

It's Time to Reform Transfer Pricing Benchmarking

by Michael C. Durst

Reprinted from *Tax Notes International*, June 6, 2022, p. 1271

It's Time to Reform Transfer Pricing Benchmarking

by Michael C. Durst



Michael C. Durst

Michael C. Durst is a senior fellow of the International Centre for Tax and Development (ICTD) in Brighton, U.K., and a senior adviser to the International Lawyers Project, based in London. He is a tax lawyer and tax commentator in Washington, has taught at several law schools, and from 1994 to 1997

directed the IRS advance pricing agreement program. He is grateful for the generous and helpful comments of Joseph Andrus, Reuven Avi-Yonah, John Bush, Andrew Hughes, Sol Picciotto, and Scott Wilkie.

In this article, Durst considers how the transfer pricing benchmarking of taxpayers' profitability under the transactional net margin method and the comparable profits method has damaged international tax administration.

Any shortcomings of the article are entirely the responsibility of the author, and any views stated in the article are solely the author's.

Copyright 2022 ICTD.
All rights reserved.

The malfunction of commonly used transfer pricing methods has been central to the problem of base erosion and profit shifting that has motivated the ongoing tax reform efforts of the OECD and its inclusive framework. Nevertheless, the pending pillar 1 and 2 proposals would effect little change in the OECD's transfer pricing guidelines.¹

¹ See Joseph L. Andrus and Richard S. Collier, "Transfer Pricing and the Arm's-Length Principle After the Pillars," *Tax Notes Int'l*, Jan. 31, 2022, p. 543 (expressing surprise at extent to which the international tax system under pillars 1 and 2 would continue to rely on current, manifestly problematic, transfer pricing methods).

This article explores the damage that has been inflicted on international tax administration by one widely used category of transfer pricing methods, the benchmarking of taxpayers' profitability under the transactional net margin method (TNMM) of the OECD transfer pricing guidelines and the similar comparable profits method of the U.S. transfer pricing regulations.² Unless corrective measures are taken soon by the OECD and the inclusive framework, even after pillars 1 and 2 have been implemented the defective elements of these transfer pricing methods will impair the administration of corporate tax in countries at all levels of economic development.

This article seeks to explain the problems caused by benchmarking methods and to suggest relatively simple, path-dependent measures that might be taken now to improve the effectiveness of benchmarking. In particular, the article:

- describes how intractable defects in benchmarking methods, especially the problem of wide "arm's-length ranges," have for many years facilitated BEPS and threaten to do so even after implementation of pillars 1 and 2;
- offers evidence of grossly wide arm's-length ranges based on research recently performed by Andrew Hughes and published by *Tax Notes*; and
- suggests modifications to benchmarking methods that would remain consistent with the arm's-length principle and with the overall framework of current practices but would determine arm's-length levels of income with much more certainty.

² Rules for the TNMM are prescribed in chapter 2, Part III of the OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations; rules for the CPM in the United States are provided at reg. section 1.482-5.

I. Benchmarking, BEPS, and GLOBE Income

Deficiencies in transfer pricing rules have been central to the flowering in recent decades of BEPS around the world. These structures typically are built upon the treatment of an in-country affiliate of a multinational — for example, a manufacturing or service-provider subsidiary — as a limited-risk entity, which in return for the limited business risks that it bears is entitled to only low, routine levels of income.³ To achieve this, much of the income of the limited-risk entity is paid out during the year to foreign affiliates in the form of management fees, royalties, and other deductible payments. The amount of income left in the limited-risk subsidiary is typically quite low.

