CHAPTER 1

Introduction

After reading this chapter you will understand:

- What the field of financial econometrics covers.
- The three steps in applying financial econometrics: model selection, model estimation, and model testing.
- What is meant by the data generating process.
- How financial econometrics is used in the various phases of investment management.

Financial econometrics is the science of modeling and forecasting financial data such as asset prices, asset returns, interest rates, financial ratios, defaults and recovery rates on debt obligations, and risk exposure. Some have described financial econometrics as the econometrics of financial markets. The development of financial econometrics was made possible by three fundamental enabling factors: (1) the availability of data at any desired frequency, including at the transaction level; (2) the availability of powerful desktop computers at an affordable cost; and (3) the availability of off-the-shelf econometric software. The combination of these three factors put advanced econometrics within the reach of most financial firms such as banks and asset management firms.

In this chapter, we describe the process and the application of financial econometrics. Financial econometrics is applied to either time series data, such as the returns of a stock, or cross-sectional data such as the market capitalization[1] of all stocks in a given universe at a given moment. With the progressive diffusion of high-frequency financial data and ultra high-frequency financial data, financial econometrics can now be applied to

[1]A firm's market capitalization, popularly referred to as "market cap," is a measure of the firm's size in terms of the total market value of its common stock. This is found by multiplying the number of common stock shares outstanding by the price per share of common stock.

larger databases making statistical analysis more accurate as well as providing the opportunity to investigate a wider range of issues regarding financial markets and investment strategies.[2]

FINANCIAL ECONOMETRICS AT WORK

Applying financial econometrics involves three key steps:

Step 1. Model selection

Step 2. Model estimation

Step 3. Model testing

For asset managers, traders, and analysts, the above three steps should lead to results that can be used in formulating investment strategies. Formulating and implementing strategies using financial econometrics is the subject of the final chapter of this book, Chapter 15.

Below we provide a brief description of these three steps. More details are provided in later chapters. Model selection is the subject of Chapter 14 and model estimation is covered in Chapter 13.

Step 1: Model Selection

In the first step, model selection, the modeler chooses a family of models with given statistical properties. This entails the mathematical analysis of the model properties as well as financial economic theory to justify the model choice. It is in this step that the modeler decides to use, for example, an econometric tool such as regression analysis to forecast stock returns based on fundamental corporate financial data and macroeconomic variables.

In general, it is believed that one needs a strong economic intuition to choose models. For example, it is economic intuition that might suggest what factors are likely to produce good forecasting results, or under what conditions we can expect to find processes that tend to revert to some long-run mean. We can think of model selection as an adaptive process where

[2] Engle provides the following distinction between high-frequency financial data and ultra high-frequency data. Observations on financial variables such as prices that are taken daily or at a finer time scale are referred to as high-frequency financial data. Typically, such observations are regularly spaced over time. Ultra high-frequency financial data refers to time stamped transaction-by-transaction or tick-by-tick data which are irregularly spaced. See Robert F. Engle, "The Econometrics of Ultra-High Frequency Data," *Econometrica* 69, no. 1 (2000), 1–22.

economic intuition suggests some family of models which need, however, to pass rigorous statistical testing.

On the other hand, financial econometrics might also use an approach purely based on data. “Let the data speak” is the mantra of this approach. An approach purely based on data is called *data mining*. This approach might be useful but must be used with great care. Data mining is based on using very flexible models that adapt to any type of data and letting statistics make the selection. The risk is that one might capture special characteristics of the sample which will not repeat in the future. Stated differently, the risk is that one is merely “fitting noise.” The usual approach to data mining is to constrain models to be simple, forcing models to capture the most general characteristics of the sample.

Hence, data mining has to be considered a medicine which is useful but which has many side effects and which should be administered only under strict supervision by highly skilled doctors. Imprudent use of data mining might lead to serious misrepresentations of risk and opportunities. On the other hand, a judicious use of data mining might suggest true relationships that might be buried in the data.

