

See everything available through the O'Reilly learning platform and sta

Search

Probabilistic Machine Learning for Finance and Investing by Deepak K. ...

[BUY ON AMAZON](#)

[BUY ON EBOOKS.COM](#)

Chapter 1. The Need for Probabilistic Machine Learning

Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind.

—George Box, eminent statistician

A map will enable you to go from one geographic location to another. It is a very useful mathematical model for navigating the physical world. It becomes even more useful if you automate it into a GPS system using artificial intelligence (AI) technologies. However, neither the mathematical model nor the AI-powered GPS system will ever be able to capture the human experience and richness of the terrain it represents. That's because all models have to simplify the complexities of the real world, thus enabling us to focus on some of the features of a phenomenon that interest us.

George Box, an eminent statistician, famously said, "all models are wrong, but some are useful." This deeply insightful quip is our mantra. We accept that all models are wrong because they are inadequate and incomplete representations of reality. Our goal is to build financial systems based on models and supporting

All financial models, whether derived theoretically or discovered empirically by humans and machines, are not only wrong but are also at the mercy of three types of errors. In this chapter, we explain this trifecta of errors with an example from consumer credit and explore it using Python code. This exemplifies our claim that inaccuracies of financial models are features, not bugs. After all, we are dealing with people, not particles or pendulums.

Finance is not an accurate physical science like physics, dealing with precise estimates and predictions, as academia will have us believe. It is an inexact social study grappling with a range of values with varying plausibilities that change continually, often abruptly.

We conclude the chapter by explaining why AI in general and probabilistic machine learning (ML) in particular offers the most useful and promising theoretical framework and technologies for developing the next generation of systems for finance and investing.

WHAT IS A MODEL?

AI systems are based on models. A model maps functional relationships among its inputs and outputs variables based on assumptions and constraints. In general, input variables are called independent variables and output variables are called dependent variables.

In high school, you learned that the equation of any line in the XY plane can be expressed as $y = mx + b$, where m is the slope and b is the y-intercept of the line. For example, if you assume that consumer spending—the output/dependent variable y —has a linear relationship with personal income—the input/independent variable x —the equation for the line is called a model for consumer spending. Moreover, the slope m and the intercept b are referred to as the model's parameters. They are treated as constants, and their specific values define unique functional relationships or models.

Depending on the type of functional relationships, the parameters, and the nature of inputs and outputs variables, models may be classified as deterministic or probabilistic. In a deterministic model, there are no uncertainties about the type of functional relationships, the parameters, or the inputs or outputs of the model. The exact opposite is true for probabilistic models discussed in this book.

Finance Is Not Physics

Adam Smith, generally recognized as the founder of modern economics, was in awe of Newton's laws of mechanics and gravitation.¹ Since then, economists have endeavored to make their discipline into a mathematical science like physics. They aspire to formulate theories that accurately explain and predict the economic activities of human beings at the micro and macro levels. This desire gathered momentum in the early 20th century with economists like Irving Fisher and culminated in the econophysics movement of the late 20th century.

Despite all the complicated mathematics of modern finance, its theories are woefully inadequate, almost pitiful, especially when compared to those of physics. For instance, physics can predict the motion of the moon and the electrons in your computer with jaw-dropping precision. These predictions can be calculated by any physicist, at any time, anywhere on the planet. By contrast, market participants—traders, investors, analysts, finance executives—have trouble explaining the causes of daily market movements or predicting the price of an asset at any time, anywhere in the world.

THE POLITICAL ECONOMICS OF MISREPRESENTING a Nobel Prize

In his will, Alfred Nobel did not create a prize in economics or mathematics or any other discipline besides physics, chemistry, medicine, literature, and peace. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, now commonly and mistakenly referred to as the Nobel Prize in Economics, was created by the Swedish Central Bank in 1968. The central bank funds the award in perpetuity and pays the Nobel Foundation to administer it like it does the Nobel prizes willed by its benefactor.

By elevating the status of economics to that of the natural sciences and by buying the ongoing support of the prestigious Nobel Foundation, the Swedish central bank was able to gain independence in its decision making from the country's politicians to pursue its market-friendly policies. Economic policy decisions were to be left to the economic "scientists," just as health policy decisions were left to medical scientists.² However, by doing this, the Foundation disregards the will of Alfred Nobel and misrepresents the fundamental nature of economics as a social science.

In his 1974 acceptance speech, Friedrich Hayek, a pioneer of libertarian economics and advocate for free-market policies, clearly understood how the newly established economics prize could be misused when he said, "The Nobel Prize confers on an individual an authority which in economics no man ought to possess...This does not matter in the natural sciences. Here the influence exercised by an individual is chiefly an influence on his fellow experts; and they will soon cut him down to size if he exceeds his competence. But the influence of the economist that mainly matters is an influence over laymen: politicians, journalists, civil servants and the public generally."³

Perhaps finance is harder than physics. Unlike particles and pendulums, people are complex, emotional, creative beings with free will and latent cognitive biases. They tend to behave inconsistently and continually react to the actions of others in unpredictable ways. Furthermore, market participants profit by beating or gaming the systems that they operate in.

After losing a fortune on his investment in the South Sea Company, Newton remarked, "I can calculate the movement of the stars, but not the madness of men."⁴ Note that Newton was not a novice investor. He served as the warden of the Mint in England for almost 31 years, helping put the British pound on the gold standard, where it would stay for over two centuries.

wrong but also dangerous. The veneer of a physical science lulls adherents of economic models into a false sense of certainty about the accuracy of their predictive powers.⁵ This blind faith has led to many disastrous consequences for their adherents and for society at large.⁶ Nothing better exemplifies the dangerous consequences of academic arrogance and blind faith in analytical financial models than the spectacular disaster of LTCM, discussed in the sidebar.

The Disaster of Long-Term Capital Management (LTCM)

LTCM was a hedge fund founded in 1994 by Wall Street veterans and academics Myron Scholes and Robert Merton, inventors of the famous Black-Scholes-Merton option pricing formula. The LTCM team was so confident in its investment models, overseen by two future "Nobel laureates," that it leveraged its portfolios to dangerously high levels. The team intended to magnify the tiny profits that LTCM was making on its various investments. In the first four years, LTCM had very impressive annual returns and had to turn away investor money.

However, the unpredictable complexity of social systems reared its ugly head in 1998, when the Russian government defaulted on its domestic local currency bonds. Such an event was not anticipated by LTCM's models, since a government can always print more money rather than default on its debt. This shocked global markets and led to the rapid collapse of LTCM—and showed that leverage magnifies losses, as it does gains. To prevent the crisis of LTCM from spreading and crashing the global financial markets, the Federal Reserve and a consortium of large banks bailed out LTCM. See [Figure 1-1](#), which compares the value of \$1,000 invested separately in LTCM, Dow Jones (DJIA), and US Treasury bonds.

