Wiley Trading Series

POSITIONAL OPTION TRADING

An Advanced Guide

EUAN SINCLAIR

WILEY

Table of Contents

List of Tables

List of Illustrations

Chapter 9

Founded in 1807, John Wiley & Sons is the oldest independent publishing company in the United States. With offices in North America, Europe, Australia, and Asia, Wiley is globally committed to developing and marketing print and electronic products and services for our customers' professional and personal knowledge and understanding.

The Wiley Trading series features books by traders who have survived the market's ever-changing temperament and have prospered—some by reinventing systems, others by getting back to basics. Whether a novice trader, professional, or somewhere in-between, these books will provide the advice and strategies needed to prosper today and well into the future.

For more on this series, visit our website at www.WileyTrading.com.

POSITIONAL OPTION TRADING

An Advanced Guide

Euan Sinclair

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762–2974, outside the United States at (317) 572–3993, or fax (317) 572–4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at http://booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Names: Sinclair, Euan, 1969- author.

Title: Positional option trading : an advanced guide / Euan Sinclair.

Description: Hoboken, New Jersey : Wiley, [2020] | Series: Wiley trading series | Includes bibliographical references and index.

Identifiers: LCCN 2020008849 (print) | LCCN 2020008850 (ebook) | ISBN 9781119583516 (hardback) | ISBN 9781119583523 (adobe pdf) | ISBN 9781119583530 (epub)

Subjects: LCSH: Options (Finance) | Financial futures.

Classification: LCC HG6024.A3 S56225 2020 (print) | LCC HG6024.A3 (ebook) | DDC 332.64/53—dc23

LC record available at https://lccn.loc.gov/2020008849

LC ebook record available at https://lccn.loc.gov/2020008850

Cover Design: Wiley

Cover Image: © blackred/Getty Images

INTRODUCTION

You know nothing, Jon Snow.

—Ygritte in *A Storm of Swords* by George R. R. Martin.

He is not the only one.

We are not in a time where reason is valued. In economics, the idea that marginal tax cuts pay for themselves is still advanced, even though all evidence says they don't.

Forty percent of Americans do not believe in evolution. Forty-five percent believe in ghosts. These beliefs are not based on any evidence. They are manifestations of another philosophy, whether it is economic, religious, or sociological. Usually these opinions reveal more about what people want to be true rather than any facts that they know. And many people know few facts anyway. Evidence is seen as irrelevant and arguments are won by those who shout loudest and have the best media skills.

The idea that opinions are as valid as facts also affects trading and investing. Many investors rely on methods that are either unproven or even proven to be ineffective. The few of these investors who keep records will see that they are failing but rely on cognitive dissonance to continue to believe in their theories of how the markets behave. One would think that losing money would prompt reexamination, but the persistence of losers is remarkable. And even when some people give up or are forced out, there is always new money and new participants to replace the old.

The only way to learn about anything is through the scientific method. This is an iterative procedure where theory is modified by evidence. Without evidence we are just in the realm of opinion. Most of what I present here is backed by evidence. There are also some opinions. My justification for this is that experience is also a real thing. But I'm no less prone to self-delusion than anyone else, so feel free to pay less attention to these ideas.

Trading is fundamentally an exercise in managing ignorance. Our ability to judge whether a situation presents a good opportunity will always be based on only a simplified view of the world, and it is impossible to know the effects of the simplification. Our pricing model will be similarly compromised. It will be a simplification and possibly a very unrealistic simplification. Finally, the parameters the models need will have estimation errors and we generally won't know how large these are.

It is impossible to understand the world if you insist on thinking in absolute terms. The world is not black and white. Everything has shades of gray. You won't learn much from this book is you aren't comfortable with this.

This is clear for “risk.” Everyone has a different risk tolerance, whether this is personal or imposed by management or investors. But more important, risk is multidimensional. We are comfortable with this idea in some areas of life. Imagine you have a choice of going on vacation to either Costa Rica or Paris. Both are nice places and any given person could reasonably choose either one. But Paris has no beaches and Costa Rica doesn't have the Pompidou Centre. There is no one correct choice in this situation. And that is also the case in most investing decisions.

Many of the ideas I write about can be extrapolated to a ludicrous level. But if you do, don't blame me or credit me with the resulting conclusions.

In no particular order here are some facts that are often misrepresented:

- There is usually a variance premium. This does not mean there is always a variance premium.
- There is usually a variance premium. This does not mean you should always be short volatility.
- Short volatility can be risky. This does not mean that short volatility has to have unlimited risk.
- Some theories (e.g., GARCH, BSM, EMH, returns are normally distributed) have limitations. This does not mean the theories are stupid or useless.
- The Kelly criterion maximizes expected growth rate. This does not mean you should always invest according to it.

If what I write is unclear or incorrect, that is a problem of my making. But if you choose to ignore nuance, that is your issue.

Trading as a Process

I have made no attempt here to write a comprehensive option trading book. I don't cover the definitions and specifications of various types of options. There are no derivations of option pricing models. I expect the reader to know about the common option structures such as straddles, spreads, and strangles. Many books cover these topics (e.g., Sinclair, 2010). A very brief summary of the theory of volatility trading is provided in Chapter One, but this is not a book for beginners.

This is a book for experienced traders who want the benefits of including options in their strategies and portfolios but who are unwilling or unable to perform high-frequency, low-cost dynamic hedging. Again, there are many books on this type of positional option trading, but none are theoretically rigorous, and most ignore the most important part of trading anything: having an edge.

One of the things that distinguishes professionals from amateurs in any field is that professionals use a consistent process. Trading should be a process: find a situation with edge, structure a trade, then control the risk. This book documents these steps.

The book's first section explores how to find trades with positive expectation.

In Chapter Two, we look at the efficient market hypothesis and show that the idea leaves plenty of room for the discovery of profitable strategies. This insight lets us categorize these "anomalies" as either inefficiencies or risk premia. These will behave differently and should be traded differently. Next, we briefly review how behavioral psychology can help us and also its limitations. We examine two popular methodologies for finding edges: technical analysis and fundamental analysis.

Chapter Three looks at the general problem of forecasting. No matter what they say, every successful trader forecasts. The forecast may not be one that predicts a particular point value, but probabilistic forecasting is still forecasting. We introduce a classification of forecasting methods. Forecasts are either model based, relying on a generally applicable model, or situational, taking advantage of what happens in specific events. We very briefly look at predicting volatility with time-series models before moving on to our focus: finding specific situations that have edge.

The most important empirical fact that an option trader needs to know is that implied volatility is usually overpriced. This phenomenon is called the variance premium (or the volatility premium). Chapter Four summarizes the variance premium in indices, stocks, commodities, volatility indices, and bonds. We also present reasons for its existence.

Having established the primacy of the variance premium, Chapter Five gives eleven specific phenomena that can be profitably traded. The observation is summarized, the evidence and reasons for the effect are given, and a structure for trading the idea is suggested.

The second section examines the distributional properties of some option structures that can be used to monetize the edge we have found. We need to have an idea of what to expect. It is quite possible to be right with our volatility forecasts and still lose money. When we hedge, we become exposed to path dependency of the underlying. It matters if a stock move occurs close to the strike when we have gamma or away from a strike when we have none. If we don't hedge, we are exposed to only the terminal stock price, but we can still successfully forecast volatility and lose because of an unanticipated drift. Or we can successfully forecast the return and lose because of unanticipated volatility.

Chapter Six discusses volatility trading structures. We look at the P/L distributions of straddles, strangles, butterflies, and condors, and how to choose strikes and expirations.

In Chapter Seven we look at trading options directionally. First, we extend the BSM model to incorporate our views on both the volatility and return of the underlying. This enables us to consistently compare strikes on the basis of a number of risk measures, including average return, probability of profit, and the generalized Sharpe ratio. Chapter Eight examines the P/L distributions of common directional option structures.

The final section is about risk. Good risk control can't make money. Trading first needs edge. However, bad risk management will lead to losses.

Chapter Nine discusses trade sizing, specifically the Kelly criterion. The standard formulation is extended to allow for parameter estimation uncertainty, skewness of returns, and the incorporation of a stop level in the account.

The most dangerous risks are not related to price movement. The most dangerous risks are in the realm of the unknowable. Obviously, it is impossible to predict these, but Chapter Ten explores some historical examples. We don't know when these will happen again, but it is certain that they will. There is no excuse for blowing up due to repeat of a historical event.

It is inevitable that you will be wrong at times. The most dangerous thing is to forget this.

Summary

- Find a robust source of edge that is backed by empirical evidence and convincing reasons for its existence.
- Choose the appropriate option structure to monetize the edge.
- Size the position appropriately.
- Always be aware of how much you don't know.

CHAPTER 1
Options: *A Summary*

Option Pricing Models

> *Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.*
>
> —Box (1976)

Some models are wrong in a trivial way. They clearly don't agree with real financial markets. For example, an option valuation model that included the return of the underlying as a pricing input is trivially wrong. This can be deduced from put-call parity. Imagine a stock that has a positive return. Naively this will raise the value of calls and lower the value of puts. But put-call parity means that if calls increase, so do the values of the puts. Including drift leads to a contradiction. That idea is trivially wrong.

Every scientific model contains simplifying assumptions. There actually isn't anything intrinsically wrong with this. There are many reasons why this is the case, because there are many types of scientific models. Scientists use simplified models that they know are wrong for several reasons.

A reason for using a wrong theory would be because the simple (but wrong) theory is all that is needed. Classical mechanics is still widely used in science even though we now know it is wrong (quantum mechanics is needed for small things and relativity is needed for large or fast things). An example from finance is assuming normally distributed returns. It is doubtful anyone ever thought returns were normal. Traders have long known about extreme price moves and Osborne (1959), Mandelbrot (1963), and others studied the non-normal distribution of returns from the 1950s. (Mandelbrot cites the work of Mitchell [1915], Olivier [1926], and Mills [1927], although this research was not well known.) The main reason early finance theorists assumed normality was because it made the equations tractable.

Sometimes scientists might reason through a stretched analogy. For example, Einstein started his theory of the heat capacity of a crystal by first assuming the crystal was an ideal gas. He knew that this was obviously not the case. But he thought that the idea might lead to something useful. He had to start somewhere, even if he knew it was the wrong place. This model was metaphorical. A metaphorical model does not attempt to describe reality and need not rely on plausible assumptions. Instead, it aims to illustrate a non-trivial mechanism, which lies outside the model.

Other models aim to mathematically describe the main features of an observation without necessarily understanding its deeper origin. The GARCH family of volatility models are phenomenological, and don't tell us why the GARCH effects exist. Because these models are designed to describe particular features, there will be many other things they totally ignore. For example, a GARCH process has nothing to say about the formation of the bid-ask spread. The GARCH model is *limited,* but not *wrong*.

The most ambitious models attempt to describe reality as it truly is. For example, the physicists who invented the idea that an atom was a nucleus around which electrons orbited thought this was actually what atoms were like. But they still had to make simplifying assumptions. For example, when formulating the theory, they had to assume that atoms were not subject to gravity. And, in only trivial situations could the equations be analytically solved. The Black-Scholes-Merton (BSM) model was meant to be of this type.

But it isn't used that way at all.

The inventors of the model envisaged that the model would be used to find a fair value for options. Traders would input the underlying price, strike, interest rate, expiration date, and volatility and the model would tell them what the option was worth. The problem was that the volatility input needed to be the volatility over the life of the option, an unknown parameter. Although it was possible for a trader to make a forecast of future volatility, the rest of the market could and did make its own forecast. The market's option price was based on this aggregated estimate. This is the implied volatility, which became the fundamental parameter. Traders largely didn't think of the model as a predictive valuation tool but just as an arbitrage-free way to convert the quickly changing option prices into a slowly changing parameter: implied volatility. For most traders, BSM is not a predictive model; it is just a simplifying tool.

This isn't to say that BSM can't be used as a pricing model to get a fair value. It absolutely can. But even traders who do this will think in volatility terms. They will compare the implied volatility to their forecast volatility, rather than use the forecast volatility to price the option and compare it to the market value. By using the model backwards, these traders still benefit from the way BSM converts the option prices into a slowly varying parameter.

We need to examine the effects of the model assumptions in light of how the model is used. Although the assumptions make the model less realistic, this isn't important. The model wasn't used because it was realistic; it was used because it was useful.

Obviously, it is possible to trade options without any valuation model. This is what most directional option traders do. We can also trade volatility without a model. Traders might sell a straddle because they think the underlying will expire closer to the strike than the value of the straddle. However, to move beyond directional trading or speculating on the value of the underlying at expiration we will need a model.

The BSM model is still the benchmark for option pricing models. It has been used since 1973 and has direct ancestors dating to the work of Bachelier (1900) and Bronzin (1906). In terms of scientific theory, this age makes it a dinosaur. But just as dinosaurs were the dominant life form for about 190 million years for a reason, BSM has persisted because it is good.

We want an option pricing model for two reasons.

The first is so we can reduce the many, fast-moving option prices to a small number of slow-moving parameters. Option pricing models don't really *price* options. The market prices options though the normal market forces of supply and demand. Pricing models convert the market's prices into the parameters. In particular, BSM converts option prices to an implied volatility parameter. Now we can do all analysis and forecasting in terms of implied volatility, and if BSM was a perfect model, we would have a single, constant parameter.

The second reason to use a pricing model is to calculate a delta for hedging. Model-free volatility trading exists. Buying or selling a straddle (or strangle, butterfly, or condor, etc.) gives a position that is primarily exposed to realized volatility. But it will also be exposed to the drift. The most compelling reason to trade volatility is that it is more predictable than returns (drift) and the only way to remove this exposure is to hedge. To hedge we need a delta and for this we need a model. This is the most important criterion for an option trader to consider when deciding if a model is good enough. Any vaguely sensible model will reduce the many option prices to a few parameters, but a good model will let us delta hedge in a way that captures the volatility premium.

In this chapter we will examine the BSM model and see if it can meet this standard. By *BSM model* I mean the partial differential equation rather than the specific solution for European vanilla options. The particular boundary conditions and solution methods aren't a real concern here.

Derivations of BSM can be found in many places (see Sinclair, 2013, for an informal derivation). Here we will look at how the model is used.

Option Trading Theory

Here we will very briefly summarize the theory of option pricing and hedging. For more details refer to Sinclair (2010; 2013).

An option pricing model must include the following variables and parameters:

- Underlying price and strike; this determines the moneyness of the option.
- Time until expiration.
- Any factors related to carry of either the option or the underlying; this includes dividends, borrow rates, storage costs, and interest rates.
- Volatility or some other way to quantify future uncertainty.

A variable that is not necessary is the expected return of the underlying. Clearly, this is important to the return of an option, but it is irrelevant to the instantaneous value of the option. If we include this drift term, we will arrive at a contradiction. Imagine that we expect the underlying to rally. Naively, this means we would pay more for a call. But put-call parity means that an increase in call price leads to an increase in the price of the put with the same strike. This now seems consistent with us being bearish. Put-call parity is enough to make the return irrelevant to the current option price, but (less obviously perhaps) it is also enforced by dynamic replication.

This isn't an option-specific anomaly. There are many situations in which people agree on future price change, but this doesn't affect current price. For example, Ferrari would be justified in thinking that the long-term value of their cars is higher than their MSRP. But they can build the car and sell it at a profit right now. Their replication value as a manufacturer guarantees a profit without taking future price changes into account. Similarly, market-makers can replicate options without worrying about the underlying return. And if they do include the return, they can be arbed by someone else.

The canonical option-pricing model is BSM. Ignoring interest rates for simplicity, The BSM PDE for the price of a call, C, is

$$\frac{\sigma^2 S^2}{2}\frac{\partial^2 C}{\partial S^2} + \frac{\partial C}{\partial t} = 0 \tag{1.1}$$

where S is the underlying price, σ is the volatility of the underlying, and t is the time until expiration of the option.

Or using the standard definitions where Γ is the second partial derivative of the option price with respect to the underlying and θ is the derivative of the option price with respect to time,

$$\frac{\sigma^2 S^2}{2}\Gamma + \theta = 0 \tag{1.2}$$

This is then solved using standard numerical or analytical techniques and with the final condition being the payoff of the particular option.

This relationship between Γ and θ is crucial for understanding how to make money with options. Imagine we are long a call and the underlying stock moves from S_t to S_{t+1}.The delta P/L of this option will be the average of the initial delta, Δ, and the final delta, all multiplied by the size of the move. Or

$$P/L = \left(\frac{\Delta}{2} + \frac{(\Delta + \Gamma(S_{t+1} - S_t))}{2} \right) (S_{t+1} - S_t) \tag{1.3}$$

If this option was initially delta-hedged, the P/L over this price move would be

$$P/L = \frac{\Gamma(S_{t+1} - S_t)^2}{2} \tag{1.4}$$

Next note that

$$(S_{t+1} - S_t)^2 \simeq \sigma^2 S_t^{\ 2} \tag{1.5}$$

so that the profit in hedging over each time interval is

$$P/L = \frac{\Gamma \sigma^2 S_t^{\ 2}}{2} \tag{1.6}$$

(Although equation 1.6 is only asymptotically true, if we worked with an infinitesimal price change, this derivation would be exact.

This is the first term of the BSM differential equation. Literally, BSM says that these profits from rebalancing due to gamma are exactly equal to the theta of the option. Expected movement cancels time decay. The only way a directionally neutral option position will make money is if the option's implied volatility (which governs theta) is not the same as the underlying's realized volatility (which determines the rebalancing profits). This is true no matter which structure is chosen and the particulars of the hedging scheme.

If we can identify situations where this volatility mismatch occurs, the expected profit from the position will be given by

$$P/L = \text{Vega} |\sigma_{implied} - \sigma_{realized}| \tag{1.7}$$

This is the fundamental equation of option trading. All the “theta decay” and “gamma scalping” profits and losses are tied up in this relationship.

Note also that this vega P/L will affect directional option trades. If we pay the wrong implied volatility level for an option, we might still make money but we would have been better off replicating the option in the underlying.

The BSM equation depends on a number of financial and mathematical assumptions.

- The underlying is a tradable asset.
- There is a single, risk-free interest rate.
- The underlying can be shorted.
- Proceeds from short sales can be invested at the risk-free rate.
- All cash flows are taxed at the same rate.
- The underlying's returns are continuous and normally distributed with a constant volatility.

Traders have devised various workarounds to address these limiting assumptions (see Appendix One). The most important of these is the concept of the implied volatility surface. If the BSM were an accurate descriptive theory, all options on a given underlying would have one volatility. This is not true. For the BSM equation to reproduce market option prices, options with different strikes have different implied volatilities (the *smile*) and options with different maturities have different implied volatilities (the *term structure*). These implied volatilities make up the IV surface. An example is in shown in Figure 1.1.

The IV surface exists partially because the BSM is mathematically misspecified. The underlying does not have returns that are continuous and normally distributed with a constant volatility. However, even a model that perfectly captured the underlying dynamics would need a fudge factor like the implied volatility surface. Some of the reasons for its existence have nothing to do with the underlying. Different options have different supply and demand, and these distort option prices. Because of this, there is often an edge in selling options with high volatilities relative to others on the same underlying (see the section on the implied skewness premium in Chapter Four).

Equation 1.7 gives the average PL of any hedged option position, but there is a wide dispersion of results for this mean, and the spread of this distribution decreases with the number of hedges. Figure 1.2 shows the PL distribution of a short straddle that is never re-hedged, and Figure 1.3 shows the distribution when the straddle is re-hedged every day. The underlying paths were generated from 10,000 realizations of a GBM. The implied and realized volatility were equal so we expect an average PL of zero.

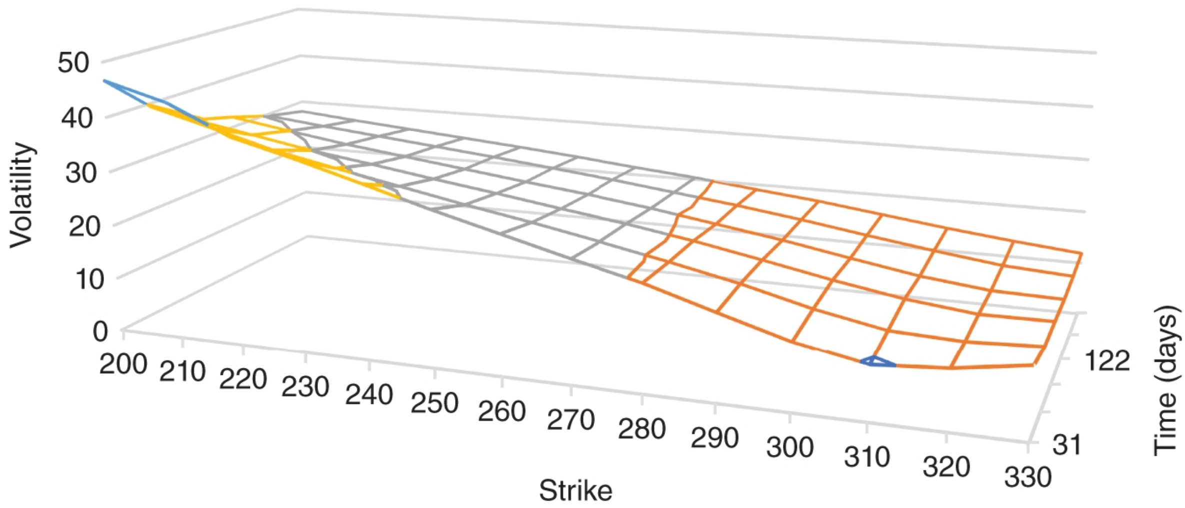


FIGURE 1.1 The implied volatility surface for SPY on September 10, 2019.

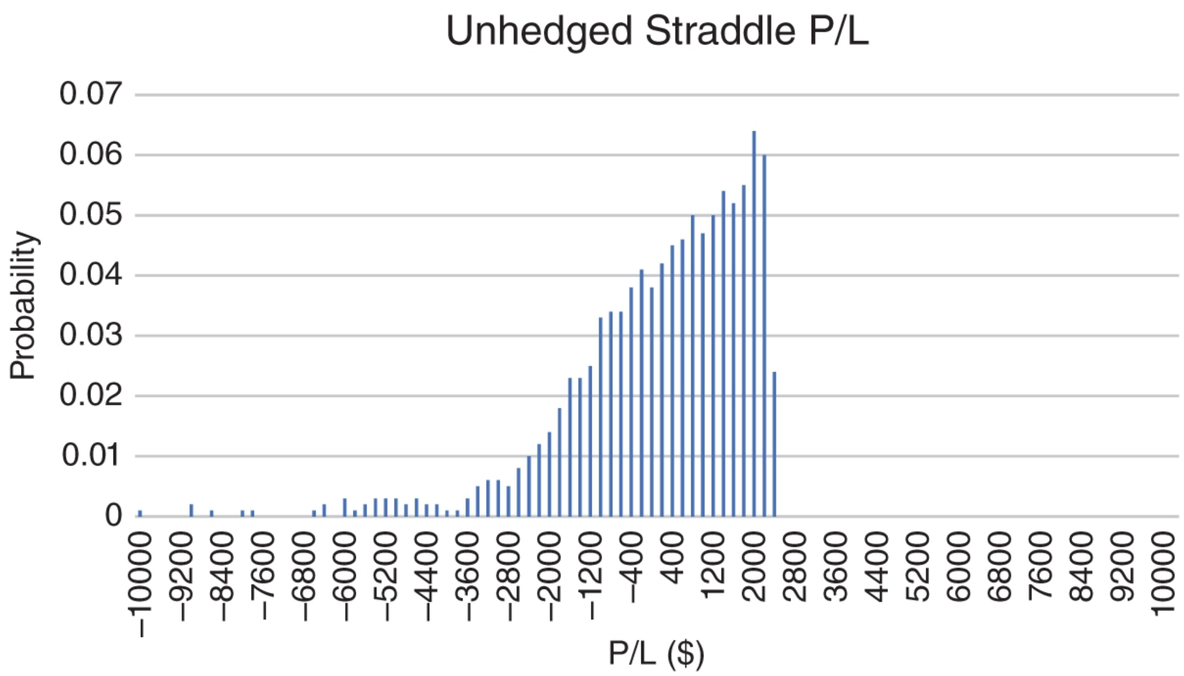


FIGURE 1.2 The terminal PL distribution of a single short one-year ATM straddle that is never re-hedged. Stock price is $100, rates are zero, and both realized and implied volatilities are 30%.

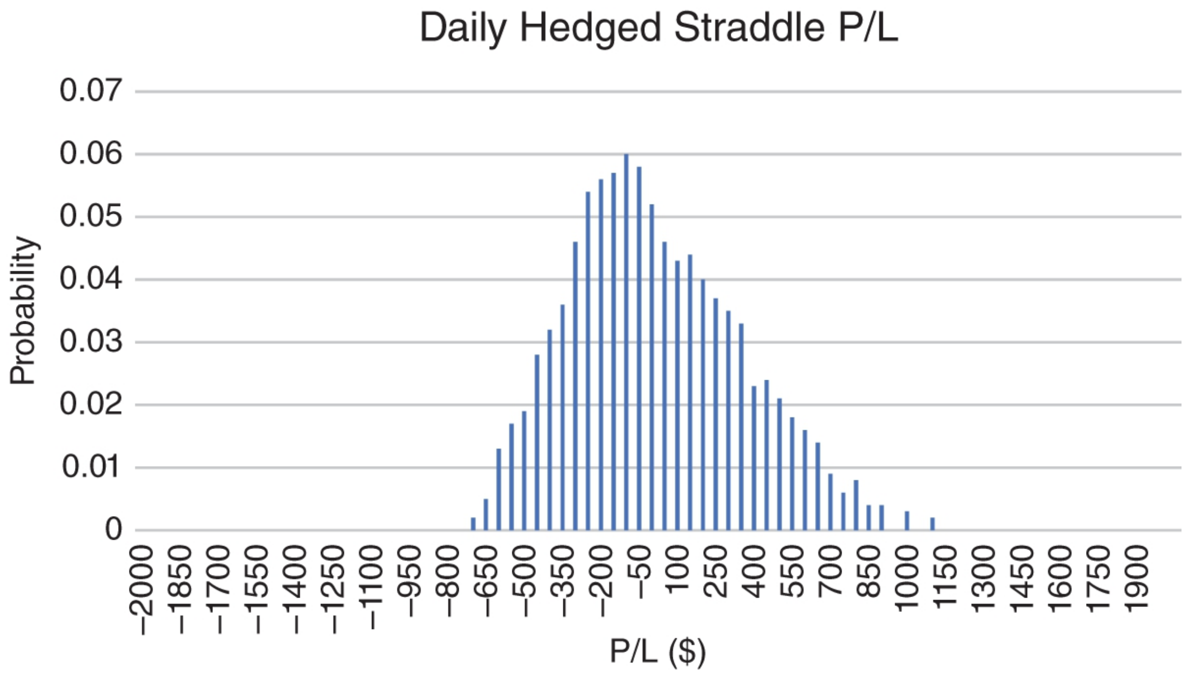


FIGURE 1.3 The terminal PL distribution of a single one-year ATM straddle that is hedged daily. Stock price is $100, rates are zero, and both realized and implied volatilities are 30%.

The dependence of the standard deviation of the PL distribution on the number of hedges is shown in Figure 1.4.