Step 2: Model Estimation

In general, models are embodied in mathematical expressions that include a number of parameters that have to be estimated from sample data, the second step in applying financial econometrics. Suppose that we have decided to model returns on a major stock market index such as the Standard & Poor’s 500 (S&P 500) with a regression model, a technique that we discuss in later chapters. This requires the estimation of the regression coefficients, performed using historical data. Estimation provides the link between reality and models. We choose a family of models in the model selection phase and then determine the optimal model in the estimation phase.

There are two main aspects in estimation: finding estimators and understanding the behavior of estimators. Let’s explain. In many situations we simply directly observe the magnitude of some quantity. For example, the market capitalization of firms is easily observed. Of course there are computations involved, such as multiplying the value of a stock by the number of outstanding stocks, but the process of computing market capitalization is essentially a process of direct observation.

When we model data, however, we cannot directly observe the parameters that appear in the model. For example, consider a very simple model of trying to estimate a linear relationship between the weekly return on General Electric (GE) stock and the return on the S&P 500. When we discuss the econometric technique known as simple linear regression analysis in Chapter 2, we will see the relationship of interest to use would be[3]

$$\text{Return on GE stock} = \alpha + \beta\ (\text{Return on S\&P 500}) + \text{Error term}$$

The two parameters in the above relationship are α and β and are referred to as regression coefficients. We can directly observe from trading data the information necessary to compute the return on both the GE stock and the S&P 500. However, we cannot directly observe the two parameters. Moreover, we cannot observe the error term for each week. The process of estimation involves finding estimators. Estimators are numbers computed from the data that approximate the parameter to be estimated.

Estimators are never really equal to the theoretical values of the parameters whose estimate we seek. Estimators depend on the sample and only approximate the theoretical values. The key problem in financial econometrics is that samples are generally small and estimators change significantly from sample to sample. This is a major characteristic of financial econometrics: samples are small, noise is very large, and estimates are therefore very uncertain. Financial econometricians are always confronted with the problem of extracting a small amount of information from a large amount of noise. This is one of the reasons why it is important to support econometric estimates with financial economic theory.

Step 3: Model Testing

As mentioned earlier, model selection and model estimation are performed on historical data. As models are adapted (or fitted) to historical data there is always the risk that the fitting process captures characteristics that are specific to the sample data but are not general and will not reappear in future samples. For example, a model estimated in a period of particularly high returns for stocks might give erroneous indications about the true average returns. Thus there is the need to test models on data different from the data on which the model was estimated. This is the third step in applying financial econometrics, model testing. We assess the performance of models on fresh data. This is popularly referred to as "backtesting."

A popular way of backtesting models is the use of moving windows. Suppose we have 30 years of past weekly return data for some stock and we want to test a model that forecasts one week ahead. We could estimate the model on the past 30 years minus one week and test its forecasting

[3]As explained in Chapter 2, this relationship for a stock is referred to as its characteristic line.

abilities on the last week. This method would have two major drawbacks. First, we would have only one forecast as a test; second, the model would be estimated on data that do not reflect the market situation today.

A sensible way to solve the problem of backtesting is to use samples formed from a shorter series of data (say, three or four years), estimate the model on the sample data, and then test the forecast on the week immediately following the sample data. We then move the window forward one week and we repeat the process. In this way, we can form a long series of test forecasts. Note two things about this procedure. First, for each window there is a strict separation of sample and testing data. Second, we do not test a single model, but a family of models that are reestimated in each window.

The choice of the length of the estimation window is a critical step. One must choose a window sufficiently long to ensure a reasonable estimation of the model. At the same time, the window must be sufficiently short so that the parameters don't change too much within the window.

