bhojraj, Lee, and

[2]Central identification systems include: North American Industry Classification System (NAICS) developed by USA, Canada and Mexico; Standard Industry Classification System (SIC) delivered by U.S. Central Statistical Board, superseded by NAICS and has not updated since 1987; Industry Classification Benchmark (ICB) developed by Dow Jones Indexes and FTSE.

[3]https://www.msci.com/gics

Oler (2003) found the identification of comparable firms through the GICS system yields a lower valuation error compared to an identification based on alternative classification systems.

2.2.2 Industry and the value drivers of multiples

Various theoretical arguments can be put forward as to why firms competing in an identical industry necessarily must converge in profitability, risk and growth. Porter (2008) argues industry structure drives competition and profitability within a given industry. Taking the perspective of a firm present in an industry, the threat of entry from new competitors can limit the potential profit in this industry. Thus, barriers of entry are sources of intra-industry similarities affecting the overall profitability for all firms in the industry through the threat of entry[4]. For instance, lower unit cost through economies of scale reduces the industry competition from aspiring entrants not entering on a large scale (i.e. high initial investment). In contrast, established firms will, all things being equal, enjoy the benefits of higher cost margins and lower price competition. Thus, concentration of competition is expected to define the gap between revenues and costs for firms within industries. The industrial economics literature generally accepts that this positive correlation exists (Peltzman, 1977; Collins & Preston, 1966; Clarke, Davies, & Waterson, 1984).

Industry affiliation leads to similar properties in regard to both external- and strategic risk factors influencing the industry members (Petersen et al., 2017). External risk factors typically originate from risks beyond the control of firms. Examples of these could be commodity price changes, overall economic conditions, political instability or commercial law. Second, strategic risks stem from the overall competitive environment in the industry. For instance, intense price competition or dependency on too few customers or suppliers. Additionally, firms delivering identical output (i.e. product or service) will naturally share some market risks on the demand side. A sudden event that temporarily increases or decreases the demand for a specific good or service should cause correlating revenues, leading to some similarity in the uncertainty for future cash flows for industry firms. These transitory shocks could also occur on the input level, where intra-industry correlation for the cost of common supplies would affect all firms in the industry. Shared input risks could also stem from either labor or capital requirements. Salary expenses as well as the cost of machinery (i.e. interest levels) is expected to converge for firms delivering an identical product or service. Finance literature generally acknowledge that the required return across industries must vary due to differences in risk[5] (Petersen et al., 2017).

Products undergo various stages of demand over time. These stages are often denoted

[4]These include, but are not limited to, supply-side economies of scale, demand side economies of scale, customer switching costs, capital requirements, unequal access to distribution channels and restrictive government policies.

[5]Damodaran's homepage (NYU) models industry risk through unlevered beta levels.

as the life cycle, which is often presented through four distinct phases: introduction, growth, maturity and decline (Petersen et al., 2017). When similar considerations are made for businesses in general, some growth prospects in the short term should theoretically be similar at the industry level. In relation to this, Aghion, Bloom, Blundell, Griffith, and Howitt (2005) argue that market concentration has a direct influence on the innovation within a sector (i.e. market or industry). When competition is low, the incremental profits from innovation leads to higher levels of innovation, whereas when competition is already high, the rate of innovation seems to decrease. Thus, some convergence in regard to the short-term growth of future cash flows should be expected for industry firms. In the long run however, the potential growth will be affected by various uncertain factors such as the underlying market growth, rivalry among competitors and relative competitive strengths arising from innovation and corporate strategy. Ultimately, most literature conclude, that no firm can grow more than the total growth in the world economy in the long run (Petersen et al., 2017). In addition to the presumed convergence of value drivers for profitability, risk and growth within industries, the comparability of these variables might also be enhanced since firms in the same industry often use similar accounting methods (Alford, 1992).

Chapter 3

Related literature

This chapter reviews the related existing literature on peer group selection. Initially, the central studies supporting both the fundamental- and industry-based schools of thought on peer selection is reviewed. Hereafter, the original study proposing the SARD approach is presented and related to the existing literature.

3.1 Coverage of literature on peer group selection

3.1.1 Evidence in favor of industry classification

One of the very first investigations within the area of peer selection was Alford (1992), who tested the accuracy of various methods when selecting comparable firms based on industry SIC codes and proxies for earnings and risk. His findings were based on a cross-sectional analysis of a broad US data set extracted from NYSE, ASE and OTC. His sampling period included three years 1978, 1982 and 1986. The performance of selection models was investigated by estimating the absolute percentage error compared to the observable market prices for all firms in each year. Overall, accuracy was then summarized by averages of the yearly cross-sectional median errors and non-parametric statistical tests. Alford (1992) concluded that comparable firms should be selected through either industry membership or a combination of industry and proxies for risk or earnings. Individually, the proxies for valuation fundamentals did not seem to capture as much information as industry classification alone. When selecting firms based on either ROE or total assets (TA) individually, the precision of estimates was considerably worse, than when industry classification was applied. Furthermore, selections based on analysts' growth forecasts led to an improvement compared to ROE or TA. Also, the performance of selection models combining industry and ROE or TA was statistically indistinguishable from the single industry models. Little to non-incremental predictive accuracy was attained when combining analysts' growth forecasts and industry. Thus, industry classification capture much of the same information as the valuation fundamentals. Alford (1992) also tested for the importance of industry fineness. The 3- and 4-digit SIC codes increased estimation precision significantly compared to fewer SIC codes, while being statistically indistinguishable from each other. Finally, Alford (1992) examined whether larger firms yield higher accuracy than smaller firms. The results implied an increase in estimation accuracy for larger firms for all selection models examined.

The findings of Alford (1992) was limited to the P/E multiple only. These findings were confirmed and expanded through Cheng and McNamara (2000). Through a large US sample from 1973 to 1992, the authors found that the combination of industry classification and ROE was more accurate than all other selection rules tested for both the P/E and P/B multiple. Selections based solely on fundamentals were generally more inaccurate than when industry classification was used. As an extension to Alford (1992), the authors provided statistical evidence for the superiority of the combination of industry and ROE compared to the selection based on industry alone. This result was explained trough the larger sample (twenty years of data) used in Cheng and McNamara (2000). Furthermore, Cheng and McNamara (2000) also investigated the accuracy of a combined pricing method, averaging the estimates from the P/E and P/B valuations altogether (the P/E-P/B method). This method was generally more accurate than when assessing P/E or P/B individually. The P/E-P/B method was most accurate through selections based on industry classification alone. Thus, Cheng and McNamara (2000) concluded that industry classification was the superior single firm-peer identifier.

3.1.2 Evidence in favor of fundamental-based identification

Recalling the proposition depicted in the expression for EV/EBIT through equation **2.9**, the underlying value drivers should be highly relevant in the identification of peers in an efficient capital market. Bhojraj and Lee (2002) revise the idea of selecting comparable firms through industry classification and argue the ideal peer selection should rather be based on value drivers approximating expected profitability, growth and cost of capital (i.e. risk). In order to do so, they apply a regression approach to predict the 'warranted multiple' for each firm and then, in a second step, select comparable firms based on the relative closeness to this. They perform their study on a large cross-sectional sample of US firms from the S&P 1500 in a period ranging from 1982 to 1998. They limit their work to include the P/B and EV/SALES multiples only. Contrary to Alford (1992) and Cheng and McNamara (2000), Bhojraj and Lee (2002) measure the accuracy of the warranted multiple approach by estimating multiples and analyzing their predictive power on future multiples. This is done by regressing the forward value of the observable multiple (one-, two- and three years ahead) by the value of the predicted multiple estimated through the fundamental value drivers. Accuracy is then perceived as the amount of cross-sectional variation captured by the regression. The authors find the warranted multiple performs significantly better than both industry and size when predicting the forward value of the observable multiples in the cross-sectional regressions. Furthermore, through the second step of the analysis, they find estimations based on the nearest warranted multiples offer a sharp decrease in the percentage estimation errors as compared to industry-based multiples. Interestingly, in the second step, errors were lower when estimations were based on the actual multiples of the peers selected vis-a-vís when estimates were based on the warranted multiples of the selected peer group. In the context of valuations in an international setting, Bhojraj et al. (2003) suggest the warranted multiple method also work in a sample consisting of firms from the G7 countries within a 10-year period (1990-2000). Apart from fundamentals, they suggest that both industry- and country membership capture some of the cross-country variation that exist. They find strong explanatory power for fundamentals and industry controls[1], while controls for country membership yield insignificant informative value. Collectively, the findings of Bhojraj and Lee (2002) and Bhojraj et al. (2003) confirm that fundamentals capture a significant proportion of cross-sectional variation for the regressed P/B and EV/SALES multiples.

More comparable to the methods applied in this thesis are Herrmann and Richter (2003), S. Nel, Bruwer, and le Roux (2014) and Dittmann and Weiner (2005). All of these studies focus on international data and assume accuracy to be the ability to predict the observable price in the market at the time of valuation (assumption in line with Alford (1992)).

Herrmann and Richter (2003) examine a sample of large European- and US firms. They show that peer groups based on firms with a deviation less than 30% from the target firm in regard to ROE and analysts' growth forecast are superior to groups based on industry selection. They do not find any statistically significant improvement in the valuation accuracy when combining industry information and fundamentals implying that industry does not offer additional information beyond that already controlled for through the fundamental selection.

In order to validate whether the fundamental school of thought holds in an emerging market setting, S. Nel et al. (2014) conduct a comprehensive analysis on peer group selection in the South African market. In total, the authors examine seven fundamental-based selection rules through three variables: ROE, TA and revenue growth. As such, three single models and four combined models are examined. The authors generally find the combined models are superior to single models. S. Nel et al. (2014) do not compare results with a selection based on industry classification.

Dittmann and Weiner (2005) denote the only study that, to my knowledge, includes a broad global sample of OECD firms. Thus, this particular paper is of high relevance to the

[1] The harmonic mean of multiples for all the firms with the same two-digit SIC code (Bhojraj et al., 2003).

methods applied in this thesis. The authors examine whether the results of Alford (1992) stand when estimating the EV/EBIT multiple for firms from 16 different OECD countries (15 EU-member countries and USA) in a period from 1993 to 2002. The authors generally find that estimation errors are lowest when firms are selected based on the return on assets (ROA) or a combination of ROA and TA[2] for all 16 countries. Both selection rules are found to be superior to the industry-selection based on SIC codes. The authors stress that 'a selection method is a combination of the pool of firms from which comparables are chosen and the rules that describe how comparables are selected from this pool' (Dittmann & Weiner, 2005, p. 5). Hence, the authors test for the impact of various selection pools for all country-firms examined. They generally focus on three types of groupings: firms from the same country, from the same region, or from all OECD countries. The authors find the choice of peer pool genuinely affects estimation precision across countries. For four countries (the US, UK, Denmark and Greece), median errors as well as statistical tests indicate the optimal selection occurs when the peer pool is restricted to the home country only. For the remaining twelve countries, estimation errors are smaller when selection happens at the region- or OECD level according to the median errors incurred. The mean error generally seems higher in smaller peer pools. The opposing effects detected between mean- and median estimation errors are ascribed to the effect from outliers, that are perceived more likely in small samples than in large samples. Opposed to Alford (1992), Dittmann and Weiner (2005) do not examine selection rules based on the combination of industry and fundamental value drivers.

3.2 The SARD approach

Knudsen et al. (2017) initially agree with the fundamental school of thought, arguing comparable firms should be selected through similarities in proxies for profitability, growth and risk. The authors argue that prior selection methods lack in practical efficiency because peer firms are always selected at the '*intersection*' of the most comparable firms in terms of the fundamental proxies. This '*intersection*' narrows as more proxies are included and limits the flexibility and efficacy of selection models due to data restrictions and potential peers in the market. S. Nel et al. (2014) confirm these issues in their South African study. Here, the peer groups produced by a three-variable combination of ROE, TA and revenue growth were highly insufficient and excluded from the analysis.

The SARD approach offers a selection method that, in principle, allows for an infinite number of proxies for profitability, growth and risk (i.e. selection variables). If we consider a matrix of sum ranked differences for given set of variables between each firm, the natural peers are classified as the ones with the lowest sum of ranked differences across the total

[2]Alford (1992) applied ROE instead of ROA as the measure for profitability. Dittmann and Weiner (2005) do not address this difference. Once again, TA denotes total assets.

peer pool:

$$SARD_{i,j} = |r_{X,i} - r_{X,j}| + |r_{Y,i} - r_{Y,j}| + ... + |r_{Z,i} - r_{Z,j}| \tag{3.1}$$

Equation **3.1** denotes the 'SARD-score' which is the sum of ranked differences between firm i and j. $r_{X,i}$ is the rank of firm i in terms of selection variable X, $r_{X,j}$ is the rank of firm j in terms of X, and so on. When a potential peer yields a low SARD-score compared to a given target firm, the SARD approach suggests this peer and the target firm share similarities in regard to the designated set of selection variables (Knudsen et al., 2017).

The authors note the flexibility and simplicity of the model adds to the usefulness of the approach opposed to prior peer selection models. They examine the superiority of the model through a research design based on valuations conducted on a large cross-sectional US sample over twenty years (1995-2014) and evaluate the central tendency of error terms based on the axiom that the observed market price is the true price. Model performance is validated by non-parametric statistical tests between the estimation errors of SARD-models compared to an industry benchmark. The benchmark is a random selection of firms within the identical GICS industry (i.e. GICS 6).

The SARD approach offers a significant reduction in estimation error opposed to the industry benchmark. The authors use the following selection variables: ROE, Net Debt/EBIT, Size, Growth and EBIT-margin. By adding the selection variables sequentially and testing for error differences in-between models, they confirm each individual selection variable captures useful information when selecting the natural peers of a firm. By conducting their analysis on P/E, P/B, EV/SALES and EV/EBIT, the authors demonstrate that different multiples are influenced differently by each selection variable. However, in all cases, the SARD approach offers a significant reduction in estimation error opposed to the industry benchmark. EV/EBIT is most accurate of the four multiples examined.

Through model robustness checks, the precision of SARD is found to be independent of firm size (measured by the sub-indices in S&P 1500), but heavily influenced by industries. Thus, the authors suggest 'that selection variables should be tailored to both the multiple applied and the industry examined to achieve more accurate valuation estimates' (Knudsen et al., 2017, p. 101). Additionally, the authors perform robustness checks on the number of estimation peers. An estimation peer group of six firms is found to be adequate to minimize the estimation errors.

In line with the findings of Cheng and McNamara (2000) and Bhojraj et al. (2003), the authors propose that SARD should be used in conjunction with industry. Based on the comparison of median errors, the authors show how the estimation precision of P/E and EV/EBIT increase when the SARD approach is used within industries.

Chapter 4

Data, methodology and selection models

This chapter describes the methods applied in relation to the sampling of data, development of selection models and the evaluation of estimation errors. The research design is similar to the ones of Knudsen et al. (2017) and Dittmann and Weiner (2005). I value all firms as of March 31st each year in a period ranging from 2000 to 2019. In the first section, I address central implementation issues such as the choice of multiple, the number of estimation peers and the averaging procedure. Second, I describe the selection models used. Third, I describe my sample selection. The chapter finishes off by describing how model errors are evaluated and compared against each other.

4.1 Estimating multiples

4.1.1 The choice of multiple

No best practice exists in regard to one 'obvious' choice of multiple. My analysis is limited to the EV/EBIT multiple only. This choice is primarily made in order to attain comparable results with both Knudsen et al. (2017) and Dittmann and Weiner (2005). However, the choice of multiple is also theoretically related. Compared to P/E, EV/EBIT is expected to be less susceptible by the impact of non-operating items, such as restructuring charges and write-offs. Additionally, the value of P/E is systematically affected by firms' capital structure. If the required return on equity is higher than the capital costs of debt, firms can artificially increase the P/E ratio by swapping debt for equity (Koller, Goedhart, & Wessels, 2005). EV/EBIT represents the relative value of the total firm, independent of capital structure decisions. Furthermore, both Herrmann and Richter (2003) and J. Liu, Nissim, and Thomas (2002) found that earnings-based multiples leads to lower forecast errors than the ones based on book values. Acknowledging P/E is widely used in prior literature, I use this multiple for a model robustness check.

4.1.2 Size of peer group and measurement of averages

All selection methods are based on identical implementation decisions in regard to both the number of peers selected as well as the averaging measure of the predicted multiple. A predominant part of prior literature has arbitrarily agreed on six peers as an adequate amount of estimation peers (Alford, 1992; Cheng & McNamara, 2000; Knudsen et al., 2017). Thus, the default number of estimation peers is set to six. Cooper and Cordeiro (2008) examined how the accuracy of a multiples-based valuation changes as the number of comparable firms used for estimation varies. Here, they found that ten comparable firms lead to more accurate results. Knudsen et al. (2017) argue that six estimation peers are suitable for SARD. Thus, I later perform robustness checks on the number of estimation peers.

Different studies apply different averaging procedures. Alford (1992) and Cheng and McNamara (2000) applied the median when predicting multiples from the located peer group. Baker and Ruback (1999) suggest the harmonic mean is most accurate when estimating industry multiples opposed to both median and the arithmetic- and value-weighted mean. The usefulness of the harmonic mean was further confirmed in J. Liu et al. (2002). Theoretically, the usefulness of the harmonic mean is primarily driven by the fact, that the measure generally mitigates the risk that the estimated value is affected by outliers (Petersen et al., 2017). The harmonic mean is expressed as

$$\widehat{M} = \frac{N_{np}}{\sum_{p=1}^{N_{np}} \frac{1}{M_{np}}} \tag{4.1}$$

Where $\widehat{M}$ is the estimated multiple, N_{np} is the number of the natural peers found through SARD (always equal to six) and M_{np} is the multiple for each of the natural peers located. Both Dittmann and Weiner (2005) and Knudsen et al. (2017) apply the harmonic mean when estimating multiples. In order to retain comparability with these related studies, all estimates in this thesis are based on the harmonic mean.

4.2 Identification of peers through SARD

In essence, the developed SARD algorithm is a functional implementation of equation **3.1** presented in **Section 3.2**. The natural peers are found by minimizing the following

$$\text{SARD peers}_{x,y} = \min \sum_{i=1,y=1}^{N_i,N_y} rZi_{x-y} \tag{4.2}$$

Where rZ denotes each ranked (r) selection variable (Z), i. Thus, for a given target firm x, the SARD selection happens through the sum of N absolute ranked differences of i selection variables across an array of y comparable firms. N varies in regard to the number of designated selection variables as well as the potential peers in a given sample. I obtain Zi_{x-y} by ranking N_i arrays of selection variables in descending order, assigning 1 to the largest value across each array and vice versa. Identical ranks are assigned through tiebreakers if selection variables are equal. Peers are always identified at SARD $\text{peers}_{x,y} > 0$ since the SARD-score of the target firm, x, will always attain a value of 0. Thus, each target firm will be excluded from the estimation sample in the SARD selection (i.e. out of sample). Minimization happens through the selection of the six firms with the lowest SARD-score (i.e. lowest absolute difference). Equation **4.2** is implemented using Python scripts. All essential programming code can be accessed through my **GitHub repository**[1].

4.2.1 SARD selection variables

In principle, the implemented SARD algorithm allows for an indefinite number of ranked selection variables and comparable firms. The selection parameters included in my models correspond fully with Knudsen et al. (2017). In the following, I provide the motivation for the inclusion of each selection variable.

ROE approximates profitability. Alford (1992) motivates the choice of ROE as a predictor of future growth, since earnings and reinvestment are theoretically correlated. In relation to the theoretical derivations presented in **Section 2.1**, one could argue the usage of ROIC instead of ROE would be theoretically more correct when valuing the EV/EBIT multiple. However, this would also offset practical implications such as the derivation of the invested capital for all sample firms. The definitions and classifications necessary seem just as theoretically challenging as the usage of ROE on EV/EBIT. Furthermore, both Herrmann and Richter (2003) and Knudsen et al. (2017) show that a selection based on ROE have a decreasing effect on estimation errors of EV/EBIT.

EBIT-margin is applied as a second profitability measure. Knudsen et al. (2017) found that this variable was especially important for EV/SALES. However, conceptually EBIT-margin should also be useful when comparing firm profitability across borders. Recall, that EBIT is a non-tax amount that is less likely to be affected by the variation in corporate tax rates in different countries. Thus, when used in conjunction with ROE, EBIT-margin should add to the informative power of selection models for EV/EBIT in a broad global setting with different national markets.

[1]Can be accessed through: https://github.com/ebhen/sard_thesis_journey. All material is also available on the external appendix. **Appendix B** provides an introduction to GitHub. Essential code snippets are also documented.

Net debt/EBIT is used as a proxy for risk. Knudsen et al. (2017) argue that 'Net debt/EBIT is an integrated component of credit analysis'. Intuitively, there should be a link between market perception of firm risk and the ability to pay back on interest bearing liabilities. However, little empirical evidence exists on the specific measure. In relation, Bhojraj and Lee (2002) did not find any statistically significant correlation between EV/SALES and book leverage, which theoretically should be expected to somehow transcend to the EV/EBIT multiple (Petersen et al., 2017, p. 321).

Size denotes the second proxy for risk. As initially observed by Banz (1981), smaller firms tend to have higher returns than larger firms after controlling for market risk (Crain, 2011). Size has consistently been used as a surrogate for risk in numerous prior peer selection models (Alford, 1992; Cheng & McNamara, 2000; Dittmann & Weiner, 2005). In all cases, total assets (TA) was used as a proxy of firm size with mixed results. Knudsen et al. (2017) use firm market capitalization and show that the inclusion of this variable increases estimation accuracy[2]. In order to retain comparability of results with the original SARD approach, the SARD-models developed in this thesis proxy size through firm market capitalization.

Growth is approximated by the calculated ratio of the median of analysts one- and two-year forecasts from the Institutional Brokers Estimates System (I/B/E/S). As shown in **Section 2.1**, growth is a fundamental driver of firm value. Empirical evidence suggests the I/B/E/S estimates provide a suitable proxy for this very intangible value driver. For instance, Bhojraj and Lee (2002) found strong statistical evidence on the positive correlation between analysts' growth forecasts and the value of multiples.

Summarizing the sections above, **Table 4.1** provides an illustrative example of the SARD algorithm applied. Assuming a whole sample universe consisting of only ten firms as of March 2019, I set out to select the natural peers through SARD by the inclusion of selection parameters: ROE, Net Debt/EBIT and Size. **Panel A** and **B** fully capture the essence of the selection procedure. **Panel A** reports the ranked selection variables for all firms. As an example, Heineken NV is ranked the highest (1) on Net Debt/EBIT with an estimated ratio of 3.7 while ranking the fifth (5) on Size with an estimated market capitalization of 61 billion USD. **Panel B** illustrates the SARD approach applied for Unilever Plc. The firms enclosed (excluding Unilever) is the six natural peers based on the SARD-score. For instance, the SARD-score between LVMH and Unilever is obtained by $|1-6|+|2-1|+|3-4|=7$. The algorithm predicts Unilever EV/EBIT to be 13.936 through the harmonic mean of the six natural peers, which results in a absolute valuation error of 5.20%. **Table 4.1** is only intended to serve as an illustrative example of the implemented SARD algorithm described.

[2]This leaves a logical iteration problem for the current SARD approach performed on firms where prices are not known. Ultimately, the discussion on this matter is beyond to theoretical scope of this thesis.

Table 4.1: SARD selection model, exemplified

Panel A: Firm ranks based on ROE, Size and Net Debt/EBIT									
Total firms in sample universe		EV/EBIT	*Z1 ROE*	*Z2 Size ($B)*	*Z3 Net Debt/ EBIT*	r_{Z1}	r_{Z2}	r_{Z3}	
Bridgestone Corp	Consumer Discretionary	7.5%	12.2%	30	-0.5	9	7	6	
H Lundbeck A/S	Health Care	9.284	27.4	9	-1.2	3	10	7	
Heineken NV	Consumer Staples	19.774	13.3	61	3.7	8	5	1	
Ind De Diseno Textil SA	Consumer Discretionary	16.621	23.8	92	-1.5	4	4	8	
KIA Motors Corp	Consumer Discretionary	10.553	4.2	13	-1.6	10	9	9	
LVMH Moet Hennesey Louis V	Consumer Discretionary	16.862	19.7	186	0.6	6	1	4	
Texas Instruments Inc	Information Technology	14.951	62	100	0.1	2	3	5	
Unilever Plc	Consumer Staples	13.248	81.1	166	1.6	1	2	3	
Volvo AB	Industrials	11.641	20.2	33	2.6	5	6	2	
Wirecard AG	Information Technology	25.017	18.1	15	-2.8	7	8	10	
Panel B: Peers selected for Unilever Plc, March 2019									
Comparable firms		EV/EBIT	ROE	Size ($B)	Net Debt/ EBIT	*Z1* $_{i-j}$	*Z2* $_{i-j}$	*Z3* $_{i-j}$	SARD score
Unilever Plc	Consumer Staples	13.248	81.1%	166	1.6	0	0	0	0
Texas Instruments Inc	Information Technology	14.951	62.0	100	0.1	1	2	1	4
LVMH Moet Hennesey Louis V	Consumer Discretionary	16.862	19.7	186	0.6	5	1	1	7
Volvo AB	Industrials	11.641	20.2	33	2.6	4	1	4	9
Ind De Diseno Textil SA	Consumer Discretionary	16.621	23.8	92	-1.5	3	5	2	10
Heineken NV	Consumer Staples	19.774	13.3	61	3.7	7	2	3	12
H Lundbeck A/S	Health Care	9.284	27.4	9	-1.2	2	4	8	14
Prediction, harmonic mean	13.936								
Absolute valuation error $\lvert(13.936 - 13.248)/13.248\rvert$	5.20%								

Source: Own display based on input from Compustat, Datastream and I/B/E/S and the SARD selection algorithm developed

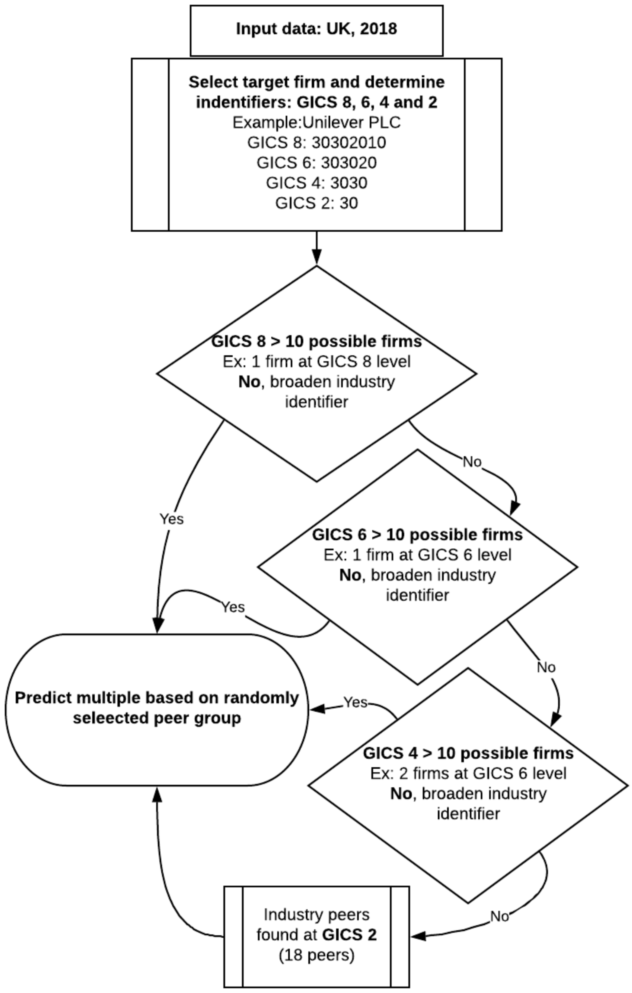
4.2.2 Cross validation of SARD algorithm

In order to ensure both reliability and validity of my results, a great deal of effort has gone into cross validation of the SARD selection models developed. Since the algorithm itself is mathematically formulated, the results of my models should correspond one-to-one with the ones of Knudsen et al. (2017) when applied on exactly identical data (Andersen, 2008). Thanks to the gratitude of the authors, providing me with their data, **Table A.1** in **Appendix A** shows that correspondence between selection algorithms is indeed the case when measuring the pooled estimation errors obtained. This provides validity to my method and results on the performance of SARD in broad global setting.

4.3 Industry selection benchmark

Knudsen et al. (2017) apply a random sample of six industry peers at the GICS 6 level as the continued benchmark in their analysis. Thus, they leave out all firms that do not satisfy a critical amount of sixteen GICS industry peers in a given year of valuation. The problem area in this thesis is concerned with the performance of SARD in a broad global setting, where smaller, local markets often do not include sixteen industry peers at the second most-detailed GICS level. Thus, I set an industry benchmark that accommodates that some firms might include a sufficient number of industry peers, while other firms do not. This builds on ideas put forward by Alford (1992), who develops an industry matching model that is progressively broadened until a sufficient amount of industry peers are located. This model accommodates the issues arising when some industries contain fewer observations than others. Combining the industry broadening, a random number generator and the four levels of GICS, I apply a randomized level-up industry selection model as my industry benchmark. The algorithm runs through the total number of possible industry peers for a given target firm. It begins at the GICS 8 level and 'broadens' by moving 'up' to the following GICS level (reducing 2 code digits: from GICS 8 to GICS 6 and so on) until a sufficient number of peers are found. **Figure 4.1** illustrates the randomized level-up industry selection model procedure as described. As an illustrative example, the randomized level-up process for Unilever Plc as of March 2018 in the UK market is provided. It is seen that the GICS 8, 6 and 4 levels all contain an insufficient number of comparable firms. Thus, the algorithm 'levels-up' until the GICS 2 default is reached. Here, 18 possible peers are identified, which means the algorithm breaks and selects a random sample of industry peers. Literature provides little evidence on an adequate cutoff for industry level-up. Cheng and McNamara (2000) note that 'if an industry definition based on n-digit SIC codes cannot generate at least six other firms (which gives seven in total), the industry should be defined with fewer SIC digits until at least six firms can be identified' (Cheng & McNamara, 2000, p. 363). Ultimately, the

Figure 4.1: Illustration of randomized level-up industry selection model



Source: Own display based on the randomized industry level-up algorithm developed sufficient number of peers for the level-up is set arbitrarily to 10 (eleven in total)[3]. Thus, if 10 peers (in addition to the firm in question) are located through the algorithm, it breaks and select the natural peers through the random number generator. **Figure B.7** in **Appendix B** provides example code of the industry level-up benchmark developed.

The randomized level-up model is arguably the hardest industry benchmark possible for my data. All firms will be selected at the most-detailed GICS level possible. Through the criteria of 10 before level-up, most industry peers will be found at the GICS 6 level aligning my benchmark with the one applied by Knudsen et al. (2017)[4].

4.3.1 Industry-based SARD-models

The industry based SARD-models applied are essentially a combination of SARD (**Section 4.2**) and the industry level-up (**Section 4.3**). First, the adequate GICS level is found for a given firm. Second, the SARD score is calculated for all industry peers located at the particular GICS level (**Table B.5** and **B.6** in **Appendix B** show example code of the industry-based SARD-model). The combination of SARD and industry level-up allows me to assess the performance of an industry-based SARD selection for firms with a limited number of industry peers.

4.4 Sample selection

A great deal of effort has gone into constructing a broad global sample that contains a sufficient amount of country-specific observations in order to apply peer selection models both locally and across borders. I use Compustat and Datastream as data sources. Compustat is generally used for the extraction of accounting fundamentals. Market data as well as forecasts originating from I/B/E/S are obtained through Datastream. Compustat includes two fundamental data resources consisting of either US data or global data. A broad global sample is attained by the combination of the two. In the following, I motivate the sample selection for both US- and global data.

US sample selection

The US sample is based on the constituents of the S&P 1500 Super Composite Index in each year throughout the sampling period. There are several arguments for the selection of S&P 1500 as the basis for the US sample. First of all, the sample selection is similar to the ones applied in related literature on multiple valuation (Knudsen et al., 2017; Bhojraj & Lee, 2002). Second, the index covers 90% of the total US market capitalization capturing

[3]The variation compared to Cheng and McNamara (2000) is due to the fact, that the level-up should accommodate some randomness when selecting the industry peer group.

[4]**Table A.2** in **Appendix A** shows that a predominant proportion of all firms will be selected at GICS 6 level or lower.

a myriad of different firms in regard to size and industries. Third, all firms are located in America and must follow the Generally Accepted Accounting Principles in the US (US GAAP)[5] (IFRS, 2018a). Finally, the S&P 1500 Index is often partitioned into sub-indices based on clear market capitalization rules. These definitions allow me to control for size effects in the US sample.

S&P 1500 constituents can have ideal peers outside the index that will not be identified due to the sample limitation. However, for the purpose of this thesis, 1500 US firms seem adequate when combined with global firms from all around the world. The US is by far the largest single country measured by the number of total observations in the combined sample.

