Chapter 10
Model Risk and Uncertainty—Illustrated with Examples from Mathematical Finance

Karl F. Bannör and Matthias Scherer

Stochastic modeling techniques have become increasingly popular during the last decades, particularly in mathematical finance since the groundbreaking work of Bachelier (Théorie de la spéculation, Gauthier-Villars, Paris, 1900), Samuelson (Ind. Manag. Rev. 6(2):13–39, 1965), and Black and Scholes (J. Polit. Econ. 81(3):637–654, 1973). Essentially, all models are wrong in the sense that they simplify reality. However, there are numerous models available to model particular phenomena of financial markets and calculated option prices, hedging strategies, portfolio allocations, etc. depend on the chosen model. This gives rise to the question which model to choose from the rich pool of available models and, second, how to determine the correct parameters after having selected some specific model class. Thus, one is exposed to both model and parameter risk (or uncertainty). In this survey, we first provide an inside view into the principles of stochastic modeling, illustrated with examples from mathematical finance. Afterwards, we define model risk and uncertainty according to Knight (Risk, uncertainty, and profit, Hart, Schaffner & Marx, Chicago, 1921) and present some methods how to deal with model risk and uncertainty.

The Facts

- In natural sciences as physics, chemistry, and biology, laws of nature often support model building. In social sciences like economics, there may be no natural laws offering models.

K.F. Bannör (✉)
Deloitte & Touche GmbH, Rosenheimer Platz 4, 81669 Munich, Germany
e-mail: kbannoer@deloitte.de

M. Scherer
Mathematical Finance, Center for Mathematical Sciences, Technische Universität München, Boltzmannstrasse 3, 85748 Garching bei München, Germany

- Stochastic modeling tries to capture the stylized facts of the distribution of outcomes in concern.
- Often, there is considerable ambiguity which model (or, equivalently, which probability distribution) to choose.
- One distinguishes between model risk and model uncertainty, following the terminology of Knight [9].
- Model risk is a situation where one can quantify the likelihood of the validity of the different models to choose from, i.e. a probability distribution on the set of models is known.
- Model uncertainty is a situation where one does not have any additional information about the different models, i.e. a probability distribution on the set of models is unknown.

1 Stochastic Modeling of Real-World Phenomena

Die Theorie liefert viel, aber dem Geheimnis des Alten bringt sie uns doch nicht näher. Jedenfalls bin ich überzeugt davon, dass der nicht würfelt.[1]—*Albert Einstein, Nobel Laureate in Physics*

Models from classical mechanics, as illustrated in Chap. 4 of Mainzer [38], often describe effects that have fully been studied. Hence, a deterministic functional relationship can be taken as a mathematical model for description.[2] In contrast, there exist many real-world phenomena that exhibit deterministic behavior, but the description of the deterministic behavior is much too complex, or the behavior is difficult to observe. In such cases, it has turned out to be a tractable way to move from deterministic modeling to stochastic modeling, enriching a deterministic functional relationship by accounting for different random states which may occur. These different random states are gathered in a stochastic basis, which is mathematically described by a probability space $(\Omega, \mathcal{F}, P)$.

Simplification Due to Stochasticity Stochasticity is often used to model deterministic phenomena in a tractable way such that the model still describes the *outcomes* of real-world phenomena (that might actually be deterministic in nature). Instead of modeling the deterministic and possibly complicated procedure which leads to the outcome, one focuses only on data concerning the outcome, analyzes the "distribution" of the outcomes, and finally one sets up a stochastic model which captures the distribution of the outcomes as realistic as possible.

A very easy but vivid example of a situation where specifying the deterministic behavior may be awkward is modeling the result of throwing a (fair) dice: obviously, throwing a dice is an action which can be described completely by classical

---

[1]Translation: the theory yields a lot, but it does not bring us closer to the secret of the old one [god]. Anyway, I am convinced that he [god] does not throw the dice.

[2]One prominent exception is the statistical approach to quantum physics.

mechanics. Shaking the dice in a dice cup is a mechanic procedure, where the dice turns when touching the walls of the dice cup, falls and rolls on the table, and eventually displays some number. But the whole procedure of shaking the dice and rolling is extremely complicated to model in the world of classical mechanics, since many different influences have to be taken into account (like, e.g., the shape and size of the dice cup and the dice, the different directions and magnitudes of the shaking, etc.). Such a deterministic model would be hard to determine, to set up, and even more difficult to evaluate.

If one, however, is only interested in the result, i.e. the thrown number, one might imagine a model which is much more simple and circumvents the difficulties of modeling such a situation with classical mechanics. The mixing procedure cannot be reproduced easily and as a result, every side of the dice occurs similarly often. Mathematically spoken, the relative share $r(j)$ of obtaining a fixed number $j \in \{1, 2, 3, 4, 5, 6\}$ is independent of the number j and since the relative shares have to add up to one, it follows that $r(j) \approx 1/6$ for all $j \in \{1, 2, 3, 4, 5, 6\}$. Hence, a probabilistic model describing the result of throwing a dice, which both models reality feasibly and yields a tractable situation, is to provide a stochastic basis in the following way: let $\Omega := \{1, 2, 3, 4, 5, 6\}$ be the state space of possible dice throw outcomes, $\mathcal{F} := \mathfrak{P}(\Omega)$ all possible combinations of outcomes, and $P : \mathcal{F} \to [0, 1]$ a probability measure defined via $P(\{j\}) = 1/6$ for all $j \in \{1, 2, 3, 4, 5, 6\}$. Then the probability space $(\Omega, \mathcal{F}, P)$ sufficiently describes the possible outcomes of a dice throw in an abstract, easy, and tractable manner.

Contrary to modeling the dice throw by classical mechanics, the stochastic model has simplified and abstracted tremendously from the original situation. The whole procedure of throwing the dice *physically* is completely disregarded. Instead, the stochastic model only focuses on the *result* of the dice throw and models it directly, which turns out to be much more tractable and also feasible from an empirical point of view.

A Detailed Excursion: Stochastic Modeling in Finance   In physics and engineering, mathematical modeling of real-world phenomena goes back to Isaac Newton, Gottfried Wilhelm Leibniz, and even to the ancient Greeks. In contrast, in finance, mathematical and particularly stochastic modeling is a rather recent trend, starting with the seminal dissertation of Bachelier [16].

When regarding the financial world instead of modeling phenomena from classical mechanics, one immediately recognizes that the whole system is much more complex in the sense that many different forces drive the market, and their influence is of non-negligible order. When describing the fall of a stone to the ground in a laboratory, there are undoubtly also many different forces apart from earth gravitation that actually have some influence (e.g. the aerodynamic resistance, the gravitation of different objects in the laboratory). But their magnitude is so small compared to the magnitude of earth gravitation that not considering them eventually does not matter for a realistic model.

In contrast, when modeling financial markets (e.g. stock markets for the purpose of, e.g., option pricing), there are many different market participants that influence

asset prices by their trade decisions. Hence, a model trying to capture the whole market microstructure with all interactions of market participants would be a monstruous, extremely complicated attempt with myriads of parameters. Thus, such an approach is only tractable under severe simplifications (similar to the dice example). But, additionally, there are several other reasons not to model the microstructure of financial markets.

- First, different to the dice example, financial markets cannot be put under laboratory conditions and therefore models cannot be tested reliably, i.e. experiments cannot be repeated.
- Second, due to the complexity of the operations, it is impossible to observe all market participant's behavior and interaction simultaneously.
- Third, many market participants exhibit irrational and erratic behavior which may be difficult to model even when modeling only a single market participant. There have been approaches as the celebrated "Prospect Theory" of Kahneman and Tversky [34][3] trying to provide a scope for such a kind of behavior, which still is ongoing research.
- Finally, and maybe most crucial, the whole system is dynamic, with new market participants entering and leaving the system. Even if one could observe the market participants' behavior and collect huge amounts of data, in every second, new market participants enter the financial markets and behave differently, such that predictions relying on historical data might not explain future market situations successfully.[4]

Hence, the typical approach to model stock markets is to disregard the market microstructure (which is, e.g., forgetting about the market participants action and interaction,[5] analog to forgetting about the mechanics when rolling the dice) and to model asset prices statistically.

To set up a sensible stochastic model for the price of, e.g., a stock or an index, one typically scrutinizes stylized facts of time series of the price process and tries to mimic these properties with stochastic models fulfilling as many of these stylized facts as possible. Compared to an ansatz focusing more on data (an extreme ansatz may be a non-parametric one only exploiting data), such a modeling paradigm allows to capture general movements. Furthermore, a stochastic model for a stock price should be tractable enough in the sense that it costs moderate effort to simulate the stock price and prices of related financial instruments (e.g. futures and options, see Hull [8] for an introduction into financial instruments) may be calculated in a (semi-)analytic way. With these requirements for a model, one starts to collect some stylized facts of time series of stock prices and obtains as first observations:

---

[3]Daniel Kahneman was awarded the Nobel Memorial Prize in Economic Sciences 2002 for his work on irrational behavior in economics.

[4]In financial markets, one can even argue that relying too much on collected data may result in overconfidence, since the data may not be representative any more to model future events.

[5]One should note that there are some approaches trying to capture the microstructure.