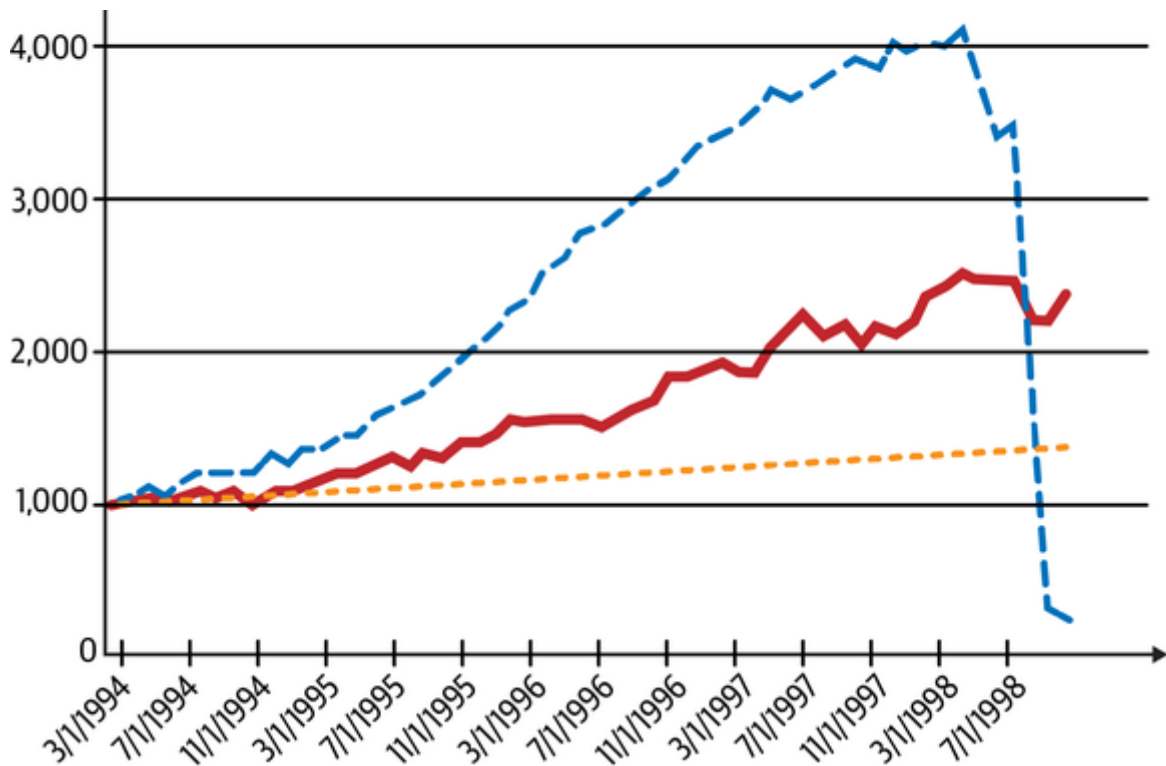


Figure 1-1. The epic disaster of Long Term Capital Management (LTCM)⁷

Taking a diametrically different approach from hedge funds like LTCM, Renaissance Technologies, the most successful hedge fund in history, has put its critical views of financial theories into practice. Instead of hiring people with a finance or Wall Street background, the company prefers to hire physicists, mathematicians, statisticians, and computer scientists. It trades the markets using quantitative models based on nonfinancial theories such as information theory, data science, and machine learning.

The Trifecta of Modeling Errors

Whether financial models are based on academic theories or empirical data-mining strategies, they are all subject to the trifecta of modeling errors. Errors in analysis and forecasting may arise from any of the following modeling issues: using an inappropriate functional form, inputting inaccurate parameters, or failing to adapt to structural changes in the market.⁸

Errors in Model Specification

Almost all financial theories use the Gaussian or normal distribution in their mod-

bonds, currencies, and commodities have fat-tailed return distributions that are distinctly non-Gaussian.¹⁰ In other words, extreme events occur far more frequently than predicted by the normal distribution. In [Chapter 3](#) and [Chapter 4](#), we will actually do financial data analysis in Python to demonstrate the non-Gaussian structure of equity return distributions.

If asset price returns were normally distributed, none of the following financial disasters would occur within the age of the universe: Black Monday, the Mexican peso crisis, the Asian currency crisis, the bankruptcy of LTCM, or the Flash Crash. “Mini flash crashes” of individual stocks occur with even higher frequency than these macro events.

Yet, finance textbooks, programs, and professionals continue to use the normal distribution in their asset valuation and risk models because of its simplicity and analytical tractability. These reasons are no longer justifiable given today's advanced algorithms and computational resources. This reluctance to abandon the normal distribution is a clear example of “the drunkard's search”: a principle derived from a joke about a drunkard who loses his key in the darkness of a park but frantically searches for it under a lamppost because that's where the light is.

Errors in Model Parameter Estimates

Errors of this type may arise because market participants have access to different levels of information with varying speeds of delivery. They also have different levels of sophistication in processing abilities and different cognitive biases. Moreover, these parameters are generally estimated from past data, which may not represent current market conditions accurately. These factors lead to profound epistemic uncertainty about model parameters.

Let's consider a specific example of interest rates. Fundamental to the valuation of any financial asset, interest rates are used to discount uncertain future cash flows of the asset and estimate its value in the present. At the consumer level, for example, credit cards have variable interest rates pegged to a benchmark called the prime rate. This rate generally changes in lockstep with the federal funds rate, an

one year from now. Suppose the current prime rate is 2% and your credit card company charges you 10% plus prime. Given the strength of the current economy, you believe that the Federal Reserve is more likely to raise interest rates than not. Based on our current information, we know that the Fed will meet eight times in the next 12 months and will either raise the federal funds rate by 0.25% or leave it at the previous level.