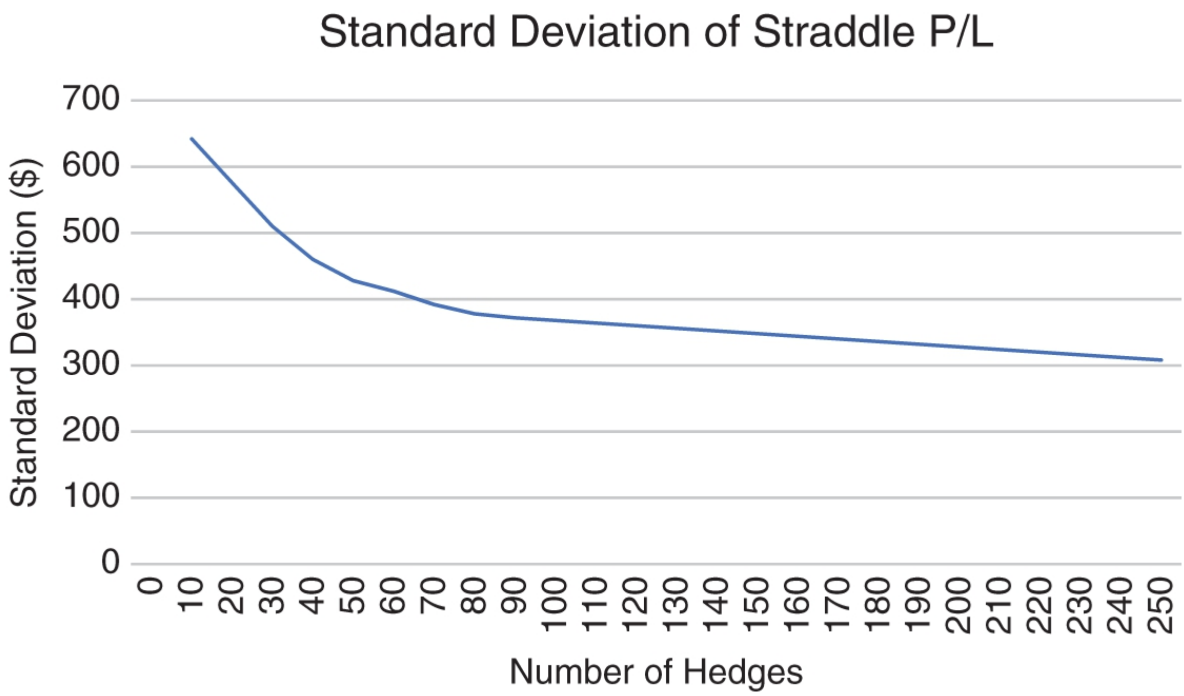


FIGURE 1.4 The standard deviation of the terminal PL distribution of a single one-year ATM straddle as a function of the number of hedges. Stock price is $100, rates are zero, and both realized and implied volatilities are 30%.

The reason to hedge less frequently and accept a wider standard deviation of results is that hedging costs money. All hedges incur transaction costs (brokerage, exchange fees, and infrastructure costs). Costs like this are an easily forgotten drain on a portfolio. Individually they are small, but they accumulate. To emphasize this point, Table 1.1 compares the summary statistics of results for the daily hedged short straddle when there is a transaction cost of $.10 per share and when hedges are costless.

The difference between these two cases is roughly equivalent to misestimating volatility by two points.

In practice, aggressive re-hedging is done by market-making firms and some volatility specialists. The vast majority of retail and buy-side users seldom or never hedge. The relevant theory for those hoping to approximate continuous hedging is discussed in Sinclair (2013). In this book we will generally assume that no re-hedging takes place. These results are also applicable to those who hedge infrequently. They can just assume that the original position has been closed and a new one opened. So, a one-year position that is hedged after a month would thereafter have the expected distribution of an 11-month option.

TABLE 1.1 Statistics for the Short One-Year ATM Daily Hedged Straddle With and Without Hedging Costs (stock price is $100, rates are zero, and both realized and implied volatilities are 30%.)

Statistic	Costless Hedges	$.10/Share Hedges
Average	−$6.10	−$121.54
Median	−$49.85	−$111.68
Percent profitable	44%	30%

Conclusion

The BSM model gives the replication strategy for the option. The expected return of the underlying is irrelevant to this strategy. The only distributional property of the underlying that is used in the BSM model is the volatility. A hedged position will, on average, make a profit proportional to the difference between the volatility implied by the option market price (by inverting the BSM model) and the subsequent realized volatility. The choice of the option structure and hedging scheme can change the shape of the PL distribution, but not the average value. These choices are far from immaterial, but successful option trading depends foremost on finding situations in which the implied volatility is mispriced.

Summary

- Arbitrage-free option pricing models do not include the underlying return. BSM includes only volatility.
- Inverting the pricing model using the option's market price as an input gives the implied volatility.
- The average profit of a hedged option position is proportional to the difference between implied volatility and the subsequent realized volatility.
- Practical option hedging is designed to give an acceptable level of variance for a given amount of transaction costs.

CHAPTER 2
The Efficient Market Hypothesis and Its Limitations

A lot of trading books propagate the myth that successful trading is based on discipline and persistence. This might be the worst advice possible. A trader without a real edge who persists in trading, executing a bad plan in a disciplined manner, will lose money faster and more consistently than someone who is lazy and inconsistent. A tough but unskilled fighter will just manage to stay in a losing fight longer. All she will achieve is being beaten up more than a weak fighter would.

Another terrible weakness is optimism. Optimism will keep losing traders chasing success that will never happen. Sadly, hope is a psychological mechanism unaffected by external reality.

Emotional control won't make up for lack of edge. But, before we can find an edge, we need to understand why this is hard and where we should look.

The Efficient Market Hypothesis

The traders' concept of the efficient market hypothesis (EMH) is “making money is hard.” This isn't wrong, but it is worth looking at the theory in more detail. Traders are trying to make money from the exceptions to the EMH, and the different types of inefficiencies should be understood, and hence traded, differently.

The EMH was contemporaneously developed from two distinct directions. Paul Samuelson (1965) introduced the idea to the economics community under the umbrella of “rational expectations theory.” At the same time, Eugene Fama's studies (1965a, 1965b) of the statistics of security returns led him to the theory of “the random walk.”

The idea can be stated in many ways, but a simple, general expression is as follows:

A market is efficient with respect to some information if it is impossible to profitably trade based on that information.

And the “profitable trades” are risk-adjusted, after all costs.

So, depending on the information we are considering, there are many different EMHs, but three in particular have been extensively studied:

- The strong EMH in which the information is anything that is known by anyone
- The semi-strong EMH in which the information is any publicly available information, such as past prices, earnings, or analysts' studies
- The weak EMH in which the information is past prices

The EMH is important as an organizing principle and is a very good approximation to reality. But, it is important to note that no one has ever believed that any form of the EMH is strictly true. Traders are right. Making money is hard, but it isn't impossible. The general idea of the theory and also the fact it isn't perfect is agreed on by most successful investors and economists.

> “I think it is roughly right that the market is efficient, which makes it very hard to beat merely by being an intelligent investor. But I don't think it's totally efficient at all. And the difference between being totally efficient and somewhat efficient leaves an enormous opportunity for people like us to get these unusual records. It's efficient enough, so it's hard to have a great investment record. But it's by no means impossible.”

—Charlie Munger

Even one of the inventors of the theory, Eugene Fama, qualified the idea of efficiency by using the word *good* instead of *perfect*.

> “In an efficient market, at any point in time, the actual price of a security will be a good estimate of its intrinsic value.”

—Eugene Fama

There is something of a paradox in the concept of market efficiency. The more efficient a market is, the more random and unpredictable the returns will be. A perfectly efficient market will be completely unpredictable. But the way this comes about is through the trading of all market participants. Investors all try to profit from any informational advantage they have, and by doing this their information is incorporated into the prices. Grossman and Stiglitz (1980) use this idea to argue that perfectly efficient markets are impossible. If markets were efficient, traders wouldn't make the effort to gather information, and so there would be nothing driving markets toward efficiency. So, an equilibrium will form where markets are mostly efficient, but it is still worth collecting and processing information.

(This is a reason fundamental analysis consisting of reading the *Wall Street Journal* and technical analysis using well-known indicators is likely to be useless. Fischer Black [1986] called these people “noise traders.” They are the people who pay the good traders.)

There are other arguments against the EMH. The most persuasive of these are from the field of behavioral finance. It's been shown that people are irrational in many ways. People who do irrational things should provide opportunities to those who don't. As Kipling (1910) wrote, “If you can keep your head when all about you are losing theirs, … you will be a man, my son.”

In his original work on the EMH, Fama mentioned three conditions that were sufficient (although not necessary) for efficiency:

- Absence of transaction costs
- Perfect information flow
- Agreement about the price implications of information

Helpfully for us, these conditions do not usually apply in the options market. Options, particularly when dynamically hedged, have large transaction costs. Information is not universally available and volatility markets often react slowly to new information. Further, the variance premium cannot be directly traded. Volatility markets are a good place to look for violations of the EMH.

Let's accept that the EMH is imperfect enough that it is possible to make money. The economists who study these deviations from perfection classify them into two classes: risk premia and inefficiencies. A risk premium is earned as compensation for taking a risk, and if the premium is mispriced, it will be profitable even after accepting the risk. An inefficiency is a trading opportunity caused by the market not noticing something. An example is when people don't account for the embedded options in a product.

There is a joke (not a funny one) about an economist seeing a $100 bill on the ground. She walks past it. A friend asks: “Didn't you see the money there?” The economist replies: “I thought I saw something, but I must've imagined it. If there had been $100 on the ground, someone would've picked it up.” We know that the EMH is not strictly true, but the money could be there for two different reasons. Maybe it is on a busy road and no one wants to run into traffic. This is a risk premium. But maybe it is outside a bar where drunks tend to drop money as they leave. This is an inefficiency. There is also the possibility that the note was there purely by luck.

It is often impossible to know whether a given opportunity is a risk premium or an inefficiency, and a given opportunity will probably be partially both. But it is important to try to differentiate. A risk premium can be expected to persist: the counterparty is paying for insurance against a risk. They may improve their pricing of the insurance, but they will probably continue to pay something.

By contrast, an inefficiency will last only until other people notice it. And failing to differentiate between a real opportunity and a chance event will only lead to losses.

Some traders will profit from inefficiencies. Not all traders will. A lot of traders will use meaningless or widely known information. Many forecasts are easy. I can predict the days the non-farm payroll will be released. I can predict what days fall on weekends. I can predict the stock market closes at 4 p.m. eastern time. In many cases, making a good prediction is the easy part. The hard part is that the forecast has to be better than the market's, which the consensus of everyone else's prediction is. For developed stock indices the correlation between the daily range on one day and the next is roughly between 65% and 70%. So a very good volatility estimator is that it will be what it was the day before (a few more insights like this will lead you to GARCH). It is both hard and profitable to make an even slightly better one-day forecast. And whether it is because the techniques that are used are published, employees leave and take information with them, or just that several people have a similar idea at the same time, these forecast edges don't last forever.

Aside: Alpha Decay

The extinction of floor traders is an example of a structural shift in markets destroying a job. Similar to most people, traders tend to think that their skills are special, and their jobs will always be around. This isn't true. The floors have gone. Fixed commissions have gone. Investment advisors are being replaced by robo-advisors. There are fewer option market-makers, each trading many more stocks than in the past. Offshoring will definitely come to trading, and it is quite possible that a market structure such as a once-a-day auction could replace continuous trading.

But as well as these structural changes, the alpha derived from market inefficiencies (as opposed to the beta of exposure to a mispriced risk factor) doesn't last forever. Depending on how easy it is to trade the effect, the half-life of an inefficiency-based strategy seems to be between 6 months and 5 years. Mclean and Pontiff (2016) showed that the publication of a new anomaly lessens its returns by up to 58%. And publication isn't the only thing that erodes alpha. Chordia et al. (2014) showed that increasing liquidity also reduces excess returns by about 50%. Sometimes the anomaly exists only because it isn't worth the time of large traders to get involved. A similar effect is that the easy access to data will kill strategies. Sometimes the alpha isn't due to a wrinkle in the financial market. It is due to the costs of processing information.

Just as some traders will profit by using a stupid idea like candlestick charting, some traders will succeed for a while with an overfit model. I'm in no way using this to condone data-mining, but we can learn a valid lesson from this. As Guns and Roses pointed out, “nothing lasts forever.” Lucky strategies will never last but even the best, completely valid strategy will have a lifetime. So, when you are making money don't think that being “prudent” is a good idea. The right thing to do is to be as aggressive as possible. Amateurs go broke for a lot of reasons, but professionals often suffer in bad times because they didn't fully capitalize on good times, instead thinking that making steady but small profits was the best thing to do.

They also spend too much in good times, forgetting that they won't last. I've had a floor trader tell me about his new Ferrari about an hour before laughing about the stupid spending habits of NFL and NBA players (the last I heard he was selling houses). Many times, traders have short careers because a valid strategy dies. Amateurs blow up, but professionals don't allow for alpha-decay. For example, many floor traders didn't survive the death of the open-outcry pits. Their edge disappeared, and their previous spending habits left them with little. (In this case "trickle-down" economics was correct, as profits from market-making trickled down to prostitutes, strippers, and cocaine dealers. At least it wasn't wasted.)

Behavioral Finance

> *Think about how stupid the average person is, then realize half of them are stupider than that.*
>
> —George Carlin

The history of markets is nowhere near as big as we often assume. For example, equity options have only been traded in liquid, transparent markets since the CBOE opened in 1973. S&P 500 futures and options have only been traded since 1982. The VIX didn't exist until 1990 and wasn't tradable until 2004. And the average lifetime of an S&P 500 company is only about 20 years. In the long term, values are related to macro variables such as inflation, monetary policy, commodity prices, interest rates, and earnings. And these change on the order of months and years. Even worse, they are all co-dependent.

So, what might seem like a decent length of history that we can study and look for patterns, quite possibly isn't (this does not apply to HFT or market-making where a huge number of data points can be collected in what is essentially a stationary environment). When it comes to volatility markets, I think that although there appear to be many thousands of data points, there might only be dozens. A better way to think of market data might be that we are seeing a small number of data points, and that they occur a lot of times.

I think this makes quantitative analysis of historical data much less useful than is commonly thought.

But there is something that has been constant: human nature.

Humans have been essentially psychologically unchanged for 300,000 years when *Homo sapiens* (us) first appeared. This means that any effect that can conclusively be attributed to psychology will effectively have 300,000 years of evidence behind it. This seems to be potentially a much better source for gaining clarity.

The problem with psychological explanations (for anything) is that they are incredibly easy to postulate. As the baseball writer Bill James was reported to say, "Twentieth-century man uses psychology exactly like his ancestors used witchcraft; anything you don't understand, it's psychology." The finance media is always using this kind of pop psychology to justify what happened that day. "Traders are exuberant" when the market goes up a lot; "Traders are cautiously optimistic" when it goes up a little, and so on. I try not to do this, but I'm as guilty as anyone else. I think psychology could be incredibly helpful, but we have to be very careful in applying it. Ideally, we want several psychological biases pointing to one tradeable anomaly, and we want them to have been tested on a very similar situation to the one we intend to trade.

Further, traders aren't psychologists and reading behavioral finance at any level from pop psychology to real scientific journals is probably just going to lead to hunches and guesses. To be fair, traders currently make the same mistakes from reading articles about geopolitics or economics. One week, traders will be experts on the effects of tariffs on soybeans and the next week they will be talking about Turkish interest rates. It is far easier to sound knowledgeable than to actually be so. It isn't obvious that badly applied behavioral psychology is any more useful than badly applied macroeconomics. And it is obvious that traders can't do better than misapply, either.

After I explained this nihilistic view to an ex-employer he said, "Well, I have to do something." And what we do is exactly what I've said isn't very good: we apply statistics and behavioral finance. These are far from perfect tools, but they are the best we have. The edges they give will be small, but some edges can be found. We will always know only a small part of what can be known. Making money is hard.

Proponents of behavioral finance contend that various psychological biases cause investors to systematically make mistakes that lead to market inefficiencies. Behavioral psychology was first applied to finance in the 1980s, but for decades before that psychologists were studying the ways people actually made decisions under uncertainty.

The German philosopher Georg Hegel is famous (as much as any philosopher can be famous) for his triad of thesis, antithesis, and synthesis. A thesis is proposed. An antithesis is the negation of that idea. Eventually, synthesis occurs, and the best part of thesis and antithesis are combined to form a new paradigm. Ignoring the fact that Hegel never spoke about this idea, the concept is quite useful for describing the progress of theories. A theory is proposed. Evidence is found that supports the theory. Eventually it becomes established orthodoxy. But after a period, either for theoretical reasons or because new evidence emerges, a new theory is proposed that is strongly opposed to the first one. Arguments ensue. Many people become more dogmatic and hold on tightly to their side of the divide, but eventually aspects of both thesis and antithesis are used to construct a new orthodoxy.

From the early 1960s until the late 1980s the EMH was the dominant paradigm among finance theorists. These economists modeled behavior in terms of rational individual decision-makers who made optimal use of all available information. This was the thesis.

In the 1980s an alternative view developed, driven by evidence that the rationality assumption is unrealistic. Further, the mistakes of individuals may not disappear in the aggregate. People are irrational and this causes markets to be inefficient. Behavioral finance was the antithesis.

Synthesis hasn't yet arrived, but behavioral finance is now seen as neither an all-encompassing principle nor a fringe movement. It augments, not replaces, traditional economics.

What have we learned from behavioral finance?

First, behavioral finance has added to our understanding of market dynamics. Even in the presence of rational traders and arbitrageurs, irrational “noise” traders will prevent efficiency. And although it is possible to justify the existence of bubbles and crashes within a rational expectations framework (for example, Diba and Grossman, 1988), a behavioral approach gives more reasonable explanations (for example, Abreu and Brunnermeier, 2003, and De Grauwe and Grimaldi, 2004).

Second, we are now aware of a number of biases, systematic misjudgments that investors make. Examples include the following:

- **Overconfidence:** Overconfidence is an unreasonable belief in one's abilities. This leads traders to assign too narrow a range of possibilities to the outcome of an event, to underestimate the chances of being wrong, to trade too large, and to be too slow to adapt.
- **Overoptimism:** Overconfidence compresses the range of predictions. Overoptimism biases the range, so traders consistently predict more and better opportunities than really exist.
- **Availability heuristic:** We base our decisions on the most memorable data even if it is atypical. This is one reason teeny options are overpriced. It is easy to remember the dramatic events that caused them to pay off, but hard to remember the times when nothing happened and they expired worthless.
- **Short-term thinking:** This thinking shows the irrational preference for short-term gains at the expense of long-term performance.
- **Loss aversion:** Investors dislike losses more than they like gains. This means they hold losing positions, hoping for a rebound even when their forecast has been proven wrong.
- **Conservatism:** Conservatism is being too slow to update forecasts to reflect new information.
- **Self-attribution bias:** This bias results from attributing success to skill and failure to luck. This makes Bayesian updating of knowledge impossible.
- **Anchoring:** Anchoring occurs when relying too much on an initial piece of information (the "anchor") when making a forecast. This leads traders to update price forecasts too slowly because the current price is the anchor and seems more "correct" than it should.

And there are at least 50 others.

It is these types of biases that traders have tried to use to find trades with edge. Results have been mixed. There are so many biases that practically anything can be explained by one of them. And sometimes there are biases that are in direct conflict. For example, investors underreact, but they also overreact. Between these two biases you should be able to explain almost any market phenomena. The psychologists and finance theorists working in the field are not stupid. They are aware of these types of difficulties and are working to disentangle the various effects. The field is a relatively new one and it is unfair and unrealistic to expect there to be no unresolved issues. The problem is not really with the field or the serious academic papers. The problem is with pop psychology interpretations and investors doing “bias mining” to justify ideas.

It is common in science for a new idea to be overly hyped, particularly those that are interesting to lay people (traditional finance is not interesting). In the 1970s there were popular books about catastrophe theory, a branch of physics that was meant to explain all abrupt state changes and phase transitions. It didn't. In the 1990s, chaos theory was meant to explain practically everything, including market dynamics. It didn't. Behavioral finance is being overexposed because it is interesting. It provides plenty of counterintuitive stories and also a large amount of schadenfreude. We can either feel superior to others making stupid mistakes or at least feel glad that we aren't the only ones who make these errors.

And people love intuitive explanations. We have a great need to understand things, and behavioral finance gives far neater answers than statistics of classical finance theory. Even though behavioral finance doesn't yet have a coherent theory of markets, the individual stories give some insight. They help to demystify. This is reassuring. It gives us a sense of control over our investments.

A science becoming interesting to the general public doesn't necessarily mean it is flawed. For example, there have been hundreds of popular books on quantum mechanics. However, behavioral finance does have some fairly serious problems to address.

Just as in conventional finance theory, behavioral finance studies individual decision-making despite the fact that people do not make investing decisions independently of the rest of society. Everyone is influenced by outside factors. Most people choose investments based on the recommendations of friends (Katona, 1975). And professionals are also influenced by social forces (Beunza and Stark, 2012). Over the last 30 years the sociology of markets has been an active research field (for example, Katona, 1975, Fligstein and Dauter, 2007, and references therein), but this work hasn't yet been integrated into behavioral finance. Because behavioral finance largely ignores the social aspects of trading and investing, we don't have any idea of how the individual biases aggregate and their net effect on market dynamics. This is necessary because, even though we don't understand how aggregate behavior emerges, it is very clear that markets cater to irrational behavior rather than eradicate it. For example, the services of financial advisors, stock brokers, and other financial intermediaries made up 9% of the US GDP (Philippon, 2012) despite the fact that they are almost all outperformed by much cheaper index funds and ETFs.

Next, behavioral finance has largely limited itself to the study of cognitive errors. There are many other types of nonrational behavioral inputs into decision-making, including emotion, testosterone levels, substance abuse, and the quest for status.

And behavioral finance gives no coherent alternative theory to the EMH. A catalog of biases and heuristics—the mistakes people make—is not a theory. A list of facts does not make a theory. Of course, sometimes observations are necessary before a theory can be formulated. Mendeleev drew the periodic table well before the atomic structure of matter was understood. We knew species existed well before we understood the process of speciation by natural selection. Still, to be scientific, behavioral finance eventually needs to lead to a unifying theory that gives explanations of the current observations and makes testable predictions.

Behavioral finance can still help. Whenever we find something that looks like a good trading idea we need to ask, “Why is this trade available to me?” Sometimes the answer is obvious. Market-makers get a first look in exchange for providing liquidity. Latency arbitrage is available to those who make the necessary investments in technology. ETF arbitrage is available to those with the capital and legal status to become authorized participants. But often a trade with positive edge is available to anyone who is interested. Remembering the joke about the economists, “Why is this money sitting on the ground?” Risk premia can often be identified by looking at historical data, but behavioral finance can help to identify real inefficiencies. For example, post-earnings announcement drift can be explained in terms of investor underreaction. Together with historical data, this gives me enough confidence to believe that the edge is real. The data suggest the trade, but the psychological reason gives a theoretical justification.

High-Level Approaches: Technical Analysis and Fundamental Analysis

Technical analysis is the study of price and volume to predict returns.

Technical Analysis

Aronson (2007) categorized technical analysis as either subjective or objective. It is a useful distinction.

Subjective technical analysis incorporates the trader's discretion and interpretation of the data. For example, “If the price is over the EWMA, I might get long. It depends on a lot of other things.” These methods aren't wrong. They aren't even methods. Subjectivity isn't *necessarily* a problem in science. A researcher subjectively chooses what to study and then subjectively chooses the methods that make sense. But if subjectivity is applied as part of the trading approach, rather than the research, then there is no way to test what works and what doesn't. Do some traders succeed with subjective methods? Obviously. But until we also know how many fail, we can't tell if the approach works. Further, the decisions different traders who use ostensibly the same method make won't be the same or even based on the same inputs. There is literally no way to test subjective analysis.

Some things that are intrinsically subjective are Japanese candlesticks, Elliot waves, Gann angles, trend lines, and patterns (flags, pennant, head, and shoulders, etc.). These aren't methods. In the most charitable interpretation, they are a framework for (literally) looking at the market. It is *possible* that using these methods can help the trader implicitly learn to predict the market. But more realistically, subjective technical analysis is almost certainly garbage. I can't prove the ideas don't work. No-one can. They are unfalsifiable because they aren't clearly defined. But plenty of circumstantial evidence exists that this analysis is worthless. None of the large trading firms or banks has desks devoted to this stuff. They have operations based on stat arb, risk arb, market-making, spreading, yield curve trading, and volatility. No reputable, large firm has a Japanese candlestick group.

As an ex-boss of mine once said, “That isn't analysis. That is guessing.”

Any method can be applied subjectively, but only some can be applied objectively. Aronson (2007) defines objective technical analysis as “well-defined repeatable procedures that issue unambiguous signals.” These signals can then be tested against historical data and have their efficacy measured. This is essentially quantitative analysis.

It seems likely that some of these approaches can be used to make money in stocks and futures. But each individual signal will be very weak and to make any consistent money various signals will need to be combined. This is the basis of statistical arbitrage. This is not within the scope of this book.

However, we do need to be aware of a bad classic mistake when doing quantitative analysis of price or return data: data mining.

This is where we sift through data using many methods, parameters, and timescales. This is almost certain to lead to some strategy that has in-sample profitability. When this issue is confined to choosing the parameters of a single, given strategy it is usually called *overfitting*. If you add enough variables, you can get a polynomial to fit data arbitrarily well. Even if you choose a function or strategy in advance, by “optimizing” the variables you will the get the best in-sample fit. It is unlikely to be the best out of sample. Enrico Fermini shared that the mathematician and economist John von Neumann said, “With four parameters I can fit an elephant, and with five I can make him wiggle his trunk” (Dyson, 2004).

This mistake isn't only made by traders. Academics also fall into the trap. The first published report of this was Ioannidis (2005). Subsequently, Harvey et al. (2016) and Hou et al. (2017) discussed the impact of data mining on the study of financial anomalies.

There are a few ways to avoid this trap:

- The best performer out of a sample of back-tested rules will be positively biased. Even if the underlying premise is correct, the future performance of the rule will be worse than the in-sample results.
- The size of this bias decreases with larger in-sample data sets.
- The larger the number of rules (including parameters), the higher the bias.
- Test the best rule on out-of-sample data. This gives a better idea of its true performance.
- The ideal situation is when there is a large data set and few tested rules.

Even after applying these rules, it is prudent to apply a bias correcting method.

The simplest is Bonferroni's correction. This scales any statistical significance number by dividing by the number of rules tested. So, if your test for significance at the 95% confidence level (5% rejection) shows the best rule is significant, but the rule is the best performer of 100 rules, the adjusted rejection level would be 5%/100 or 0.005%. So, in this case, a *t*-score of 2 for the best rule doesn't indicate a 95% confidence level. We would need a *t*-score of 2.916, corresponding to a 99.5% level for the single rule. This test is simple but not powerful. It will be overly conservative and skeptical of good rules. When used for developing trading strategies this is a strength.

A more advanced test is White's reality check (WRC). This is a bootstrapping method that produces the appropriate sampling distribution for testing the significance of the best strategy. The test has been patented and commercial software packages that implement the test can be bought. However, the basic algorithm can be illustrated with a simple example.

We have two strategies, A and B, which produce daily returns of 2% and 1% respectively. Each was developed by looking at 100 historical returns. We can use WRC to determine if the apparent outperformance of strategy A is due to data mining:

- Using sampling with replacement, generate a series of 100 returns from the historical data.

- Apply the strategies (A and B) to this ahistorical data to get the pseudo-strategies A' and B'.
- Subtract the mean return of A from A' and B from B'.
- Calculate the average return of the return-adjusted strategies, A" and B".
- The larger of the returns of A" and B" is the first data point of our sample distribution.
- Repeat the process *N* times to generate a complete distribution. This is the sampling distribution of the statistic, maximum average return of the two rules with an expected return of zero.
- The *p*-value (probability of our best rule being truly the better of the two) is the proportion of the sampling distribution whose values exceed the returns of A, that is, 2%.

A realistic situation would involve comparing many rules. It is probably worth paying for the software.

There is also a totally different and complementary way to avoid overfitting. Forget about the time series of the data and study the underlying phenomenon. A hunter doesn't much care about the biochemistry of a duck, but she will know a lot about their actual behavior. In this regard a trader is a hunter, rather than a scientist. Forget about whether volatility follows a GARCH(1,1) or a T-GARCH(1,2) process; the important observation is that it clusters in the short term and mean reverts in the long term. If the phenomenon is strong enough to trade, it shouldn't be crucial what exact model is used. Some will always be better in a sample, but that is no guarantee that they will work best out of a sample.

As an example, this is the correct way to find a trading strategy.

There is overwhelming evidence that stocks have momentum. Stocks that have outperformed tend to continue outperforming. This has been observed for as long as we have data (see Geczy and Samnov [2016], Lempérière et al. [2014], and Chabot et al. [2009]) and in many countries (for example, Fama and French, 2010). The observation is robust with respect to how momentum is defined and the time scales over which it is measured. In the trading world, the evidence for stock momentum is overwhelming. Starting from this fact, design a simple model to measure momentum (e.g., 6-month return). Then sort stocks by this metric and buy the ones that score well.

The worst thing to do is take a predefined model and see if it works. Has a 30-day, 200-day moving average crossover been predictive of VIX futures? What if we change the first period to 50 days? I don't know or care.

Fundamental Analysis

Fundamental analysis aims to predict returns by looking at financial, economic, and political variables. For example, a fundamental stock analyst might look at earnings, yield, sales, and leverage. A global macro trader might consider GDP, currency levels, trade deficit, and political stability.

