

# Die Vergleichbarkeit von Unternehmen bei pan-europäischen Benchmarkstudien

Von Dr. Christian Schwarz, Dr. Stefan Stein, Dr. Nils Holinski und Sebastian Hoffmann\*

*Zur Planung oder Überprüfung von Verrechnungspreisen werden häufig sogenannte pan-europäische Benchmarkstudien durchgeführt. Die darin enthaltene große Anzahl an Unternehmen ermöglicht die Berücksichtigung von weiteren Vergleichbarkeitsfaktoren (Branchen oder Funktionen) und stellt somit einen hohen Grad an Vergleichbarkeit sicher. Im Rahmen des vorliegenden Beitrags wird analysiert, ob der pan-europäische Benchmarkansatz grundsätzlich Gültigkeit hat. Aktuelle Daten zeigen, dass ein pan-europäischer Benchmarkansatz grundsätzlich Gültigkeit hat und somit als wichtiger Baustein bei der Planung und Überprüfung steuerlicher Verrechnungspreise dienen kann.*

## 1. Einleitung

Das Benchmarking eigener Schlüsselkennzahlen mit den Finanzkennzahlen geeigneter Vergleichsunternehmen (z. B. Wettbewerber) stellt eine wichtige betriebswirtschaftliche Disziplin dar. Durch den Vergleich im Rahmen eines Benchmarks mit anderen Unternehmen können Rückschlüsse für die eigene finanzielle Performance gezogen und in der Folge betriebswirtschaftlich notwendige Maßnahmen für das eigene Unternehmen gezogen werden. Aufgrund der hohen Brisanz dieser Benchmark Daten sind die einbezogenen Vergleichsunternehmen hinsichtlich ihrer Vergleichbarkeit zu validieren.

Auch das Steuerrecht bedient sich dieser betriebswirtschaftlichen Disziplin, um bei konzerninternen Transaktionen zu überprüfen, ob die angesetzten Verrechnungspreise eine Profitabilität ermöglichen, welche auch zwischen unabhängigen Vergleichsunternehmen am freien Markt mit vergleichbaren Transaktionen erzielt werden. Dies ist die fundamentale Prämisse des sogenannten Fremdvergleichsgrundsatzes (englisch: „Arm's Length Principle“).

Die zur Erfüllung des Fremdvergleichsgrundsatzes steuerlich erforderlichen Benchmarkanalysen werden in der Regel auf öffentlich zugängliche Datenbanken gestützt, welche relevante Finanzzahlen für eine Vielzahl von Unternehmen einer bestimmten Region (z. B. Europa) beinhalten. Allerdings ist zu beobachten, dass sich die Datenverfügbarkeit zwischen einzelnen Ländern einer Region erheblich unterscheiden kann und es daher in der Regel notwendig ist, hinreichend einheitliche Märkte zu aggregieren. Die somit erzielte größere Grundgesamtheit an Unternehmen stellt unter Berücksichtigung weiterer Vergleichbarkeitsfaktoren, wie beispielsweise Branchen oder ausgeübte Funktionen, einen hohen Grad an Vergleichbarkeit sicher.

\* Dr. Christian Schwarz, StB/FBISr, Dr. Stefan Stein, StB, Dr. Nils Holinski und Sebastian Hoffmann sind Mitarbeiter der KPMG AG Wirtschaftsprüfungsgesellschaft. Dr. Christian Schwarz ist daneben Vertretungsprofessor für quantitative Methoden an der Hochschule Düsseldorf.

Vor diesem Hintergrund haben sich in der Verrechnungspreispraxis für die Region Europa sog. pan-europäische Benchmarkstudien etabliert, welche in ihrer Grundgesamtheit sämtliche verfügbare Unternehmen der 28 EU-Mitgliedstaaten sowie der Länder der Europäischen Freihandelsassoziation („EFTA“)<sup>1</sup> berücksichtigen.

Im Rahmen des vorliegenden Beitrags wird analysiert, ob der pan-europäische Benchmarkansatz grundsätzlich Gültigkeit hat. Die hohe Aktualität dieser Fragestellung wird durch das jüngst am 25.01.2017 veröffentlichte OECD Diskussionspapier zu genau dieser Fragestellung unterstrichen.<sup>2</sup> Für Unternehmen ist dies von großer Bedeutung, da pan-europäische Benchmarkstudien häufig einen zentralen Baustein bei der Planung und Verprobung steuerlicher Verrechnungspreise bilden.

