INSIGHT: Disrupting Tax Processes with Artificial Intelligence Technology

By Michel Braun, Steven Galginaitis, Rebecca Reitz, Jillian Schleicher and Brian Shea

Artificial Intelligence (AI) is redefining the future of business and is substantially helping companies to deliver robust intelligent automation solutions. While AI—an area of computer science dealing with the simulation of intelligent behavior in a computer—is relevant to any business field, applying AI to tax processes is in its early stages.

The current Covid-19 pandemic intensifies the need for technological innovation as tax departments increasingly work remotely.

The use case detailed in this article is the application of AI to the development of benchmarking sets used for tax transfer pricing (TP). The development of this use case is a joint, multi-year effort of tax TP and data and analytics professionals.

Traditional benchmarking is a very manual and time-consuming process, and also a process that can be highly subjective. AI can streamline this process while expanding the number of companies to be examined beyond what is practical for a human to handle, and in a more objective and consistent manner.

AI has the additional benefit of enabling analysts to move away from restrictive traditional processes and create benchmark sets that are based firmly and consistently on tested party attributes. By design, AI has the capability to identify more potential comparables than the traditional approach, because it can give consideration to companies that fall outside the traditional search criteria—that is, the quantitative and qualitative screens that are necessary to restrict the set of potential comparables to a number that is practical for a person or team to review.

Introduction Information technology is disrupting the way organizations work, replacing subjective heuristics with data-driven models and automating tedious tasks to allow workers to focus more on policy and related, more interesting, aspects of their work. Furthermore, data and analytics can discern new patterns in the data that can be leveraged to improve the quality, and expand the application, of the analysis.

Tax departments are not immune to this disruption and can benefit from these developments. This article examines a specific use case, benchmarking for TP, and how AI can be used to improve that process.

The Use Case–Background Multinational enterprises (MNEs) must establish transfer prices (i.e., the prices charged by one company of the MNE group to another company within the same MNE group) that meet the arm's-length standard (Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations, OECD, 2017). Essentially, inter-company prices must

Michel Braun is a Director in the tax practice of KPMG Germany; Steven Galginaitis and Brian Shea are Managing Directors with the Washington National Tax practice of KPMG US; Rebecca Reitz is a Senior Associate, Data Scientist and Jillian Schleicher is an Associate, D&A Modeler with the Lighthouse practice of KPMG US.

be consistent with what third parties would charge each other for similar goods and services under similar conditions as those in the subject inter-company transaction.

Under some approaches or pricing methods employed to apply the arm's-length standard, TP benchmarking is to identify independent companies that perform similar functions, own similar assets, and incur similar risks as the company or company sector (the "tested party") under review. The profitability of these "comparable" companies is used to test whether the tested party's profitability is appropriate and, therefore, by implication, its inter-company transactions were consistent with the arm's-length standard.

Currently, this benchmarking effort for TP analysis is a very labor-intensive process (many consulting companies that offer this service are using lower cost skilled labor offshore (e.g., in India) to deliver these services just because they are so labor-intensive).

The current "state of the art" for creating benchmark sets was developed to satisfy two objectives, one regulatory and one practical:

■ to meet the relevant comparability factors of the Organization for Economic Co-operation and Development (OECD) Guidelines (which provide guidance on the main principles that arise in the tax TP area); and

■ to reduce the number of companies to be examined by an analyst to an amount that can be accomplished in a "reasonable" amount of time—that is, the amount of time that enables the analyst to complete their work within budget. Briefly, the current set of tasks to be performed for a TP benchmarking exercise is:

■ Access a vendor-provided database of independent companies. One such database of North American companies is Standard & Poor's, which gathers financial and descriptive information from the U.S. Securities and Exchange Commission public filings. For Europe, Bureau Van Dijk's Amadeus database is often used. Depending on the database, one can start with tens of thousands, up to millions, of company profiles.

■ Reduce the number of companies to be examined using "objective" quantitative screening criteria, such as:

■ selecting companies in a specific industry according to Standard Industrial Classification (SIC) or *Nomenclature des Activités Économiques dans la Communauté Européenne* (NACE) codes (SIC and NACE codes are standard industry classification systems);

■ selecting only those companies with sufficient financial information;

■ selecting companies within a range of certain financial characteristics such as the ratio of expense-to-sales.

■ Once the population of potentially comparable companies is established, an analyst reads the provided business descriptions and often additional information such as public filings or websites, in order to select only those companies with functions, assets, and risks similar to those of the tested party.

This is a labor-intensive, and arguably subjective process—two highly qualified professionals could come up with significantly different benchmarks based on how they set their criteria for determining "comparability," which is often influenced by the amount of experience they have in performing such work. In the authors' experience, a manual benchmarking analysis can average 16 hours for North American companies and 26 hours for European companies.

In addition, for the analysis to be manageable, the process needs to ignore huge amounts of data contained in the companies that are "screened out"—that is, companies that do not meet the quantitative screens described above.

Disrupting the Process Through AI The manual nature of TP benchmarking presents a clear opportunity to apply information technology to improve a tax process. KPMG built a team of TP professionals, data scientists, and user interface programmers to incorporate AI into the process of performing benchmarking analyses. It is important to note that the "human in the loop" makes all final decisions on the appropriateness of the benchmark—the AI only makes recommendations: that is, it does not make final decisions.