Fundamental analysis, particularly global macro, is particularly susceptible to subjectivity. It also tempts otherwise intelligent people to make investment decisions based on what they read in the *Wall Street Journal* or *The Economist*. It is exceedingly unlikely that someone can consistently profit from these public analyses, no matter how well the story is sourced or how smart the reader is.

Consider these statements from “experts”:

> “Financial storm definitely passed.”
>
> **—Bernard Baruch, economic advisor to presidents Woodrow Wilson and Franklin Roosevelt in a cable to Winston Churchill, November 1929**

Stocks dropped for the next 3 years, with the Dow losing 33% in 1930, 52% in 1931, and 23% in 1932.

> “The message of October 1987 should not be taken lightly. The great bull market is over.”
>
> **—Robert Prechter, prominent Elliot wave theorist and pundit, in November 1987. The Dow rallied for 11 of the next 12 years, giving a return (excluding dividends) of over 490%.**

> “A bear market is likely… It could go down 30% or 40%.”
>
> **—Barton Biggs, chief strategist for Morgan Stanley, October 27, 1997**

The Dow had its largest one-day gain on October 28 and continued to rally hard for the next 6 months.

In most situations it is just mean to make fun of people's mistakes. We all make mistakes. But the people I have quoted have proclaimed themselves experts in a field where real expertise is very, very rare.

And evidence of this is more than anecdotal.

The poor prediction skill of experts is a general phenomenon. Gray (2014) summarizes the results of many studies that show that simple, systematic models outperform experts in fields as diverse as military tactics, felon recidivism, and disease diagnosis. Expertise is needed to build the models, but experts should not make case-by-case decisions.

Koijen et al. (2015) show that surveys of economic experts (working for corporations, think tanks, chambers of commerce, and NGOs) have a negative correlation to future stock returns. They were also contraindicative for the returns of currencies and bonds. This effect applies across 13 equity markets, 19 currencies, and 10 fixed income markets. A simple “fade the experts” strategy would have given a Sharpe ratio of 0.78 from 1989 to 2012.

Financial advisors are equally bad. Jenkinson et al. (2015) look at the performance of advisors in picking mutual funds. They conclude with, “We find no evidence that these recommendations add value, suggesting that the search for winners, encouraged and guided by investment consultants, is fruitless.” And fund managers themselves can't consistently beat the averages. Due to costs, most managers underperform and there is no correlation between performances from one year to the next. So, managers can't pick stocks and it is pointless to try to pick good managers.

It is also likely that much of the alpha generated by fundamental analysis is smart beta, compensation for exposure to a certain risk factor. There is absolutely nothing wrong with this. Trading profits are profits, no matter whether they are due to smart beta or alpha. But before we ascribe a trader's results to skill, we should know what is causing the profits. Beta should cost a lot less than alpha.

Conclusion

It is difficult to make money in financial markets. The EMH isn't completely true, but it is closer to being correct than to being wrong. If a trader can't accept this, she will see edges in noise and consequently overtrade. Behavioral finance, technical analysis, and fundamental analysis can all be used as high-level organizing principles for finding profitable trades, but each of these needs to be believed only tentatively, and the most robust approach is to look for phenomena that are independently clear. For example, momentum can be discovered through technical analysis but also understood as a behavioral anomaly. The observable phenomenon must come before any particular method.

Summary

- Exceptions to the EMH exist but they are rare.
- Exceptions will either be inefficiencies, temporary phenomena that last only until enough people notice them, or poorly priced risk premia.
- Risk premia will persist and can form the core of a trader's operations but the profits due to inefficiencies will decay quickly and need to be aggressively exploited as soon as they are found.
- A promising trading strategy is one whose basis is independent of the specific methods used to measure it. Start with observation, then move to quantification and justification.

CHAPTER 3
Forecasting Volatility

All successful trading involves making a forecast. Some traders (for example, trend followers) say they don't forecast, they react. I don't know why they say this, but in any case, they are wrong. The moment a trader enters an order, she has implicitly made a forecast. Why would you get long if you didn't think the market was going up? No matter how it was arrived at, the forecast is, “the market is going up.” Except for a pure arb (which are practically extinct), to get positive expectation we need to make a forecast that is both correct and more correct that the consensus.

In this chapter we will concentrate on making correct forecasts of volatility. But, first, here are some principles that are applicable to any financial forecasting:

- Pick a good problem. Some things are impossible to forecast. No one can predict the price of AAPL in 25 years. Some things are hard to predict. Forecasting the S&P 500 index in two days is a hard problem. Some things are trivial to predict. The FED funds rate in the next day is almost certainly going to be unchanged. Aim to find problems that are solvable but are hard enough that you will be able to profit from the predictions. Volatility is a perfect candidate for this.
- Actively look for comparable historical situations. What happens when the government shuts down? What is the link between recessions and the stock market? This is a good general principle, but it is also vital if you are looking for catalysts that could lead to volatility explosions. Good periods to be short volatility can often be deduced from financial data alone, but long trades generally need a catalyst (that isn't priced in) to be successful. Don't trust your intuition or what you think is true. These will be biased by your experiences, environment, and political persuasion. If you don't have data, you don't have knowledge.

> “When my information changes, I alter my conclusions. What do you do, sir?”
>
> **—J. M. Keynes**

- Aim to balance being conservative and reactive. All good investors have a Bayesian model in their head and update their forecasts as new information arrives but you also shouldn't update too aggressively.
- Actively look for counterarguments. Every person has biases. If you are convinced that every article you read is a harbinger of chaos, be open to the possibility that you are wrong. And the same holds if you are a habitual volatility seller.

> "It is impossible to lay down binding rules, because two cases will never be exactly the same."
>
> **—Field Marshall Helmuth von Moltke**

Remember that there are no certainties when predicting the future.

Model-Driven Forecasting and Situational Forecasting

An effective way to learn is to organize our current knowledge. Sometimes this makes it clear that we don't understand certain things, but even if no such gaps become apparent the thought that goes into a classification scheme is helpful. Science often starts by classifying knowledge. We knew about species before we knew about speciation through natural selection. We knew that elements could be grouped into the periodic table before we knew about atomic structure. We knew about dominant and recessive genes before we knew about DNA structure.

We have already classified trading opportunities into inefficiencies and risk premia. This distinction is important on a strategic level. Mispriced risk premia can last for a long time. A business can be built on harvesting risk premia. Inefficiencies aren't likely to be as persistent. These need to be aggressively traded while they last, and we can assume that they won't last long.

There is also an important classification of trades at the tactical level (strategy defining the high-level goals and tactics being the methods we use to reach them). Trades are either model driven or event driven.

With a model-driven trade, we have a theoretical model of a situation that lets us calculate a fair value or edge. At any moment, we will have an opinion based on the model. For example, if we have an option pricing model, we can continually generate a theoretical value for all of the options on a given underlying. An event-driven trade is based on a specific unusual situation. Noticing that implied volatility declines after a company releases earnings would be the basis for an event-driven trade.

All types of trading, investing, or gambling can be classified like this. In blackjack, card counting is model driven. The player's counting scheme assigns a value for each card that is dealt. As cards are played, the player updates the count, and modifies her edge estimate accordingly. At any point in the dealing, she will know what her edge is. But there is an event-driven method as well: ace tracking. Ace tracking is based on the fact that shuffles aren't perfect randomizers. Cards that are close together in one shoe will tend to stay close together in the next shoe. So an ace tracker notes the cards that are located close to aces and when those same cards are dealt in the next shoe she knows that there is a heightened chance of an ace being close. Because the player advantage in blackjack is due to the 2–1 blackjack payout, having a better than random idea where aces are is enough to give a significant edge. Ace tracking can be more effective than card counting.

Stock investing can be similarly classified. We can rank stocks with a factor model such as Fama-French-Carhart, or we could buy stocks that have had positive earnings surprises. In horse racing, an example of a model would be Beyer's speed figures, whereas an event-driven strategy would be to back horses in certain traps.

Both of these approaches have strengths and weaknesses.

For a model-driven approach to be effective we need a good model. Some situations lend themselves to this more than others. For example, there are very good option valuation models, but stock valuation methods are crude. Sometimes, the work required to build a model is not worth it. But if we have a model, we will always be able to trade. We will have a theoretical value for every trade opportunity. This means the approach scales very well and has great breadth of applicability. We will also be able to scale our trades based on our perceived edge.

The largest problem with this approach is that models have to be vastly simplified views of reality. Often a model's apparent effectiveness isn't due to its effectiveness but more because data gathering and processing is being rewarded. In the 1980s collecting daily closing prices and calculating volatility from those was enough to give a volatility edge in the options markets. Now that this data is free and easy to automatically process, it seems like there is no edge left in this volatility arbitrage model. But there never was any edge in the model at all: the edge was in data collection and processing.

Event-driven trades have two great strengths. The process for finding and testing them is very simple. What happens to stocks in the three days after a FED meeting? What does the VIX do on Mondays? Are teeny options overpriced? All we need to test these ideas is data and a spreadsheet.

Most important, trades that are based on specific events or situations can be very profitable. I have one trade that has literally never lost money. It only sets up a few times a year and is quite liquidity constrained, but it has an unblemished record. This profitability is probably linked to the fact that there is huge uncertainty about why this situation is lucrative.

A drawback of event-based trades is that we must wait for the event, and some events don't occur very often. It is hard to structure a business based on a strategy that might not trade for years at a time. Also, it is often hard to know why the trade exists. This isn't always true. For example, ace tracking is profitable for a very clear reason. But sometimes even a trade with compelling statistics has no obvious reason. I don't do trades if I have no idea why they exist, but sometimes it is easy to come up with a post-hoc reason. For example, many sports fans have convinced themselves that home field advantage is due to travel fatigue. This seems plausible, but it is wrong. Even when teams share the same ground the home team has an advantage. There is no magic answer to this dilemma. The weaker the evidence for a cause is, the stronger the statistical evidence needs to be.

Related to this problem is that if we have only a vague idea for why a trade works, we will have a hard time knowing if it has stopped working or if we are just experiencing a bad period. This is particularly true if the proposed reason is a psychological one. It is always tempting to ascribe any anomaly to psychology. This inevitably leads to overconfidence in the trade. After all, human psychology isn't going to change so why would these trades ever stop working?

Finally, we will often have no way to differentiate between "good" and "bad" trades of the same class. If all we know is that selling options over earnings is profitable, we won't know if it is better to sell Apple options or IBM options. This makes sizing difficult. We only have statistics for the entire class of trades. We need to be very conservative.

Models can give a false sense of security. No model can account for everything. No event has a single cause. Most events have many causes. A situational strategy directly acknowledges this uncertainty and generally traders who are comfortable with uncertainty will do best. So, although creating models is not a bad idea, you also need to become comfortable trading with the ambiguity inherent with specific events.

Our focus in this book is finding *situations* where we can do better than the consensus. This is covered in detail in Chapter Five. The ease of finding and manipulating financial data has considerably lessened the efficacy of forecasting volatility using time series models (the primary forecasting tool used in *Volatility Trading* [Sinclair, 2013]), but measuring and forecasting volatility in this way is still necessary for trade sizing and allocation. Econometricians are still writing endless papers about different members of the GARCH family, but there have been no fundamentally different advances in volatility measurement and forecasting in the last 20 years.

For a summary of these time-series methods refer to Sinclair (2013) and for more detail refer to Poon (2005). Here, I have two general observations.

The GARCH Family and Trading

The simplest forecasting model is to assume that the volatility over the next N days will be the same as it was over the previous N days. Mathematically,

$$\sigma_t^2 = \sigma_{t-1}^2 \tag{3.1}$$

This has two major problems. First is the "windowing" effect where a large single return affects the volatility calculation for N

days, then drops out of the sample. This creates jumps in the volatility measurements and hence the forecast. An example is given in Figure 3.1, where we calculate the 30-day volatility of Maximus, Inc. (MMS) from June, 15, 2019, to September, 30, 2019.

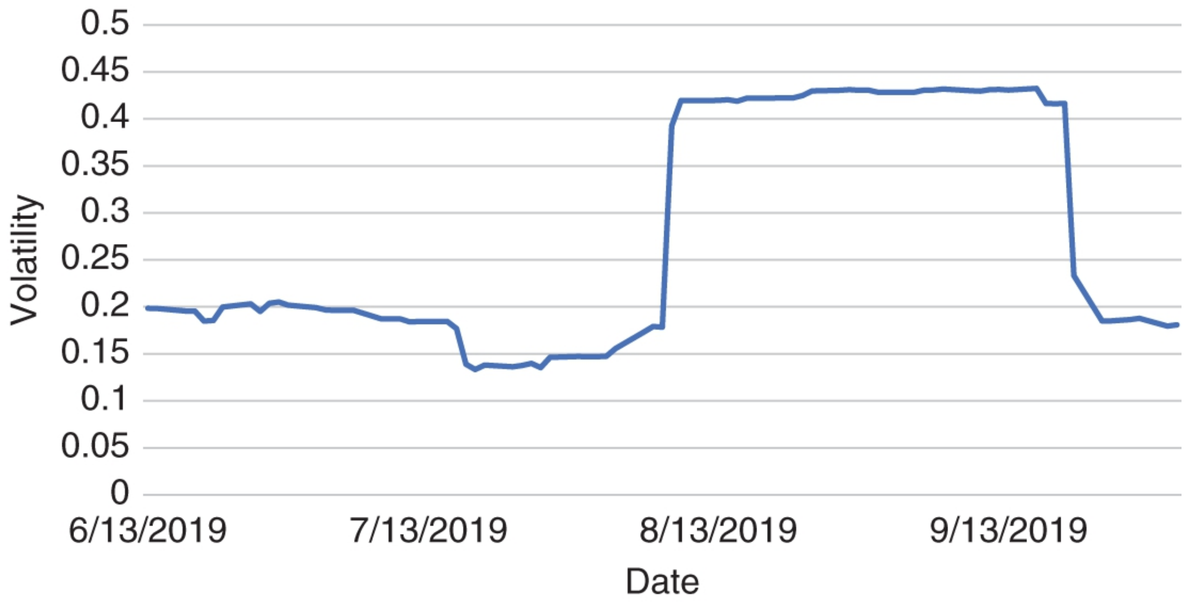


FIGURE 3.1 The rolling 30-day close-to-close volatility of Maximus, Inc.

The typical daily move of this stock was about 0.7% but on August 8 it jumped by 12% because of earnings. This caused the 30-day volatility to jump from 17.8% to 39.3%. Thirty days later, the earnings day dropped out of the calculation and volatility again dropped to 23.3%. If we can know what events are outliers, we can avoid this problem by removing them from the data. We can just throw out the earnings day return.

A bigger problem is that this method doesn't take volatility clustering into account. Periods of exceptionally high or low volatility will persist for only a short time. The exponentially weighted moving average (EWMA) model takes this into account. This says variance evolves as

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) r_{t-1}^2 \tag{3.2}$$

where λ is usually chosen to be between 0.9 and 1.

The GARCH (generalized autoregressive conditional heteroskedasticity) family of models extend this idea to allow for mean reversion to the long-term variance. The GARCH(1,1) model (so-called because it contains only first-order lagged terms) is

$$\sigma_t^2 = \alpha\sigma_{t-1}^2 + \beta r_{t-1}^2 + \gamma V \tag{3.3}$$

where α, β, and γ sum to 1 and γV is the long-term variance.

GARCH is both an insightful improvement on naively assuming future volatility will be like the past and also wildly overrated as a predictor. GARCH models capture the essential characteristics of volatility: volatility tomorrow will probably be close to what it is today and volatility in the long term will probably be whatever the historical long-term average has been. Everything in between is interpolation, and it is in the interpolation that the models in the family differ. As an example, Figure 3.2 shows the term structure of forecast volatility for SPY on August 1, 2019, using GARCH(1,1) and GJR-GARCH(1,1), which also accounts for the asymmetry of positive and negative returns. Both models are estimated from the previous four years of daily returns using MLE.

From a practical perspective, the difference is negligible. And this is what has led to the proliferation of GARCH-type models. They are all roughly the same. No model is clearly better than the others. In any situation where there are many competing theories it is a sign that all of the theories are bad. There is one Schrödinger equation. It works very well. There are thousands of GARCH variants. None work very well.

In fact, it has been shown that the forecasts from GARCH generally are no better than a simple EWMA model, and most professional traders are reluctant to use GARCH. Part of the reticence is due to the instability of the MLE parameters. These can change considerably week to week. MLE also requires about a thousand data points to get a good estimate. This means that if we are using daily data, our forecast will be using information from four years ago. This isn't good.

But there is a practical way to combine the robustness of EWMA and the decay to a long-term mean that GARCH allows. When a trader uses EWMA, he arbitrarily chooses the decay parameter instead of fitting to historical data and using MLE. We can do the same with GARCH. Choose a model, choose the parameters, and use it consistently. This means that eventually we will develop intuition by "seeing" the market through the lens of this model.

For indices, choosing α in the range of 0.9 and β between 0.02 and 0.04 seems to work.

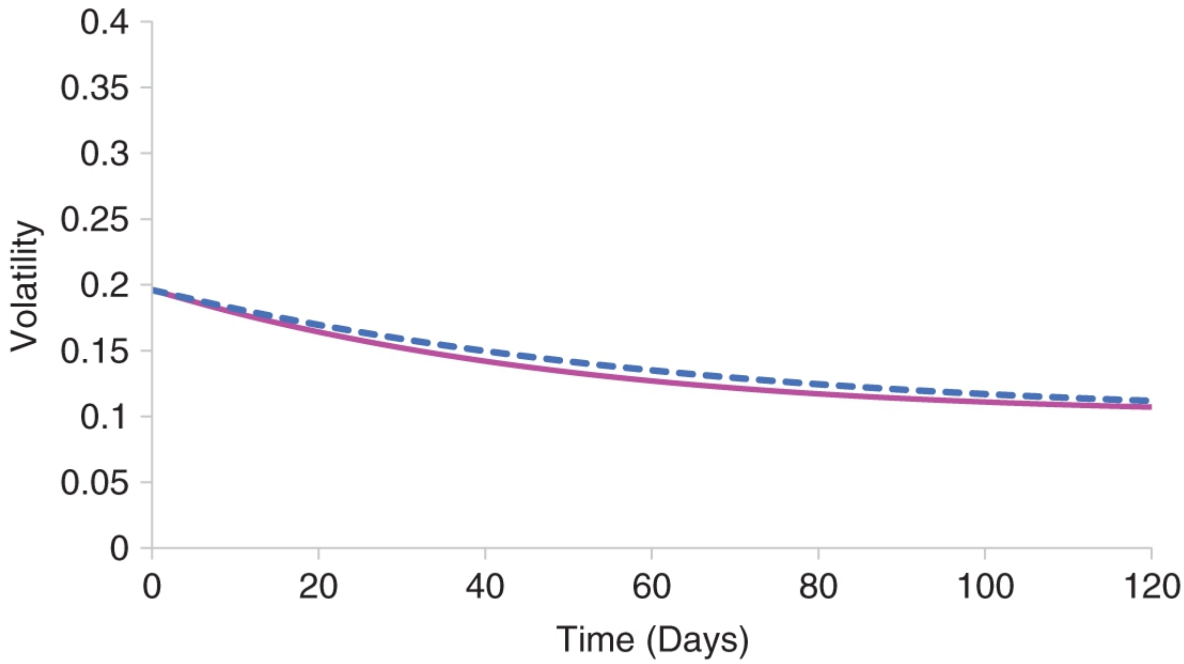


FIGURE 3.2 Term structure of forecast volatility for SPY using GARCH(1,1) (solid line) and GJR-GARCH (dashed line).

Implied Volatility as a Predictor

Implied volatility can be used to predict future realized volatility if we account for the variance premium. So a forecast of the 30-day volatility for the S&P 500 would be given by subtracting the appropriate variance premium for the current VIX level (refer to Table 4.3) from the VIX.

Most underlying products do not have a calculated VIX index. The first way to deal with this is to follow the CBOE's published methodology and construct a VIX. An easier way is to create a weighted average of the appropriate ATM volatilities and use that as a proxy. This methodology was used to create the original VIX (ticker symbol VXO). VXO and the VIX returns have an 88% correlation and the average difference between their values is about 0.5% of the VIX level. This approximation isn't ideal but will usually be the best there is.

Ensemble Predictions

The volatility market is now mature enough that any time series–based volatility method will probably not provide forecasts that are good enough to profit in the option market. A better approach is to combine a number of different forecasts. This idea of the usefulness of information aggregation is far from new. One of the earliest advocates for the “wisdom of crowds” was Sir Francis Galton. In 1878, he photographically combined many different portraits to show that “all the portraits are better looking than their components because the averaged portrait of many persons is free from the irregularities that variously blemish the look of each of them.” His experiment has been repeated and his conclusions validated numerous times using more advanced equipment.

An aggregate forecast can be better than any of the components that make it up. This can be demonstrated with a simple example. Imagine that we ask 100 people the multiple-choice question, “What is the capital of Italy?” with the possibilities being Rome, Milan, Turin, and Venice. Twenty of the group are sure about the correct answer (Rome). The remaining 80 just guess so their choices are equally divided among all the choices, which get 20 votes each. So, Rome receives 40 votes (the 20 people who knew and 20 votes from guesses) and the other cities get 20 votes each. Even though only a small proportion of the people had genuine knowledge, the signal was enough to easily swamp the noise from the guesses of the guessers.

This example also shows that for forecast combinations to be most useful they need to contain diverse information. We need the people who are wrong to be uncorrelated sources of noise. That isn't the case with volatility time series models. Most models will have very high correlation with each other. However, simply averaging the predictions from a number of simple models will still improve predictions. I used five volatility models to predict subsequent 30-day S&P 500 volatility from 1990 to the end of 2018. Table 3.1 shows the summary statistics for each model and also for a simple average.

The error of the average is only beaten by that of the simple 30-day average (the least sophisticated model) but it beats it when we consider the dispersion of results. Interestingly averaging the 0.9 and 0.95 EWMA models also leads to a slight improvement. This is shown in Table 3.2.

Even very similar models can be usefully averaged. This is probably the best way to apply this concept. Average over every GARCH model possible, a wide range of time scales, and a wide range of parameters. Ideally, the models that are averaged would be based on totally different ideas or data, but with volatility this won't happen.

TABLE 3.1 Thirty-Day Volatility Forecasts for the S&P 500 from 1990 to the End of 2018

	Average	30-Day Historical Volatility	EWMA (λ = 0.9)	EWMA (λ = 0.95)	VIX	GARCH (1,1)
Average Error (volatility points)	0.27	−0.10	−0.92	−0.76	0.30	0.44
SD of Error	5.2	6.0	5.8	5.9	5.2	5.9
10th Percentile	−5.1	−5.8	−7.3	6.6	−6.4	−6.3
90th Percentile	5.2	5.8	4.1	5.0	8.9	5.3
R-Squared	0.65	0.62	0.62	0.60	0.64	0.60

TABLE 3.2 Thirty-Day Volatility EWMA Forecasts for the S&P 500 from 1990 to the End of 2018

	Average	EWMA (λ = 0.9)	EWMA (λ = 0.95)
Average Error (volatility points)	−1.1	−1.5	−0.76
SD of Error	5.7	5.8	5.9
10th Percentile	−6.9	−7.3	6.6
90th Percentile	4.5	4.1	5.0
R-Squared	0.61	0.62	0.60

Conclusion

Realized volatility is reasonably forecastable for a financial time series. Unfortunately, this means that it is hard to make a good forecast that differs significantly from the market's consensus. However, volatility predictions are essential even when they are not the basis for finding edge. In particular, any sensible sizing scheme will need a prediction of future volatility.

Summary

- All trading strategies can be categorized as either model driven or based on special situations. Each type has weaknesses and strengths.
- An ensemble prediction of volatility will usually outperform time series methods.

CHAPTER 4
The Variance Premium

> *In finance, everything that is agreeable is unsound and everything that is sound is disagreeable.*
>
> —Winston Churchill

The variance premium (also known as the volatility premium) is the tendency for implied volatility to be higher than subsequently realized volatility.

This is not a recent phenomenon. In his 1906 book *The Put and Call,* Leonard Higgins writes how traders on the London Stock Exchange first determine a statistical fair value for options, then "add to the 'average value' of the put and call an amount which will give a fair margin of profit." That is, a variance premium was added.

The variance premium exists in equity indices, the VIX, bonds, commodities, currencies, and many stocks. It is probably the most important factor to be aware of when trading options. Even traders who are not trying to directly monetize the effect need to know of it and understand it. It is the tide that long option positions need to overcome to be profitable. Even traders who only use options to trade directionally need to take this into account. Even if directional predictions are correct, it is very hard to make money if one is consistently paying too much for options (see Chapters Six and Seven for more discussion of this point).

This effect can be monetized in many ways. The size and persistence of the variance premium is so strong that the precise details of a strategy often aren't very important. Practically any strategy that sells implied volatility has a significant head start on being profitable if the premium is there.

In this chapter we will discuss the characteristics of the variance premium in various products; look at the relationships among the variance premium, correlation, and skewness; and give some possible reasons for the existence of the effect.

Aside: The Implied Variance Premium

The *variance premium* refers to the difference between implied volatility, which can be defined by either BSM implied volatilities or variance swaps, and subsequent realized volatility. There is a related phenomenon that occurs entirely in the implied space. Being short VIX futures is generally a profitable strategy (although not a wildly successful one). Figure 4.1 shows the results of always being short the VIX front month future from June 2015 to October 2019.

The VIX itself doesn't decay in the same way (refer to Figure 4.2). This is really a term-structure effect in the futures.

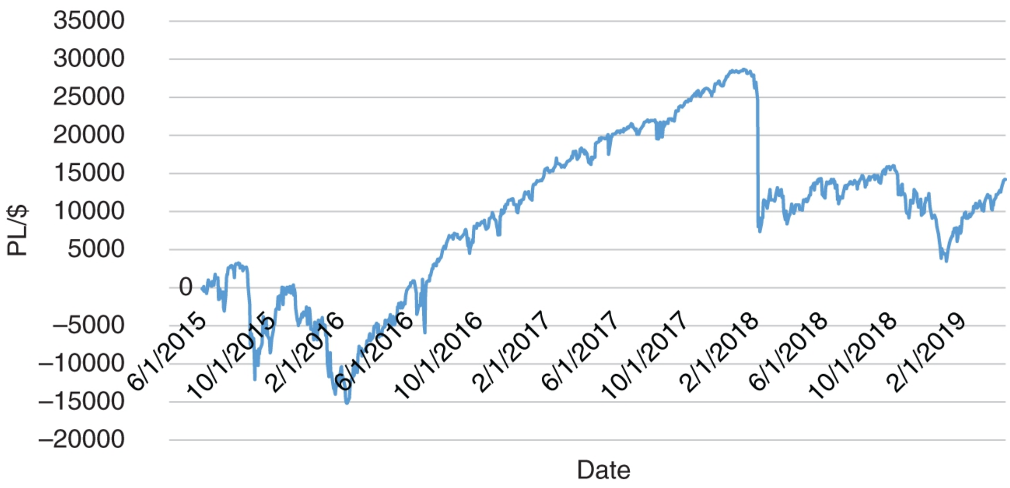


FIGURE 4.1 Profit from selling 1 front-month VIX future.

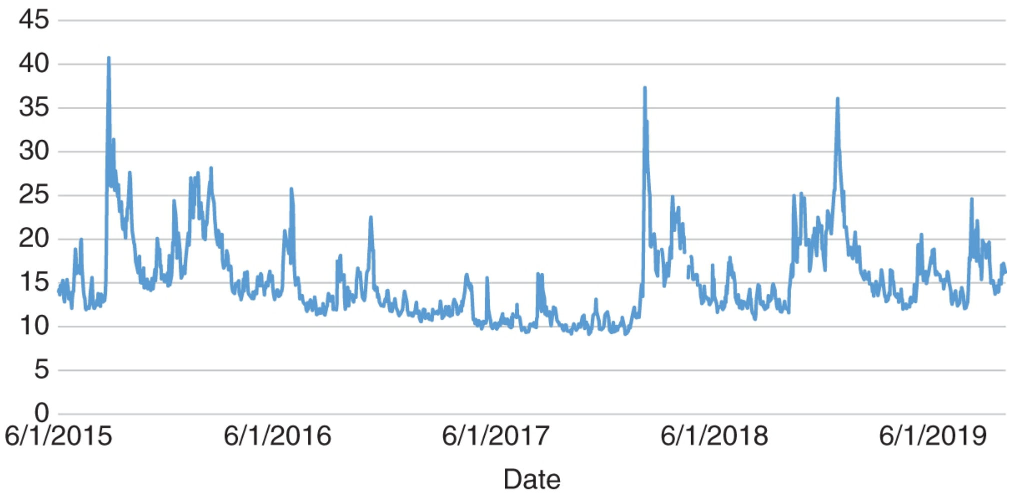


FIGURE 4.2 The VIX index from June 2015 to October 2019.