Global sample selection

The global data resource does not include one broad index similar to S&P 1500 that constitutes a broad set of global firms of different sizes across countries. One solution was to include global firms through a combined list of global stock indices and various international composites. This sample strategy was applied by Schreiner and Spremann (2007) in their study focused on the constituents of Dow Jones STOXX 600. This procedure naturally excludes a high amount of potential peer firms at extraction which contradicts my overall idea of a broad global sample. Furthermore, it could lead to a severe size bias if only larger global firms were included. It would jeopardize the comparative element in my analysis, if the Global sample only consisted of the very largest non-US firms in STOXX 600, opposed to the US sample which includes both small-, mid- and large sized US firms. Finally, it would be challenging to sustain an adequate number of home-market peers in smaller countries, if the initial number of firms in each year was 600 across 17 European countries before data cleaning and removal. Instead, I apply a similar sampling process as the one used in Dittmann and Weiner (2005). I extract firm data for the 28 OECD countries as of year 2000 (excl. USA). Reason for this selection is that all of these are developed countries with high-income and open economies. This should add to the overall correspondence in-between firms and mitigate most of the potential biases as explained above. Furthermore, within this specific selection some convergence in accounting information should be expected. Only Japan and Switzerland are not required to apply the International Financial Reporting Standards (IFRS) as of today. Japanese firm-year observations are kept in the data because the Japanese stock market, as a whole, weighs considerably in the overall world market index[6]. In addition to this, the Japanese financial

[5]Foreign private issuers report in line with the International Financial Reporting Standards (IFRS). However, it is later shown than dual listings are accommodated in data cleaning. Furthermore, it is found that 99.999% of the observations in the US sample apply US GAAP through accounting standard identifiers.

[6]Japan market weighs 7.61% in the MSCI world index: https://www.msci.com/world

regulation has allowed Japanese firms to use the IFRS since 2010 resulting in a significant increase in the application of IFRS for Japanese firms[7]. Switzerland allows reporting through both the US GAAP, IFRS, Swiss GAAP or Bank law. A predominant part of all listed Swiss firms applies IFRS (IFRS, 2018b) and should therefore be included in the Global data. A large proportion of the global sample stems from 17 EU-member countries, that have been obliged to report in line with the IFRS as of the mandatory adoption in the EU in 2005. Ultimately, this leaves a global sample that is reasonable broad, but still presumed comparable in regard to the financial information reported. Accounting identifiers indicate that 72% of the global data is reported through IFRS throughout the total sampling period from 2000-2019. After the mandatory adoption in the EU in 2005, this proportion increases to 86%. A predominant part of instances not in line with the IFRS is found to be Japanese firms. Thus, effects from varying accounting standards will naturally affect the results to some degree. Some global firms can have natural peers outside the limitations of the OECD selection. However, when combined with the US data, the selection rules for Global firms seem adequate in order to address the problem area in this thesis.

In summary, the combination of the sampling processes for both global- and US firms, allows me to verify whether the accuracy of SARD stands in a broad global setting. The size and nature of the combined sample allows me to assess whether model precision increases with the number of potential peers included in the peer pool or if results are blurred through cross-country variation.

4.4.1 Data items

All data was imported through Structured Query Language (SQL) scripts from WRDS[8]. In the following, I report the data items included in the analysis. In most cases, data items were identical across global- and US data. However, some additional items had to be included in the global data because of the necessary currency translation of the reported values into USD. Finally, market data for global firms was obtained through Datastream. **Table 4.2** reports all the data items applied in this thesis as well as the calculation of multiples and selection variables. **Panel A** encloses data entries from the two data sources, whereas **Panel B** reports the calculation of selection variables and multiples. In order to fully grasp the nature of both estimation multiples, selection variables and firm identifiers, the reader is advised to study both panels closely. The variation in data items in-between the global- and US sample is explicitly noted in the table. These differences will be emphasized in the following.

Panel B shows that the market capitalization for global firms was obtained through

[7] https://advisory.kpmg.us/articles/2017/ifrs-adoption-japan-affect-us-subsidiaries.html
[8] Example code of import scripts can be found in **Appendix B**.

Table 4.2: Variables overview

Panel A: Data Import Entries	
Data items	**Description**
Resource 1, COMPUSTAT (compd.fundq/compg.fundq)	
Fundamentals and market prices	
SALEQ	Quarterly Sales
OIADPQ	Quarterly Operating Earnings before Amortization and Depreciation
IBQ	Quarterly Net Earnings before Extraordinary Items
DLTTQ	Long-term Debt
DLCQ	Debt in Current Liabilities
CHEQ	Cash and Short-term Investments
CEQQ	Book Value of Equity
PRCCQ (US data only)	Share price at Quarter-end
CSHOQ (US data only)	Number of shares outstanding at Quarter-end
Firm identifiers and GICS codes	
CONM	Company Name
GVKEY	Unique Compustat Company Identifier Code
LOC	Country Code of Incorporation (ISO)
GSECTOR	GICS2 code
GGROUP	GICS4 code
GIND	GICS6 code
GSUBIND	GICS8 code
ACCTSTDQ	Accounting Standard applied in Report
CURCDQ	Currency Identifier
DATADATE	Balance Sheet Date of Report
FYR	Fiscal Year-end Month
RDQ (US data only)	Publication date of Report
Currency translation (compg.g_exrt_dly)	
FROMCURD	From Currency, Local
TOCURD	To Currency, (GBR,USD)
EXRATD	Daily Exchange Rate
Resource 2, DATASTREAM **Market data and I/B/E/S estimates**	
Company Market Cap	Market Capitalization (Global data only)
Earning Per Share Median (FY2)	Consensus Estimate, FY2
Earnings Per Share Median (FY3)	Consensus Estimate, FY3
Panel B: Calculation of multiples and selection variables	
Variable	**Calculation***
Market Cap (MKTCAP)	PRCCQ×CSHOQ (US data) or Company Market Cap (Global data only)
Net Debt (NIDB)	DLTTQ+DLCQ−CHEQ
Net Debt/EBIT	NIBD/OIADP
Enterprise Value (EV)	MKTCAP+NIBD
EBIT-margin	OIADP/SALES
Return on Equity (ROE)	IB/CEQ
Implied Growth	(EPSFY3/EPSFY2)−1
EV/EBIT	EV/OIADP
P/E	MKTCAP/IB

*All calculations are based on trailing financial statements. Thus, annualized amounts of quarterly data items.

Source: Own display based on input from Compustat, Datastream and I/B/E/S

Datastream as the Global Compustat data file did not include either PRCCQ or CSHOQ. Second, I translated global currencies into USD. I followed the official guidelines from WRDS when performing the currency translation (Standard & Poor's, 2011). Using the Compustat daily exchange rate file, I used the exchange rate at the balance sheet date or the last date of trading within the specific quarter for all quarterly reports denominated in 'foreign' currencies (i.e. not USD). Third, global firms did not include the variable RDQ. This was somehow problematic since this variable provided a nice proxy for the publication date of financial information. As a result of this, I used a proxy originating from the US sample. The issue of including the most relevant information for valuations is further addressed in the next section. Finally, it is important to note that challenges arise on two data items concerning global firms from the financial- and real estate sectors. SALEQ as well as CHEQ were consistently missing in all years throughout the entire sampling period. As a result of this, no financial- or real estate firms are present in the global data. This leaves some bias when ascribing the global sample as 'fully broad'. Thus, any potential findings concerning these particular sectors should be treated with caution since these will hold for US firms only.

4.4.2 Construction of dataset

A great effort has been made to use the most recent financial information when performing the yearly valuations throughout the 20-year sample period. March 31st is the consistent date of valuation. On this date, I apply information originating from the four most recent financial quarters to calculate the trailing financial statement. The balance sheet date is not identical to the time of publication for quarterly accounting figures. Instead, I apply RDQ, which represents the date in which quarterly earnings are first publicly reported in the market, as a proxy for the 'true' publication date of financial information. When RDQ is missing for a particular observation, I assume the publication date to be 28 days after the balance sheet date[9]. As pointed out earlier, the Global Compustat file does not include RDQ. Thus, publication date of financial information for global firms is consistently assumed to be 28 days after the balance sheet date.

March 31st is strategically chosen in order to retain as many company-year observations as possible. The financial year for most firms follows the calendar year (i.e. annual balance sheet date at December 31st). When this holds, the next quarterly report (Q1) cannot be released before March 31st. Thus, for a predominant part of the firms in my sample, the most recent financial information at the time of valuation is an annual report with balance sheet date in December. This allows me to substitute the annual reports in the case of intermittent missing quarterly data values for firms with financial year-end in December.

[9] 28 is the median number of days between the balance sheet date, DATADATE, and the approximated publication date, RDQ for US firms.

Financial information for firms with a 'skewed' financial year-end is attained through the calculated trailing financial statement. The construction of trailing financial statements corresponds with the methodology of Knudsen et al. (2017). Thus, my data construction approach is directly related to the central study of the SARD approach.

Table 4.3 reports each central step in the data construction process alongside the number of firm-year observations as a result of the following actions. Thus, it provides the reader with an overview of the impact on observations from the various data requirements set for my sample. Step 1 trough 4 have the largest impact on the total number of firm-year

Table 4.3: Articulation of total firm-year observations, two samples

Steps:	**Description** Sum of obs. after:	**US data**	**Global data**	**Total**
1. Initial data	WRDS extraction	924,374	2,312,539	3,235,913
2. Sample selection filters	filtering for OECD and S&P 1500	179,764	1,052,895	1,232,659
3. Time select	time specific selection through RDQ	121,046	1,084,157	1,205,203
4. Trail reduct	calculating trailing from quarterly reports	30,375	331,825	362,200
5. Missing values	removal of intermittent missing data	23,493	155,672	179,165
6. Missing I/B/E/S forecasts	removal of missing Implied growth	21,371	42,134	63,505
7. Mkt cap restrict (global firms only)	removing smallest firms in terms of size (Mkt. cap $\geq$ 20mio. USD)	21,371	40,665	62,036
8. Non negative req.	removal of negatives in Equity, EV, Mkt. Cap and Sales and duplicates	18,065	32,165	50,230
9. Industry cleaning	cleaning insufficient GICS sector firms (num. GICS2 $\geq$ 16)			50,193

Source: Own display based on input from Compustat, Datastream and I/B/E/S

observations in the two samples. Note that step 4 shows the effect from the calculation of trailing financial statements through the quarterly reports. This is an iterative process performed in each year for each individual firm where four quarterly reports are reduced to annual figures. Steps 6 through 9 represent the specific data requirements set for my sample. Step 6 and 8 are logical requirements that need to be in place in order to perform my analysis. I drop observations when firms do not have a positive value in either of the following accounting figures: earnings before extraordinary items, EBIT (OIADP), book value of equity, EV or market capitalization. Furthermore, instances not including both the one- and two-year median analyst earnings forecast from I/B/E/S are dropped.

Naturally, I am concerned with possibility for dual listings and private foreign issuers across the global- and US data. In order to accommodate this, I remove duplicated instances of the GVKEY, which is the unique company identifier generally applied across the Compustat databases. Thus, data construction rests on the assumption that all unique GVKEYs included in the sample each represent one unique firm[10]. Step 7 is only applied for global firms. The global sample generally contains a higher proportion of 'small firms' compared to the US sample. I do not have specific index identifiers for any of the global firms. Ideally, the size of the global firms included would replicate the size of US firms. However, this is clearly not the case. The first quantile of the market capitalization in the global sample before cleaning is 142 mio. USD contrary to the US sample with a first quantile of 961 mio. USD. Application of the official guidelines for construction of S&P 1500 does not seem as a preferable data criterion. The stock market in the US is generally larger than in the rest of the world, meaning that a small-cap US firm could easily be perceived as mid- or large-cap firm globally. Thus, a large proportion of small global firms would be dropped if official S&P 1500 guidelines were applied. Instead, the global sample is restricted to only include firms with a market capitalization above 20 mio. USD which corresponds to the minimum level detected for the US sample before cleaning. This data restriction implies the smallest firms in the final sample will be independent of the country origin. 1,469 firm-year observations are left out of the global sample because of this restriction. It should be expected that some small-firm bias will naturally affect the results of my analysis. On the contrary, this should be preferred over an analysis focused only on larger global- and US firms.

Step 9 denotes an industry restriction set in order to attain a sufficient amount of GICS sector peers. The critical value is set to 16 GICS sector peers for a given firm each year throughout the sampling period (i.e. GICS 2). This correction is done for two reasons: first, the critical value is necessary when developing the industry-based selection algorithms. Second, I purposely seek to preserve firms with a limited number of industry peers. In relation to this, it is worth mentioning that the 50,193 observations denote the broadest sample analyzed throughout the thesis. I also test the accuracy of selection models on alternative sub samples. These will all originate from the 50,193 observations depicted above but vary in regard to the number of observations and GICS sector peers available. Thus, the GICS 2 criteria will naturally weigh more in smaller sub samples than seen in **Table 4.3** for the combined sample. Each sub sample analyzed will be explained as the thesis progresses.

In some rare cases, the cleaned data contains extreme values of the observed multiples that heavily influence overall statistic efficiency in my analysis. In order to minimize the effect from spurious outliers, the observed multiples are winsorized at the 99% and 1% level each year. Knudsen et al. (2017) apply an identical strategy to increase the statistical

[10] GVKEYs seem identical across Compustat libraries. For instance, the dual listing of Novo Nordisk in the US attains the exact same GVKEY as Novo Nordisk found in the global database.

robustness of their study.

4.5 Evaluation of errors

Valuation accuracy is assessed by the ability to minimize the estimation error between the value of the estimated multiple and the observed multiple in the market. In line with prior research, this is done through calculation of the percentage error

$$e_{i,t} = \frac{\widehat{M_{i,t}} - M_{i,t}}{M_{i,t}} \tag{4.3}$$

where $\widehat{M_{i,t}}$ and $M_{i,t}$ are the predicted and observed EV/EBIT multiple for a given firm i at time t. Dittmann and Maug (2008) criticize the use of percentage errors since overvaluation errors are penalized more than undervaluation errors. Instead, they suggest the application of logarithmic errors because the error distributions are closer to satisfying the normality assumptions often made for statistical inference. On this matter, Alford (1992) argues that absolute percentage errors put equal weight on positive and negative errors. Furthermore, logarithmic errors are generally difficult to interpret in economic sense compared to the more 'intuitive' percentage errors. Knudsen et al. (2017) and Dittmann and Weiner (2005) both apply absolute percentage errors in their studies. Thus, the evaluation of all model errors happens through the absolute percentage errors as expressed in equation **4.3**[11]. All selection models are programmed to identify an equal amount of firm-year valuation errors. Thus, each selection model produces an ordered vector of absolute percentage errors for all firms in my data. The absolute percentage errors will always attain values between zero and 1 when the estimated multiple, $\widehat{M_{i,t}}$, is lower than the actual value, $M_{i,t}$ (i.e. undervaluations). On the contrary, no theoretical upper bound exists for the errors incurred when overvaluations ($\widehat{M_{i,t}} > M_{i,t}$) occur. As a result, firms with smaller multiple values (the undervaluation cases) will have larger prediction errors than the firms with higher multiple values (the overvaluation cases) and the distribution of estimation errors will be right skewed. Outlier effects break assumptions of errors following a Gaussian distribution even further. These statistical issues correspond with the majority of prior studies within the field of peer selection (Dittmann & Maug, 2008; Alford, 1992; Knudsen et al., 2017). Thus, all related studies apply a similar set of key measures in order to evaluate the central tendency of the estimation errors. When selecting comparable firms, Knudsen et al. (2017), Dittmann and Weiner (2005), Herrmann and Richter (2003) all assume that the central tendency of selection model estimation errors is captured by the mean, median and interquatile range (IQ range) of the errors incurred. In addition

[11] Absolute percentage errors $\left|\frac{\widehat{M_{i,t}} - M_{i,t}}{M_{i,t}}\right|$

to this, the fraction of estimation errors below 15% is used as an indicator for performance across models (Lie & Lie, 2002; Herrmann & Richter, 2003; Dittmann & Weiner, 2005). My analysis includes all of these key measures. Altogether these should give a fair representation of the central tendency of errors for the various selection models examined.

Statistical tests focus on the comparison of central tendency across selection models. I calculate the pairwise differences between all error vectors. **Figure A.1a** in **Appendix A** reports a density plot for the pairwise differences between two selection models. It is evident, the distributions of paired differences are affected by heavy tails which leads to a 'pointy' distribution. This implies the normality assumption is violated, and non-parametric tests should be used to reach validity[12]. Thus, I apply the Wilcoxon signed rank test (Newbold, Carlson, & Thorne, 2013, p. 624) to test for the difference between the median errors of two models. I supplement with Student's t-test for differences in means between matched pairs (Newbold et al., 2013, p. 387). The inclusion of this test is reasoned by the large sample size. Applying the central limit theorem, it can be argued, that the mean of a sample drawn from heavy-tailed distributions, will be approximately normally distributed given a large enough sample size (Newbold et al., 2013, p. 254). The application of two different statistical tests is expected to increase the overall validity of results.

[12]The normality assumption of the pairwise differences are also tested through untabulated Jarque-Bera tests and Quantile-quantile plots (QQ-plot). Both indicate that extreme outliers violate the normality assumption. An exemplary QQ-plot is reported in **Figure A.1b** in **Appendix A**.

Chapter 5

Empirical results

The following chapter is devoted to the presentation of empirical results. The chapter is structured as follows: **Section 5.1** presents the descriptive statistics of all variables applied in the various selection models for the three large samples: EU firms, global firms (excl. US firms) and all firms. **Section 5.2** describes the main findings on the overall accuracy of both the SARD- and industry-based SARD approach in a broad global setting. **Section 5.3** examines the performance of SARD across 16 OECD countries with a varying number of firm-year observations. Purpose is to investigate the use-case of the SARD approach across countries as well the dependency of the designated peer pools in which valuation estimates arise. **Section 5.4** reports the results of the robustness checks performed in order to validate the empirical results. These checks include time variation of errors, the impact of the number of estimation peers, error groupings in regard to industry and size and the application of the P/E multiple.

5.1 Summary statistics

Table 5.1 reports the mean, median and IQ range for all variables used in modeling and analysis. The table includes three panels referring to either all-, global- (excl. USA) and EU firms respectively. The all firm sample includes all trailing financial statements, whereas global firms exclude US firms. EU firms are restricted to EU-member countries only. All measures are based on the pooled trailing financial statements over the sampling period (2000-2019). The difference between median and mean measures within samples indicate that distributions are generally right skewed across all measures. These differences can be attributed to the nature of the data. Large firms are expected to cause the heavy skewness in regard to Size. The distributions of financial ratios such as EBIT-margin, ROE, Net Debt/EBIT and Implied growth imply that some dispersion in the financial performance of sample firms exists. Comparing the three panels, it seems that distribution mean and median generally increase when US firms are included in the all firm sample.

Table 5.1: Summary statistics for SARD variables and estimated multiples

All firms, **Global firms (excl. USA)** and **EU firms** include 50,193, 31,899 and 20,112 firm-year observations respectively. The GICS2>16 restriction holds for all samples. Size is market capitalization denoted mio. USD ($).

Sub sample	**Mean**	**Median**	**IQ range**
Panel A: All firms			
ROE	0.178	0.127	0.116
NET DEBT/EBIT	2.263	1.257	4.054
Size (mio. $)	7,885	1,378	4,410
Implied growth	0.167	0.120	0.129
EBIT margin	0.140	0.110	0.117
EV/EBIT	17.156	13.244	9.071
P/E	26.411	18.875	15.558
Panel B: Global firms (excl. USA)			
ROE	0.163	0.127	0.119
NET DEBT/EBIT	2.729	1.255	3.913
Size (mio. $)	5,207	800	2,864
Implied growth	0.151	0.109	0.131
EBIT margin	0.126	0.097	0.105
EV/EBIT	16.841	12.926	9.172
P/E	25.329	17.888	15.103
Panel C: EU firms			
ROE	0.171	0.131	0.119
NET DEBT/EBIT	3.195	1.439	3.881
Size (mio. $)	5,701	832	3,240
Implied growth	0.151	0.108	0.128
EBIT margin	0.119	0.094	0.097
EV/EBIT	16.295	12.817	8.620
P/E	25.018	17.717	14.374

Source: Own calculations on data from Compustat and I/B/E/S

This indicates US firms exhibit higher levels of valuation multiples relative to other global firms. The values for ROE and EBIT-margin suggest the higher pricing is not caused by lower earnings. Size exhibits high variation across samples. These summery statistics are not surprising. USA alone accounts for more than 50% of the MSCI World Index[1]. Here, the S&P 500 component itself contains only the largest US firms, but accounts for approximately 80% of the total market capitalization of the S&P 1500 Super Composite (Knudsen et al., 2017). The increase in mean levels in **Panel A** relative to the levels in the other two panels could be explained by the inclusion of these large US firms. Additionally, we know both the global- and EU-firm sub samples include more 'small' firms relative to the US[2]. This is explicitly made clear through the differences in sample medians across the three panels. No significant differences are observed when comparing summary statistics of global- and EU-firms.

Table A.3 in **Appendix A** confirms the combined sample used for modeling and analysis is truly a broad global sample. The all firm sample contains firm-year observations

[1] https://www.msci.com/world

[2] The first quantile of Market cap in the cleaned Global- and EU sample is 231 and 228 mio. USD respectively, whereas the first quantile is 1,105 mio. USD in the US sample.

from 27 out of 29 OECD countries as of year 2000. Data does not include any Icelandic or Irish firms. The number of observations for each country every year indicates that the amount of data is sufficient in order to perform cross-country analysis of model errors. USA is by far the largest country sample including 18,034 firm-year observations.

5.2 Overall performance of SARD

Before moving on to the empirical results, **Table 5.2** provides an overview of SARD-model abbreviations used throughout the remainder of this thesis. I distinguish between regular SARD-models and industry-based SARD-models denoted by SARD and INDSARD respectively. The selection variables included are identical regardless of whether the selection happens within industries or not. Selection variables are added sequentially starting with ROE. I label each SARD-model by the number of selection variables included. Thus, a SARD-model including ROE is abbreviated to SARD1, a SARD-model based on both ROE and Net Debt/EBIT is abbreviated to SARD2 and so on. The randomized industry level-up algorithm is generally referred to as the Industry benchmark. Analysis is limited to examine these models and selection variables only.

Table 5.2: Definition of selection models, selection variables and abbreviations

Model	**SARD selection variables included:**				
SARD1/ INDSARD1	ROE				
SARD2/ INDSARD2	ROE	NET DEBT/ EBIT			
SARD3/ INDSARD3	ROE	NET DEBT/ EBIT	Size		
SARD4/ INDSARD4	ROE	NET DEBT/ EBIT	Size	Implied growth	
SARD5/ INDSARD5	ROE	NET DEBT/ EBIT	Size	Implied growth	EBIT margin

Source: Own production

5.2.1 Accuracy of SARD-models

The following section evaluates whether peer groups based on SARD is more precise vis-a-vís estimations whose peer groups are based on the Industry benchmark. Model accuracy is assessed through the central tendency of estimation errors. Statistical tests are performed to verify whether the pattern observed through the key measures are statistically significant or not. **Table 5.3** presents the key measures of valuation errors incurred for SARD1 through 5 and the Industry benchmark for the three different sub samples analyzed. It is generally seen that adding more selection variables improve the valuation accuracy across

Table 5.3: Valuation Errors, Five SARD Combinations and Industry Benchmark

Median, mean, IQ range and proportion less 15% of valuation errors incurred for each peer selection models. The parentheses indicate the ranking based on the estimation errors across the models. Red parentheses indicate lowest ranking, whereas blue coloring indicates top ranking. **Bold numbers** are addressed in the text.

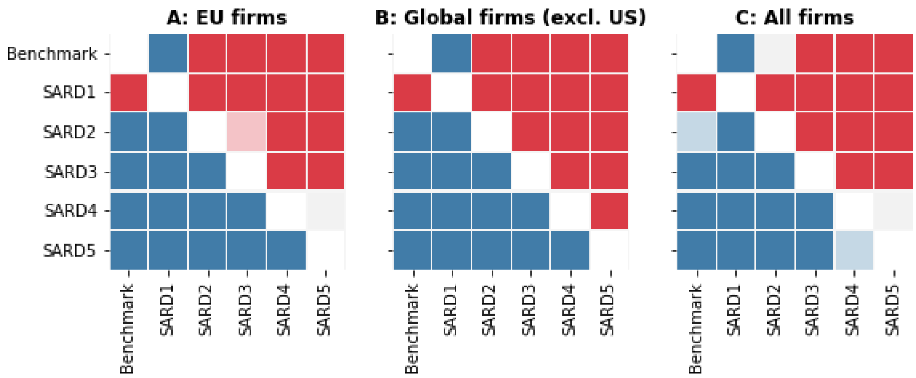
	Benchmark	**SARD1**	**SARD2**	**SARD3**	**SARD4**	**SARD5**
	Industry	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
Panel A: All firms						
Median	**0.312 (4)**	**0.348 (6)**	0.314 (5)	0.301 **(3)**	0.283 **(2)**	**0.280 (1)**
Mean	**0.482 (5)**	**0.528 (6)**	0.480 (4)	0.456 **(3)**	0.430 **(2)**	**0.427 (1)**
IQ range	**0.402 (5)**	**0.425 (6)**	0.392 (4)	0.384 **(3)**	0.368 **(2)**	**0.366** (1)
Prop<15%	**0.253 (5)**	**0.228 (6)**	0.255 (4)	0.265 **(3)**	0.281 **(2)**	**0.282 (1)**
Panel B: Global firms (excl. US firms)						
Median	**0.335 (5)**	0.356 (6)	0.324 (4)	0.311 **(3)**	0.297 **(2)**	**0.294 (1)**
Mean	**0.530 (5)**	0.548 (6)	0.494 (4)	0.468 **(3)**	0.453 **(2)**	**0.444 (1)**
IQ range	**0.433 (6)**	0.426 (5)	0.401 (4)	0.397 **(3)**	0.386 **(2)**	**0.376 (1)**
Prop<15%	**0.236 (5)**	0.217 (6)	0.247 (4)	0.258 **(3)**	0.269 (1)	**0.268 (2)**
Panel C: EU firms						
Median	**0.312 (5)**	0.331 (6)	0.299 (4)	0.291 **(3)**	0.278 **(2)**	**0.271 (1)**
Mean	**0.476 (5)**	0.502 (6)	0.451 (4)	0.440 **(3)**	0.417 **(2)**	**0.409 (1)**
IQ range	**0.403 (6)**	0.401 (5)	0.379 (4)	0.371 **(3)**	0.363 **(2)**	**0.353 (1)**
Prop<15%	**0.255 (5)**	0.235 (6)	0.270 (4)	0.277 **(3)**	0.284 **(2)**	**0.289 (1)**

Source: Own calculations based on errors incurred from the SARD algorithm and Industry benchmark

the three sub samples. SARD1 seems to be the worst-performing SARD-model, but both mean and median errors as well as the IQ range decrease as more selection variables are included. The proportion of valuation errors less 15% exhibits an identical pattern. For example, in the all firms sample, the median absolute percentage error is 34.8% when ROE is used, decreasing to 28.0% when all five selection variables are applied. The mean error decreases from 52.8% to 42.7%. The dispersion of errors decreases from an IQ range of 42.5% at SARD1 to 36.6% in SARD5. Finally, SARD5 yields a proportion of valuation errors less 15% of 28.3% opposed to SARD1 where 22.8% of valuation errors are lower than the 15% cutoff. The ranking and color codes indicate that the trend and direction of errors are similar across the three different sub samples. This finding indicates the selection variables approximate the underlying value drivers for the EV/EBIT multiple and are useful for peer selection in a broad global setting. Comparing the results of the SARD-models with the Industry benchmark, the SARD-approach generally seems superior to industry classification as more selection variables are included. SARD1 seems more inaccurate through the key measures. However, all key measures indicate that SARD3, 4 and 5 yield more accurate estimates than the Industry benchmark across all three sub samples. Some variation in the level of valuation errors across samples is detected. For instance, the median error of SARD5 is 27.1% when selection happens within the EU,

Figure 5.1: Statistical tests, Five SARD Combinations and Industry Benchmark

Results of the two-tailed Wilcoxon signed rank test are presented below the white diagonal and the two-tailed t-test are placed above the white diagonal. Blue fields indicate that the selection model in the row obtains a lower valuation error than the selection model in the column, whereas red fields indicate the opposite. Solid color fields indicate significance at the 1%, whereas light color fields indicate significance at 5%. The grey fields indicate an insignificant difference between a set of models.



Source: Own calculations based on errors incurred from the SARD algorithm and Industry benchmark

compared to 29.4% when the SARD identification are performed for global firms excluding the US. This could either indicate that the explanatory power of selection variables varies from country to country or that the overall performance of the SARD approach is also related to the designated peer pool in which estimates are based on. SARD5 seems as the preferable peer selection model across the three samples. Only for global firms (excl. US) there seems to be some ambiguity in regard to the proportion less 15% with a top ranking at SARD4.

Figure 5.1 reports the results of the Wilcoxon signed rank test and t-test[3] presented in a heat map matrix[4]. The results of the heat maps support the pattern detected in **Table 5.3** and reveal the differences in absolute percentage valuation errors are significant at conventional levels (i.e. p-value < 0.05). Thus, the incremental gain in explanatory power when adding selection variables seems statistically robust. This is exemplified by the consistent solid blue color fields when comparing the different SARD-models through the Wilcoxon test. In most cases, the two tests presented on each side of the diagonal yield identical conclusions. No opposing results are detected. However, in the field comparing SARD4 and 5 in the all firm sample (**Sub figure C**), the Wilcoxon test provides a probability of 95%, that SARD5 is more accurate than SARD4 (p-value $<$ 0.05), but attains a grey field indicating indistinguishable mean errors through the corresponding t-test. The same holds when comparing SARD2 and the Industry benchmark. The Wilcoxon test yields a significant difference at the 95%-level but accepts the hypothesis of paired

[3] $H0 : \pi_r - \pi_c = 0$ versus $H1 : \pi_r - \pi_c \neq 0$ where π denotes either median/mean of a given error vector and r and c denote the model in the row and column respectively.

[4] Numeric values of the test results are enclosed in **Table A.6** in **Appendix A**.

mean difference equal to zero through the t-test. The EU-firm sample (**Sub figure A**) also includes some statistical ambiguity on the superiority of SARD5 compared to SARD4. The t-test suggests that SARD4 and 5 is indistinguishable from each other in this sub sample. Hence, evidence suggests the inclusion of EBIT-margin in SARD5 does not add significant additional explanatory power over SARD4. This is not the case in the global firm sample (**Sub figure B**). Here, SARD5 is significantly more accurate than SARD4. The upper- and lower panels in **Sub figure B** articulates fully with one another.

The field comparing SARD1 and the Industry benchmark consistently attains a solid red color at Wilcoxon (below diagonal) and a solid blue color at the t-test (above the diagonal) providing strong evidence that the Industry benchmark outperforms SARD1 only (p-value < 0.01). The more comprehensive SARD-models (3+ selection variables) are superior to industry selection. This pattern is found for all three sub figures (i.e. sub samples) and is aligned with the results of the key measures in **Table 5.3**.

In summary, the results support acceptance of the hypothesis that peer selection through SARD rather than the Industry benchmark leads to more accurate valuations for the EV/EBIT multiple. This holds for all three sub samples examined. Some ambiguity arises regarding the incremental informative value of EBIT-margin. Ultimately, SARD5 seems to be the optimal model as no statistical evidence points in favor of the opposite. Sample specific differences are observed in regard to the level of absolute valuation error. This suggests SARD estimates are affected by the pool of potential peers in which estimates are based on.

5.2.2 Accuracy of industry-based SARD-models

In this section, the results from the industry-based SARD approach is examined and compared to the results from the regular SARD-models. **Table 5.4** reports the key measures of valuation errors incurred for INDSARD1 through 5 estimating the EV/EBIT multiple in the three different sub samples analyzed.

The results of **Table 5.4** imply the SARD approach used within industry groups produces an incremental increase in accuracy compared to the SARD approach used across industries as presented in the previous section. The explanatory power of selection variables is still evident. For example, in **Panel A** displaying the errors incurred at the all firm sample, the median absolute valuation error is 29.9% when selections are based on ROE and industry level-up alone, while decreasing to 25.9% when all five selection variables are applied within the same industry. Both mean error, IQ range and proportion of valuation errors less than 15% move in a similar direction. Knowing the baseline results are equal to the benchmark results from **Table 5.3**, this finding reinforces the preliminary conclusion that the selection variables approximate the underlying value drivers of

Table 5.4: Performance of Five INDSARD Combinations

Median, mean, IQ range and proportion of valuation errors less than 15% incurred for each industry-based SARD model. Color ranking scheme as outlined in **Table 5.3**, **Bold numbers** are addressed in the text.