- The stock price process, abbreviated by $S = (S_t)_{t \geq 0}$, is always positive.
- Returns (yields) of stock prices are symmetrically scattered around 0 (or around somewhere close to 0) and behave roughly similar and uncorrelated of each other.

Taking the second stylized fact as a starting point, a possible tool for modeling stock returns seems to be the normal distribution, which is widely understood, mathematically tractable, and plays a prominent role in asymptotic statistics (cf. the central limit theorem). Furthermore, for small periods Δt, the discrete return

$$\frac{S_{t+\Delta t} - S_t}{S_t}$$

may comfortably be approximated by the difference of the logarithm $\log S_{t+\Delta t} - \log S_t$. Hence, a first idea might be to model logarithmic differences by i.i.d. normally distributed random variables. With this motivation and the notion of Brownian motion (we omit the formal definition due to technicalities, see Øksendal [10] for details), one arrives at modeling stock prices with a geometric Brownian motion (which goes back to Samuelson [44]), also often called the Black–Scholes model.[6]

Example 1.1 (Black–Scholes Model)   A stock price $(S_t)_{t \geq 0}$ is modeled by a Black–Scholes model if it follows a geometric Brownian motion, i.e. its dynamics follow the stochastic differential equation[7]

$$\mathrm{d}S_t = \mu S_t \, \mathrm{d}t + \sigma S_t \, \mathrm{d}W_t, \quad S_0 > 0,$$

where $(W_t)_{t \geq 0}$ is a standard Brownian motion. The parameter $\mu \in \mathbb{R}$ is called the *drift* of the stock price and the parameter $\sigma > 0$ is called the *volatility* of the stock price.

The Black–Scholes model allows for an easy and comprehensive interpretation: the whole model is parameterized by the drift and the volatility of the process. Since the model implies normally distributed stock returns, everyone who is familiar with the normal distribution can apply and handle the model. The drift parameter μ controls the average stock return, which grows linearly in μ. In terms of stock prices, μ is the (exponential) growth rate of the stock price. The higher the drift μ, the faster the stock price grows *on average*. On the other hand, the volatility parameter σ describes how the returns scatter around the average returns. When regarding the stock price instead of the returns, the volatility controls how much the stock price

---

[6]Actually, the model was not developed by Black and Scholes, but by Samuelson and was inspired by the seminal PhD thesis Bachelier [16]. Fischer Black and Myron Scholes derived tractable formulae for European options in this model and introduced the idea of replication in their seminal paper Black and Scholes [3]. This work, together with the inspired work of Robert Merton, resulted in awarding the Nobel Memorial Prize in Economic Sciences to Robert Merton and Myron Scholes in 1997. Fischer Black died already in 1995, thus he did not receive the prize.

[7]Stochastic processes are often described via stochastic differential equations (SDEs). For readers that are unfamiliar with SDEs, we recommend the introductory book of Øksendal [10].
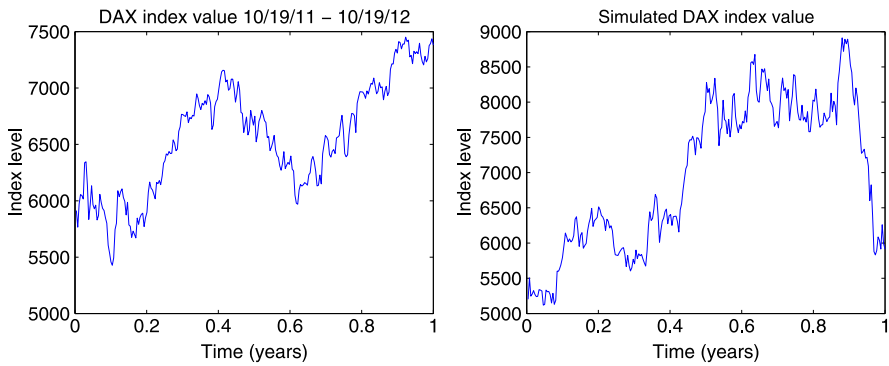
Fig. 1 Comparison: a time series of the DAX index level compared with a simulated path of the DAX in the Black–Scholes model

moves non-directionally. The higher the volatility σ, the more fluctuations the stock price exhibits.

For the pricing of options on the stock, one applies the *risk-neutral version* of the Black–Scholes model, where the drift equals the interest rate of a risk-free investment.

Obviously, the dynamics imposed by the Black–Scholes model are rough simplifications of real stock price dynamics. While the only source of randomness in the Black–Scholes model is the Brownian motion and all other ingredients (i.e. drift and volatility) are deterministic, real stock prices are driven by an extremely complex market microstructure. Instead of modeling the whole market microstructure with the dynamics of action and interaction, one simply assumes that it suffices to reduce the complexity to the determination of two parameters—the drift and the volatility. In case of risk-neutral dynamics (which is the standard assumption when pricing options), the complexity is further reduced to the determination of one single parameter—the Black–Scholes volatility. On the other hand, trajectories which are simulated in the Black–Scholes model look somewhat like plots of time series of real stock prices (cf. Fig. 1). Furthermore, the simple structure ensures the tractability of the model, in particular, there exist closed-form pricing formulas for various kind of options, like the classical Black–Scholes formula for European calls and puts.

Taking a closer look on stock price time series as well as on stock price related data (e.g. option prices), one clearly sees that the Black–Scholes model is oversimplifying reality and some stylized facts may not be explained by the Black–Scholes model like the following (which are not exhaustive):

- Extremely high and low returns are more likely to occur in reality than the normal distribution implies ("heavy tails of returns").
- Volatility is not constant, different market periods (high and low volatility) can be observed ("volatility clustering").

- Downward price movements are typically accompanied by large undirectional movements ("leverage effect").
- Option prices do not follow the Black–Scholes model, implied volatilities[8] are non-constant ("smile effect").

Hence, different alternatives to (and extensions of) the Black–Scholes model have been developed to tackle the shortcomings of using simple geometric Brownian motion, introducing models based on different processes with heavier tails or stochastic volatility and/or jumps. One model that has become popular in practice is the Heston model, see Heston [7], it uses a Cox–Ingersoll–Ross square-root process[9] as stochastic volatility. We briefly sketch the ingredients of the Heston model.

Example 1.2 (Heston Model)   A stock price $(S_t)_{t \geq 0}$ is modeled by a Heston model if its dynamics follow the coupled stochastic differential equations

$$dS_t = \mu S_t \, dt + \sigma_t S_t \, dW_t^{(1)}, \quad S_0 > 0,$$

$$d\sigma_t^2 = \kappa \big(\sigma_{\text{long}}^2 - \sigma_t^2\big) \, dt + \xi \sigma_t \, dW_t^{(2)}, \quad \sigma_0^2 > 0,$$

$$dW_t^{(1)} dW_t^{(2)} = \rho \, dt,$$

where $(W_t^{(j)})_{t \geq 0}$, $j = 1, 2$ are correlated Brownian motions with correlation $\rho \in [-1, 1]$.

For further explanation, one can see that the general stock price dynamics resemble closely the dynamics of the Black–Scholes model, except for one fact: the volatility σ is not assumed to be constant any more, but is now a stochastic process itself (due to technical reasons, one models the "variance process" $(\sigma_t^2)_{t \geq 0}$ instead of the volatility process $(\sigma_t)_{t \geq 0}$). In particular, the noise in the stock price process is now time-dependent and has its own dynamics.

Assuming the dynamics of a Cox–Ingersoll–Ross square-root process for the variance process, one may see the following behavior of the variance:

- The variance process $(\sigma_t^2)_{t \geq 0}$ exhibits non-constant noise, which is governed by the parameter $\xi > 0$. This parameter is usually called the *vol-of-vol*.
- In the long run, the variance fluctuates around a fixed number, the *long-term variance*, which is controlled by the parameter $\sigma_{\text{long}}^2 > 0$.
- The variance process is mean-reverting to the long-term variance, i.e. if the variance is dragged away from its long-term level, it drifts back to the long-term variance. The *speed of mean reversion* is controlled by the parameter $\kappa > 0$.

---

[8]In the Black–Scholes model, for a given European option, there is a one-to-one relationship between volatilities and option prices. Hence, for options with known market prices, one can recalculate the *implied volatility* from the market prices. Usually, one can exhibit that for different options, the recalculated implied volatilities differ, which is a hint that the Black–Scholes model cannot explain the observed option prices.

