CHAPTER 1

Models for Discontinuous Markets

The broadening and deepening of markets for risk transfer has marked the development of financial services perhaps more than any other trend. The past 30 years have witnessed the development of secondary markets for a wide variety of financial assets and the explosion of derivative instruments made possible by financial engineering. The expansion of risk transfer markets has liquefied and transformed the business of traditional financial firms such as banks, asset managers, and insurance companies. At the same time, markets for risk transfer have enabled nontraditional players to enter financial services businesses, invigorating competition, driving down prices, and confounding the efforts of regulators. Such specialist risk transfer firms occupy a number of niches in which they can outperform their more diversified counterparts in the regulated financial system by virtue of their specialized knowledge, transactional advantages, and superior risk management.

For all firms operating in risk transfer markets, traditional and nontraditional alike, the ability to create, calibrate, deploy, and refine risk models is a core competency. No firm, however specialized, can afford to do without models that extract information from market prices, measure the sensitivity of asset values to any number of risk factors, or forecast the range of adverse outcomes that might impact the firm's financial position.

The risk that a firm's models may fail to capture shifts in market pricing, risk sensitivities, or the mix of the firm's risk exposures is thus a central operational risk for any financial services business. Yet many, if not most, financial services firms lack insight into the probabilistic structure of risk models and the corresponding risk of model failures. My thesis is that most firms lack insight into model risk because of the way they practice statistical modeling. Because generally accepted statistical practice provides thin means for assessing model risk, alternative methods are needed to take model risk seriously. Bayesian methods allow firms to take model risk seriously—hence a book on Bayesian risk management.

RISK MODELS AND MODEL RISK

Throughout this book, when I discuss risk models, I will be talking about parametric risk models. Parametric risk models are attempts to reduce the complexity inherent in large datasets to specific functional forms defined completely by a relatively low-dimensional set of numbers known as parameters. Nonparametric risk models, by contrast, rely exclusively on the resampling of empirical data, so no reduction of the data is attempted or accomplished. Such models ask: Given the risk exposures I have today, what is the distribution of outcomes I can expect if the future looks like a random draw from some history of market data? Nonparametric risk models lack model specification in the way we would normally understand it, so that there is no risk of misspecification or estimation error by construction. Are such models therefore superior? Not at all. A nonparametric risk model cannot represent any outcome different from what has happened, including any outcomes more extreme than what has already happened. Nor can it furnish any insight into the ultimate drivers of adverse risk outcomes. As a result, nonparametric risk models have limited use in forecasting, though they can be useful as a robustness check for a parametric risk model.

Parametric risk models begin life as a probability distribution, which is a statement of the likelihood of seeing different values conditional only on the parameters of the distribution. Given the parameters and the form of the distribution, all possibilities are encompassed. More parameters create more flexibility: A Weibull distribution is more flexible than an exponential distribution. Many risk models rely heavily on normal and lognormal distributions, parameterized by the mean and variance, or the covariance matrix and mean vector in the multivariate case. A great deal has been written on the usefulness of heavier-tailed distributions for modeling financial data, going back to Mandelbrot (1963) and Fama (1965).

Undoubtedly, the unconditional distributions of most financial returns have heavier tails than the normal distribution. But to solve the problem of heavy tails solely through the choice of a different family of probability distributions is to seek a solution at a very low level of complexity.

More complex risk models project a chosen risk distribution onto a linear system of covariates that helps to articulate the target risk. Regression models such as these seek to describe the distribution of the target variable conditional on other available information. The functional form of the distribution is subsumed as an error term. Familiar examples include the following:

- *Linear regression with normally distributed errors*, widely used in asset pricing theory and many other applications.

- *Probit and logit models*, which parameterize the success probability in binomial distributions.
- *Proportional hazard models from insurance and credit risk modeling*, which project a series of gamma or Weibull distributions onto a linear system of explanatory factors.

Parameters are added corresponding to each of the factors included in the projection. The gain in power afforded by projection raises new questions about the adequacy of the system: Are the chosen factors sufficient? Unique? Structural? What is the joint distribution of the system parameters, and can that tell us anything about the choice of factors?