	INDSARD1	INDSARD2	INDSARD3	INDSARD4	INDSARD5
	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
Panel A: All firms					
Median	**0.299(5)**	0.277 (4)	0.270 (3)	0.260 (2)	**0.259(1)**
Mean	**0.463(5)**	0.425 (4)	0.411 (3)	0.402 (2)	**0.398(1)**
IQ range	**0.388(5)**	0.368 (4)	0.360 (3)	0.352 (2)	**0.347(1)**
Prop<15%	**0.268(5)**	0.291 (4)	0.298 (3)	0.309 (2)	**0.308(1)**
Panel B: Global firms (excl. US firms)					
Median	**0.323(5)**	0.298 (4)	0.292 (3)	0.286 (2)	**0.285(1)**
Mean	**0.504(5)**	0.460 (4)	0.447 (3)	0.440 (2)	**0.436(1)**
IQ range	**0.408(5)**	0.388 (4)	0.384 (3)	0.379 (2)	**0.375(1)**
Prop<15%	**0.246(5)**	0.263 (4)	0.273 (3)	0.279 (2)	**0.279(1)**
Panel C: EU firms					
Median	**0.302(5)**	0.275 (4)	0.271 (3)	0.270 (2)	**0.266(1)**
Mean	**0.454(5)**	0.413 (4)	0.407 (3)	0.405 (2)	**0.402(1)**
IQ range	**0.376(5)**	0.361 (4)	0.357 (3)	0.354 (2)	**0.348(1)**
Prop<15%	**0.265(5)**	0.291 (4)	0.297 (3)	0.302 (2)	**0.302(1)**

Source: Own calculations based on errors incurred from the industry-based SARD algorithm

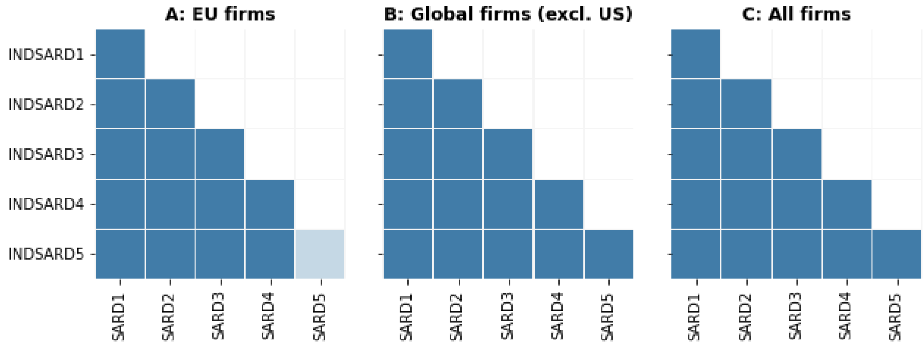
the EV/EBIT multiple[5]. Model error tendencies articulate across the three sub samples. However, contrary to the results of the regular SARD-models, it is seen that the lowest level of absolute percentage valuation errors for INDSARD-models is obtained at the all firm sub sample. Recalling that EU firms incurred the lowest level of valuation errors using the regular SARD-models, the incremental benefits from industry classification seem relatively small compared to the impact observed in the all firm sample. This finding suggests the performance of the INDSARD approach is somehow correlated with the size of the peer pool, as the level of detail for GICS classification presumably will be higher when the number of potential industry peers increases. Furthermore, this dynamic could imply the eventual effect from country-specific differences is mitigated when the SARD approach is applied within industry groups in the all firm sample.

In order to assess the overall benefits when applying INDSARD opposed to the regular SARD-models, statistical tests for differences in the absolute percentage valuation errors have been performed. **Figure 5.2** reports the results of the Wilcoxon tests performed for the differences in model errors on the three sub samples analyzed. The results of the t-test are enclosed in **Figure A.2** in **Appendix A**. The heat maps show that INDSARD-

[5]**Table A.7** in **Appendix A** encloses statistical tests proving the incremental gain in model accuracy when adding more selection variables in INDSARD-models. Furthermore, INDSARD is superior to the Industry benchmark.

Figure 5.2: Wilcoxon tests, SARD versus Industry SARD-models

Statistical tests of differences in absolute percentage valuation errors in between SARD- and INDSARD-models. Results of the two-tailed Wilcoxon signed rank test are displayed. Blue fields indicate that the selection model in the row obtains a lower valuation error than the selection model in the column. Solid color fields indicate significance at the 1%, whereas light color fields indicate significance at 5%. The grey fields indicate an insignificant difference between a set of models.



Source: Own calculations based on errors incurred from the industry-based SARD algorithm

models are more precise when identifying peers compared to identification through the regular SARD-models (e.g. p-value < 0.01). This finding applies for all three sub samples examined. EU firms (**Sub figure A**) deviate slightly as the benefit of INDSARD5 is only significant at the 95%-level (e.g. p-value < 0.05). Similar conclusions emerge when analyzing the results of the t-tests addressing the mean errors. INDSARD is generally superior to SARD. However, the incremental informative value of INDSARD5 opposed to SARD5 is found to be insignificant for both the global- and EU-firm sub samples. Regardless of this, most evidence support the hypothesis stating the industry-based SARD approach is superior to the regular SARD approach when identifying peers in a broad global setting.

5.2.3 Summary of overall performance of SARD

Summarizing the first part of the empirical analysis, it seems the SARD approach should be considered as superior to a selection based solely on industry classification in a broad global setting. The selection variables applied seem to hold explanatory power in regard to the fundamentals underlying the EV/EBIT multiple. Furthermore, it is evident that the combination of industry classification and SARD increases estimation accuracy even further. As a result of these findings, both **H1** and **H2** find support. However, the results also indicate, that the accuracy of both SARD- and INDSARD varies in regard to the size and nature of the peer pool. These findings suggest the selection methods could be affected by cross-country differences. In relation to this, the pre-defined peer pool should be just as important as the selection variables used in the model. These effects relate to the assessment of **H3**, which will be analyzed in the following section.

5.3 Cross-country performance of SARD

This section examines whether cross-country differences affect the precision of model estimates through the size of the peer pool. Specific country errors need to be assessed to provide a more nuanced assessment on the overall performance of SARD in a broad global setting. This relates to the examination of **H3**. I perform country analysis for 16 OECD countries: USA, Japan, South Korea, Australia, Norway, Switzerland, France, Germany, Great Britain, Netherlands, Italy, Spain, Denmark, Finland, Sweden and Belgium. Out of a total of 27 OECD countries present in the sample, these 16 countries have been selected because they include a sufficient number of observations in each year over the sampling period. I define 'the sufficient amount of observations' as a minimum of 10 home country peers available for a yearly valuation[6]. The cross-country analysis of INDSARD5 is performed on data restricted through the GICS 2 requirement. This limits the number of countries examined at the home country peer pool. The reason for this will be explained later.

I consider four selection pools of varying sizes for both the INDSARD5- and SARD5-model estimating the EV/EBIT multiple. The previous section did not provide any statistical evidence that any other selection model should be preferred over either SARD5 or INDSARD5. INDSARD5 was superior to SARD5, but these results seemed to depend on the size of the peer pool. I test how the country-specific estimation accuracy differs based on selection restricted to four distinct peer pools: **1)** Home country peers only, **2)** Regional firm peers only, **3)** Global firm peers only (excl. USA) and **4)** All firms. In order to attain as many potential peers as possible for the SARD5 analysis, I add back the industry peers initially removed, when the GICS 2 restriction was applied in the former analysis. The bullets below clarify the selection pools examined. I recommend readers of this thesis to study these carefully, as it provides the general intuition of both size and geography of the various peer pools:

- **Regional firms**: Denoted by 'Region'. Denotes either of **1)** firms from EU-member countries[7] or **2)** firms from the Asia-pacific including Australia, Japan or South Korea. EU-firms include 20,367[8] firm-year observations, whereas the Asia-pacific firms include 7,656[9].

[6]**Table A.4 in Appendix A** presents the count of observations for all countries in the all firm sample without the GICS 2 restriction.

[7]These include Austria, Czech Republic, France, Germany, Great Britain, Greece, Hungary, Luxembourg, Netherlands, Italy, Poland, Spain, Denmark, Finland, Sweden and Belgium.

[8]255 firm-year observations are added back.

[9]671 firm-year observations are added back.

- **Global firms excluding US (i.e. global firms)**: Articulates with the Global firms without the GICS 2 restriction. Number of firm-year observations: 32,165[10].

- **All firms**: Corresponds to the all firm sample as defined earlier without the GICS 2 restriction. Number of firm-year observations: 50,230[11].

- **Home country firms**: Peer pool is limited to the home country peers only. Individual sub samples of each of the 16 OECD countries (without the GICS 2 restriction).

Thirteen countries are examined through all four peer pools (EU-member countries, Australia, Japan and South Korea). Switzerland and Norway will not be evaluated through the regional peer pool. USA is only considered at the home country- and all firm level.

5.3.1 Accuracy of SARD across countries and peer-pools

Table 5.5 and **5.6** report all key measures of the errors incurred estimating EV/EBIT trough SARD5 in 10 EU countries and 6 non-EU countries respectively. The right-hand side of the two tables encloses the statistical tests performed for the differences of errors for each country. Because of the nature of **H3**, tests are performed with the all firm selection as the benchmark.

Table 5.5 generally exhibits large variation of absolute percentage valuation errors across EU countries. The largest median error is found at the home selection for Germany at 32%, whereas the lowest level is obtained for Sweden at 23.1%. The differences of valuation errors generally seem statistically insignificant across the larger peer pools: regional-, global- and all firms. Belgium is the only country in which some statistical evidence is provided of a preference for the all firm selection compared to the home-country peer pool (p-value = 0.023 for Wilcoxon, p-value = 0.039 for *t*-test). However, no statistical support is found that all firm selection performs better than either of the global firm- or regional firm peer pools. Key measures for Denmark, Italy and Spain all seem to be slightly lower when peers are identified at either the global firm- or regional-firm level (i.e. selection should exclude the US). The differences in median error seem reasonably strong for Italy and Denmark through the Wilcoxon test (p-value = 0.005 for Italy, p-value = 0.012 for Denmark), whereas the results for Spain yields insignificant results (p-value = 1.00). In all three cases, the results of the *t*-test indicate that there is not enough statistical evidence to conclude that the EU selection pool is superior to the all firm selection pool (p-value = 0.131 for Italy, p-value = 0.164 for Denmark and p-value = 0.425 for Spain). For all other countries, valuation errors seem relatively stable across the three larger peer pools.

[10]266 firm-year observations are added back.

[11]37 firm-year observations are added back.

Table 5.5: Valuation accuracy across countries and peer-pool sizes, EU firms

Median, mean, IQ range and proportion less 15% for absolute estimation errors at the EV/EBIT multiple of each EU member country with a sufficient amount of country peers to perform SARD5 selection (i.e peer group ≥ 10). Number of firm-year observations is presented in the left parentheses. Test results are displayed in the right panel. p-values are given in the parentheses below each test size. A red colored (blue) test size implies that selection at the all firm level is more inaccurate (accurate) than the alternative at conventional levels of significance (p-value < 0.05). **Bold letters** are addressed in the text.

Country		**Peer pool selection**				**Statistical tests**					
		Home	Region	Global	All	HomevsAll		GlobalvsAll		RegionvsAll	
				ex.USA		Wilcoxon	t-test	Wilcoxon	t-test	Wilcoxon	t-test
France	Mean	**0.361**	0.395	0.396	**0.393**	**-5.1**	**-3.9**	0.7	0.5	-1.2	0.3
(3,140)	Median	**0.253**	0.271	0.280	**0.278**	(0.000)	(0.000)	(0.495)	(0.625)	(0.214)	(0.783)
	IQ range	**0.319**	0.358	0.362	**0.363**						
	Prop<15%	**0.304**	0.289	0.284	**0.281**						
Germany	Mean	0.523	0.553	0.537	0.532	1.6	-0.5	1.1	1.0	1.2	1.6
(2,928)	Median	**0.320**	0.305	0.305	0.309	(0.105)	(0.586)	(0.284)	(0.320)	(0.238)	(0.101)
	IQ range	0.400	0.402	0.406	0.399						
	Prop<15%	0.242	0.259	0.268	0.266						
Great Britain	Mean	0.404	0.399	0.396	0.395	-0.8	1.5	-0.3	0.2	-1.1	0.8
(5,074)	Median	0.278	0.280	0.282	0.284	(0.406)	(0.133)	(0.795)	(0.871)	(0.280)	(0.443)
	IQ range	0.365	0.358	0.355	0.363						
	Prop<15%	0.285	0.285	0.275	0.276						
Netherlands	Mean	0.404	0.394	0.412	0.382	0.7	1.1	1.7	1.2	-0.5	0.5
(893)	Median	0.258	0.256	0.265	0.255	(0.492)	(0.260)	(0.082)	(0.232)	(0.614)	(0.619)
	IQ range	0.351	0.311	0.330	0.327						
	Prop<15%	0.310	0.300	0.288	0.282						
Italy	Mean	0.361	0.378	0.384	0.392	-1.9	**-3.6**	-0.7	-1.4	**-2.8**	-1.5
(1,253)	Median	0.273	**0.259**	**0.265**	0.275	(0.061)	(0.000)	(0.503)	(0.160)	(0.005)	(0.131)
	IQ range	0.312	0.340	0.380	0.367						
	Prop<15%	0.283	0.304	0.291	0.288						
Spain	Mean	0.327	0.305	0.312	0.310	1.4	2.1	0.6	0.4	0.0	-0.8
(885)	Median	0.257	**0.243**	**0.249**	0.261	(0.174)	(0.035)	(0.548)	(0.679)	(1.000)	(0.425)
	IQ range	0.325	0.303	0.301	0.327						
	Prop<15%	0.301	0.305	0.290	0.307						
Denmark	Mean	0.431	0.379	0.386	0.392	1.0	**2.4**	-1.2	-1.0	**-2.5**	-1.4
(613)	Median	0.304	**0.272**	**0.276**	0.300	(0.325)	(0.019)	(0.249)	(0.339)	(0.012)	(0.164)
	IQ range	0.378	0.352	0.360	0.343						
	Prop<15%	0.269	0.282	0.287	0.284						
Finland	Mean	**0.341**	0.328	0.332	0.334	**-3.0**	0.8	-0.7	-0.4	-1.2	-0.9
(1,123)	Median	**0.232**	0.265	0.266	0.272	(0.002)	(0.448)	(0.486)	(0.722)	(0.241)	(0.372)
	IQ range	**0.336**	0.303	0.296	0.290						
	Prop<15%	**0.340**	0.281	0.268	0.257						
Sweden	Mean	**0.339**	0.333	0.333	0.332	**-3.7**	0.6	-1.6	0.3	-1.1	0.1
(1,775)	Median	**0.231**	0.245	0.254	0.251	(0.000)	(0.576)	(0.113)	(0.798)	(0.252)	(0.894)
	IQ range	**0.317**	0.335	0.323	0.319						
	Prop<15%	**0.347**	0.320	0.307	0.305						
Belgium	Mean	**0.464**	0.428	0.423	**0.409**	**2.3**	**2.1**	-0.2	1.1	-0.2	0.7
(669)	Median	**0.268**	0.252	0.250	**0.252**	(0.023)	(0.039)	(0.830)	(0.268)	(0.831)	(0.504)
	IQ range	**0.343**	0.300	0.317	**0.315**						
	Prop<15%	**0.300**	0.317	0.311	**0.330**						

Source: Own calculations based on errors incurred from the SARD algorithm

The most striking results are found when comparing the errors incurred when selection happens at the home-level opposed to the resulting errors at the all firm peer pool. Most evident is France, that yields the most accurate estimates when the peer selection is

Table 5.6: Valuation accuracy across countries and peer-pool sizes, non-EU firms

Should be analyzed and interpreted as outlined in **Table 5.5** above.

Country		**Peer pool selection**				**Statistical tests**					
		Home	Region	Global ex.USA	All	HomevsAll		GlobalvsAll		RegionvsAll	
						Wilcoxon	t-test	Wilcoxon	t-test	Wilcoxon	t-test
Australia	Mean	**0.443**	0.437	0.438	**0.450**	**-4.6**	-1.1	**-2.8**	**-2.2**	-1.5	-1.8
(3,722)	Median	0.301	0.307	0.311	0.319	(0.000)	(0.285)	(0.005)	(0.002)	(0.137)	(0.080)
	IQ range	0.394	0.404	0.406	0.406						
	Prop<15%	0.272	0.251	0.257	0.247						
South Korea	Mean	**0.509**	0.581	0.614	**0.611**	**-4.0**	**-5.6**	1.3	0.2	-0.8	-1.4
(1,996)	Median	**0.355**	0.378	0.383	**0.384**	(0.000)	(0.000)	(0.205)	(0.847)	(0.432)	(0.155)
	IQ range	**0.423**	0.478	0.478	**0.464**						
	Prop<15%	**0.215**	0.212	0.206	**0.206**						
Japan	Mean	**0.489**	0.520	0.550	**0.597**	**-3.8**	**-4.5**	**-2.5**	**-2.6**	**-2.1**	**-3.4**
(1,998)	Median	**0.344**	0.363	0.357	**0.376**	(0.000)	(0.000)	(0.011)	(0.009)	(0.040)	(0.001)
	IQ range	**0.424**	0.442	0.464	**0.455**						
	Prop<15%	**0.225**	0.220	0.226	**0.206**						
Norway	Mean	**0.560**		**0.497**	0.512	**2.6**	**2.2**	-0.9	-1.8		
(833)	Median	**0.346**		**0.317**	0.328	(0.010)	(0.028)	(0.351)	(0.076)		
	IQ range	0.401		0.401	0.393						
	Prop<15%	**0.232**		**0.274**	0.246						
Switzerland	Mean	**0.409**		0.403	**0.408**	**-3.7**	0.1	-1.3	-0.3		
(1,642)	Median	0.258		0.296	0.290	(0.000)	(0.947)	(0.183)	(0.739)		
	IQ range	0.357		0.318	0.325						
	Prop<15%	0.329		0.249	0.253						
USA	Mean	0.376			0.389	**-5.9**	**-2.0**				
(18,065)	Median	**0.245**			**0.253**	(0.000)	(0.041)				
	IQ range	0.334			0.342						
	Prop<15%	0.324			0.318						

Source: Own calculations based on errors incurred from the SARD algorithm

restricted to the home peer pool. All key measures support this finding. Median error is 25.3% opposed to 27.8% through the all firm selection. Both the IQ range and mean error obtain the lowest levels of 31.9% and 36.1% respectively. Finally, only at the home level, the proportion of valuation errors less than 15% realize a value above 30%. Both statistical tests find strong support for the pattern detected through the key measures (p-value = 0.000). This suggests the superior peer group for French firms is other French firms only. The results of Sweden and Finland also indicate the SARD selection should be limited to home country peers only. Across the panel, these two countries generally obtain the most accurate estimates at home based on the key measures. In both cases, median errors decrease significantly when moving to the home-level from the all firm selection (p-value = 0.002 for Finland, p-value = 0.000 for Sweden). In addition to this, the highest proportion of valuation errors less 15% are realized through the home selection for both countries. On the contrary, opposing results arise when investigating the mean errors and IQ range. In both cases, mean errors increase at the home firm peer pool relative to the all firm selection. In Finland, the IQ range also rise from 29% to 33.6% when the selection pool changes from all firm to home market. The reason for the differences between median and mean absolute valuation errors can be that outliers have more impact in smaller samples (home firms) than in large samples (all firms). This could also be the reason for the higher dispersion of errors detected through the IQ range in Finland.

Table 5.6 reports the results of the country analysis for the non-EU countries. The results for Japan, South Korea and USA all indicate a clear preference for the home firm peer pool as compared to the all firm pool. Japan and South Korea generally exhibit larger levels of absolute errors compared to the ones for the US. However, for all three countries, median errors are lowest at the home level with strong statistical evidence (p-value = 0.000) through the Wilcoxon test. This also holds for the mean errors where the results of the t-tests indicate a probability greater than 99% that mean errors are lower when selection happens in the home country as compared to the all firm selection for both Japan and South Korea (p-value = 0.000). This probability is found to be 95% (p-value = 0.041) in the US. For the Asian-pacific firms, it seems the smaller peer pools should be preferred. Most evident is Japan where both the regional- and global selections are significantly more accurate than the all firm selection. Similar tendencies seem to be present for Australian- and Swiss firms. In both cases, median errors are lowest at the home-country level with high statistical evidence. In Switzerland, the mean error increases slightly at home, whereas this decreases in Australia. In both cases, results of the t-test are supported by weak statistical power (p-value = 0.285 for Australia, p-value = 0.947 for Switzerland). The all firm selection seems as the better alternative for Norwegian firms when compared to the home selection. However, key measures indicate the preferable peer pool for Norway is the global peer pool attaining the lowest absolute valuation errors across all key measures.

The analysis presented above yields two principal results that are relevant in the assessment of the SARD approach in a broad global setting consisting of different national markets. First, cross-country differences clearly exist in regard to the level of absolute valuation errors incurred. Countries such as Australia, Japan and South Korea yield considerably larger error measures than the ones for USA, France and Sweden. European- and US firms generally seem to incur lower valuation errors as compared to firms from the Asia-pacific. This suggests that firms from the Western part of the world are more homogeneous when compared with other global firms through SARD. Japanese firms seem more heterogeneous with a clear preference for the home peer pool. Considering the fundamental idea of SARD, it is somehow surprising the errors from the smallest European markets such as Denmark and Belgium, including considerably fewer yearly observations, are clearly lower than the valuation errors obtained in the large Asian markets. This indicates the selection variables might not work as well approximating the underlying fundamentals of EV/EBIT in the Asian countries.

Second, country-wise variation in regard to the preferable peer pool seems to exist.

Interestingly, none of the 16 countries examined elicit a significantly better peer selection when identification happens at the largest sample. In fact, most of the evidence points in the opposite direction. The results of the larger country-wise sub samples (e.g. France, USA, Japan) generally indicate selection should be restricted to the home market only, whereas smaller sub samples (e.g. Belgium, Denmark) seem to be better off at the broader selection levels. Test results for the preference of alternative peer pools are enclosed in **Table A.8** in **Appendix A** and prove the home market selection is consistently preferred for French firms and EU selection seems superior over the home selection for Denmark and Spain. Furthermore, home selection is clearly the preferred peer pool for both South Korean- and Japanese firms.

It is beyond the scope of thesis to identify the optimal peer pool as well as selection model for each individual country. However, the findings above clearly points in the favor of rejection of the hypothesis that the accuracy of SARD increases with the number of peers included in the peer pool. France, USA, Japan and South Korea provide strong evidence that peer selection should be limited to home country firms only. To a lesser degree, Sweden, Finland, Switzerland and Australia also seem better off when restricted to the home country. Most of these instances include a relatively high amount of firm-year observations over the sampling period. This suggests firms originating from relatively large local markets (e.g. ± 50 firm-year observations pr. year[12]) are generally better identified with comparable firms from the same country. Conversely, it seems that firms from smaller country markets (e.g. num. home peers < 50) benefit from a peer selection across borders (i.e. Denmark, Belgium and Spain). Finally, some large countries seem relatively stable across the varying peer pools. Germany and Great Britain consistently yield indistinguishable test results and similar key measures when selection pools change. This indicates these particular countries remain indifferent in regard to the home country preference, that seems more dominant in other countries (i.e. France, USA, Japan).

5.3.2 Accuracy of INDSARD across countries and peer-pools

This analysis is in direct relation to the assessment of **H3**; if firm characteristics are captured by the combination of industry-classification and SARD selection variables alone, a larger pool of potential peers should be preferable. In order to conduct this analysis, the GICS 2 restriction holds again. This has implications for the number of countries able to run the INDSARD algorithm at the home country level throughout most of the sampling period. **Table A.5** in **Appendix A** displays the number of firm-year observations for each country after applying the GICS 2 restriction in the home market. Australia, France, Finland, Germany, Great Britain, Japan, South Korea, Sweden, Switzerland and the US

[12]Countries with 1000+ firm-year observations seem to be better off when selection is restricted to the home market. Given this $\frac{1000}{20} = 50$ should roughly add to a sufficient peer pool for home selection.

all include at least ten years of data after removing insufficient industry observations. This is considered to be sufficient in order to statistically assess and compare the effects of varying peer pool sizes through the INDSARD-model. The amount of home peers in the Netherlands, Italy, Spain, Denmark, Belgium and Norway is insufficient. Thus, these countries are only analyzed through selections based on the larger peer pools.

Table 5.7 reports the key measures of the estimation errors using INDSARD5 for the 10 countries able to select peers in the home country. High dispersion of country errors is observed. For instance, the lowest median error is obtained at the home peer pool for Sweden at 19.8%, whereas South Korean firms yield a median error of 37.9% at the regional peer pool. Valuation errors seem to decrease overall when applying the INDSARD5-model as compared to the SARD5-model in the previous section. Most evident are the errors obtained in Great Britain, Sweden and the US where median errors have fallen 2-3% from the ones incurred when applying SARD5. Other key measures such as the IQ range and proportion less 15% exhibit an identical pattern. Country errors for France and Finland have remained unchanged compared to SARD5. This could indicate the industry characteristics are more dominant in some countries than others and that INDSARD5 is not always the superior selection model.

Home firm selection generally seems less attractive when INDSARD5 is used. Most countries exhibit insignificant differences among the various peer pools. For instance, the US median error is 22.1% at the home peer pool while being 22.4% at the all firm peer pool. Both statistical tests show the two pools are indistinguishable from one another. This pattern is also observed for Sweden and Finland. Interestingly for Sweden is that the all firm peer pool is superior to both the global- and regional peer pool (p-value < 0.05). This indicates that industry classification mitigates some of the cross-country differences that previously affected the comparison of Swedish firms with other global peers in the larger selection pools. France stands out with strong statistical support in favor of the home country peer pool. As noted earlier, French firms do not experience the same increase in accuracy through INDSARD5 as seen in other countries. Both the median error and IQ range have slightly increased compared to the levels incurred when applying the regular SARD5-model. Regardless of this, both tests indicate home selection should be preferred over any of the other selection pools examined (p-value = 0.013 at Wilcoxon, p-value = 0.000 in the *t*-test)[13]. Switzerland stands out compared to prior findings. Recall that the previous section showed slight evidence that Swiss firms should preferably be compared with other Swiss firms. After correcting the Swiss sample for home industry peers, Swiss firms are clearly better off when compared to other firms across the world. Errors from the home selection have increased, while the errors incurred from the global- and all

[13]Additional statistical tests performed for the alternative peer pools enclosed in **Table A.9** in **Appendix A** prove that home selection is the better peer pool for France at conventional levels of significance.

Table 5.7: INDSARD valuation accuracy across countries and peer-pool sizes

The table displays estimation errors of INDSARD5 on all countries with sufficient GICS 2 peers in a given valuation year (num. GICS 2 peers ≥ 16). Mean, median, IQ range and proportion less 15% of the absolute estimation error are further displayed. Right hand parentheses present the number of firm-year observations for a given country as well as the count of valuation years with sufficient restricted data. The left side displays results of statistical tests (should be analyzed as pointed out in **Table 5.5**).

Country		**Peer pool selection**				**Statistical tests**					
		Home	Region	Global	All	HomevsAll		GlobalvsAll		RegionvsAll	
				ex.USA		Wilcoxon	t-test	Wilcoxon	t-test	Wilcoxon	t-test
France	Mean	0.343	0.387	0.382	0.393	**-2.5**	**-6.2**	-1.8	**-2.5**	-1.0	-1.0
(2,327)	Median	**0.258**	**0.269**	**0.263**	0.271	(0.013)	(0.000)	(0.069)	(0.011)	(0.299)	(0.332)
(19)	IQ range	0.341	0.363	0.371	0.352						
	Prop<15%	0.306	0.304	0.300	0.298						
Germany	Mean	0.543	0.568	0.570	0.560	**2.0**	-1.0	0.9	0.5	0.3	0.6
(2,303)	Median	0.324	0.308	0.304	0.293	(0.042)	(0.302)	(0.393)	(0.600)	(0.745)	(0.531)
(20)	IQ range	0.409	0.400	0.414	0.409						
	Prop<15%	0.250	0.262	0.255	0.258						
Great Britain	Mean	0.363	0.354	0.360	0.357	0.2	1.2	1.4	1.1	-0.1	-0.6
(4,395)	Median	**0.249**	**0.246**	**0.255**	**0.253**	(0.831)	(0.217)	(0.153)	(0.266)	(0.890)	(0.553)
(20)	IQ range	0.349	0.337	0.345	0.333						
	Prop<15%	0.327	0.325	0.314	0.314						
Sweden	Mean	0.294	0.294	0.288	0.280	-1.1	1.5	**2.7**	**2.0**	**2.3**	**2.3**
(1,122)	Median	**0.198**	**0.226**	**0.228**	**0.219**	(0.284)	(0.126)	(0.008)	(0.044)	(0.019)	(0.019)
(19)	IQ range	0.278	0.279	0.273	0.270						
	Prop<15%	0.383	0.344	0.347	0.346						
Finland	Mean	**0.350**	**0.324**	**0.322**	**0.320**	-0.4	1.8	0.1	0.4	0.7	0.4
(325)	Median	0.237	0.265	0.255	0.255	(0.679)	(0.081)	(0.933)	(0.704)	(0.473)	(0.717)
(13)	IQ range	0.309	0.272	0.291	0.291						
	Prop<15%	0.357	0.280	0.308	0.314						
Australia	Mean	0.460	0.463	0.453	0.450	-1.7	0.9	**2.3**	1.2	0.5	1.1
(2,907)	Median	0.315	0.329	0.324	0.320	(0.094)	(0.389)	(0.023)	(0.212)	(0.603)	(0.251)
(19)	IQ range	0.389	0.399	0.392	0.391						
	Prop<15%	0.256	0.242	0.231	0.230						
South Korea	Mean	0.534	0.545	0.593	0.591	-0.7	**-3.0**	1.7	0.2	-0.3	**-2.7**
(1,351)	Median	0.368	**0.379**	0.376	0.365	(0.475)	(0.003)	(0.095)	(0.821)	(0.762)	(0.008)
(10)	IQ range	0.432	0.445	0.446	0.449						
	Prop<15%	0.207	0.212	0.212	0.213						
Japan	Mean	0.529	0.589	0.581	0.599	0.4	**-2.3**	-1.1	**-2.0**	1.5	-0.6
(1,290)	Median	0.361	0.368	0.352	0.366	(0.698)	(0.020)	(0.279)	(0.042)	(0.131)	(0.527)
(16)	IQ range	0.448	0.471	0.484	0.487						
	Prop<15%	0.212	0.220	0.242	0.238						
Switzerland	Mean	0.388		0.348	0.343	**3.0**	**4.0**	1.8	1.0		
(629)	Median	0.280		0.260	0.258	(0.003)	(0.000)	(0.068)	(0.319)		
(20)	IQ range	0.325		0.310	0.315						
	Prop<15%	0.289		0.310	0.312						
USA	Mean	0.327			0.325	-0.8	0.9				
(18,034)	Median	**0.221**			**0.224**	(0.435)	(0.363)				
(20)	IQ range	0.304			0.304						
	Prop<15%	0.362			0.358						

Source: Own calculations based on errors incurred from the INDSARD5 algorithm

firm selection have fallen to the same levels as found for the home pool using SARD5. While acknowledging firm-year observations have been more than halved compared to the prior analysis, it seems that some trade-off exists between industry- and national characteristics for the Swiss firms able to be identified through INDSARD5 in the home market. **Table 5.8** displays the errors obtained for INDSARD5 at the three larger peer

Table 5.8: INDSARD valuation accuracy across countries and peer-pool sizes, home-country selection not available

Results for countries with insufficient GICS 2 peers in a given valuation year (num. GICS 2 peers < 16). Should be interpreted and analyzed as outlined in **Table 5.7** and **5.5**.