In the following Python code example, we use the binomial distribution to model your credit card's interest rate at the end of the 12-month period. Specifically, we'll use the following parameters for our range of estimates about the probability of the Fed raising the federal funds rate by 0.25% at each meeting: `fed_meetings = 8` (number of trials or meetings); `probability_raises = [0.6, 0.7, 0.8, 0.9]`:

```
# Import binomial distribution from scipy library
from scipy.stats import binom
# Import matplotlib library for drawing graphs
import matplotlib.pyplot as plt

# Total number of meetings of the Federal Open Market Committee (FOMC) in any
# year
fed_meetings = 8
# Range of total interest rate increases at the end of the year
total_increases = list(range(0, fed_meetings + 1))
# Probability that the FOMC will raise rates at any given meeting
probability_raises = [0.6, 0.7, 0.8, 0.9]

fig, axs = plt.subplots(2, 2, figsize=(10, 8))

for i, ax in enumerate(axs.flatten()):
    # Use the probability mass function to calculate probabilities of total
    # raises in eight meetings
    # Based on FOMC bias for raising rates at each meeting
    prob_dist = binom.pmf(k=total_increases, n=fed_meetings,
                          p=probability_raises[i])
    # How each 25 basis point increase in the federal funds rate affects your
    # credit card interest rate
    cc_rate = [j * 0.25 + 12 for j in total_increases]
```



```
ax.legend()  
ax.set_ylabel('Probability of credit card rate')  
ax.set_xlabel('Predicted range of credit card rates after 12 months')  
ax.set_title(f'Probability of raising rates at each meeting:  
{probability_raises[i]}')
```

```
# Adjust spacing between subplots  
plt.tight_layout()
```

```
# Show the plot  
plt.show()
```

In [Figure 1-2](#), notice how the probability distribution for your credit card rate in 12 months depends critically on your estimate about the probability of the Fed raising rates at each of the eight meetings. You can see that for every increase of 0.1 in your estimate of the Fed raising rates at each meeting, the expected interest rate for your credit card in 12 months increases by about 0.2%.

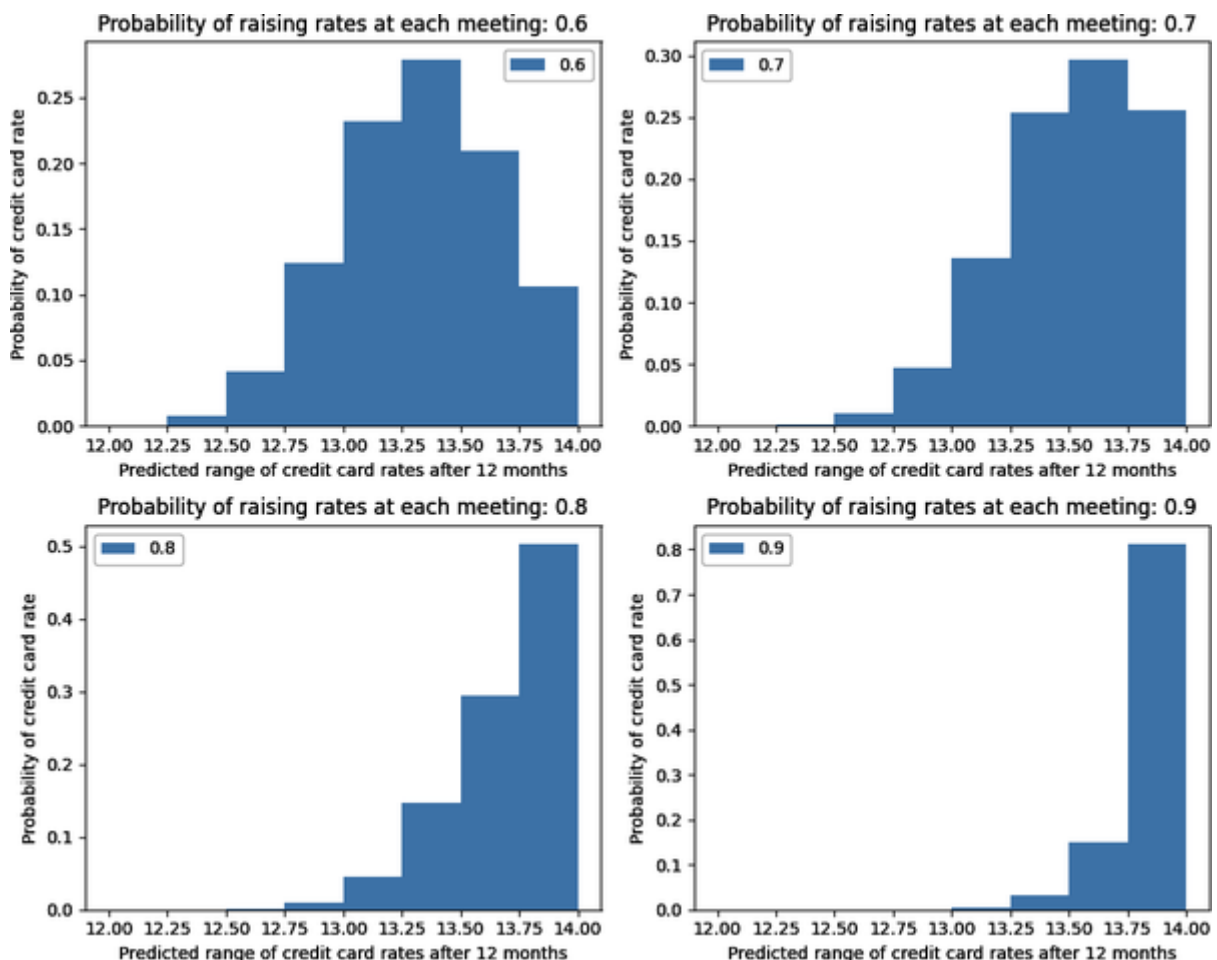


Figure 1-2. Probability distribution of credit card rates depends on your parameter estimates

this parameter is hard to estimate. Many institutions have dedicated analysts, including previous employees of the Fed, analyzing the Fed's every document, speech, and event to try to estimate this parameter. This is because the Fed funds rate directly impacts the prices of all financial assets and indirectly impacts the employment and inflation rates in the real economy.

Recall that we assumed that this parameter, `probability_raises`, was constant in our model for each of the next eight Fed meetings. How realistic is that? Members of the Federal Open Market Committee (FOMC), the rate-setting body, are not just a set of biased coins. They can and do change their individual biases based on how the economy changes over time. The assumption that the parameter `probability_raises` will be constant over the next 12 months is not only unrealistic, but also risky.

Errors from the Failure of a Model to Adapt to Structural Changes

The underlying data-generating stochastic process may vary over time—i.e., the process is not stationary ergodic. This implies that statistical moments of the distribution, like mean and variance, computed from sample financial data taken at a specific moment in time or sampled over a sufficiently long time period do not accurately predict the future statistical moments of the underlying distribution. The concepts of stationarity and ergodicity are very important in finance and will be explained in more detail later in the book.

We live in a dynamic capitalist economy characterized by technological innovations and changing monetary and fiscal policies. Time-variant distributions for asset values and risks are the rule, not the exception. For such distributions, parameter values based on historical data are bound to introduce error into forecasts.