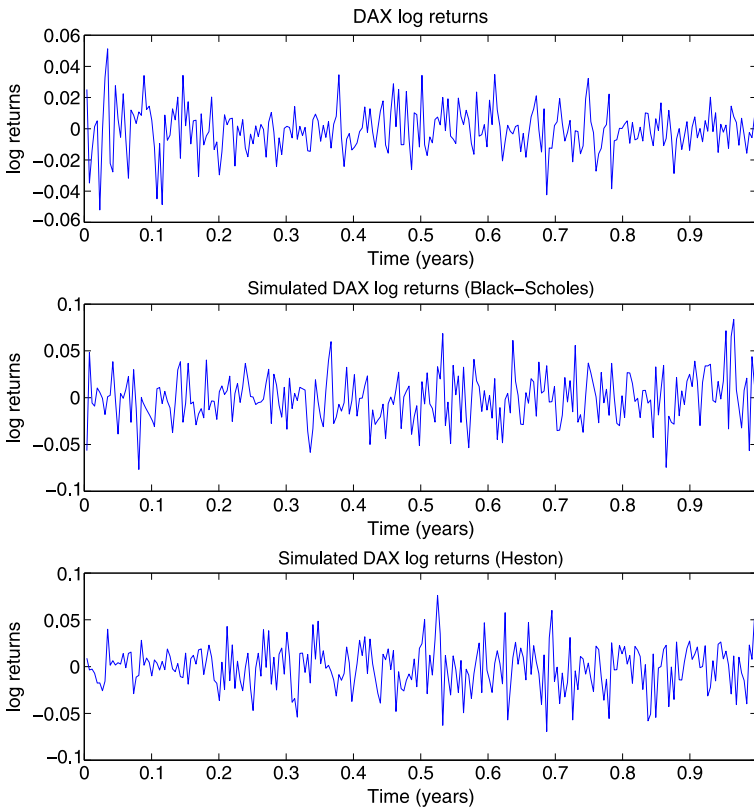[9]The name stems from the Cox–Ingersoll–Ross interest rate model.

Fig. 2 Comparison: logarithmic returns of the DAX compared with simulated logarithmic returns from the Black–Scholes and the Heston model. One can see that the Black–Scholes model produces returns with regular noise, while the Heston model incorporates volatility clustering, i.e. there exist time periods of high and low fluctuations in the returns

- The correlation $\rho \in [-1, 1]$ describes the co-movement of the stock price and its variance. As described above, this can be used to account for the so-called "leverage effect", establishing that volatility movements and stock price movements have negative correlation.

The Heston model is a relatively simple extension of the Black–Scholes framework (replacing constant volatility by a variance process following a Cox–Ingersoll–Ross model) to model stock prices. But, unarguably, the Heston model overcomes some of the shortcomings of the Black–Scholes model that have been described above (cf. Fig. 2). By making volatility stochastic and time-dependent, it captures the non-constant behavior of volatility. Furthermore, incorporating correlation between the drivers of the stock price and variance processes allows to account for the leverage effect, i.e. for negative correlations ρ. One has to remark that these additional stylized facts come at the price of losing mathematical tractability: prices for some important options (e.g. European put and call options) cannot be calculated

with simple formulae any more as in the Black–Scholes model, instead one has to rely on numerical algorithms as, e.g., techniques from Fourier analysis to obtain semi-analytic formulae as described in Carr and Madan [23].

2 Model Risk and Uncertainty

[T]here are known knowns; there are things we know that we know. There are known un-knowns; that is to say there are things that, we now know we don't know. But there are also unknown unknowns—there are things we do not know, we don't know.—Donald Rumsfeld, United States Secretary of Defence 1975–1977, 2001–2006

In the previous section, we have roughly outlined the main principles of mathematical modeling, in particular stochastic modeling where we will focus on below. Hence, if we refer to *modeling* in the remaining part of this survey, we always mean *stochastic modeling*.

When setting up a stochastic model, one often observes a complicated situation where the outcome in concern behaves in a more or less erratic manner. In some cases (like the dice example), a simple and accurate description may be provided easily. But, typically, the object to model is much more complicated (like the price process of a stock). Hence, it is not clear from the beginning that the choice of one stochastic model P is a good choice or a different model $\tilde{P}$ might be more suitable, like choosing either a Black–Scholes or a Heston model for stock prices. Typically, the quantity of interest is modeled by a random variable X or some stochastic process $(S_t)_{t \geq 0}$. Hence, a situation where modeling may be complex can be mathematically described as a situation where a whole set of probability measures $\mathcal{P}$ (which may typically be infinite) is available for modeling. Sometimes, the set of possible probability measures (i.e. different stochastic models) $\mathcal{P}$ may be parameterized in a canonical way by a parameter space Θ, i.e. $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$.

To provide a concise wording to different situations that may occur if different models $\mathcal{P}$ are available, we first make a short excursion into the literature. The seminal dissertation of Knight [9] analyzes the situation where different states $x_1, \ldots, x_N$ are possible outcomes for X. Knight [9] distinguishes between two possible situations that may occur:

1. One knows the probability of each possible outcome $x_1, \ldots, x_N$.
2. One does not know the probability of each possible outcome $x_1, \ldots, x_N$.

The ladder situation, where hardly any information is available, is called *uncertainty* by Knight [9]. The former one, which at least allows for a probabilistic description, is called *risk*. Obviously, facing risk is a special case of uncertainty (since one could always forget about the probabilities) and a more comfortable situation compared to facing real uncertainty. One can try to deal with a risky situation by *risk management*, i.e. exploiting the information about the probabilities of the different outcomes $x_1, \ldots, x_N$ and acting such that a certain risk functional is minimized.

Research from economics, but also from behavioral sciences like psychology and cognitive science, has shown that most people exhibit aversion towards both risk and

uncertainty (often subsumed under the term *risk aversion*). A mathematical concept covering risk aversion (prefering situations of certainty over situations of risk) is described by the foundations of utility theory by von Neumann and Morgenstern [48] and furthermore by the introduction of the axioms of subjective expected utility by Savage [45]. Arrow [14] and Pratt [41] analyze risk aversion from an economic perspective. Concerning uncertainty, it has been shown that the concept of uncertainty aversion is available, describing that a situation of risk is generally prefered to a situation where true uncertainty is exhibited. This idea was promoted by Ellsberg [29], challenging the axioms of Savage, which was later reconciled in the works of Gilboa and Schmeidler [31].

Transfering the concepts of risk and uncertainty to stochastic modeling, the situation of having a whole set of models $\mathcal{P}$ to choose from for modeling is generally referred to as *model uncertainty*. If each model $P \in \mathcal{P}$ can be identified by a parameter θ from some parameter space Θ, one speaks about *parameter uncertainty*.[10] If we additionally have given a probability measure R on the set of possible models $\mathcal{P}$ (resp. on the parameter space Θ) which quantifies the probability of each model (resp. parameter) to be the right choice, then we are in a setting of *model risk* (resp. *parameter risk*), which can be considered as a special case of model (resp. parameter) uncertainty.

This is illustrated in Fig. 3.

Examples   Model and parameter uncertainty arise in numerous situations. If one faces a complex situation where a stochastic model is applied, one is often ambiguous between different models to choose from. Even after having decided for a specific parametric model, the correct determination of the model's parameters is not straightforward and may result in different obstacles.

When stochastically modeling financial objects, there are myriads of possibilities to simplify, thus many different models are competing with each other. In option pricing, model risk (resp. uncertainty) should not be underestimated, as pointed out by Figlewski [5]. During the financial crisis of 2008, where massive misvaluation of portfolio credit instruments played an important role, this has been discussed in quite some detail among experts, but also in popular media as, e.g., Salmon [11].

Example 2.1 (Parameter Uncertainty in Financial Market Models)   All models treated in Sect. 1 are exposed to parameter uncertainty. We will discuss later whether we experience *true parameter uncertainty* in the sense that no information about the parameters is known or we have *parameter risk*, i.e. we are able to quantify whether certain parameters are more likely than others.[11]

---

[10]From a purely mathematical point of view, distinguishing between model and parameter uncertainty is just up to a mapping $\Theta \to \mathcal{P}$ which may always be obtained for some set Θ. Often, the set Θ can be chosen such that treating different parameters $\theta \in \Theta$ allows for more convenient interpretation in the real world than treating the corresponding model P_θ.

[11]Besides parameter uncertainty, the chosen parametric models can also be incorrect, i.e. model uncertainty can occur.

Fig. 3 Relationship between model uncertainty and risk. One can regard model risk as a "special case" of model uncertainty, since one can always ignore the probability measure R quantifying the likelihood of the different models

model uncertainty $- \mathcal{P}$
contains ≥ 2 models

model risk $-$ additional probability measure R on $\mathcal{P}$
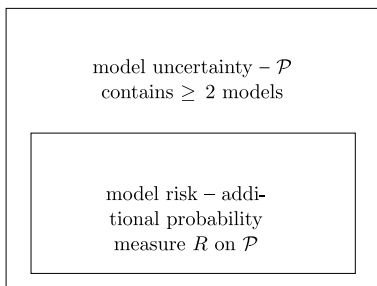
1. Examining the risk-neutral version of the Black–Scholes model, the dynamics of a stock price follow the stochastic differential equation

$$dS_t = r\, S_t\, dt + \sigma\, S_t\, dW_t, \quad S_0 > 0,$$

with $(W_t)_{t \geq 0}$ being Brownian motion, r the risk-free interest rate, and σ the stock's volatility. While the initial stock price S_0 and the risk-free rate r are usually available from market information, one does not have direct information about the volatility σ. Hence, a priori every positive number $\sigma > 0$ can be taken. Usually, one uses market data (e.g. estimation based on time series of stock prices, or fits the model to the prices of traded instruments) to specify the volatility σ.