Country		Peer pool selection			Statistical tests			
		Region	Global ex.USA	All	GlobalvsAll		RegionvsAll	
					Wilcoxon	t-test	Wilcoxon	t-test
Netherlands	Mean	0.380	0.390	0.381	1.8	1.4	0.4	-0.2
(891)	Median	0.251	0.252	0.245	(0.073)	(0.156)	(0.717)	(0.854)
(20)	IQ range	0.331	0.343	0.329				
	Prop<15%	0.304	0.299	0.330				
Italy	Mean	0.343	0.348	0.347	-1.1	0.2	-0.9	-0.6
(1,224)	Median	0.259	0.262	0.260	(0.275)	(0.839)	(0.363)	(0.552)
(20)	IQ range	0.328	0.329	0.334				
	Prop<15%	0.306	0.313	0.303				
Spain	Mean	0.291	0.293	0.292	0.4	0.3	0.2	-0.3
(871)	Median	0.238	0.243	0.243	(0.675)	(0.743)	(0.852)	(0.788)
(20)	IQ range	0.295	0.304	0.312				
	Prop<15%	0.325	0.317	0.320				
Denmark	Mean	0.362	0.360	**0.356**	1.6	0.6	1.8	0.6
(609)	Median	**0.275**	**0.259**	**0.261**	(0.114)	(0.579)	(0.074)	(0.524)
(20)	IQ range	0.353	0.330	0.327				
	Prop<15%	0.305	0.296	**0.322**				
Belgium	Mean	0.521	0.537	0.541	1.1	-0.3	-1.4	-1.2
(655)	Median	0.248	0.277	0.261	(0.263)	(0.780)	(0.165)	(0.214)
(20)	IQ range	0.361	0.356	0.334				
	Prop<15%	0.337	0.308	0.310				
Norway	Mean		0.436	0.476	0.3	-1.2		
(832)	Median		0.316	0.290	(0.744)	(0.239)		
(20)	IQ range		0.385	0.410				
	Prop<15%		0.254	0.267				

Source: Own calculations based on errors incurred from the INDSARD5 algorithm

pools. Denmark, Italy, the Netherlands, Norway and Spain all seem to benefit slightly from using the INDSARD5-model. Key measures of errors generally seem to decrease in the span of 1-4% for these countries. Furthermore, the proportion of valuation error less than 15% seem to increase with identical levels. For instance, Denmark obtained a median error of 30% applying SARD5 at the all firm selection, while now obtaining a median error of 26.1% through INDSARD5. The firm-year observations for each country in this analysis are more or less identical to the amount of observations included when applying SARD5 in the previous section. Thus, the differences in key measures confirm the overall finding, that INDSARD-models are generally superior to SARD-models in the all firm sample for these countries. However, only Danish firms seem to be better off overall when comparing errors across the different peer pools with the results obtained in **Section 5.3**; when using SARD5, the lowest valuation error for Danish firms was obtained at the regional level with a median error of 27.2%, whereas the tendencies of the key measures for

INDSARD5 points in favor of the all firm pool with a median error of 26.1%, an increased proportion of valuation errors less 15% of 32.2% and mean error of 35.6%. No significant difference is found between the various peer pools using INDSARD5 for Denmark. For all other countries presented in **Table 5.8** it holds, that similar levels of key measures trough INDSARD5 can also be found when applying SARD5 at the various peer pools (**Table 5.5** and **5.6**). This goes along prior indications that the designated peer pool seems highly relevant when applying either of the SARD- or INDSARD approaches for peer selection in a broad global setting.

5.3.3 Summary of cross-country performance of SARD

Comparing the results of **Section 5.3.1** and **5.3.2** it seems evident that cross-country differences and distinct peer pool characteristics are highly influential when performing either SARD- or INDSARD-models in a broad global setting. These findings lead to rejection of **H3**, stating that larger peer pools should lead to more accurate SARD estimates. Thus, SARD selection variables are not the only underlying drivers for the EV/EBIT multiple in a broad global setting. Some indications are provided, that firms from countries with a relatively large amount of home peers (num.tot.obs > 1000) should be compared to other firms from the same country when using the SARD5-model. However, this preference seems less evident when applying the INDSARD5-model. Thus, industry characteristics might offset national differences when comparing firms across borders. Smaller local markets (num.tot.obs < 1000) generally seem to benefit from larger peer pools. However, these results are supported by weaker statistical power. Additionally, regional differences seem to be present. Firms from the Asia-pacific region seem less suitable for SARD compared to firms from the Western part of the world.

As mentioned earlier, it is beyond the scope of this thesis to identify the optimal peer selection model and peer pool for each of the 16 countries examined. However, findings clearly indicate that no single SARD- or INDSARD-model works best for all countries. This corresponds with the findings in Dittmann and Weiner (2005) and nuances the fundamental idea of thought underlying the SARD approach: the best choice for French firms might be a sub-optimal choice when valuing US firms and so on.

5.4 Robustness checks

This section reports the robustness checks performed for the results obtained. Except when explicitly stated otherwise, the methodological setup presented in **Section 4** holds across all robustness checks performed. I perform robustness checks in regard to the usage of an alternative multiple, error groupings, time variation of errors and the number of estimation peers used for valuation.

5.4.1 Alternative multiple

In the following, I provide a robustness check centered on analysis performed for the P/E multiple. Purpose of this is to analyze whether the results would have been different if another multiple had initially been selected. Thus, the methodological setup used is fully identical to the one used in **Section 5.2** and **5.3**, but for the usage of P/E alternative to EV/EBIT.

Overall performance when estimating P/E

Figure 5.3 displays a condensed version of the full error analysis performed on estimates of the P/E multiple. Three sub samples are presented in three bar charts: EU firms, global firms (excl. USA) and all firms. Each bar chart represents one sub sample and reports the median error incurred from one specific SARD/INDSARD-model. Alternative key measures are enclosed in **Table A.10** in **Appendix A**. The estimation errors of SARD are consistently below the horizontal lines representing the errors of the Industry benchmark. This supports the finding that peer groups based on SARD are generally more accurate vis-a-vís estimates whose peer groups are based on industry classification independent of the choice of multiple[14]. As seen for the EV/EBIT multiple, median errors decrease as more selection variables are included. The lowest median errors are incurred at the most comprehensive models. In relation to this, note that no significant difference is detected between SARD4 and SARD5. This suggests EBIT-margin is less useful when estimating P/E.

Comparing the differences between the left- (grey blue) and right bars (light blue), INDSARD-models seem more accurate than the regular SARD-models when estimating the P/E multiple[15]. Examining other key measures such as the mean error, IQ range and proportion of valuation errors less 15%, similar conclusions emerge (**Table A.10** in **Appendix A**). Finally, no clear pattern on median error differences across sub samples is observed. This is investigated in more detail in the following section.

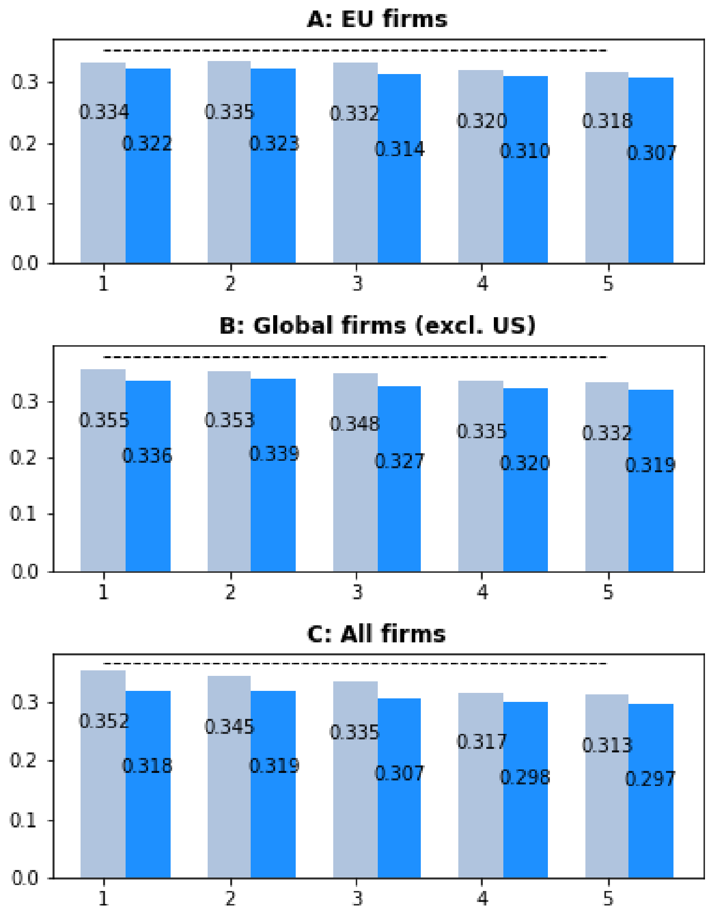
The estimation errors of P/E are higher than the ones found for EV/EBIT. Overall, this implies the existence of inter-multiple differences on accuracy when using SARD. EV/EBIT seems better suited for SARD compared to P/E. I do not pursue the question on inter-multiple accuracy any further since this is beyond the overall focus of this thesis.

[14]Test results confirm SARD-models are superior to the Industry benchmark. Results for both tests are enclosed in **Table A.11** in **Appendix A**.

[15]Statistical tests enclosed in **Figure A.3a** and **A.3b** in **Appendix A** confirm that INDSARD are superior at the all- and global sub samples. For EU-firms no significant difference is found between SARD5 and INDSARD5.

Figure 5.3: Median errors incurred for SARD- and INDSARD-models

Median errors for the P/E multiple. The x-axis denotes the number of selection variables included in the SARD/INDSARD-model. The left bars (grey blue) denote the median errors incurred when using the regular SARD-models, whereas the right bars (light blue) are the median errors obtained when applying INDSARD-models. The dashed horizontal lines illustrate the accuracy of the Industry benchmark. Each bar is labelled with the numerical value of the median absolute percentage error.



Source: Own calculations based on errors incurred from the SARD- and INDSARD algorithm

Cross-country differences when estimating P/E

Table 5.9 reports the results from a cross-country analysis performed for P/E. The analysis has been limited to include France, Sweden, South Korea and USA since these countries exhibited relatively distinct patterns in the analysis for EV/EBIT. Great Britain is additionally included because this country did not exhibit any significant peer pool dependency earlier. All country results can be found in **Appendix A**[16]. Country-wise differences still apply when estimating P/E. The highest errors are obtained for South Korea, whereas the lowest errors are found for the US. The key measures are generally smaller when applying INDSARD5 in **Panel B**. This generally corresponds with the initial analysis for EV/EBIT.

Examining the results across the various peer pools in **Panel A**, France, South Korea and the US all seem to be better off when selected in the home country opposed to the larger selection levels. Statistical evidence is very strong for South Korea (p-value < 0.000) but seems slightly weaker for the US and France due to ambiguous and insignificant test results through the Wilcoxon- and *t*-tests performed. Recalling the indistinguishable results for EV/EBIT, Great Britain seems to be benefit more from the all firm pool when using P/E. This is especially the case when examining the mean error, where British firms are clearly better off through the all firm selection as compared to the home country. Sweden also varies from the initial results. No indication is provided in favor of home firm selection when estimating P/E. This is opposed to prior findings and suggests the impact of peer pools can also be attributed to differences in-between multiples.

Examining **Panel B**, the preference for home firm selection seems to decrease through INDSARD5. The preference for home firm selection still holds for South Korean firms. The results in the US are ambiguous with a significantly strong preference for all firm selection through the *t*-test (p-value $= 0.003$) and an opposing preference for home selection through the Wilcoxon-test (p-value $= 0.055$). Estimates for Swedish firms are most accurate at the all firm pool using INDSARD5. In the case of these three countries, results generally align with the ones obtained earlier for EV/EBIT. However, this is not the case for France. Estimating EV/EBIT, France was clearly better off when restricted to the home peer pool. No evidence supports this tendency when using P/E. As seen in **Panel A**, British firms are better off in the larger peer pools. Here, no significant difference is found when evaluating differences across the larger peer pools.

The results obtained in **Table 5.9** confirm the inferiority of the all firm peer pool. Thus, the rejection of **H3** seems robust regardless of the choice of multiple. However, country-specific differences in regard to preferred peer pools are more ambiguous. The results of Sweden contrast findings from earlier. Furthermore, the pattern for both Great Britain and France also adds some ambiguity to prior findings. It generally adds to the

[16]**Table A.12**, **A.13**, **A.15**, **A.14**

Table 5.9: Combined results of country specific analysis, P/E multiple

Panel A displays the results using SARD5. **Panel B** reports the results using INDSARD5. Mean, median, IQ range and proportion less 15% of the absolute estimation error are displayed. Left hand parentheses display the number of firm-year observations for a given country as well as the count of valuation years with sufficient restricted data. The right hand side displays test sizes as well as p-values in parentheses below. Red colored test sizes indicate that all firm is significantly less accurate, blue colored test sizes the opposite. **Bold numbers** are addressed in the text.

Country		**Peer pool selection**				**Statistical tests**					
		Home	Region	Global	All	HomevsAll		GlobalvsAll		RegionvsAll	
				ex.USA		Wilcoxon	t-test	Wilcoxon	t-test	Wilcoxon	t-test
Panel A: Regular SARD-models											
France	Mean	0.397	0.404	0.400	0.403	**-2.4**	-0.9	-0.3	-0.8	0.7	0.2
(3,140)	Median	0.289	0.307	0.303	0.304	(0.018)	(0.376)	(0.767)	(0.434)	(0.476)	(0.808)
(20)	IQ range	0.379	0.376	0.357	0.372						
	Prop<15%	0.284	0.264	0.254	0.259						
Sweden	Mean	0.416	0.395	0.382	0.386	0.7	**2.9**	0.3	-1.0	0.0	1.2
(1,775)	Median	0.291	0.298	0.303	0.299	(0.491)	(0.004)	(0.774)	(0.313)	(0.969)	(0.245)
(20)	IQ range	0.382	0.367	0.360	0.364						
	Prop<15%	0.268	0.274	0.269	0.269						
Great Britain	Mean	0.459	0.452	0.442	**0.437**	1.6	**3.8**	**2.1**	1.5	1.6	**3.2**
(5,074)	Median	0.336	0.335	0.338	0.337	(0.103)	(0.000)	(0.039)	(0.139)	(0.103)	(0.001)
(20)	IQ range	0.412	0.403	0.400	0.392						
	Prop<15%	0.239	0.228	0.232	0.236						
South Korea	Mean	0.494	0.610	0.685	0.691	**-8.5**	**-10.5**	-0.7	-0.7	**-3.5**	**-5.8**
(1,996)	Median	**0.391**	**0.430**	**0.449**	**0.469**	(0.000)	(0.000)	(0.499)	(0.512)	(0.000)	(0.000)
(20)	IQ range	0.458	0.527	0.582	0.576						
	Prop<15%	0.206	0.184	0.176	0.170						
USA	Mean	0.373			0.373	**-3.7**	0.3				
(18,065)	Median	**0.276**			**0.282**	(0.000)	(0.797)				
(20)	IQ range	0.352			0.358						
	Prop<15%	0.290			0.279						
Panel B: INDSARD-models											
France	Mean	0.413	0.410	0.409	0.421	1.5	-0.7	0.8	**-2.7**	1.2	-1.8
(2,327)	Median	0.301	0.291	0.284	0.283	(0.144)	(0.472)	(0.421)	(0.008)	(0.217)	(0.080)
(19)	IQ range	0.393	0.370	0.377	0.384						
	Prop<15%	0.272	0.275	0.273	0.276						
Sweden	Mean	0.375	0.367	0.346	0.353	**2.0**	**2.3**	0.8	-1.2	**3.5**	**2.3**
(1,122)	Median	0.266	0.272	0.248	0.248	(0.050)	(0.021)	(0.448)	(0.231)	(0.001)	(0.023)
(19)	IQ range	0.378	0.338	0.331	0.345						
	Prop<15%	0.316	0.296	0.316	0.318						
Great Britain	Mean	0.439	0.422	0.418	**0.420**	**4.4**	**3.3**	1.7	-0.5	1.8	0.4
(4,395)	Median	0.319	0.310	0.310	0.311	(0.000)	(0.001)	(0.087)	(0.595)	(0.076)	(0.697)
(20)	IQ range	0.409	0.402	0.395	0.398						
	Prop<15%	0.251	0.266	0.261	0.267						
South Korea	Mean	0.477	0.523	0.569	0.580	**-3.8**	**-6.2**	0.0	-1.6	**-2.4**	**-4.7**
(1,351)	Median	**0.390**	**0.395**	**0.408**	**0.409**	(0.000)	(0.000)	(0.962)	(0.109)	(0.018)	(0.000)
(10)	IQ range	0.435	0.477	0.506	0.511						
	Prop<15%	0.204	0.207	0.196	0.200						
USA	Mean	0.357			0.351	**-1.9**	**3.0**				
(18,034)	Median	**0.259**			**0.266**	(0.055)	(0.003)				
(20)	IQ range	0.356			0.351						
	Prop<15%	0.316			0.304						

Source: Own calculations based on errors incurred from the SARD5 and INDSARD5 algorithm

complexity of the global dynamics detected, that the choice of the peer pool might also depend on the multiple in question. As noted earlier, an evaluation on the optimal peer pool and selection model for specific countries is beyond the scope the thesis. However, the discussion of these dynamics is highly relevant when addressing the SARD approach in a global setting. Thus, I will return to this discussion in the next chapter.

5.4.2 Error groupings

In this section, estimation errors are grouped and analyzed in regard to size and industry. Purpose is to examine whether the results could be driven by certain types of firms in the sample. In order to focus on the explicit effects from firm groupings, the analysis is restricted to focus on the all firm sample only. Note that these checks are simply different groupings of the overall estimation errors presented earlier.

Size

Size potentially influences the estimation errors for certain types of firms. The sample includes a high proportion of relatively small firms, where the growth and risk characteristics are likely to differ from those of the larger, publicly traded firms (Alford, 1992). To assess how firm size affects valuation accuracy, I partition the all firm sample into four distinct portfolios. As noted in the methodology chapter, my global sample does not include a clear size index identification. Instead, I group firms by the market capitalization rules provided in the official guidelines for the structuring of the S&P 1500. The Large cap portfolio consists of firms with a market capitalization greater or equal to 8,200 mio. USD. Firm values ranging from 1,600 to 8,200 mio. USD are included in the Mid cap portfolio and firms worth more than 450, but less than 1,600 mio. USD are defined as Small cap (Standard & Poor's, 2019)[17]. Finally, all firms with a market capitalization less than 450 mio. USD are denoted as the 'Rest index'.

Table 5.10 reports the median errors obtained through the SARD- and INDSARD approach as well as the Industry benchmark. The results indicate that accuracy increases with firm size across all models. The level of error is clearly lowest at the Large cap portfolio. The estimation errors exhibit a similar pattern when selection variables are added sequentially across all indices. Hence, SARD1 and INDSARD1 generally seem less accurate than SARD5 and INDSARD5. Statistical tests confirm the superiority of SARD4 and 5 compared to the Industry benchmark across all size indices[18]. Comparing the results in **Panel A** with the ones in **Panel B**, Large-, Mid- and Small cap firms generally benefit from the usage of INDSARD-models. For instance, Large cap incurs a median error of 20.9% when INDSARD5 is used opposed to a median error of 24.9% when SARD5 is used.

[17]These criteria are consistently updated over time. Indexing is based on the 2015-level and applied each year throughout over the sampling period.

[18]**Sub figure A.4a**, **A.4b**, **A.4c** and **A.4d** display heat maps of test results.

Table 5.10: Median errors grouped by size indices, SARD- and INDSARD combined

The ranking of each model is presented in the parentheses next to the error measure. Red coloring indicates lowest rank, Blue coloring implies best rank. **Bold numbers** are addressed in the text.

Panel A: Regular SARD-models						
	Benchmark	**SARD1**	**SARD2**	**SARD3**	**SARD4**	**SARD5**
Large cap	0.274 (3)	0.333 (6)	0.296 (5)	0.276 (4)	0.246 (2)	**0.243 (1)**
Mid cap	0.284 (5)	0.334 (6)	0.293 (4)	0.284 (3)	0.265 (2)	0.263 (1)
Small cap	0.312 (5)	0.338 (6)	0.310 (4)	0.301 (3)	0.286 (1)	0.288 (2)
Rest index	0.381 (5)	0.391 (6)	0.357 (4)	0.340 (3)	0.332 (2)	**0.328 (1)**
Panel B: Industry based SARD-models						
		INDSARD1	**INDSARD2**	**INDSARD3**	**INDSARD4**	**INDSARD5**
Large cap		0.261 (5)	0.232 (4)	0.218 (3)	0.209 (2)	**0.209 (1)**
Mid cap		0.275 (5)	0.254 (4)	0.244 (3)	0.235 (2)	0.234 (1)
Small cap		0.294 (5)	0.278 (3)	0.279 (4)	0.267 (2)	0.263 (1)
Rest index		0.373 (5)	0.348 (4)	0.340 (3)	0.336 (2)	**0.335 (1)**

Source: Own calculations based on errors incurred from the SARD- and INDSARD algorithm

Test results confirm this finding at conventional levels of significance[19]. The estimation errors obtained for the Rest index are significantly higher than the three other indices. No clear preference for INDSARD-models as compared to SARD-models is observed. The median error through INDSARD5 is 33.5% opposed to 32.8% at SARD5.

Summarizing, the median errors prove the influence of size on the precision of SARD estimates. This leaves some questions in regard to whether some of the cross-country differences observed would stand without the distortion from the 'super small' firms in the 'Rest index'. Larger firms generally seem more suited for SARD- and INDSARD-models.

Industry sectors

Table 5.11 reports median estimation errors grouped by GICS sectors on the all firm sample. A high degree of industry-specific variation exists in regard to the estimation precision across the eleven GICS sectors. The comprehensive models do not consistently yield the most accurate estimates. For instance, in **Panel A**, the lowest median error of 27.7% for the Telecommunication Services sector is obtained through the Industry benchmark. This generally contradicts with the overall finding in this thesis.

Analyzing the median errors obtained through the INDSARD-models reported in **Panel B**, accuracy generally seem to increase. Here, the Telecommunication Services sector obtains the lowest median error of 24.1% at INDSARD5. This is an improvement of almost 4% from the Industry benchmark in **Panel A**. This finding implies the informative value of selection variables might be reinforced when kept within industries.

Due to the industry-specific variation observed, prior results seem to be less general across industry sectors. In relation to prior findings on cross-country differences in various

[19]Heat maps of test results are provided in **Figure A.5a**.

Table 5.11: Median errors grouped by sectors, SARD- and INDSAD combined

Median errors grouped by sector. Rankings of each model is presented in the parentheses next to the error measures. Coloring scheme as pointed out in **Table 5.10**. **Bold numbers** are addressed in the text.

Panel A: Regular SARD-models	**Benchmark**	**SARD1**	**SARD2**	**SARD3**	**SARD4**	**SARD5**
Energy	0.371 (5)	0.409 (6)	0.350 (3)	0.352 (4)	0.329 (2)	0.329 (1)
Materials	0.322 (5)	0.338 (6)	0.291 (4)	0.281 (3)	0.269 (2)	0.263 (1)
Industrials	0.290 (5)	0.311 (6)	0.281 (4)	0.263 (3)	0.251 (1)	0.254 (2)
Consumer Discretionary	0.320 (4)	0.342 (6)	0.322 (5)	0.300 (3)	0.281 (1)	0.282 (2)
Consumer Staples	0.277 (5)	0.308 (6)	0.275 (4)	0.247 (3)	0.238 (1)	0.244 (2)
Health Care	0.316 (2)	0.380 (6)	0.372 (5)	0.356 (4)	0.324 (3)	0.316 (1)
Financials	0.389 (4)	0.444 (6)	0.389 (3)	0.421 (5)	0.371 (1)	0.377 (2)
Information Technology	0.370 (3)	0.405 (6)	0.403 (5)	0.383 (4)	0.354 (2)	0.352 (1)
Telecommuncation Services	**0.277 (1)**	0.313 (6)	0.275 (2)	0.281 (5)	0.278 (3)	0.279 (4)
Utilities	0.229 (5)	0.272 (6)	0.191 (3)	0.197 (4)	0.191 (2)	0.180 (1)
Real Estate	0.257 (1)	0.521 (6)	0.284 (2)	0.295 (4)	0.331 (5)	0.289 (3)

Panel B: Industry-based SARD-models	**INDSARD1**	**INDSARD2**	**INDSARD3**	**INDSARD4**	**INDSARD5**
Energy	0.345 (5)	0.318 (4)	0.296 (1)	0.299 (3)	0.298 (2)
Materials	0.304 (5)	0.276 (4)	0.268 (3)	0.266 (1)	0.268 (2)
Industrials	0.279 (5)	0.258 (4)	0.251 (3)	0.242 (1)	0.244 (2)
Consumer Discretionary	0.307 (5)	0.291 (4)	0.285 (3)	0.267 (2)	0.263 (1)
Consumer Staples	0.269 (5)	0.250 (4)	0.237 (3)	0.230 (2)	0.229 (1)
Health Care	0.302 (5)	0.288 (4)	0.283 (3)	0.262 (2)	0.258 (1)
Financials	0.376 (5)	0.328 (4)	0.326 (3)	0.315 (2)	0.312 (1)
Information Technology	0.352 (5)	0.346 (4)	0.332 (3)	0.318 (2)	0.317 (1)
Telecommuncation Services	**0.268 (5)**	0.251 (4)	0.242 (2)	0.243 (3)	**0.241 (1)**
Utilities	0.208 (5)	0.164 (1)	0.160 (3)	0.170 (4)	0.168 (2)
Real Estate	0.236 (5)	0.204 (1)	0.222 (4)	0.213 (3)	0.210 (2)

Source: Own calculations based on errors incurred from the SARD- and INDSARD algorithm

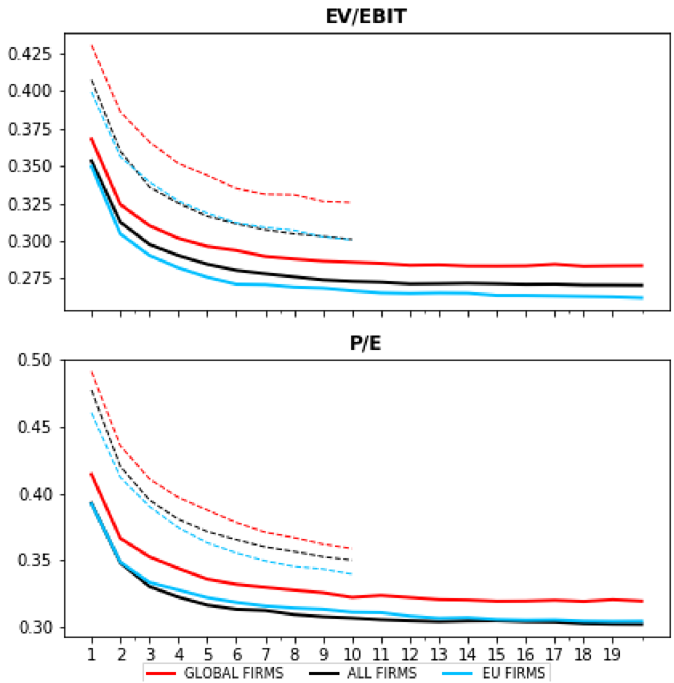
national markets, one could argue that some of the sector-specific variation might also influence specific country errors and the preference for distinct peer pools.

5.4.3 Number of comparable peers

In this thesis, all selection algorithms have consistently used six peers as the base for estimation. However, this amount might not be optimal when the sample is very large and presumably also affected by cross-country differences. In this subsection, I test whether the results differ when an alternative number of estimation peers is used. I perform a

Figure 5.4: Effect of the number of estimation peers

The graph displays the effect on the median estimation error when varying the amount of estimation peers for all firms, global firms (excl. USA) and EU firms. The dashed lines denote the Industry benchmark, whereas the solid lines denote SARD5.



Source: Own calculations based on errors incurred from the SARD algorithm and Industry level-up

sensitivity analysis on the median estimation error from the SARD5-model and Industry benchmark when varying the number of estimation peers in the algorithm. The analysis is performed on the three sub samples analyzed throughout this thesis. In order to increase robustness, I expand the analysis to include median errors for both the P/E- and EV/EBIT multiple. **Figure 5.4** displays the results of the sensitivity analysis. The tendencies observed seem similar across the two multiples and the three different sub samples examined. Benchmark errors are clearly larger than SARD5. This confirms the overall superiority of the SARD approach over the Industry benchmark regardless of the number of estimation peers. Interestingly, errors do not seem fully minimized when six estimation peers are applied. The trend lines seem to flatten entirely around 10 peers for both multiples. However, the incremental benefits from an increased number of estimation peers seem too small to represent a clear methodological bias underlying the prior results obtained. This finding is in line with Knudsen et al. (2017).

5.4.4 Time variation of errors

Naturally, I am concerned whether the results obtained are robust over time. This relates to the findings on the overall performance of the SARD approach compared to the Industry benchmark and whether the results of specific country peer-pools seem robust over time. **Figure 5.5** reports subplots containing the median error over time displayed for both the larger sub samples analyzed and selected country errors. **Subplot A** compares the median errors of SARD5 with the ones for the Industry benchmark for all-, global-, and EU firms respectively. The plot confirms the superiority of SARD5 over the Industry benchmark over time. Time patterns seem to correlate across models and sub samples. Distinct peaks in valuation errors are found in year 2000 and again in the period 2007-2009. This corresponds with the timing of the dot.com bubble in 2000 and the global credit crisis starting in 2007. Finally, an ascending trend can be observed in recent years. **Subplot B** reports the median errors for France, Great Britain, South Korea, Sweden and the US. Comparing the variation across countries little correlation is observed. Most striking is South Korea, where median errors often move opposite to the direction of the errors obtained for the US and EU-countries. In comparison, the time pattern for Sweden, France, Great Britain and the US generally seem more aligned. The effect of peer pools varies across countries. The median errors for US- and British firms generally follow an identical pattern regardless of the peer pool. For US firms, this somehow contradicts the pattern detected through the statistical tests in **Section 5.3**. Opposite, Great Britain is in line with the indistinguishable results obtained through both key measures and tests. The dashed lines (i.e. home peer pool) for Sweden and France exhibit lower error over time as compared to the solid lines (i.e. all peer pool). This supports prior findings implying the preference for the home peer pool for French- and Swedish firms. Finally, South Korea exhibits a highly surprising pattern. The trends of the dashed- and solid purple lines seem to move in opposite directions in the period 2000 to 2015, while converging in recent years. This confirms the importance of the choice of peer pool when using SARD in a broad global setting with different national markets.

In summary, the superiority of the SARD approach over the Industry benchmark seems robust over time. Estimation errors are time dependent. Cross-country differences clearly affect the level of error for particular country firms. Trend lines vary from country to country implying cross-country differences affect the precision of SARD estimates. Furthermore, the choice of peer pool seems particularly important for France, Sweden and South Korea. Overall, this is in line with prior findings.

Figure 5.5: Median estimation error in each year, selected countries and analyzed sub samples

Subplot A displays the median errors of EV/EBIT through SARD5 versus the Industry benchmark over time. SARD5 is the solid lines, whereas the benchmark is dashed. **Subplot B** plots SARD5 all firm selection versus SARD5 home country selection over time for France, Great Britain, South Korea, Sweden and USA. All firm selection is the solid lines and home country is dashed.

Source: Own calculations based on errors incurred from the SARD5- and Industry level-up algorithm

Chapter 6

Discussion

This chapter is devoted to the discussion and interpretation of the results obtained in **Section 5**. First, I relate my findings to previous studies on peer selection. Second, I interpret own results and discuss these in regard to the practical relevance of the SARD approach in a broad global setting. I present the limitations of my work and propose some areas that could be relevant for future research.

6.1 Relation to prior studies

Through the large sample analysis, the peer groups identified through the SARD approach are found to be more accurate than peer groups based on industry classification alone. This is in line with the fundamental school of thought and corresponds with the findings of Knudsen et al. (2017). The level of median absolute percentage errors in my analysis of the all firm sample varies from 34.8% (SARD1) to 28% (SARD5). In comparison, the variation of errors across models for Knudsen et al. (2017) ranges from 29.2% to 22.2%. Thus, SARD errors generally seem to increase through the effects of a broad global sample with different national markets. The same pattern holds when examining the results from the P/E multiple. Knudsen et al. (2017) did not find a significant difference between SARD4 and SARD5. The inclusion of EBIT-margin in SARD5 seems slightly more useful through my results. Compared to the broader literature results are similar to the findings of Bhojraj and Lee (2002) and Herrmann and Richter (2003) implying peers should be selected through similarities in profitability, risk and growth. The superiority of the SARD approach over industry classification seems in particular evident through the more comprehensive SARD-models. This finding exemplifies the usefulness of SARD opposed to prior fundamental-based selection models. S. Nel et al. (2014) note that the combination of valuation fundamentals exhibits lower valuation errors relative to multiples based on single-factor valuation fundamentals in the South African market. However, due to a limited sample size the number of comparable peers found when combining more than three valuation fundamentals was negligible and consequently excluded from the analysis (S. Nel et al., 2014, p. 634). SARD is independent of sample size and therefore seems more suitable for a fundamental selection in smaller global markets compared to 'traditional' selection models in prior literature.

Industry-based SARD-models are found to be superior to the regular SARD-models. This is in conformity with ideas proposed by Knudsen et al. (2017) and explicitly tested in my analysis. The INDSARD-models used in this thesis vary from the ones suggested by Knudsen et al. (2017), since I combine the level-up rule from Alford (1992) to accommodate the potential issue of limited industry peers. The incremental benefits in estimation accuracy through the combination of industry and fundamentals are in line with Cheng and McNamara (2000), who found that peer groups based on industry and ROE were significantly more accurate than peer groups based on ROE only.

Robustness checks provided evidence that the precision of SARD varies with industry sector and the size of firms. The large variation of errors across size groups is not in conformity with the original paper, where the SARD approach was found to be relatively robust across firms of different sizes (Knudsen et al., 2017). My results are more in line with Alford (1992) and Cheng and McNamara (2000), who both suggest larger firms incur lower estimation errors than smaller firms. In regard to the industry grouping of error, Knudsen et al. (2017) observed some industries seem more suited for SARD than others. W. Nel and Le Roux (2017) identified a similar pattern for South African firms when analyzing the optimal selection method for 28 different industry sectors. Furthermore, both papers suggest the choice of multiple is highly sector specific. Following the delimitations of the thesis, this issue has not been examined further. Finally, the robustness checks on the optimal number of estimation peers as well as the time dependency of SARD errors correspond with prior findings. Knudsen et al. (2017) only found a marginal increase of precision when more estimation peers were included. The time-error variation detected for selection models is aligned with both Knudsen et al. (2017) and Dittmann and Weiner (2005).