It seems the pinnacle in financial risk modeling is achieved when parameters governing several variables—a yield curve, a forward curve, a volatility surface—may be estimated from several time series simultaneously, where functional forms are worked out from primitives about stochastic processes and arbitrage restrictions. Such models pass over from the physical probability measure P to the risk-neutral probability measure Q. In terms of the discussion above, such models may be seen as (possibly nonlinear) transformations of a small number of factors (or state variables) whose distributions are defined by the nature of the underlying stochastic process posited for the factors. When the number of time series is large relative to the parameters of the model the parameters are overidentified, permitting highly efficient inference from the data. Such models are the ultimate in powerful description, offering the means to capture the dynamics of dozens of interest rates or forward contracts with a limited number of factors and parameters.

Our hierarchy of risk models thus includes as elements probability distributions, parameters, and functional forms, which may be linear or nonlinear, theoretically motivated or *ad hoc*. Each element of the description may not conform to reality, which is to say that each element is subject to error. An incorrect choice of distribution or functional form constitutes specification error on the part of the analyst. Errors in parameters arise from estimation error, but also collaterally from specification errors. The collection of all such opportunities for error in risk modeling is what I will call model risk.

TIME-INVARIANT MODELS AND CRISIS

The characteristics enumerated above do not exhaust all dimensions of model risk, however. Even if a model is correctly specified and parameterized inasmuch as it produces reliable forecasts for currently observed data, the possibility remains that the model may fail to produce reliable forecasts in the future.

Two assumptions are regularly made about time series as a point of departure for their statistical modeling:

1. Assuming the joint distribution of observations in a time series depends not on their absolute position in the series but only on their relative position in this series is to assume that the time series is *stationary*.
2. If sample moments (time averages) taken from a time series converge in probability to the moments of the data-generating process, then the time series is *ergodic*.

Time series exhibiting both properties are said to be ergodic stationary. However, I find the term *time-invariant* more convenient. For financial time series, time-invariance implies that the means and covariances of a set of asset returns will be the same for any T observations of those returns, up to sampling error. In other words, no matter when we look at the data, we should come to the same conclusion about the joint distribution of the data, and converge to the same result as T becomes large.

Standard statistical modeling practice and classical time series analysis proceed from the underlying assumption that time series are time-invariant, or can be made time-invariant using simple transformations like detrending, differencing, or discovering a cointegrating vector (Hamilton 1994, pp. 435–450, 571). Time series models strive for time-invariance because reliable forecasts can be made for time-invariant processes. Whenever we estimate risk measures from data, we expect those measures will be useful as forecasts: Risk only exists in the future.

However, *positing* time-invariance for the sake of forecasting is not the same as *observing* time-invariance. Forecasts from time-invariant models break down because time series prove themselves not to be time-invariant. When the time-invariance properties desired in a statistical model are not found in empirical reality, unconditional time series models are no longer a possibility: Model estimates must be conditioned on recent history in order to supply reasonable forecasts, greatly foreshortening the horizon over which data can be brought to bear in a relevant way to develop such estimates.

In this book, I will pursue the hypothesis that the greatest obstacle to the progress of quantitative risk management is the assumption of time-invariance that underlies the naïve application of statistical and financial models to financial market data. A corollary of this hypothesis is that extreme observations seen in risk models are not extraordinarily unlucky realizations drawn from the extreme tail of an unconditional distribution describing the universe of possible outcomes. Instead, extreme observations are manifestations of inflexible risk models that have failed to adapt to shifts in the market data. The quest for models that are true for all time and for all eventualities actually frustrates the goal of anticipating the range of likely adverse outcomes within practical forecasting horizons.

Ergodic Stationarity in Classical Time Series Analysis

To assume a financial time series is ergodic stationary is to assume that a fixed stochastic process is generating the data. This data-generating process is a functional form combining some kind of stochastic disturbance summarized in a parametric probability distribution, with other parameters known in advance of the financial time series data being realized. The assumption of stationarity therefore implies that if we know the right functional form and the values of the parameters, we will have exhausted the possible range of outcomes for the target time series. Different realizations of the target time series are then just draws from the joint distribution of the conditioning data and the stochastic disturbance. This is why a sample drawn from any segment of the time series converges to the same result in an ergodic stationary time series. While we cannot predict where a stationary time series will go tomorrow, we can narrow down the range of possible outcomes and make statements about the relative probability of different outcomes. In particular, we can make statements about the probabilities of extreme outcomes.