## 2. Umsetzung des Fremdvergleichsgrundsatzes mithilfe von Datenbankstudien

### 2.1. Darlegung der Fremdüblichkeit konzerninterner Verrechnungspreise

Verrechnungspreise beschreiben Wertansätze für den Güter- und Dienstleistungsaustausch innerhalb von Unternehmen und Konzernen, welche nicht über Angebot und Nachfrage freier Märkte koordiniert werden, sondern vielmehr Subjekt einer zweckorientierten Bewertung durch Unternehmen sind.<sup>3</sup>

Durch den Fremdvergleichsgrundsatz fordert das Steuerrecht von dieser zweckorientierten Bewertung zu abstrahieren, so dass die konzerninterne Verrechnung zu einer Gewinnverteilung führt, die vergleichbar zu jener bei marktkoordinierten Preisen ist. Aufgrund der beschränkten Ermittlungsmöglichkeiten der Finanzbehörden im Ausland sind Steuerpflichtige bei Auslandssachverhalten zu einer erhöhten Mitwirkung verpflichtet. Im Zusammenhang mit konzerninternen Geschäftsbeziehungen müssen diese ihr ernsthaftes Bemühen darlegen, dass Verrechnungspreise im Einklang mit dem Fremdvergleichsgrundsatz ausgestaltet werden.<sup>4</sup> Hierbei sollen sich die Steuerpflichtigen auf objektive Kriterien, soweit möglich auf

1 Dabei handelt es sich um Island, Norwegen und die Schweiz. Aufgrund geringer Datenverfügbarkeit wurde Liechtenstein nicht berücksichtigt.

2 Vgl. The Platform for Collaboration on Tax, A Toolkit for Addressing Difficulties in Accessing Comparables Data for Transfer Pricing Analyses, OECD and International Monetary Fund, United Nations and World Bank Group.

3 Vgl. etwa Baumhoff, in: Festaussgabe Wassermeyer, Doppelbesteuerung, Kapitel 34, Rn. 1.

4 Vgl. § 1 Abs. 1 Satz 1 GAufzV; Tz. 3.4.12.1 VerwGr.-Verf. 2005.

(betriebsexterne<sup>5</sup>) Fremdvergleichsdaten, stützen.<sup>6</sup> Soweit geeignet können diese aus öffentlich zugänglichen Datenbanken entnommen werden.

## 2.2. Geeignetheit von Datenbankanalysen zur Darlegung der Fremdüblichkeit

Die Akzeptanz einer Verprobung der Fremdüblichkeit konzerninterner Verrechnungspreise unter Rückgriff auf Fremdvergleichsdaten aus Datenbanken hängt im Wesentlichen von der Vergleichbarkeit der Geschäftsvorfälle des betrachteten Unternehmens mit denen der vergleichbaren Unternehmen aus der Datenbank ab. Mithin von der Frage, ob unter dieser Prämisse hinreichend viele Vergleichsunternehmen in den Datenbanken vorzufinden sind. Hierzu sind die konkreten Geschäftsbedingungen zu identifizieren und einer Vergleichbarkeitsprüfung zu unterziehen.<sup>7</sup> Ist als Ergebnis dieser Prüfung eine Unvergleichbarkeit gegeben, d. h. unterscheiden sich Geschäftsbedingungen, die eine wesentliche Auswirkung auf den Preis und Unternehmensgewinn haben, so wesentlich, dass diese nicht durch eine Anpassungsrechnung nivelliert werden können, so ist nach Auffassung der Finanzverwaltung die Darlegung der Fremdüblichkeit anhand von Unternehmensinformationen aus Datenbanken abzulehnen. Die Vergleichbarkeit soll insbesondere dann nicht gegeben sein, wenn das betrachtete Unternehmen besonders wertvolle immaterielle Wirtschaftsgüter einsetzt oder unternehmensspezifische Funktionen ausübt und Risiken trägt, die keine Entsprechung bei den potentiellen Vergleichsunternehmen finden.<sup>8</sup> Hieraus leitet die Finanzverwaltung ab, dass Datenbankstudien nur für die Verprobung der Gewinne von Routineunternehmen zur Anwendung kommen können.<sup>9</sup> Diese Unternehmen üben in der Regel einfache Funktionen aus, setzen nur in geringem Umfang Wirtschaftsgütern ein und sind nur geringen Geschäftsrisiken ausgesetzt.

Vor diesem regulatorischen Hintergrund werden Datenbankstudien in der Praxis häufig zum Nachweis der Fremdüblichkeit von (i) konzerninternen Dienstleistungen, (ii) einfachen Vertriebstätigkeiten eines Kommissionärs, Handelsvertreters oder Vertragshändlers (low risk distributor)<sup>10</sup> und (iii) Fertigungstätigkeiten im Rang eines Auftrags- oder Lohnfertigers eingesetzt.

## 2.3. Datenbankanalyse als Quelle betriebsexterner Fremdvergleichsdaten

Die betriebsexternen Fremdvergleichsdaten zur Überprüfung der Angemessenheit steuerlicher Verrechnungspreise werden in der Praxis im Rahmen von Benchmarkanalysen aus öffent-

lich zugänglichen Datenbanken erhoben.<sup>11</sup> Hierbei werden oftmals Renditekennziffern (d. h. das Verhältnis aus EBIT und einer geeigneter Bezugsgröße wie z. B. Umsatzerlöse oder Kosten) vergleichbarer unabhängiger Unternehmen als Referenz herangezogen und mit den Renditekennziffern des zu überprüfenden Unternehmens (der sogenannten „Tested Party“) verglichen.