In our previous example, if the economy were to show signs of slowing down, the Fed might decide to adopt a more neutral stance in its fourth meeting, making you change your `probability_raises` parameter from 70% to 50% going forward. This change in your parameter will, in turn, change the forecast of your credit card

ously or abruptly, as in the Mexican peso crisis. For either continuous or abrupt changes, the models used will need to adapt to evolving market conditions. A new functional form with different parameters might be required to explain and predict asset values and risks in the new market regime.

Suppose after the fifth meeting in our example, the US economy is hit by an external shock—say a new populist government in Greece decides to default on its debt obligations. Now the Fed may be more likely to cut interest rates than to raise them. Given this structural change in the Fed's outlook, we will have to change the binomial probability distribution in our model to a trinomial distribution with appropriate parameters.

Probabilistic Financial Models

Inaccuracies of financial models are features, not bugs. It is intellectually dishonest and foolishly risky to represent financial estimates as scientifically precise values. All models should quantify the uncertainty inherent in financial inferences and predictions to be useful for sound decision making and risk management in the business world. Financial data are noisy and have measurement errors. A model's appropriate functional form may be unknown or an approximation. Model parameters and outputs may have a range of values with associated plausibilities. In other words, we need mathematically sound probabilistic models because they accommodate inaccuracies and quantify uncertainties with logical consistency.

There are two ways model uncertainty is currently quantified: forward propagation for output uncertainty, and inverse propagation for input uncertainty. [Figure 1-3](#) shows the common types of probabilistic models used in finance today for quantifying both types of uncertainty.

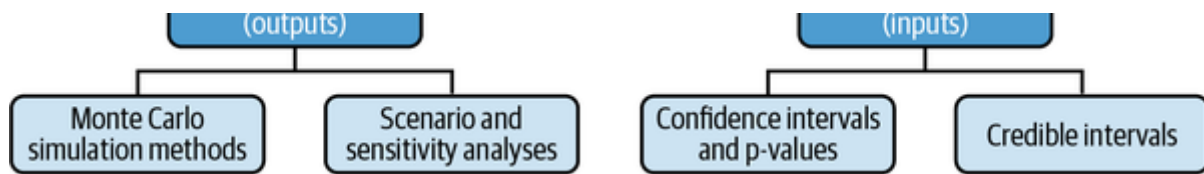


Figure 1-3. Quantifying input and output uncertainty with probabilistic models

In forward uncertainty propagation, uncertainties arising from a model's inexact parameters and inputs are propagated forward throughout the model to generate the uncertainty of the model's outputs. Most financial analysts use scenario and sensitivity analyses to quantify the uncertainty in their models' predictions. However, these are basic tools that only consider a few possibilities.

In scenario analysis, only three cases are built for consideration: best-case, base-case, and worst-case scenarios. Each case has a set value for all the inputs and parameters of a model. Similarly, in sensitivity analysis, only a few inputs or parameters are changed to assess their impact on the model's total output. For instance, a sensitivity analysis might be conducted on how the value of a company changes with interest rates or future earnings. In [Chapter 3](#), we will learn how to perform Monte Carlo simulations (MCS) using Python and apply it to common financial problems. MCS is one of the most powerful probabilistic numerical tools in all the sciences and is used for analyzing both deterministic and probabilistic systems. It is a set of numerical methods that uses independent random samples from specified input parameter distributions to generate new data that we might observe in the future. This enables us to compute the expected uncertainty of a model, especially when its functional relationships are not analytically tractable.

In inverse uncertainty propagation, uncertainty of the model's input parameters is inferred from observed data. This is a harder computational problem than forward propagation because the parameters have to be learned from the data using dependent random sampling. Advanced statistical inference techniques or complex numerical computations are used to calculate confidence intervals or credible intervals of a model's input parameters. In [Chapter 4](#), we explain the deep flaws and limitations of using p-values and confidence intervals, statistical techniques that are commonly used in financial data analysis today. Later in [Chapter 6](#), we explain Markov chain Monte Carlo, an advanced, dependent, random sampling method, which can be used to compute credible intervals to quantify the uncertainty of a

and inverse uncertainty propagation seamlessly. We don't want the piecemeal approach that is currently in practice today. That is, we want our probabilistic models to quantify the uncertainty in the outputs of time-variant stochastic processes, with their inexact input parameters learned from sample data.

Our probabilistic framework will need to update continually the model outputs or its input parameters—or both—based on materially new datasets. Such models will have to be developed using small datasets, since the underlying environment may have changed too quickly to collect a sizable amount of relevant data. Most importantly, our probabilistic models need to know what they don't know. When extrapolating from datasets they have never encountered before, they need to provide answers with low confidence levels or wider margins of uncertainty.

Financial AI and ML

Probabilistic machine learning (ML) meets all the previously mentioned requirements for building state-of-the-art, next-generation financial systems.¹¹ But what is probabilistic ML? Before we answer that question, let's first make sure we understand what we mean by ML in particular and AI in general. It is common to see these terms bandied about as synonyms, even though they are not. ML is a subfield of AI. See [Figure 1-4](#).