Put differently, when a statistical model is specified, stationarity is introduced as *an auxiliary hypothesis about the data* that allows the protocols of statistical sampling to be applied when estimating the model. Stationarity implies that parameters are constant and that further observations of the data improve their estimates. Sampling-based estimation is so widely accepted and commonplace that the extra hypothesis of stationarity has dropped out of view, almost beyond criticism. Consciously or unconsciously, the hypothesis of stationarity forms a basic part of a risk manager's worldview—if one model fails, there must be another encompassing model that would capture the anomaly; some additional complication must make it possible to see what we did not see in the past.

Yet stationarity remains an assumption, and it is important to understand its function as the glue that holds together classical time series analysis. The goal in classical time series econometrics is to estimate parameters and test hypotheses about them. Assuming stationarity ensures that the estimated parameter values converge to their "correct" values as more data are observed, and tests of hypotheses about parameters are valid.

Both outcomes depend on the law of large numbers, and thus they both depend on the belief that when we observe new data, those data are sampled from the same process that generated previous data. In other words, only if we assume we are looking at a unitary underlying phenomenon can we apply the law of large numbers to ensure the validity of our estimates and hypothesis tests. Consider, for the example, the discussion of 'Fundamental Concepts in Time-Series Analysis' in the textbook by Fumio Hayashi (2000, pp. 97–98) concerning the 'Need for Ergodic Stationarity':

> *The fundamental problem in time-series analysis is that we can observe the realization of the process only once. For example, the sample on the U.S. annual inflation rate for the period from 1946 to 1995 is a string of 50 particular numbers, which is just one possible outcome of the underlying stochastic process for the inflation rate; if history took a different course, we would have obtained a different sample ....*
>
> *Of course, it is not feasible to observe many different alternative histories. But* if the distribution of the inflation rate remains unchanged *[my emphasis] (this property will be referred to as stationarity), the particular string of 50 numbers we do observe can be viewed as 50 different values from the* same *distribution.*

The discussion is concluded with a statement of the ergodic theorem, which extends the law of large numbers to the domain of time series (pp. 101–102).

The assumption of stationarity is dangerous for financial risk management. It lulls us into believing that, once we have collected enough data, we have completely circumscribed the range of possible market outcomes, because tomorrow will just be another realization of the process that generated today. It fools us into believing we know the values of parameters like volatility and equity market beta sufficiently well that we can ignore any residual uncertainty from their estimation. It makes us complacent about the choice of models and functional forms because it credits hypothesis tests with undue discriminatory power. And it leads us again and again into crisis situations because it attributes too little probability to extreme events.

We cannot dismiss the use of ergodic stationarity as a mere simplifying assumption, of the sort regularly and sensibly made in order to arrive at an elegant and acceptable approximation to a more complex phenomenon. A model of a stationary time series approximates an object that can never be observed: a time series of infinite length. This says nothing about the model's ability to approximate a time series of any finite length, such as the lifetime of a trading strategy, a career, or a firm. When events deemed to occur 0.01 percent of the time by a risk model happen twice in a year, there may be no opportunity for another hundred years to prove out the assumed stationarity of the risk model.

Recalibration Does Not Overcome the Limits of a Time-Invariant Model

Modern financial crises are intimately connected with risk modeling built on the assumption of stationarity. For large actors like international banks, brokerage houses, and institutional investors, risk models matter a lot for the formation of expectations. When those models depend on the assumption of stationarity, they lose the ability to adapt to data that are inconsistent with the assumed data-generation process, because any other data-generation process is ruled out by fiat.

Consider what happens when an institution simply recalibrates the same models, without reexamining the specification of the model, over a period when economic expansion is slowing and beginning to turn toward recession. As the rate of economic growth slows the assumption of ergodicity dissolves new data signaling recession into a long-run average indicating growth. Firms and individuals making decisions based on models are therefore unable to observe the signal being sent by the data that a transition in the reality of the market is under way, even as they recalibrate their models. As a result, actors continue to behave as if growth conditions prevail, even as the market is entering a process of retrenchment.

Thinking about a series of forecasts made during this period of transition, one would likely see forecast errors consistently missing in the same direction, though no information about the forecast error would be fed back into the model. When models encompass a large set of variables, small changes in the environment can lead to sharp changes in model parameters, creating significant hedging errors when those parameters inform hedge ratios. Activity is more at odds with reality as the reversal of conditions continues, until the preponderance of new data can no longer be ignored; through successive recalibrations the weight of the new data balances and overtakes the old data. Suddenly actors are confronted by a vastly different reality as their models catch up to the new data. The result is a perception of discontinuity. The available analytics no longer support the viability of the financial institution's chosen risk profile. Management reacts to the apparent discontinuity, past decisions are abruptly reversed, and consequently market prices show extreme movements that were not previously believed to be within the realm of possibility.