Zur Überprüfung der Vergleichbarkeit der beobachteten Marktdaten definieren sowohl die OECD-Verrechnungspreisleitlinien als auch die deutsche Finanzverwaltung Vergleichbarkeitsfaktoren.<sup>12</sup> Allerdings kann im Rahmen von Benchmarkanalysen nicht per se von einer vollständigen Vergleichbarkeit der Referenzunternehmen ausgegangen werden, da die verfügbaren Informationen nicht ausreichen, damit alle definierten Vergleichbarkeitsfaktoren gleichermaßen Anwendung finden können.<sup>13</sup>

Den daraus resultierenden Unschärfen wird regelmäßig dadurch begegnet, dass die Bandbreite an beobachteten Fremdvergleichswerten durch Ausschluss der oberen und unteren 25 % der Beobachtungen auf die Interquartilsbandbreite von Fremdvergleichswerten eingegrenzt wird.<sup>14</sup>

Die Finanzverwaltung vertritt die Auffassung, dass Informationen aus Datenbanken alleine in aller Regel nicht für die Überprüfung der Angemessenheit ausreichen und erachtet die Ergebnisse eines reinen „Datenbankscreening“ regelmäßig als unverwertbar.<sup>15</sup> Vielmehr sei es erforderlich, die in den Datenbanken nicht enthaltenen Informationen aber für den Vergleich erheblicher Umstände auf andere Weise (z. B. Internetauftritte) zu ermitteln.

Allerdings ist in Bezug auf die Verwendung eines reinen quantitativen „Datenbankscreenings“ aus statistischer Sicht anzumerken, dass die Ergebnisse und mithin die Datenqualität einer Datenbankstudie aufgrund der großen Anzahl an Vergleichsunternehmen regelmäßig zu robusten und repräsentativen Ergebnissen führen. Daher ist der Auffassung der Finanzbehörde<sup>16</sup> statistisch zu entgegnen, dass bei einer geeigneten rein quantitativen Suchstrategie die Datenqualität ebenfalls zu einer fremdvergleichskonformen Bandbreite führt und Ausreißer durch die Interquartilsbandbreitenbetrachtung ausgeschlossen werden. Dass die Ergebnisse breit angelegter Datenbankstudien zu repräsentativen Ergebnissen führen, zeigen diverse Studien, deren Ergebnisse konsistent zu den praktischen Erfahrungen von Datenbankstudien mit einem zusätzlichen qualitativen Screening stehen.<sup>17</sup>

5 Vgl. Tz. 3.4.12.2 VerwGr.-Verf. 2005.

6 Vgl. § 1 Abs. 3 Satz 3 GAufzV; Tz. 3.4.12.3 VerwGr.-Verf. 2005.

7 Vgl. Tz. 3.4.12.7 VerwGr.-Verf. 2005.

8 Vgl. Tz. 3.4.12.7 Buchst. b VerwGr.-Verf. 2005.

9 Vgl. Tz. 3.4.10.3 Buchst. b VerwGr.-Verf. 2005.

10 Siehe für eine Unterscheidung dieser typischen Vertriebsmodelle und möglicher Implikationen durch die OECD Initiative zur Bekämpfung von Gewinnverkürzungen und Gewinnverlagerungen etwa *Freudenberg/Stein/Weskamp*, Ubg 2016, S. 603 ff.

11 Zur Nutzung von Datenbankstudien siehe VerwGr.-Verf. 2005, Tz. 3.4.12.4 f.; OECD-Verrechnungspreisleitlinien Tz. 3.30 ff.; *Wahl/Preisser*, IStR 2008, S. 51 ff.

12 Vgl., Tz. 3.4.12.7 VerwGr.-Verf. 2005; OECD-Verrechnungspreisleitlinien Tz. 1.36 ff.

13 Beispiele für üblicherweise nicht verfügbare Unternehmensinformationen sind die Vertragsverhältnisse von Vergleichsunternehmen.

14 Vgl. § 1 Abs. 3 Satz 3 AStG; Tz. 3.4.12.5 VerwGr.-Verf. 2005.

15 Vgl. Tz. 3.4.12.3 VerwGr.-Verf. 2005.

16 Vgl. Tz. 3.4.12.3 VerwGr.-Verf. 2005.

17 Siehe hierzu *Schwarz/Holinski/Stein/Hoffmann*, DB 2016, S. 2857 bis 2861; *Schwarz/Holinski/Stein/Hoffmann*, DB 2016, S. 2307 bis 2311; *Krüger/Nientimp/Schwarz*, Ubg 2016, S. 221 bis 225.

