edgeworth | economics

February 2020

Taking a "Byte" Out of Antitrust Data Discovery: The Top 10 "Bits"

George Korenko, PhD
Partner | Washington, DC
gkorenko@edgeworthe conomics.com
+1 202 559 4408

Matthew Milner
Partner | Washington, DC
mmilner@edgeworthe conomics.com
+1 202 559 4387

In many antitrust cases, economists use transactional data and related information to conduct economic analyses and form opinions on antitrust impact and damages. For example, in price fixing cases, detailed data are used to measure the effects (if any) of the alleged conspiracy and to quantify damages. In monopolization cases, transactional data are often used to define relevant antitrust markets, to measure the effects of the alleged conduct on prices, and to assess procompetitive benefits of that conduct. Moreover, transactional data can be useful at every stage of antitrust litigation from class certification—to assess whether reliable statistical models can be used to demonstrate that all or substantially all members of the proposed class are injured from the alleged conduct—through the merits and damages stages, where injury must be demonstrated and quantified. However, before any economic expert can apply econometric methods using transactional data, these data are collected from the respective parties.

A challenge facing the parties to antitrust litigation can be the identification, coding, and extraction of relevant transactional data from the plaintiff and defendant computing systems. The data and information requested by the parties can span several years, multiple systems, and cover a myriad of data tables and fields. To help practitioners overcome the hurdles of gathering detailed data in antitrust cases, here are 10 bits (or best practices) to make the process more cost efficient and effective for litigation.

Originally published in ABA *Economics Committee Newsletter*, Winter 2020.

1. It sounds simple but start with the parties involved in the lawsuit.

Upon starting discovery for data and information it is imperative to understand which parties are bringing the lawsuit and identifying the relevant entities that are named as defendants. While apparent on paper, the implications on the scope of data can vary depending on the entities in the complaint. For example, the companies' subsidiaries with relevant sales, the physical locations of facilities, and where customers purchase relevant products can affect which databases and systems are pertinent for discovery.

2. The nature of the allegations dictates what information may be pertinent to the case.

The purported conduct specified in the complaint provides guidance on the types of information that may be important to a case. For instance, in a direct purchaser lawsuit the focus is on whether an alleged conspiracy resulted in elevated prices for the defendants' customers. Economists can use transactional sales data to test:

- Which customers experienced a price increase relative to the world absent the alleged conspiracy (*i.e.*, the "but-for" world)?
- How much were prices elevated compared to the but-for world?
- What are the estimated economic damages resulting from the purported anticompetitive conduct?

To answer these questions requires analytical rigor and reliance on data containing sales and prices to customers of relevant products, and other fact-specific information related to the industry. A great feature of these data is the depth of information often contained in them. For instance, transactional data consist of detailed invoice-level prices of the products sold to individual customers. This data typically contains specifics on each sale, including the sales amount, quantity sold, customer name and addresses, and product characteristics. This information (along with other relevant economic factors) can provide inputs into an econometric model seeking to explain how prices are determined under competitive conditions. Accounting for relevant non-collusive factors in an econometric model is critical to isolate the effects of the alleged conduct.

3. The period of interest is influenced by the complaint.

A complaint lays out when the alleged antitrust conduct purportedly started and ended. This window can be used to guide the time period selected for data collection. For example, if allegations of attempted monopolization involve conduct that took place between 2012 and 2018, the focus is on the nature of competition and the potential elevation in prices paid during that period.

An economist approaches these issues by asking questions about the appropriate comparisons to the world absent the alleged conduct. What period could be used as an appropriate competitive benchmark? Are data available prior to 2012 (the pre-conduct period)? Are data during the post-conduct period an appropriate benchmark? Are there other closely related products sold during the relevant period that were not subject to the allegations?

From a discovery perspective, the decision to collect data from certain periods may be influenced by what is readily available from the parties during the ordinary course of business. To the economist, the data available may influence the decisions regarding which period(s) can be used as a competitive benchmark and how to reliably control for varying supply and demand conditions.

4. Focus on data systems that contain relevant information.

The proliferation of "big data" collected and maintained by businesses has resulted in staggering growth in the amount of data available over the past decade. Technological gains in data processing, the advent of real time analytics, and cloud-based storage solutions have created opportunities for businesses to leverage both structured and unstructured data in new ways.[1]

Moreover, the shift towards big data has required companies to make choices on how data are collected, used, and integrated into enterprise management systems. For example, transactional sales data across offices and subsidiaries can now be managed in a more cohesive manner. Cloud-based platforms facilitate data access across geographic locations and have also made it easier to export data into other formats. Notwithstanding, the adaptation of cloud solutions can create other challenges as it relates to e-discovery.

Data variety is a key issue. Migrations from older systems across an enterprise can introduce heterogeneity within a centralized data warehouse. Transactional sales data may be maintained in the same database as web clicks, social media engagements, and logistics. Which of these data are relevant? What about the sales from non-relevant subsidiaries that are maintained in the same database? The need for data from legacy systems may be influenced by the period in the complaint, but are data on web clicks relevant? If the answer to each of these questions is "no," then these data should not be extracted simply because they are maintained, even if they are in the same system. Rather, it will be useful to isolate the data that have information that relates directly to the economic analysis of the issues at hand.