Models staked on stationarity thus sow the seeds of their own destruction by encouraging poor decision making, the outcomes of which later register as a realization of the nearly-impossible. *Crises are therefore less about tail events "occurring" than about model-based expectations failing* *to adapt*. As a result, perennial efforts to capture extreme risks in stationary models as if they were simply given are, in large part, misguided. They are as much *effect* as they are *cause*. Financial firms would do much better to confront the operational task of revising risk measurements continuously, and using the outputs of that continuous learning process to control their business decisions. Relaxing the assumption of stationarity within one's risk models has the goal of enabling revisions of expectations to take place smoothly, to the extent that our expectations of financial markets are formed with the aid of models, in a way that successive recalibrations cannot.

BAYESIAN PROBABILITY AS A MEANS OF HANDLING DISCONTINUITY

The purpose of this book is to set out a particular view of probability and a set of statistical methods that untether risk management calculations from the foundational assumption of time-invariance. Such methods necessarily move away from the classical analysis of time series, and lay bare the uncertainties in statistical and financial models that are typically papered over by the assumption of ergodic stationarity. Thus, our methods will allow us to entertain the possibilities that we know the parameters of a model only within a nontrivial range of values, multiple models may be adequate to the data, and different models may become the best representation of the data as market conditions change. It is the author's conjecture (and hope) that introducing flexibility in modeling procedures along these multiple dimensions will reduce or even eliminate the extreme discontinuities associated with risk models in crisis periods.

Efforts to deal with nonstationarity within the realm of classical time series have centered around—and foundered on—the problems of unit roots, cointegration, and structural change (Maddala and Kim 1998). Unit roots and cointegration both deal with nonstationary time series by transforming them into stationary time series. Unit root econometrics achieves stationarity by differencing, whereas the analysis of cointegrated time series depends on the discovery of a linear combination of nonstationary series which becomes stationary. Still other methods rely on fractional differencing or other methods of removing deterministic or seasonal trends. Yet all of these classically-motivated methods for dealing with nonstationarity run into the problem of structural change. The possibility of structural change means unit root processes and cointegrating relations, among other data relationships, may not persist over the entirety of an observed period of data. When estimated models fail to detect and cope with structural changes, forecasts based on those models can become completely unreliable.

Bayesian probability methods may be used to overcome the assumptions that render classical statistical analysis blind to discontinuities in market conditions. As a result, we anticipate that firms operating in risk transfer markets can remain more sensitive to shifts in the market landscape and better understand the risks which form their core business focus by adopting a Bayesian modeling regime.

The choice of a Bayesian toolkit will tempt many readers to dismiss out of hand the alternatives presented here. I will plead pragmatism and try to mollify such readers by showing the conditions under which Bayesian results converge with classical probability. These skeptical readers can then decide whether they prefer to remain within the bounds of classical time series analysis or, better yet, choose to adapt their deployments of classical time series models to remain more sensitive to weaknesses in those models. For readers who are not burdened by such preconceptions, I will be unashamed of showing where Bayesian methods allow for possibilities ruled out a priori by classical probability and statistics.

In the previous section, we identified a taxonomy of model risks, which included parameter uncertainty, model specification uncertainty, and breakdowns in forecasting performance. In other words, models can lead us to incorrect conclusions because unknown parameters are known imprecisely, because the form of the model is incorrect, or because the form of the model no longer describes the state of affairs in the marketplace. Bayesian probability is predicated on the existence and irreducibility of all of these forms of model risk, and as a result, it furnishes resources for quantifying and monitoring each of these aspects of model risk.

Accounting for Parameter and Model Uncertainty

Let's consider a basic model. Denote the data by $\{x_t\}$ (we can assume they are continuously compounded large-cap equity returns) and the unknown parameters within the model as θ. If the model were the normal distribution, for example, we would have

$$p(x_t \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{\left(x_t - \mu\right)^2}{\sigma^2}\right],$$

with $\theta = \{\mu, \sigma^2\}$, the unknown mean and variance of the return series. Classical statistics would treat θ as unknown constants to be found by computing sample moments from $\{x_t\}$. Any uncertainty about θ is held to arise from sampling error, which implies that uncertainty can be reduced to a negligible amount by observing ever more data.