Des Weiteren sollen Datenbankstudien nach Verwaltungsauffassung nur für die Verprobung der Gewinne von Routineunternehmen zur Anwendung kommen. Für Entrepreneurseien Datenbankstudien unter Verwendung von Renditekennziffern funktional vergleichbarer Vergleichsunternehmen nicht anwendbar, da solche Unternehmen erhebliche individuelle Risiken tragen, für die eine ausreichende Entsprechung in den Renditekennziffern der Vergleichsunternehmen oftmals nicht hinreichend sicher feststellbar ist.<sup>18</sup>

Zusammenfassend bleibt trotz einiger Kritik<sup>19</sup> an der Datenverfügbarkeit und der -qualität öffentlich zugänglicher Datenbanken festzuhalten, dass Benchmarkanalysen sowohl zur Dokumentation der steuerlichen Angemessenheit der steuerlichen Einkunftsabgrenzung zwischen Konzerngesellschaften als auch zur Planung von steuerlichen Verrechnungspreisen sachgerecht und international etabliert sind.

#### 2.4. Europäischer Binnenmarkt als Grundlage pan-europäischer Studien

Die Qualität einer Benchmarkanalyse hängt oftmals von der Verfügbarkeit einer hinreichend großen Datengrundgesamtheit ab. Wie bereits oben dargelegt, steigen die Robustheit und damit verbunden auch der statistische Aussagegehalt einer Benchmarkanalyse mit einer steigenden Anzahl vergleichbarer Referenzdaten. Es ist allerdings zu beobachten, dass in den öffentlich zugänglichen Datenbanken oftmals nicht genügend Vergleichsunternehmen für einzelne europäische Länder (z. B. den Ansässigkeitsstaat der Tested Party) verfügbar sind. Als Lösung wird in der Praxis die Grundgesamtheit für Benchmarkstudien auf andere Mitgliedstaaten der EU ausgeweitet (sog. „pan-europäische“ Benchmarkstudie), damit zusätzlich zu dem Länderkriterium auch die anderen Vergleichbarkeitskriterien hinreichend berücksichtigt werden können. Hierfür bestehen aus ökonomischer Sicht zahlreiche Argumente, welche im Folgenden diskutiert werden.

Historisch befindet sich Europa seit 1957 und dem „Vertrag von Rom“ in einem Prozess der gesellschaftlichen, aber vor allem auch ökonomischen Harmonisierung. Die Grundpfeiler der EU und Ergebnis dieses Harmonisierungsprozesses und der Schaffung eines europäischen Binnenmarkts sind der freie Warenaustausch, die freie Mobilität von Personen, die Kapitalfreiheit und die Dienstleistungsfreiheit. Durch die Einführung des Euros als gemeinsame „Binnenmarktwährung“ im Jahr 2002 im Zusammenspiel mit einer Integration der Finanzmärkte ist de facto das Wechselkursrisiko für Transaktionen innerhalb des Euro-Raums vollständig eliminiert worden. Empirisch kann für den europäischen Binnenmarkt beobachtet werden, dass die Handelsvolumina durch den Abbau von Handelsbeschränkungen in der EU (z. B. Zölle oder der Abbau administrativer Anforderungen) sehr stark angestiegen sind.<sup>20</sup>

18 Vgl. Tz. 3.4.10.3 Buchst. b VerwGr.-Verf. 2005.

19 Vgl. etwa Naumann, IStR 2013, S. 617.

20 Vgl. statistisches Amt der Europäischen Kommission (Eurostat) abrufbar unter: [http://ec.europa.eu/eurostat/statistics-explained/index.php/Intra-EU\\_trade\\_in\\_goods\\_-\\_recent\\_trends](http://ec.europa.eu/eurostat/statistics-explained/index.php/Intra-EU_trade_in_goods_-_recent_trends)

Trotz aller Bemühungen zur Schaffung eines einheitlichen Binnenmarkts sind auch weiterhin innerhalb der EU strukturelle Unterschiede zu beobachten. So unterscheidet sich die Kaufkraft zwischen verschiedenen Regionen zum Teil noch immer erheblich. Dies ist auf verschiedene Faktoren, wie z. B. Preisbindungen aber auch regulatorische Rahmenbedingungen, zurückzuführen. Bezugnehmend auf empirische Studien ist allerdings festzustellen, dass die Preisunterschiede zwischen einzelnen EU-Mitgliedstaaten nicht höher sind als die Preisunterschiede zwischen einzelnen Regionen innerhalb der EU-Mitgliedstaaten.<sup>21</sup> Daher ist es nicht verwunderlich, dass trotz des Harmonisierungsprozesses und weiteren Anstrengungen innerhalb der EU wie auch in anderen voll integrierten Märkten (z. B. Bundesstaaten in den USA) Unterschiede zwischen den einzelnen Ländern zu beobachten sind.