According to the rational expectations hypothesis, the VIX futures curve should be an unbiased predictor of where the VIX index will be on the expiration date. The narrowing of the basis as time approaches the expiration date should be more dependent on the cash index moving toward the future's price.

The theory of rational expectations has been tested on many different commodity futures and it is generally a poor description of price movements. Futures tend to move toward the cash. Alternatively, the cash VIX is a better predictor of future VIX levels than the futures are.

It is probably not surprising that this also occurs in the VIX. VIX futures are unusual. Generally, futures are priced by first assuming that they are forwards, then constructing an arbitrage-free portfolio of the underlying and the future. However, the VIX index cannot be traded so this method is not useful for pricing VIX futures. Given that VIX futures are not constrained by tight, no-arbitrage bounds, there is even more room for inefficiencies.

On its own this doesn't mean short positions have to be profitable. But the VIX term structure is usually in contango (from the time VIX futures were listed in 2006 to the start of 2019, the term structure has been in contango 81% of the time). This means that the futures are above the cash and tend to decline toward it. The best discussion of this effect is in Simon and Campasano (2014). Selling a future only when the previous day's prices were in contango considerably improves this strategy. Figure 4.3 shows the results of being short the VIX front-month future from June 2015 to October 2019, when the term structure is in contango.

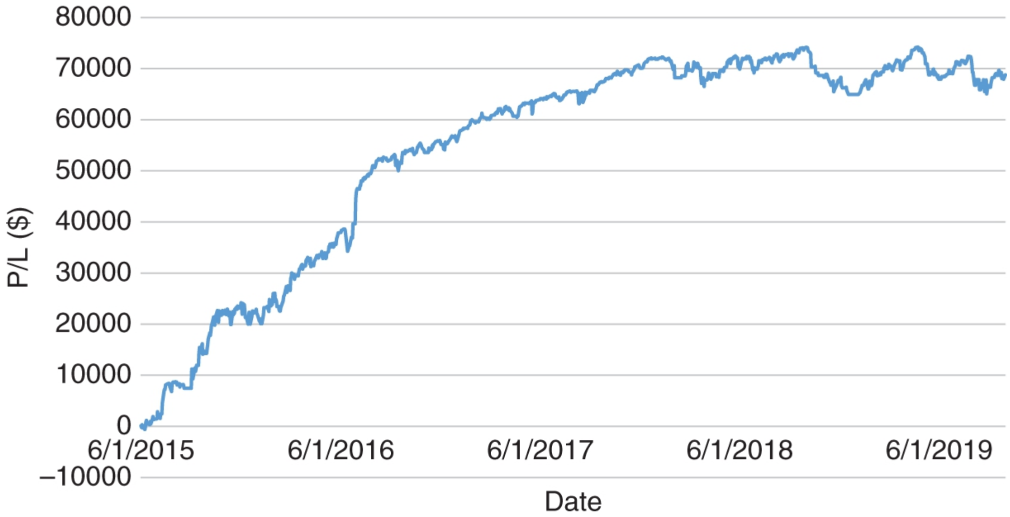


FIGURE 4.3 Profit from selling 1 front-month VIX future when the term structure is in contango.

Variance Premium in Equity Indices

Figure 4.4 shows the VIX and the subsequent 30-day realized volatility of the S&P 500 from 1990 to the end of 2018.

On average the VIX was four volatility points higher than the realized volatility and the premium is positive 85% of the time. Figure 4.5 shows the premium in volatility points. Figure 4.6 shows the distribution of the daily premia and Table 4.1 gives the summary statistics.

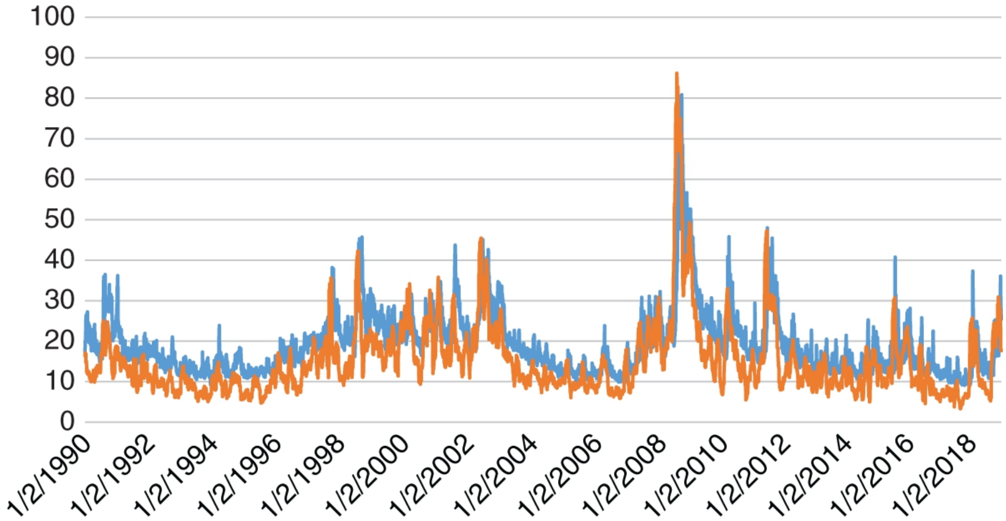


FIGURE 4.4 The VIX and the subsequent 30-day realized S&P 500 volatility.

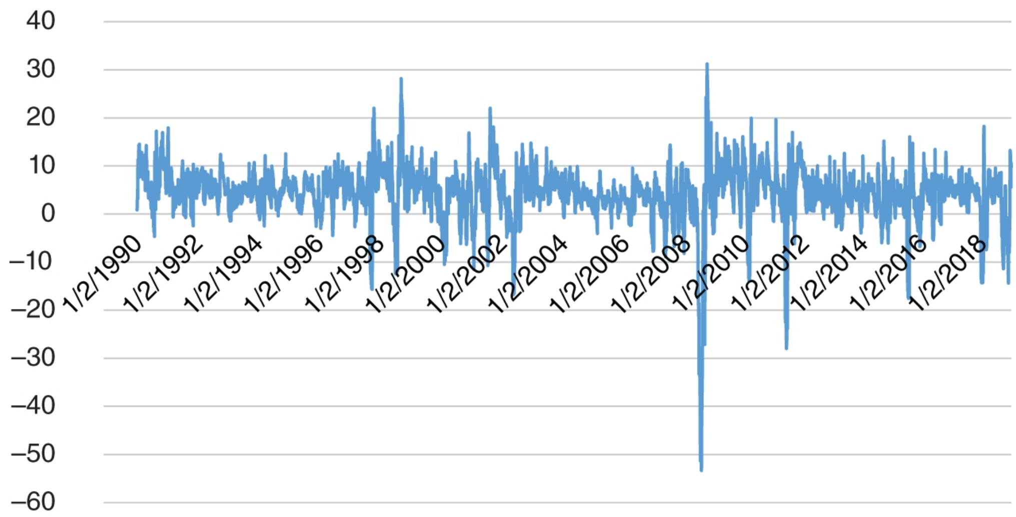


FIGURE 4.5 The S&P 500 variance premium (VIX minus realized volatility).

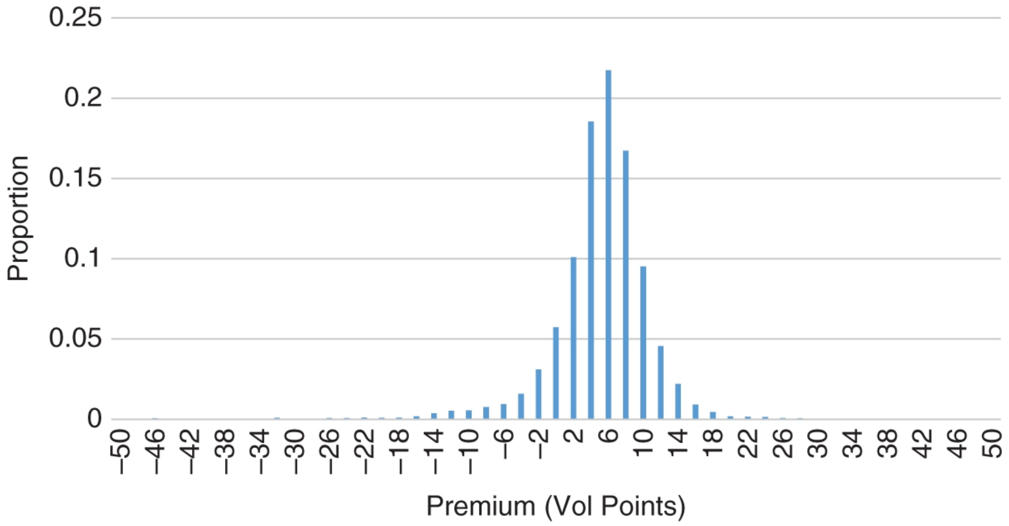


FIGURE 4.6 The S&P 500 variance premium distribution.

TABLE 4.1 Summary Statistics for the S&P 500 Variance Premium

Mean	4.08
Standard deviation	5.96
Skewness	−2.33
Maximum	31.21
Minimum	−53.34
Median	4.63
90th percentile	9.62
10th percentile	−1.45

The Dow Jones 30, NASDAQ 100, and Russell 2000 indices have similar variance premia. The summary statistics for these are given in Table 4.2.

TABLE 4.2 **Summary Statistics for the Dow Jones, NASDAQ 100, and Russell 2000 Variance Premia**

Index	Dow Jones (from 1998)	NASDAQ 100 (from 2001)	Russell 2000 (from 2004)
Mean	3.50	3.41	3.24
Standard deviation	6.18	6.99	6.58
Skewness	−2.03	−2.02	−2.72
Maximum	28.90	26.48	27.28
Minimum	−49.42	−53.79	−46.08
Median	4.22	3.91	4.05
90th percentile	9.46	10.37	9.10
10th percentile	−2.60	−3.11	−2.81

The variance premium varies considerably with the market volatility. Table 4.3 shows the variance premium statistics for the five VIX quintiles. Here, as in Figure 4.6, the premium is expressed in volatility points.

So, broadly speaking, the typical size of the variance premium (measured by either mean or median) increases with volatility levels. The variability of the variance premium also increases with volatility. When trading the variance premiums with options, a mid-level VIX is probably the best environment. Low-volatility regimes give lower returns as a percentage of margin and require more frequent hedging due to the higher gamma. High-volatility regimes have good average returns but a bad risk profile.

TABLE 4.3 **Summary Statistics for the VIX Sorted by Quintiles**

	VIX <13.02	13.02 <VIX<15.89	15.89 <VIX<19.42	19.42 <VIX<24.15	VIX >24.15
Mean	2.61	3.37	4.35	4.19	5.87
Standard deviation	3.54	3.60	4.58	6.53	9.06
Skewness	−2.36	−1.67	−2.23	−2.21	−2.46
Maximum	8.33	10.17	13.84	14.82	31.21
Minimum	−17.54	−16.84	−27.95	−37.66	−53.34
Median	3.25	4.05	5.25	5.44	7.23
90th percentile	5.16	7.21	8.73	10.45	13.81
10th percentile	−0.81	−0.80	−0.74	−3.15	−2.80

There is also more direct evidence of the premium. Short positions in delta-neutral equity index option positions (straddles, strangles, butterflies, and condors) have been profitable, as have short positions in variance swaps. Sharpe ratios vary between 0.4 and 1.0 depending on details of the implementation.

The effect has been studied in many countries and time periods (see, for example, Driessen and Maenhout, 2006; Londono, 2011).

The CBOE publishes two separate short option volatility indices: CNDR and BFLY. CNDR tracks the performance of a strategy that sells a 1-month SPX iron condor (short the 20-delta strangle and long the 5-delta strangle). BFLY tracks a short iron butterfly. The hypothetical performance of these strategies is shown in Figures 4.7 and 4.8.

It is worth discussing why neither of these indices has been particularly successful since 2008.

The variance premium was virtually the same before the end of 2008 (median of 4.63 points) and since (median of 4.61). However, the median VIX level dropped from 18.3 to 16.3, and after the chaos of 2009, the median VIX dropped further to only 15.4. The lower volatility level gives options higher gamma, which drastically increases the effects of path dependency and drift.

The second reason is the effect of skew. Although lower ATM volatility usually means all other options will also trade at lower volatilities, the relative implied volatility of teeny puts tends to increase. So as volatility decreased, the prices of the 5-delta puts increased relative to the premium from selling the options closer to the ATM. This effect is illustrated by the CBOE SKEW index having an average value of 116.2 in the earlier period and 125.7 afterwards. These strategies were paying more for their put hedges. And these teeny puts are always the most options with the highest variance premium. Unless buying them for a hedge they are the ultimate sucker bets. Refer to Hodges et al. (2003) for a study of this fact.

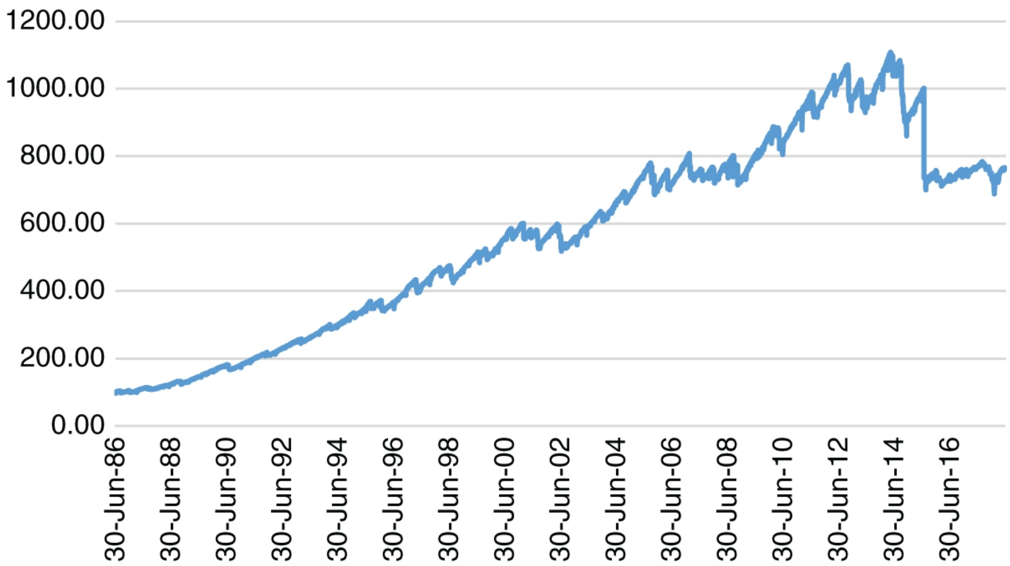


FIGURE 4.7 Performance of the CNDR index.

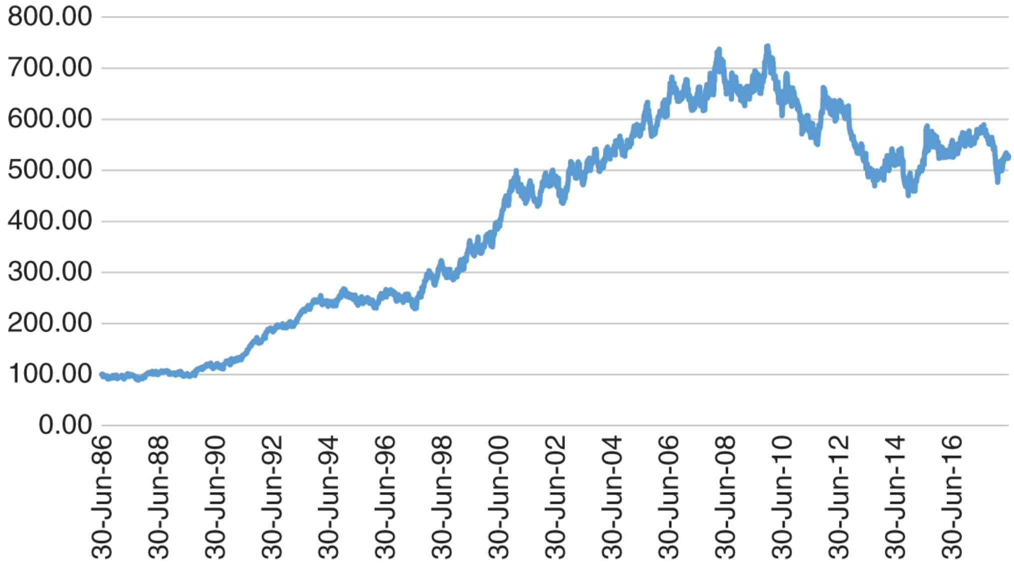


FIGURE 4.8 Performance of the BFLY index.

So, although the variance premium is a strong and consistent phenomenon, attention needs to be paid to the capture strategy. This will be discussed further in Chapter Six.

The Implied Skewness Premium

A large part of the index variance premium is due to the OTM (out-of-the-money) puts being overpriced. Intuitively this should be the case, because much of the value of a variance swap is driven by the value of the OTM puts. And it is well known that index put options are overpriced (see, for example, Bondarenko, 2003).

The implied volatility curve predicts very high negative skewness in realized returns. Although many pricing models can reproduce such a curve, their parameters are not consistent with the absence of substantial negative skewness in stock index returns. To misquote Samuelson, option markets predict nine out of the past five market corrections. Also, the predictions are reactive, with the implied skew being steepest after crashes.

Kozhan et al. (2011) examined the profitability of selling skew swaps. The skew swap is a model-free skewness contract whose payoff is equal to the difference between realized skew and implied skew. As with variance swaps, the skew swap can be replicated from a portfolio of vanilla options. Using this contract enables the researchers to investigate the risk premium implicit in implied skew without having to worry about the misspecification of any specific option pricing model.

They show that for S&P 500 options (from 1996 to 2009) about half of the excess return from selling out-of-the-money puts is due to the correlation between returns and volatility: the realized skewness.

The Implied Correlation Premium

Selling index volatility and buying component volatility is a short correlation position. All the index components dropping is the same as correlations rising.

Short correlation is also short implied volatility.

This strategy has Sharpe ratios comparable to selling index variance swaps. Although the returns are lower, the variance is also considerably reduced (Driessen et al., 2009). Dispersion trading suffers during market turmoil when the correlations increase, that is, the same periods when selling index variance suffers.

Commodities

The variance premium probably exists in commodities. Prokopczuk and Simen (2014) used option prices to construct synthetic variance swaps and found significantly negative variance risk premia in nearly all commodity markets. They examined 21 commodity markets between 1989 and 2011 and found that 18 of the commodities had statistically significant returns to short variance swaps. Table 4.4 summarizes the results for the 60-day swaps (a premium of 10% would mean the 60-day implied variance was 10% above the 60-day realized variance).

The correlation of variance premia for commodities in the same sector was positive but small. And the correlation across sectors was also small enough to make harvesting the premium in commodities a good diversifier. The correlations are shown in Tables 4.5 and 4.6.

TABLE 4.4 The Size and Significance of the Variance Premium in Commodity Options

Commodity	Variance Premium (%)	*T*-Score
Crude oil	3.4	6
Heating oil	3	7
Natural gas	10.2	9
Corn	2.3	8
Cotton	−0.6	−11
Beans	0.8	2
Bean meal	0	0
Bean oil	1	4
Sugar	2.6	6
Wheat	0.7	3
Hogs	1.2	3
Cattle	1	11
Copper	2.4	4
Gold	1	4
Silver	0.2	1
Cocoa	3	8
Coffee	1.7	1
Oats	6.2	8
OJ	2.3	3

Commodity	Variance Premium (%)	*T*-Score
Rice	3	8
Lumber	3.5	10

Trading options on commodities requires some fundamental knowledge. An equity option trader can trade any equity, often without even knowing more than the ticker symbol. This is largely because the movement of equities is almost random and the effect of any real fundamental knowledge is small (indices are even more ignorance driven: the best situation for a statistically driven trader). But people actually know things about commodities. There are different crops, pipeline bottlenecks, storage squeezes, and weather effects. There is no guarantee that a good wheat trader can become a good corn trader. Different things affect different commodities.

TABLE 4.5 The Correlation of the Variance Premium Within Commodity Sectors

Sector	Correlation of 60-day VP
Energy	33.4%
Grains	24.2%
Livestock	31.4%
Metals	30.1%
Tropicals	5.7%
Wood	N/A

TABLE 4.6 The Correlation of the Variance Premium Between Commodity Sectors

	Energy	Grains	Livestock	Metals	S&P 500	*T*-bonds	Tropicals	Wood
Energy	100%							
Grains	9.0%	100%						
Livestock	13.2%	14.9%	100%					
Metals	22.1%	14.5%	6.5%	100%				
S&P 500	26.5%	2.0%	16.4%	30.7%	100%			

	Energy	Grains	Livestock	Metals	S&P 500	*T*-bonds	Tropicals	Wood
T-bonds	20.2%	16.0%	5.3%	23.1%	39.3%	100%		
Tropicals	8.9%	24.7%	11.3%	11.6%	5.9%	0.3%	100%	
Wood	7.3%	0.9%	7.8%	3.9%	9.5%	6.5%	11.7%	100%

We can see this in the variance premia measured by Prokopczuk and Simen (2104). Traders who have fundamental insight might know why the corn premium is so much greater than that of wheat and why the premium in silver options is so different from that in gold options.

Bonds

Choi et. al (2017) looked at bond options from 1990 to 2012. They found that 1-month variance swaps on US 5-year notes, 10-year bonds, and 30-year bonds are overpriced by about 20% (18.7% for the 5-year, 27.6% for the 10-year, and 21.2% for the 30-year). The Sharpe ratio for a strategy that sells all these swaps is about 2. Selling 1-month ATM straddles was also profitable, although less so.

Most bond uncertainty (and hence most of the variance premium) is clustered around the release of macroeconomic data. Jones et al. (1998) showed that from 1979 to 1995 about 90% of excess returns to Treasury bonds accrued on days with either an employment or PPI announcement. Interestingly, Andersson et al. (2009) showed that German government bonds reacted far more strongly to US data releases than German ones. German unemployment releases had almost no effect at all, leading them to conclude that the number was widely leaked.

The VIX

There is also a variance premium in VIX options. Hogan (2011) used VIX options to construct synthetic VIX variance swaps and showed that these had a premium to the realized variance of the VIX futures that was similar in size to the premium in equity indices. Kaeck (2017) found similar results using data from 2006 to 2014.

Currencies

Lo and Zhang (2005) found direct evidence of a variance premium in OTC currency options. They found that a strategy of selling straddles was profitable for options on USD versus GBP, YEN, CHF, and the euro for terms between 1 month and 1 year. They also found that the variance premium increased as options got closer to expiration.

Londono and Zhou (2017) reported significant positive variance premia for variance swaps on USD/GBP, USD/Yen, and USD/euro at 1-, 3-, and 6-month durations. However, the variance premia for some currencies, notably NZD and

AUD, was large and negative. As the antipodean currencies are generally seen as "safe havens," this suggests some linkage between the priced variance risk and real-world political and macroeconomic risk perceptions.

Equities

The results of selling options on equities are not clear-cut. First, because we know that implied correlation accounts for a portion of the index variance premium, we would expect the returns to selling stock volatility to be lower than selling index volatility. This is true. Also, as with commodities, there are specific fundamental factors that affect the variance premia in stock options. This is examined in more depth in Chapter Five.

TABLE 4.7 The Average Return to a Short 1-Month Variance Swap for Stock Options in Different Industries

Industry	Average Return	*T*-statistic
Utilities	31.25%	10.39
Nondurable consumer goods	17.71%	7.89
Other	9.83%	3.71
Durable consumer goods	8.56%	3.36
Energy	8.56%	3.55
Manufacturing	8.05%	3.29
Retail	4.68%	1.93
Health care	3.27%	1.39
Telecommunications	−1.40%	−0.48
Technology	−7.66%	−2.97

Di Pietro and Vainberg (2006) show how size and book-to-market firm characteristics are linked to the expensiveness of equity options. They find that options on small stocks are more expensive than options on large stocks and that options on value stocks are more expensive than options on growth stocks.

This was partially corroborated by Vilkov (2008). He did not find a robust size effect but found that value stocks had more of a variance premium than growth stocks. He also found that illiquid stock had high variance premia (this won't be practically useful because these stocks will also have illiquid options). He also looked at the variance premia of various industry sectors (as defined by French). These results are shown in Table 4.7.

Anyone thinking of trading a portfolio of equity options should probably study the interrelationships among these various factor exposures. I'm not aware of any published work on this.

Reasons for the Variance Premium

There are many reasons for the existence of the variance premium. This means it is likely to persist, because it is doubtful that all of these reasons would disappear at the same time, even if they all vary across different market and economic regimes.

Insurance

The most compelling reason for a variance premium is that people are willing to pay for insurance. The most obvious insurance buyers are those who are long stocks and buy puts for downside protection. This demand is a major driver of the implied skew and the resulting skew premium. However, there are also investors who will buy calls to insure against missing out on large rallies. Obviously, for every option buyer there is also a seller, but the customer demand sets the price. Hedgers are prepared to pay for insurance. The sellers take on the risk and also demand an extra premium. The variance premium is reflective of a risk premium, but it is also a mispriced risk premium.

Jump Risk

If underlying prices were continuous, options could be perfectly replicated. This would make their existence redundant. However, underlying prices can jump and options give protection against these jumps in a way that a dynamic hedging scheme cannot, making them more attractive to buyers (hedgers or speculators). The non-redundancy of options can be seen as both a cause and consequence of the variance premium.

Trading Restrictions

Many retail traders have restrictions placed on them by their brokers. It is common for them to be forbidden from selling naked options. This means that an entire class of speculators can only be long volatility, thus driving the variance premium higher.

Market-Maker Inventory Risk

Contrary to popular belief, bookies do not try to completely balance their risk. They often speculate on a team to win. Similarly, option market-makers usually have speculative views on the markets and use the income from market-making operations to smooth the losses from these positions. There are far fewer market-makers now than there were 10 years ago. Increasing automation means fewer people are needed. On the floor, a trader could cover two or three stocks but now it is routine for a single trader to trade hundreds of stocks. But the overall profits to the community are still enormous. Spreads have narrowed, but this has been compensated for by increasing volumes. As long as a market-maker can stay in business, she will eventually be successful.

The best trading opportunities for liquidity providers are in times of turmoil. Spreads widen. Volume soars. Customers are not as price sensitive as usual. So, it is imperative that market-makers can trade aggressively in these situations. And this can only happen if they aren't stuck with a loss from the move. Ideally, they will have a nice profit and can trade freely and with no restrictions. This means that they are almost always long teeny options. A standard risk-management heuristic for market-makers is always to be net long options. They are aware that options are overpriced, but they need them as insurance. Not inventory insurance: business insurance.

Path Dependency of Returns

Grosshans and Zeisberger (2018) show that investors also care about exactly how their returns are realized. They performed surveys that asked people to imagine that they had six stocks. Three made 10% and three lost 10%. But each of the three had different paths: up-down, a linear path, or down-up. These are shown stylistically in Figures 4.9 and 4.10.

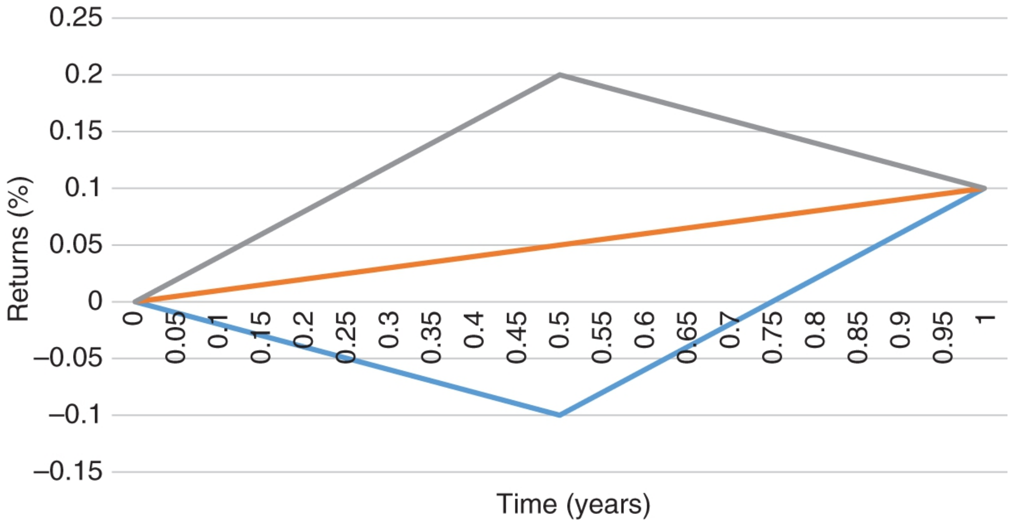


FIGURE 4.9 The three different positive return paths.