The basic insight of Bayesian probability comes from Bayes' rule, a simple theorem about conditional probability. If we consider the joint probability of x and θ, both of the following statements are true:

$$\begin{aligned} p(x,\theta) &= p(x \mid \theta)p(\theta) \\ &= p(\theta \mid x)p(x). \end{aligned}$$

Hence, if we equate the two statements on the right-hand side with each other and rearrange, we obtain another true statement, which is Bayes' rule:

$$\begin{aligned} p(\theta \mid x)p(x) &= p(x \mid \theta)p(\theta) \\ p(\theta \mid x) &= \frac{p(x \mid \theta)p(\theta)}{p(x)}. \end{aligned}$$

Setting aside the unconditional probability of the data $p(x)$ for the time being, we have the following expression of Bayes' rule as a statement about proportionality:

$$p(\theta \mid x) \propto p(x \mid \theta)p(\theta).$$

A particular interpretation is attached to this last expression. The term on the right $p(\theta)$ is a probability distribution expressing beliefs about the value of θ before observing the data. Rather than treating θ as a set of unknown constants, uncertainty about θ is explicitly recognized by assigning a probability distribution to possible values of θ. Here, we might break $p(\theta) = p(\mu, \sigma^2)$ into $p(\mu \mid \sigma^2)\ p(\sigma^2)$. Since μ can be anywhere on the real line, an appropriate prior distribution $p(\mu \mid \sigma^2)$ could be another normal distribution with parameters μ_0 and σ_0^2. An inverse-gamma distribution is useful as a model for $p(\sigma^2)$ because it is defined on the interval $[0, \infty)$ and variance cannot be negative in a normal distribution.

At the same time $p(\theta)$ recognizes uncertainty about the values of θ, it also provides a vehicle for introducing knowledge we already have about θ. Such knowledge is 'subjective' in that it is not based on the data. But that does not mean that it is arbitrary. For a lognormal model of large-cap equity returns at daily frequency, we may believe μ is centered on zero, with some greater or lesser degree of confidence, expressed through the specification of σ_0^2. The mode for the distribution of σ^2 might be $(50\%)^2/252$.

Specifying $p(\theta)$ also places useful restrictions on the parameter space, such as requiring σ^2 to be positive, while also indicating which values for θ would be surprising. A large nonzero value for the mean of a large-cap equity return series would be surprising, as would a volatility of 5 percent or 5,000 percent. Results such as these would lead a classical statistician to question his computations, data, and methods. The information in $p(\theta)$ may be interpreted as an indication of which results would be so contrary to sense as to be nearly impossible. However $p(\theta)$ has a failsafe, since $p(\theta) > 0$ everywhere along its support; no valid value of θ can be completely excluded a priori.

Because $p(\theta)$ is determined before seeing x on the basis of nondata knowledge, it is known as the *prior distribution* for the parameters θ. The prior distribution captures knowledge obtained from sources other than the data at hand, while recognizing the provisional and imperfect nature of such knowledge.

The distribution $p(x \mid \theta)$ will be familiar to students of statistics as the *likelihood* of seeing the data x conditional on the parameters θ. Maximum-likelihood techniques search for the θ that maximizes $p(x \mid \theta)$, bypassing prior information $p(\theta)$. Working the other way around, for fixed θ the likelihood is a statement about how surprising x is. The likelihood captures the information contained in the data. Moreover, most statisticians subscribe to the *likelihood principle*, which states that *all* information in the data is captured by the likelihood.

Given these elements, Bayes' theorem tells us that the *posterior distribution* of parameter values $p(\theta \mid x)$ is proportional to the prior distribution of parameter values $p(\theta)$ times the likelihood $p(x \mid \theta)$. The posterior distribution refines the knowledge introduced by the prior distribution on the basis of information contained in the likelihood. Thus, for unknown parameters within a given statistical model, we begin and end with a probabilistic expression for the model parameters that acknowledges the uncertainty of our knowledge about the parameters. We know before and after seeing the data what degree of uncertainty applies to our parameter estimates. If the data are consistent with our prior estimates, the location of the parameters will be little changed and the variance of the posterior distribution will shrink. If the data are surprising given our prior estimates, the variance will increase and the location will migrate. In Chapter 2, we explore the consequences of introducing prior information in this way, and compare the Bayesian approach to classical methods for handing prior information via hypothesis tests.