5. All tables in databases are not created equal.

There are often thousands of data tables that are running in an enterprise resource management system. For example, when limiting the data to sales and distribution information alone, there may be up to 700 data tables in an SAP system that are running in the background.[2] These tables are connected through relational variables (*i.e.*, common fields) and database processes that allow users to query the data across multiple dimensions and create reports used to support the business.

For instance, customer information may be found in multiple tables. To illustrate with an example, suppose the customer name field is in one table and the address field is kept in another table—the tables are linked with a customer identification code so the system can retrieve information from both tables when running a report. The customer name, address, and identification code data may be all that is needed from these tables.

1. Randy Bean, "How Big Data Is Empowering AI and Machine Learning at Scale," *MIT Sloan Management Review,* May 8, 2017. See: https://sloanreview.mit.edu/article/how-big-data-is-empowering-ai-and-machine-learning-at-scale/.

2. See for example, https://www.se80.co.uk/sapmodules/s/sd-t/sd-tables.htm.

While there could be many other tables that contain fields that relate to customers, that information may be used for internal processing or contain non-alphanumeric characters that are not useful for an economic analysis. It is important to recognize which tables in a database contain relevant fields and which other tables may contain information not pertinent to discovery.

6. Keep in mind that relevant pricing data does not mean all the data.

There can be wide variation in the types of pricing data and information kept at a company. That is, invoice prices found in transactional data may differ from the actual prices that customers paid due to factors such as off-invoice discounts and rebates based on periodic sales volumes. Data containing these rebates, credits, discounts, and other price adjustments can be critical in the calculation of actual prices paid and, therefore, the assessment of antitrust impact and damages.

There can also be thousands of product numbers in transactional data systems. However, depending on the nature of the allegations, only a subset of the products may be relevant to the antitrust claims. For example, depending on the how the data are organized, transactions may be limited to relevant products based on information such as product descriptions and codes that delineate product lines. A careful assessment of the pricing data and product information can influence which transactions and adjustments are relevant to assess the antitrust allegations.

7. Let the invoice be your guide.

In an era where sales are increasingly shifting to digital platforms, there is still tremendous value in the paper trail that is left from a transaction. Whether it is a printout of an electronic invoice or the paper invoice itself, the receipt contains valuable information on pricing and sales. Specifically, the invoice provides details on each transaction including the customer, products purchased, unit prices, quantities sold, and relevant dates of sale and shipment. Circling back to the complexities of relational databases—the information on the invoice links to data tables and fields, providing a roadmap to the important information for extraction.

8. A small data extract should be reviewed before the data are pulled *en masse*.

Resist the temptation. It may seem like a good idea that once you have identified the relevant information in a data system to request the data be produced for the *entire* period of interest. In order to extract the data from a system into a more manageable format, coding and queries are run by IT professionals to export relevant tables and fields. Unfortunately, the approach of trying to take a single bite at the proverbial apple does not always go as planned.

Efficient extraction of these data files can be done using a measured approach. Before a party begins generating gigabytes or terabytes of new files for turnover, it is prudent to review a smaller extract of the data. Depending on the size of the data, the extract may represent a few months or up to a year of transactions. A smaller file requires less computing resources and allows counsel and/or consulting experts to examine the data before it is finalized. This intermediate step allows for changes to the queries supporting the larger extraction and reduces the chance of errors with the final export.

9. Validation can be a safeguard.

To ensure the data are reliable, the client or consulting experts should check to make sure the data in the extract accurately reflect the experience of the business and customers. That means examining the data to check if there were errors during the extraction, such as truncated variables, deleted values, or missing records.

Another best practice is to plot the prices and sales data over time to see how the number of observations vary, how sales fluctuate, and the how the values of key variables change. Lastly, if contemporaneous sales reports exist, comparing the totals derived from the transactional data to the line items in the sales report can serve to confirm that the data are complete and accurate.

10. Do it once.

The nature of discovery in each antitrust case is unique to the parties involved and the allegations at hand. Taking a principled approach to data discovery will increase the likelihood that reliable structured data are extracted only once, and that they contain the fields that are relevant to address the economic issues at hand. In the end that is something that benefits all parties to the litigation. ■

edgeworth | economics

About Edgeworth

Edgeworth Economics is an economic and quantitative consulting firm that provides economic analysis and expert testimony for clients facing complex litigation, regulatory, and other challenges involving antitrust, class certification, intellectual property, and labor and employment. Edgeworth's expert economists, statisticians, data analysts, and other professionals assist clients with innovative solutions rooted in the rigorous application of economic principles and hard data. Organizations including leading law firms, Fortune 500 companies, and government agencies rely on Edgeworth Economics to help them navigate through their most critical legal disputes and other challenges.

www.edgeworthеconomics.com