THE DATA GENERATING PROCESS

The basic principles for formulating quantitative laws in financial econometrics are the same as those that have characterized the development of quantitative science over the last four centuries. We write mathematical models—relationships between different variables and/or variables in different moments and different places. The basic tenet of quantitative science is that there are relationships that do not change regardless of the moment or the place under consideration. For example, while sea waves might look like an almost random movement, in every moment and location the basic laws of hydrodynamics hold without change. Similarly, in financial markets, asset price behavior might appear to be random, but financial econometric laws should hold in every moment and for every asset class.[4]

There are similarities between financial econometric models and models of the physical sciences but there are also important differences. The physical sciences aim at finding immutable laws of nature; financial econometric

[4]In most developed countries, the four major asset classes are (1) common stocks, (2) bonds, (3) cash equivalents, and (4) real estate. Typically, an asset class is defined in terms of the following three investment characteristics that the members of an asset class have in common: (1) the major economic factors that influence the value of the asset class and, as a result, correlate highly with the returns of each member included in the asset class, (2) a similar risk and return characteristic, and (3) a common legal or regulatory structure. Based on this way of defining an asset class, the correlation between the returns of different asset classes would be low.

models model the economy or financial markets—artifacts subject to change. For example, U.S. financial markets in the form of stock exchanges have been in operation since May 1792 (the origin of the New York Stock Exchange). Since that time, stock exchanges in the United States—as well as throughout the world—have changed significantly both in the number of stocks listed and the type of trading. And the information available on transactions has also changed. Consider that in the 1950s, market participants had access only to daily closing prices and this typically was available the next day rather than at the close of the trading day; now we have instantaneous information on every single transaction. Because the economy and financial markets are artifacts subject to change, financial econometric models are not unique representations valid throughout time; they must adapt to the changing environment.

We refer to the mathematical model that represents future data in function of past and present data as the *data generating process* (DGP). If we know the DGP, we can generate data with the same statistical characteristics as our empirical data. If we know a DGP as a mathematical expression, we can implement computer programs that simulate data. These simulated data can be used to compute statistical quantities that would be difficult or even impossible to compute mathematically. Methods based on simulation techniques are generally called *Monte Carlo methods*.

APPLICATIONS OF FINANCIAL ECONOMETRICS TO INVESTMENT MANAGEMENT

Researchers investigating important issues in finance employ financial econometrics in their empirical analysis. The issues that they have tackled in finance cover critical issues in the fields of financial markets, corporate finance, and investment management. Many of the studies on financial markets have helped either formulate or discredit policies used by investors and regulators. Empirical studies of the impact of capital structure (i.e., the mix of debt and equity in the financing of a firm) decision, the dividend decision, and the stock-buyback decision using financial econometrics have provided a useful guide to senior corporate management and boards of directors in formulating corporate financial policy.

The most significant use of financial econometrics since the early 1990s has been in the field of investment management. It is an important part of the arsenal of tools used by quantitative asset management firms. Within the real world of investment management, financial econometrics has been used in the following tasks: asset allocation, portfolio construction, and portfolio risk management. Since the key real-world use of financial econometrics has been in investment management and many of the illustrations in this book are from this field, we conclude this chapter with a brief explanation of asset allocation, portfolio construction, and portfolio risk management.

Asset Allocation

A major activity in the investment management process is establishing policy guidelines to satisfy a client's investment objectives. Setting policy begins with the asset allocation decision. That is, a decision must be made as to how the funds to be invested should be distributed among the major asset classes.

The term "asset allocation" means different things to different people in different contexts. One can divide asset allocation into three types: (1) policy asset allocation, (2) dynamic asset allocation, and (3) tactical asset allocation.[5] The policy asset allocation decision can loosely be characterized as a long-term asset allocation decision, in which the investor seeks to assess an appropriate long-term "normal" asset mix that represents an ideal blend of controlled risk and enhanced return. In dynamic asset allocation, the asset mix is mechanistically shifted in response to changing market conditions. Once the policy asset allocation has been established, the investor can turn attention to the possibility of active departures from the normal asset mix established by policy. That is, suppose that the long-run asset mix is established as 40% stocks and 60% bonds. A departure from this mix under certain circumstances may be permitted. If a decision to deviate from this mix is based upon rigorous objective measures of value, it is often called tactical asset allocation. Tactical asset allocation broadly refers to active strategies that seek to enhance performance by opportunistically shifting the asset mix of a portfolio in response to the changing patterns of reward available in the capital markets. Notably, tactical asset allocation tends to refer to disciplined processes for evaluating prospective rates of return on various asset classes and establishing an asset allocation response intended to capture higher rewards.