2. In the (risk-neutral) Heston model, the stock price dynamics follow the coupled stochastic differential equations

$$dS_t = r\, S_t\, dt + \sigma_t\, S_t\, dW_t^{(1)}, \quad S_0 > 0,$$

$$d\sigma_t^2 = \kappa\left(\sigma_t^2 - \sigma_{\text{long}}^2\right) dt + \xi \sigma_t\, dW_t^{(2)}, \quad \sigma_0^2 > 0,$$

with $(W_t^{(j)})_{t \geq 0}$, $j = 1, 2$, being Brownian motions with correlation $\rho \in [-1, 1]$. Contrary to the Black–Scholes model, the number of unknown parameters is higher. Again, the initial stock price S_0 and the risk-free rate r are known by market quotation. On the other hand, the initial volatility σ_0, the mean reversion speed $\kappa > 0$, the long-term volatility $\sigma_{\text{long}}^2 > 0$, the vol-of-vol $\xi > 0$, and the correlation $\rho \in [-1, 1]$ are typically not given and—different from the Black–Scholes case—their interpretation is more complicated. Hence, we face parameter risk concerning the parameters $\sigma_0, \kappa, \sigma_{\text{long}}^2, \xi, \rho$.

Even across different models and when establishing perfect fits to market prices[12] of standard instruments (e.g. European call options), one obtains that there is still ambiguity and different models may cause different prices for non-standard options (as pointed out in Schoutens, Simons, and Tistaert [12]).

---

[12]One possibility to estimate the model parameters is to fit the parameters to known market prices of options.

3 Dealing with Model Risk

If history repeats itself, and the unexpected always happens, how incapable must Man be of learning from experience?—George Bernard Shaw, dramatist

Scrutinizing the available mathematical objects in presence of model (resp. parameter) risk, there exists more than only the set of different possible models $\mathcal{P}$. Additionally, one assumes that $\mathcal{P}$ is the state space of a probability space $(\mathcal{P}, \mathcal{F}^{\mathcal{P}}, R)$ where the probability measure R quantifies the probabilities that the different models $P \in \mathcal{P}$ are the correct models to choose. This delivers a lot of information which has to be analyzed carefully: first, for each stochastic model $P \in \mathcal{P}$, there are given probabilities for the different outcomes one has to deal with. Second, among all these models there is a second probability measure R assigning "weights" to the different models collected in the set $\mathcal{P}$.[13] In this case, one has numerous mathematical obstacles to tackle and to find the right way to incorporate model risk into quantities which may be of interest to be calculated like, e.g., prices of options.

From a statistical perspective, model risk can be regarded as an ansatz in the tradition of *Bayesian statistics*, where one main assumption is that the chosen model (or parameter) itself is random and the probability distribution on the possible models reflects subjective beliefs about the likelihood of the model. Opposed to this view, so-called *frequentist statistics* (going back to the seminal work of Fisher [30]) assumes that a true, but unknown, model (resp. parameter) exists and one cannot assign probabilities to different "candidate models". In history, there has been major dissent between these two philosophical approaches to statistics. A detailed critique and discussion of Bayesian and frequentist methods in statistics is beyond the scope of this article and we refer the interested reader to the books of Samaniego [43] and Bertsch McGrayne [18], but we give a short insight into the foundations of Bayesian statistics later in Sect. 3.2.

One situation where parameter risk traditionally occurs is parameter estimation from given data (e.g. time series of stock prices). In a standard procedure, disregarding parameter risk, one computes the derived estimators from the given data, i.e. calculates point estimates for the parameters. But from estimation theory, one knows that an estimator is a random object itself. Furthermore, an estimator may be biased. Hence, procedures that solely rely on using the point estimate disregard the parameter risk which arises through the estimator's distribution, e.g. its bias and variance.

Parameter estimation is a key step in every application where real data is analyzed. Hence, we present an example employing the Black–Scholes model where the estimator's distribution quantifies the parameter risk.

Example 3.1 (Parameter Risk from Estimation of the Black–Scholes Volatility) We consider a Black–Scholes setting as given in Example 1.1, where the volatility σ is

---

[13]Due to technical reasons, it may occur that the probability for all single models is zero, i.e. $R(\{P\}) = 0$ for all $P \in \mathcal{P}$.

the key parameter for option pricing. This parameter is not directly given by the
market (different from the current stock price S_0 and the risk-free rate r). Hence,
the determination of the volatility is a situation where one is exposed to parameter
uncertainty. If the stock price actually follows a Black–Scholes model, it may be
a sensible idea to estimate the volatility from time series data. Taking the logarith-
mic returns $x_1, \ldots, x_N$, $x_j = \log S_{t_j + \Delta t} - \log S_{t_j}$, $j = 1, \ldots, N$, one may choose
the classical estimator for the variance (it may be more convenient to estimate the
returns' variance), corrected for the frequency of the data Δt, which results in the
estimator

$$\hat{\sigma}_N^2 = \frac{1}{\Delta t (N-1)} \sum_{j=1}^{N} (x_j - \bar{x})^2, \qquad \bar{x} = \frac{1}{N} \sum_{j=1}^{N} x_j$$

for the variance corresponding to the Black–Scholes volatility, which is consistent
and asymptotically normal under very weak assumptions. Applying general theory
from statistics, one obtains that, under the assumption of independent normally dis-
tributed returns and a true variance $\sigma_0^2 > 0$ (as the Black–Scholes model does), the
distribution of the estimator is a χ^2-distribution up to some scaling. Hence, the dis-
tribution determining the parameter risk arising from the estimation risk of volatility
(resp. variance) is essentially determined by the χ^2-distribution, provided that the
true model is a Black–Scholes model with variance σ_0^2. The parameter space is given
by $\Theta = \mathbb{R}_{>0}$ and the estimator's distribution R has density r given by

$$r(x) = \frac{(\Delta t (N-1))^{\frac{N-1}{2}}}{\Gamma(\frac{N-1}{2})(2\sigma_0^2)^{\frac{N-1}{2}}} x^{\frac{N-3}{2}} \exp\left(-\frac{x \Delta t (N-1)}{2\sigma_0^2}\right) \mathbf{1}_{\{x>0\}}.$$

3.1 Measuring and Quantifying Model Risk

As defined by Knight [9], the exposure to model risk is a situation where probabili-
ties of different possible models are available. Hence, one should have mathematical
instruments at hand to measure and/or to quantify model risk. Fortunately, for the
general situation of the measurement and quantification of risk, a rich and mathe-
matically rigorous theory of *risk measures*[14] has been developed, yielding numerous
interesting results. For the specific purpose of treating model risk, the theory of risk
measures can be transferred, specifically tailored, and applied to the model risk set-
ting under concern. The theory of (convex) risk measures was originally designed
for treating financial and actuarial risk, headed by the seminal paper Artzner, Del-
baen, Eber, and Heath [1], we follow the red line of this survey and the model risk
framework in a financial context.

---

[14]The terminology "risk measure" may be misleading from a mathematical point of view, since the
functions that are proposed to be risk measures are not measures from a measure-theoretical point
of view, but functionals.

To ensure a concise understanding, we recapitulate the proper definition of risk measures in a slightly more general setup. A special case of the definition can be found in the textbook Föllmer and Schied [6].

Definition 3.2 (Risk Measure, cf. Biagini, Meyer-Brandis, and Svindland [19], Chap. 5) Let $\mathcal{X}$ be a collection of random variables on a probability space $(\Omega, \mathcal{F}, P)$, i.e. risk-exposed quantities, let $\pi : \mathcal{H} \to \mathbb{R}$ be a linear mapping on a subcollection of random variables $\mathcal{H} \subset \mathcal{X}$ and let $\rho : \mathcal{X} \to \mathbb{R}$ be a function.

ρ is called a *risk measure* w.r.t. π, if ρ fulfills the following axioms:

- ρ is monotone, i.e. for $X, Y \in \mathcal{X}$ and $X \geq Y$, $\rho(X) \geq \rho(Y)$ holds;
- ρ is π-translation invariant, i.e. for $X \in \mathcal{X}$ and $Y \in \mathcal{H}$ the equality $\rho(X + Y) = \rho(X) + \pi(Y)$ holds.[15]

Furthermore, ρ may have additional properties which are often postulated:

- ρ is called *convex*, if for $X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$, $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda \rho(X) + (1 - \lambda)\rho(Y)$ holds;
- ρ is called *coherent*, if it is convex and positively homogeneous, i.e. for $X \in \mathcal{X}$ and $c > 0$, $\rho(cX) = c\rho(X)$ holds;
- ρ is called *P-law-invariant*,[16] if the value of $\rho(X)$ only depends on the P-distribution of X, i.e. $\rho(X) = \rho(Y)$ holds if X and Y have the same distribution under P.

In practice, several "risk measures" are used. A traditional risk measure is, e.g., the variance, which was suggested for quantifying the risk of investments in portfolio theory in the seminal work of Markowitz [39].[17] However, the variance is not a risk measure in the sense of Definition 3.2, since it fails to be monotone.

The idea behind a risk measure is to compress all risk modeled by a random variable X into a single number $\rho(X)$. Obviously, this means that some information (i.e. the whole distribution of X) is lost and complexity is reduced, but it is a helpful and popular method to provide insight into risk for professional risk managers and to communicate to external audience. The convexity property translates into risk diversification: combining different risky quantities should not be penalized, i.e. the combined position cannot be riskier than the combination of the single positions. Furthermore, at first glance, the notion of π-translation invariance is rather unintuitive and difficult to understand: the interpretation is that the elements from $\mathcal{H}$ do not exhibit the kind of risk which is supposed to be measured ("risk-less positions"). Its risk quantification is solely determined by the linear mapping π, which is not risky

---

[15]In many cases, as discussed below, it is sufficient to think of $\mathcal{H}$ as the constants and of π as the identity function.

