Chapter 1. Time Series: An Overview and a Quick History

Time series data and its analysis are increasingly important due to the massive production of such data through, for example, the internet of things, the digitalization of healthcare, and the rise of smart cities. In the coming years we can expect the quantity, quality, and importance of time series data to grow rapidly.

As continuous monitoring and data collection become more common, the need for competent time series analysis with both statistical and machine learning techniques will increase. Indeed, the most promising new models combine both of these methodologies. For this reason, we will discuss each at length. We will study and use a broad range of time series techniques useful for analyzing and predicting human behavior, scientific phenomena, and private sector data, as all these areas offer rich arrays of time series data.

Let's start with a definition. *Time series analysis* is the endeavor of extracting meaningful summary and statistical information from points arranged in chronological order. It is done to diagnose past behavior as well as to predict future behavior. In this book we will use a variety of approaches, ranging from hundred-year-old statistical models to newly developed neural network architectures.

None of the techniques has developed in a vacuum or out of purely theoretical interest. Innovations in time series analysis result from new ways of collecting, recording, and visualizing data. Next we briefly discuss the emergence of time series analysis in a variety of applications.

The History of Time Series in Diverse Applications

Time series analysis often comes down to the question of causality: how did the past influence the future? At times, such questions (and their answers) are treated strictly within their discipline rather than as part of the general discipline of time series analysis. As a result, a variety of

disciplines have contributed novel ways of thinking about time series data sets.

In this section we will survey a few historical examples of time series data and analysis in these disciplines:

- Medicine

- Weather

- Economics

- Astronomy

As we will see, the pace of development in these disciplines and the contributions originating in each field were strongly tied to the nature of the contemporaneous time series data available.

Medicine as a Time Series Problem

Medicine is a data-driven field that has contributed interesting time series analysis to human knowledge for a few centuries. Now, let's study a few examples of time series data sources in medicine and how they emerged over time.

Medicine got a surprisingly slow start to thinking about the mathematics of predicting the future, despite the fact that prognoses are an essential part of medical practice. This was the case for many reasons. Statistics and a probabilistic way of thinking about the world are recent phenomena, and these disciplines were not available for many centuries even as the practice of medicine developed. Also, most doctors practiced in isolation, without easy professional communication and without a formal recordkeeping infrastructure for patient or population health. Hence, even if physicians in earlier times had been trained as statistical thinkers, they likely wouldn't have had reasonable data from which to draw conclusions.

This is not at all to criticize early physicians but to explain why it is not too surprising that one of the early time series innovations in population health came from a seller of hats rather than from a physician. When you think about it, this makes sense: in earlier centuries, an urban hat seller

would likely have had more practice in recordkeeping and the art of spotting trends than would a physician.

The innovator was John Graunt, a 17th-century London haberdasher. Graunt undertook a study of the death records that had been kept in London parishes since the early 1500s. In doing so, he originated the discipline of demography. In 1662, he published *Natural and Political Observations . . . Made upon the Bills of Mortality* (See Figure 1-1).

Figure 1-1. John Graunt's actuarial tables were one of the first results of time-series-style thinking applied to medical questions. Source: Wikipedia.

In this book, Graunt presented the first *life tables*, which you may know as *actuarial tables*. These tables show the probability that a person of a given age will die before their next birthday. Graunt, as the first person known to have formulated and published life tables, was also the first documented statistician of human health. His life tables looked something like Table 1-1, which is taken from some Rice University statistics course notes.

Age	Proportion of deaths in the interval	Proportion surviving unti
0–6	0.36	1.0

Age	Proportion of deaths in the interval	Proportion surviving unti
7–16	0.24	0.64
17–26	0.15	0.40
27–36	0.09	0.25

Table 1-1. A sample of John Graunt's life tables

Unfortunately, Graunt's way of thinking mathematically about human survival did not take. A more connected and data-driven world began to form—complete with nation states, accreditation, professional societies, scientific journals, and, much later, government-mandated health recordkeeping—but medicine continued to focus on physiology rather than statistics.