Models used in each type of asset allocation described above rely on the forecasting of returns for the major asset classes and the expected future relationship among the asset classes. Broad-based market indexes are used to represent major asset classes. For U.S. common stock, this would typically mean forecasting returns for the S&P 500 index, and for bonds, the returns for the Barclays U.S. Aggregate Bond index.

---

[5] Robert D. Arnott and Frank J. Fabozzi, "The Many Dimensions of the Asset Allocation Decision," in *Active Asset Allocation*, ed. Robert D. Arnott and Frank J. Fabozzi (Chicago: Probus Publishing, 1992).

Forecasting for asset allocation goes beyond just forecasting returns. A fundamental principle of finance is that investors must accept a trade-off between risk and returns. Hence in asset allocation modeling, one must forecast risk and not only returns. The most fundamental ingredient to forecast risk is the covariance matrix. Hence, a fundamental component of portfolio formation is the estimation of the covariance matrix between the major asset classes.

Portfolio Construction

Selecting a portfolio strategy that is consistent with the investment objectives and investment policy guidelines of a client or an institution is a major activity in the investment management process. Portfolio strategies can be classified as either active or passive.

An *active portfolio strategy* uses available information and forecasting techniques to seek a better performance than a portfolio that is simply diversified broadly. Essential to all active strategies are expectations about the factors that have been found to influence the performance of an asset class. For example, with active common stock strategies this may include forecasts of future earnings, dividends, or price-earnings ratios. With bond portfolios that are actively managed, expectations may involve forecasts of future interest rates and sector spreads. Active portfolio strategies involving foreign securities may require forecasts of local interest rates and exchange rates.

Portfolio construction and optimization in active portfolio strategies require models for forecasting returns: There is no way to escape the need to predict future returns. In stock portfolios, we would need a forecast of the return for every candidate stock that a portfolio manager wants to consider for inclusion into the portfolio. Moreover, as explained in our discussion of asset allocation, risk must be forecasted in constructing a portfolio. The covariance matrix for the candidate assets must therefore be estimated.

A *passive portfolio strategy* involves minimal expectational input, and instead relies on diversification to match the performance of some market index. In effect, a passive strategy assumes that the marketplace will efficiently reflect all available information in the price paid for all assets.[6] Passive strategies eschew the need to forecast future returns of individual asset classes by investing in broad indexes. These strategies effectively shift the need to forecast to a higher level of analysis and to longer time horizons. Active strategies, however, form portfolios based on forecasts of future returns.

[6] Between these extremes of active and passive strategies, several strategies have sprung up that have elements of both. For example, the core of a portfolio may be passively managed while the balance is actively managed.

The most sophisticated models used in portfolio construction are factor risk models (or simply factor models) using the financial econometric tools of factor analysis and principal components analysis described in Chapter 12.

Portfolio Risk Management

Portfolio risk management can be broadly defined as a set of methods and techniques to set portfolio risk objectives, estimate the risk of a portfolio strategy, and take appropriate corrective measures if necessary. Portfolio risk itself can be defined in many different ways but essentially is a measurement of the uncertainty related to future returns. There is risk when there is the possibility that future returns, and therefore the value of future wealth, will deviate from expectations.

Portfolio management is essentially the management of the trade-off between risk and return. There are various analytical measures that can be used to identify the various risks of a portfolio such as standard deviation, value-at-risk, or conditional value-at-risk, tracking error, to name just a few. (These measures are described later in this book.) Often these measures must be estimated using the financial econometrics tools described in the chapters to follow. The larger asset management firms have an in-house risk group that monitors portfolio risk and provides at least daily the portfolio's risk exposure.