[16]If there is no ambiguity between different probability measures, the reference to the probability measure P is omitted.

[17]Harry M. Markowitz received the Nobel Memorial Prize in Economic Sciences 1990 for his groundbreaking research on portfolio theory.

by definition (since it does not exhibit risk diversification). In the original definition of convex risk measures, the subspace $\mathcal{H}$ only consists of the constant functions ("no risk") and the linear mapping π is simply the identity, i.e. $\pi(c) = c$.

The notion of risk measures was developed due to the shortcoming of classical risk measures as, e.g., quantiles (Value-at-Risk, often abbreviated by VaR), which in many cases did not exhibit desirable properties (e.g. VaR does not always support diversification). (Convex) risk measures provide a mathematically precise and rich framework for the measurement of risk, thus, it may also be adapted to measure model (resp. parameter) risk. The most popular non-trivial example of a convex risk measure is the Average-Value-at-Risk, which averages over the tails of a distribution and overcomes the shortfall of the Value-at-Risk being not convex.

The concrete implementation of the adaptation of the general framework of risk measures always depends on the setting what has to be measured, but, as a first idea, when a certain number $f(P)$ has to be calculated which depends on the probability measure $P \in \mathcal{P}$, it may be a sensible idea to apply the risk measure framework to the function f to provide a number accounting for the model (resp. parameter) risk.

Example: Option Pricing Incorporating Parameter Risk   A canonical example where model/parameter risk arises is option pricing. For this task, one uses financial market models as described in Sect. 1 which heavily rely on parameters that are not directly observable on the markets. Hence, those parameters have to be estimated, either via time series analysis of financial data or via fitting to market prices of available instruments (e.g. call and put options). As pointed out in Example 3.1, the procedure of obtaining the parameters exposes one to parameter risk. If one wants to state a price for some option using a certain model, e.g. the Heston model, one should account for parameter risk in the chosen model.[18] For some option X, each parameter vector θ in a financial market model yields the risk-neutral price of the option X w.r.t. the parameter vector θ as an expectation $\mathbb{E}_\theta[X]$. But, different from the usual model output, option traders typically state two prices—a bid price (to which she or he is willing to buy the option) and an ask price (to which she or he sells the option). Hence, the key idea is that parameter risk is a crucial determinant for the width and location of the bid-ask spread.

Thus, for option pricing purposes, the notion of a (model) risk-capturing functional and risk-captured (ask and bid) prices are developed in Bannör and Scherer [17] using the theory of convex risk measures.

Definition 3.3 (Model Risk-Capturing Functional, Risk-Captured Prices) Let $\mathcal{Q}$ be a family of option pricing models[19] and let R be a probability measure on $\mathcal{Q}$. Let $\mathcal{D}$ denote all options X we seek to price, which additionally satisfy some technical conditions. Let furthermore ρ be a normalized, law invariant convex risk measure

---

[18]Actually, one should also account for model risk, but this may not be tractable any more.

[19]To remain consistent with the usual terminology from mathematical finance, we denote risk-neutral measures by Q and a set of different risk-neutral measures by $\mathcal{Q}$.

on some functions on $\mathcal{Q}$. Then the mapping $\Gamma : \mathcal{D} \to \mathbb{R}$, defined by

$$\Gamma(X) := \rho\big(Q \mapsto \mathbb{E}_Q[X]\big), \tag{3.1}$$

is called a *model risk-capturing functional* w.r.t. the distribution R. $\Gamma(X)$ is called the *risk-captured* (*ask*) *price of X* given the functional Γ. Furthermore, $\bar{\Gamma}(X) := -\Gamma(-X)$ is called the *risk-captured bid price of X*.

The definition of risk-captured prices is somewhat technical and involves many different requirements (mainly to ensure the existence of the objects we deal with), but, in principle, the concept of treating the number of interest—the option price—as a function of the random model and applying a risk measure to it remains the same. In this case, since the methodology is supposed to be used for option pricing purposes, some additional quantities are required (e.g. normalization) to ensure that the number $\Gamma(X)$ makes sense. Furthermore, convexity is crucial since model risk should be a risk that profits from risk diversification. Option traders always regard their positions from a portfolio point of view, quoting bid-ask prices according to their portfolio position (e.g. they give better prices for options fitting to their present position).

The definition of model (resp. parameter) risk-captured prices is related to the idea behind some other non-linear pricing ideas that were mainly used for pricing in incomplete markets (like, e.g., Carr, Geman, and Madan [24], Cherny and Madan [25]).

3.2 Bayesian Treatment of Model Risk

A popular mathematical tool, when confronted with model risk, is Bayesian statistics. The basic idea behind Bayesian statistics is that the relationship between distributions of different models and samples thereof is not static, but is a dynamic process where the knowledge of the model distribution is constantly enhanced/updated. In this case, the model (resp. the parameter) is regarded to be random as well. Hence, one of the key results of Bayesian statistics we will present here is how the model (resp. parameter) distribution is updated and learns from the collected samples. Summarizing, Bayesian methodology is about how to obtain a proper distribution on the models incorporating information about the data into the construction process. A standard reference on Bayesian theory is Bernardo and Smith [2], one can find more about Bayesian methods in Chap. 8 of Czado and Brechmann [27].

Bayes's theorem, going back to the English minister of the Presbyterian church Thomas Bayes, is—in its most basic form—a relationship of conditional probabilities. Interchanging the conditioning set with the set which is evaluated, the conditional probability can be easily derived. Formulated in a mathematically precise manner, Bayes's theorem states the following result:

Theorem 3.4 (Bayes's Theorem, General Version) *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A, B \in \mathcal{F}$ some events with $P(A), P(B) > 0$. Then the following relationship between the conditional probabilities of the considered events holds*:

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)}.$$

At first glance, Bayes's theorem does not seem to have any interconnection with model risk and the application of Bayes's theorem towards model risk is not obvious. But when a distribution on the set of possible probability measures $\mathcal{P}$ is at hand, Bayes's theorem delivers an interesting interpretation of the relationship between the probability of outcomes and the probability of having the right model.

Therefore, let R be a probability distribution on the set of probability measures $\mathcal{P}$ quantifying the model risk, a joint probability measure Π living on the Cartesian product of the state space and the possible probability measures $\Omega \times \mathcal{P}$ may be defined on the "rectangle sets" via

$$\Pi(A \times B) := \int_B P(A)R(\mathrm{d}P) \tag{3.2}$$

for $A \times B \in \Omega \times \mathcal{P}$ (this measure may be extended to the whole product σ-algebra). The product measure Π can be interpreted as a probability measure which both incorporates possibilities of the outcomes and the different models. If we then apply Bayes's theorem to this situation, we obtain the following "model risk version" of Bayes's theorem.

Theorem 3.5 (Bayes's Theorem, Model Risk Version) *Let Π be defined as in* (3.2) *and $\Pi(A \times \mathcal{P}) > 0$. Then*

$$\Pi(\Omega \times B | A \times \mathcal{P}) = \frac{\Pi(\Omega \times B)\Pi(A \times \mathcal{P}|\Omega \times B)}{\Pi(A \times \mathcal{P})} = \frac{R(B)\Pi(A \times \mathcal{P}|\Omega \times B)}{\int_{\mathcal{P}} P(A)R(\mathrm{d}P)}$$

holds.

Defining suggestively $\Pi(A|B) := \Pi(A \times \mathcal{P}|\Omega \times B)$ as well as $\Pi(B|A) := \Pi(\Omega \times B|A \times \mathcal{P})$, one may summarize Theorem 3.5 via the handy expression

$$\Pi(B|A) = \frac{R(B)\Pi(A|B)}{\int P(A)R(\mathrm{d}P)}. \tag{3.3}$$

If we have a closer look on this formula, (3.3) reveals an interesting relationship between *model-intrinsic risk* (which is inherent in the different possible stochastic models $P \in \mathcal{P}$) and *model risk* (which is quantified by the probability measure R on the possible models $\mathcal{P}$). The probability that a set of stochastic models $B \subset \mathcal{P}$ is correct, given that a certain outcome $A \subset \Omega$ arrives, can be calculated by a fraction of the raw probability $R(B)$, corrected by a fraction which consists of the probability of the outcome A given the models B and the probability of A averaged over all

possible models $\mathcal{P}$. Hence, starting with a probability measure R on $\mathcal{P}$ quantifying model risk, one may obtain some further information and correct for the outcome A. In particular, if $B = \{P_0\}$ consists only of the probability measure P_0 (with positive probability $R(\{P_0\}) > 0$), (3.3) reduces to the even simpler form

$$\Pi(P_0|A) = \frac{R(P_0)P_0(A)}{\int P(A)R(\mathrm{d}P)}. \tag{3.4}$$

In a model risk framework based on continuous risk, one often has that the probability for a single model P_0 is zero (i.e. risk that comes from Lebesgue-a.c. probability measures), so the convenient representation (3.4) is usually not available. But there is a way out to find a nice form for Bayes's theorem treating model risk: if we assume that a parameterization of the set of possible models $\mathcal{P} = (P_\theta)_{\theta \in \Theta}$ with $\Theta \subset \mathbb{R}^n$ is at hand, the model risk probability measure R has a density $r(\theta)$, and the random variable of interest $X : \Omega \to \mathbb{R}^d$ has density $p(x|\theta)$ under P_θ for all $\theta \in \Theta$, we obtain the classical model risk version of Bayes's theorem involving densities.