The analysis of cross-country differences led to rejection of the hypothesis that more potential peers yield more accurate estimates. This overall conclusion is in line with the findings of Dittmann and Weiner (2005) suggesting the choice of the peer pool is just as important as the selection rules. However, when comparing the country-specific results high variation exists. Dittmann and Weiner (2005) found most European countries were better off when the peer pool included EU-member firms. Conversely, my analysis only finds minor support in favor of the EU peer pool (i.e. Region) for Danish- and Italian firms. In all other cases, the larger peer pools seem to be statistically indistinguishable from each other. More interestingly, French firms clearly seem better off in the home market. In relation to this, Dittmann and Weiner (2005) suggest that Denmark, Great

Britain and the US should be restricted to the home market only. My results on Danish firms are in direct opposition to this. The findings on Great Britain was indistinguishable in the choice of peer pool. In fact, only the US results seem to correspond with Dittmann and Weiner (2005) favoring home country selection. Country-specific results do not seem fully robust when estimating P/E. Thus, it seems fair to conclude that high ambiguity of country-specific results exists[1].

The differences of findings compared to those found in prior studies should primarily be explained by variation in the sample selection and research design. As mentioned in **Section 4.2.2**, the SARD algorithm developed corresponds one-to-one with the algorithm of the original paper. Thus, differences should predominantly be attributed to the characteristics of the broad global sample. Over a twenty-year period, the all firm sample includes 50,193 firm-year observations from 27 countries, opposed to 12,350 firm-year observations from the US only in the original paper (Knudsen et al., 2017). The combination of global- and US data most likely drives the higher errors in my analysis. Furthermore, I do not restrict my data at the GICS 6 level, which means more firms are kept in the data used for modeling. This could be part of the explanation why errors are generally higher. However, looking isolated at the errors incurred for the US in the home market, this implication seems less relevant explaining the differences in errors from the original paper[2].

The variation of findings from Dittmann and Weiner (2005) cannot be explained by differences in data only. In this thesis the dependency of peer pools is evaluated through the SARD approach, whereas Dittmann and Weiner (2005) used the selection rules as put forward by Alford (1992). The 'most comprehensive' selection model in their study included ROA and TA, whereas I apply either SARD5 or INDSARD5. Variation in results could occur because of the clear difference in-between the selection models applied. Furthermore, Dittmann and Weiner (2005) explicitly aim to locate the optimal selection model for each country examined. This includes evaluation of both selection rules and peer pools. I do not claim to have found an optimum in my analysis, but rather provided evidence that the two models examined are clearly affected by country-specific factors. Finally, the nature of samples could also affect differences in results. The sampling period of Dittmann and Weiner (2005) ranges from 1993 to 2002. Thus, my sample only overlap in two years. Furthermore, my sample includes the mandatory adoption of the IFRS in 2005 for EU-members, which means a predominant part of countries in my sample applies identical accounting standards. The results of Dittmann and Weiner (2005) is expected to have been influenced more heavily by differences in accounting standards across countries.

[1] Dittmann and Weiner (2005) did not perform robustness checks on alternative multiples in their study.

[2] Recall US firms incur a median error of 24.5% at SARD5 in the home market (2000-2019), whereas Knudsen et al. (2017) incurred a median error at SARD5 of 22.2% (1995-2014). Considering potential time effects these two measures articulate fairly well regardless of the different GICS restriction criteria.

6.2 Interpretations and practical relevance

Up until now the performance of SARD in a broad global setting has only been assessed through the key measures and the relation to prior literature. Evidence suggests the SARD approach offers a more accurate selection model than when peer groups are based solely on industry classification. In addition to this, accuracy is further increased when SARD is performed within industries. However, when assessing the country-specific performance it clearly shows the performance of the SARD approach varies from country to country. The all firm peer pool does not yield more accurate estimations than the smaller ones. This is not in line with the fundamental idea of thought and suggests country-specific factors weigh just as much as the SARD selection variables when selecting peers across borders. In order to address the practical relevance of my findings these eventual idiosyncrasies need to be discussed. Petersen et al. (2017) argue an ideal valuation model includes unbiased estimates and realistic assumptions. The pattern detected throughout this thesis indicates this might not be the case for SARD when applied globally. In the following subsections, I highlight possible explanations for why country-specific idiosyncrasies could affect the performance of SARD in a broad global setting consisting of different national markets. These are also intended to serve as practical key insights which should be considered by a global analyst using the SARD approach across country borders.

6.2.1 Value relevance of accounting figures across borders

According to Schreiner (2009) value relevance reflects the ability of accounting information to capture or summarize the information that affects market values or stock returns. Following the fundamental idea of SARD, the selection variables capture the underlying value drivers of multiples. Hence, selection variables should be value relevant. Furthermore, it goes along the identification based on the absolute ranked differences that selection variables should be fully comparable across the matrix of ranks. Meaning biased estimates will occur if the fundamental proxies do not exhibit identical economic characteristics. The ambiguity of results presented in the cross-country analysis suggests the informative power of SARD selection variables varies across countries.

One potential explanation for the difference in results can be attributed to accounting differences. A predominant part of my sample applies either US GAAP or IFRS. All things being equal, the higher estimation errors in the all firm sample compared to home country-markets could be explained by the differences in the requirements of the two accounting entities. For instance, the Last-in-First-out (LIFO) costing methodology is accepted under US GAAP but precluded under IFRS applying either the First-in-First-out (FIFO) or weighted average method (pwc, 2018). Under the LIFO method the newest inputs are expensed before the older inputs in the inventory, whereas the FIFO method incur the cost of the newest inputs. Thus, with increasing input prices, two fully identical firms (that would have been perceived as natural peers) would elicit completely different operating margins and taxable incomes when one applies LIFO and the other uses FIFO. The SARD approach would not capture these differences and select alternative peers that seem more similar based on the outlook of the accounting fundamentals. The choice of costing methodology is just one example of the potential biases stemming from accounting differences in a broad global setting. Since the all firm sample includes a high proportion of US firms, one should expect these biases would be most dominant in other global countries when peer selection happens through the all firm sample. However, while results indicate home-country selection should be preferred for some countries, little statistical evidence is provided for an exclusion of US firms entirely. In most cases, the all firm peer pool (US GAAP and IFRS combined) appears statistically indistinguishable from the regional pool (IFRS only). For EU-member countries in particular, a preference for the regional peer pool in favor of the all firm peer pool should be expected, if idiosyncrasies primarily exist because of accounting differences in-between the IFRS and US GAAP. This is clearly not the case. Alternatively, Japanese firms using domestic GAAP could also drive estimation biases following similar arguments as above. However, the effect from these are presumed to have less impact due to the smaller proportion of total firm-year observations.

The differences in-between US GAAP and IFRS are not the only area in which accounting methods potentially influence the precision of the SARD approach. Firms reporting under IFRS enjoy a great deal of accounting flexibility which can easily disturb the economic interpretation of the selection variables included in SARD. Issues could include the extent to which certain R&D costs should be expensed as incurred or capitalized as assets subject to future depreciation (Petersen et al., 2017, p. 438). Another example is firms' decision on whether lease contracts should be classified as *operating leases* recognized as expenses on the income statement or whether these should be treated as *financial leases* recognized as leased assets with offsetting liabilities on the balance sheet (Petersen et al., 2017, p. 516). In both cases, these decisions do not change the underlying economics of the firm, however the outlook of profitability measures and financial ratios will vary dependent on the accounting decision. Ultimately, accounting differences will affect the comparability in-between SARD selection variables. This bias is presumed to exist regardless of the boundaries set for the pool of potential peers. Based on the empirical results, it is difficult to conclude whether accounting differences drive idiosyncrasies alone. Most likely these underlie some of the ambiguous results obtained.

Another interpretation is that the value relevance of accounting information varies from country to country. Accounting flexibility can be motivated for many reasons. Often this stems from firms seeking to optimize their particular circumstances in regard to the local rules and regularities in their home country. For instance, the corporate tax bill is strictly country-specific and directly related to the level of operating income for the firm. Maybe firms from countries with higher tax rates actively seek to expense more R&D in order to minimize the tax bill? If firm decisions are driven by these local incentives it decreases the value relevance. One could argue the level of comparability across selection variables decreases proportionately with the number of different countries included in the sample. From an accounting perspective this could explain some of the ambiguity detected in the cross-country analysis. The institutional circumstances in some countries favors that home firms should be treated separately opposed to a selection across country borders. The results related to Asian-pacific countries could be interpreted as a result of this. The precision of SARD in both Japan and South Korea were considerably more inaccurate compared to most Western countries. The home country peer pools were consistently preferred to the regional pools for both countries, which indicates that these particular countries are generally more heterogeneous than EU-member firms for instance. In fact, even when restricted to the home market only, estimation errors were still considerably higher when compared to other countries. Hence, from a value relevance perspective, accounting information from this region seems less informative in regard to the market value of firms relative to what was generally observed for other OECD countries. Alternatively, the relative precision of SARD for Western firms compared to Asian-pacific firms can be attributed to a higher degree of capital market efficiency in the European- and US markets[3]. The estimation errors for France and the US were much lower. However, in both cases it seemed sub-optimal to include firms from foreign countries in the peer pool. Along the value relevance perspective, this generally suggests the link between accounting fundamentals and market prices is blurred by country-specific idiosyncrasies for these particular countries.

Since SARD treats all potential peers equally based on the selection variables, this points in the favor of biased estimates when value relevance is not addressed for a particular country prior to the selection. Furthermore, accounting differences will naturally cause some bias if selection variables are not fully comparable. The opposing results observed when using P/E alternative to EV/EBIT suggest the choice of multiple also influences the precision of SARD estimates. Ultimately, the selection method does not fully capture all eventual complexities present in a global setting with different national markets. This implies country-specific value relevance should always be assessed when applying SARD globally. If not, analysts- and managers might make sub-optimal decisions when certain groups of countries clearly stand out as 'less-suited for modeling'. Alternatively, new selection variables could be added to screen for regional- or country specific differences as suggested in Bhojraj et al. (2003). The fundamental idea underlying SARD is contradicted

[3]A similar argument was put forward by Herrmann and Richter (2003) when they compared the relative performance of US firms compared to European firms in their 1997-1999 study.

in both cases.

6.2.2 Size as a proxy for cross-country differences?

The cross-country analysis shows how the size of the home country peer pool, measured by the number of firm-year observations in a particular country, somehow correlates with the preference for home country selection. It was primarily countries with a relatively high number of observations which were better off using the home country peer pool as compared to the alternative larger peer pools. This is line with prior discussions that country-specific factors influence the value relevance of the SARD selection variables. If this is indeed the case, it should always be preferable to restrict the peer pool to the country market only and include as many domestic firms as possible. These results were nuanced through the robustness check on size indices in **Section 5.4.2**. Here, estimation errors were highly dependent of the size of firms as measured by the market capitalization. The Large cap index incurred considerably lower estimation errors than the Rest index. Even when compared to both Mid- and Small cap firms the median errors from the Large cap firms were roughly 2% lower when applying SARD4 and 5. In line with the classic theoretical argument, that smaller firms have greater risk than larger firms (Banz, 1981), one could add these very large firms are less susceptible towards the influence from cross-country differences. Following a phenomenon observed by Zhang (2006), the value relevance could be affected by the 'information uncertainty', which exists for smaller firms that often provide poorer financial information to markets and investors. With famous brands and operations all over the world, the Large cap firms presumably hold significantly less 'information uncertainty' compared to smaller firms. In addition to this, the stocks of smaller firms are often presumed to be less liquid than the stocks of larger firms. Hence, the market value of smaller firms could also reflect the additional liquidity risk for which global investors require a return (W. Liu, 2006). Combined with these theoretical considerations, the findings from **Section 5.4.2** imply SARD might not work as well for very small firms. Thus, the country variation observed could be attributed to the general size of firms in the respective domestic market opposed to actual number of home-country firms available. Apart from the US, many global markets only include few very large firms. For large firms in 'smaller' national markets, it will be more challenging to locate the natural peers based on economic fundamentals when selection is restricted to the home market only. It seems more likely that the natural peers of the largest firms are better identified in the world market regardless of the number of small firms in the home country. Opposite, the natural peers for very small firms should be selected from the same country. These firms most likely will be more dependent on the economic situation of the region in which they operate (Schreiner, 2009). For instance, a small firm from a region in an economic downturn would trade at lower multiples than the same firm would do when the local economy is booming. Data does not provide any concluding answer. Indeed, the proportion of Large cap firms vary across the different countries examined. In Japan, only 9% of all firm-year observations are included in the artificial Large cap index, whereas this proportion is found to be 15 and 20% for Germany and France respectively. However, both France and Japan are better off through the home selection. The results for Germany were indistinguishable. Thus, direct causality between the size of firms and the preferred selection pool cannot be inferred through the empirical data.

The trade-off between the size of firms and the number peers available in the national market represents an interesting dynamic related to the performance of SARD in a broad global setting. If both country-specific factors as well as the size of firms influence the optimal choice of selection method for a particular national market, these cross-effects should be analyzed carefully by the analyst using SARD. It is not an easy task to asses these cross-effects. For instance, how should a portfolio manager covering LVMH[4] set the peer pool knowing both country- and firms size dependency influence estimation accuracy in the French capital market? or should the smaller Danish firms be compared with other small firms from the rest of Europe because of an insufficient number of domestic comparable firms? Raising these questions generally contradicts the fundamental school of thought but seems highly relevant to assess the performance of the SARD approach in a broad global setting. I leave it up to future research to examine the impact of these potential cross-effects.

6.2.3 Industry classification as a safeguard for cross-country variation?

One of the central findings in this thesis is that INDSARD-models generally increase estimation precision. As previously mentioned, this is in line with ideas suggested by Knudsen et al. (2017). Overall, this implies the combination of both fundamentals- and industry classification should be preferred when identifying comparable firms. In **Section 4.3.1** the all firm sample incurred the lowest level of errors across the analyzed key measures when INDSARD-models were used. This was opposed to the findings related to the regular SARD-models in **Section 5.2.1** where the lowest estimation errors were incurred at the smaller sub sample consisting of EU-firms only. Industry classification generally seems to benefit more from a larger peer pool as it is more likely to locate comparable industry peers when the sample size increases. A similar pattern was detected in the cross-country analysis. For a predominant part of the countries examined, estimation errors from the all firm peer pool were considerably lower when using INDSARD5 compared to when SARD5 was applied. Furthermore, the preference for home country peer pools was decreased. In fact, France was the only country in which home selection was consistently preferred across the two models examined. The pattern observed throughout the thesis indicates

[4]LVMH Moët Hennessy Louis Vuitton

the industry affiliation of firms - to some degree - captures some of the variation affecting estimates more heavily trough the regular SARD-models. One could perceive industry classification as a 'safeguard' towards some of the cross-country differences present in a broad global setting. Proponents of the industry-based selection argue firm comparability within industries is enhanced since industry firms often use similar accounting methods (Alford, 1992). For instance, logistic firms might monitor accounting methods of industry competitors in regard to the decision on whether lease contracts should be treated as *operating*- or *financial* leases (**cf. Section 6.2.1**) in order to attain comparability for shareholders. Also, the differences between IFRS and US GAAP weigh less when the SARD selection is initially filtered to include similar GICS firms only. Two identical firms applying FIFO and LIFO respectively would offer significantly different operating margins when input prices are increasing. However, if the two firms share an industry, they would still be included in the same initial cluster used for the final peer selection. Following this idea of thought, biases existing because of differences in accounting methods could be mitigated when selection happens within the GICS level-up. This interpretation rests on the notion that the overall explanatory power of SARD increases when the selection algorithm is initially 'pointed in the right direction' being industry firms. For instance, this was seen for the Telecommunication Services sector in **Section 5.4.2**. Here, the Industry benchmark was more accurate than all other SARD-models, suggesting industry classification alone capture more information than any of the SARD selection variables in this particular sector. However, when used in conjunction with industry, the median estimation error was 4% lower at INDSARD5 including all selection variables. Thus, some accounting fundamentals might be redundant when applied across industries, but value relevant within identical GICS levels. The analysis of size indices provided a similar pattern. INDSARD-models significantly decrease median errors for firms located in the both the Large-, Mid- and Small cap indices. Overall, the relatively 'simplistic' treatment of the ranked selection variables in SARD benefits from industry classification working as an initial screener for some of the possible variation described in prior sections. Hence, the term 'safeguard' seems plausible when the sample is large and cross-country differences exist.

Some important aspects need to be addressed when evaluating the usefulness of industry-based SARD-models opposed to regular SARD-models. First of all, the firms included in the Rest index exhibited a completely different pattern. Median errors only decreased marginally using INDSARD1 and 2, whereas both INDSARD4- and 5 were dominated by the corresponding regular SARD-models. This suggests the 'strength' of the industry safeguard is more questionable when selecting peers for smaller firms. It also adds to the overall finding that the SARD approach generally seems less suited for small firms. Second, INDSARD-models can only be performed when a sufficient number of industry peers are available. Even when applying the industry level-up, a high proportion of domestic firms were eliminated from the home country peer pool. Thus, the tendencies observed in the cross-country analyses of SARD5- and INDSARD5 are not fully comparable. It cannot be denied the results on the preferred peer pools arise from the exclusion of particular firms exhibiting distinct country-specific tendencies. Finally, the industry error groupings imply high variation across industry sectors. For instance, the Utilities sector incurs estimation errors less than half of the estimation errors obtained in the Information Technology sector. As a result, the 'strength' of the safeguard must necessarily rely on the actual industry in question. One should be careful ascribing cross-country differences to what is really industry related variation. Various domestic markets are dominated by different industries. For instance, the Danish market relies more heavily on Health Care contrary to the Swedish market more influenced by Information Technology firms[5]. All prior research implies the performance of selection models is highly influenced by the industry in question (Knudsen et al., 2017; W. Nel & Le Roux, 2017). Maybe SARD-models should be tailor made to accommodate both industry- and country-specific characteristics? Alternatively, the optimal peer pool should be linked to the particular industry affiliation of the global firm being valued.

In summary, industry classification should be perceived as a 'safeguard' for large firms in large samples only. However, truly optimal choices only arise when both the peer pool and the selection model are assessed in conjunction with one another.

6.3 Limitations

One important limitation is that results are influenced by accounting differences. In order to clarify potential biases from accounting differences between IFRS and US GAAP, I consistently analyze various sub samples that is expected to offer incremental convergence in the accounting standard applied. By examining specific country errors and the dependency of peer pools, I address the particular countries which could be affected by accounting biases. The methodological setup is in line with Dittmann and Weiner (2005) and Herrmann and Richter (2003) who purposely compare the accounting information for both global- and US firms. The purpose is to investigate the performance of the SARD approach in a broad global setting. Thus, accounting differences are generally perceived as a premise central to the problem area in this thesis. I do not argue to have found the optimal selection model for any of the 16 countries examined. Thus, another limitation is, that results may stem from sub-optimal decisions in specific country markets. It cannot be denied that both the peer pool preference and estimation errors could change through the usage of other SARD- or INDSARD-models. My results are limited to the EV/EBIT

[5]Based on own data: Swedish observations from Health Care and Information Technology are 9.6 and 18.0% respectively. Danish observations are 23 and 6% respectively.

multiple only. Robustness checks have shown the tendencies observed for particular countries might change with the choice of multiple. Thus, some bias in regard to the choice of multiple exists. Finally, my data is restricted to include the OECD countries as of year 2000 only. Thus, results might not be applicable for firms in developing countries, emerging markets etc.

6.4 Relevance for future research

Several aspects still need to be addressed regarding the performance of the SARD approach in a broad global setting. First of all, the suggested cross-effects between the size of firms and the number potential peers within the home market should be examined further. A predominant part of literature suggests that larger firms generally yield lower valuation errors (Alford, 1992). Future studies could focus on a SARD selection restricted to specific country markets (not being the US). By partitioning the country sample into quintiles on the basis of firm size, some of the cross-effects between the number of available peers and the size of firms could be analyzed on a country level. The optimal choice of the peer pool could also be analyzed through a country-specific analysis. It was suggested that peer pools could be selected based on existing links between country firms. Thus, future studies could address whether the performance of SARD increases when peer pools are methodologically selected based on economic, regulatory or industry-specific characteristics. My results suggest that some regions are more suited for an application of SARD than others. In direct extension of my results, the higher errors incurred for Asian-pacific countries should be addressed. Additionally, one could examine the performance of the SARD approach in emerging markets such as South Africa or Latin America. Finally, one could examine the potential biases from accounting differences across borders. This could be addressed through an event study linked to the mandatory IFRS adoption in the EU. Alternatively, future studies could examine the performance of SARD based on selection variables less susceptible towards accounting differences. These alternatives could include cash flow figures, additional 'objective' estimates from I/B/E/S or the total number of employees from the annual report.

Chapter 7

Conclusion

The objective of this thesis was threefold. On the basis of **H1** it was examined whether the peer selection of firms through SARD yields higher estimation accuracy than when peers are selected based on industry classification alone. Furthermore, **H2** investigated whether the combination of the SARD approach and industry classification offer a higher a degree of valuation accuracy vis-a-vís EV/EBIT multiples whose peer groups are based on the regular SARD approach. Finally, the purpose of **H3** was to analyze whether the accuracy of the two selection methods was more precise when the pool of potential peers is as large as possible compared to estimations originating from smaller peer pools.

The empirical results confirm **H1**. The evidence suggests the combination of more than two fundamental selection variables consistently yields significantly more accurate estimates of EV/EBIT. Thus, the applied SARD approach is superior to industry classification in a broad global setting. On average, the median errors of the predicted EV/EBIT multiples whose peer firms were based on the most comprehensive SARD-models were 4% lower than when estimated through industry alone. When SARD-models are based on ROE alone industry classification is a more accurate selection strategy.

H2 is also confirmed. The combination of SARD and industry classification significantly increases estimation precision compared to the corresponding regular SARD-model. Thus, the industry-based SARD-models applied are superior to the regular SARD-models in a broad global setting. In the all firm sample the average increase in precision is 3.2% vis-a-vís estimates from the regular SARD approach.

H3 is clearly rejected. Prediction accuracy is not positively correlated with the number of potential peers included in the peer pool. Thus, larger global peer pools are inferior when using the SARD approach. Using the regular SARD approach France, Japan, South Korea and the US are all better off when the pool of potential peers is restricted to home country firms only. The larger peer pools seem statistically indistinguishable from one another. The preference for home country peer pools decreases when using industry-based SARD-models. France is the only country in which clear home country preference is persistent across the two methods examined. However, no evidence is provided that estimation precision is higher through the largest peer pool when using the industry-based SARD-model.

The assessment of **H3** further implies large variation in estimation precision across countries. Results from the regular SARD approach reveals country-specific median estimation errors vary from 23.1 to 38.4%. Several factors potentially underlie the cross-country variation. Hence, industry-specific characteristics in individual country markets might affect the precision of selection variables and the preference for particular peer pools. Also, larger firms are generally more suited for SARD than smaller firms as estimations seem more robust to cross-country differences. Asian-pacific firms generally seem less suited for SARD as compared to Western firms, implying the geography of firms influences model accuracy. Finally, differences in accounting methods in-between firms can cause biased estimates when using SARD.

Median errors are generally lower applying industry-based SARD-models. The combination of lower estimation errors and indistinguishable peer pools leads to the interpretation that industry classification might work as a 'safeguard' against cross-country differences in broad global samples when used in conjunction with SARD. However, this interpretation only applies for larger firms.

Ultimately, the research results from a broad global setting consisting of different national markets concur with prior evidence. The SARD approach is superior to industry when identifying comparable firms for EV/EBIT multiple valuations. Incremental benefits arise when SARD is used in conjunction with industry. However, the performance of SARD seems highly country dependent. Acknowledging cross-country differences affect SARD estimations contradicts the fundamental school of thought underlying the SARD approach. Therefore, the SARD approach should capture other factors than sole fundamentals in order to maximize estimation accuracy in a broad global setting. The best-performing SARD-model in France might not be the most preferable one in the US. Thus, the global analyst might make sub-optimal decisions if she does not consider country-specific factors and the choice of peer pool before applying the SARD approach globally.

Bibliography

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, *120*(2), 701–728.

Alford, A. W. (1992). The effect of the set of comparable firms on the accuracy of the price-earnings valuation method. *Journal of Accounting Research*, *30*(1), 94–108.

Andersen, I. (2008). *Den skinbarlige virkelighed: Om vidensproduktion inden for samfundsvidenskaberne*. Samfundslitteratur.

Baker, M. & Ruback, R. (1999). Estimating industry multiples. *Harvard University*.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, *9*(1), 3–18.

Bhojraj, S. & Lee, C. M. (2002). Who is my peer? a valuation-based approach to the selection of comparable firms. *Journal of accounting research*, *40*(2), 407–439.

Bhojraj, S., Lee, C. M., & Oler, D. K. (2003). What's my line? a comparison of industry classification schemes for capital market research. *Journal of Accounting Research*, *41*(5), 745–774.

Cheng, C. A. & McNamara, R. (2000). The valuation accuracy of the price-earnings and price-book benchmark valuation methods. *Review of Quantitative Finance and Accounting*, *15*(4), 349–370.

Clarke, R., Davies, S., & Waterson, M. (1984). The profitability-concentration relation: Market power or efficiency? *The Journal of Industrial Economics*, 435–450.

Collins, N. R. & Preston, L. E. (1966). Concentration and price-cost margins in food manufacturing industries. *The Journal of industrial economics*, 226–242.

Cooper, I. A. & Cordeiro, L. (2008). Optimal equity valuation using multiples: The number of comparable firms. *Available at SSRN 1272349*.

Crain, M. A. (2011). A literature review of the size effect. *Available at SSRN 1710076*.

Damodaran, A. (2016). *Damodaran on valuation: Security analysis for investment and corporate finance*. John Wiley & Sons.

Demirakos, E. G., Strong, N. C., & Walker, M. (2004). What valuation models do analysts use? *Accounting horizons*, *18*(4), 221–240.

Dittmann, I. & Maug, E. G. (2008). Biases and error measures: How to compare valuation methods. *ERIM Report Series Reference No. ERS-2006-011-F&A*, 2006–07.

Dittmann, I. & Weiner, C. (2005). Selecting comparables for the valuation of european firms. *Available at SSRN 644101*.

Gode, D. D. K. & Ohlson, J. A. (2006). A unified valuation framework for dividends, free cash flows, residual income, and earnings growth based models. *James A., A Unified Valuation Framework for Dividends, Free Cash Flows, Residual Income, and Earnings Growth Based Models (February 14, 2006)*.

Guba, E. G. (1990). The paradigm dialog. In *Alternative paradigms conference, mar, 1989, indiana u, school of education, san francisco, ca, us*. Sage Publications, Inc.

Heldbjerg, G. (1997). *Grøftegravning i metodisk perspektiv: Et videnskabsteoretisk og metodologisk overblik*. Samfundslitteratur.

Herrmann, V. & Richter, F. (2003). Pricing with performance-controlled multiples. *Schmalenbach Business Review*, *55*(3), 194–219.

IFRS. (2018a). IFRS Foundation use of ifrs standards around the world. https://www.ifrs.org/-/media/feature/around-the-world/adoption/use-of-ifrs-around-the-world-overview-sept-2018.pdf. Accessed: 2019-03-07.

IFRS. (2018b). Jurisdictional profile switzerland ifrs. https://www.ifrs.org/-/media/feature/around-the-world/jurisdiction-profiles/japan-ifrs-profile.pdf. Accessed: 2019-03-07.

Knudsen, J. O., Kold, S., & Plenborg, T. (2017). Stick to the fundamentals and discover your peers. *Financial Analysts Journal*, *73*(3), 85–105.

Koller, T., Goedhart, M., & Wessels, D. (2005). The right role for multiples in valuation. *McKinsey on Finance*, (15), 7–11.

Kotler, P., Keller, K. L., Brady, M., Goodman, M., & Hansen, T. (2016). *Marketing management*. Pearson Education Ltd.

Lee, C. M., Ma, P., & Wang, C. C. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, *116*(2), 410–431.

Lie, E. & Lie, H. J. (2002). Multiples used to estimate corporate value. *Financial Analysts Journal*, *58*(2), 44–54.

Liu, J., Nissim, D., & Thomas, J. (2002). Equity valuation using multiples. *Journal of Accounting Research*, *40*(1), 135–172.

Liu, W. (2006). A liquidity-augmented capital asset pricing model. *Journal of financial Economics*, *82*(3), 631–671.

Malkiel, B. G. & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, *25*(2), 383–417.

Nel, S., Bruwer, W., & le Roux, N. (2014). An emerging market perspective on peer group selection based on valuation fundamentals. *Applied Financial Economics*, *24*(9), 621–637.

Nel, W. & Le Roux, N. (2017). An analyst's guide to sector-specific optimal peer group variables and multiples in the south african market. *Economics, Management and Financial Markets*, *12*(1), 25.

Newbold, P., Carlson, W. L., & Thorne, B. (2013). *Statistics for business and economics*. Pearson Boston, MA.

Nightingale, J. (1978). On the definition of industry and market. *The Journal of Industrial Economics*, 31–40.

Peltzman, S. (1977). The gains and losses from industrial concentration. *The Journal of Law and Economics*, *20*(2), 229–263.

Petersen, C. V., Plenborg, T., & Kinserdal, F. (2017). *Financial statement analysis: Valuation-credit analysis-performance evaluation*. Fagbokforlaget.

Pinto, J. E., Henry, E., Robinson, T. R., & Stowe, J. D. (2015). *Equity asset valuation*. John Wiley & Sons.

Plenborg, T. & Pimentel, R. C. (2016). Best practices in applying multiples for valuation purposes. *The Journal of Private Equity*, *19*(3), 55.

Porter, M. E. (2008). The five competitive forces that shape strategy. *Harvard business review*, *86*(1), 25–40.

pwc. (2018). Ifrs and us gaap: Similarities and differences. https://www.pwc.com/us/en/cfodirect/assets/pdf/accounting-guides/pwc-ifrs-us-gaap-similarities-and-differences.pdf. Accessed: 2019-08-24.

Schreiner, A. (2009). *Equity valuation using multiples: An empirical investigation*. Springer Science & Business Media.

Schreiner, A. & Spremann, K. (2007). Multiples and their valuation accuracy in european equity markets. *Available at SSRN 957352*.

Standard & Poor's. (2011). *Standard & poor's compustat xpressfeed using the data*. The McGraw-Hill Companies Inc.

Standard & Poor's. (2018). *Global industry classification standard - a guide to the gics methodology*. S&P Global Market Intelligence.

Standard & Poor's. (2019). S&p u.s. indices methodology. https://us.spindices.com/documents/methodologies/methodology-sp-us-indices.pdf. Accessed: 2019-08-12.

Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance*, *61*(1), 105–137.

Appendix A

Tables and figures

Table A.1: Replication of original SARD estimation errors

The table encloses the estimation errors incrured from my model on the original data provided by the authors. Overall, estimation errors articulate fully with the values presented in Panel A of Table 5 in Knudsen, Kold, and Plenborg (2017). Industry benchmark and industry-based SARD are not included, since these purposely differ from the original article. Note that rounding errors can occur.

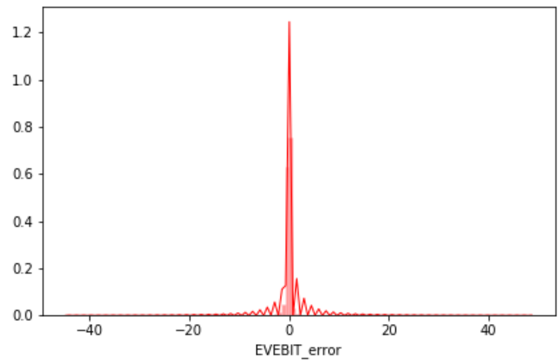
	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
EV/EBIT					
Median	0.292 (5)	0.260 (4)	0.250 (3)	0.228 (2)	0.222 (1)
Mean	0.390 (5)	0.351 (4)	0.343 (3)	0.309 (2)	0.307 (1)
IQ range	0.364 (5)	0.335 (4)	0.330 (3)	0.297 (2)	0.291 (1)
EV/SALES					
Median	0.531 (5)	0.502 (4)	0.478 (3)	0.470 (2)	0.254 (1)
Mean	0.761 (5)	0.719 (4)	0.693 (3)	0.670 (2)	0.360 (1)
IQ range	0.499 (5)	0.485 (4)	0.479 (2)	0.480 (3)	0.332 (1)
P/B					
Median	0.298 (5)	0.283 (4)	0.274 (3)	0.241 (2)	0.240 (1)
Mean	0.377 (5)	0.360 (4)	0.348 (3)	0.316 (2)	0.313 (1)
IQ range	0.352 (5)	0.347 (4)	0.335 (3)	0.312 (1)	0.313 (2)
P/E					
Median	0.297 (5)	0.287 (4)	0.279 (3)	0.244 (2)	0.240 (1)
Mean	0.375 (5)	0.364 (4)	0.355 (3)	0.325 (2)	0.325 (1)
IQ range	0.352 (5)	0.346 (4)	0.340 (3)	0.321 (2)	0.321 (1)

Source: Data from auhtors of original SARD paper and the developed SARD algortihm.