Under existing transfer pricing practices, companies and their advisers are supposed to test whether the income left in the subsidiary is adequate through benchmarking based on comparables. A transfer pricing expert tries to locate, from publicly available financial databases, financial information from independent businesses with operations allegedly similar to those of the limited-risk subsidiary being tested. Ideally, the comparables selected from the databases should operate in the same geographic area as the tested party. Once the comparables data have been collected, the arm's-length range of acceptable income results for the tested party is determined, typically based on the interquartile range (the range of results extending from the 25th to the 75th percentile) of the comparables' observed income levels.⁴

This facially elegant means of benchmarking a tested party's arm's-length income rarely, if ever, provides useful information to tax inspectors. The problem is that even in the most economically developed regions of the world, it is difficult to find, in publicly available financial databases, more than a few companies that are plausibly similar in function to the tested party. In practice, transfer pricing benchmarking virtually always relies on too few data points to support a statistically reliable determination of an arm's-length range. The problem is aggravated by the tendency of financial data of different companies, even operating in the same industry, to demonstrate a relatively low degree of statistical convergence.⁵ Since the inception of today's benchmarking practices, the result of the small sizes of available data sets, and relatively nonconvergent financial data, has been arm's-length ranges typically far too wide to be of much use in tax enforcement.

For example, as shown below, it is not unusual for an arm's-length range of, say, net operating margins to extend from a lower-quartile level of 2 percent to an upper-quartile level of 5 percent. According to this range, for a taxpayer with \$200 million of revenue, any level of income between \$4 million and \$10 million should be treated as consistent with the arm's-length principle. A range of this width offers little confidence to the tax authority that it is enforcing a realistic measure of the taxpayer's arm's-length income.⁶ Inevitably, tax inspectors find themselves compelled to allow income levels for local subsidiaries that are far below economic plausibility. The phenomenon of very wide transfer pricing ranges has been essential in the dramatic growth of BEPS over the years.

³The kinds of BEPS structures that are common around the world have been described at many places in the tax literature. See, e.g., Michael C. Durst, *Taxing Multinational Business in Lower-Income Countries: Economics, Politics and Social Responsibility* 31-55 (2019), and authorities cited therein.

⁴The tax professional who performs a benchmarking study must select a profit-level indicator (sometimes called a net profit indicator) against which to perform the benchmarking. Commonly used profit-level indicators include the net operating margin (a company's ratio of net profit to sales), the markup on total costs, the Berry ratio (the ratio of gross profit to operating expense), and the return on assets. The choice of a profit-level indicator for a particular benchmarking exercise is a matter for the tax professional's judgment, with the return on assets often used for capital-intensive businesses like manufacturers, and the other profit-level indicators typically used for distribution and service businesses.

⁵For discussion of the lack of convergence among companies' income data, see John Wills, "Risk Measurement: Applying Financial Theory to Transfer Pricing," *Tax Notes*, Sept. 9, 1991, p. 1311.

⁶OECD guidelines and U.S. regulations raise the possibility that a tax authority might insist on an arm's-length range that differs from the interquartile range. A determination of this kind is highly subjective, however, and very difficult for a tax inspector to make with any degree of persuasiveness. See also note 20, *infra*, describing India's practice of using a somewhat narrower range.

Although the prevalence of excessively wide ranges has been known and discussed among practitioners for many years,⁷ the problem has not been faced squarely by policymakers. This may partly reflect reluctance to confront the inconvenient implications of a conclusion that frequently used transfer pricing methods are not functioning as originally intended. Also, the most obvious way of empirically testing the efficacy of benchmarking would entail analysis of taxpayers' transfer pricing documentation in government files, potentially raising confidentiality concerns over taxpayer information.

Andrew Hughes, a Brussels-based transfer pricing practitioner, has recently published a large volume of simulated transfer pricing benchmarking studies, relying on publicly available financial data.⁸ The simulated studies rely on enough data to provide statistically useful insights. Some methodological caveats should be recognized: Hughes's articles are intended in part to show the feasibility of simplifying current benchmarking practices, and his method differs in some respects from that typically used in practice. Hughes does not use commercially maintained financial databases for his benchmarking, but instead uses individual company filings downloaded directly from the internet. Also, Hughes's use of relatively broadly defined taxpayer business categories (like IT services) may lead to greater willingness to accept some comparables than would be the case if a specific, identified taxpayer were the subject of the benchmarking. Because of these differences, studies based on anonymized taxpayer data would be more definitive concerning typical range widths under TNMM than Hughes's measurements.