Now consider alternative model specifications. Instead of x and θ, we could just as easily consider (formally) the joint probability of x and M_i, where M_i is a candidate model specification:

$$p(M_i \mid x) \propto p(x \mid M_i)p(M_i)$$

The notation suppresses θ, but that is not to say that $p(\theta)$ and $p(\theta \mid x)$ do not matter to the determination of model probabilities. We defer dealing with these subtleties for now.

The explicit recognition of multiple models with competing claims to being true on the basis of the data is in sharp contrast to classical statistical practice, which merely permits the acceptance or rejection of individual model specifications against unspecified alternatives, or as hypothesis tests within the context of an encompassing model. If the idea of attaching numeric probabilities to models seems unnatural, think of the probabilities as expressions of the odds model i is a better representation of the data than model j, $p(M_i)/p(M_j)$, subject to the constraint $\sum_i p(M_i) = 1$.

Just as in the specification of $p(\theta)$, $p(M_i)$ for each model i expresses existing beliefs about the adequacy of different models, which recognizes uncertainty about the best representation of the data-generation process. The axioms of probability ensure that $p(M_i) > 0$ for any model within the set of models being considered. On the other hand, closure of the set of models is not as straightforward as closure of the set of possible parameter values: We implicitly set $p(M_i) \equiv 0$ for any model not entertained. However, efforts to close the set of models are neither possible nor practical.

Posterior model probabilities are updated on the basis of the data from their prior values, also as in the case of prior and posterior parameter distributions. As a result, the adequacy of the model is explicitly evaluated on the basis of the available data, relative to the adequacy of other models. Further, with the aid of posterior model probabilities, an expectation may be computed over an ensemble of models, so that forecasts need not depend exclusively on a particular specification. Instead of assuming stationarity, a Bayesian approach with multiple models admits the possibility that a variety of data-generating processes—all of which are known only approximately—may be responsible for current observations. The manner in which Bayesian probability accounts for uncertainty about data-generating processes is discussed at greater length in Chapter 3.

Responding to Changes in the Market Environment

The passage from prior to posterior probability via the likelihood suggests a sequential approach to modeling in which inferences are progressively updated as more data are observed. Sequential model implementation in turn suggests a means of coping with the third aspect of model risk, the risk of discontinuity. Part Two of the book is concerned with extending the Bayesian framework for handling parameter and model uncertainty to a dynamic form, which allows for ongoing monitoring and updating. The goal of Part Two is to construct adaptive models that remain sensitive to anomalous data and learn from their forecasting mistakes, and to identify metrics that will show the evolution of parameter and model uncertainty as new data are encountered.

With the construction of adaptive models, our approach to modeling financial time series switches from the *batch analysis* perspective of classical time series to an *online* perspective that updates inferences each time new data are observed. The shift in perspective is essential to abandoning the assumption of a time-invariant data-generation process. When we model a financial time series with classical methods, time series data are batched without knowing if the same process generated all of the data. If transitions between processes go undetected, the time series model will average parameter values on either side of the transition, glossing over the change in the process. The resulting model would not have produced valid forecasts on either side of the transition, and its ability to forecast out-of-sample would be anyone's guess.

The primary obstacle to making Bayesian analysis sequential is its own tendency in the face of accumulating data to reproduce the *ignorant certainty* inherent in classical statistics. If all observed data are regarded as being sampled from the same data-generation process, the relative weight on the likelihood converges to unity, while the weight on the prior goes to zero. Asymptotically, Bayesian estimates are equivalent to maximum-likelihood estimates unless we explicitly recognize the possibility that current observations are not sampled from the same process as past observations. The technique of discounting, introduced in Chapter 4, ensures that current observations have a greater role in reevaluating parameter distributions and model probabilities than the accumulated weight of observations in the distant past.

Discounting past data is already common practice and is implemented in standard risk management software. When new data enter the observation window, models are recalibrated on the reweighted data set. However, reweighting the data introduces new problems and does not nullify the problems associated with recalibration. First, the weight that current data deserve relative to past data cannot be specified a priori. Efforts to estimate "optimal" discount rates via maximum likelihood are once again misguided, because the result will be sensitive to the data set and may paper over important differences. Second and more important, model recalibration fails to carry any useful information about parameters or models from one date to the next. The results from previous calibrations are simply thrown away and replaced with new ones. Often, the result is that model parameters jump discontinuously from one value to another.