Theorem 3.6 (Bayes's Theorem, Parameter Risk Version with Densities) *Let r, $(p_\theta)_{\theta \in \Theta}$ be as above. Then the conditional density $r(\cdot|x)$ can be calculated via*

$$r(\theta|x) = \frac{r(\theta)p(x|\theta)}{\int_\Theta p(x|\theta)r(\theta)\,\mathrm{d}\theta}. \tag{3.5}$$

Theorem 3.6 suggests particularly that the distribution on the parameters (represented by the density r) can be updated and adjusted, given the information from the samples $x = (x_1, \ldots, x_d)$. This can be regarded as follows: one starts with a parameter distribution r[20] (which is usually called a priori distribution or *prior distribution*, since it is the distribution imposed without any further information) and observes samples $x_1, \ldots, x_d$ on the market. Now, the distribution r is adjusted to the observation of the sample $x = (x_1, \ldots, x_d)$. Roughly speaking, the weights on the parameters are adjusted according to the likelihood of the sample outcome $x = (x_1, \ldots, x_d)$. As a result, one obtains a new distribution represented by the density $r(\cdot|x)$ incorporating both the information which was given by the a priori distribution and the additional information contained in the samples $x_1, \ldots, x_d$. Consequently, the obtained distribution $r(\cdot|x)$ is called the *a posteriori distribution* or *posterior distribution* on the parameters given $x = (x_1, \ldots, x_d)$. The whole procedure is referred to as *Bayesian updating* or *Bayesian inference*, since the new information contained in the samples $x_1, \ldots, x_d$ causes the old beliefs of the parameter distribution (summarized in the a priori density r) to be updated, resulting in the a posteriori density $r(\cdot|x)$. Bayesian updating can be done constantly when new data is available. Often, the old posterior distribution then comes into play as the new prior distribution, which is again updated with information from new samples $\tilde{x} = (x_{d+1}, \ldots, x_{d+\tilde{d}})$. Figure 4 illustrates this updating procedure.

---

[20]In the following, we use the word *distribution* for abbreviation and mean the distribution induced by the respective density.

Fig. 4 This diagram illustrates the process which is done in Bayesian updating (here as a mathematical "black box"). The information from the prior density (*top left*) is merged with data samples (*top right*), resulting in a unified distribution (*bottom*)
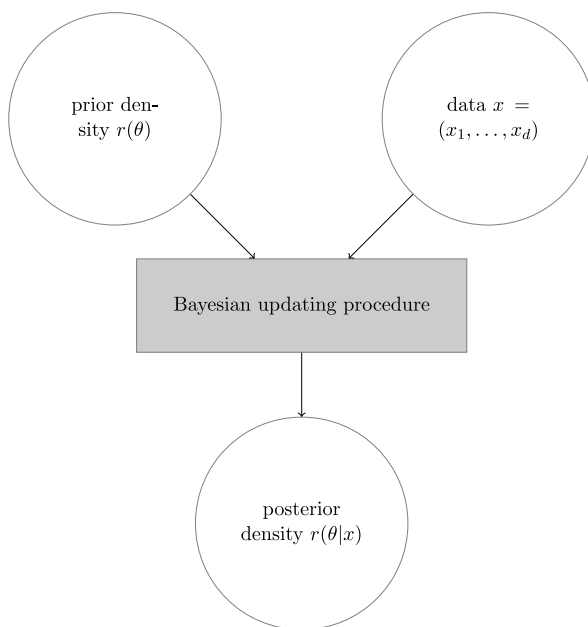
Merging Expert Knowledge and Data Evidence into a Unified Framework
A common application of the Bayesian updating process is when the input source is twofold: first, one has real-world data available for estimating parameters. A classical statistic paradigm would now solely rely on the given data, estimating the parameters and—if required—calculating the (asymptotic) distribution by using theory from mathematical statistics or resampling methods. But, in some cases, one wants to incorporate some expert judgement as well, particularly in case that the data may be difficult to judge (e.g. the data only reflects the recent past and some events not reflected in the past may happen in the future). Another case where one would like to incorporate expert judgements is when only very few data is available (like, e.g., operational risk events or corporate defaults) or a large fraction of data is outdated. For example, an option trader with long experience might impose a distribution on the parameters of a financial market model (e.g. Heston model) being subject to parameter risk (compare Example 2.1). Using a Bayesian updating procedure, one would use this distribution being the result of expert judgement as the *a priori distribution*. As a second step, one may use the Bayesian updating procedure and samples from financial market data (e.g. option prices) to adjust the expert view to real-world data.

One could also interpret Bayesian updating the other way round and start with a prior distribution that may be a sensible idea without having a closer look (prior distribution as "default distribution"). One can then use data or "expert estimates" to tilt the distribution towards results that are more in line with the data/the expert estimates.

As a result of applying Bayes's theorem in the version stated in Theorem 3.6, one obtains the *a posteriori distribution* integrating both the expert judgement as well as the data. Hence, loosely speaking, the a posteriori distribution may be regarded as a "merger" between the expert opinion and information extracted from data.

Examples   The methodology of Bayesian updating has widely been exploited in practice. Due to its handyness in terms of mathematical formulae and its mathematical rigorousity, it is one of the first choices to obtain distributions on models and particularly parameters.

Example 3.7 (Black–Litterman Portfolio Selection) A popular application of Bayesian updating is the Black–Litterman approach to portfolio optimization, as described in Black and Litterman [20]. In classical Markowitz portfolio optimization, risk and return characteristics of different investments are purely estimated from data (e.g. time series, option prices). A clear drawback of this procedure is that the used data is backward looking and does not carry information about future developments. Hence, one would like to introduce some procedure where data is one input, but on the other hand some subjective market opinion may influence the result. One way to incorporate some "market opinion" additionally is to use a Bayesian approach. In this case, both subjective views of investment performance and risk (a priori distribution) as well as financial market data (typically time series of financial instrument prices) can be integrated by means of Bayesian updating. As a result, one obtains a new distribution for risks and returns which is used for portfolio optimization purposes, called Black–Litterman portfolio selection.

Also in option pricing, Bayesian methodology provides a framework to obtain a distribution on the parameters such that today's option prices can be merged with an external view, e.g. coming from expert judgement or exploiting "more probable" market information.

Example 3.8 (Bayesian Option Pricing) There have been several attempts to incorporate Bayesian ideas into option pricing, we only sketch few of them (a complete overview would be out of scope). As described above, option pricing is a situation where one is exposed to parameter risk (and, presumably, model risk). Hence, Bunnin, Guo, and Ren [22] and Gupta and Reisinger [32] both suggest to compute the posterior distribution via Bayesian updating incorporating new data like realizations from time series and (more forward-looking) prices of European options. Gupta and Reisinger [32] assume that put and call option prices follow a true model that is noised by independent error terms. A mathematical framework is suggested how this assumption can be interpreted in terms of a parameter prior distribution. In particular, a local volatility framework is used and it is assumed that in the short run, the market-implied Black–Scholes volatilities of the most popular options[21] are concise approximations for the local volatility.

---

[21]Usually, the options which are most traded are the options with a strike close to today's stock price, the so-called at-the-money options.

An interesting question remains from choosing the prior distribution. Once having done the Bayesian updating procedure several times, one may use the old obtained posterior density as the new prior to start with as described above. Later, we refer to the Bernstein–von Mises theorem treating the asymptotic impact of the prior distribution.

In insurance applications, one is often more involved with using time-series data due to more stationary conditions (e.g. fire claims or other insurance losses observe more stationary behavior as financial markets). Many textbooks as, e.g., Böcker [21], Klugman [35], Wüthrich and Merz [49] address Bayesian methods for risk management in insurance and finance.

4 Dealing with Model Uncertainty

In nichts zeigt sich der Mangel an mathematischer Bildung mehr, als in einer übertrieben genauen Rechnung.[22]*—Carl Friedrich Gauss, mathematician*

In some cases, it is a hard task to quantify the probability of certain models to be the true model. It may be even impossible to impose a probability measure R on the set of different models $\mathcal{P}$ from which one may choose. In these situations, one experiences *true model uncertainty*. In such a situation, one has much fewer alternatives than in case of model risk, where quantification may be done via different risk measures, as we have described earlier. Conversely, in case of model uncertainty, one is typically restricted to consider worst-case scenarios: if there is no additional information and we have complete ambiguity between different stochastic models represented by the set of probability measures $\mathcal{P}$, one has little choice to boil the "degree of model uncertainty" down to one number as we have done it in case of model risk.