Hughes's studies, nevertheless, generally conform closely to benchmarking in actual practice, and the studies strongly suggest that current methods are not offering income measurements that are of practical use in tax

administration. It is hoped that revenue authorities, in consultation with the OECD, will conduct more definitive studies using anonymized data from taxpayers' actual transfer pricing documentation, large volumes of which should be available in tax administration files. The performance and dissemination of this kind of analysis is essential for adequate effectiveness and transparency in international tax policymaking.

II. Empirical Evidence

For his source of comparables, Hughes uses financial data from North American companies obtained from the internet. I have reviewed the results, including the interquartile ranges, that are presented in 11 of Hughes's articles.⁹ The results of my review are contained in the table and the accompanying scatter plot in the appendix to this article.¹⁰

The interquartile ranges computed in virtually all of Hughes's studies are far too wide to be useful in tax administration. The question of when an arm's-length range is too wide for practical use is, concededly, somewhat subjective. For example, most tax examiners probably would consider an arm's-length range extending 5 percent above the bottom of the range to offer

⁹ The articles reviewed include Andrew Hughes, "Transfer Pricing Benchmark: North American IT Services," *Tax Notes Int'l*, Dec. 7, 2020, p. 1315, and 10 subsequent articles covering other North American industry groupings including strategic consulting services, *Tax Notes Int'l*, Jan. 18, 2021, p. 299; engineering services, *Tax Notes Int'l*, Feb. 8, 2021, p. 757; food distribution, *Tax Notes Int'l*, Mar. 1, 2021, p. 1141; electronic manufacturing services, *Tax Notes Int'l*, Apr. 5, 2021, p. 53; healthcare distribution, *Tax Notes Int'l*, May 3, 2021, p. 623; IT and electrical components distribution, *Tax Notes Int'l*, June 21, 2021, p. 1671; construction distribution, *Tax Notes Int'l*, July 5, 2021, p. 51; contract research services, *Tax Notes Int'l*, Aug. 9, 2021, p. 715; metal manufacturing services, *Tax Notes Int'l*, Sept. 13, 2021, p. 1485; and HR services, *Tax Notes Int'l*, Oct. 18, 2021, p. 317.

¹⁰ The 11 articles by Hughes each contain three years of data for each industry group studied; I have employed three-year averages of the data points to make the analysis more tractable. Hughes computes three different profit-level indicators for each industry group: (1) either operating margin or markup on total costs; (2) the Berry ratio (ratio of gross profit to operating income); and (3) return on assets. The method I used to determine the ratio of upper to lower quartile values varied to some extent to accommodate differences in how the profit-level indicators are defined. For the operating margin, markup on total costs, and return on assets, I divided the upper-quartile value over the lower-quartile value; but for the Berry ratio, I divided the upper-quartile number to the right of the decimal point by the lower-quartile number to the right of the decimal point. For example, when the upper-quartile observation for the Berry ratio was 1.78 and the lower-quartile observation was 1.39, I divided 78 by 39 to compute a ratio of 2.

⁷ See Durst and Robert E. Culbertson, "Clearing Away the Sand: Retrospective Methods and Prospective Documentation in Transfer Pricing Today," *57 Tax L. Rev.* 37, 112-113 (2003).

⁸ See *infra* note 9.

reasonable precision for use in tax enforcement. The same is probably true of ranges extending 10, or even 15 percent, above the lower bound.

Views might begin to differ when the range extends to as much as 25 percent above the bottom: Many might consider a transfer pricing method that pins down a taxpayer's arm's-length income within the range of, say, \$10 million to \$12.5 million as being too wide to provide reasonable certainty of the taxpayer's correct level of income. But the table and chart in the appendix show that none of the arm's-length ranges computed by Hughes extend above the bottom by less than 125 percent. Indeed, only a few ranges extend above the lower quartile by less than 150 percent. Most of the ranges extend by at least 200 percent from the bottom, and a significant number extend above the bottom by more than 300 percent. A 200 percent range means, for example, that a particular taxpayer's income might fall anywhere between \$10 million and \$20 million; a 300 percent range, of course, leads to even greater uncertainty. A transfer pricing method yielding these kinds of results is not suited for effective use in tax administration; the method leaves open the possibility that taxpayers are paying far less than an appropriate level of tax.