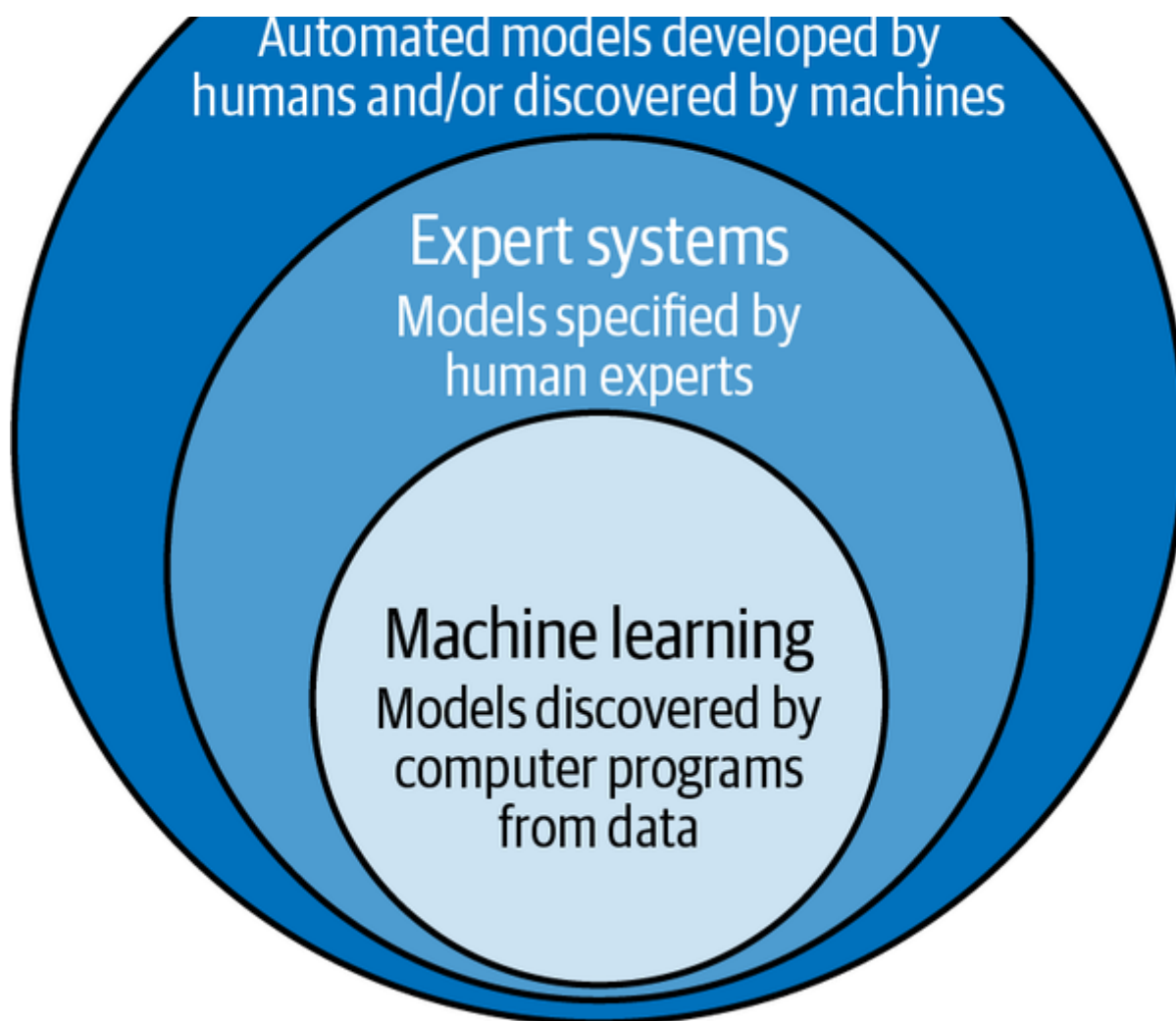


Figure 1-4. ML is a subfield of AI

AI is the general field that tries to automate the cognitive abilities of humans, such as analytical thinking, decision making, and sensory perception. In the 20th century, computer scientists developed a subfield of AI called symbolic AI (SAI), which included methodologies and tools to embed into computer systems, symbolic representations of human knowledge in the form of well-defined rules or algorithms.

SAI systems automate the models specified by domain experts and are aptly called expert systems. For instance, traders, finance executives, and system developers work together to explicitly formulate all the rules and the model's parameters that are to be automated by their financial and investment management systems. I have managed several such projects for marquee financial institutions at one of my previous companies.

However, SAI failed in automating complex tasks like image recognition and natural language processing—technologies used extensively in corporate finance and

rithms, abundant data, and cheap computing resources.

ML turns the SAI paradigm on its head. Instead of experts specifying models to process data, humans with little or no domain expertise provide general-purpose algorithms that learn a model from data samples. More importantly, ML programs continually learn from new datasets and update their models without any human intervention for code maintenance. See the next sidebar for a simple explanation of how parameters are learned from data.

Training a Linear ML System to Learn

Recall the deterministic linear model discussed earlier and expressed by the equation $y = mx + b$. A unique line crosses at least two distinct points in the XY plane. The two points enable us to solve analytically for the exact values of parameters m and b using simple algebra. Once you have computed the parameters, you can use your model to make accurate predictions; given any point x , you can predict exactly what y will be.

However, financial models are not deterministic but probabilistic. For instance, if you were to plot a company's stock price returns on the y -axis and the growth rate of its quarterly earnings on the x -axis, you would see stock returns generally increase with earnings growth of a company. If you assume the relationship between stock price returns and quarterly earnings growth is *approximately* linear, you can use an analytical statistical technique to solve for the model's parameters m and b that gives you the line that best fits the company's sample data. If the linear approximation persists in the future, your model's predictions are not going to be precise, but they are going to be better than making random guesses or relying on luck.

Alternatively, you could use ML software to do similar calculations for you. In ML systems, the independent variable x is called a feature or predictor, and the dependent variable y is called the target or response variable. Feeding sample data to the ML system is referred to as training the system. When our linear ML system computes the values of the parameters m and b , we say that the ML system has learned the model from the in-sample data. The objective in ML is to predict the target values on out-of-sample data, which the system has not been trained on. This is where predictions become challenging.

We will get into the details of modeling, training, and testing probabilistic ML systems in the second half of the book. Here is a useful definition of ML from Tom

Figure 1-5.