Worst-Case Approaches   Mostly, one seeks to calculate a number $f(P)$ (e.g. the price of some option) which depends on the chosen model $P \in \mathcal{P}$. Not having any further information at hand, the easiest way (and maybe the only feasible one—since everything in the scope of the model set $\mathcal{P}$ is possible) to quantify model uncertainty (as described for option pricing by Cont [4]) is to take the worst cases (resp. best cases) between the different models, namely

$$u = \sup_{P \in \mathcal{P}} f(P), \qquad l = \inf_{P \in \mathcal{P}} f(P).$$

Hence, the whole model uncertainty may be quantified by the difference of the two numbers

$$u - l = \sup_{P \in \mathcal{P}} f(P) - \inf_{P \in \mathcal{P}} f(P).$$

---

[22]Translation: nothing shows the lack of mathematical education more than an exaggeratedly exact calculation.

The difference between the extremes is an appropriate number to measure the (maximal) impact of model uncertainty on the quantity f. In case of model risk, i.e. the knowledge about the likelihood of each model, worst-case approaches can also be done. But, due to the additional knowledge, many other alternatives (as, e.g., convex risk measures) are possible.

Often, the number of interest $f(P)$ is the expectation of some random variable X w.r.t. the probability measure P (as in the case of option pricing). If this holds, the theory of convex risk measures (a standard reference is Föllmer and Schied [6]) immediately yields that the quantity

$$u(X) = \sup_{P \in \mathcal{P}} \mathbb{E}_P[X]$$

fulfills all the axioms of a coherent risk measure (without law invariance). One can go even further and define the *upper envelope* of a set of probability measures by defining

$$\mu_{\mathcal{P}}(A) := \sup_{P \in \mathcal{P}} P(A), \quad A \in \mathcal{F}.$$

In general, the upper envelope $\mu_{\mathcal{P}}$ is not a probability measure any more, but a submodular set function. Here, we can still define some integral, the Choquet integral w.r.t. $\mu_{\mathcal{P}}$, and the quantity $u(X)$ can be represented as a Choquet integral

$$u(X) = \int X \, d\mu_{\mathcal{P}}.$$

The Choquet integral is a generalization of the regular integral and relaxes some properties, e.g. it is not linear any more in general, but preserves features as, e.g., monotonicity. The rich theory of Choquet integration, delivering many tools to work with, can be found in the compendium of Denneberg [28].

Examples

Worst-Case Option Pricing Cont [4] describes the situation when a set of risk-neutral probability measures $\mathcal{Q}$ is available, but one does not have any information which one to pick for the valuation of some option X. As described above, it is suggested to use a worst case ansatz and to deliver two prices

$$u(X) = \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[X] \quad \text{and} \quad l(X) = \inf_{Q \in \mathcal{Q}} \mathbb{E}_Q[X],$$

which can again be interpreted as bid-ask prices. As described above, the functional u fulfills the axioms of a coherent risk measure. Conversely, if there is a coherent risk measure ρ which is defined on a suitable collection of random variables, general theory immediately yields that it can be represented as the supremum of the expectation w.r.t. some "stress-test measures" $\mathcal{Q}$, i.e.

$$\rho(X) = \sup_{P \in \mathcal{Q}} \mathbb{E}_Q[X]$$

holds for a set of "stress-test measures" $\mathcal{Q}$ which are absolutely continuous w.r.t. the original measure P. Hence, in this sense, convex risk measures as treated in Sect. 3.1 can also provide a framework to measure model uncertainty.

In some cases (as, e.g., the calibration to market prices), one might have additional information about the trustworthyness of a model, contained in some "penalty function" $\alpha : \mathcal{Q} \to [0, \infty]$. In this case, Cont [4] suggests "penalized worst-case pricing" by setting the two option prices via

$$u(X) = \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[X] - \alpha(Q) \quad \text{and} \quad l(X) = \inf_{Q \in \mathcal{Q}} \mathbb{E}_Q[X] - \alpha(Q). \qquad (4.1)$$

On the other hand, it can be shown that, in principle, every convex risk measure can be represented in the style of (4.1) (cf., e.g., Föllmer and Schied [6]). The very general framework developed by Cont [4] may be understood best by stating an example. One prominent example incorporating a rich class of pricing models is the uncertain volatility model by Avellaneda, Levy, and Paras [15].

Example 4.1 (Pricing with Uncertain Volatility) As described earlier, in a Black–Scholes model (cf. Example 1.1), the assumption of volatility being constant has caused numerous critique. Hence, stochastic volatility models (e.g. the Heston model presented in Example 1.2) have been developed. Again, these models assume certain characteristics of the volatility process. Another approach, leaving many degrees of freedom, was suggested by Avellaneda et al. [15]: volatility is introduced to be a stochastic process, living on a compact interval; i.e. the volatility process $(\sigma_t)_{t \geq 0}$ has its range in an interval $[\sigma_l, \sigma_u]$ with $\sigma_u > \sigma_l > 0$. The bounds σ_u, σ_l may be obtained from expert judgements or data like, e.g., available implied volatilities of liquid options. With these implicitly imposed models, Avellaneda et al. [15] develop an approach based on control theory methods to calculate model-free upper and lower bounds for the price of options.

Dependence Modeling Another situation where model uncertainty may arise is dependence modeling: often, different stochastic quantities that are related to each other (e.g. weight and height of persons) should be modeled jointly. Typically, this is modeled by assuming the realizations to come from a random vector $X = (X_1, \ldots, X_d)$. Assuming that the univariate distributions of the random variables $X_1, \ldots, X_d$ are known, one still has to determine the interconnection between the random variables, i.e. the dependence structure. Fortunately, Sklar's theorem provides that the dependence structure of any multivariate distribution may be separated from the univariate marginal distributions and any dependence structure corresponds to some copula, which is a multivariate distribution function with uniform marginals (see, e.g., Nelsen [40]). However, the set of copulas provides a broad and rich source with numerous dependence structures like, e.g., elliptical copulas or Archimedean copulas. Hence there are still infinitely many copulas to choose from, and sometimes there is little evidence about how the dependence structure may look like. More on this class of functions can be found in Chap. 9 of Klüppelberg and Stelzer [36].

In many cases, the choice of dependence structure is crucial for modeling events correctly, a vividly discussed example being portfolio default risk. On the eve of the financial crisis of 2008, there existed massive misvaluation of financial products called CDO (collateralized debt obligations) that were structured from, e.g., housing mortgages. The key principle of these products was to bundle several credits and redistribute the credit repayments and interest payments into different slices (so-called "tranches") in the following manner: in case of default, all defaults first reduce the notional of the most junior tranche. After elimination of the most junior tranche through defaults, the notional of the second-most junior tranche is reduced by occuring defaults and so on. As pointed out by Heitfield [33], the valuation of CDO tranches heavily relies on the imposed model of the dependence structure between the credit defaults. Predominantly, Gaussian copulas were used to account for the dependence, but Gaussian copulas were not able to capture important stylized facts like, e.g., contagion effects and tail dependence.

Typically, in case of *dependence uncertainty*, one wants to calculate a quantity $f_P(X_1, \ldots, X_d)$ and is uncertain about the dependence structure (represented by a copula model P) of the random vector $(X_1, \ldots, X_d)$. This means that the set of possible models $\mathcal{P}$ is constructed such that the univariate distributions of the random variables $X_1, \ldots, X_d$ do not vary, but the dependence structure, which can be summarized by

$$\mathcal{P} := \big\{ P \text{ probability measure on } (\Omega, \mathcal{F}) \text{ with}$$
$$\text{fixed marginal distributions } P^{X_j} \sim F_j \big\}.$$

The optimization problem to solve is to find upper and lower bounds (as in the proposal of Cont [4])

$$u(X_1, \ldots, X_d) = \sup_{P \in \mathcal{P}} f_P(X_1, \ldots, X_d),$$
$$l(X_1, \ldots, X_d) = \inf_{P \in \mathcal{P}} f_P(X_1, \ldots, X_d)$$

for functions f, which may include numerous applications, e.g. the calculation of risk measures of portfolios of financial instruments $X_1, \ldots, X_d$. An important result from copula theory is that the set of copulas has upper and lower natural bounds, called the Fréchet–Hoeffding bounds. These can be interpreted (at least in dimension $d = 2$) as "complete positive dependence" (comonotonicity) and "complete negative dependence" (countermonotonicity). But the Fréchet–Hoeffding copula bounds are not necessarily the copulas[23] which produce the upper and lower bounds $u(X_1, \ldots, X_d)$ resp. $l(X_1, \ldots, X_d)$ for the quantity $f(X_1, \ldots, X_d)$. Hence, the problem of determining the right dependence structure to approximate the upper and lower bounds has to be tackled mathematically.

---

[23]For $d > 2$, the lower Fréchet–Hoeffding bound is not even a copula.

Puccetti and Rüschendorf [42] present numerical and computational techniques to calculate upper and lower bounds for special functions f, including important examples like the Value-at-Risk (VaR) of portfolios. Using the fact that the empirical equivalent of copulas can be regarded as rearrangements, an algorithm is developed to calculate the bounds $u(X_1, \ldots, X_d)$ resp. $l(X_1, \ldots, X_d)$. In particular, it turns out that the comonotonicity copula (the upper Fréchet–Hoeffding bound) usually produces not the largest Value-at-Risk, but a copula that manages concentrating mass to the tail in a uniform manner.

In the bivariate case, there is another approach by Tankov [47], which refines the upper and lower bounds for a functional $f(X_1, X_2)$ when some information about the dependence (i.e. Kendall's Tau, a standardized association measure which is often more suitable than the correlation) is given. This is used to compute model-free bounds for bivariate options (e.g. best-of-two options), given a certain level of association measured by Kendall's Tau.