III. Benchmarking and the Pillars

As envisioned, pillar 2 would establish a minimum tax on members of multinational enterprises, on a country-by-country basis, at a rate of 15 percent of the MNE member's GLOBE income.¹¹ The starting point in computing a local entity's GLOBE income is its MNE group's consolidated financial income, which must be disaggregated to determine the specific group member's local income. Because each member's local income is likely to involve a wide range of transactions with other group members, a final step in the determination of local GLOBE income

requires the use of transfer pricing rules, presumably based on the OECD's existing transfer pricing guidelines.¹² Thus, it appears that today's widely used benchmarking techniques will be used in evaluating the arm's-length level of a local taxpayer's GLOBE income, as well as the taxpayer's regular taxable income. This would seem to inject a high level of uncertainty into taxpayers' reporting of their GLOBE income, yielding a high incidence of transfer pricing controversy between taxpayers and tax authorities, and between different countries' tax authorities seeking to avoid double taxation.

It might be argued that inaccurate transfer pricing in determining GLOBE income would pose relatively little practical difficulty, because the GLOBE minimum tax is meant to be an internationally seamless device under which income shifted out of one country's GLOBE income will inevitably be shifted to another country's GLOBE income and subjected to at least a 15 percent tax. But that neat result cannot be assured. Not every country is likely to participate in the GLOBE network, and even among countries that do, assuring seamless coverage of the minimum tax will require expensive and uncertain double-tax negotiations. It will be far better if each country's GLOBE income is reported reasonably accurately in the first instance. Importing today's transfer pricing rules into the GLOBE arena would, without significant improvement, appear to be a serious and costly mistake.

The OECD offers several explanations for its importation of existing transfer pricing practices into the determination of GLOBE income.¹³ First, it suggests that nontax regulatory pressure will encourage entities to use, for their book transfer prices, amounts that satisfy the arm's-length

¹² OECD, "Tax Challenges Arising From the Digitalisation of the Economy — Global Anti-Base Erosion Model Rules (Pillar Two)," at article 3.2.3 (Dec. 20, 2021).

¹³ *Id.* at 61: Constituent Entities of an MNE Group typically maintain a transfer pricing policy based on the Arm's-length Principle and this standard is used to determine the transfer price that is reflected in their financial accounts and in computing the local taxable income. Therefore, it is generally expected that Constituent Entities' financial accounts will reflect transactions between Group Entities based on the Arm's-length Principle and at the same price.

¹¹ GLOBE stands for global anti-base-erosion proposal and is another name for the pillar 2 plan. The minimum tax might be imposed in the jurisdiction of the local constituent entity, under the income inclusion rule of pillar 2, or it might be imposed in the jurisdiction of the constituent entity if that jurisdiction has enacted a qualified domestic minimum top-up tax.

principle. However, the transfer prices that entities might use for book purposes typically have no legal significance outside the field of taxation, so no local regulatory body is likely to be motivated to police these prices.

The OECD also appears to assume that business factors — the desire to maintain arm’s-length internal transfer pricing for business purposes — will encourage companies to use arm’s-length pricing for book purposes.¹⁴ Business groups, however, for management accounting purposes tend to divide their operations into cost centers and profit centers.¹⁵ Goods and services sold in intragroup transactions typically are priced at cost for management purposes, not at an arm’s-length level. In short, for management and book purposes, as for tax purposes, there is no reason to expect a multinational group to naturally set its internal transfer prices at an arm’s-length level. If arm’s-length pricing is desired, it must be enforced by the tax authorities, and the tax transfer pricing rules need to make available statistically sound tools for enforcement.