### 3. Empirische Analyse

Vor dem skizzierten Hintergrund der europäischen Harmonisierung wurde im Jahr 2004 eine empirische Untersuchung<sup>22</sup> durchgeführt, im Rahmen derer die regionalen bzw. länderspezifischen Unterschiede in den operativen Margen unabhängiger Vergleichsunternehmen im Kontext steuerlicher Datenbankstudien analysiert wurden. Die Studie enthält unterstützende Argumente für die Nutzung pan-europäischer Datensets, damit auch die anderen Vergleichbarkeitsfaktoren (z. B. Funktionen oder Branche) hinreichend berücksichtigt werden können. Zwar kommt die Studie zu dem Ergebnis, dass in bestimmten Fällen systematische Unterschiede zwischen den Industrien für einzelne Länder bestehen können. Allerdings vermag dies nicht die Möglichkeit einer besseren Berücksichtigung anderer Vergleichbarkeitsfaktoren sowie die statistische Vorteilhaftigkeit einer höheren Grundgesamtheit im Rahmen einer pan-europäischen Datenbankstudie aufzuwiegen. Die Ergebnisse dieser Studie haben Eingang in ein von der Generaldirektion für Steuern und Zollunion der Europäischen Kommission veröffentlichtes „Whitepaper“ gefunden, welches als Referenz für die Verwendung pan-europäischer Datenbankstudien gilt.

In der nachfolgenden ökonomischen Analyse werden die Ergebnisse der Studie aus dem Jahr 2004 auf der Grundlage aktueller Daten validiert. Hierbei werden die länderspezifischen Unterschiede methodisch mithilfe komplementärer statistischer Verfahren untersucht. Wie in der Einleitung motiviert, ist dies für Unternehmen von besonderem Interesse, da pan-europäische Benchmarkstudien häufig einen wichtigen Baustein bei der Planung und Verprobung steuerlicher Verrechnungspreise darstellen.

21 Vgl. etwa AC Nielsen, A report on International Price Comparisons, 2000.

22 Vgl. EU Joint Transfer Pricing Forum, 2004, Is Europe One Market? A Transfer Pricing Economic Analysis of Pan European Comparables Sets, Deloitte White Paper. Siehe ebenso Appendix 12 in „A Toolkit for Addressing Difficulties in Accessing Comparables Data for Transfer Pricing Analyses, OECD und International Monetary Fund, United Nations and World Bank Group für eine direkte Replikation der Studie aus dem Jahr 2004.