5 Food for Thoughts

This chapter intends to give a brief survey about model risk and uncertainty with a tilt towards financial topics, but, obviously, there are several questions that naturally arise.

- In this chapter, model risk and uncertainty is discussed in the context of mathematical finance. Obviously, also in natural sciences, model risk and uncertainty plays an important role. As a detailed example for a discussion of model risk and uncertainty in a natural sciences context, we refer to the book of Cooke [26].
- Convex risk measures are a tractable and well-studied class of risk functionals, but convexity (resp. subadditivity) may be an assumption that is too strong for real-life applications. Thus, there have been numerous generalizations and enhancements of convex risk measures incorporating weaker properties, like quasi-convexity or comonotone convexity (resp. subadditivity), studied in Song and Yan [46].
- When incorporating model (resp. parameter) risk by using convex risk measures, one might think about continuity properties of the computed numbers when imposing different kind of distributions on the parameters. In particular, one might want that if there is a sequence of distributions $(R_N)_{N \in \mathbb{N}}$ on the parameter set Θ converging to some limit distribution R_∞, the sequence of numbers capturing the model risk w.r.t. the distributions $(R_N)_{N \in \mathbb{N}}$ should eventually converge to the number capturing the model risk w.r.t. the distribution R_∞. An application would, e.g., be the distribution induced by some consistent estimator $\hat{\theta}_N$ converging to the "true" parameter. It turns out that, dependent on the risk measure, different types of convergence yield convergence for different classes of risk measures. Some ideas which risk measures behave as desired with weak convergence can be found in Bannör and Scherer [17], a detailed technical analysis about different topologies on probability measures that induce convergence of the risk measures is given in Krätschmer, Schied, and Zähle [37].

- In case of Bayesian methodology, one key problem is the choice of prior distribution. In some cases, the Bernstein–von Mises theorem states that in the asymptotics, the choice of prior distribution does not matter any more (e.g. van der Vaart [13]). Hence, the more iterations one does in the Bayesian updating procedure, one obtains more stable results (in case of drawing the sample from a stationary situation). Conversely, there are also situations where the Bernstein–von Mises theorem does not hold, which lead to criticism of the Bayesian methodology.

6 Summary

We presented an introduction to stochastic modeling and highlighted some problems concerned with model specification and the decision process which model to select. We defined and distinguished model uncertainty and risk, both are situations one typically faces when modeling complex objects as, e.g., financial markets, in a stochastic manner. We mentioned various examples, primarily from mathematical finance, where model and parameter risk and uncertainty play a prominent role. We have outlined methods based on convex risk measures dealing with both model risk and uncertainty, furthermore, we gave insight into Bayesian updating, which can be a helpful tool to refine parameter distributions in case of parameter risk.

References

Selected Bibliography

1. P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk. Math. Finance **9**(3), 203–228 (1999)
2. J.M. Bernardo, A.F.M. Smith, *Bayesian Theory*, 2nd edn. Wiley Series in Probability and Statistics (2007)
3. F. Black, M. Scholes, The pricing of options and corporate liabilities. J. Polit. Econ. **81**(3), 637–654 (1973)
4. R. Cont, Model uncertainty and its impact on the pricing of derivative instruments. Math. Finance **16**(3), 519–547 (2006)
5. S. Figlewski, Derivatives risks, old and new, in *Wharton-Brookings Papers on Financial Services* (1998)
6. H. Föllmer, A. Schied, *Stochastic Finance*, 2nd edn. (De Gruyter, Berlin, 2004)
7. S.L. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. **6**(2), 327–343 (1993)
8. J. Hull, *Options, Futures, & Other Derivatives* (Prentice Hall, New York, 2000)
9. F.H. Knight, *Risk, Uncertainty, and Profit* (Hart, Schaffner & Marx, Chicago, 1921)
10. B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications* (Springer, New York, 2003)
11. F. Salmon, Recipe for disaster: the formula that killed Wall Street. Wired Mag. (2009)

12. W. Schoutens, E. Simons, J. Tistaert, A perfect calibration! Now what? Wilmott Mag. **3** (2004)
13. A.W. van der Vaart, *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics (2000)

Additional Literature

14. K.J. Arrow, *Aspects of the Theory of Risk-Bearing* (Yrjö Jahnssonin Säätiö, Helsinki, 1965)
15. M. Avellaneda, A. Levy, A. Paras, Pricing and hedging derivative securities in markets with uncertain volatilities. Appl. Math. Finance **2**, 73–88 (1995)
16. L. Bachelier, *Théorie de la spéculation* (Gauthier-Villars, Paris, 1900)
17. K.F. Bannör, M. Scherer, Capturing parameter risk with convex risk measures. Eur. Actuar. J. 1–36 (2013)
18. S. Bertsch McGrayne, *The Theory that Would Not die* (Yale University Press, New Haven, 2011)
19. F. Biagini, T. Meyer-Brandis, G. Svindland, The mathematical concept of measuring risk, in *Risk – A Multidisciplinary Introduction*, ed. by C. Klüppelberg, D. Straub, L. Welpe (2014)
20. F. Black, R. Litterman, Global portfolio optimization. Financ. Anal. J. **48**(5), 28–43 (1992)
21. K. Böcker, *Rethinking Risk Measurement and Reporting Volume I—Uncertainty, Bayesian Analysis and Expert*. Risk Books (2010)
22. F.O. Bunnin, Y. Guo, Y. Ren, Option pricing under model and parameter uncertainty using predictive densities. Stat. Comput. **12** (2000)
23. P. Carr, D. Madan, Option valuation using the fast Fourier transform. J. Comput. Finance **2**, 61–73 (1999)
24. P. Carr, H. Geman, D. Madan, Pricing and hedging in incomplete markets. J. Financ. Econ. **62**, 131–167 (2001)
25. A. Cherny, D. Madan, Markets as a counterparty: an introduction to conic finance. Int. J. Theor. Appl. Finance **13**(8), 1149–1177 (2010)
26. R.M. Cooke, *Uncertainty Modeling in Dose Response: Bench Testing Environmental Toxicity*. Statistics in Practice (Wiley, New York, 2009)
27. C. Czado, E. Brechmann, Bayesian risk analysis, in *Risk – A Multidisciplinary Introduction*, ed. by C. Klüppelberg, D. Straub, L. Welpe (2014)
28. D. Denneberg, *Non-additive Measure and Integral* (Kluwer Academic, Norwell, 1994)
29. D. Ellsberg, Risk, ambiguity, and the Savage axioms. Q. J. Econ. **75**(4), 643–669 (1961)
30. R.A. Fisher, On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. A **222**, 309–368 (1922)
31. I. Gilboa, D. Schmeidler, Maxmin expected utility with non-unique prior. J. Math. Econ. **18**(2), 141–153 (1989)
32. A. Gupta, C. Reisinger, Robust calibration of financial models using Bayesian estimators. J. Comput. Finance (2012, to appear)
33. E. Heitfield, Parameter uncertainty and the credit risk of collateralized debt obligations. Available at SSRN 1190362 (2009)
34. D. Kahneman, A. Tversky, Prospect theory: an analysis of decision under risk. Econometrica **47**, 263–291 (1979)
35. S.A. Klugman, *Bayesian Statistics in Actuarial Science: With Emphasis on Credibility* (Springer, Berlin, 2011)
36. C. Klüppelberg, R. Stelzer, Dealing with dependent risks, in *Risk – A Multidisciplinary Introduction*, ed. by C. Klüppelberg, D. Straub, L. Welpe (2014)
37. V. Krätschmer, A. Schied, H. Zähle, Comparative and qualitative robustness for law-invariant risk measures (2012)
38. K. Mainzer, The new role of mathematical risk modelling and its importance for society, in *Risk – A Multidisciplinary Introduction*, ed. by C. Klüppelberg, D. Straub, L. Welpe (2014)

39. H.M. Markowitz, The utility of wealth. J. Polit. Econ. **60**, 151 (1952)
40. R.B. Nelsen, *An Introduction to Copulas*, 2nd edn. (Springer, Berlin, 2006)
41. J.W. Pratt, Risk aversion in the small and in the large. Econometrica **32**, 122–136 (1964)
42. G. Puccetti, L. Rüschendorf, Computation of sharp bounds on the distribution of a function of dependent risks. J. Comput. Appl. Math. **236**(7), 1833–1840 (2012)
43. F.J. Samaniego, *A Comparison of the Bayesian and Frequentist Approaches to Estimation* (Springer, Berlin, 2010)
44. P.A. Samuelson, Rational theory of warrant pricing. Ind. Manage. Rev. **6**(2), 13–39 (1965)
45. L.J. Savage, *The Foundations of Statistics* (Wiley, New York, 1954)
46. Y. Song, J.-A. Yan, Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. Insur. Math. Econ. **45**(3), 459–465 (2009)
47. P. Tankov, Improved Fréchet bounds and model-free pricing of multi-asset options. J. Appl. Probab. **48**(2), 389–403 (2011)
48. J. von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, 1944)
49. M.V. Wüthrich, M. Merz, *Stochastic Claims Reserving Methods in Insurance* (Wiley, New York, 2008)