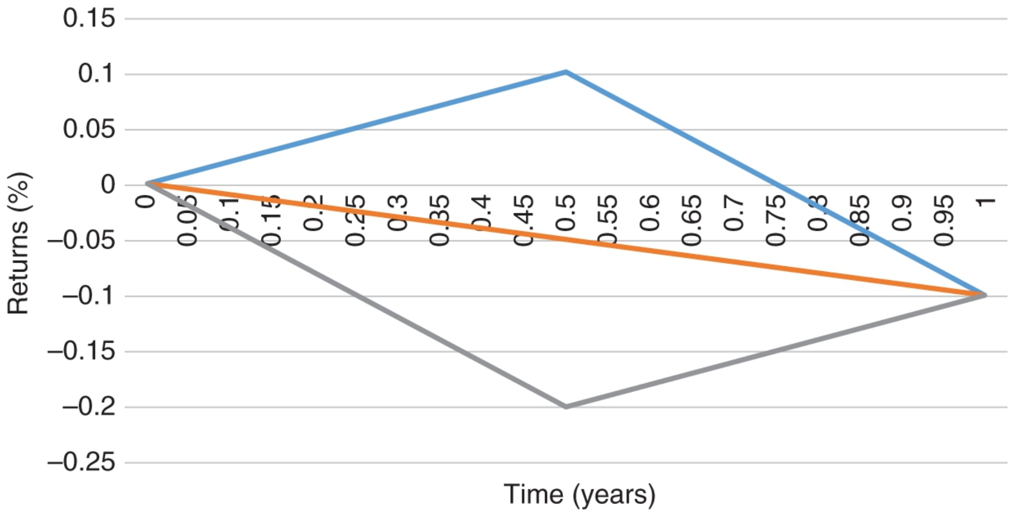


FIGURE 4.10 The three different negative return paths.

For both winners and losers, the participants were happiest when prices first declined and then rose. People were even slightly disappointed when stocks rose and then fell but were still winners! People are happiest when they feel that they have recovered from adversity, snatching victory from the jaws of defeat.

Although this is only survey data it gives another plausible reason for the variance premium.

Think about a $100 stock and the $100 strike call and put. Assuming no interest rates and a volatility of 30%, both the 1-month call and put will each be worth $3.43. One trader sells the put, and another buys the call. Consider what happens if the stock jumps from $100 to $106.86 right before expiration. Each trader makes $3.43 at expiry, but the P/L evolves slightly differently over time for the buyer and seller. These return paths are shown in Figures 4.11 and 4.12.

The largest PL difference is $5.09 on the day before the jump.

The long option maintains the ability to "snatch victory from the jaws of defeat." People prefer this, so options will tend to be overpriced, hence the variance premium.

This is really a gamma effect. The short put steadily collects theta, but the long call wins because of gamma. Redemption comes from the possibility of extreme price moves due to high gamma. This explains why the variance premium is greatest for short-dated options and also why it tends to be higher when volatility is low.

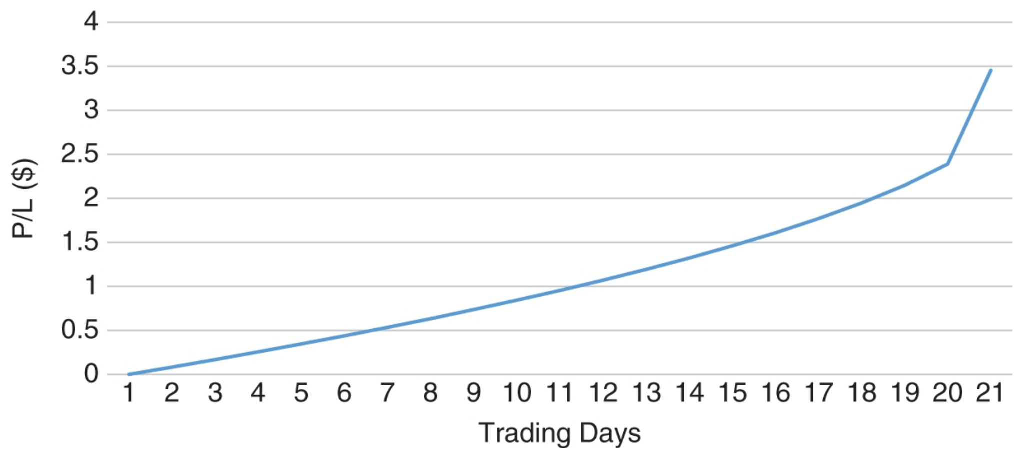


FIGURE 4.11 The P/L for a short put, with a stock jump at expiration.

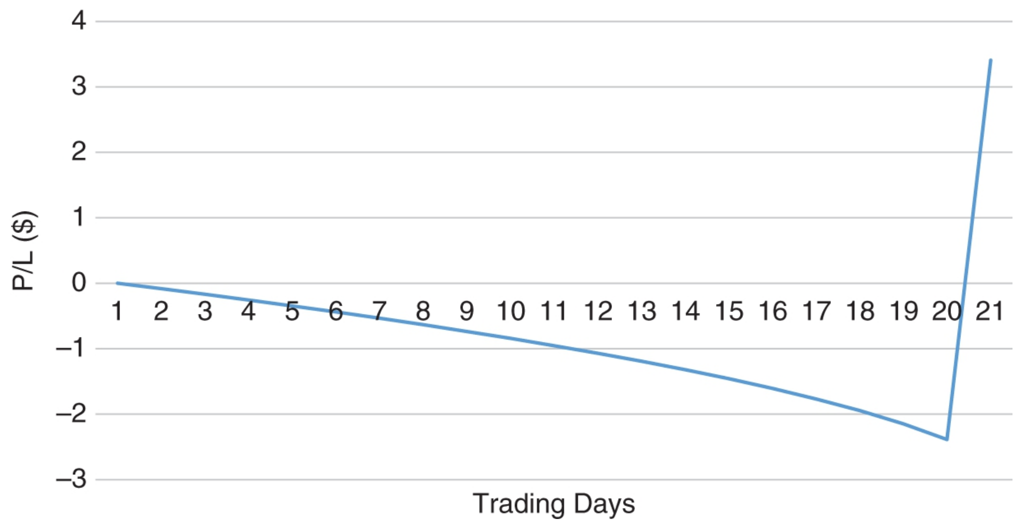


FIGURE 4.12 The P/L for a long call, with a stock jump at expiration,

The Problem of the Peso Problem

It is possible that options are actually not overpriced. Perhaps volatility and the skew are fairly priced, and the apparent variance premium is due to the fact that the events option buyers are insuring against haven't occurred in our sample period but will at some point in the future.

It is hard to see how an argument based on “just you wait” can ever be refuted. We certainly know that in the history we have, implied volatility has been overpriced. It is possible that someday an event will occur that is so large that all option-selling profits will be eradicated. It also seems unlikely. As with all trading decisions, we can either assume history will be an accurate guide to the future or it won't. “This time is different” is an appealing idea because it means we don't have to do any studying of how things have behaved in the past, but it is rarely true.

Conclusion

Implied volatility tends to be an upward-biased estimator of the future realized volatility. This is the most important empirical fact to a volatility trader. The effect applies to most underlying situations and has existed for as long as we have data to look at. There are a large number of economic, distributional, and psychological reasons for the variance premium. Although it could be diminished by the emergence of more institutions trying to capture it, I doubt that it will ever disappear completely.

Summary

- Historically it has been profitable to be short options. There are persuasive arguments that suggest that this will continue.
- Floor trader sayings: “If in doubt, hands out” and “Whenever an option trades good market-makers will know if they want to buy or sell at that price, and if they are unsure they should sell.”

CHAPTER 5
Finding Trades with Positive Expected Value

In *Volatility Trading* I wrote, “There is no attempt here to give a list of trading rules. Sorry, but markets constantly evolve, and rules rapidly become obsolete. What do not become obsolete are general principles. These are what I attempt to provide. This approach isn't as easy to digest as a list of magic rules, but I do not claim markets are easy to beat, either.”

This is still true.

But I am going to give a list of edges. Edges aren't rules. And the difference is important. A rule is a defined way to monetize an edge, whereas the edge is an observed phenomenon that could potentially be profited from in many ways. Edges can persist.

Aside: Crowding

There is a common perception that crowded trades will have diminishing performance. Whether this is always true is unclear, and it was addressed in the context of factor investing by Baltas (2019).

First, it is unclear what exactly is meant by *crowded*. Sometimes it is taken to mean the size of a particular subindustry (e.g., high-frequency trading firms), who are then exposed to the effects of a broad unwinding (such as the “quant meltdown” of August 2007 or the volatility ETN problems of February 2018). It can also relate to market capacity. Generally, the idea encapsulates a number of situations in which the unintentional coordinated actions of traders create feedback effects.

Exactly how these feedback events unfold depends on the dynamics of the specific strategy. Baltas classifies strategies into either convergence or divergence strategies. Convergence strategies, such as value investing, have a natural target price. This creates a stabilizing effect. Crowding will help the realization of profits in convergence strategies.

Divergence strategies, such as trend-following, have no anchoring price. The more these assets go up, the more are bought. This creates a destabilizing feedback effect. In the short term this will help profits, but it will eventually lead to a bubble and a subsequent crash.

Baltas postulated that crowdedness could be quantified by looking for increases in co-movements of assets beyond their beta (an idea first stated by Cahan and Luo, 2013). For example, if all value stocks started more strongly moving together, this would be an indicator of a crowding increase in this factor.

Baltas tested his convergence/divergence idea by using global equities data from September 2004 to May 2018, the constituents of the S&P GSCI Commodity Index from January 1999 to May 2018, and 26 currency pairs from January 2000 to May 2018. For the stocks, he studied the value, size, momentum, quality (return on assets), and low beta factors. For commodities he looked at momentum. And for currencies he examined momentum and value (defined by purchasing power parity). His thesis about how crowding affects various strategies was broadly confirmed. In the presence of a natural target, crowding stabilized markets. If there was no natural target value, crowding was a destabilizing factor.

An important caveat to these results is that the presence of leverage can be overwhelming regardless of the specific trade dynamics. For example, LTCM employed an ostensibly mean-reverting strategy, trading the spread between off-the-run and on-the-run treasuries. They employed huge leverage, and so did their prime brokers who copied the trade. When the spread widened, traders were forced to deleverage. This pushed the spread further against them creating a destabilizing feedback loop. What should have been a stable trade was turned into an unstable one by excess leverage.

An example of products being caught in a destabilizing feedback loop were the VIX ETN's in February 2018.

On Friday, February 2, the S&P 500 dropped 2.1% and the VIX rallied 28.5% from 13.47 to 17.31. This was a large move: at the time it was the 34th largest in history. But Monday, February 5, was truly exceptional. The S&P 500 dropped 4.2%, but the VIX rallied 115.6%. This was the largest VIX move ever and nearly twice the previous record of 64% (and in that case the move was only from 11.15 to 18.31). As the VIX was only implying a daily move of about 1% the S&P 500 return was a genuinely large event. However, the response of the VIX was well out of line with an underling move of that size. A linear regression model linking the S&P 500 returns to the VIX returns predicted that a 4.2% drop in the equity index would correspond to a 12.8% move in the VIX. The actual VIX increase was nine times the expected amount.

On the next day the VIX had its largest range ever, with a low of 22.42 and a high of 50.3 (the 70th highest value recorded).

I've seen nonpublic results for volatility fund returns in these few days ranging from up 25% to down 95%. And some of the losers were large funds. One operation with assets close to a billion dollars lost 89%. The median was a loss of about 30%. Some of these volatility funds lost money by being short index options but the option space losses appear to have been relatively benign compared to those experienced in the VIX ETNs.

The first volatility ETN was VXX, which was listed in 2009. It was designed to match the returns of a (hypothetical) 30-day VIX future. In 2010, XIV was launched with the stated aim of delivering the returns of a short position in the 30-day VIX future. These, and the other similar products that followed, proved immensely popular. First, they gave investors who were unable to trade futures a way to speculate on implied volatility. Second, their returns were relatively predictable. This is due to the contango effect in the VIX futures. To maintain a 30-day notional exposure, the manager of VXX must sell front month futures and buy second month futures. As about 80% of the time the second future trades at a premium to the first, this means that the VXX rebalancing process usually must sell low and buy high. Conversely XIV benefits from this effect. As a result, since its launch until the end of 2017, VXX had decayed from a split-adjusted price of 107090 to 27.92. XIV had increased from 9.56 to 134.44.

Short volatility products had become more popular in 2016 and 2017 because the realized volatility of the S&P 500 index was very low and the contango decay was very high. Open interest increased enormously and resulted in crowding in these products prior to the crash in February 2018. In VXX alone, short interest increased nearly 1300% from the end of 2013 to the end of 2017.

It is obvious that a 100% rally in the VIX 30-day future would drive XIV to zero; however, conditions for termination (named *acceleration* in the prospectus) didn't even require this. From the prospectus,

> "...an Acceleration Event includes any event that adversely affects our ability to hedge or our rights in connection with the ETNs, including, but not limited to, if the Intraday Indicative Value is equal to or less than 20% of the prior day's Closing Indicative Value."

In this eventuality,

> "... you will receive a cash payment in an amount (the 'Accelerated Redemption Amount') equal to the Closing Indicative Value on the Accelerated Valuation Date."

It is important to stress that XIV is an ETN, not an ETF. An ETF owns the stocks, bonds, or commodities that make up the portfolio, whereas the ETN is merely a note that pays the return on the portfolio. Whether and how the issuers hedge their obligations is up to them. This means that there is no direct way to create an arbitrage between the ETN and its fair value. This led to severe dislocation between the fair value of XIV and its price on Monday afternoon. By the close, the one-month future had risen by 45%, yet XIV had only dropped by 15%. Directly after the stock market close the VIX futures spiked to an increase of 100% on the day, which triggered the acceleration event in XIV.

This after-hours jump was not due to any nefarious manipulation. Around the close, ETNs rebalance their exposures to the VIX. So, on this day, short VIX ETNs needed to buy futures to reduce their exposure and long VIX ETNs needed to buy VIX to increase exposure. This severe buying pressure created a large imbalance and drove the price higher. Even if the actual product issuers were hedged with swaps, the counterparties to those agreements would have needed to hedge.

Lessons

- It is never good to be so leveraged that you are *forced* to exit a trade.
- ETNs are more dangerous than ETFs.
- It is dangerous to be in a strategy or product that has to make trades and has no discretion in doing so.

- It is even more dangerous when other people know what you need to do and can push the market against you.

> "The day you say you have to do something, you're screwed. Because you are going to make a bad deal."
>
> **—Billy Beane, general manager of the Oakland As from *Moneyball* by Michael Lewis**

Trading Strategies

I assign each edge a rating on a scale of one to three with three being the best. This is a subjective score based on the amount of empirical and theoretical support, the longevity of the effect, and the volatility of the associated trade results. I also suggest ways to implement each effect. These are also subjective.

Confidence Level Three

These strategies are based on effects that have been well documented over either many markets or time periods. They also have convincing theoretical bases. These can form the majority of a trader's strategy portfolio.

Implied Volatility Term Structure as a Predictor It is well known that the term structure of commodity futures is a predictor of future returns (see, for example, Erb and Harvey, 2006; Gorton and Rouwenhorst, 2006; Gorton et al., 2013). When a commodity has a term structure in contango (long-dated futures more expensive than short-dated futures), it is profitable to short the futures. And when futures are in backwardation (long-dated futures cheaper than short-dated futures), it is profitable to buy futures. Essentially, the cash price is a better predictor of futures prices than the futures are. The way convergence happens at expiration is that the futures move toward the cash price. This is not what is predicted by the rational expectations hypothesis. Rational expectations say that the futures prices are unbiased predictors of the future value of the cash price. If this was true, the basis would have no predictive power.

The effect also exists for VIX futures. Simon and Campasano (2014) present evidence that the futures can be predicted by looking at the basis. That is, if the futures are trading over the index, VIX, the futures will tend to fall, and if the futures are trading below the index, they will tend to rise (see Chapter Four).

The VIX term structure is usually in contango at low VIX levels and in backwardation when the VIX is high. This means the term-structure anomaly is often in conflict with the mean reversion of the VIX. Mean reversion would encourage you to buy futures at low VIX levels even in the presence of contango. However, it seems that the term structure is usually a stronger predictor of futures' returns. The exception would be when the VIX is at very high levels.

This is probably why the term-structure effect fails to apply to the cross-section of equity option returns. Vasquez (2017) found that long straddle portfolios with high contango in the volatility term structure outperformed straddle portfolios with low contango or backwardation. My guess is that this is because the entire universe of stocks contained enough examples of extreme cases for mean reversion to dominate.

This effect has existed in commodity futures for at least as long as we have had data to look at. This points to it being a mispriced risk premium. Keynes's (1930) theory of normal backwardation says that producers hold short futures positions to hedge against price drops. They are prepared to pay a premium for this insurance. However, you can also argue exactly the opposite side: consumers take long futures positions in order to hedge against unexpected future price rises. So, they should pay an insurance premium. The debate continues.

Trading Strategy

Sell VIX futures or index options when the term structure is in contango. Buy VIX futures or index options when the term structure is in backwardation.

Options and Fundamental Factors

It is now well accepted that certain factors predict future stock returns. This is the idea behind “smart beta.” Value stocks outperform growth stocks. Small-cap stocks outperform large-cap stocks. Low-beta stocks outperform high-beta stocks. High-momentum stocks outperform low-momentum stocks. High-quality stocks outperform low-quality stocks.

Factor investing is not a new idea. Academic studies began with the development of the capital asset pricing model (CAPM) by Treynor (1962), Sharpe (1964), Lintner (1965), and Mossin (1966), which suggested that individual stock returns were driven by the broad market. Expanding this idea, the arbitrage pricing theory (APT) of Ross (1976) modeled stock returns as driven by many different factors. Unfortunately, the model did not say what these factors were, but nonetheless APT gave a solid theoretical underpinning to the idea of factor investing. Once a theoretical basis was established its implications could be tested and explored. This led to academics discovering a number of different investing factors or “anomalies” (so named because they didn't fit into the world of CAPM).

What is less well known is that similar fundamental factors also predict volatility returns. The number of studies that relate factors to volatility is much smaller than the literature on smart beta stock returns and unfortunately researchers haven't studied exactly the same factors as each other.

I did a study constructing option trading strategies based on P/E (the price-to-earnings ratio or the current stock price divided by the previous year's earnings per share), P/B (the price-to-book ratio or the current stock price divided by the previous year's balance sheet assets), RoA (the return on assets), RoE (the return on equity or the return on assets after all liabilities have been settled), market capitalization (the dollar value of the company's outstanding shares), D/E (the debt-to-equity ratio), and P/CF (the ratio of the stock price to the previous year's cash flow per share).

The universe of stocks considered was the S&P 100 after excluding financial companies. Naive application of accounting metrics can give a misleading picture of financial companies because their business is based on providing loans. So, their “assets” are actually liabilities, which makes things confusing. The time period we looked at was from the start of 2000 through to the end of 2012. On the Friday of each week I ranked all the stocks according to the valuation metrics and then formed an option portfolio based on this ranking by selling straddles on the top quartile of stocks and buying straddles on the bottom quartile. Specifically, we traded at the money straddles in the second monthly expiry. Each trade is done in a notional size of $10,000. For example, on a $100 stock we would trade one straddle (each straddle controls 100 shares and each share is worth $100). Correspondingly, we would trade two straddles on a $50 stock. Under standard strategy-based margin this sizing choice gives a margin of roughly $100,000. Portfolio margins would be considerably smaller. This portfolio is held for a week, then the process is repeated.

P/E Straddle Trading Results

- Long options on low P/E stocks and short options on high P/E stocks
- Average weekly PL: $294
- Best week: $21,111
- Worst week: –$8,111
- Sharpe ratio: 1.03

P/B Straddle Trading Results

- Long options on high P/B stocks and short options on low P/B stocks
- Average weekly PL: $48
- Best week: $12,094
- Worst week: –$7,246
- Sharpe ratio: 0.24

Market Capitalization Trading Results

- Long options on high-cap stocks and short options on low-cap stocks
- Average weekly PL: $355
- Best week: $15,532
- Worst week: –$23,077
- Sharpe ratio: 1.17

P/CF Straddle Trading Results

- Long options on low P/CF stocks and short options on high P/CF stocks

- Average weekly PL: $198
- Best week: $22,235
- Worst week: −$13,637
- Sharpe ratio: 0.71

D/E Straddle Trading Results

- Long options on high D/E stocks and short options on low D/E stocks
- Average weekly PL: $338
- Best week: $25,382
- Worst week: −$5,292
- Sharpe ratio: 1.16

RoE Straddle Trading Results

- Long options on high-RoE stocks and short options on low-RoE stocks
- Average weekly PL: $361
- Best week: $16,922
- Worst week: −$19,102
- Sharpe ratio: 1.20

RoA Straddle Trading Results

- Long options on high-RoA stocks and short options on low-RoA stocks
- Average weekly PL: $220
- Best week: $16,539
- Worst week: −$21,702
- Sharpe ratio: 0.72

The results of the P/B strategy are disappointing, but the others are good. The results can be summarized simply. Try to be long volatility (or options) in stocks with these characteristics:

- Low P/E
- Low P/CF
- High market cap
- High RoE
- High RoA
- High D/E

These are classic indicators of “value” stocks apart from the high debt-to-equity ratio. This is a strange anomaly. High debt to equity (or high leverage) implies a higher risk of bankruptcy. This would clearly be a high-volatility event for the stock. However, these are all large companies that have presumably a very low chance of going bankrupt. Also, high leverage could imply that the bond market sees value in the company. Disentangling this conundrum would involve further work. Does the effect hold in smaller companies? Does it matter how the leverage reached current levels? It could well be that high leverage resulting from a declining stock price is very different from high leverage from new bond issuance. Does the creditworthiness of the debt matter more than the absolute size of the debt?

We can combine our findings to construct a portfolio of options that is long “value” volatility and short “growth” volatility. To do this we first create a combined ranking consisting of one-sixth the P/E ranking plus one-sixth the P/CF ranking and so on. This portfolio has these results:

Portfolio Straddle Trading Results

- Average weekly PL: $476
- Best week: $18,963
- Worst week: −$9,429
- Sharpe ratio: 1.44

A similar study was done by Cao et al. (2015). They studied delta-hedged option returns for US equity options from 1996 to 2012. They found that long volatility positions profits increase with size, momentum, and company profitability and decrease with cash holding, analyst forecast dispersion, and new issuance. These option returns are independent of any associated underlying predictability.

Their study showed the profitability of portfolios sorted into deciles. The long-short 10/90 portfolios had monthly returns of the order of 3% to 5%. (As usual this was calculated on the notional value of the positions. This makes sense for long positions but not for short positions, where a risk-based margin number would be a more appropriate denominator.) Sharpe ratios were between 0.6 and 2.0.

There is a lot more work that could be done along these lines and this research raises a lot of questions. For example, how independent are these various measures of value when applied to option trading? Are these results independent of the results from time series analysis? How do these results change across industry groups?

As always, the phenomena are more important than the particular implementation. The best way to trade this is probably to build a factor portfolio instead of doing single factor sorts. Alternatively, the fundamental factors could be used to directly forecast volatility. Trading options based on smart beta is still a new and unexplored idea, but it is probably a more promising avenue than trying to improve time series–based forecasts.

Generally speaking, there are two views as to why factor returns exist: risk and investor behavior. The risk school of thought says that the return is simply reward for bearing some sort of risk. For instance, the equity market premium is a result of the greater uncertainty associated with owning stocks as opposed to bonds. This explanation fits well with mainstream economic theory but sometimes it is a little hard to figure out what exact risk is being compensated for. As an example of this, volatility is usually associated with risk, but historically low volatility stocks have outperformed high-volatility stocks. The behavioral explanation contends that investors systematically make decisions that cause these anomalies. Again, these arguments seem more persuasive in some cases than in others. Momentum, for example, has plausible behavioral roots but in other cases trying to identify a psychological reason for a factor seems like merely searching for a tidy explanation, rather than doing real science and following the data. These explanations are summarized in Table 5.1.

Whether stock smart beta is inefficiency due to behavioral biases or risk premia is still being debated. And even if that dispute is resolved, the corresponding volatility effect will probably still be argued over. However, applying factor analysis to options trading is where I will be spending most of my future research efforts. It is a very new field so the risk and uncertainty is high, but I'm sure it is better to deal with this than to try to scratch another tiny improvement from a time-series method.

TABLE 5.1 Postulated Risk and Behavioral Reasons for the Smart Beta Factors

Factor	Risk Explanation	Behavioral Explanation
Size	Smaller firms are less able to weather bad periods and have less diversified businesses.	Smaller firms are generally poorly covered by analysts, which leads to investor uncertainty.
Value	Cheap stocks tend to be those of the companies that perform worst in periods when the overall economy is suffering.	Investors pay too much attention to recent stock performance and overly fear distressed situations.
Momentum	Momentum can lead to bubbles and crashes.	Underreaction: new information is not incorporated into prices instantaneously.
Quality	It is harder for a high-quality company to improve. It may only have downside.	High-quality firms may be "too good to be true."

Post-Earnings Announcement Drift (PEAD)

Post-earnings announcement drift (PEAD) is the tendency for a stock to continue to move in the direction caused by an earnings surprise. The first study of this effect was by Beaver (1968), who showed that prices react to the information content in earnings reports. Ball and Brown (1968) found evidence that stock returns continue to drift in the same direction as unexpected earnings. That is, stock prices tend to increase (decrease) after earnings announcements with positive (negative) earnings surprises. This wasn't greatly shocking. What was unexpected was that the outperformance didn't happen abruptly but accumulated over three months. The stock prices did incorporate the news but did so slowly. In fact, the prices reacted so slowly that the effect was tradable. This effect is inconsistent with the concept of efficient markets, where the information contained in an earnings report should be quickly incorporated in prices, but even one of the inventors of EMH acknowledges the existence of the anomaly (Fama, 1998).

He wrote in the abstract, “Most long-term return anomalies tend to disappear with reasonable changes in technique” or when “alternative approaches are used to measure them” (p. 288), but concluded, “Which anomalies are above suspicion? The post-earnings announcement...has survived robustness checks, including extension to more recent data.” He has also written that PEAD is the only anomaly that is “above suspicion” (p. 304).

Early research focused on the change in year-over-year earnings. But in the 1980s the consensus of analyst forecasts became available. This enabled the magnitude of the earnings surprise to be defined as the percentage difference between the reported EPS and the consensus forecast. Sorting stocks on the basis of this surprise gave the same result. Further, it seems that the effect of the earnings report isn't limited to the headline EPS number. The momentum created by the earnings report is a follow-through of the initial price move based on the entire report, rather than being due to any single accounting number.

Bernard and Thomas (1989, 1990) show that the drift is robust to risk adjustments, cannot be attributed to market frictions, and is not due to model misspecification, standard risk factors, or flaws in research design. These conclusions have been verified in many subsequent studies. PEAD seems more likely to be an inefficiency than a risk premium.

There is some evidence that it is most pronounced in firms that report after their direct competitors and have earnings that outperform those firms by a surprising amount (Lee, 2017).

The effect is remarkably robust. Long-short stock portfolios based on the size of the surprises have been shown to return 9% to 27% annually, depending on the sorting method (deciles, quartiles, etc.), the definition of surprise, and whether returns are raw or adjusted with respect to sector, beta, size, value, and momentum.

Many studies (for example, Foster et al., 1984; Bernard and Thomas, 1989; Watts, 1978) have shown that the anomalous drift occurs over a period of months. Almost all of the drift occurs within 9 months for small firms and within 6 months for large firms. And a disproportionate amount of the return occurs within 5 days of the earnings announcement (13% for small companies, 18% for medium companies, and 20% for large companies).

Unfortunately, what is not clear is why this anomaly exists.

A number of behavioral explanations have been proposed. If the effect is predominantly caused by psychological biases, that would point to PEAD being an inefficiency rather than a risk premium. Some research (for example, Daniel et al., 1998) suggests that investor overconfidence causes investors to be anchored to the pre-earnings price and incorporate the new information only slowly.