It might also be argued that the other component of the pending reform efforts, pillar 1, by removing some income from the coverage of the arm’s-length principle, will mitigate the damage from defective benchmarking methods. Even if countries ultimately implement it, however, pillar 1 is designed to apply to only a portion of the income generated in only some intragroup transactions. Arm’s-length transfer pricing methods will remain central to global tax administration, and the task of repairing serious defects in those methods remains important.

An additional note should be made here about how transfer pricing benchmarking affects developing countries. It has been suggested over the years that the TNMM is mainly problematic for smaller, developing economies with access to relatively limited local comparables data and little

experience working with financial databases. Therefore, it has been hoped that the problem can be redressed through capacity building, and through the acquisition by developing country tax administrations of additional commercial financial databases.¹⁶

Hopes for marked improvement from capacity-building and additional databases may not, however, be realistic. The Hughes studies, which are based on North American data, suggest that transfer pricing benchmarking in even the wealthiest countries has failed to solve the twin problems of small numbers of useful comparables and insufficient statistical convergence among different comparables’ financial data. It seems unlikely that these problems can be fixed for developing countries by only capacity building. More workable transfer pricing methods are essential as well. The OECD and inclusive framework should recommend improved methods in connection with the current reform efforts.

IV. Repairing Transfer Pricing Benchmarking

One possibly helpful approach would be to adopt a simplified benchmarking method that does not rely on searches for case-specific comparables. This approach seems currently to be envisioned for the determination of a company’s appropriate routine marketing income under amount B of pillar 1. Under the amount B method, the tax authority would prescribe minimum arm’s-length profit levels for different categories of activity based on a review of comparables data from databases. The use of inevitably broad categories of activity would lead to some degree of approximation but would eliminate the uncertainties and administrative difficulties of excessively wide arm’s-length ranges.

The envisioned amount B approach represents a desirable balance between achieving greater certainty for taxpayers and tax administrations, without abandoning the concept

¹⁴ *Id.*

¹⁵ For classic expositions of this point, see Jack Hirshleifer, “On the Economics of Transfer Pricing,” 29 *J. Bus.* 172 (1956); and Hirshleifer, “Economics of the Divisionalized Firm,” 30 *J. Bus.* 96 (1957). See also Durst, “Management vs. Tax Accounting in Intercompany Transfer Pricing,” *Tax Mgmt. Int’l J.* 95 (Feb. 8, 2002), and other authorities cited therein.

¹⁶ See Platform for Collaboration on Tax, “A Toolkit for Addressing Difficulty in Accessing Comparables Data for Transfer Pricing Analysis” (2017).

of arm's-length benchmarking, on which tax practice has now relied for a considerable time.¹⁷ As of this writing, however, there remains within the OECD's inclusive framework political disagreement over the degree of approximation that should be allowed under the amount B method.¹⁸ Given political perceptions, it is not clear that the proposed amount B approach to benchmarking ultimately will be accepted even for pillar 1, let alone for wider application under the OECD guidelines.

Several additional solutions to the benchmarking problem might succeed in securing reasonable approximations of arm's-length results. First, a benchmarking system might be based not on comparables searches, but instead on a multinational group's global level of profitability. For example, a limited-risk subsidiary might be expected to earn, say, at least 30 percent of the group's global operating margin. Transfer pricing methods based on this concept were proposed earlier in the consideration of amount B by the multinational groups Johnson & Johnson Services Inc. and Procter & Gamble Co.,¹⁹ but the proposals do not seem to have made much headway in subsequent negotiations. I agree that a transfer pricing method based not on comparables, but instead on groupwide profitability, offers promise for international tax administration. But I fear that this approach is unlikely to attract much political support currently.

Still another possibility would be to leave the practice of comparables-based benchmarking as it is but prescribe an arm's-length range that is substantially narrower than the interquartile range — say, a range of plus-or-minus 10 percent around a median value.²⁰ This should greatly reduce the danger of the unrealistically low income levels that have historically facilitated BEPS. The approach of narrowing the range would have the benefit of path-dependency: It would leave in place much of the infrastructure of transfer pricing benchmarking, in both the private and public sectors, that has grown over the past quarter-century.