There were understandable reasons for this. First, the study of anatomy and physiology in small numbers of subjects had provided the major advances in medicine for centuries, and most humans (even scientists) hew to what works for them as long as possible. While a focus on physiology was so successful, there was no reason to look further afield. Second, there was very little reporting infrastructure in place for physicians to tabulate and share information on the scale that would make statistical methods superior to clinical observations.

Time series analysis has been even slower to come into mainstream medicine than other branches of statistics and data analysis, likely because time series analysis is more demanding of recordkeeping systems. Records must be linked together over time, and preferably collected at regular intervals. For this reason, time series as an epidemiological practice has only emerged very recently and incrementally, once sufficient governmental and scientific infrastructure was in place to ensure reasonably good and lengthy temporal records.

Likewise, individualized healthcare using time series analysis remains a young and challenging field because it can be quite difficult to create data sets that are consistent over time. Even for small case-study-based research, maintaining both contact with and participation from a group of individuals is excruciatingly difficult and expensive. When such studies are conducted for long periods of time, they tend to become canonical in their fields—and repeatedly, or even excessively researched—because their data can address important questions despite the challenges of funding and management.[1]

MEDICAL INSTRUMENTS

Time series analysis for individual patients has a far earlier and more successful history than that of population-level health studies. Time series analysis made its way into medicine when the first practical electrocardiograms (ECGs), which can diagnose cardiac conditions by recording the electrical signals passing through the heart, were invented in 1901 (see Figure 1-2). Another time series machine, the electroencephalogram (EEG), which noninvasively measures electrical impulses in the brain, was introduced into medicine in 1924, creating more opportunities for medical practitioners to apply time series analysis to medical diagnosis (see Figure 1-3).

Both of these time series machines were part of a larger trend of enhancing medicine with repurposed ideas and technologies coming out of the second Industrial Revolution.

Figure 1-3. The first human EEG recording, from 1924. Source: Wikipedia.

ECG and EEG time series classification tools remain active areas of research for very practical purposes, such as estimating the risk of a sudden cardiac crisis or a seizure. These measurements are rich sources of data, but one "problem" with such data is that it tends to be available only for patients with specific ailments. These machines do not generate long-range time series that can tell us more broadly about human health and behavior, as their measurements are seldom applied for long periods of time or before a disease has emerged in a patient.

Luckily, from a data analysis point of view, we are moving past the era where ECGs and the like are the dominant medical time series available. With the advent of wearable sensors and "smart" electronic medical devices, many healthy humans take routine measurements automatically or with minimal manual input, leading to the ongoing collection of good longitudinal data about both sick and healthy people. This is in stark contrast to the last century's medical time series data, which was almost exclusively measured on sick people and which was very limited in access.

As recent news coverage has shown, a variety of nontraditional players are entering the medical field, ranging from enormous social media companies to financial institutions to retail giants.[2] They likely all plan to use large data sets to streamline healthcare. There aren't just new players in the healthcare field—there are also new techniques. The personalized DNA-driven medicine means that time series *data* is increasingly measured and valued. Thanks to burgeoning modern healthcare data sets, both healthcare and time series analysis will likely evolve in the coming years, particularly in response to the lucrative data

sets of the healthcare sector. Hopefully this will happen in such a way that time series can benefit everyone.

Forecasting Weather

For obvious reasons, predicting the weather has long been a preoccupation to many. The ancient Greek philosopher Aristotle delved into weather with an entire treatise (*Meteorology*), and his ideas about the causes and sequencing of the weather remained dominant until the Renaissance. At that time, scientists began to collect weather-related data with the help of newly invented instruments, such as the barometer, to measure the state of the atmosphere. They used these instruments to record time series at daily or even hourly intervals. The records were kept in a variety of locations, including private diaries and local town logbooks. For centuries this remained the only way that Western civilization tracked the weather.

Greater formalization and infrastructure for weather recording arrived in the 1850s when Robert FitzRoy was appointed the head of a new British government department to record and publish weather-related data for sailors.[3] FitzRoy coined the term *weather forecast*. At the time, he was criticized for the quality of his forecasts, but he is now regarded to have been well ahead of his time in the science he used to develop them. He established the custom of printing weather forecasts in the newspaper; they were the first forecasts printed in *The Times* of London. FitzRoy is now celebrated as the "father of forecasting."