Another possibility is that investors don't have the time and resources to process the new information quickly. DellaVigna and Pollet (2009) postulate that this effect is even more pronounced on Fridays when investor attention is at its lowest. Although this seems like a reach or a post-hoc justification of data mining, a long-short portfolio of stocks announcing on Friday did outperform a portfolio based on stocks that announced on other days (9.76% over the subsequent 75 trading days versus 5.14%). Hirshleifer et al. (2009) also reason that information processing capacity is a reason for the anomaly. They show that a hedged portfolio constructed from stocks that announce on days with few other releases underperform a portfolio of stocks that announce on busy days (2.81% quarterly compared to 5.37%).

Bartov et al. (2000) show that firms with high institutional ownership have lower PEAD. They suggest that this is because these supposedly more sophisticated investors are faster to absorb the new information. Similarly, Taylor (2011) shows that the anomaly is largest for stocks where retail traders trade against the direction of the initial price spike. Further, he also finds a positive correlation between PEAD size and transaction costs, which presumably are proportionally a lower deterrent to larger investors due to their more sophisticated execution methods and access to a wider range of liquidity sources.

There are also explanations that ascribe PEAD to a risk factor. For this to be true firms with positive (negative) earnings surprises must become riskier (less risky) after the release. It is well accepted in classical finance that stocks with higher returns are riskier, and if we are looking for a rational expectation explanation we need to accept this as true. The harder thing to explain is why stocks become riskier immediately after the earnings report is published. Despite attempts to do this (Bernard and Thomas, 1989) no convincing supporting evidence has been discovered.

In summary, the available evidence suggests that PEAD is an inefficiency, albeit a very persistent one.

Trading Strategy

Because implied volatility tends to be comparatively cheap after earnings releases (although this isn't enough to profit from delta-neutral volatility strategies; see Chung and Lewis, 2017), this is an effect that can be traded from the long side. For stocks with positive earnings surprises, being long a short-dated 40 delta/10 delta call spread is usually a cheap way to leverage PEAD. For bearish positions a 50 delta/20 delta put spread achieves the same effect while also selling one of the most expensive parts of the implied volatility curve.

An alternative approach is to sell a covered call or put. Although the variance premium being collected won't be significant, this method is guaranteed to profit if any drift occurs at all.

Confidence Level Two

These strategies are not as promising as those of level three. They may be lacking in empirical evidence or theoretical justification. Or they could provide more limited opportunities in some way. Nonetheless, a trader can be reasonably confident when using these strategies.

Trading Equity Options over Earnings Announcements

The front-expiration implied volatility will almost always rise significantly in the weeks before a company's earnings release.

After the release the implied volatility will collapse. This effect has been known since at least 1979 (Patell and Wolfson, 1979).

Buying options before the release is profitable (Chung and Lewis, 2017). And even in cases when the options do not increase in price, the rise in implied volatility gives a very cheap long position where the vega profits negate most of the time decay.

I conducted a test of this effect in Sinclair (2013). Using data for 73 large cap stocks from Q1 2005 to Q3 2010, a strategy of buying front expiration, ATM straddles 10 trading days before the earnings release had these results:

- Forty-three percent of trades were profitable.
- Win size/loss size was 1.70.
- Results were skewed positively (skewness of 5.4) and fat-tailed (excess kurtosis of 76.6).

Putting a $10,000 notional size bet on each trade led to the cumulative P/L curve of Figure 5.1.

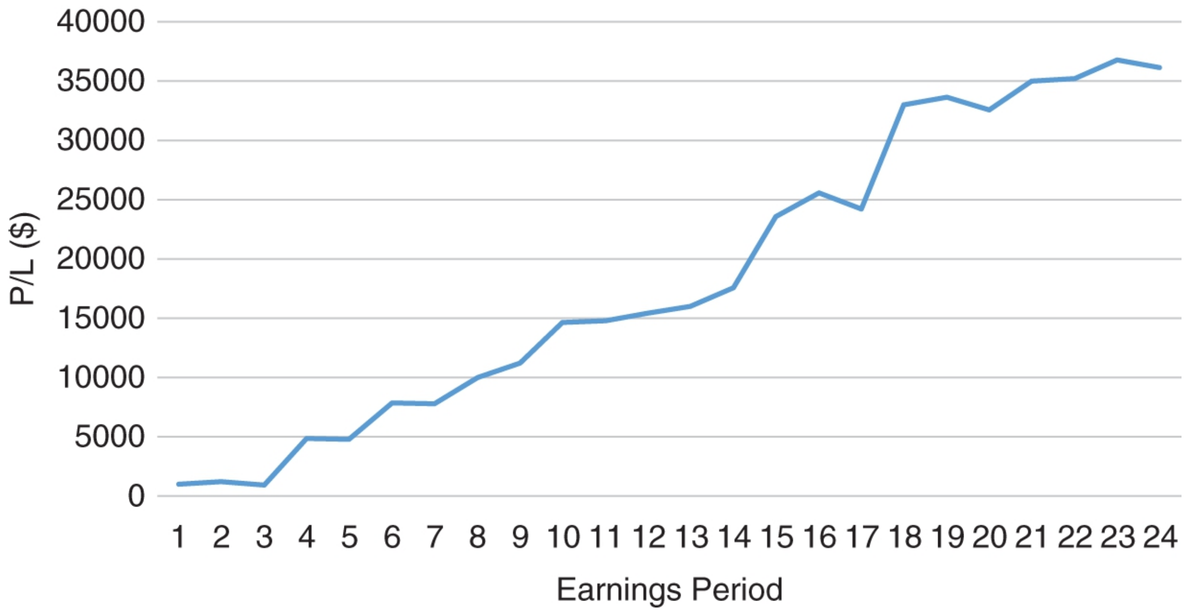


FIGURE 5.1 Results of the long straddle strategy.

A far more comprehensive study was completed by Gao et al. (2017). They bought delta-neutral straddles three days before the earnings announcement. In the period 1996 through 2013, selling a day before the release gave a return of 1.9%, selling on the release day returned 2.60%, and holding for a day after earnings had a return of 1.98%. (All of these returns were calculated using mid-market prices.) These time periods are not entirely independent. Some companies on the earnings date report before the open and some after the close. So the second group includes both pre- and post-report companies. Notably, holding the straddle for a full day after earnings had a lower return than selling on the release date. This indicates that the return on the straddle is negative after earnings are announced.

The authors don't present a separate study for the post-earnings period. But using my smaller sample, I directly tested a short volatility strategy. As close as possible before earnings, sell the front straddle and hope that the collapse in implied volatility compensates for the movement of the underlying.

The results are as follows:

- Sixty-four percent of trades were profitable.
- Win size/loss size was 0.69.
- Results were skewed negatively (skewness of negative 5.4) and fat-tailed (excess kurtosis of 12.2).

FIGURE 5.2 Results of the short straddle strategy.

Putting a $10,000 notional size bet on each trade led to the P/L curve of Figure 5.2.

There haven't been many published studies of this effect and the sample size of my study is very small, so it is dangerous to draw any conclusions about the profitability of subcategories of stocks. However, it seems that there is a small tendency for this strategy to fare poorly with high PE stocks. Studies have shown that such growth and value stocks do react to earnings differently and this effect might be related (He et al., 2010). However, the difference is so weak it can probably be ignored.

There was no significant relationship was between profitability of the option strategies and industry or company size. However, there is a link between the dispersion of analyst estimates (defined as high estimate–low estimate)/average estimate) and profitability. Trading the effect on stocks with high dispersion is more profitable. High dispersion literally means that the analysts disagree about the unreleased earnings, a situation of uncertainty. The profitability of option trades with edge, either from the long or short side, is generally tied to uncertainty.

Other signifiers of uncertainty are with higher historical volatilities, larger historical earnings surprises, and more volatile earnings surprises. The strategies also perform better in these cases. Gao et al. (2017) also found that the strategy was more profitable in stocks that were covered by fewer analysts.

There are likely two reasons for this. High uncertainty leads to a greater increase in implied volatility, so we will be selling the options at a good price. But we also benefit from the uncertainty in the actual move of the stock price that happens due to the release. Stocks move when news comes out. News is something contrary to expectations. Uncertainty implies no coherent expectations. There can be no “surprise” if no one agrees on what to expect.

The fact that pre-earnings long straddles are profitable makes it unlikely that this effect is due to a risk premium. Gao et al. (2017) postulate that one reason for the underestimation of uncertainty is the difficulty of extracting a signal from sparse, noisy past information. This seems plausible.

Separating the effect into the pre-earnings implied volatility increase and the post-earnings implied volatility collapse and stock price adjustment makes finding possible explanations easier. The pre-earnings situation seems to be driven by similar factors to PEAD. That is, investors can process the information only from previous earnings results slowly and inefficiently. This is consistent with the link to analyst coverage and estimate dispersion. However, unlike the case of PEAD, there have been no further studies to test this idea. The profitability of the post-earnings straddle sales could be due to ambiguity aversion. But, again, this has not been independently verified or studied.

The short side of this strategy hasn't done well recently (2016–2018). I think this is probably a temporary, variance-driven period of poor performance. Usually, selling options before situations of uncertainty is a good trade. I still think that is the case here. As any casino knows, sometimes the customers have to win. Otherwise they won't come back.

Trading Strategy

On the long side, buy the shortest-dated straddles that span the earnings date. Or, if the smile is relatively flat, strangles are a higher risk/higher reward play.

On the short side, sell the shortest-dated straddles that span the earnings date. After the earnings release, either buy the straddle back (generally preferable if the trade is a winner), or, if the stock has a big move, hedge the losing leg as 100 delta. Over-hedging one side of the trade usually works due to PEAD, and the short-dated options not giving the stock enough time to reverse. Stocks in which the earnings number is outside of the most extreme estimates are most likely to not reverse.

The Overnight Effect

The index variance premium is realized overnight. Muravyev and Ni (2018) wrote a paper that studies this. Although it is very well known that returns to index options are negative (about −0.7% a day in terms of actual premium decay for the S&P), it turns out that all of this comes from overnight decay. Specifically, delta-hedged option returns are −1% overnight, and intraday returns are positive at 0.3% per day. The anomaly holds for options of all maturities and moneyness and also equity options.

This can be partially explained by the fact that overnight returns are less volatile than intraday returns (since 2000 the S&P has had an annualized volatility of 10.4% overnight and 15.9% during the day). The effect is also consistent with the variance premium being a mispriced risk premium. The risk is much higher during the untraded (or illiquid) hours. We get paid for taking the risk of having a bet that we can't get out of or hedge if it goes against us.

Trading Strategy

Being short options overnight and flattening during the day is impractical due to transaction costs. Instead the best way to profit from this is to sell very short-term options that include as many overnight periods as possible. For example, instead of selling daily options on the morning of expiration, sell them the night before.

FOMC and Volatility

The most important news for an equity is the earnings release. The equivalent for the broad market is the monetary policy decision of the Federal Reserve Open Market Committee (FOMC). And, just as company earnings releases lead to predictable pre-event price drift and post-event volatility collapse, analogous phenomena occur in the broad market around the FOMC announcements.

In “When No News Is Good News—The Decrease in Investor Fear after FOMC Announcements” Fernandez-Perez et al. (2017) show that the VIX and VIX-Futures drop significantly after the announcements of the meeting.

From their abstract, “We find that the VIX and the VIX futures start to decline immediately after the FOMC announcement, and this decline persists for about 45 minutes after the announcement. The VIX declines by about 3% on announcement days, whereas the nearest term VIX futures contract declines about 1.4% around the announcement.”

Similar results have been obtained by others. Nikkinen and Sahlström (2004) showed that the VIX index declined after FOMC announcements, CPI and PPI releases, and the Non-farm Payroll report. Chen and Clements (2007) and Vähämaa and Äijö (2011) replicated the results for the VIX. Gospodinov and Jamali (2012) found both implied and realized volatility dropped after the release and any initial market reaction. Fuss et al. (2011) showed that German equity implied volatility decreased after the release of GDP, CPI, and PPI data, and Shaikh and Padhi (2013) confirmed the effect in the Indian market.

The stock earning's related volatility collapse has to be played with options. Implied volatility will reliably collapse following the news release, but this will generally be partially offset by an increase in realized volatility. Trading the FOMC-related implied volatility collapse with VIX futures avoids this problem. We can trade an implied effect with an implied contract and have no realized volatility exposure.

Why does implied volatility collapse? Ederington and Lee (1996) argued that the release of prescheduled news can be viewed as the resolution of uncertainty, and an announcement reduces the degree of future uncertainty, and hence implied volatility, left in the market.

This effect can be combined with a directional anomaly. Equity market returns have been about 30 times larger on announcement days than on normal days. Surprisingly most of these excess returns occur *before* the announcement (Lucca and Moench, 2011). Stocks tend to rally strongly in the day and a half before the announcement.

It is difficult to think of a convincing risk-based reason for these returns. Although investors holding stocks over the announcement period are exposed to jump risk, the excess return occurs mainly before the announcement. So, it is possible to sell the stocks before the news and earn the return without taking this risk. Further, realized volatility is lower than usual before the release. Market risk is actually lower than normal in the period when the excess returns are generated.

It seems more likely that this is a market-wide version of the pre-earnings drift seen in individual equities. As in that case, investors buy stocks because of an attention-grabbing event (see Barber and Odean, 2008; Lamont and Frazzini, 2007).

Trading Strategy

Buy S&P 500 futures two days before the FOMC release. Sell these before the announcement and at the same time sell VIX futures. Cover the short VIX futures 45 minutes after the announcement.

The Weekend Effect

Equity options decay more than expected over weekends. The variance premium is probably the most important effect in option trading. On average options are overpriced. But the premium doesn't accrue consistently: some periods are better than others. In particular, options lose more value over weekends than over other equivalent time periods.

Market-makers have been arguing about this for a long time. One group said it was smart to sell options to “collect theta.” The other side said that idea was stupid because an edge could only exist when one side had an information disadvantage. Everyone had a calendar, so why would there be an edge purely due to the passing of time? The market would price the risk of holding options over the weekend correctly.

The “smarter” traders were in the second group. They were wrong. The options market does not correctly price weekend decay. It is profitable to sell options over the weekend.

Christopher Jones and Joshua Shemesh studied this issue (2017). They looked at the returns of long option portfolios on US equities from 1996 to 2007 and found the average return over the weekend was negative (0.62%) while the returns for all other days were slightly positive (0.18% a day).

Having established that weekend returns are significantly lower than those of other days, the authors went on to study other holidays, including *long* weekends. Their hypothesis was the effect was directly related to non-trading, which would imply lower returns would also be associated with other holidays, and the effect would be stronger over long weekends. This all seems to be true: returns on equity options are negative whenever the market is closed. It seems the effect exists because market-makers are not correctly adjusting the implied volatilities on Fridays to account for the upcoming weekend.

This effect is significant. There is no general edge in selling many stock options (unlike index options, when being short is normally the way to lean), so this is a totally new effect, rather than a matter of just timing an entry. Also, the effect was consistent across years and was robust with respect to exactly how the portfolios were constructed.

It seems likely that the same effect exists with index, bond, and commodity options, but this is untested.

Trading Strategy

On a Friday, sell the options that expire the next Monday. In general, time any short-volatility strategies to include as many non-trading periods as possible.

Volatility of Volatility Risk Premia

Options on products with high volatility of volatility tend to be overpriced. This is true both in the cross-section—options on stocks with high volatility of volatility are overpriced relative to options on stocks with low volatility of volatility—and in the time series—the VIX tends to decline (rise) after very high (low) values of the VVIX index.

The first empirical study of this effect was done by Ruan (2017). Using data on US equity options from 1996 to 2016, he found that ranking stocks by the volatility of their implied ATM volatility showed that there was a strong and consistent negative relationship between delta-neutral long option positions and volatility of volatility.

A similar study was independently carried out by Cao et al. (2018), who also studied US equity options. Again, using data from 1996 to 2016, they found that the delta-hedged returns of long option positions decreased in uncertainty of volatility. This was true whether they used implied volatility, time series volatility from daily returns (specifically EGARCH), or high-frequency volatility. Their results were robust with respect to idiosyncratic volatility, jumps, term structure, the implied–realized spread, liquidity, analyst coverage, and the Fama-French factors. They also showed that the effect was largely driven by volatility of positive volatility moves, and that volatility of negative volatility moves had a negligible effect.

These studies leave little room to interpret this effect as anything other than a separate volatility of volatility premium. Ruan (2017) just states, "Investors indeed dislike uncertainty about volatility of individual stocks, so that they are willing to pay a high premium to hold options with high VOV [sic]," with no supporting argument. Cao et al. (2018) speculates that market-makers were charging a higher premium for options with high uncertainty of volatility, because those were more difficult to hedge. This might be a partial reason, but it doesn't take into account the time-series result that shows high volatility of volatility predicts a fall in subsequent implied volatility. This effect is independent of hedging issues.

The relationship between high VVIX (the model-free implied volatility derived from VIX options) and subsequent lower VIX levels is very strong. Using VVIX data from 2007 through 2018, I calculated the rolling 1-year 90th percentile of VVIX. Going forward, if the VVIX crossed above this level, I "sold" the VIX and "held" until VVIX reached its rolling 1-year median. This produced 31 hypothetical trades. The total "profit" was 108 points. Twenty-seven trades were winners. "Buying" the 10th percentile was also "profitable," making 62 points over 35 trades, 26 of which were winners. Clearly this particular idea cannot be implemented because the VIX is not a traded product, but I've included it to show that extreme VVIX is a strong predictor of the VIX (however, if we traded VIX futures, the idea is still profitable). No optimization was attempted. The idea also works if we use different look-back periods or moving averages instead of medians.

This effect has been studied (far more rigorously) by others. Huang et al. (2018) showed that volatility of volatility significantly and negatively predicts delta-hedged long option payoffs. Park (2015) showed that high levels of VVIX raised the prices of S&P 500 puts and VIX calls and lowered their subsequent returns over the next three to four weeks (a similar time period to the average holding period in my simple test). He speculates that the effect is caused by either "risk premiums for a time-varying crash risk factor or uncertainty premiums for a time-varying uncertain belief in volatility." Both of these are plausible but at this point there is no independent evidence for these causes.

Trading Strategy

When VVIX reaches extremely high (low) levels either sell (buy) VIX futures or sell (buy), and dynamically hedge, S&P 500 straddles.

Confidence Level One

The confidence-level-three strategies should form the core of a trading operation. But the ideas that I think are true but only give a confidence rating of one are also important. Trades based on market inefficiencies will be most profitable when the evidence for them is still underwhelming. Many inefficiencies will not survive long enough to reach my level three. So, although I wouldn't allocate a great deal of my portfolio to these ideas, they can still be very profitable.

They also offer a way to deal with the desire to gamble. Many traders overtrade and need to always be involved in the market. Instead of denying this tendency, it is better to accept it and learn to accommodate this need by tinkering with small trades that still have expected edge. Level-one trades are perfect for this. This is like the idea of a “cheat meal” when dieting. Instead of trying to religiously stick to a diet it is better to accept that temptation exists and schedule regular times when you can eat garbage. Dieting increases cravings (Massey and Hill, 2012). There is solid psychological research that shows that dieters who include cheats do better than those who don't (do Vale et al., 2016). I expect that active traders are tempted to over-trade and that cheating helps them as well.

Remember that cheats, whether in dieting or in trading, need to be kept small. If every meal is a cheat meal, you aren't on a diet. You will just get fat. And if every trade is a speculative one, you aren't a disciplined trader. You will just lose money.

Earnings-Induced Reversals

Earnings-induced reversals are the tendency of stocks that have drifted a lot before their earnings announcement to reverse the pre-announcement drift when the news breaks. This effect was first studied by So and Wang (2014). Using US equity data from 1996 through 2011, they created a trading strategy that shorted stocks with high market–adjusted returns in the period from four days before earnings through two days before earnings and went long those stocks with the worst pre-earnings market adjusted returns. (This seemingly odd time period was so that they could trade on the close of the day before earnings without using the trade price when choosing the portfolio. Obviously, a trader using intra-day data could use a different time period without “cheating.”) Liquidating the portfolio on the close after earnings they found this portfolio made 145 bps compared to the 22 bps earned by a similarly constructed portfolio during non-earnings periods.

A similar study was done by Jansen and Nikiforov (2016). Simply fading stocks with large percentage moves in the week before earnings would have averaged 1.3% over 2-day periods.

Jansen and Nikiforov (2016) speculate that the effect is due to investor overreaction in the pre-earnings period. Individual investors fear that they are missing information and trade in the direction of price changes, fueling the trend. After the announcement, the fear of being ignorant of information goes away and the pre-earning return is seen as excessive. This might be true. A similar effect is seen in sports gambling, when “steam chasers” bet on teams that have shortening odds on the suspicion that smart money is driving the price changes. But much more would need to be done before I am confident that this is an inefficiency. Currently the statistics are inarguable, but the reasons for them are close to a mystery.

Trading Strategy

I'm more confident in the collapse of implied volatility when earnings are released than I am of this reversal effect. So, when I trade both of these effects together I sell a straddle but shade the delta if I want to also bet on the reversal.

Pre-Earnings Announcement Drift

Pre-earnings announcement drift is the tendency of stocks to move in the direction of any earnings-related abnormal returns experienced by stocks in the same industry that reported earlier. This effect was first studied by Ramnath (2002), who investigated how information from the very first earnings announcer within each industry (the 30 industries identified by Fama and French, 1997) affects the prices of later announcers. He found that the earnings information for the earliest announcing firm within an industry predicts both the earnings surprise and the returns of other firms within the industry.

This effect was later confirmed by Easton et al. (2010), who used not just the first reporter in each industry but also the effect of all the earlier announcing peers.

The drift begins as the results from the early announcers are reported and continues up until the later announcing stock releases its earnings. The effect is above the industry beta, which measures the normal relationship between returns. If earlier reporting stocks all rally, we would expect later reporting stocks to also rally just due to industry exposure. Pre-announcement drift is a separate effect.

Pre-earnings anomalies have not been studied nearly as much as post-earnings anomalies, so the evidence is comparatively weak, and it is not clear what causes the drift. As with PEAD, the pre-earnings move is plausibly due to underreaction to new information: here the earnings of the related companies. It could be that investor overconfidence causes them to be anchored to the pre-earnings price and incorporate the new information only slowly. A lot more study would be needed before we could be confident in this explanation. But there is no obvious risk factor that could explain the drift, so I would say, tentatively, that this is a market inefficiency.

Trading Strategy

As we also expect implied volatility to increase in the time leading to the earnings release, any long volatility directional strategy would be appropriate. For example, if we expect a rally, we could buy a call or call spread. My preference is for a 50 delta/20 delta, 1-month call spread. But tastes vary.

Conclusion

The idea that trading edges disappear as soon as they become public is an oversimplification. Markets vary in their ability to absorb new volume. A published edge will persist longer in the S&P 500 than in soybeans. Further, crowding affects different strategies in different ways. And risk premia will survive longer than inefficiencies.

But unless noted, the edges listed in this chapter have been robust until now. It is quite possible that their size will diminish and even disappear but we have a fairly basic choice: go with the effect that has worked in the past and hope it continues or choose to do the thing that would have lost money in the past. Your choice.

Summary

- It is worthwhile to search SSRN periodically to find new trading ideas.
- Many volatility trading edges involve selling options in situations of uncertainty. This can be viewed as an extra, situational variance premium.
- Because of the variance premium, long-volatility strategies are unlikely to have as much edge as those that involve selling options.

CHAPTER 6
Volatility Positions

One of the things that make options great is that there are many ways to express an opinion. But this is also one of the things that make options tricky. Just because there are many ways to express an opinion doesn't mean they will all be equally good. The differences are not trivial. Some will be a lot worse than others.

In this section we will compare some option positions that are primarily used to express views on volatility. We will look at the possible distribution of returns by using both GBM returns and historical data. We will also look at the effects of the underlying having a drift. This will generally be done from the perspective of a volatility seller, but the case of long volatility is a trivial extension.

All of the simulations will assume that we initiate the position and then leave it alone until expiration. In reality, we will usually have opportunities to trade out of the position before then. But it is important to understand the terminal distribution of the P/L for several reasons:

- Even an adjusted (or hedged) position is instantaneously subject to the same issues as one that won't be adjusted in the future.
- Very short-dated options (depending on the market liquidity this could be weekly, daily, or hourly) can't meaningfully be adjusted.
- The actual adjustment procedure will be different for different traders so will be impossible to simulate.

Aside: Adjustment and Position "Repair"

"Repair" is a dangerous misnomer. First, in any other situation to repair something is to return it to its previous condition. But in the trading world it is usually taken to mean turning a losing trade into a winning trade. This is a falsely reassuring idea, but it can't be done. The loss is already in your account. That money is gone. Forget about the original trade and ask yourself, "Given what I now know, what position do I want?" Then put that position on. This is completely independent of the original trade. This should also be done when examining winning positions. Their profits are also in the past. Do you like the position now? If not, do something else.

You should adjust a position when it no longer matches your forecast or opinion. This is true whether it has previously made money or lost money.

Straddles and Strangles

The two most basic ways to short volatility are to sell either a straddle or a strangle. It is easy to work out the expected profit of an option position. It is just the position value at the volatility we sold at, minus the position value evaluated at the realized volatility. Or in terms of vega:

$$P/L = vega(\sigma_i - \sigma_r) \tag{6.1}$$

But this doesn't tell the whole story. Unhedged option positions are not exposed to path dependency in the underlying, but their profitability is still tied to the returns of the underlying. Returns are not relevant when pricing options, but the final option return is enormously dependent on returns. And even a truly driftless process will have periods when it *appears* to trend.

Further, when selling options our upside is capped by the collected premium, but our downside is infinite. This means that all short option positions will have significant negative skew in their P/L.

Because of all this I evaluated the strategies by running simulations. I sold a 1-year ATM straddle on a $100 stock at 30% implied volatility and then simulated 10,000 paths of the stock where the realized volatility was also 30%. The drift and interest rates were both zero.

There are two ways to quantify these results. We could either track returns on the margin required (for the short straddle, strategy-based margin would be $2,000) or in terms of dollars. Most professional option traders think in terms of dollars so that is what we will do, but the overall conclusions would be very similar if we considered returns.

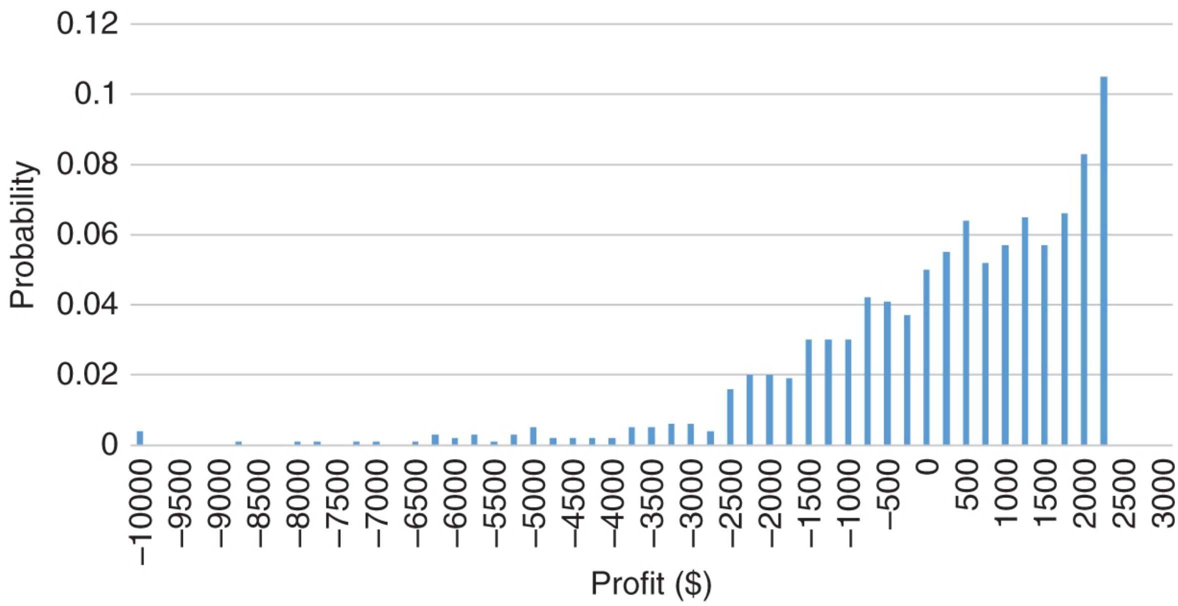


FIGURE 6.1 The profit distribution of the short straddle.

The distribution of profits (in dollars) is shown in Figure 6.1.