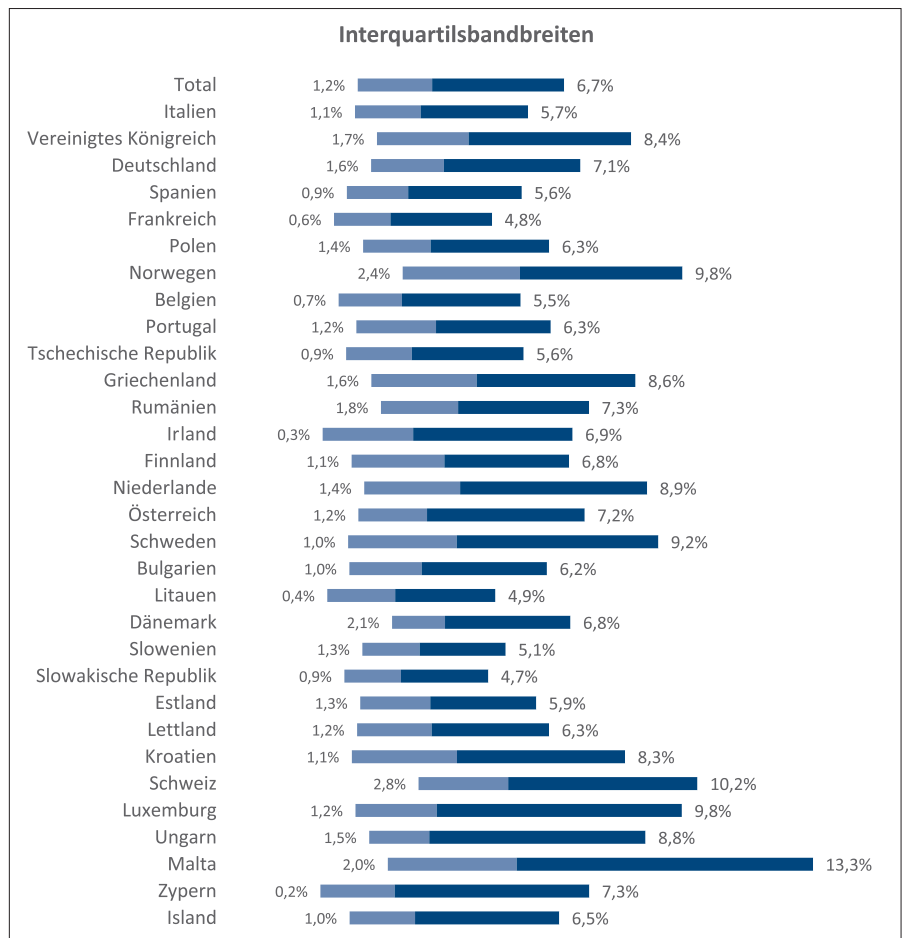
### 3.1. Datensatz – Beschreibung

Die nachfolgende Analyse wurde auf Basis der AMADEUS Neo Datenbank<sup>23</sup> der Bureau van Dijk Electronic Publishing GmbH erstellt. Der Datensatz umfasst bis zu 439.348 unabhängige<sup>24</sup> Unternehmen, die in den EU-28 Ländern<sup>25</sup> sowie Island, Norwegen und der Schweiz tätig sind, verschiedenen NACE Codes<sup>26</sup> zugeordnet werden können, als auch ausgewählte Finanzkennzahlen dieser Unternehmen für die Jahre 2012 bis 2014.

Nach der Anwendung eines Umsatzfilters (Unternehmen mit Umsatz größer oder gleich 10 Millionen EUR) sowie dem Ausschluss von extrem hohen und niedrigen operativen Gewinnen und Umsätzen infolge von Fehleinträgen in der Datenbank (obere und untere 1%), jeweils im Mittelwert über die drei Jahre, verbleiben noch 35.288 Unternehmen in unserem Datensatz.

Um den Anforderungen der Finanzbehörden an Datenbankstudien zu genügen, wurden Unternehmen mit hohem Anteil<sup>27</sup> an immateriellen Vermögensgegenständen als Surrogat für Enterpreneure ausgeschlossen. Zusätzlich wurden Unternehmen mit verfügbaren Finanzdaten für weniger als drei Jahre mit dem Ziel ausgeschlossen, lediglich aktive Unternehmen im Datenset zu berücksichtigen. Nach diesem zusätzlichen Filterschritt verbleiben 29.442 Unternehmen.

Die hohe Unternehmensanzahl des Datensatzes ermöglicht eine Indikation über Funktionen (NACE Codes) und Länder für einen Zeitraum von drei Jahren, welche als breit angelegte Querschnittsübersicht robust gegenüber dem Ausschluss einzelner Unternehmen ist.



Grafik 1: Interquartilsbandbreiten der operativen EBIT Mergen

Anmerkung: Sortierung in absteigender Unternehmensanzahl, äußere linke Zahl = 1. Quartil, äußere rechte Zahl = 3. Quartil pro Land.

### 3.2. Länderübersicht über die operativen EBIT Mergen

Grafik 1 stellt für die einzelnen Länder die Interquartilsbandbreiten der realisierten durchschnittlichen<sup>28</sup> EBIT Mergen für den Zeitraum von 2012 bis 2014 dar.<sup>29</sup>

Über alle Länder hinweg ergibt sich eine totale Interquartilsbandbreite von 1,2 % bis 6,7 % mit einem Median von 3,2%. Schon die Interquartilsbandbreite von über 5 Prozentpunkten über alle Länder hinweg ist recht breit. Allerdings sind auf rein deskriptiver Ebene noch weit breitere Interquartilsbandbreiten für kleinere Länder wie beispielsweise Malta oder auch Luxemburg zu beobachten. Südliche EU Länder wie Spanien, Frankreich, aber auch Italien hingegen weisen tendenziell geringere Werte auf, beispielsweise liegt das erste Quartil in diesen Ländern unterhalb des ersten Quartils der totalen Interquartilsbandbreite. Die zugehörigen Interquartilsbandbreiten sind dabei nicht breiter als die auf totaler Ebene.

23 AMADEUS Neo Datenbank in der Version 13.06, Update 260.

24 Unternehmen mit den folgenden Bureau van Dijk Unabhängigkeitsindikatoren wurden ausgeschlossen: C+, C, C-, D+, D, D-, U.

25 Die folgenden Länder der „EU 28“ werden berücksichtigt: Belgien, Bulgarien, Dänemark, Deutschland, Estland, Finnland, Frankreich, Griechenland, Irland, Italien, Kroatien, Lettland, Litauen, Luxemburg, Malta, Niederlande, Österreich, Polen, Portugal, Rumänien, Schweden, Slowakei, Slowenien, Spanien, Tschechische Republik, Ungarn, das Vereinigte Königreich und Zypern.

26 Die enthaltenen Wirtschaftszweige nach der NACE Code Klassifikation (Rev. 2) sind für die Distributionsfunktion 4511 bis 4540, 4600 bis 4799, für die Dienstleistungsfunktion 1800, 4900 bis 7120, 7410 bis 8299, 7311,7312 und für die Produktionsfunktion 0111 bis 0150, 0170 bis 0899, 1010 bis 3299, 4110 bis 4399.