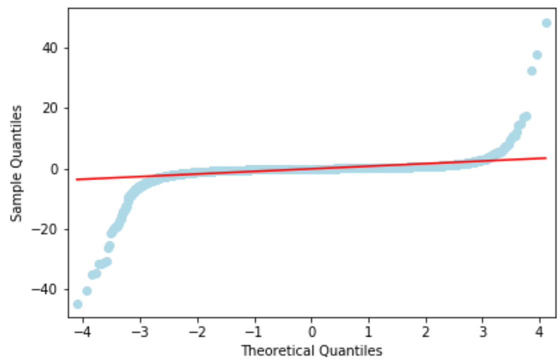
Figure A.1: Statistical of distribution for pairwise differences between two models

A displays the density distribution of the paired differences between a SARD-model based on ROE, Net Debt/EBIT, Size, Implied growth and EBIT-margin and a SARD-model based on ROE only for the EV/EBIT multiple. Errors are based on the all firm sample. The distribution displayed is generally representative of all distributions analyzed. **B** reports QQ plots for the sample- and theoretical quantiles originating from the pairwise differences between a SARD-model based on ROE, Net Debt/EBIT, Size, Implied growth and EBIT-margin and a SARD-model based on ROE only. If the sample quantiles are placed on the theoretical straight line normally assumption should. However, distribution heavy tails seem to exist. The outlook generally holds for all pairwise differences across every selection model examined.

(a) Density plot of distribution



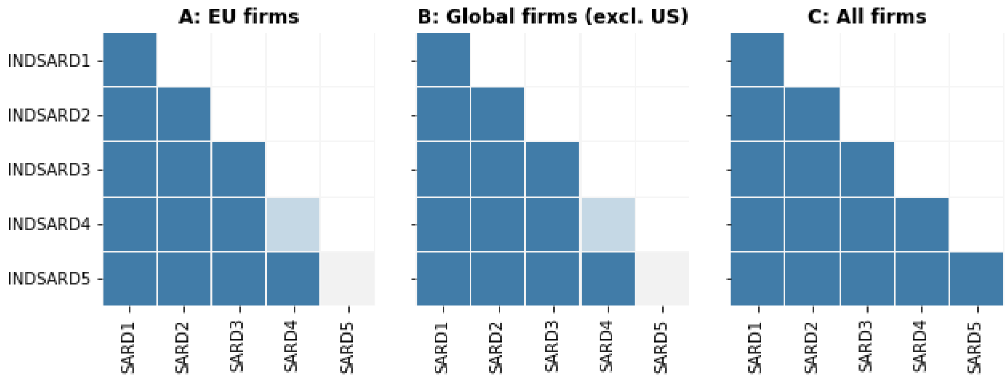
(b) Quantile-quantile probability plot



Source: Own calculations based on errors incurred from the SARD algorithm

Figure A.2: *t*-tests, SARD versus Industry SARD-models

Results of the two-tailed *t*-tests are displayed below the diagonals. Blue fields indicate that the selection model in the row obtain a lower valuation error than the selection model in the column, whereas red fields indicate the opposite. Solid color fields indicate significance at the 1%, whereas light color fields indicate significance at 5%. The grey fields indicate an insignificant difference between a set of models.

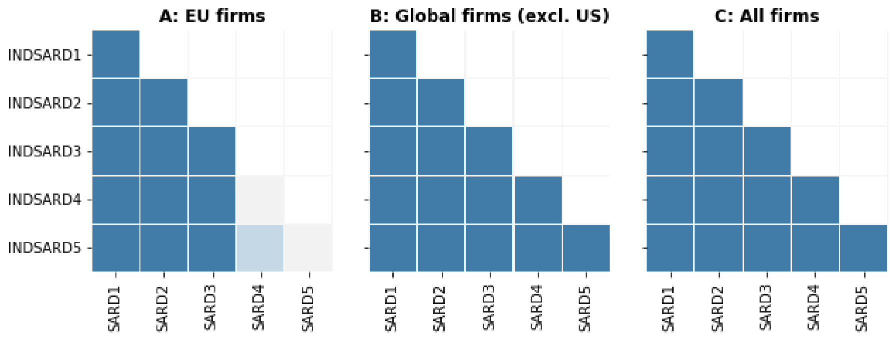


Source: Own calculations based on errors incurred from the Industry SARD algorithm

Figure A.3: Statistical-tests, SARD versus Industry SARD-models, P/E multiple

Results of the two-tailed Wilcoxon signed rank tests and *t*-tests are displayed in **A** and **B** respectively. Blue fields indicate that the selection model in the row obtain a lower valuation error than the selection model in the column, whereas red fields indicate the opposite. Solid color fields indicate significance at the 1%, whereas light color fields indicate significance at 5%. The grey fields indicate an insignificant difference between a set of models. **Source**: Own calculations based on errors incurred from the SARD algorithm.

(a) Results from the Wilcoxon tests



(b) Results from the *t*-tests

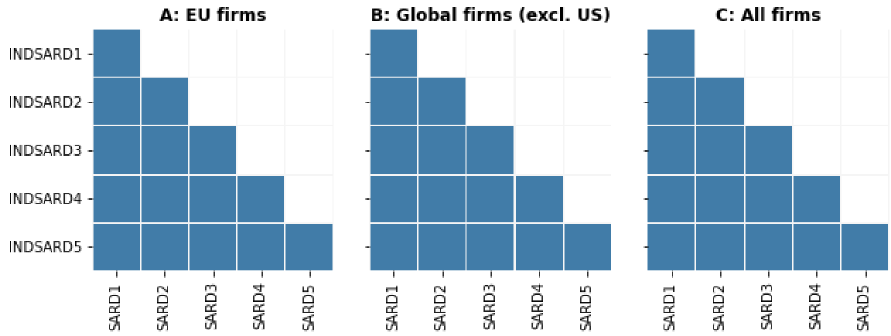


Table A.2: Industry level selection for three sub samples

The table encloses the proportion of firms able to be selected at certain GICS levels. Note that the two sub samples vary in size. All firms being the largest and EU-firms the smallest.

	Firms able to identify required amount of peers within:			
Required number of peers	**Sub-industry** GICS8	**Industry** GICS6	**Industry group** GICS4	**Sector** GICS2
Panel A: Percentage of identification, All firms				
6	93%	99%	100%	100%
7	91%	98%	100%	100%
8	89%	98%	100%	100%
9	87%	97%	100%	100%
10	85%	97%	100%	100%
11	82%	96%	99%	100%
12	79%	95%	99%	100%
13	77%	94%	99%	100%
14	74%	94%	99%	100%
15	71%	93%	99%	100%
16	69%	92%	99%	100%
Panel B: Percentage of identification, EU firms				
6	66%	89%	100%	100%
7	60%	87%	100%	100%
8	55%	85%	99%	100%
9	49%	82%	99%	100%
10	58%	86%	99%	100%
11	41%	77%	97%	100%
12	37%	74%	97%	100%
13	33%	69%	96%	100%
14	28%	66%	96%	100%
15	25%	62%	95%	100%
16	23%	59%	95%	100%

Source: Own calculations based on errors incurred from the level-up industry algorithm. Format inspired by Cheng and McNamara (2000)

Figure A.4: Statistical-tests, SARD-models versus Industry benchmark, Size indices

Results of the two-tailed Wilcoxon signed rank tests and t-tests are displayed above and below the diagonals respectively. Blue fields indicate that the selection model in the row obtain a lower valuation error than the selection model in the column, whereas red fields indicate the opposite. Solid color fields indicate significance at the 1%, whereas light color fields indicate significance at 5%. The grey fields indicate an insignificant difference between a set of models. **Source**: Own calculations based on errors incurred from the SARD algorithm.

(a) Large cap **(b)** Mid cap **(c)** Small cap **(d)** Rest index

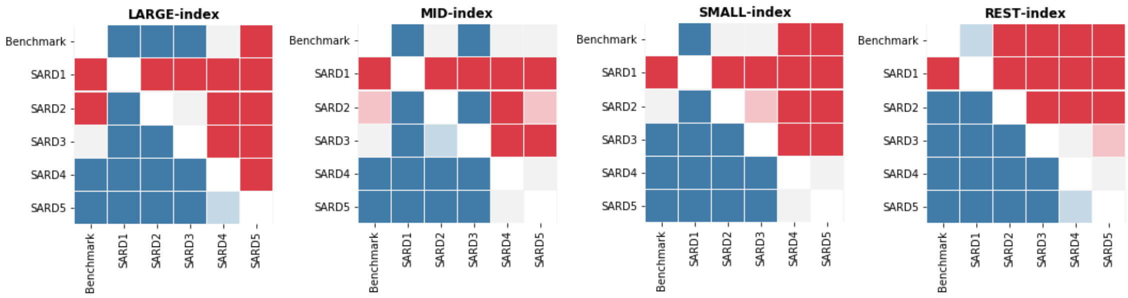


Table A.3: Number of observations by firm-year, GICS2 restriction

The table displays the annual number of observations in the total sample after correcting for GICS sector peers pr year. In total, the sample includes 27 OECD member states, including 17 European Union member countries, the USA, Japan, South Korea, Mexico, Australia, Turkey, Norway, New Zealand, Canada and Switzerland.

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Australia	18	81	88	109	138	148	197	221	230	192	197	211	214	229	249	240	230	238	241	248
Austria	10	13	9	6	10	9	13	18	25	23	17	22	24	21	21	17	22	19	15	21
Belgium	18	24	20	24	34	36	34	41	52	39	33	49	39	22	35	38	42	41	20	27
Canada	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1
Switzerland	57	78	58	65	69	69	85	86	93	93	80	89	87	85	88	96	90	95	83	96
Czech Republic	1	3	3	2	3	2	2	4	4	4	3	4	3	3	3	3	4	4	3	2
Germany	90	135	111	84	90	111	131	163	179	157	144	185	170	178	172	171	163	179	136	178
Denmark	11	29	22	29	28	30	34	30	39	31	32	37	35	33	28	30	34	34	34	33
Spain	20	39	52	52	32	47	32	31	55	36	40	47	52	46	44	49	52	57	51	51
Finland	16	37	49	41	38	49	68	72	77	60	45	67	58	54	56	62	68	68	67	71
France	25	140	156	129	137	157	163	188	208	177	171	188	167	185	176	163	161	177	103	169
Great Britain	132	96	107	152	166	222	216	273	299	246	269	324	335	312	336	339	333	344	245	327
Greece	5	19	23	24	22	33	31	32	27	17	26	16	13	15	8	14	14	17	8	11
Hungary	2	10	10	3	8	7	7	7	6	5	7	6	5	5	5	3	2	3	3	4
Italy	36	45	47	55	30	44	42	60	92	73	66	81	70	66	80	87	75	74	55	75
Japan	26	39	45	53	81	60	76	72	83	78	79	90	88	101	128	139	150	186	204	220
South Korea	43	117	137	32	29	14	22	34	37	31	26	38	108	111	131	139	183	275	222	207
Luxembourg	0	1	2	0	1	6	4	6	10	9	6	11	14	11	12	15	12	14	13	15
Mexico	7	12	6	12	8	12	13	11	21	18	32	31	30	37	35	37	39	39	33	41
Netherlands	19	35	38	43	45	64	58	64	61	48	42	49	40	42	43	42	39	41	38	42
Norway	18	30	28	27	27	24	45	59	56	47	52	45	45	52	53	46	40	49	43	47
New Zealand	2	13	17	18	29	35	36	40	38	34	34	33	38	35	42	45	44	41	34	41
Poland	5	13	8	8	10	14	18	29	39	37	38	71	59	46	61	74	72	43	24	43
Portugal	7	10	8	7	17	19	15	8	19	18	14	16	15	17	12	13	16	12	7	12
Sweden	29	57	51	47	53	52	75	85	90	94	85	105	89	115	107	107	112	139	126	157
Turkey	0	0	0	7	3	14	18	39	41	30	33	46	42	55	39	40	30	30	35	37
USA	813	804	711	764	824	902	918	941	960	839	874	998	1002	997	1003	1001	929	944	944	866

Source: Data from trailing financial statements, Compustat and I/B/E/S

Table A.4: Number of observations by firm-year, no GICS2 restriction

The table displays the annual number of observations in the total sample before correcting for GICS sector peers pr year. Restriction set for specific country analysis is that the country holds at least 10 home country peers in a given year of valuation in order to perform home-based SARD.

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Australia	18	82	89	109	139	148	197	221	230	192	197	211	214	229	249	240	230	238	241	248
Austria	10	13	9	6	10	9	13	18	25	23	17	22	24	21	21	17	22	19	15	21
Belgium	18	24	21	24	34	36	34	41	52	39	33	49	39	22	35	38	42	41	20	27
Canada	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1
Switzerland	57	78	58	65	69	69	85	86	93	93	80	89	87	85	88	96	90	95	83	96
Czech Republic	1	3	3	2	3	2	2	4	4	4	3	4	3	3	3	3	4	4	3	2
Germany	91	135	111	84	90	111	131	163	179	157	144	185	170	178	172	171	163	179	136	178
Denmark	11	29	22	29	28	30	34	30	39	31	32	37	35	33	28	30	34	34	34	33
Spain	20	39	52	52	32	47	32	31	55	36	40	47	52	46	44	49	52	57	51	51
Finland	16	37	49	41	38	49	68	72	77	60	45	67	58	54	56	62	68	68	67	71
France	25	140	156	129	137	157	163	188	208	177	171	188	167	185	176	163	161	177	103	169
Great Britain	132	97	107	152	166	222	216	273	299	246	269	324	335	312	336	339	333	344	245	327
Greece	5	19	23	24	22	33	31	32	27	17	26	16	13	15	8	14	14	17	8	11
Hungary	2	10	10	3	8	7	7	7	6	5	7	6	5	5	5	3	2	3	3	4
Italy	36	45	47	55	30	44	42	60	92	73	66	81	70	66	80	87	75	74	55	75
Japan	26	39	45	53	81	60	76	72	83	78	79	90	88	101	128	139	150	186	204	220
South Korea	43	117	137	32	29	14	22	34	37	31	26	38	108	111	131	139	183	275	222	207
Luxembourg	0	1	2	0	1	6	4	6	10	9	6	11	14	11	12	15	12	14	13	15
Mexico	7	12	6	12	8	12	13	11	21	18	32	31	30	37	35	37	39	39	33	41
Netherlands	19	35	38	43	45	64	58	64	61	48	42	49	40	42	43	42	39	41	38	42
Norway	18	30	28	27	27	24	45	59	56	47	52	45	45	52	53	46	40	49	43	47
New Zealand	2	13	17	18	29	35	36	40	38	34	34	33	38	35	42	45	44	41	34	41
Poland	5	13	8	8	10	14	18	29	39	37	38	71	59	46	61	74	72	43	24	43
Portugal	7	10	8	7	17	19	15	8	19	18	14	16	15	17	12	13	16	12	7	12
Sweden	29	57	51	47	53	52	75	85	90	94	85	105	89	115	107	107	112	139	126	157
Turkey	0	0	0	7	3	14	18	39	41	30	33	46	42	55	39	40	30	30	35	37
USA	816	806	714	772	839	902	918	941	960	839	874	998	1002	997	1003	1001	929	944	944	866

Source: Data from trailing financial statements, Compustat and I/B/E/S

Table A.5: Number of observations by firm-year, GICS2 restriction set at the country-level

The table displays the annual number of observations in the total sample after correcting for GICS sector peers pr year at the country level. Restriction set for specific country analysis is that the country holds at least 16 home country GICS2 peers in a given year of valuation in order to perform industry-based SARD at the home market.

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Great Britain	79	51	57	88	102	177	169	229	269	220	252	306	317	301	320	310	309	318	217	304
Australia		37	38	71	86	95	141	179	188	154	160	162	159	188	198	206	207	214	215	209
Japan					41	19	41	27	28	41	42	63	43	67	107	119	127	161	176	188
South Korea		90	87										71	59	100	110	161	262	210	201
Switzerland	21	30	20	24	27	22	37	34	32	37	27	33	35	32	36	40	35	37	33	37
France		96	122	103	95	120	129	162	176	134	145	147	117	126	121	107	123	125	56	123
Germany	66	102	77	41	42	78	95	140	150	125	116	174	139	148	143	143	130	144	97	153
Denmark									17											
Finland							43	41	28	23		22	21	19	19	18	23	23	23	22
Netherlands			21	20	21	25	25	27	25	20		18								
Sweden		28	20	22	22	24	35	52	39	56	36	82	36	87	85	77	96	124	73	128
Norway														18	19					
Spain														3	3					
Italy			9	13	5	8	8	10	20	20		18								
USA	813	804	711	764	824	902	918	941	960	839	874	998	1002	997	1003	1001	929	944	944	866

Source: Data from trailing financial statements, Compustat and I/B/E/S

Table A.6: Statistical test sizes, SARD selection models and benchmark

Statistical tests for difference in absolute percentage valuation errors across SARD-models. Test results of the two-tailed Wilcoxon signed rank test are presented below the white diagonal and the two-tailed *t*-test are placed above the white diagonal.

	Benchmark	**SARD1**	**SARD2**	**SARD3**	**SARD4**	**SARD5**
	Industry	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
Panel A: EU firms						
Benchmark		5.248	-4.292	-6.101	-8.753	-10.448
SARD1	-6.790		-12.661	-12.589	-13.533	-15.977
SARD2	7.799	17.073		-2.545	-5.204	-7.783
SARD3	10.938	20.469	4.268		-4.150	-6.286
SARD4	16.670	25.244	10.106	6.979		-1.938
SARD5	19.993	28.687	13.811	10.396	4.758	
Panel B: Global firms (excl. US)						
Benchmark		4.104	-5.533	-10.871	-11.274	-13.761
SARD1	-7.964		-10.859	-16.807	-16.814	-19.249
SARD2	9.908	19.921		-4.770	-6.189	-7.989
SARD3	16.554	26.826	7.933		-3.245	-5.384
SARD4	23.390	33.627	15.498	9.844		-2.783
SARD5	25.736	35.726	17.450	11.614	3.028	
Panel C: All firms						
Benchmark		12.569	-0.514	-6.052	-11.472	-12.351
SARD1	-20.309		-12.905	-18.884	-25.990	-26.221
SARD2	2.234	25.747		-6.296	-10.620	-11.263
SARD3	10.365	34.239	10.176		-6.336	-7.267
SARD4	22.646	46.535	23.413	15.776		-1.207
SARD5	24.502	48.722	24.805	17.171	2.552	

Source: Data from trailing financial statements, Compustat and I/B/E/S

Table A.7: Statistical test sizes, INDSARD selection models and benchmark

Statistical tests for difference in absolute percentage valuation errors across INDSARD-models. Test results of the two-tailed Wilcoxon signed rank test are presented below the white diagonal and the two-tailed *t*-test are placed above the white diagonal.

	Benchmark	**INDSARD1**	**INDSARD2**	**INDSARD3**	**INDSARD4**	**INDSARD5**
	Industry	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
Panel A: EU firms						
Benchmark		-5.952	-16.958	-18.206	-19.237	-20.449
INDSARD1	9.426		-13.146	-15.052	-16.830	-18.363
INDSARD2	23.243	18.237		-2.752	-3.411	-4.385
INDSARD3	26.393	20.599	4.177		-0.925	-2.120
INDSARD4	28.677	22.276	6.539	3.293		-1.886
INDSARD5	30.249	23.355	8.508	5.078	2.938	
Panel B: Global firms (excl. US)						
Benchmark		-6.745	-18.459	-22.291	-23.587	-24.764
INDSARD1	10.743		-17.345	-18.478	-19.923	-21.092
INDSARD2	26.142	20.314		-6.862	-8.239	-9.484
INDSARD3	31.416	24.990	6.759		-3.729	-5.193
INDSARD4	35.972	29.422	12.498	7.491		-3.122
INDSARD5	37.448	30.725	14.511	8.512	3.750	
Panel C: All firms						
Benchmark		-7.403	-18.648	-25.583	-29.349	-30.291
INDSARD1	14.934		-14.014	-21.071	-27.346	-28.421
INDSARD2	34.207	25.159		-8.805	-9.641	-10.626
INDSARD3	41.896	32.003	10.228		-5.425	-6.945
INDSARD4	49.654	40.731	19.713	13.198		-4.270
INDSARD5	51.822	42.410	22.141	15.160	4.668	

Source: Data from trailing financial statements, Compustat and I/B/E/S

Table A.8: Statistical tests for alternative selection pools trough the SARD approach

The table displays Wilcoxon- and *t*-tests test sizes as well as p-values in parentheses below. Tests are performed for the alternative peer pools recommended trough the initial analysis with All-firm selection as the benchmark. Selection happens through SARD5 for the EV/EBIT multiple.

Country	**HomevsGlobal**		**GlobalvsRegion**		**HomevsRegion**	
	Wilcoxon	t-test	Wilcoxon	t-test	Wilcoxon	t-test
Panel A: EU member countries						
France	-5.206	-4.419	2.447	0.070	-3.637	-5.480
(3,140)	(0.000)	(0.000)	(0.014)	(0.944)	(0.000	(0.000)
Germany	1.268	-0.905	-0.332	-1.315	1.744	-2.184
(2,928)	(0.205)	(0.366)	(0.740)	(0.189)	(0.081)	(0.029)
Great Britain	-1.060	1.365	1.571	-0.747	0.079	0.980
(5,074)	(0.289)	(0.172)	(0.116)	(0.455)	(0.937)	(0.327)
Netherlands	-0.226	-0.668	1.986	2.847	1.184	0.810
(893)	(0.821)	(0.504)	(0.047)	(0.005)	(0.236)	(0.418)
Italy	-1.339	-2.448	2.598	0.531	0.689	-1.801
(1,253)	(0.181)	(0.015)	(0.009)	(0.596)	(0.491)	(0.072)
Spain	0.857	1.843	1.046	1.325	1.758	2.943
(885)	(0.391)	(0.066)	(0.295)	(0.186)	(0.079)	(0.003)
Denmark	1.852	2.819	1.014	0.820	2.739	3.119
(613)	(0.064)	(0.005)	(0.310)	(0.413)	(0.006)	(0.002)
Finland	-2.553	0.894	1.392	0.684	-2.269	1.325
(1,123)	(0.011)	(0.371)	(0.164)	(0.494)	(0.023)	(0.186)
Sweden	-3.079	0.541	0.955	0.093	-2.831	0.583
(1,775)	(0.002)	(0.588)	(0.339)	(0.926)	(0.005)	(0.560)
Belgium	2.690	1.718	-0.429	-0.209	2.829	1.306
(669)	(0.007)	(0.086)	(0.668)	(0.835)	(0.005)	(0.192)
Australia	-3.026	0.679	-0.477	0.071	-4.592	0.842
(3,722)	(0.002)	(0.497)	(0.634)	(0.943)	(0.000)	(0.400)
South Korea	-4.569	-5.580	0.767	3.015	-3.487	-4.404
(1,996)	(0.000)	(0.000)	(0.443)	(0.003)	(0.000)	(0.000)
Japan	-2.057	-3.644	0.305	1.868	-2.393	-3.267
(1,998)	(0.040)	(0.000)	(0.761)	(0.062)	(0.017)	(0.001)
Norway	3.269	2.447				
(833)	(0.001)	(0.015)				
Switzerland	-3.296	0.447				
(1,642)	(0.001)	(0.655)				

Source: Own calculations based on errors incurred from the SARD5 algorithm

Table A.9: Statistical tests for alternative selection pools trough the INDSARD approach

The table displays Wilcoxon- and *t*-tests test sizes as well as p-values in parentheses below. Tests are performed for the alternative peer pools recommended trough the initial analysis with All-firm selection as the benchmark. Selection happens through SARD5 for the EV/EBIT multiple.

Country	**GlobalvsRegion**		**HomevsRegion**		**HomevsGlobal**	
	Wilcoxon	t-test	Wilcoxon	t-test	Wilcoxon	t-test
France	-0.066	-0.902	-1.723	-5.699	-2.127	-5.412
(2,327)	(0.947)	(0.367)	(0.085)	(0.000)	(0.033)	(0.000)
Germany	0.162	0.073	1.227	-1.373	1.245	-1.105
(2,303)	(0.871)	(0.942)	(0.220)	(0.170)	(0.213)	(0.269)
Great Britain	1.456	1.326	0.410	1.933	-0.436	0.620
(4,395)	(0.145)	(0.185)	(0.682)	(0.053)	(0.663)	(0.536)
Sweden	-1.015	-1.402	-3.104	-0.084	-2.195	0.666
(1,122)	(0.310)	(0.161)	(0.002)	(0.933)	(0.028)	(0.506)
Finland	-0.990	-0.163	-0.767	1.536	-0.462	1.656
(325)	(0.322)	(0.087)	(0.443)	(0.125)	(0.644)	(0.099)
Netherlands	0.946	1.760				
(891)	(0.344)	(0.079)				
Italy	0.109	0.901				
(1,224)	(0.913)	(0.368)				
Spain	-0.217	0.600				
(871)	(0.828)	(0.549)				
Denmark	-0.623	-0.266				
(609)	(0.533)	(0.790)				
Belgium	1.977	0.913				
(655)	(0.048)	(0.362)				
Australia	0.379	-0.868	-2.985	-0.477	-2.800	0.588
(2,907)	(0.705)	(0.385)	(0.003)	(0.634)	(0.005)	(0.556)
South Korea	0.462	2.766	-0.845	-0.852	-1.314	-2.928
(1,351)	(0.644)	(0.006)	(0.398)	(0.394)	(0.189)	(0.003)
Japan	-2.543	-0.540	-0.190	-2.393	1.000	-1.913
(1,290)	(0.011)	(0.589)	(0.849)	(0.017)	(0.317)	(0.056)
Switzerland					2.143	3.339
(629)					(0.032)	(0.001)

Source: Own calculations based on errors incurred from the SARD5 algorithm

Table A.10: Performance of Five SARD- and INDSARD Combinations, P/E multiple

Median, mean interquartile range and proportion less 15% of valuation errors incurred for each of the industry-based- and regular SARD-models. The parentheses indicate the ranking based on the estimation error across the models applied for the P/E multiple. Results are presented for all three sub samples analyzed.

Panel A: Regular SARD-models						
	Industry	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
	Benchmark	**SARD1**	**SARD2**	**SARD3**	**SARD4**	**SARD5**
Panel A1: All firms						
Median	0.365 (6)	0.352 (5)	0.345 (4)	0.335 (3)	0.317 (2)	0.313 (1)
Mean	0.509 (6)	0.487 (5)	0.476 (4)	0.451 (3)	0.436 (2)	0.430 (1)
IQ range	0.458 (6)	0.419 (5)	0.403 (4)	0.394 (3)	0.386 (1)	0.391 (2)
Prop<15%	0.220 (6)	0.229 (4)	0.224 (5)	0.234 (3)	0.249 (2)	0.255 (1)
Panel A2: Global firms (excl. US)						
Median	0.378 (6)	0.355 (5)	0.353 (4)	0.348 (3)	0.335 (2)	0.332 (1)
Mean	0.537 (6)	0.499 (5)	0.490 (4)	0.471 (3)	0.465 (2)	0.461 (1)
IQ range	0.476 (6)	0.422 (5)	0.406 (3)	0.411 (4)	0.403 (1)	0.406 (2)
Prop<15%	0.208 (6)	0.223 (4)	0.217 (5)	0.227 (3)	0.235 (2)	0.237 (1)
Panel A3: EU firms						
Median	0.356 (6)	0.334 (4)	0.335 (5)	0.332 (3)	0.320 (2)	0.318 (1)
Mean	0.507 (6)	0.466 (5)	0.463 (4)	0.452 (3)	0.443 (2)	0.443 (1)
IQ range	0.463 (6)	0.404 (5)	0.396 (4)	0.396 (3)	0.392 (1)	0.394 (2)
Prop<15%	0.227 (6)	0.240 (3)	0.230 (5)	0.237 (4)	0.244 (2)	0.247 (1)
Panel B: Industry-based SARD-models						
		INDSARD1	**INDSARD2**	**INDSARD3**	**INDSARD4**	**INDSARD5**
Panel B1: All firms						
Median		0.318 (4)	0.319 (5)	0.307 (3)	0.298 (2)	0.297 (1)
Mean		0.439 (5)	0.437 (4)	0.419 (3)	0.415 (1)	0.416 (2)
IQ range		0.391 (3)	0.401 (5)	0.394 (4)	0.391 (2)	0.391 (1)
Prop<15%		0.245 (5)	0.251 (4)	0.264 (3)	0.272 (2)	0.270 (1)
Panel B2: Global firms (excl. US)						
Median		0.336 (4)	0.339 (5)	0.327 (3)	0.320 (2)	0.319 (1)
Mean		0.463 (4)	0.466 (5)	0.448 (3)	0.446 (2)	0.446 (1)
IQ range		0.408 (1)	0.419 (5)	0.411 (3)	0.411 (2)	0.413 (4)
Prop<15%		0.229 (5)	0.229 (4)	0.242 (3)	0.246 (2)	0.248 (1)
Panel B3: EU firms						
Median		0.322 (4)	0.323 (5)	0.314 (3)	0.310 (2)	0.307 (1)
Mean		0.442 (4)	0.447 (5)	0.433 (3)	0.432 (2)	0.432 (1)
IQ range		0.396 (1)	0.408 (5)	0.400 (2)	0.401 (3)	0.405 (4)
Prop<15%		0.249 (4)	0.247 (5)	0.256 (3)	0.261 (2)	0.265 (1)

Source: Own calculations on data from Compustat and I/B/E/S

Table A.11: Statistical test sizes, SARD selection models and benchmark, P/E multiple

Statistical tests for difference in absolute percentage valuation errors across SARD-models. Test results of the two-tailed Wilcoxon signed rank test are presented below the white diagonal and the two-tailed *t*-test are placed above the white diagonal.

	Industry	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
	Benchmark	**SARD1**	**SARD2**	**SARD3**	**SARD4**	**SARD5**
Panel A: All firms						
Benchmark		-9.266	-9.840	-11.425	-13.961	-13.930
SARD1	9.076		-0.931	-4.066	-6.248	-5.995
SARD2	10.932	-0.340		-3.821	-6.269	-5.981
SARD3	14.316	3.531	2.384		-3.472	-3.061
SARD4	18.084	7.697	7.018	5.003		-0.049
SARD5	19.097	8.472	7.714	5.210	0.359	
Panel B: Global firms (excl. USA)						
Benchmark		-9.680	-13.293	-17.973	-18.220	-20.828
SARD1	11.975		-2.647	-8.144	-10.693	-10.761
SARD2	14.417	1.048		-7.004	-6.987	-9.752
SARD3	20.026	7.737	6.066		-2.323	-4.284
SARD4	25.867	13.705	12.555	7.321		-1.548
SARD5	26.398	13.981	12.910	6.767	-0.374	
Panel C: EU firms						
Benchmark		-7.176	-11.578	-19.762	-24.191	-28.255
SARD1	8.666		-4.378	-14.596	-20.977	-22.170
SARD2	13.505	4.713		-12.338	-15.731	-20.458
SARD3	23.726	15.262	11.158		-7.697	-11.368
SARD4	33.612	26.190	22.386	13.871		-3.549
SARD5	36.091	28.022	23.657	14.735	1.938	

Source: Own calculations based on errors incurred from the SARD algorithm

Table A.12: Valuation accuracy across countries and peer-pool sizes, EU firms, P/E multiple

Median, mean, IQ range and proportion less 15% for absolute estimation errors at the P/E multiple of each EU member country with sufficient amount of country peers to perform SARD5 selection (i.e peer group ≥ 10). Number of firm-year observations is presented in the left parentheses. Test sizes and p-values of statistical tests between difference in central tendency of absolute estimation errors are displayed in the right panel. P-values are presented in parentheses. The number of country-specific firm-year observations for the analysis is given in the parentheses below the country names.