The dollar value of the straddle was $2,385 (assuming the convention of the options being on 100 shares). So, the maximum profit is $2,385. As the implied and subsequent realized volatilities were the same, the average profit should have been zero. The simulation confirms both of these figures. Statistics are summarized in Table 6.1. (The minimum is included as an indicative number and should not be relied on as the sampling error is huge for extrema. The 10th percentile is a better measure of downside risk.)

Next we run the same simulation for a short 70/130 strangle (corresponding to shorting the 9-delta put and the 23-delta call). To get the same vega exposure as we had with the short straddle, we need to sell 1.68 strangles. This position has an initial value of $841. The distribution of profits (in dollars) is shown in Figure 6.2.

TABLE 6.1 Summary Statistics for the Returns of a Fairly Priced Short Straddle

Average	$8.04
Standard deviation	$1,882
Skewness	−1.83
Excess kurtosis	6.41

Median	$384
10th percentile	−$2,274
Minimum	−$15,321
Percent profitable	57%

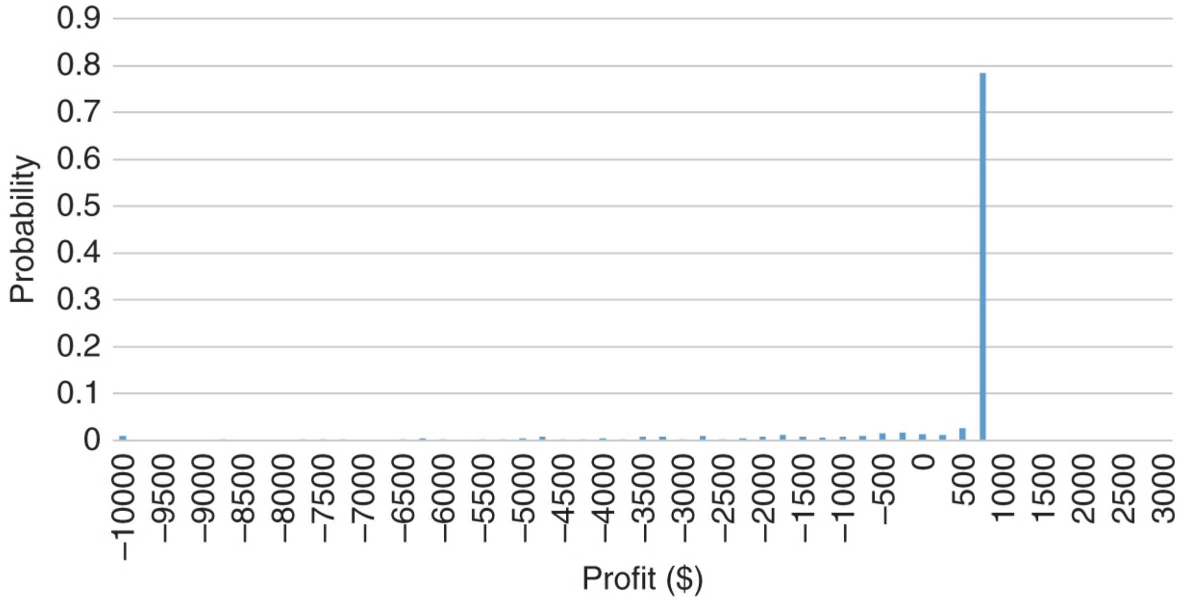


FIGURE 6.2 The profits of the short strangle.

The dollar value of the position was $841. So, the maximum profit is $841. Again, the average return should have been zero. As before, the simulation confirms both of these figures. But the results are also much more negatively skewed than for the straddle. And even if you don't look at the numbers, a quick glance at the histograms tells us a lot. The strangle hits its maximum profit about 60% of the time, but, because the total premium we took in is lower than with the straddle, the losses can be much greater. Statistics are summarized in Table 6.2.

Because there are many situations in which selling volatility has a positive expected value, when we enter short volatility positions, we should focus on controlling our risk. If we can keep plugging along, the profits should eventually take care of themselves. So instead of just comparing the straddle and strangle in the case where we had no edge, we now look at the results where we are completely wrong. Specifically, realized volatility was 70%. The straddle returns for this scenario are shown in Figure 6.3, and statistics are summarized in Table 6.3.

TABLE 6.2 **Summary Statistics for the Returns of a Fairly-Priced Short Strangle**

Average	−\$6.12
Standard deviation	\$2,140
Skewness	−4.8
Excess kurtosis	24.2
Median	\$841
10th percentile	−\$1,994
Minimum	−\$27,683
Percent profitable	78%

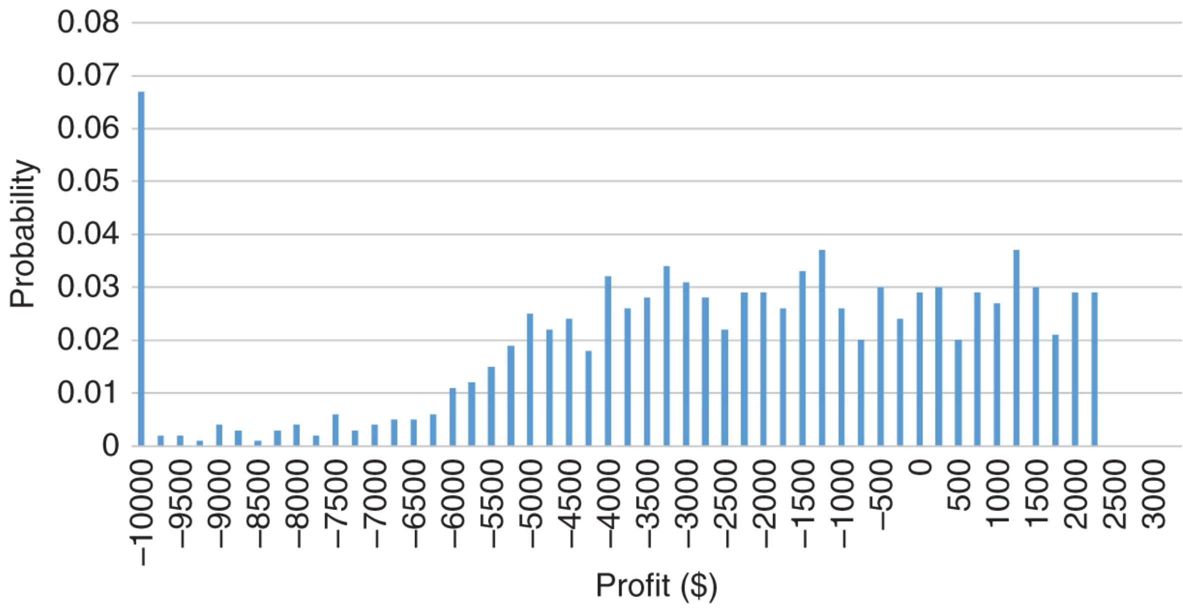


FIGURE 6.3 The returns of the short straddle when our forecast was poor.

Not a good set of results. But we were very wrong in our volatility forecast, so we can't really expect great results.

But now look at the returns of the strangle in Figure 6.4 and Table 6.4.

Again, we can see that the extreme results (minimum and 10th percentile) were worse than the corresponding straddle returns. However, generally speaking, the poor volatility forecast had similar effects on both the straddle and strangle results.

TABLE 6.3 Summary Statistics for the Returns of a Mispriced Short Straddle

Average	−$3,071
Standard deviation	$5,425
Skewness	−4.31
Excess kurtosis	29.01
Median	−$2,143
10th percentile	−$7,069
Minimum	−$94,084
Percent profitable	25%

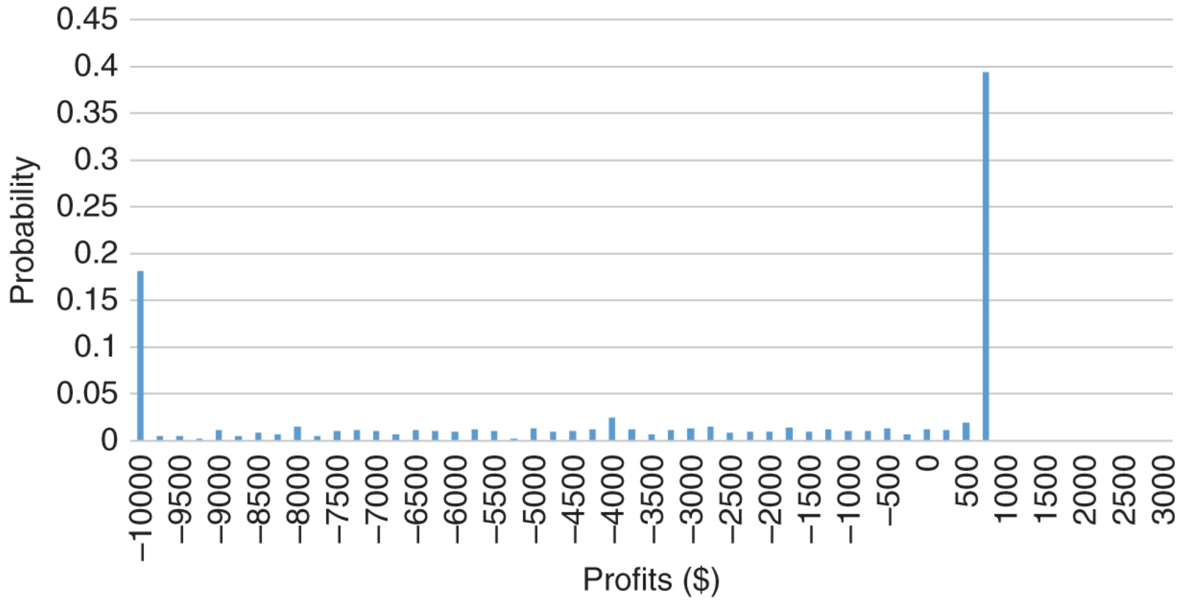


FIGURE 6.4 The returns of the short strangle when our forecast was poor.

TABLE 6.4 Summary Statistics for the Returns of a Poorly Priced Short Strangle

Average	−$3,230
Standard deviation	$9,190
Skewness	−4.2
Excess kurtosis	21.7
Median	−$1,650

10th percentile	−$10,085
Minimum	−$197,526
Percent profitable	36%

Next, we look at the case where our volatility forecast was neutral (i.e., 30%) but there was also an unanticipated 20% drift in the underlying. The returns of the two option structures in this scenario are summarized in Tables 6.5 and 6.6.

Again, the strangle is profitable more often than the straddle, but it can go more badly wrong. The difference between the average profits is largely due to the initial delta of the positions. The straddle had a delta of −12, and the strangle delta was −15. When scaled by the size of each position, this leads to an expected PL difference of $264.

Differences between straddle and strangle results are not dependent on the actual process that generates the underlying returns. Skewed and fat-tailed distributions will create more dramatic results, but the straddle will still have fewer disasters than the strangle.

TABLE 6.5 Summary Statistics for the Returns of a Short Straddle When Our Directional Forecast Was Poor

Average	−$800
Standard deviation	$2,914
Skewness	−1.62
Excess kurtosis	6.14
Median	$12
10th percentile	−$4,712
Minimum	−$28,143
Percent profitable	50%

TABLE 6.6 Summary Statistics for the Returns of a Short Strangle When Our Directional Forecast Was Poor

Average	−$1106
Standard	$3,925

deviation	
Skewness	−4.72
Excess kurtosis	25.74
Median	\$841
10th percentile	−\$6,127
Minimum	−\$38,830
Percent profitable	65%

Instead of showing this by using another postulated distribution, we will look at the returns of the S&P 500 from January 1990 to December 2018. Over this entire period, volatility has been 17.6%, skewness of daily returns has been −0.08, and the excess kurtosis has been 8.9. We sample 252 returns from this population to find the returns of a 1-year straddle and strangle on an imaginary \$100 stock. Options are priced at a 17.6% volatility (this approach ignores autocorrelation in the returns so it isn't exactly what would have happened in the real index). To give each strategy the same vega we sell one straddle and 1.68 strangles. The simulation was run 1,000 times. Results are shown in Figures 6.5 and 6.6.

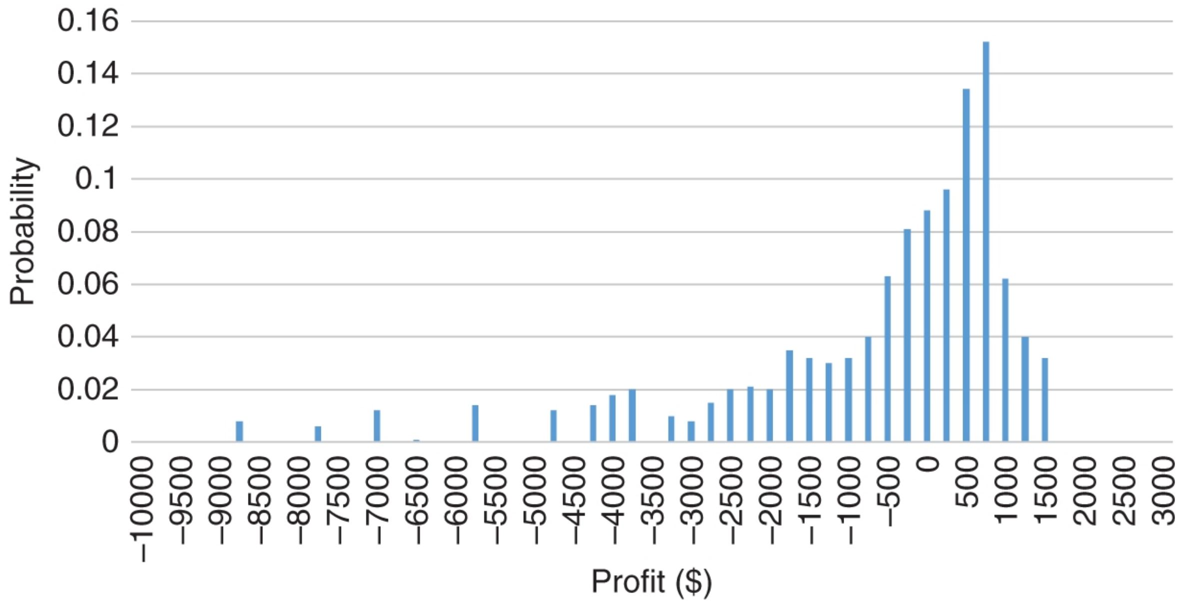


FIGURE 6.5 The returns of the short 100 straddle when the underlying has the S&P 500 return distribution.

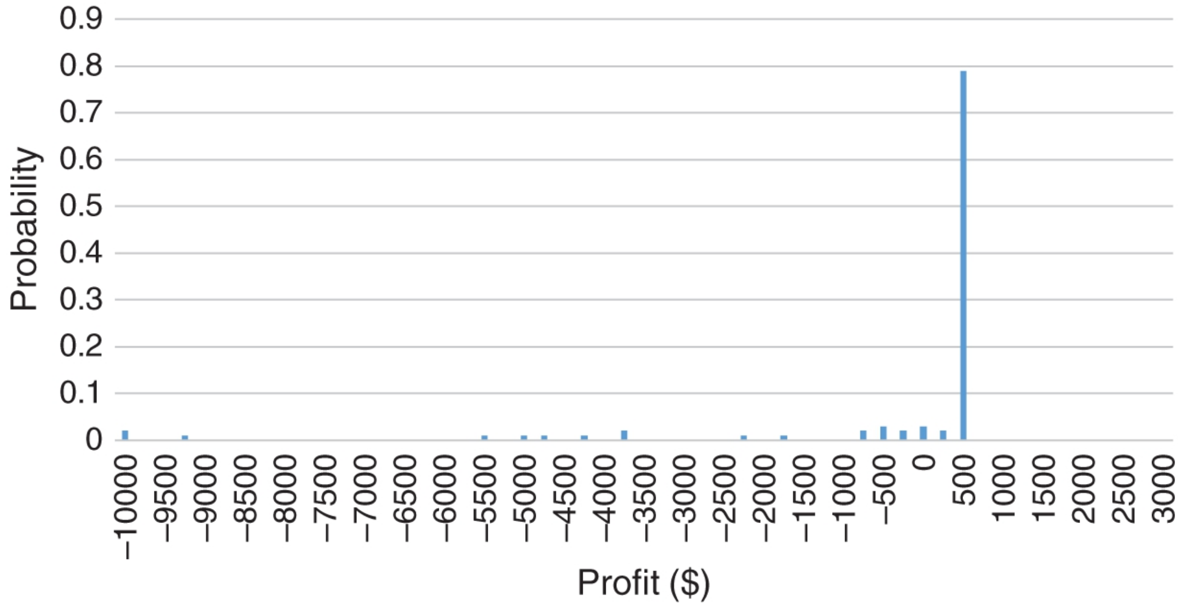


FIGURE 6.6 The returns of the short 85/134 strangle (10-delta call and put) when the underlying has the S&P 500 return distribution.

Again, the straddle has less downside. This is summarized in Table 6.7.

In practice selling a strangle will often collect an implied skewness premium. Because the implied skew overstates the actual skewness of returns, this effect will raise the average profit of the strangle relative to the straddle, but it won't affect the relative risk conclusions.

The strangle's win percentage is a very powerful piece of feedback that can trick us into doing trades like this even when they have negative expectation. The straddle has a better correspondence between correctness of forecast and profits. Hence you will be far less likely to fool yourself into thinking you have a volatility edge than you would with a strangle.

TABLE 6.7 Comparing Results for Straddles and Strangles if the Underlying Has the Same Historical Returns

Result	Straddle	Strangle
Skewness	−1.32	−3.27
Excess kurtosis	1.36	9.93

Result	Straddle	Strangle
Worst case	−$9,428	−$16,522
Worst decile	−$3,356	−$4,211

This is where many people who sell options “for income” go wrong. There is no magic in selling strangles, even if they are struck a long way out of the money. If you don't have an edge in volatility, you will lose eventually.

The straddle has a payoff that is less sensitive either to extreme moves or to making a poor forecast. It won't be profitable as often as the strangle, nor will it practically ever make its theoretical maximum, but it also won't go as badly wrong as a strangle can.

By choosing to sell a strangle instead of a straddle, a trader is gaining an increased median return in exchange for greater extreme risks. It is impossible for someone else to say that a choice like this is wrong as it depends on individual risk preferences, but by most risk metrics the straddle might initially appear to be the riskier position, but it really isn't.

Aside: Delta-Hedged Positions

This book has focused on trading options from the perspective of speculators and end users. These investors tend not to delta hedge much or at all. However, it is worth repeating the straddle and strangle comparison assuming we hedge daily. Both implied and realized volatilities are 30%, so we again expect an average profit of zero. The results of 10,000 GBM simulations are shown in Figures 6.7 and 6.8 and Table 6.8.

It is clear that delta hedging does exactly what it is supposed to do: reduces risk. Extreme results are far more palatable than for unhedged positions. We also see that the differences between straddles and strangles are greatly minimized. The positions are not exactly the same because the strangle will have a more concentrated vega profile (shown in Figure 6.9). This means that when things are going very badly (i.e., the underlying has moved a lot) the straddle risk will start to decrease as we move away from the option strike. This need not happen for strangles. Also note that although the positions were scaled to have the same vega at initiation, if the stock rallies, the strangle picks up more vega. This increases risk during adverse events.

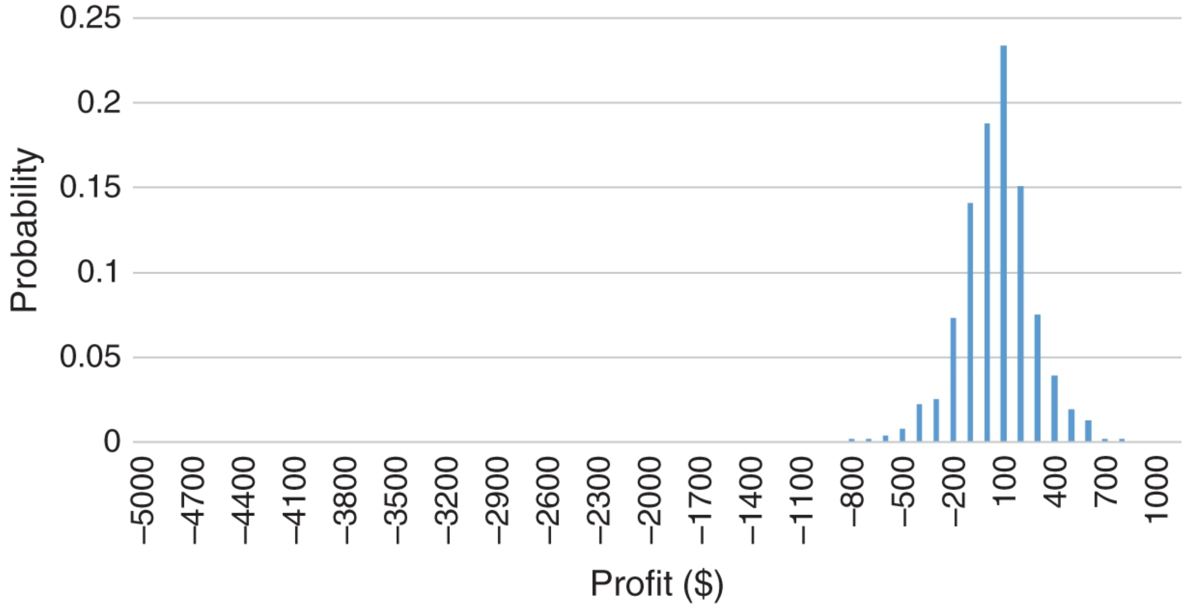


FIGURE 6.7 The returns of the short straddle when hedging daily.

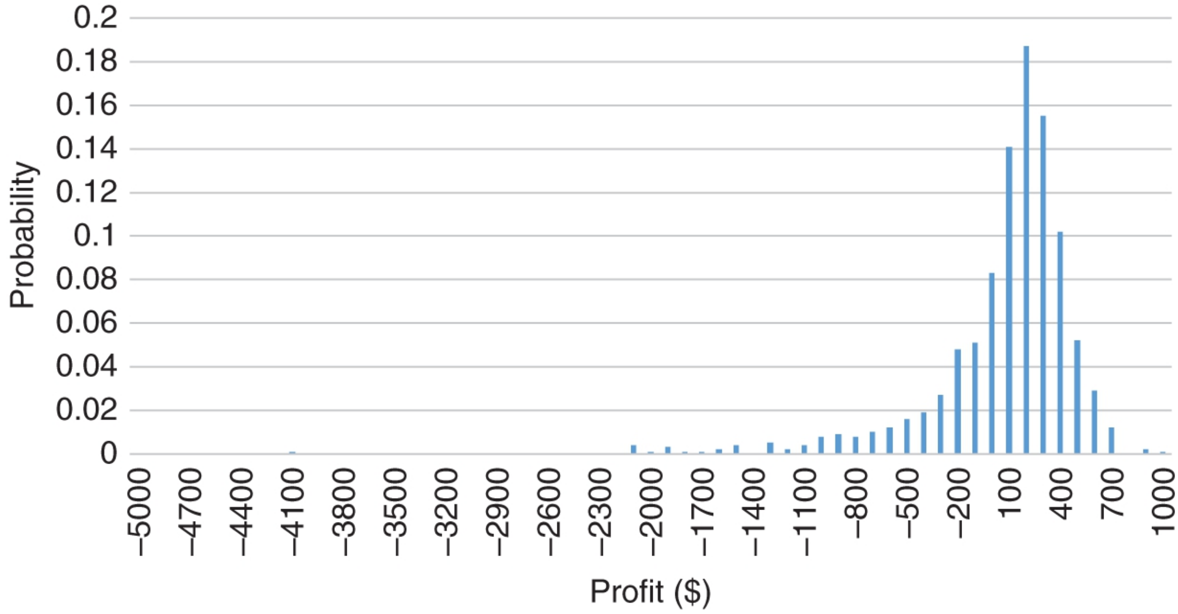


FIGURE 6.8 The returns of the short 70/130 strangle when hedging daily

TABLE 6.8 Comparing Results for Straddles and Strangles When Hedging Daily

Result	Straddle	Strangle

Result	Straddle	Strangle
Average	$7	$22
Median	$12	$126
Skewness	−0.2	−2.71
Excess kurtosis	1.28	13.32
Worst case	−$856	−$2,170
Worst decile	−$242	−$450

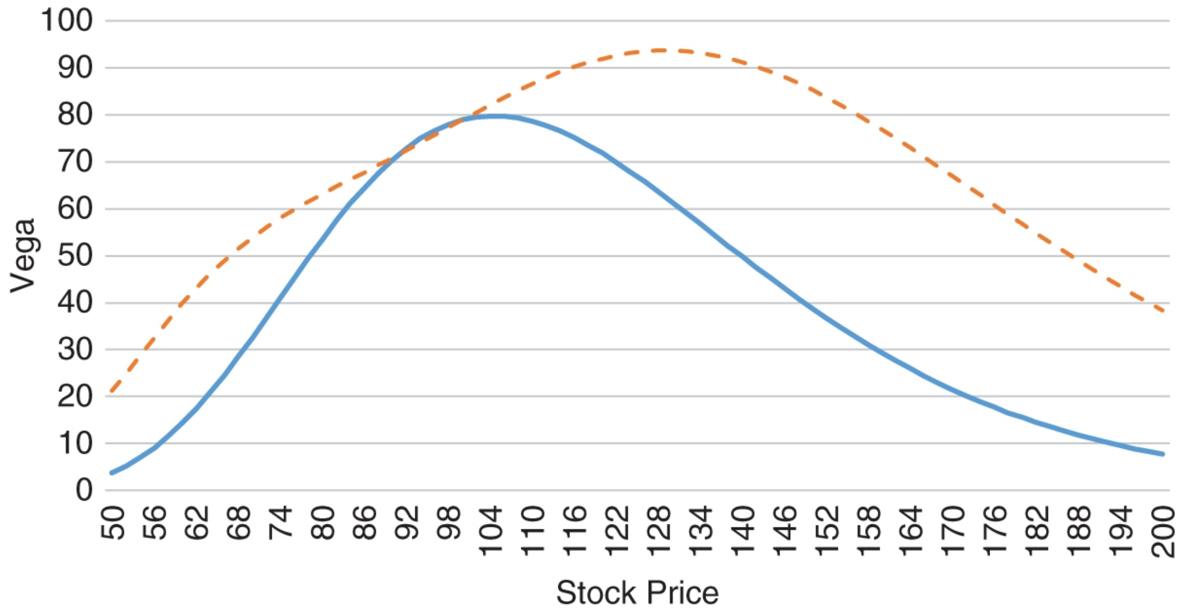


FIGURE 6.9 Vega as a function of underlying price for the straddle (solid line) and 70/130 strangle (dashed line).

Hedging has the effect of reducing the variance of returns, but hedging drastically increases transactions costs and introduces operational issues associated with position monitoring. Most people who are not broker-dealers should probably not be dynamically hedging. For those who are interested, refer to Sinclair (2013) for the theory and practicalities of actively delta hedging.

Butterflies and Condors

These are the more conservative versions of straddles and strangles respectively. In each case, exposure to adverse moves is capped. Technically, these strategies are constructed from either all calls or all puts but in practice traders use the equivalent iron-butterfly and iron-condor structures. These are constructed from out-of-the-money options. Put-call parity means these positions are exactly the same.

We repeat the straddle/strangle analysis to compare butterflies to condors. Results are shown in Figures 6.10 through 6.13 and Tables 6.9 through 6.12. (As before, we weight the positions so they have the same vega as a short straddle.)