In Chapters 5 and 6, dynamic state-space models are introduced as a means of carrying inferences through time without the profligate waste of information imposed by recalibration. Dynamic models thus allow discounting to take place without erasing what has been learned from earlier data. Indexing models by alternative discount rates then allows for uncertainty about discount rates to be handled through the computation of posterior model probabilities. When the world looks more like a long-run average, models that give less relative weight to current data should be preferred, whereas models that forget the past more rapidly will be preferred in times of rapid change.

TIME-INVARIANCE AND OBJECTIVITY

Bayesian methods view probability as a degree of justified belief, whereas classical methods define probability as frequency, or the expected number of outcomes given a large number of repetitions. Classical statisticians trumpet the "objectivity" of their approach against the "subjectivity" of Bayesians, who speak unabashedly about belief, rather than "letting the data speak." So-called objective Bayesians aim to split the difference by using *uninformative priors,* which have minimal influence on inferences, though disagreements exist about which priors are truly uninformative. To the extent that classical statisticians will arrive at the same result if they apply the same protocol, their process is "objective."

However, it is rarely the case that everyone with access to the same data draws the same conclusion from it—they test different hypotheses, use different models, and weigh the results against other knowledge (justified belief?) before coming to a (provisional) conclusion. Bayesian probability makes these subjective prior commitments explicit and produces an outcome which weighs the prior commitment and the data in a completely transparent way. Two Bayesians applying the same prior and model to the same data will arrive at the same result. So is the Bayesian process "subjective" because it makes a summary of non-data-based knowledge explicit, whereas "objective" statistics leave such things unstated?

Given an unlimited amount of data, any prior belief expressed by a Bayesian will be swamped by the evidence—the relative weight accorded to the prior belief goes to zero. Hence, from the point of view of Bayesian probability, objectivity is a kind of limit result that is only possible under the strong assumption of unlimited data drawn from a time-invariant data-generating process. In the realm of classical time series analysis, objectivity requires stationarity, as well as a possibly unlimited amount of time to permit ergodicity (the law of large numbers) to take hold. We should be wary of a protocol that requires everyone to ignore the possibility that the world does not accord with our modeling assumptions, and to suspend our disbelief about short-term results in the faith that in the limit, our measurements of relative frequency will be correct. If accounting for these possibilities introduces subjectivity, then so be it.

Dispersion in prior probabilities is the essence of trading and entrepreneurship. New trading ideas and new ventures do not get under way without a strong prior belief that something is the case within a market. These ideas and ventures are new and entrepreneurial precisely because they are out of sync with what is generally accepted. Different prior probabilities will most certainly generate different posterior probabilities, particularly when data are scarce, and when decisions are being made on the basis of posterior probabilities, dispersion in beliefs will generate dispersion in actions. Competition and speculation both depend on the heterogeneity of opinions and actions.

A protocol that encourages market participants to agree on "objective" estimates and take identical actions in response to those estimates enforces homogeneity, crowding, and ossification (Chincarini 2012). Multiple firms acting on the same "objective" conclusion from the same data herd into markets pursuing the same value proposition. Consider the universe of statistical arbitrage hedge funds that mine the CRSP and Compustat databases—among other standard data sources—to discover asset-pricing anomalies and build "riskless" portfolios to exploit them. Starting from the same data and the same set of models, they should buy and sell the same universes of securities in pursuit of value. When results break down, as they did in August 2007, it is impossible for all traders to obtain liquidity at a reasonable price, and an entire segment of the asset management industry can get crushed at once (Khandani and Lo 2007, Section 10). Objectivity does not lead to robustness at a systemic level, and objective statistics cannot generate competition or support new ideas, so their enduring value within the financial firm is circumscribed, at best.

It is also striking, on deeper examination, how "objective" statistical practice buries subjective elements deep within methodology as *ad hoc* choices and rationalizing simplifications. So-called objective classical statistics not only rely on the dogma of uniform data-generation processes already discussed; they also enforce certain beliefs about nondata knowledge and loss functions about which most people would express different views, if they were free to do so.

We already know that risk is subjective. Different people have different risk tolerances, and their willingness to bear risk depends crucially on their relative knowledge endowments. Thus, if the goal of a risk specialist firm is to identify and exploit a particular opportunity within the universe of financial risks, a modeling framework that provides a vehicle for that firm's particular knowledge is to be greatly preferred to a modeling framework that enforces the same conclusions on all users. On the other hand, if a firm, its management, and its regulators are eager to follow the herd, even if it means going over the precipice, they are welcome to take refuge in the "objective" of their conclusions.