Country		**Peer pool selection**				**Statistical tests**					
		Region	Global ex.USA	Home	All	HomevsAll		GlobalvsAll		RegionvsAll	
						Wilcoxon	t-test	Wilcoxon	t-test	Wilcoxon	t-test
France	Mean	0.404	0.400	0.397	0.403	**-2.4**	-0.9	-0.3	-0.8	0.7	0.2
(3,140)	Median	0.307	0.303	0.289	0.304	(0.018)	(0.376)	(0.767)	(0.434)	(0.476)	(0.808)
	IQ range	0.376	0.357	0.379	0.372						
	Prop<15%	0.264	0.254	0.284	0.259						
Germany	Mean	0.491	0.473	0.489	0.474	**3.5**	1.9	0.2	-0.3	1.0	1.8
(2,928)	Median	0.324	0.317	0.338	0.322	(0.001)	(0.052)	(0.850)	(0.796)	(0.314)	(0.074)
	IQ range	0.410	0.416	0.412	0.401						
	Prop<15%	0.244	0.252	0.227	0.243						
Great Britain	Mean	0.452	0.442	0.459	0.437	1.6	**3.8**	2.1	1.5	1.6	**3.2**
(5,074)	Median	0.335	0.338	0.336	0.337	(0.103)	(0.000)	(0.039)	(0.139)	(0.103)	(0.001)
	IQ range	0.403	0.400	0.412	0.392						
	Prop<15%	0.228	0.232	0.239	0.236						
Netherlands	Mean	0.382	0.392	0.413	0.377	**3.0**	**2.2**	1.5	1.7	0.1	0.5
(893)	Median	0.288	0.289	0.302	0.274	(0.002)	(0.031)	(0.135)	(0.083)	(0.935)	(0.622)
	IQ range	0.355	0.384	0.401	0.354						
	Prop<15%	0.271	0.288	0.279	0.288						
Italy	Mean	0.410	0.411	0.427	0.412	1.2	1.3	0.4	-0.2	0.3	-0.3
(1,253)	Median	0.308	0.315	0.329	0.308	(0.219)	(0.199)	(0.658)	(0.857)	(0.768)	(0.753)
	IQ range	0.376	0.378	0.413	0.407						
	Prop<15%	0.247	0.247	0.252	0.263						
Spain	Mean	0.372	0.377	0.404	0.386	1.9	1.6	-0.1	-1.3	-1.1	-1.7
(885)	Median	0.285	0.302	0.321	0.316	(0.058)	(0.117)	(0.944)	(0.211)	(0.266)	(0.091)
	IQ range	0.364	0.388	0.403	0.382						
	Prop<15%	0.258	0.279	0.243	0.266						
Denmark	Mean	0.456	0.454	0.508	0.450	**3.0**	**2.9**	0.2	0.5	0.3	0.5
(613)	Median	0.322	0.314	0.350	0.322	(0.003)	(0.004)	(0.876)	(0.628)	(0.780)	(0.621)
	IQ range	0.423	0.414	0.416	0.409						
	Prop<15%	0.233	0.253	0.207	0.258						
Finland	Mean	0.406	0.397	0.436	0.405	1.5	**2.5**	-1.1	-1.4	-0.4	0.2
(1,123)	Median	0.317	0.314	0.306	0.318	(0.141)	(0.012)	(0.289)	(0.164)	(0.701)	(0.879)
	IQ range	0.347	0.350	0.381	0.355						
	Prop<15%	0.240	0.237	0.242	0.237						
Sweden	Mean	0.395	0.382	0.416	0.386	0.7	**2.9**	0.3	-1.0	0.0	1.2
(1,775)	Median	0.298	0.303	0.291	0.299	(0.491)	(0.004)	(0.774)	(0.313)	(0.969)	(0.245)
	IQ range	0.367	0.360	0.382	0.364						
	Prop<15%	0.274	0.269	0.268	0.269						
Belgium	Mean	0.444	0.439	0.450	0.420	2.3	1.8	0.4	1.1	0.7	1.4
(669)	Median	0.305	0.289	0.311	0.291	(0.023)	(0.079)	(0.662)	(0.277)	(0.505)	(0.175)
	IQ range	0.363	0.362	0.367	0.376						
	Prop<15%	0.265	0.254	0.257	0.281						

Source: Own calculations based on errors incurred from the SARD5 algorithm

Table A.13: Valuation accuracy across countries and peer-pool sizes, non-EU firms, P/E multiple

Median, mean, IQ range and proportion less 15% for absolute estimation errors on the P/E multiple of each non-EU country with sufficient amount of country peers to perform SARD5 selection (i.e peer group $\geq$ 10). Should be analyzed as pointed out in **Table 5.5**.

Country		**Peer pool selection**				**Statistical tests**					
		Home	Region	Global ex.USA	All	HomevsAll Wilcoxon	HomevsAll t-test	GlobalvsAll Wilcoxon	GlobalvsAll t-test	RegionvsAll Wilcoxon	RegionvsAll t-test
Australia (3,722)	Mean	0.487	0.487	0.480	0.480	-0.2 (0.828)	0.9 (0.349)	0.6 (0.529)	0.0 (0.966)	**2.0** (0.050)	1.0 (0.305)
	Median	0.337	0.357	0.350	0.341						
	IQ range	0.408	0.410	0.403	0.412						
	Prop<15%	0.226	0.208	0.218	0.228						
South Korea (1,996)	Mean	0.494	0.610	0.685	0.691	**-8.5** (0.000)	**-10.5** (0.000)	-0.7 (0.499)	-0.7 (0.512)	**-3.5** (0.000)	**-5.8** (0.000)
	Median	0.391	0.430	0.449	0.469						
	IQ range	0.458	0.527	0.582	0.576						
	Prop<15%	0.206	0.184	0.176	0.170						
Japan (1,998)	Mean	0.427	0.411	0.418	0.428	**-2.3** (0.023)	-0.1 (0.950)	-1.3 (0.192)	**-2.1** (0.032)	-0.8 (0.400)	**-2.4** (0.016)
	Median	0.324	0.343	0.353	0.360						
	IQ range	0.420	0.408	0.409	0.411						
	Prop<15%	0.247	0.224	0.230	0.223						
Norway (833)	Mean	0.538	0.538	0.549	0.560	**2.0** (0.047)	-1.1 (0.266)	-0.4 (0.679)	-1.2 (0.218)		
	Median	0.404	0.404	0.384	0.388						
	IQ range	0.457	0.457	0.454	0.468						
	Prop<15%	0.182	0.182	0.226	0.221						
Switzerland (1,642)	Mean	0.393	0.393	0.387	0.378	1.2 (0.248)	1.5 (0.122)	2.3 (0.024)	1.2 (0.224)		
	Median	0.285	0.285	0.296	0.294						
	IQ range	0.352	0.352	0.340	0.344						
	Prop<15%	0.270	0.270	0 .252	0.275						
USA (18,065)	Mean	0.373	0.373		0.373	**-3.7** (0.000)	0.3 (0.797)				
	Median	0.276	0.276		0.282						
	IQ range	0.352	0.352		0.358						
	Prop<15%	0.290	0.290		0.279						

Source: Own calculations based on errors incurred from the SARD5 algorithm

Table A.14: INDSARD valuation accuracy across countries and peer-pool sizes, P/E multiple

The table displays estimation errors of industry-based SARD5 on all countries with sufficient GICS2 peers in a given valutaion year when valuing the P/E multiple (num. GICS2 peers ≥ 16). Mean, median, IQ range and proportion less 15% of the absolute estimation error are further displayed. Right hand parentheses presents the number of firm-year observations for a given country as well as the count of valuation years with sufficient restricted data. The left side displays Wilcoxon- and *t*-tests test sizes as well as p-values in parentheses below.

Country		Peer pool selection				Statistical tests					
		Region	Global ex.USA	Home	All	HomevsAll		GlobalvsAll		RegionvsAll	
						Wilcoxon	t-test	Wilcoxon	t-test	Wilcoxon	t-test
France	Mean	0.410	0.409	0.413	0.421	1.5	-0.7	0.8	**-2.7**	1.2	-1.8
(2,327)	Median	0.291	0.284	0.301	0.283	(0.144)	(0.472)	(0.421)	(0.008)	(0.217)	(0.080)
(19)	IQ range	0.370	0.377	0.393	0.384						
	Prop<15%	0.275	0.273	0.272	0.276						
Germany	Mean	0.480	0.482	0.497	0.487	**2.0**	1.0	0.7	-0.6	0.6	-0.9
(2,303)	Median	0.329	0.327	0.337	0.327	(0.047)	(0.295)	(0.476)	(0.521)	(0.528)	(0.367)
(20)	IQ range	0.431	0.417	0.442	0.427						
	Prop<15%	0.244	0.244	0.239	0.241						
Great Britain	Mean	0.422	0.418	0.439	0.420	**4.4**	**3.3**	1.7	-0.5	1.8	0.4
(4,395)	Median	0.310	0.310	0.319	0.311	(0.000)	(0.001)	(0.087)	(0.595)	(0.076)	(0.697)
(20)	IQ range	0.402	0.395	0.409	0.398						
	Prop<15%	0.266	0.261	0.251	0.267						
Sweden	Mean	0.367	0.346	0.375	0.353	**2.0**	**2.3**	0.8	-1.2	**3.5**	**2.3**
(1,122)	Median	0.272	0.248	0.266	0.248	(0.050)	(0.021)	(0.448)	(0.231)	(0.001)	(0.023)
(19)	IQ range	0.338	0.331	0.378	0.345						
	Prop<15%	0.296	0.316	0.316	0.318						
Finland	Mean	0.375	0.377	0.411	0.368	1.7	**2.2**	1.3	1.2	1.3	0.6
(325)	Median	0.297	0.272	0.296	0.290	(0.083)	(0.030)	(0.197)	(0.222)	(0.204)	(0.566)
(13)	IQ range	0.381	0.404	0.409	0.391						
	Prop<15%	0.274	0.277	0.255	0.286						
Australia	Mean	0.500	0.503	0.503	0.502	-0.5	0.1	**2.8**	0.3	-0.5	-0.3
(2,907)	Median	0.356	0.368	0.353	0.359	(0.646)	(0.908)	(0.005)	(0.772)	(0.639)	(0.793)
(19)	IQ range	0.428	0.429	0.436	0.416						
	Prop<15%	0.225	0.208	0.227	0.214						
South Korea	Mean	0.523	0.569	0.477	0.580	**-3.8**	**-6.2**	0.0	-1.6	**-2.4**	**-4.7**
(1,351)	Median	0.395	0.408	0.390	0.409	(0.000)	(0.000)	(0.962)	(0.109)	(0.018)	(0.000)
(10)	IQ range	0.477	0.506	0.435	0.511						
	Prop<15%	0.207	0.196	0.204	0.200						
Japan	Mean	0.416	0.412	0.452	0.410	1.7	**3.5**	**2.0**	0.3	0.9	0.7
(1,290)	Median	0.332	0.343	0.341	0.334	(0.095)	(0.001)	(0.050)	(0.735)	(0.379)	(0.490)
(16)	IQ range	0.414	0.427	0.449	0.418						
	Prop<15%	0.239	0.236	0.234	0.242						
Switzerland	Mean		0.335	0.356	0.336	1.9	1.6	0.9	-0.2		
(629)	Median		0.260	0.275	0.255	(0.060)	(0.105)	(0.358)	(0.861)		
(20)	IQ range		0.330	0.314	0.333						
	Prop<15%		0.307	0.281	0.304						
USA	Mean			0.357	0.351	-1.9	**3.0**				
(18,034)	Median			0.259	0.266	(0.055)	(0.003)				
(20)	IQ range			0.356	0.351						
	Prop<15%			0.316	0.304						

Source: Own calculations based on errors incurred from the INDSARD5 algorithm

Table A.15: INDSARD valuation accuracy across countries and peer-pool sizes, P/E, home-country selection not available

Country		**Peer pool selection**			**Statistical tests**			
		Region	Global	All	GlobalvsAll		RegionvsAll	
			ex.USA		Wilcoxon	t-test	Wilcoxon	t-test
Netherlands	Mean	0.400	0.403	0.403	1.4	-0.1	0.8	-0.4
(891)	Median	0.278	0.284	0.275	(0.173)	(0.941)	(0.442)	(0.712)
	IQ range	0.387	0.386	0.378				
	Prop<15%	0.291	0.283	0.281				
Italy	Mean	0.384	0.393	0.394	0.6	-0.2	-1.2	-1.4
(1,224)	Median	0.308	0.305	0.298	(0.524)	(0.875)	(0.239)	(0.149)
	IQ range	0.387	0.395	0.394				
	Prop<15%	0.275	0.258	0.266				
Spain	Mean	0.353	0.348	0.358	1.3	-1.6	1.0	-0.6
(871)	Median	0.278	0.270	0.269	(0.209)	(0.120)	(0.312)	(0.522)
	IQ range	0.378	0.353	0.366				
	Prop<15%	0.290	0.290	0.301				
Denmark	Mean	0.424	0.397	0.408	-0.4	-1.3	1.5	1.2
(609)	Median	0.294	0.295	0.292	(0.681)	(0.185)	(0.142)	(0.218)
	IQ range	0.405	0.332	0.357				
	Prop<15%	0.271	0.259	0.251				
Belgium	Mean	0.456	0.457	0.471	-0.6	-0.8	-0.6	-0.9
(655)	Median	0.293	0.290	0.299	(0.572)	(0.442)	(0.572)	(0.376)
	IQ range	0.417	0.406	0.429				
	Prop<15%	0.279	0.295	0.270				
Norway	Mean		0.488	0.501	0.1	-1.2		
(832)	Median		0.352	0.359	(0.934)	(0.242)		
	IQ range		0.463	0.465				
	Prop<15%		0.234	0.225				

Source: Own calculations based on errors incurred from the INDSARD5 algorithm

Table A.16: Median errors grouped by Size indices, SARD

	Benchmark	**SARD1**	**SARD2**	**SARD3**	**SARD4**	**SARD5**
	Industry	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
Panel A: All firms						
Large	0.274	0.333	0.296	0.276	0.246	0.243
Mid	0.284	0.334	0.293	0.284	0.265	0.263
Small	0.312	0.338	0.310	0.301	0.286	0.288
Rest	0.381	0.391	0.357	0.340	0.332	0.328
Panel B: Global firms (excl. US)						
Large	0.286	0.333	0.301	0.278	0.256	0.250
Mid	0.303	0.336	0.299	0.292	0.272	0.272
Small	0.334	0.348	0.316	0.310	0.297	0.294
Rest	0.387	0.387	0.356	0.338	0.334	0.326
Panel C: EU firms						
Large	0.274	0.318	0.279	0.266	0.234	0.232
Mid	0.282	0.308	0.274	0.267	0.260	0.251
Small	0.298	0.320	0.290	0.287	0.271	0.271
Rest	0.365	0.361	0.334	0.324	0.315	0.306

Source: Own calculations based on errors incurred from the SARD algorithm and level-up benchmark

Table A.17: Median errors grouped by Size indices, INDSARD

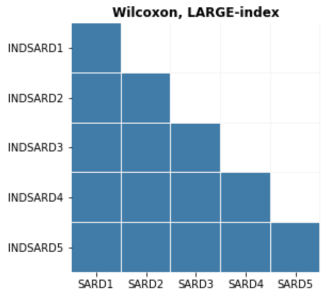
	Benchmark	SARD1	SARD2	SARD3	SARD4	SARD5
	Industry	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
Panel A: All firms						
Large	0.274	0.261	0.232	0.218	0.209	0.209
Mid	0.284	0.275	0.254	0.244	0.235	0.234
Small	0.312	0.294	0.278	0.279	0.267	0.263
Rest	0.381	0.373	0.348	0.340	0.336	0.335
Panel B: Global firms (excl. US)						
Large	0.286	0.277	0.252	0.228	0.226	0.233
Mid	0.303	0.291	0.267	0.261	0.254	0.253
Small	0.334	0.320	0.296	0.293	0.286	0.283
Rest	0.387	0.370	0.346	0.341	0.335	0.335
Panel C: EU firms						
Large	0.274	0.258	0.231	0.212	0.218	0.218
Mid	0.282	0.272	0.247	0.240	0.244	0.241
Small	0.298	0.291	0.273	0.266	0.257	0.257
Rest	0.365	0.347	0.323	0.321	0.323	0.316

Source: Own calculations based on errors incurred from the INDSARD algorithm and level-up benchmark

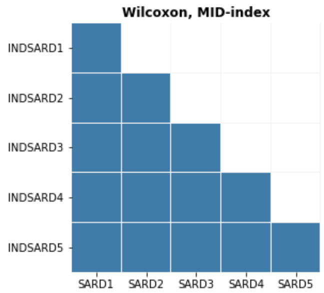
Figure A.5: Statistical-tests, SARD-models versus INDSARD-models, Size indices

Results of the two-tailed Wilcoxon signed rank tests and t-tests are displayed above and below the diagonals respectively. Blue fields indicate that the selection model in the row obtain a lower valuation error than the selection model in the column, whereas red fields indicate the opposite. Solid color fields indicate significance at the 1%, whereas light color fields indicate significance at 5%. The grey fields indicate an insignificant difference between a set of models. **Source:** Own calculations based on errors incurred from the SARD algorithm.

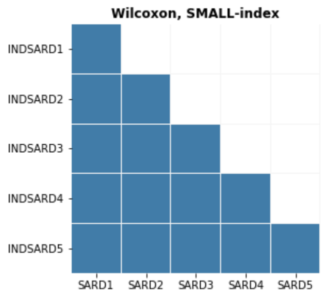
(a) Large cap, Wilcoxon



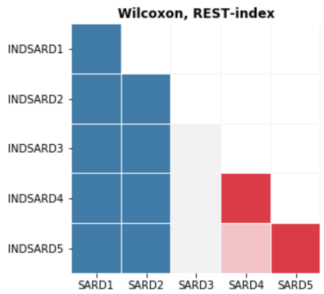
(b) Mid cap, Wilcoxon



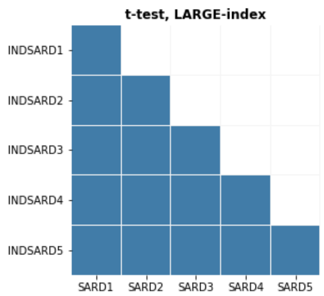
(c) Small cap, Wilcoxon



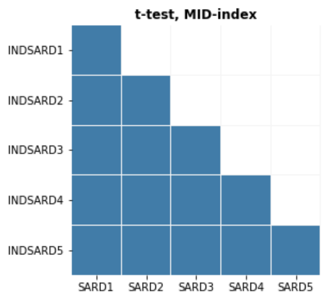
(d) Rest index, Wilcoxon



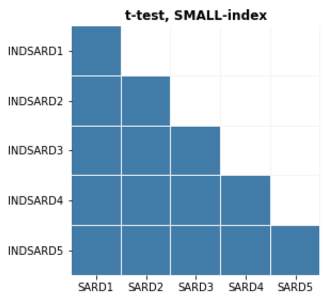
(e) Large cap, t-test



(f) Mid cap, t-test



(g) Small cap, t-test



(h) Rest index, t-test

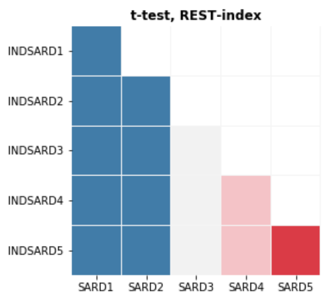


Table A.18: Median errors grouped by industries, SARD

	Benchmark	SARD1	SARD2	SARD3	SARD4	SARD5
	Industry	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
Panel A: All firms						
Energy	0.371	0.409	0.350	0.352	0.329	0.329
Materials	0.322	0.338	0.291	0.281	0.269	0.263
Industrials	0.290	0.311	0.281	0.263	0.251	0.254
Consumer Discretionary	0.320	0.342	0.322	0.300	0.281	0.282
Consumer Staples	0.277	0.308	0.275	0.247	0.238	0.244
Health Care	0.316	0.380	0.372	0.356	0.324	0.316
Financials	0.389	0.444	0.389	0.421	0.371	0.377
Information Technology	0.370	0.405	0.403	0.383	0.354	0.352
Telecommuncation Services	0.277	0.313	0.275	0.281	0.278	0.279
Utilities	0.229	0.272	0.191	0.197	0.191	0.180
Real Estate	0.257	0.521	0.284	0.295	0.331	0.289
Panel B: Global firms (excl. US firms)						
Energy	0.411	0.424	0.369	0.366	0.367	0.362
Materials	0.355	0.366	0.313	0.310	0.300	0.296
Industrials	0.314	0.328	0.293	0.281	0.269	0.269
Consumer Discretionary	0.328	0.355	0.321	0.305	0.295	0.288
Consumer Staples	0.307	0.327	0.285	0.265	0.245	0.256
Health Care	0.349	0.405	0.401	0.374	0.348	0.346
Financials	0.450	0.422	0.345	0.391	0.477	0.398
Information Technology	0.386	0.406	0.402	0.390	0.368	0.356
Telecommuncation Services	0.298	0.318	0.283	0.294	0.287	0.279
Utilities	0.321	0.326	0.253	0.258	0.254	0.242
Panel C: EU firms						
Energy	0.405	0.392	0.339	0.348	0.315	0.336
Materials	0.334	0.324	0.273	0.285	0.271	0.254
Industrials	0.293	0.306	0.270	0.257	0.252	0.249
Consumer Discretionary	0.309	0.335	0.308	0.297	0.283	0.273
Consumer Staples	0.261	0.288	0.251	0.233	0.217	0.225
Health Care	0.332	0.377	0.375	0.361	0.330	0.320
Information Technology	0.358	0.380	0.369	0.366	0.348	0.345
Telecommuncation Services	0.272	0.326	0.276	0.278	0.282	0.268
Utilities	0.285	0.315	0.236	0.237	0.224	0.226

Source: Own calculations based on errors incurred from the SARD algorithm and level-up benchmark

Table A.19: Median errors grouped by industries, INDSARD

	Benchmark	**INDSARD1**	**INDSARD2**	**INDSARD3**	**INDSARD4**	**INDSARD5**
	Industry	ROE	ROE DEBT/EBIT	ROE DEBT/EBIT Size	ROE DEBT/EBIT Size Growth	ROE DEBT/EBIT Size Growth EBIT margin
Panel A: All firms						
Energy	0.371	0.345	0.318	0.296	0.299	0.298
Materials	0.322	0.304	0.276	0.268	0.266	0.268
Industrials	0.290	0.279	0.258	0.251	0.242	0.244
Consumer Discretionary	0.320	0.307	0.291	0.285	0.267	0.263
Consumer Staples	0.277	0.269	0.250	0.237	0.230	0.229
Health Care	0.316	0.302	0.288	0.283	0.262	0.258
Financials	0.389	0.376	0.328	0.326	0.315	0.312
Information Technology	0.370	0.352	0.346	0.332	0.318	0.317
Telecommuncation Services	0.277	0.268	0.251	0.242	0.243	0.241
Utilities	0.229	0.208	0.164	0.168	0.170	0.168
Real Estate	0.257	0.236	0.204	0.222	0.213	0.210
Panel B: Global firms (excl. US firms)						
Energy	0.411	0.388	0.350	0.339	0.335	0.338
Materials	0.355	0.333	0.296	0.293	0.297	0.289
Industrials	0.314	0.306	0.280	0.275	0.271	0.269
Consumer Discretionary	0.328	0.320	0.297	0.292	0.283	0.278
Consumer Staples	0.307	0.285	0.272	0.252	0.243	0.245
Health Care	0.349	0.339	0.328	0.324	0.306	0.313
Financials	0.450	0.382	0.359	0.397	0.512	0.471
Information Technology	0.386	0.375	0.369	0.354	0.341	0.345
Telecommuncation Services	0.298	0.272	0.263	0.244	0.254	0.248
Utilities	0.321	0.300	0.239	0.241	0.249	0.252
Panel C: EU firms						
Energy	0.405	0.369	0.327	0.302	0.318	0.324
Materials	0.334	0.303	0.273	0.276	0.273	0.276
Industrials	0.293	0.289	0.256	0.249	0.252	0.252
Consumer Discretionary	0.309	0.304	0.282	0.273	0.273	0.266
Consumer Staples	0.261	0.242	0.232	0.221	0.221	0.221
Health Care	0.332	0.326	0.317	0.319	0.299	0.296
Information Technology	0.358	0.338	0.329	0.324	0.315	0.306
Telecommuncation Services	0.272	0.262	0.257	0.255	0.245	0.238
Utilities	0.285	0.290	0.223	0.222	0.244	0.238

Source: Own calculations based on errors incurred from the INDSARD algorithm and level-up benchmark

Appendix B

Source code

I apply a substantial amount of programming and scripts to import data, construct trailing financial statements, conduct yearly valuations and perform data analysis (e.g. graphs, tables and statistical tests). All essential source code and selection algorithms can be accessed through my **GitHub repository**[1]. Note that scripts require a Python 3.6 installation to run. Python as well as the the integrated development environment Spyder can be downloaded altogether trough **Anaconda**[2]. The external appendix includes the same material as provided in GitHub. I generally advise readers to examine the scripts through GitHub. Alternatively, the scripts can be examined using any regular text editor. For Mac users, I recommend **Sublime text**[3]. Windows users should apply **Notepad++**[4]. In the following, I provide snippets of how to use and access the material on GitHub. I also document the most essential parts of the scripts used in the thesis. Finally, I provide a brief overview of the material enclosed in the external appendix and GitHub.

[1] https://github.com/ebhen/sard_thesis_journey
[2] https://www.anaconda.com
[3] https://www.sublimetext.com
[4] https://notepad-plus-plus.org/download/v7.5.6.html

GitHub repository walk-trough

Sub figure B.1 document the front-page of the GitHub repository. All code, data and error vectors are enclose in the master branch (blue letter files). Scrolling down the page, the reader find an overview text on all material available. This is exemplified in the snippet provided in **Sub figure B.1**. Here, information regarding the material enclosed is provided. **Sub figure B.2** shows an example of the commit-timeline provided on GitHub. This can be accessed by clicking on the commits button (15 commits) in the top left corner in **Sub figure B.1**. Here, readers can follow all actions and uploads through the commit-timeline. **Sub figure B.2** shows the download option provided directly on GitHub. My GitHub repository includes both data and scripts. Both need to be downloaded in order to run the scripts. All *.py*-files are python scripts. The *.csv*-files are data files. The *.csv*-files including 'errors' in the file caption represent error vectors stemming directly from the algorithms provided. For instance, 'KK errors alldata 1 7.csv' denotes the errors incurred for the SARD1-model applied on the all firm sample, while 'indwctd.csv' is the all firm data as presented in **Table 4.3** and **5.1**. Error vectors are included in the repository because of the 'new heatmaps evebit subs v1.py' script providing the heat maps used in the thesis. **Sub figure B.3** shows how the reader can examine the code directly in GitHub. Clicking the file marked by the red letters in **A**, the reader access the programming code used for the regular SARD algorithm performed for all firms throughout the valuation period displayed in **B**.

Documentation of essential code

In the following, I provide snippets of the SARD selection algorithm, SQL data import and the implemented functions to perform statistical tests. All code are simple screenshots of the material available on GitHub and the external appendix.

Sub figure B.4 reports the regular SARD selection algorithm used throughout the analysis. In **A** the given target firm in the year of valuation is selected. **B** shows how a peer group is identified when using 5 selection parameters (as such in SARD5). Note that this script is the individual firm code, whereas the empirical results stem from the 'looped' solution provided in 'KKsard subs selector v1.py' (regular SARD-model), 'indSARD subs selector v1.py' (industry-based SARD-model) and 'random lvl indbenchmark sub selector v1.py' (random level-up industry benchmark).

Sub figure B.5 and **B.6** report the snippets of the script used for INDSARD valuation. The code exemplifies selection through INDSARD2 (incl. ROE and Net Debt/EBIT). This is example code from a program that iterates over all sample firms throughout the total sampling period (2000-2019). In extension to this, I also provide one snippet of the Industry benchmark in **Figure B.7**. Notice the difference between the two scripts; in the benchmark i use the random sample (line 192), whereas INDSARD uses the SARD score of ROE and Net Debt/EBIT (in **A** line 380 and 381).

Sub figure B.8 provides snippets of the data import scripts. **A** is the global SQL statement and **B** is the US SQL statement. Note that 'new trailing fin statement global5.py' and 'new trailing fin statement us2.py' were used to calculate the trailing financial statements from the imported data. The script used for data cleaning and sub setting is labelled 'new data construct allfirms v3.py'.

Overview of external appendix

GitHub includes the same material as the one provided on the external appendix. I generally advise all readers to use the GitHub solution. The enumerations below provide a brief description of the material enclosed in the appendix/GitHub.

Data import, construction and trailing financial statements

- **total data extract global and us**: Data import from WRDS, both US- and Global data. This requires a WRDS username and password which is accessible for CBS students and teachers (OBS approx. run time 12-14 hrs.).
- **new trailing fin statement global5**: Trailing financial statements for Global firms (OBS approx. run time 8-9 hrs.). The raw data extracted from the script above is necessary to re-run this script.
- **new trailing fin statement us2**: Trailing financial statements for US firms (OBS approx. run time 8-9 hrs.). The raw data extracted from the script above is necessary to re-run this script.
- **new data construct allfirms v3**: Final data construction including data cleaning and sub samples. Here, I/B/E/S estimates imported from Datastream are also included. I/B/E/S estimates are collected through the terminals at CBS library (OBS approx. run time 10-15 mins.). The trailing financial statements from both US- and global firms are required to run this script.

Selection algorithms; SARD, INDSARD and industry benchmark

- **KKsard subs selector v1**: Regular SARD-models based on Knudsen et al. (2017). Iterative process over all sampling years (2000-2019) for the three different samples; EU-firms, global firms (excl. the US) and all firms.
- **random lvl indbenchmark sub selector v1**: Random level-up benchmark used in **H1** to test accuracy of the SARD approach vis-a-vís estimates based on industry classification.

- **indSARD subs selector v1**: Industry level-up SARD selection algorithm used to test **H2**.
- **SARD individual app v1**: Single firm regular SARD used to identify peers for one particular firm. Used for the illustrative example and represent how the model could be implemented for practical purposes.
- **KKsard country sub selector**: Iterative process over all sampling years (2000-2019) for the all individual country samples. Regular SARD-models.
- **indSARD country subs selector2 v1**: Iterative process over all sampling years (2000-2019) for the all individual country samples. INDSARD-models. Note that this can only run after running the **new data construct allfirms v3**.

Error evaluation and heat maps

- **new heatmaps evebit subs v1**: Heat maps of statistical tests for EV/EBIT only. Remember to close graphs while running the code.

Financial data and error vectors

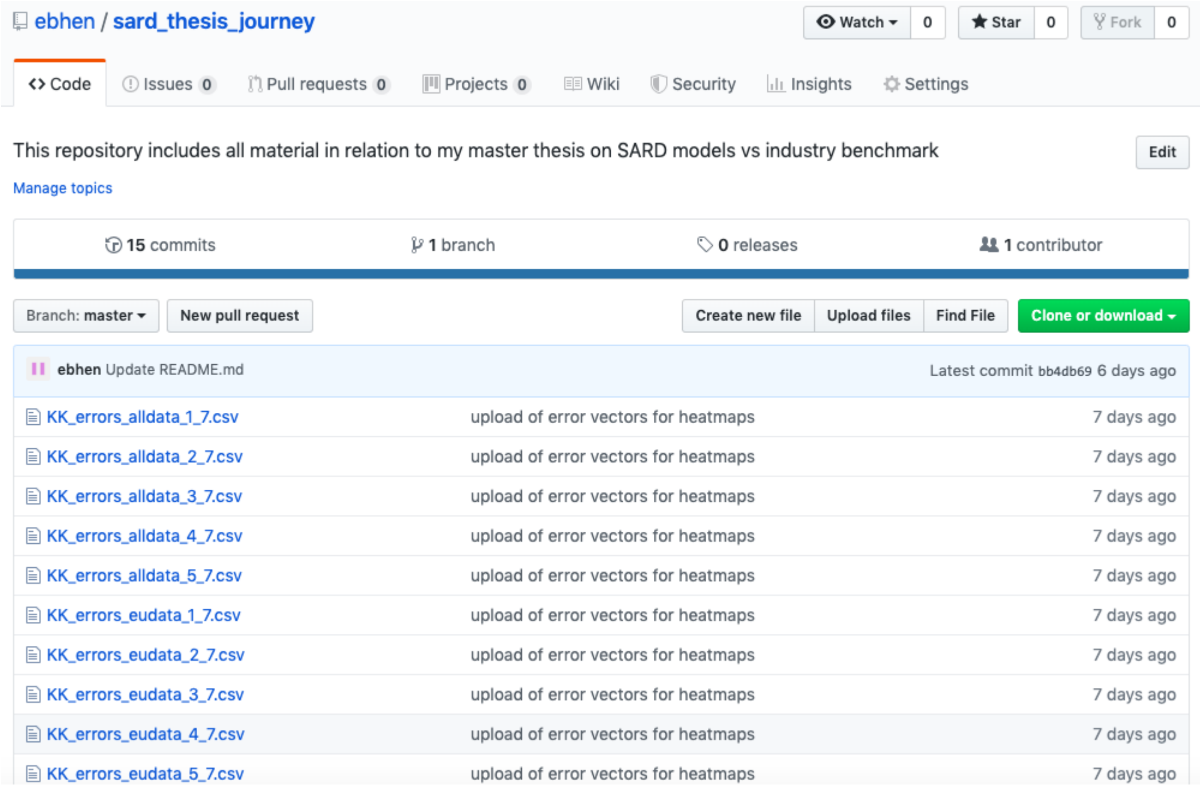
1. **Data used for modeling**
    - **induwtd.csv**: US firm sample. This is used for second part of the cross-country analysis including GICS 2 cleaning. All other country sub samples are left out of the appendix.
    - **indgwtd.csv**: global firms (excl. the US)
    - **indwctd.csv**: all firms
    - **indeuwtd.csv**: EU-firms
    - **US IBES MKTCAP.xlsx**: I/B/E/S estimates, US firms
    - **GLOBAL IBES MKTCAP4.xlsx**: I/B/E/S estimates, Global firms
2. **Error vectors (used for heat maps) OBS! All errors are automatically created when running the scripts enclosed**
    - **All firm errors**: 'KK errors alldata 1 7.csv' through 'KK errors alldata 5 7.csv' denote the estimation errors incurred from regular SARD-models 1 to 5 at the all firm sample.
    - **Global firm errors**: 'KK errors globaldata 1 7.csv' through 'KK errors globaldata 5 7.csv' denote the estimation errors incurred from regular SARD-models 1 to 5 at the global firm (excl. the US).