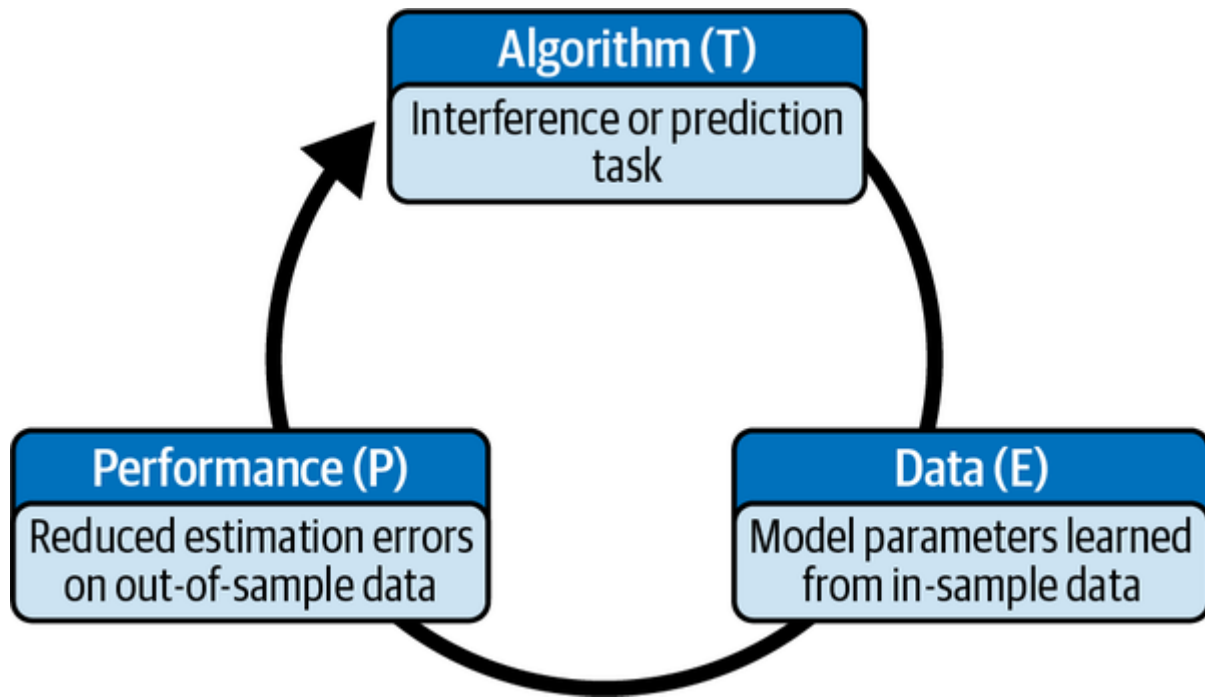


Figure 1-5. An ML model learns its parameters from in-sample data, but its performance is evaluated on out-of-sample data

Performance is measured against a prespecified objective function, such as maximizing annual stock price returns or lowering the mean absolute error of parameter estimates.

ML systems are usually classified into three types based on how much assistance they need from their human teachers or supervisors.

Supervised learning

ML algorithms learn functional relationships from data, which are provided in pairs of inputs and desired outputs. This is the most prevalent form of ML used in research and industry. Some examples of ML systems include linear regression, logistic regression, random forests, gradient-boosted machines, and deep learning.

Unsupervised learning

ML algorithms are only given input data and learn structural relationships in the data on their own. The K-means clustering algorithm is a commonly used data exploration algorithm used by investment

Reinforcement learning

An ML algorithm continually updates a policy or set of actions based on feedback from its environment with the goal of maximizing the present value of cumulative rewards. It's different from supervised learning in that the feedback signal is not a desired output or class, but a reward or penalty. Examples of algorithms are Q-learning, deep Q-learning, and policy gradient methods. Reinforcement learning algorithms are being used in advanced trading applications.

In the 21st century, financial data scientists are training ML algorithms to discover complex functional relationships using data from multiple financial and nonfinancial sources. The newly discovered relationships may augment or replace the insights of finance and investment executives. ML programs are able to detect patterns in very high-dimensional datasets, a feat that is difficult if not impossible for humans. They are also able to reduce the dimensions to enable visualizations for humans.

AI is used in all aspects of the finance and investment process—from idea generation to analysis, execution, portfolio, and risk management. The leading AI-powered systems in finance and investing today use some combination of expert systems and ML-based systems by leveraging the advantages of both types of approaches and expertise. Furthermore, AI-powered financial systems continue to leverage human intelligence (HI) for research, development, and maintenance. Humans may also intervene in extreme market conditions, where it may be difficult for AI systems to learn from abrupt changes. So you can think of modern financial systems as a complex combination of SAI + ML + HI.

Probabilistic ML

Probabilistic ML is the next-generation ML framework and technology for AI-powered financial and investing systems. Leading technology companies clearly understand the limitations of conventional AI technologies and are developing their probabilistic versions to extend their applicability to more complex problems.

ML technologies are PyMC and Stan. PyMC is written in Python, and Stan is written in C++. In Chapter 7, we use the PyMC library because it's part of the Python ecosystem.

Probabilistic ML as discussed in this book is based on a generative model. It is categorically different from the conventional ML in use today, such as linear, nonlinear, and deep learning systems, even though these other systems compute probabilistic scores. Figure 1-6 shows the major differences between the two types of systems.

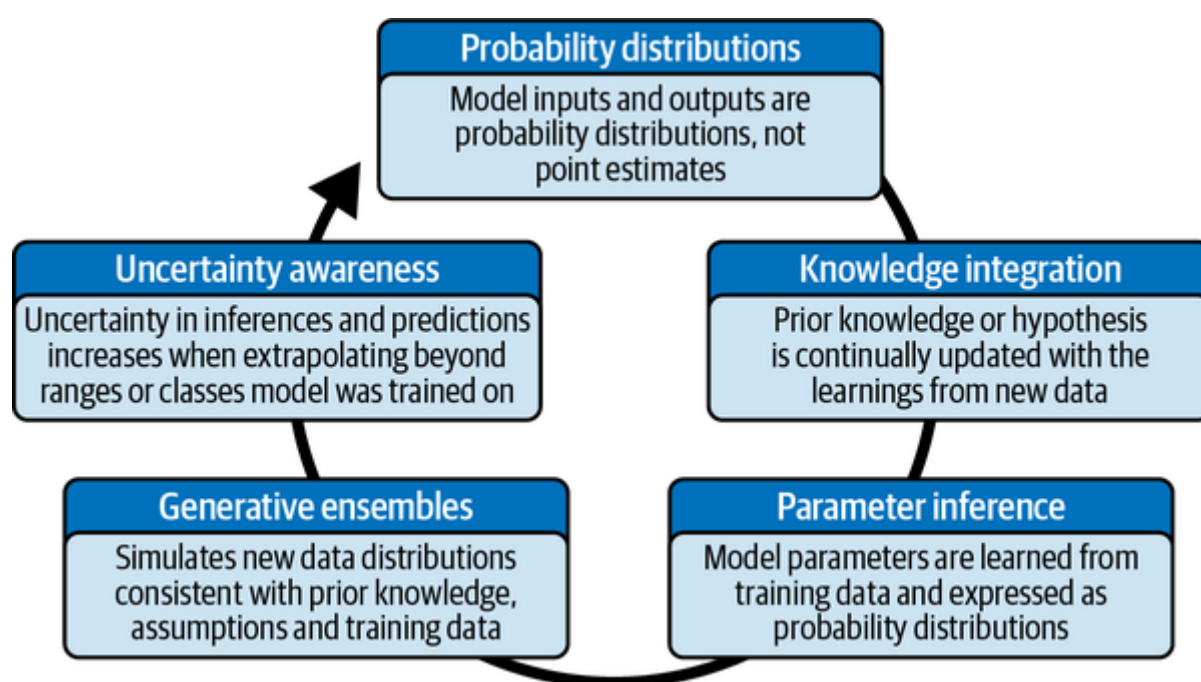


Figure 1-6. Summary of major characteristics of probabilistic ML systems

Probability Distributions

Even though conventional ML systems use calibrated probabilities, they only compute the most likely estimates and their associated probabilities as single-point values for inputs and outputs. This works well for domains, such as image recognition, where the data are plentiful and the signal-to-noise ratio is high. As was discussed and demonstrated in the previous sections, a point estimate is an inaccurate and misleading representation of financial reality, where uncertainty is very high. Furthermore, the calibrated probabilities may not be valid probabilities as the unconditional probability distribution of the data is almost never computed by MLE

Probabilistic ML systems only deal in probability distributions in their computations of input parameters and model outputs. This is a realistic and honest representation of the uncertainty of a financial model's variables. Furthermore, probability distributions leave the user considerable flexibility in picking the appropriate point estimate, if required, based on their business objectives.