In portfolio management, the key risk is that the performance of the portfolio manager is below the return earned on a client-approved benchmark after adjusting for management fees. The benchmark could be any index such as the S&P 500 index or the Barclays Capital U.S. Aggregate Bond index. The key measure used in controlling a portfolio's risk is *tracking error*. Tracking error measures the dispersion of a portfolio's returns relative to the returns of its benchmark. That is, tracking error is the standard deviation of the portfolio's *active return*, where active return is defined as:

Active return = Portfolio's actual return – Benchmark's actual return

A portfolio created to match the benchmark (referred to as an index fund) that regularly has zero active returns (i.e., always matches its benchmark's actual return) would have a tracking error of zero. But an actively managed portfolio that takes positions substantially different from the benchmark would likely have large active returns, both positive and negative, and thus would have an annual tracking error of, say, 5% to 10%. By taking positions that differ from the benchmark is where the portfolio manager is making bets. For example, in common stock portfolio management this could involve one or more of the following factors: portfolio sensitivity to the benchmark (referred to as the portfolio beta), sector allocations that differ from the benchmark, style tilt (i.e., value versus growth stocks) that differs from the benchmark, and individual stock selections whose weight in the portfolio differs from that of the benchmark.

There are two types of tracking error: backward-looking tracking error and forward-looking tracking error. The former is obtained from a straightforward calculation based on the historical returns of a portfolio over some period of time. For example, suppose 52 weeks are computed for a portfolio return and the benchmark. An active return can then be calculated for each week and the annualized standard deviation can be calculated. The result is the backwark-looking tracking error. This tracking error, also referred to as an ex-post tracking error, is the result of the portfolio manager's decisions during those 52 weeks with respect to portfolio positioning issues.

One problem with a backward-looking tracking error is that it does not reflect the effect of current decisions by the portfolio manager on the future active returns and hence the future tracking error that may be realized. If, for example, the portfolio manager significantly changes the portfolio beta or sector allocations today, then the backward-looking tracking error that is calculated using data from prior periods would not accurately reflect the current portfolio risks going forward. That is, the backward-looking tracking error will have little predictive value and can be misleading regarding the portfolio's risks going forward.

The portfolio manager needs a forward-looking estimate of tracking error to more accurately reflect the portfolio risk going forward. The way this is done in practice is by using factor risk models, discussed in Chapter 12, that have defined the risks associated with a benchmark. Financial econometric tools analyzing the historical return data of the stocks in the benchmark index are used to obtain the factors and quantify their risks. Using the portfolio manager's current portfolio holdings, the portfolio's current exposure to the various factors can be calculated and compared to the benchmark's exposures to the same factors. Using the differential factor exposures and the risks of the factors, a forward-looking tracking error for the portfolio can be computed. This tracking error is also referred to as the predicted tracking error or ex-ante tracking error.

KEY POINTS

- Financial econometrics is the science of modeling and forecasting financial data.
- The three steps in applying financial econometrics are model selection, model estimation, and model testing.

- In model selection, the modeler chooses a family of models with given statistical properties. Financial economic theory is used to justify the model choice. The financial econometric tool used is determined in this step.
- Data mining is an approach to model selection based solely on the data and, although useful, must be used with great care because the risk is that the model selected might capture special characteristics of the sample which will not repeat in the future.
- In general, models are embodied in mathematical expressions that include a number of parameters that have to be estimated from sample data. Model estimation involves finding estimators and understanding the behavior of estimators.
- Model testing is needed because model selection and model estimation are performed on historical data and, as a result, there is the risk that the estimation process captures characteristics that are specific to the sample data used but are not general and will not necessarily reappear in future samples.
- Model testing involves assessing the model's performance using fresh data. The procedure for doing so is called backtesting and the most popular way of doing so is using a moving window.
- The data generating process refers to the mathematical model that represents future data in function of past and present data. By knowing the data generating process as a mathematical expression, computer programs that simulate data using Monte Carlo methods can be implemented and the data generated can be used to compute statistical quantities that would be difficult or even impossible to compute mathematically.
- Financial econometric techniques have been used in the investment management process for making decisions regarding asset allocation (i.e., allocation of funds among the major asset classes) and portfolio construction (i.e., selection of individual assets within an asset class). In addition, the measurement of portfolio risk with respect to risk factors that are expected to impact the performance of a portfolio relative to a benchmark are estimated using financial econometric techniques.