None of these possible improvements to OECD benchmarking methods is ideal. All would involve a degree of approximation of arm's-length pricing that might strike some as insufficiently scientific. Against this, however, today's benchmarking rules, while perhaps facially scientific, have led to systematic understatement of subsidiary income levels and have fueled the BEPS problem that the OECD and the inclusive framework are now trying to solve. Some reform of benchmarking practices will be essential if a reformed system of international taxation is to be administrable.

¹⁷ See generally Durst, "Simplifying Developing Country Multinational Taxation: The 'Amount B' Proposal and the Repair of Profit Shifting," *Tax Notes Int'l*, Aug. 17, 2020, p. 917.

¹⁸ Apparently, existing comparables-based benchmarking has become so embedded in patterns of tax planning (including BEPS planning) and tax practice that even minor departures from it can be reflexively and strenuously opposed. In connection with the debate over amount B, OECD Tax Chief Pascal Saint-Amans recently said:

This [approval of the Amount B proposal] should be low-hanging fruit. It is not. It is not, because you have a number of people who still pray to the arm's-length principle god, and they think that if we slightly depart from it, the world will collapse.

Stephanie Soong Johnston, "Businesses Must Speak Up on Amount B, OECD Tax Chief Says," *Tax Notes Int'l*, Mar. 28, 2022, p. 1576.

¹⁹ See generally Durst, *supra* note 17.

²⁰ India generally employs an arm's-length range extending from the 35th to the 65th percentile, instead of the interquartile range. OECD, "Transfer Pricing Profile — India" (July 2021); Rule 10CA of the Indian Income-tax Rules. Although it is not clear whether this modification is enough to change results substantially in practice, the Indian experience — and that of any other country that has departed from the interquartile range — should be studied carefully for any lessons that might be provided.

Appendix
Data From Hughes's Studies

	Markup on Total Costs (or operating margin)	Width of Range	Berry Ratio	Width of Range	Return on Assets	Width of Range
IT Services						
Upper Quartile	16.77		1.75		12.93	
Lower Quartile	4.93		1.21		5.2	
		3.4		3.52		2.49
Strategic Consulting Services						
Upper Quartile	11.2		1.38		14.27	
Lower Quartile	6.4		1.17		5.73	
		1.75		2.24		2.49
Engineering Services						
Upper Quartile	6.57		1.72		10.7	
Lower Quartile	2.73		1.24		4.3	
		2.4		3		2.49
Food Distribution	Operating Margin					
Upper Quartile	3		1.18		10.43	
Lower Quartile	1.03		1.1		5.07	
		2.9		1.8		2.06
Electronic Manufacturing Services						
Upper Quartile	4.17		1.8		6.27	
Lower Quartile	1.67		1.24		2.5	
		2.5		3.33		2.51
Healthcare Distribution	Operating Margin					
Upper Quartile	2.43		1.36		5.4	
Lower Quartile	1.23		1.19		4.17	
		1.98		1.89		1.29
IT and Electrical Components Distribution	Operating Margin					
Upper Quartile	5.23		1.39		11.87	
Lower Quartile	3.13		1.27		6.23	
		1.67		1.44		1.91
Construction Distribution (excluding maintenance, repair, and operating)	Operating Margin					
Upper Quartile	5.53		1.36		11.33	
Lower Quartile	1.3		1.07		2.3	
		4.25		5.14		4.93

Data From Hughes's Studies (Continued)

	Markup on Total Costs (or operating margin)	Width of Range	Berry Ratio	Width of Range	Return on Assets	Width of Range
Contract Research Services						
Upper Quartile	16.8		1.78		10.3	
Lower Quartile	10.03		1.39		5.97	
		1.67		2		1.73
Metal Manufacturing Services						
Upper Quartile	8.63		2.03		9.47	
Lower Quartile	1.37		1.12		2.23	
		6.3		8.58		4.25
HR Services						
Upper Quartile	9		1.4		10.77	
Lower Quartile	5.17		1.26		7.77	
		1.74		1.54		1.39