27 Verhältnis von immateriellen Vermögensgegenständen zu Umsatz ist größer 5%.

28 Gewichteter Durchschnitt mittels Umsatz in den einzelnen Jahren.

29 Die operative EBIT Marge muss nicht für alle Unternehmen bzw. ausgeübte Funktionen ein geeigneter Vergleichsmaßstab sein. Da sich daraus aber direkt die Vollkostenaufschläge ableiten lassen, erachten wir für Vergleichszwecke die EBIT Marge als sinnvollste Finanzkennzahl.

Eine systematische und vor allem statistisch fundierte Analyse lässt sich jedoch aus einem rein deskriptiven Vergleich der in Grafik 1 dargestellten Interquartilsbandbreite der EBIT Margen explizit nicht ableiten. Die Frage, ob sich bestimmte Länder in einem statistischen Sinne von dem totalen Set über alle Länder hinweg unterscheiden, soll vielmehr mittels statistischer Tests untersucht werden.

Die Analyse unterscheidet hierbei zwischen den drei betriebswirtschaftlichen Hauptfunktionen Produktion, Vertrieb und Dienstleistungserbringung. Spezifische Vergleichbarkeitsfaktoren, die im konkreten Fall im Vergleich zur Tested Party ebenso zu berücksichtigen sind, werden vorliegend explizit nicht betrachtet.

### 3.3. Ländervergleich der EBIT Margen

Als statistischer Test zur Überprüfung, ob systematische Unterschiede in den EBIT Margen Verteilungen über die Länder hinweg vorliegen, wird im Folgenden sowohl der *Kolmogorov-Smirnov*-Test<sup>30</sup> als auch der *Wilcoxon-Mann-Whitney*-Test<sup>31</sup> angewendet. Als nicht-parametrische Tests benötigen diese beiden Testverfahren keine kritische Annahme über eine Normalverteilung der EBIT Margen, was üblicherweise bei EBIT Margen nicht erfüllt ist. Ebenso sind beide Tests auch für kleine Stichproben geeignet und setzen nicht gleiche Fallzahlen voraus. Somit können auch kleinere Länder mit größeren Ländern bzw. dem gesamten Datensatz sinnvoll verglichen werden.

Im Folgenden soll durch beide Tests überprüft werden, ob die EBIT Margen in den unterschiedlichen Ländern eine gleiche Verteilung besitzen (Null-Hypothese), d. h. uneingeschränkt für pan-europäische Benchmarkstudien genutzt werden können. Falls die Null-Hypothese verneint (bzw. verworfen) wird, soll zusätzlich mittels stochastischer Dominanz überprüft werden, ob die länderspezifische Verteilung der EBIT Margen größer (gekennzeichnet durch ein Pluszeichen) oder kleiner (Minuszeichen) ist als die Verteilung der EBIT Margen des restlichen Datensatzes ist.

Tabelle 1 fasst die Ergebnisse bei einem Signifikanzniveau von 1 % zusammen (Spalte „Total“). Tendenziell höhere bzw. niedrigere EBIT Margen werden nur dann dargestellt, wenn sowohl der *Kolmogorov-Smirnov*-Test als auch der *Wilcoxon-Mann-Whitney*-Test zum gleichen Ergebnis, d. h. zur Verwerfung der Null-Hypothese, führen. Somit sind die Ergebnisse bzgl. Abweichler als konservativ zu interpretieren.

Da im spezifischen Fall einer konkreten Benchmarkstudie weitere Vergleichbarkeitsfaktoren möglicherweise zu berücksichtigen sind, bedeutet ein Plus- oder Minuszeichen explizit nicht, dass diese Länder aus einer konkreten Benchmarkstudie ausgeschlossen werden müssen. Vielmehr sollen die Plus- und Minuszeichen als Indikation für die relative EBIT Margen Verteilung der einzelnen Länder verstanden werden und kön-

nen daher als weiterer Robustness Check oder mögliche Anpassungsrechnungen verstanden werden.

Es zeigt sich, dass für die Mehrzahl der Länder (17 von 31), eine gleiche Verteilung der operativen EBIT Margen auf Basis des *Kolmogorov-Smirnov*-Tests und des *Wilcoxon-Mann-Whitney*-Tests festzustellen ist. Lediglich für einzelne Länder sind tendenziell höhere EBIT Margen (7 von 31) bzw. niedrigere Margen (7 von 31) festzustellen. Die Anzahl der „Abweichler“ nach oben als auch nach unten ist gleich, so dass sich auch die unterschiedlichen Verteilungen approximativ ausgleichen.

### 3.4. Analyse auf Funktionsebene

Analog zur vorangestellten Analyse über Funktionsunterschiede hinweg zeigt Tabelle 1 ebenso für die drei betriebswirtschaftlichen Hauptfunktionen Produktion, Vertrieb und Dienstleistungserbringung jeweils die Länderunterschiede auf Basis des *Kolmogorov-Smirnov*-Tests und des *Wilcoxon-Mann-Whitney*-Tests.