Knowledge Integration

Conventional ML systems do not have a theoretically sound framework for incorporating prior knowledge, whether it is well-established scientific knowledge, institutional knowledge, or personal insights. Later in the book, we will see that conventional statisticians sneak in prior knowledge using ad hoc statistical methods, such as null hypothesis, statistical significance levels, and L1 and L2 regularizations, while pounding the table about letting only "the data speak for themselves."

It is foolish not to integrate prior knowledge in our personal and professional lives. It is the antithesis of learning and vitiates against the nature of the scientific method. Yet this is the basis of null hypothesis significance testing (NHST), the prevailing statistical methodology in academia, research, and industry since the 1960s. NHST prohibits the inclusion of prior knowledge in experiments based on the bogus claim that objectivity demands that we only let the data speak for themselves. By following this specious claim, NHST ends up committing the prosecutor's fallacy, as we will show in [Chapter 4](#).

NHST's definition of objectivity would require us to touch fire everywhere and every time we find it because we cannot incorporate our prior knowledge of what it felt like in similar situations in the past. That is the definition of foolishness, not objectivity. In [Chapter 4](#), we will discuss how and why several metastudies have shown that the majority of published medical research findings based on NHST are false. Yes, you read that right, and it has been an open secret since a seminal paper published in 2005.¹³

Fortunately, in this book we don't have to waste much ink or pixels on this specious argument about objectivity or the proliferation of junk science produced by

one can challenge it or change it. This is the essence of learning and the basis of the scientific method.

One of the important implications of the no free lunch (NFL) theorems is that prior domain knowledge is necessary to optimize an algorithm's performance for a specific problem domain. If we don't apply our prior domain knowledge, the performance of our unbiased algorithm will be no better than random guessing when averaged across all problem domains. There is no such thing as a free lunch, especially in finance and investing. We will discuss the NFL theorems in detail in the next chapter.

It is common knowledge that integration of accumulated institutional knowledge into a company's organization, process, and systems leads to a sustainable competitive advantage in business. Moreover, personal insights and experience with markets can lead to "alpha," or the generation of exceptional returns in trading and investing, for the fund manager who arrives at a subjectively different viewpoint from the rest of the crowd. That's how Warren Buffet, one of the greatest investors of all time, made his vast fortune. Markets mock dogmatic and unrealistic definitions of objectivity with lost profits and eventually with financial ruin.

Parameter Inference

Almost all conventional ML systems use equally conventional statistical methodologies, such as p-values and confidence intervals, to estimate the uncertainty of a model's parameters. As will be explained in [Chapter 4](#), these are deeply flawed—almost scandalous—statistical methodologies that plague the social sciences, including finance and economics. These methodologies adhere to a pious pretense to objectivity and to implicit and unrealistic assumptions, obfuscated by inscrutable statistical jargon, in order to generate solutions that are analytically tractable for a small set of scenarios.

Probabilistic ML is based on a simple and intuitive definition of probability as logic, and the rigorous calculus of probability theory in general and the inverse probability rule in particular. In the next chapter, we show how the inverse probability

it is to derive most mathematical theorems.

However, because of the normalizing constant in the inversion formula, it was previously impossible to invert probabilities analytically, except for simple problems. With the recent advancement of state-of-the-art numerical algorithms, such as Hamiltonian Monte Carlo and automatic differentiation variational inference, probabilistic ML systems are now able to invert probabilities to compute model parameter estimates from in-sample data for almost any real-world problem. More importantly, they are able to quantify parameter uncertainties with mathematically sound credible intervals for any level of confidence. This enables inverse uncertainty propagation.

Generative Ensembles

Almost all conventional ML systems are based on discriminative models. This type of statistical model only learns a decision boundary from the in-sample data, but not how the data are distributed statistically. Therefore, conventional discriminative ML systems cannot simulate new data and quantify total output uncertainty.

Probabilistic ML systems are based on generative models. This type of statistical model learns the statistical structure of the data distribution and so can easily and seamlessly simulate new data, including generating data that might be missing or corrupted. Furthermore, the distribution of parameters generates an ensemble of models. Most importantly, these systems are able to simulate two-dimensional output uncertainty based on data variability and input parameter uncertainty, the probability distributions of which they have learned previously from in-sample data. This seamlessly enables forward uncertainty propagation.

Uncertainty Awareness

When computing probabilities, a conventional ML system uses the maximum likelihood estimation (MLE) method. This technique optimizes the parameters of an assumed probability distribution such that the in-sample data are most likely to be observed, given the point estimates for the model's parameters. As we will see

What makes it worse is that these MLE-based ML systems attach horrifyingly high probabilities to these wrong estimates. We are automating the overconfidence of powerful systems that lack basic common sense. This makes conventional ML systems potentially risky and dangerous, especially when used in mission-critical operations by personnel who either don't understand the fundamentals of these ML systems or have blind faith in them.

Probabilistic ML systems do not rely on a single-point estimate, no matter how likely or optimal, but a weighted average of every possible estimate of a parameter's entire probability distribution. Moreover, the uncertainty of these estimates increases appropriately when systems deal with classes of data they have never seen before in training, or are extrapolating beyond known data ranges. Unlike MLE-based systems, probabilistic ML systems know what they don't know. This keeps the quantification of uncertainty honest and prevents overconfidence in estimates and predictions.

Summary

Economics is not a precise predictive science like physics. Not even close. So let's not pretend otherwise and treat academic theories and models of economics as if they were models of quantum physics, the obfuscating math notwithstanding.

All financial models, whether based on academic theories or ML strategies, are at the mercy of the trifecta of modeling errors. While this trio of errors can be mitigated with appropriate tools, such as probabilistic ML systems, it cannot be eliminated. There will always be asymmetry of information and cognitive biases. Models of asset values and risks will change over time due to the dynamic nature of capitalism, human behavior, and technological innovation.

Probabilistic ML technologies are based on a simple and intuitive definition of probability as logic and the rigorous calculus of probability theory. They enable the explicit and systematic integration of prior knowledge that is updated continually with new learnings.

tions, not point estimates. This makes for realistic financial inferences and predictions that are useful for decision making and risk management. Most importantly, these systems become capable of forewarning us when their inferences and predictions are no longer useful in the current market environment.