The KPMG AI relies on vendor-provided industry codes, business descriptions, and financial data as the inputs for the models. The AI reviews and analyzes this set of inputs and classifies the key attributes of each company. The financial information from the database is used in combination with the text information to determine the main activities (functions) of each company. The text and industry codes are further used to identify the products and services offered by that company as well as any relevant risks.

A critical element for the use of AI in this process is having access to the large set of historical data from KPMG's proprietary database of previous benchmark sets. This database is used to train the AI on how to make recommendations on whether a company should receive further consideration (accepted) or be ignored (rejected).

There are two main means by which the AI interacts with the user: attribute-based screening (ABS) and a "recommender." The KPMG ABS allows the analyst to screen potential comparable companies based on how their attributes align with those of the tested party. After the traditional screening, the potential comparables are characterized by their AI-predicted attributes (i.e., functions, products, services, and risks). The analyst is presented with a menu of these key attributes and can determine which attributes to screen. For example, the user may select that he/she wants service providers of administrative or staffing services. Companies for which the AI does not predict those user-selected attributes are automatically eliminated from the pool of companies to analyze.

The AI "scores" the remaining pool of companies and lists them from most likely to be comparable to least likely to be comparable. The analyst then reviews the companies and begins to either "accept" them or "reject" them. By providing such a list ranked in this manner, we have found that reviewers are able to apply a set of criteria to accept companies in a more consistent manner, resulting in sets that arguably are more internally consistent and therefore better able to withstand tax authority scrutiny.

At any time during the accept/reject process, the analyst can request that the AI run the recommender. The

best way to describe the recommender is to think about an online shopping experience. The recommender is analogous to the ''users who bought the item you like, also purchased this other item'' approach. A customer goes to an online website, enters what they are looking for—a pair of men's shoes, say—and then starts reviewing potential goods to buy. Once the customer goes to the page of an item they wish to purchase—brown dress shoes in size 10, for example—the website will start making recommendations for other items that are often purchased with those shoes (shoe polish and a brown leather belt, perhaps).

The recommender acts in a similar fashion and re-captures information that may otherwise have been disregarded through the traditional screening process. Once the analyst starts the accept/reject approval process, the AI searches the entire vendor database for companies that have similar attributes to those companies accepted by the analyst. The recommender then recommends new companies not previously considered, and re-ranks the pending companies based on the analyst's behavior. The recommender also reviews KPMG's proprietary database of historic searches and suggests companies that were commonly accepted in previously created, similar benchmark sets.

Other key aspects of the AI are:

- The AI suggests reasons for accepting or rejecting each potential comparable to help the analyst make the best decision.
- The AI reads and annotates the business descriptions of comparable companies to highlight the key components of the description, categorizing each annotation by attribute (e.g., function, product, service and risk).
- An open-source webscraping framework supplements the AI. The webscraping process crawls the websites of potential comparable companies in a non-invasive way, identifies relevant sections describing the company, and extracts and summarizes that information for the analyst to consider.
- The AI allows an analyst to consider a much broader and better range of information than is usually considered in the traditional, labor-intensive process.

Once the benchmarking analysis is complete, the final results become part of the historical set of benchmarks to create an ever-expanding set of data to retrain and improve the AI model's performance.

Conclusion Our experience is that a successful AI project with a tax focus requires:

- a multi-disciplined team that combines deep technical expertise on AI and intelligent automation with subject matter expertise on tax TP;
- a scrum-based daily meeting to align the work streams;
- a significant quantity of training data—the use of AI without sufficient training data is not recommended;
- the development of a robust validation process to ensure quality and good performance of AI in combination with a monitoring process to collect and action user feedback.

KPMG continues to expand the AI to utilize additional data sources such as publicly available regulatory documents, and to add features that improve the user experience and model performance. We are also identifying additional use cases such as trade and customs analyses, determining a company's tax and regulatory risk profile based on qualitative and financial information, and assessing the regulatory completeness of a TP documentation report.

Technology is disrupting every aspect of how companies conduct their business. Through the adaptation of AI and other information technologies, tax departments can gain efficiencies and consistency in their processes.

Michel Braun is a Director in the tax practice of KPMG Germany; Steven Galginaitis and Brian Shea are Managing Directors with the Washington National Tax practice of KPMG US; Rebecca Reitz is a Senior Associate, Data Scientist and Jillian Schleicher is an Associate, D&A Modeler with the Lighthouse practice of KPMG US.

The authors may be contacted at: mbraun@kpmg.com; sgalginaitis@kpmg.com; rreitz@kpmg.com; jillianschleicher@kpmg.com; brianshea@kpmg.com

The information in this article is not intended to be ''written advice concerning one or more Federal tax matters'' subject to the requirements of section 10.37(a)(2) of Treasury Department Circular 230. The information contained herein is of a general nature and based on authorities that are subject to change. Applicability of the information to specific situations should be determined through consultation with your tax adviser. This article represents the views of the author(s) only, and does not necessarily represent the views or professional advice of KPMG LLP.

This column does not necessarily reflect the opinion of The Bureau of National Affairs, Inc. or its owners.