Es zeigt sich, dass auch nach der Separierung nach Funktionen zwar weiterhin Länderunterschiede in den EBIT Margen zu beobachten sind (d. h. beide Tests verwerfen die Null-Hypothese gleicher Verteilungen), allerdings geht die Zahl der Abweichler stark zurück (von 14 auf 10 für Vertrieb und jeweils von 14 auf 6 für Produktion und Dienstleistungen). Somit kann festgestellt werden, dass nach der Berücksichtigung von Funktionsunterschieden die Vergleichbarkeit über die Ländergrenzen hinweg zunimmt und somit ebenso die Gültigkeit der Anwendung von pan-europäischer Benchmarkstudien.

Zwar bestehen auch nach der Berücksichtigung von betriebswirtschaftlichen Funktionen weiterhin Länderunterschiede, diese können aber durch die Berücksichtigung von weiteren Vergleichbarkeitsfaktoren, insbesondere durch die Berücksichtigung von Branchen und Umsatzunterschieden,<sup>32</sup> abgemildert werden. In der Theorie würden so alle Vergleichbarkeitsfaktoren vollständig Berücksichtigung finden, allerdings können in der Praxis auf Grund der Datenverfügbarkeit nicht immer alle definierten Vergleichbarkeitsfaktoren gleichermaßen Anwendung finden. Vereinzelt Länderunterschiede sind dann zu akzeptieren, wenn durch die Datenverfügbarkeit weitere Vergleichbarkeitsfaktoren berücksichtigt werden können, die insgesamt zu einer höheren Verlässlichkeit der Vergleichsdaten führen.

Für den dieser Analyse zugrunde liegenden Datensatz zeigt Tabelle 1, dass sich tendenziell höhere EBIT Margen beispielsweise im Vereinigten Königreich aber auch in Norwegen über alle Funktionen hinweg konsistent beobachten lassen, wohingegen im Rahmen einer Gesamtbetrachtung tendenziell niedrigere EBIT Margen stark durch einzelne Funktionen getrieben werden. So zeigt sich beispielsweise eine niedrige EBIT Marge in Belgien, Spanien und für die Tschechische Republik

30 Vgl. Michiel Hazewinkel, „Kolmogorov-Smirnov Test“, Encyclopedia of Mathematics, Springer, 2001, ISBN 978-1-55608-010-4.

31 Vgl. Henry Mann und Donald Whitney, 1947, „On a test of whether one of two random variables is stochastically larger than the other“, Annals of mathematical Statistics 18, S. 50 bis 60.

32 Siehe für die Berücksichtigung von Umsatzgrößen und Skaleneffekten bei Produktionsunternehmen: Schwarz/Holinski/Stein/Hoffmann, DB 2016, S. 2307 bis 2311; und bei Vertriebsunternehmen: Krüger/Nientimp/Schwarz, Ubg 2016, S. 221 bis 225.

nur für die Funktion Vertrieb, nicht aber für die Funktionen Produktion und Dienstleistungserbringung.

Land	Total	Vertrieb	Dienstleistung	Produktion
Italien	-	-	-	
Vereinigtes Königreich	+	+	+	+
Deutschland	+	+		
Spanien	-	-		
Frankreich	-	-	-	-
Polen				
Norwegen	+	+	+	+
Belgien	-	-		-
Portugal				
Tschechische Republik	-	-		
Griechenland	+	+		+
Rumänien	+	+		
Irland			-	
Finnland				
Niederlande				
Österreich				
Schweden				
Bulgarien				
Litauen	-			
Dänemark	+			
Slowenien				
Slowakische Republik	-		-	
Estland				
Lettland				
Kroatien				
Schweiz	+			
Luxemburg				
Ungarn				
Malta				
Zypern				
Island				+

Tabelle 1: Ländervergleich der EBIT Margen

#### 4. Zusammenfassung

Ob die Vergleichsunternehmen in einer pan-europäischen Benchmarkstudie auch hinsichtlich der länderspezifischen Charakteristika hinreichend vergleichbar sind, ist regelmäßig Streitpunkt in steuerlichen Betriebsprüfungen. Seit im Jahr 2004 in einem von der Generaldirektion für Steuern und Zollunion der Europäischen Kommission veröffentlichten „Whitepaper“ eine hinreichende Vergleichbarkeit von Unternehmen in pan-europäischen Benchmarkstudien festgestellt wurde, gilt diese Veröffentlichung als die Referenz für die Verwendung und Gültigkeit pan-europäischer Datenbankstudien.

Der vorliegende Aufsatz bestätigt mittels komplementärer statistischer Methoden und aktuellen Daten die aktuelle Gültigkeit dieser Studie und dient somit als wichtiger Baustein bei der Planung und Verprobung steuerlicher Verrechnungspreise mittels pan-europäischer Benchmarkstudien.