There are several reasons why probabilistic ML is the next-generation ML framework and technology for AI-powered financial and investing systems. Its probabilistic framework moves away from flawed statistical methodologies (NHST, p-values, confidence intervals) and the restrictive conventional view of probability as a limiting frequency. It moves us toward an intuitive view of probability as logic and a mathematically rigorous statistical framework that quantifies uncertainty holistically and successfully. Therefore, it enables us to move away from the wrong, idealistic, analytical models of the past toward less wrong, more realistic, numerical models of the future.

The algorithms used in probabilistic programming are among the most sophisticated algorithms in the AI world, which we will delve into in the second half of the book. In the next three chapters, we will take a deeper dive into why it is very risky to deploy your capital using conventional ML systems, because they are based on orthodox probabilistic and statistical methods that are scandalously flawed.

References

Géron, Aurélien. "The Machine Learning Landscape." In *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 1–34. 3rd ed. O'Reilly Media, 2022.

Hayek, Friedrich von. "Banquet Speech." Speech given at the Nobel Banquet, Stockholm, Sweden, December 10, 1974. Nobel Prize Outreach AB, 2023, <https://www.nobelprize.org/prizes/economic-sciences/1974/hayek/speech/>.

Ioannidis, John P. A. "Why Most Published Research Findings Are False." *PLOS Medicine* 2, no. 8 (2005): e124. <https://doi.org/10.1371/journal.pmed.0020124>.

Offer, Avner, and Gabriel Söderberg. *The Nobel Factor: The Prize in Economics*,

Orrell, David, and Paul Wilmott. *The Money Formula: Dodgy Finance, Pseudo Science, and How Mathematicians Took Over the Markets*. West Sussex, UK: Wiley, 2017.

Sekerke, Matt. *Bayesian Risk Management*. Wiley, 2015.

Simons, Katerina. "Model Error." *New England Economic Review* (November 1997): 17–28.

Thompson, J. R., L.S. Baggett, W. C. Wojciechowski, and E. E. Williams. "Nobels For Nonsense." *Journal of Post Keynesian Economics* 29, no. 1 (Autumn 2006): 3–18.

Further Reading

Jaynes, E. T. *Probability Theory: The Logic of Science*. New York: Cambridge University Press, 2003.

Lopez de Prado, Marcos. *Advances in Financial Machine Learning*. Hoboken, New Jersey: Wiley, 2018.

Taleb, Nassim Nicholas. *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*. New York: Random House Trade, 2005.

¹ David Orrell and Paul Wilmott, "Going Random," in *The Money Formula: Dodgy Finance, Pseudo Science, and How Mathematicians Took Over the Markets* (West Sussex, UK: Wiley, 2017).

² Avner Offer and G. Söderberg, *The Nobel Factor: The Prize in Economics, Social Democracy, and the Market Turn* (Princeton, NJ: Princeton University Press, 2016).

³ Friedrich von Hayek, "Banquet Speech," Nobel Prize Outreach AB, 2023, <https://www.nobelprize.org/prizes/economic-sciences/1974/hayek/speech>.

⁴ David Orrell and Paul Wilmott, "Early Models," in *The Money Formula: Dodgy Finance, Pseudo Science, and How Mathematicians Took Over the Markets* (West

For Nonsense," *Journal of Post Keynesian Economics* 29, no. 1 (Autumn 2006): 3–18.

⁶ Orrell and Wilmott, *The Money Formula*.

⁷ Adapted from an image from Wikimedia Commons.

⁸ Orrell and Wilmott, *The Money Formula*; M. Sekerke, *Bayesian Risk Management* (Hoboken, NJ: Wiley, 2015); J. R. Thompson, L. S. Baggett, W. C. Wojciechowski, and E. E. Williams, "Nobels for Nonsense," *Journal of Post Keynesian Economics* 29, no. 1 (Autumn 2006): 3–18; and Katerina Simons, "Model Error," *New England Economic Review* (November 1997): 17–28.

⁹ Orrell and Wilmott, *The Money Formula*; Sekerke, *Bayesian Risk Management*; and Thompson, Baggett, Wojciechowski, and Williams, "Nobels for Nonsense."

¹⁰ Orrell and Wilmott, *The Money Formula*; Sekerke, *Bayesian Risk Management*; and Thompson, Baggett, Wojciechowski, and Williams, "Nobels for Nonsense."

¹¹ Sekerke, *Bayesian Risk Management*.

¹² Aurélien Géron, "The Machine Learning Landscape," in *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd edition (O'Reilly Media, 2022), 1–34.

¹³ The paper is John P. A. Ioannidis, "Why Most Published Research Findings Are False," *PLOS Medicine* 2, no. 8 (2005): e124, <https://doi.org/10.1371/journal.pmed.0020124>. See also Julia Belluz, "This Is Why You Shouldn't Believe That Exciting New Medical Study," *Vox*, February 27, 2017, <https://www.vox.com/2015/3/23/8264355/research-study-hype>.

Get *Probabilistic Machine Learning for Finance and Investing* now with the O'Reilly learning platform.

O'Reilly members experience books, live events, courses curated by job role, and more from O'Reilly and nearly 200 top publishers.

START YOUR FREE TRIAL

ABOUT O'REILLY

[Teach/write/train](#)

[Careers](#)

[Press releases](#)

[Media coverage](#)

[Community partners](#)

[Affiliate program](#)

[Submit an RFP](#)

[Diversity](#)

[O'Reilly for marketers](#)

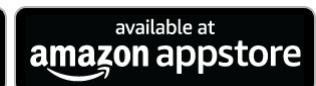
DOWNLOAD THE O'REILLY APP

Take O'Reilly with you and learn anywhere, anytime on your phone and tablet.



WATCH ON YOUR BIG SCREEN

View all O'Reilly videos, Superstream events, and Meet the Expert sessions on your home TV.



SUPPORT

[Contact us](#)

[Newsletters](#)

[Privacy policy](#)



INTERNATIONAL

[Australia & New Zealand](#)

[Hong Kong & Taiwan](#)

[India](#)

[Indonesia](#)

[Japan](#)

[DO NOT SELL MY PERSONAL INFORMATION](#)

© 2024, O'Reilly Media, Inc. All trademarks and registered trademarks appearing on oreilly.com are the property of their respective owners.

We are a participant in the Amazon Services LLC Associates Program, an affiliate advertising program designed to provide a means for us to earn fees by linking to Amazon.com and affiliated sites.

[Terms of service](#) • [Privacy policy](#) • [Editorial independence](#)