- **EU-firm errors**: 'KK errors eudata 1 7.csv' through 'KK errors eudata 5 7.csv' denote the estimation errors incurred from regular SARD-models 1 to 5 at the EU-firm.
- **Benchmark errors**: 'all errors benchmark R10E6.csv', 'global errors benchmark R10E6.csv' and 'eu errors benchmark R10E6.csv' denote the estimation errors incurred from random level-up industry benchmark for the three sub samples examined in the large sample analysis.

Figure B.1: GitHub dashboard introduction 1

GitHub is an online service to access- and storage programming code for individuals and enterprises. **Sub figure B.1** illustrates the initial overview page. **Sub figure B.1** reports the README overview provided in GitHub.

(a) Dashboard



(b) Content overview

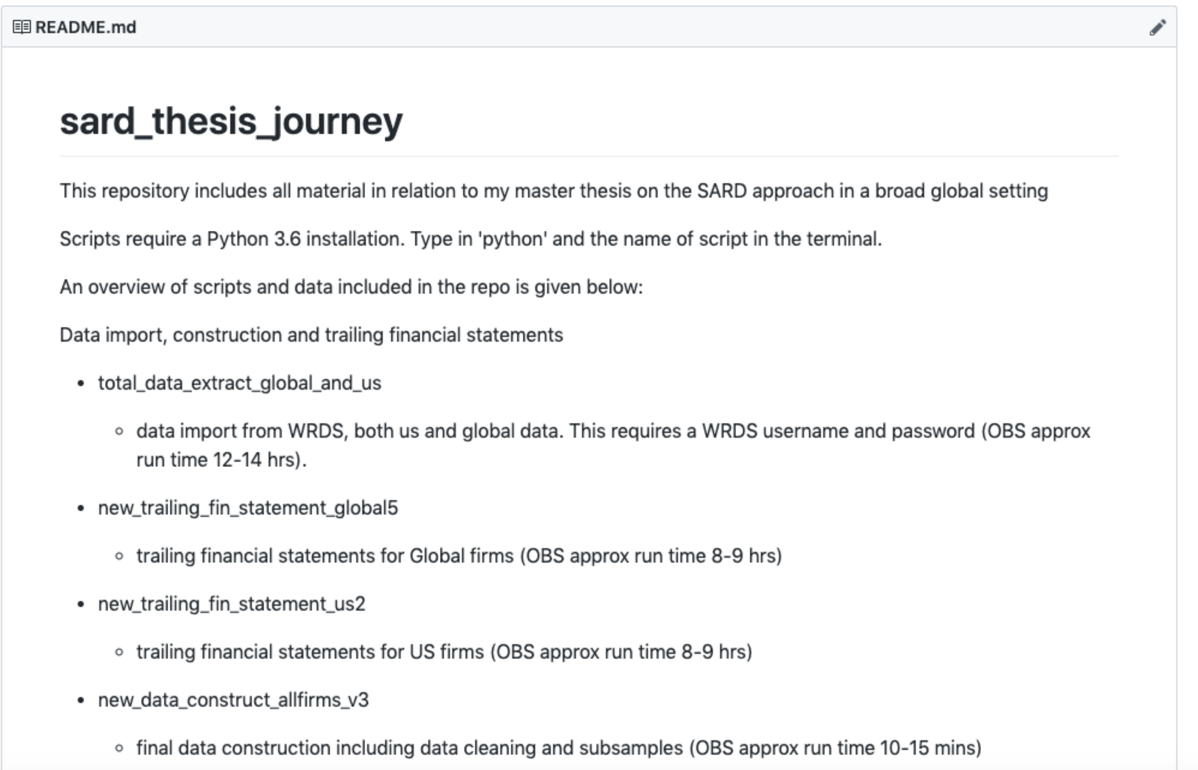
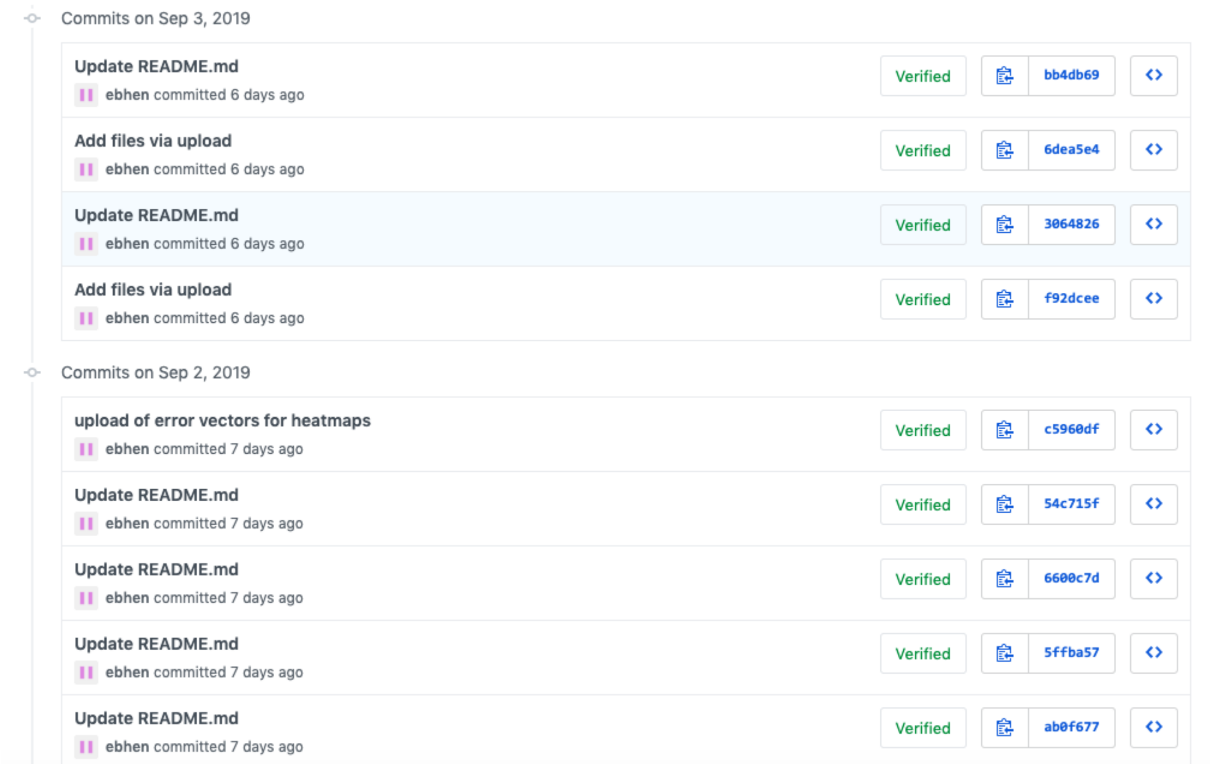


Figure B.2: GitHub dashboard introduction 2

Sub figure B.2 illustrates the commit page where readers can follow my actions on GitHub. **Sub figure B.2** shows how to download all material enclosed as a zip file.

(a) Commits



(b) Download function

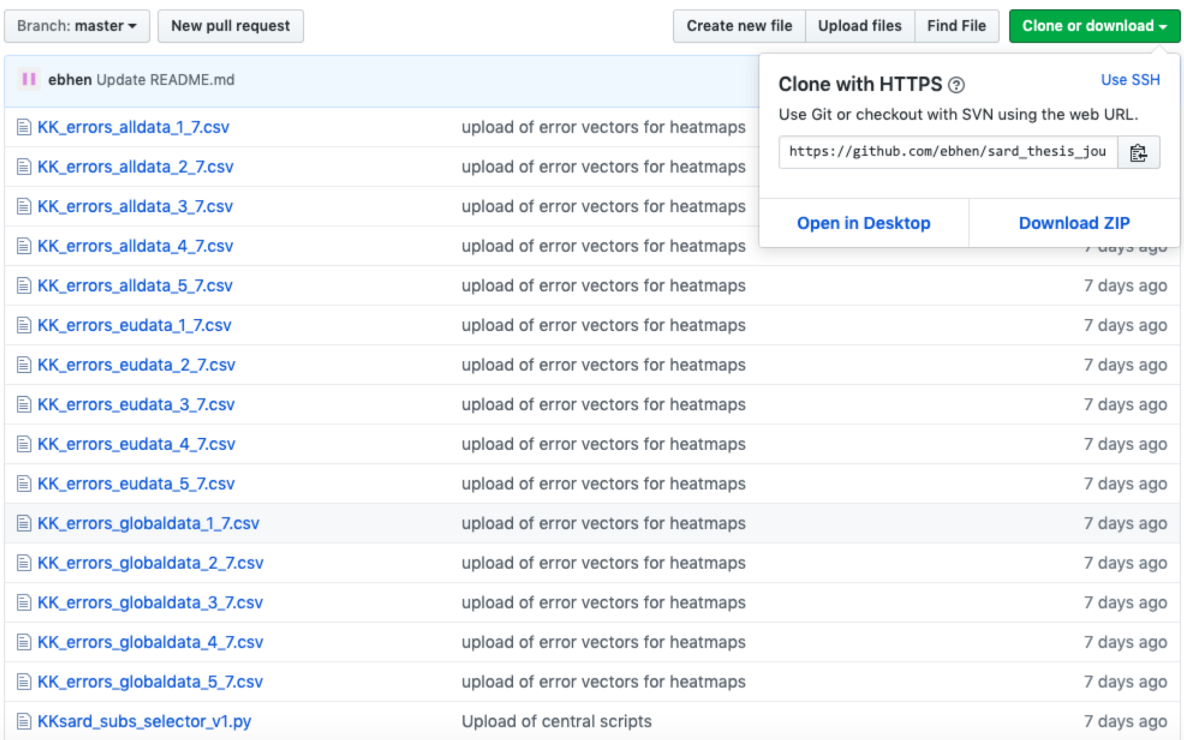
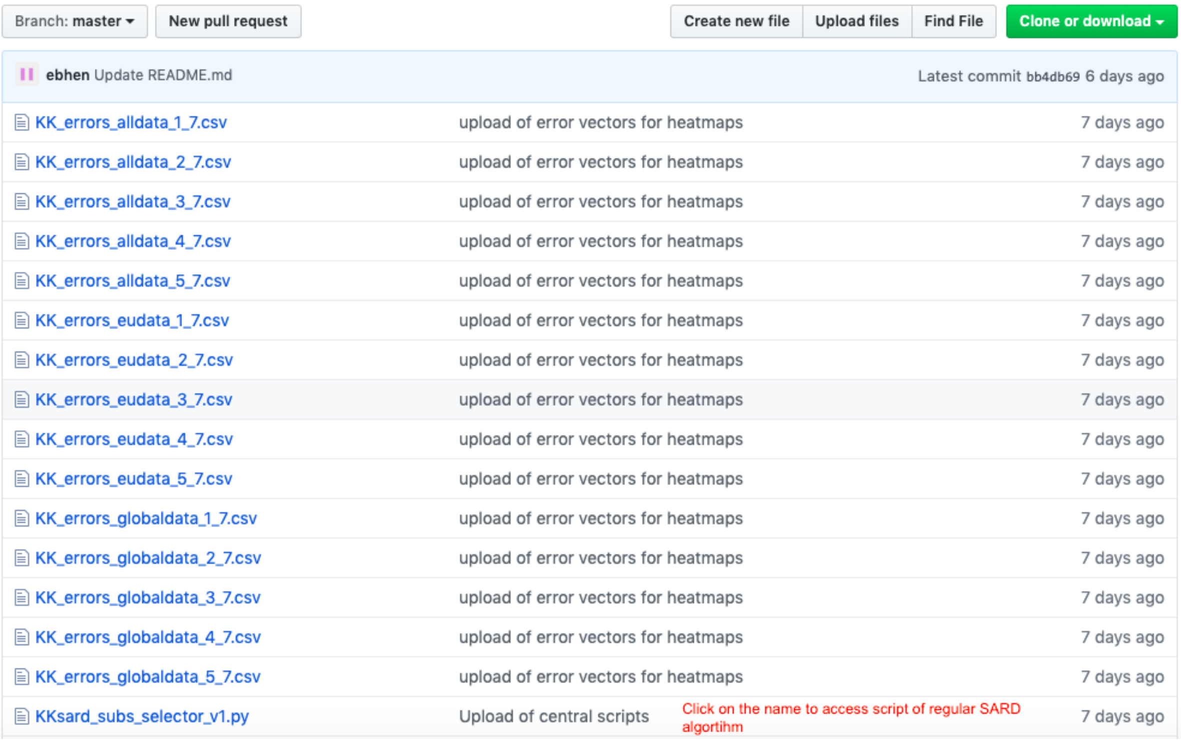


Figure B.3: Access code on GitHub

All programming code can easily be accessed through GitHub. **Sub figure B.3** and **B.3** illustrates the process.

(a) Code access 1



(b) Code access 2

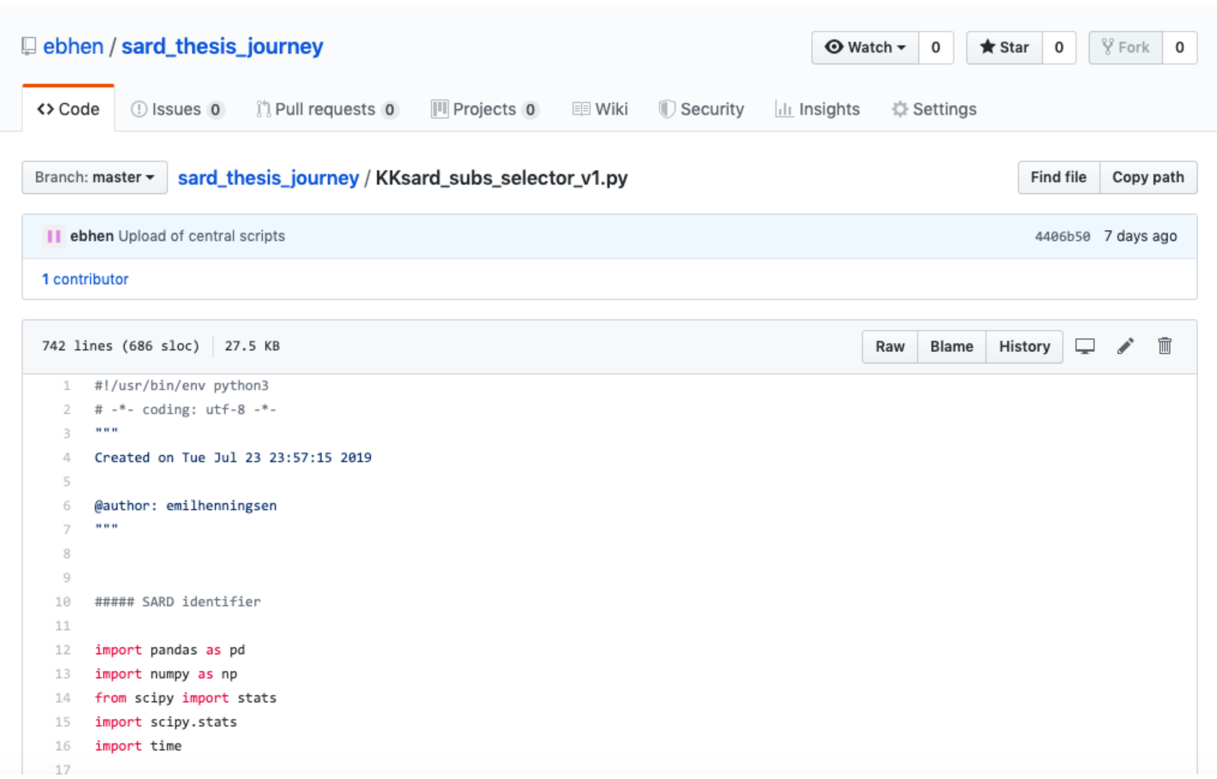


Figure B.4: Example code regular SARD5, single firm

The Sub figure illustrates some of the essential steps in the 'SARD individual app v1.py' file used for a designated target firm in a given year of valuation.

(a) Selection of target firm

```

# Ranking data for SARD algorithm
X2[rvariables] = X2[rvariables].rank(axis=0, method='dense', ascending=False)

# the selection of variables are dynamic in regard to the setting_up_data program
print('Overview of firms to conduct a valuation in the selected year: ')
print(list(X2.NAME))
value_firm = input('Please select a target firm to perform analysis: ')

vfirm = X2.loc[X2['NAME']==value_firm].copy()

# SARD selection algorithm based on input for year, variables and the designated target firm

```

(b) Identification of peer group

```
elif parameters == '5':
    # value firm specifics
    # setting up industry parameter
    cgics2 = vfirm.G_SECTOR.item()
    # setting up selection variables of target firm i
    z1_i = vfirm[[selection_parameter1]].copy() ; z1_i = float(z1_i.iloc[0])
    z2_i = vfirm[[selection_parameter2]].copy() ; z2_i = float(z2_i.iloc[0])
    z3_i = vfirm[[selection_parameter3]].copy() ; z3_i = float(z3_i.iloc[0])
    z4_i = vfirm[[selection_parameter4]].copy() ; z4_i = float(z4_i.iloc[0])
    z5_i = vfirm[[selection_parameter5]].copy() ; z5_i = float(z5_i.iloc[0])
    ##
    ##
    ##
    ##
    ##
    # calculating the SARD score
    X2['Z1ij'] = abs(X2[[selection_parameter1]] - z1_i)
    X2['Z2ij'] = abs(X2[[selection_parameter2]] - z2_i)
    X2['Z3ij'] = abs(X2[[selection_parameter3]] - z3_i)
    X2['Z4ij'] = abs(X2[[selection_parameter4]] - z4_i)
    X2['Z5ij'] = abs(X2[[selection_parameter5]] - z5_i)
    ##
    # SARD score is the sum of absolute distances for the selection variables
    X2['SARD_score'] =  X2.Z1ij + X2.Z2ij + X2.Z3ij + X2.Z4ij + X2.Z5ij
    # finding the peers based on the lowest SARD score given variables z
    temp_peers1 = X2.sort_values(['SARD_score']).copy()
    temp_peers2 = (temp_peers1.iloc[0:estimation_peers]).copy()
    peers = temp_peers2.loc[temp_peers2['NAME'] !=  value_firm].copy()
    # estimating the multiples from the natural peers
    PEest = stats.hmean(peers.PE_RATIO,axis=0)
    PBest = stats.hmean(peers.PB_RATIO,axis=0)
    EV_EBITest = stats.hmean(peers.EV_EBIT,axis=0)
    EV_SALESest = stats.hmean(peers.EV_SALES,axis=0)
```

Figure B.5: Example code INDSARD2, 1

The Sub figure illustrates some of the essential steps in the 'indSARD subs selector v1.py' file used for all firm-year valuations throughout the sampling period 2000-2019.

(a) SARD at the GICS 8 level

```
elif len(X2.columns) == 15:
    for y in years:
        print(y)
        ps = (X2.loc[X2['VALUE_YEAR']==y]).copy()
        ps[cols2] = ps[cols2].rank(axis=0, method='dense', ascending=False)
        companies = list(ps.GVKEY.unique()) # for the inner loop
        print(len(companies))
        for i in companies:
            # industry parameter for the ith company
            cgics2 = (ps.loc[ps['GVKEY']==i].G_SECTOR.item())
            cgics4 = (ps.loc[ps['GVKEY']==i].GGROUP.item())
            cgics6 = (ps.loc[ps['GVKEY']==i].GIND.item())
            cgics8 = (ps.loc[ps['GVKEY']==i].G_SUBIND.item())
            # setting up size parameter
            cindex = (ps.loc[ps['GVKEY']==i].INDEXID.item())
            # setting up sample IDs
            sampleid = (ps.loc[ps['GVKEY']==i].SAMPLE.item())
            # setting up CISO IDs
            ciso_id = (ps.loc[ps['GVKEY']==i].CISO.item())
            # setting up selection variables of target firm i
            z1_i = (ps.loc[ps['GVKEY']==i].ROE.item())
            z2_i = (ps.loc[ps['GVKEY']==i].NET_DEBT_EBIT.item())
            # applying SARD algorithm
            ps['Z1ij'] = abs(ps['ROE'] - z1_i)
            ps['Z2ij'] = abs(ps['NET_DEBT_EBIT'] - z2_i)
            # SARD score is the sum of absolute distances for the selection variables
            ps['SARD_score'] = ps.Z1ij + ps.Z2ij
            if len(ps.loc[ps['G_SUBIND']==cgics8]) - 1 >= required_peers:
                ind_peers = ps.loc[(ps['G_SUBIND'] == cgics8) & (ps['GVKEY'] != i)].copy() # the same gics8, no target firm
                temp_peers1 = ind_peers.sort_values(['SARD_score']).copy()
                peers = (temp_peers1.iloc[0:estimation_peers]).copy()
                # estimating the multiples from the natural peers
                PEest = stats.hmean(peers.PE_RATIO,axis=0)
                PBest = stats.hmean(peers.PB_RATIO,axis=0)
                EV_EBITest = stats.hmean(peers.EV_EBIT,axis=0)
```

(b) SARD at the GICS 6 level

```
            elif len(ps.loc[ps['GIND']==cgics6]) - 1 >= required_peers:
                ind_peers = ps.loc[(ps['GIND'] == cgics6) & (ps['GVKEY'] != i)].copy() # the same gics8, no target firm
                temp_peers1 = ind_peers.sort_values(['SARD_score']).copy()
                peers = (temp_peers1.iloc[0:estimation_peers]).copy()
                # estimating the multiples from the natural peers
                PEest = stats.hmean(peers.PE_RATIO,axis=0)
                PBest = stats.hmean(peers.PB_RATIO,axis=0)
                EV_EBITest = stats.hmean(peers.EV_EBIT,axis=0)
                EV_SALESest = stats.hmean(peers.EV_SALES,axis=0)
                # populating estimation lists
                sPE.append(PEest.item())
                sPB.append(PBest.item())
                sEVEBIT.append(EV_EBITest.item())
```

Figure B.6: Example code INDSARD2, 2

(a) SARD at the GICS 4 level

```
        elif len(ps.loc[ps['GGROUP']==cgics4]) - 1 >= required_peers:
            ind_peers = ps.loc[(ps['GGROUP'] == cgics4) & (ps['GVKEY'] != i)].copy() # the same gics8, no target firm
            temp_peers1 = ind_peers.sort_values(['SARD_score']).copy()
            peers = (temp_peers1.iloc[0:estimation_peers]).copy()
            # estimating the multiples from the natural peers
            PEest = stats.hmean(peers.PE_RATIO,axis=0)
            PBest = stats.hmean(peers.PB_RATIO,axis=0)
            EV_EBITest = stats.hmean(peers.EV_EBIT,axis=0)
            EV_SALESest = stats.hmean(peers.EV_SALES,axis=0)
            # populating estimation lists
            sPE.append(PEest.item())
            sPB.append(PBest.item())
            sEVEBIT.append(EV_EBITest.item())
```

(b) SARD at the GICS 2 level

```
        else:
            ind_peers = ps.loc[(ps['G_SECTOR'] == cgics2) & (ps['GVKEY'] != i)].copy() # the same gics8, no target firm
            temp_peers1 = ind_peers.sort_values(['SARD_score']).copy()
            peers = (temp_peers1.iloc[0:estimation_peers]).copy()
            # estimating the multiples from the natural peers
            PEest = stats.hmean(peers.PE_RATIO,axis=0)
            PBest = stats.hmean(peers.PB_RATIO,axis=0)
            EV_EBITest = stats.hmean(peers.EV_EBIT,axis=0)
            EV_SALESest = stats.hmean(peers.EV_SALES,axis=0)
            # populating estimation lists
            sPE.append(PEest.item())
            sPB.append(PBest.item())
            sEVEBIT.append(EV_EBITest.item())
```

Figure B.7: Example code, random industry level-up benchmark

One snippet of the random industry level-up benchmark performed at the GICS 8 level. Similar procedure at the 'higher' GICS levels.

```
for y in years:
    print(y)
    ps = (X.loc[X['VALUE_YEAR']==y]).copy()
    companies = list(ps.GVKEY.unique()) # for the inner loop
    print(len(companies))
    for i in companies:
        # industry parameter for the ith company
        cgics2 = (ps.loc[ps['GVKEY']==i].G_SECTOR.item())
        cgics4 = (ps.loc[ps['GVKEY']==i].GGROUP.item())
        cgics6 = (ps.loc[ps['GVKEY']==i].GIND.item())
        cgics8 = (ps.loc[ps['GVKEY']==i].G_SUBIND.item())
        # setting up size parameter
        cindex = (ps.loc[ps['GVKEY']==i].INDEXID.item())
        # setting up sample IDs
        sampleid = (ps.loc[ps['GVKEY']==i].SAMPLE.item())
        # setting up CISO IDs
        ciso_id = (ps.loc[ps['GVKEY']==i].CISO.item())
        # we start off in the highest detail level of industry - subind
        if len(ps.loc[ps['G_SUBIND']==cgics8]) - 1 >= required_peers:
            temp_peers = ps.loc[(ps['G_SUBIND'] == cgics8) & (ps['GVKEY'] != i)].copy() # the same gics8, no target firm
            peers = temp_peers.sample(n=estimation_peers, random_state=seed).copy() # randomly selected sub ind peers
            # estimating the multiples from the randomly selected industry peers
            PEest = stats.hmean(peers.PE_RATIO,axis=0)
            PBest = stats.hmean(peers.PB_RATIO,axis=0)
            EV_SALESest = stats.hmean(peers.EV_SALES,axis=0)
            EV_EBITest = stats.hmean(peers.EV_EBIT,axis=0)
            # populating estimation lists
            sPE.append(PEest.item())
            sPB.append(PBest.item())
            sEVEBIT.append(EV_EBITest.item())
```

Figure B.8: SQL import scripts

Sub figure B.8 illustrates the import of global accounting data, whereas **Sub figure B.8** shows the import statement for both accounting fundamentals and market prices for US firms. Both are snippets from the 'total data extract global and us.py' script.

(a) Global selection

```
# Extracting total data on compustat for both quarterly and annual data
# SQL path
global_fundq = db.raw_sql("SELECT conm,isin,gvkey,loc,fyearq,fyr,datafqtr,datacqtr,datadate,acctstdq,curcdq,saleq,oiadpq,ibq,dlttq,dlcq,che
print(len(global_fundq))

global_funda = db.raw_sql("SELECT conm,isin,gvkey,loc,fyear,fyr,datadate,acctstd,curcd,sale,oiadp,ib,dltt,dlc,che,ceq from compg.g_funda wh
print(len(global_funda))

global_fundq.to_csv('total_global_fundq_period.csv',index=False)
global_funda.to_csv('total_global_funda_period.csv',index=False)

temp = tuple(global_fundq.gvkey.unique().copy())
parm_ind = {'ID_GVKEY': temp }

# SQL path to industry groups
global_comp_ind = db.raw_sql('SELECT conm,gvkey,ggroup,gind,gsector,gsubind from compg.g_company WHERE g_company.gvkey in %(ID_GVKEY)s', pa
global_comp_ind.to_csv('global_company_GICS.csv',index=False)

# create tuple of unique currencies in subset for import
currency = tuple(global_fundq.curcdq.unique())

# creating import parameter
parm_curr = {'ID_CURR': currency }

# sql import of currency data from compustat
data_curr = db.raw_sql('SELECT datadate,fromcurd,tocurd,exratd from compg.g_exrt_dly WHERE g_exrt_dly.tocurd in %(ID_CURR)s', params=parm_c

# Now importing USD exchange rates for the period
data_curr_usd = db.raw_sql("SELECT datadate,fromcurd,tocurd,exratd from compg.g_exrt_dly WHERE g_exrt_dly.tocurd in ('USD')")

data_curr.to_csv('exchange_rates_total.csv',index=False)
data_curr_usd.to_csv('usd_rates.csv',index=False)
```

(b) US selection

```
# Extracting total data on compustat
us_fundq = db.raw_sql("SELECT conm,cusip,gvkey,fyearq,fyr,datafqtr,datacqtr,rdq,datadate,acctstdq,curcdq,saleq,oiadpq,ibq,dlttq,dlcq,cheq,c
print(len(us_fundq))
us_funda = db.raw_sql("SELECT conm,cusip,gvkey,fyear,fyr,datadate,acctstd,curcd,sale,oiadp,ib,dltt,dlc,che,ceq,prcc_c,prcc_f,csho from comp
print(len(us_funda))

us_fundq.to_csv('total_us_fundq_period.csv',index=False)
us_funda.to_csv('total_us_funda_period.csv',index=False)

# S&P 1500 constituents
SP_comp = db.raw_sql("select gvkey, iid, indexid, datadate "
           "from compd.spidx_cst "
           "where datadate in ('2005-03-31','2006-03-31','2007-03-30','2008-03-31','2009-03-31','2010-03-31',"
           "'2011-03-31','2012-03-30','2013-03-28','2014-03-31','2015-03-31','2016-03-31','2017-03-31','2018-03-29',"
           "'2019-03-29','2004-03-31','2003-03-31','2002-03-28','2001-03-30','2000-03-31','1999-03-31')")

SP_comp.to_csv('total_sp1500_comp.csv',index=False)

# creating tuple of unique company gvkeys for further data entry
us_comp_list = tuple(SP_comp.gvkey.unique())
# creating parm for sql import
parm = {'ID_GVKEY': us_comp_list }
# Industry groups
us_comp_ind = db.raw_sql('SELECT conm,gvkey,ggroup,gind,gsector,gsubind from compd.company WHERE company.gvkey in %(ID_GVKEY)s', params=par

us_comp_ind.to_csv('us_company_GICS.csv',index=False)
# marketprices for US firms at valuation dates
us_marketprices = db.raw_sql("select gvkey,cusip,conm,datadate,fyearq,fyr,datafqtr,datacqtr,prccq,cshoq "
           "from compd.fundq "
           "where datadate in ('2005-03-31','2006-03-31','2007-03-31','2008-03-31','2009-03-31','2010-03-31',"
           "'2011-03-31','2012-03-31','2013-03-31','2014-03-31','2015-03-31','2016-03-31','2017-03-31','2018-03-31',"
           "'2019-03-31','2004-03-31','2003-03-31','2002-03-31','2001-03-31','2000-03-31','1999-03-31')")
```

Figure B.9: Statistical test, functions

Sub figure B.9 reports the Wilcoxon test function, whereas **Sub figure B.9** shows the *t*-test for difference in mean between matched pairs. Snippets taken from the 'new heatmaps evebit subs v1.py' script.

(a) Wilcoxon signed rank test

```
####### Function to conduct wilcoxon signed rank test for greater and lesser respectively
# n will be corrected so that I only look at the non-zero pairwise differences

# E(T) = mu_T = n(n+1) / 4
# Var(T) = sigma_T^2 = (n(n+1)(2n+1)) / 24

# (T - E(T)) / sigma < -z_a

def wtest_greater(x):
    E_t = (len(x[x != 0]) * (len(x[x != 0]) + 1)) / 4 # expected value of length of non-zero differences
    sigma_t = math.sqrt((len(x[x != 0]) * (len(x[x != 0]) +1) * ((2*len(x[x != 0]) + 1))) / 24) # variance
    x = x.replace(0, np.nan).copy() # zero differences are removed from the smaple
    x_sign = np.sign(x).copy() # signs to conduct the signed rank
    ax = abs(x).copy() # absolute differences
    sr_x = ax.rank(method='average',ascending=True) * x_sign # ranking of abs difference
    wg = sr_x[sr_x > 0].sum().copy() # investigating greater (i.e. postive differences) in the sample
    zg = (wg - E_t) / sigma_t # Z value for t-test
    p_val = scipy.stats.norm.sf(abs(zg))*2 #two-sided
    return wg, zg, p_val
```

(b) *t*-test for mean difference between matched pairs

```
def diff_pair_means(x):
    d_bar = x.mean()
    sd = x.std()
    n = math.sqrt(len(x))
    t = d_bar / (sd / n)
    p_valx = stats.t.sf(np.abs(t), len(x)-1)*2 # two-sided
    return t, p_valx
```