In the late 19th century—hundreds of years after many atmospheric measurements had come into use—the telegraph allowed for fast compilations of atmospheric conditions in time series from many different locations. This practice became standard in many parts of the world by the 1870s and led to the creation of the first meaningful data sets for predicting local weather based on what was happening in other geographic locations.

By the turn of the 20th century, the idea of forecasting the weather with computational methods was vigorously pursued with the help of these compiled data sets. Early endeavors at computing the weather required a spectacular amount of effort but gave poor results. While physicists and chemists had well-proven ideas about the relevant natural laws, there

were too many natural laws to apply all at once. The resulting system of equations was so complex that it was a notable scientific breakthrough the first time someone even attempted to do the calculations.

Several decades of research followed to simplify the physical equations in a way that increased accuracy and computational efficiency. These tricks of the trade have been handed down even to current weather prediction models, which operate on a mix of known physical principles and proven heuristics.

Nowadays many governments make highly granular weather measurements from hundreds or even thousands of weather stations around the world, and these predictions are grounded in data with precise information about weather station locations and equipment. The roots of these efforts trace back to the coordinated data sets of the 1870s and even earlier to the Renaissance practice of keeping local weather diaries.

Unfortunately, weather forecasting is an example of the increasing attacks on science that reach even into the domain of time series forecasting. Not only have time series debates about global temperatures been politicized, but so have more mundane time series forecasting tasks, such as predicting the path of a hurricane.

Forecasting Economic Growth

Indicators of production and efficiency in markets have long provided interesting data to study from a time series analysis. Most interesting and urgent has been the question of forecasting future economic states based on the past. Such forecasts aren't merely useful for making money— they also help promote prosperity and avert social catastrophes. Let's discuss some important developments in the history of economic forecasting.

Economic forecasting grew out of the anxiety triggered by episodic banking crises in the United States and Europe in the late 19th and early 20th centuries. At that time, entrepreneurs and researchers alike drew inspiration from the idea that the economy could be likened to a cyclical system, just as the weather was thought to behave. With the right

measurements, it was thought, predictions could be made and crashes averted.

Even the language of early economic forecasting mirrored the language of weather forecasting. This was unintentionally apt. In the early 20th century, economic and weather forecasting were indeed alike: both were pretty terrible. But economists' aspirations created an environment in which progress could at least be hoped for, and so a variety of public and private institutions were formed for tracking economic data. Early economic forecasting efforts led to the creation of economic indicators and tabulated, publicly available histories of those indicators that are still in use today. We will even use some of these in this book.

Nowadays, the United States and most other nations have thousands of government researchers and recordkeepers whose jobs are to record data as accurately as possible and make it available to the public (see Figure 1-4). This practice has proven invaluable to economic growth and the avoidance of economic catastrophe and painful boom and bust cycles. What's more, businesses benefit from a data-rich atmosphere, as these public data sets permit transportation providers, manufacturers, small business owners, and even farmers to anticipate likely future market conditions. This all grew out of the attempt to identify "business cycles" that were thought to be the causes of cyclical banking failures, an early form of time series analysis in economics.

Figure 1-4. The US federal government funds many government agencies and related nonprofits that record vital statistics as well as formulate economic indicators. Source: <u>National Bureau of Economic Research</u>.

Much of the economic data collected by the government, particularly the most newsworthy, tends to be a proxy for the population's overall economic well-being. One example of such vital information comes from the number of people requesting unemployment benefits. Examples include the government's estimates of the gross domestic product and of the total tax returns received in a given year.

Thanks to this desire for economic forecasting, the government has become a curator of data as well as a collector of taxes. The collection of this data enabled modern economics, the modern finance industry, and data science generally to blossom. Thanks to time series analysis growing out of economic questions, we now safely avert many more banking and financial crises than any government could have in past centuries. Also, hundreds of time series textbooks have been written in the form of economics textbooks devoted to understanding the rhythms of these financial indicators.

TRADING MARKETS