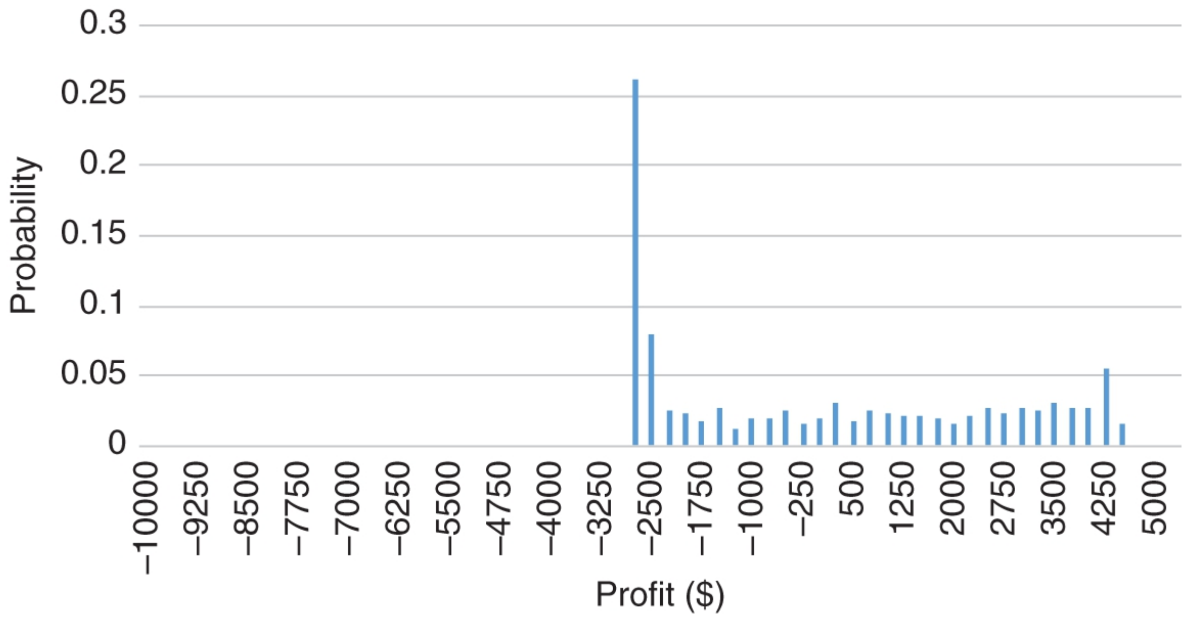


FIGURE 6.10 The profit distribution of the fairly priced butterfly (long the 70/130 strangle and short the 100 straddle).

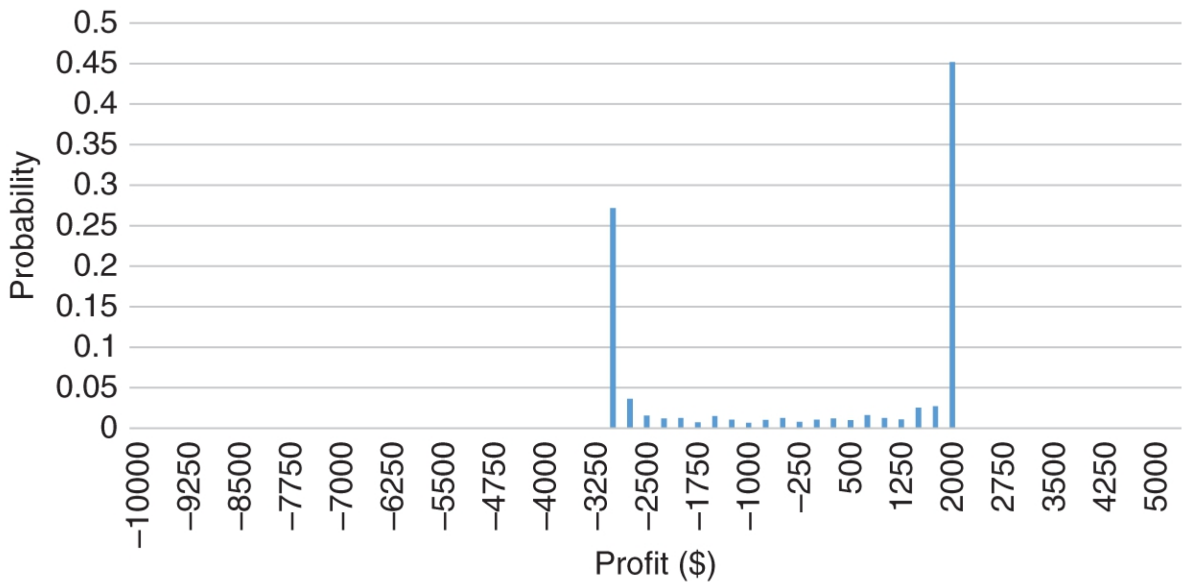


FIGURE 6.11 The profit distribution of the fairly priced condor (long the 70/130 strangle and short the 80/120 strangle).

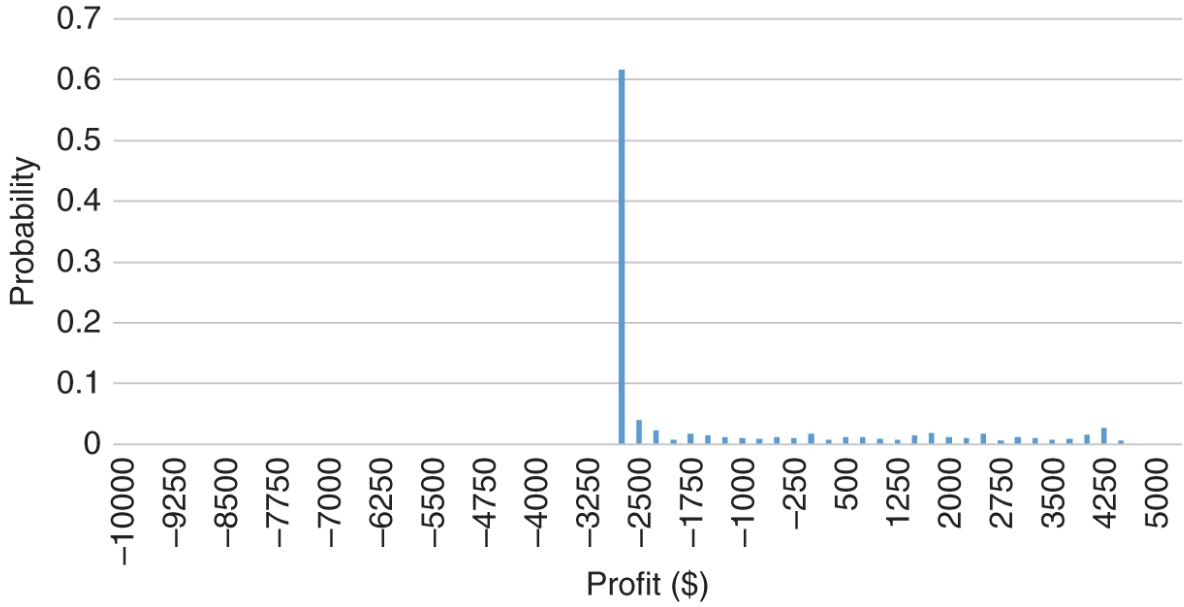


FIGURE 6.12 The profit distribution of the poorly priced butterfly (long the 70/130 strangle and short the 100 straddle).

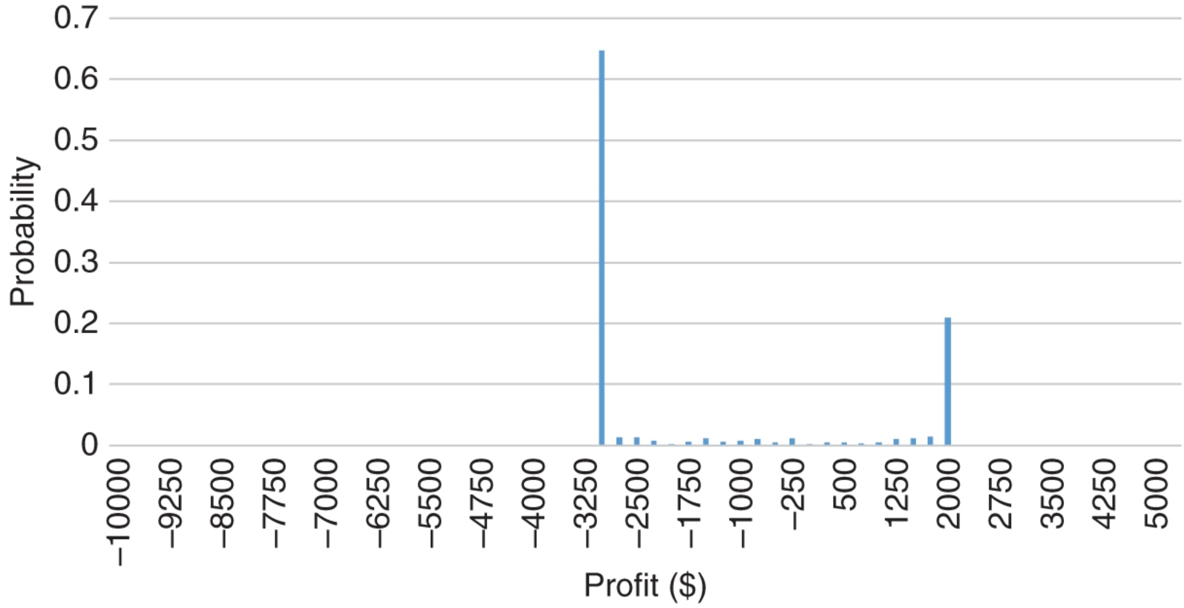


FIGURE 6.13 The profit distribution of the poorly priced condor (long the 70/130 strangle and short the 80/120 strangle).

TABLE 6.9 Summary Statistics for the Returns of a Fairly-Priced Butterfly

Average	−$25
Standard deviation	$2,460
Skewness	0.4

Excess kurtosis	−1.4
Median	−$321
10th percentile	−$2,756
Minimum	−$2,756
Percent profitable	46%

TABLE 6.10 Summary Statistics for the Returns of a Fairly Priced Condor

Average	−$9
Standard deviation	$2,252
Skewness	−0.4
Excess kurtosis	−1.7
Median	$1,524
10th percentile	−$3,032
Minimum	−$3,032
Percent profitable	58%

TABLE 6.11 Summary Statistics for the Returns of a Badly Priced Butterfly

Average	−$1,530
Standard deviation	$2,109
Skewness	1.5
Excess kurtosis	0.8
Median	−$2,756
10th percentile	−$2,756
Minimum	−$2,756
Percent profitable	20%

TABLE 6.12 **Summary Statistics for the Returns of a Badly Priced Condor**

Average	−$1643
Standard deviation	$2,109
Skewness	1.0
Excess kurtosis	−0.9
Median	−$3,032
10th percentile	−$3,032
Minimum	−$3,032
Percent profitable	26%

Using a butterfly tames the extremely bad results of a straddle but this comes at the expense of incurring the maximum possible loss 26% of the time.

Because a condor has less vega than a butterfly, we need to trade more of them to get the same volatility exposure. This means that the worst case is slightly worse than with the butterfly. And the condor realizes this worst case 30% of the time. Also, the maximum possible profit is lower than the butterfly's.

Now we look at the results where we sell implied volatility at 30% and subsequent realized volatility is 70%. The returns of the butterfly and condor in this scenario are shown in Figures 6.12 and 6.13, and statistics are summarized in Tables 6.11 and 6.12:

We can see that the condor has a higher winning percentage, but its wins are capped at the premium of $1,998. This occurred 15% of the time. By having a higher initial premium the butterfly can have larger wins. It wins 12% of the time more than $1,998. The average of these "condor-beating" wins is $3,342.

In terms of risk and reward, the butterfly is to the condor what the straddle is to the strangle: lower winning percentage but higher upside and lower downside.

Aside: Broken Wing Butterflies and Condors

A broken wing fly or condor is one with only one long strike. For example, a broken wing (iron) butterfly might be long an out-of-the-money put and short a straddle. This has the same payoff as a one- by two-call spread, and we will look at the risks of these in more detail in Chapter Eight. But for now, I want to emphasize that a common reason for implementing these strategies is wrong.

It is generally accepted that stock market down moves are more severe than up moves. So, many traders are only concerned with hedging short exposure on the downside. Be careful with this. There isn't much of a difference between daily up moves and down moves. For example, consider daily S&P 500 returns from 1990 through to the end of 2018. Summary statistics of absolute up and down returns are shown in Table 6.13. None of the differences are significant at the 5% level.

The difference between the "speed" of breaks and rallies is real, but it is due to correlation effects and is also far smaller than many traders think. It is well known that before 1987, the index option markets had almost no implied volatility skew. After the crash the markets priced in the crash risk. But markets also massively overcompensated. This is both why put options are usually the most overpriced and why, if a trader wants to hedge extreme moves, it is generally best to hedge both extreme drops and extreme rallies.

TABLE 6.13 Summary Statistics of S&P 500 Returns from 1990 to 2018

Statistic	Positive Daily Returns	Negative (Absolute) Daily Returns
Average	0.0073	0.0076
Median	0.0051	0.0049
90th percentile	0.0156	0.0177
99th percentile	0.0387	0.0384
Maximum	0.1158	0.0903

Calendar Spread

The calendar spread is used to speculate on relative implied volatility levels. Specifically, we are trying to capture the greater variance premium in the shorter-dated options. For example, we buy the 60-day ATM straddle at an implied volatility of 30% and sell the 30-day ATM straddle at an implied volatility of 40%. We hold the position until the front-month expires, and over this time realized volatility is 30%. Because the edge in this trade comes from the overpriced short-term implied volatility, we also simulate a short 1-month straddle position, which is the most direct way to capture this variance premium.

Summary statistics of the PL distributions from simulations of 10,000 paths are shown in Table 6.14.

The calendar spread has lower dispersion around the average P/L (which is the same for both positions because it is entirely due to the mispricing of the 1-month implied volatility). At an intuitive level, this is because, at the expiration of the front-month options, the spread's payoff diagram is very similar to the value of a butterfly. This is shown in Figure 6.14, which shows the P/L of the spread.

TABLE 6.14 Summary Statistics of the PL Distribution for the Straddle Spread and the Short Front-Month Straddle

Statistic	Short Straddle	Straddle Spread
Average	$261	$265
Standard deviation	$597	$210
Skewness	−1.4	0.0
Excess kurtosis	2.7	−1.4
Median	$411	$260
90th percentile	$873	$552
Maximum	$899	$578
10th percentile	−$574	−$18
Minimum	−$3,753	−$66
Percent profitable	72%	87%

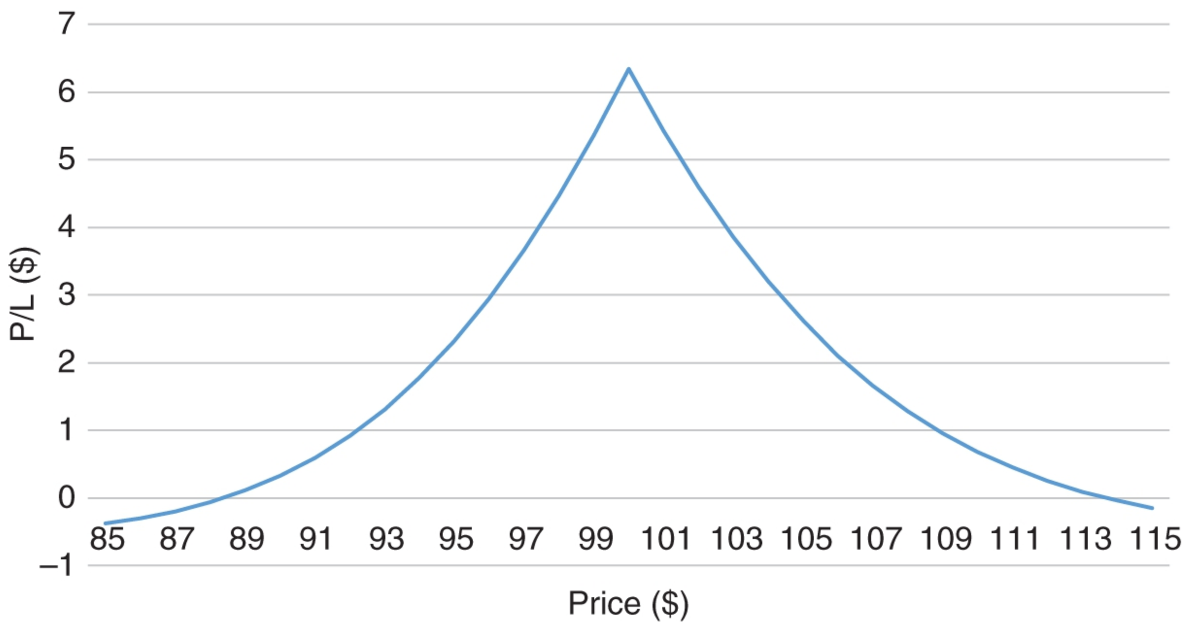


FIGURE 6.14 The P/L of the straddle spread at expiry of the front-month options.

The cost of this risk reduction is that the spread is long vega. If implied volatility drops, the spread will underperform the short straddle. This risk can also be at odds with our volatility forecast. We are projecting that the front-month volatility is too high. This will usually also imply that the second-month volatility is also overpriced (although not by as much). So, the straddle spread is a long vega position that we put on when we are hoping volatility is overpriced.

The spread still acts as a diversifier, but we also need to be correct on our view of overall, rather than relative volatility. We demonstrate this problem by simulating a situation in which the realized volatility is 20%. Summary statistics of the spread and the short straddle are shown in Table 6.15.

We can reduce this effect by weighting the spread components, so that the overall position is flat vega. This will also reduce the variance reducing properties of the spread.

Another weighting scheme is to calculate a volatility beta that relates the change in the front-month volatility to that of the second-month volatility.

The market regime in which the benefits of the calendar spread are most obvious is when there is a large variance premium and a low implied volatility.

As always, the trader's decision will depend on exactly what risk he is most concerned about.

TABLE 6.15 Summary Statistics of the PL Distribution for the Straddle Spread and the Short Front-Month Straddle (Front-month implied volatility was 40% and second-month implied volatility was 30%. Realized volatility was 20%.)

Statistic	Short Straddle	Straddle Spread
Average	\$522	\$223
Standard deviation	\$375	\$147
Skewness	−1.2	−1.0
Excess kurtosis	1.0	0.4
Median	\$630	\$266
90th percentile	\$877	\$375
Maximum	\$900	\$380
10th percentile	−\$16	\$10
Minimum	−\$2,141	−\$370
Percent profitable	89%	91%

Including Implied Volatility Skew

The preceding analysis assumed that all strikes had the same implied volatility. Of course, this is generally not the case. Usually, the OTM puts will trade at a volatility premium to the ATM options. Because ATM option-implied volatility is the most predictive of future realized volatility (Bondarenko, 2003; Poon and Granger, 2003, and further references from Chapter Four), by selling OTM puts we can collect a volatility premium. This makes selling strangles relatively more attractive than would be the case in a world where all strikes had the same implied volatility.

The shape of the implied volatility curve is fairly persistent. A given delta option's implied volatility will be a constant times the ATM implied volatility. For example, in the SPX the 10 delta put will have an implied volatility about 1.45 times the ATM volatility (Sinclair, 2013). Similarly, the 25 delta call will usually have a volatility of about 0.88 of the ATM volatility. Using this rule of thumb we can surmise that the 70 strike put (which has a 9 delta) will trade at an implied volatility of 43.5% if the ATM volatility is 30%. Similarly the 130 call (23 delta) will be expected to have an implied volatility of 0.264.

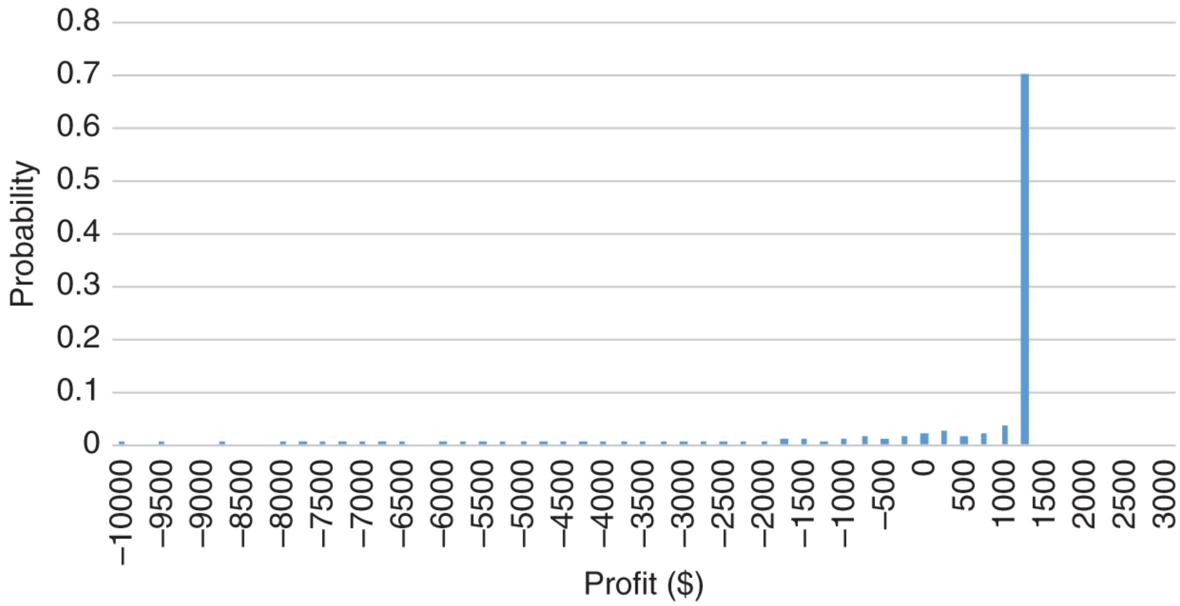


FIGURE 6.15 The profit distribution of a strangle with an implied volatility skew and a fair value for realized volatility.

Again assume that future realized volatilities are 30%. If we run the same simulations as previously but price the strangle with more realistic implied volatilities of 43.5% for the put and 26.4% for the call, we get the results shown in Figure 6.15 and Table 6.16.

These numbers are better than the strangle when all strikes are priced at a 30% volatility (the results for both situations are shown in Table 6.17 for ease of comparison).

TABLE 6.16 Summary Statistics for the Returns of a Strangle with an Implied Volatility Skew and a Fair Value for Realized Volatility

Average	$278
Standard deviation	$2,102
Skewness	−3.8
Excess kurtosis	18.0
Median	$1,123
10th percentile	−$1,765

Minimum	–$37,372
Percent profitable	79%

TABLE 6.17 The Results for Both the Flat Skew Condor and the Skewed Case

Statistic	**Strangle with Constant Strike Volatility**	**Strangle with Implied Volatility Skew**
Average	−$6.12	$278
Standard deviation	$2,040	$2,102
Skewness	−4.8	−3.8
Excess kurtosis	24.2	18.0
Median	$841	$1,123
10th percentile	−$1,994	−$1,765
Minimum	−$24,683	−$37,372
Percent profitable	78%	79%

Strike Choice

Although straddles are almost always struck at the ATM forward price, the other combinations require us to choose strikes.

To do this, start with the short strikes. For straddles and butterflies this will be the ATM strike but for strangles and condors we want to choose strikes that maximize the volatility premium due to the implied volatility curve.

The part of the implied volatility curve that is most predictive of the future realized volatility is the ATM, so when we are looking for a volatility edge we should sell a strike with the highest volatility over this value. In pure volatility terms that will almost always be the farthest out-of-the-money put strike. Consider the SPY puts in Table 6.18.

The highest implied volatility is in the 210 strike. So let's imagine we sell this option and also sell the 360 call (at an implied volatility of 11.0%) to create a delta-neutral strangle. We assume that realized volatility was what the ATM predicted (14.2%). We use Monte Carlo to simulate 10,000 instances of this strategy and get the results shown in Table 6.19. We sell $1,000 of vega.

However, this combination has very little vega per option. Contrast this with selling the put strike that has the greatest dollar premium over what the option would be worth if it were priced at the ATM implied volatility. Table 6.20 shows this choice.

TABLE 6.18 The Put Prices of the SPY June 2020 Expiration on July 30, 2019 (Down to the 5 Delta Strike) (SPY was 300.48.)

Strike	Market Price	Implied Volatility
210	1.68	0.253
215	1.95	0.248
220	2.25	0.243
225	2.58	0.237
230	2.95	0.232
235	3.35	0.226
240	3.82	0.221
245	4.33	0.215
250	4.91	0.209
255	5.55	0.203
260	6.28	0.197
265	7.07	0.192
270	7.96	0.186
275	8.95	0.179
280	10.03	0.173
285	11.26	0.167
290	12.57	0.160
295	14.10	0.153
300	15.78	0.146

TABLE 6.19 The Summary Statistics from Selling $1000 Vega of the 210/360 SPY Strangle

Statistic	Maximum Implied Volatility Strangle
Average	$6,600
Standard deviation	$105,400
Skewness	−5.5
Excess kurtosis	32.1
Median	$29,400
10th percentile	$28,670
Minimum	−$1,463,400
Percent profitable	92%

TABLE 6.20 The Dollar Premium of Options Over Their Being Priced at the ATM Volatility (14.0% in this Instance)

Strike	Market Price	Priced with ATM Volatility	Premium Due to Volatility Skew
210	1.68	0.03	1.65
215	1.95	0.06	1.89
220	2.25	0.10	2.15
225	2.58	0.17	2.41
230	2.95	0.28	2.67
235	3.35	0.43	2.92
240	3.82	0.64	3.18
245	4.33	0.94	3.39
250	4.91	1.34	3.57
255	5.55	1.86	3.69
260	6.28	2.52	3.76
265	7.07	3.36	3.71
270	7.96	4.37	3.59
275	8.95	5.60	3.35
280	10.03	7.05	2.98
285	11.26	8.73	2.53
290	12.57	10.66	1.91
295	14.10	12.84	1.26

Strike	Market Price	Priced with ATM Volatility	Premium Due to Volatility Skew
300	15.78	15.26	0.52

The greatest dollar premium is in the 260 strike. The market has them worth 6.28, but if they were priced at the ATM volatility, they would only be worth 2.52. Assume we sell 1.2 of the 260/335 strangles so we have the same vega exposure as in the previous simulation. Results are summarized in Table 6.21.

The statistics that describe typical results (average, median, percent profitable, and 10th percentile) all favor selling the highest volatility premium. That is to be expected. The edge in selling options comes from selling expensive implied volatility. However, we can also see clearly that when things go wrong, selling teeny options is far more painful than selling those with more premium.

In summary, when selling strangles or condors we want to take advantage of the skew premium, but if we do this by selling the option with the highest implied volatility, we expose ourselves to horrendous risk. Once we have found our short put strike, we choose the other call strike to give the delta or vega profile we want.

TABLE 6.21 The Summary Statistics from Selling $1000 Vega of 260/335 SPY Strangles

Statistic	Maximum Implied Volatility Strangle
Average	$3,980
Standard deviation	$32,430
Skewness	−3.4
Excess kurtosis	15.1
Median	$20,092
10th percentile	−$37,100
Minimum	−$306,928
Percent profitable	76%

Choosing a Hedging Strike

The only reason we choose a short-volatility strategy is that we think implied volatility is too high. Consistent with this, any option we buy will be one that we think is overpriced. This is true even if there is no implied volatility skew. In the presence of a skew we will be overpaying by even more. The long-hedging options have negative expected value. They are a cost of doing business. We want to choose the ones that give us our desired amount of protection for the smallest amount of money.

As an example, consider the case in which we sell a 260 put for 6.28 and we don't want a possible loss of more than $5,000. Our possible hedging options are shown in Table 6.22. The process is simple. We just find the hedging strike that gives the loss level we are comfortable with for the lowest cost.

Maximum Loss = Profit of Hedge + Loss of Short Put – Hedge Premium + Short Put Premium

So here, we would buy the 205 put for a hedge cost of $145.

Ignoring option premia, the worst P/L will occur right at the long strike. Here we will have lost money on our short and have got nothing from the long. Below the strike, the long put and short put will cancel completely. Because our loss limit is $5,000 this would put the lower put strike at 210. However, total premia received will be $628 hedge premium, so the strike will be lowered by this amount (divided by 100).

TABLE 6.22 Prices and Strikes of Possible Hedging Options for Our Short 260 Put Position

Strike	Market Price	Max Loss of Hedged Portfolio
180	0.64	$7,436
185	0.74	$6,946
190	0.88	$6,460
195	1.05	$5,977
200	1.23	$5,495
205	1.45	$5,017
210	1.68	$4,540
215	1.95	$4,067
220	2.25	$3,597

It can be tempting to buy lower premia, shorter-dated options as hedges. This is almost always a bad idea. Generally, the consecutively purchased short-dated options will have a higher total premium than the single longer-dated option (a consequence of total variance scaling with the square root of time). If we also consider the extra transaction costs, the single option becomes even more attractive.

We can illustrate this general rule and also show how exceptions to it can occur by using the Brenner and Subrahmanyam (1988) approximation. They show that for ATM options the BSM equation is approximated by

$$P \approx 0.4S\sigma\sqrt{T} \tag{6.2}$$

Take this to be our hedging benchmark and compare it to the alternative of buying two options consecutively. In that case, the hedging premium would be

$$2 \times 0.4S\sigma\sqrt{\frac{T}{2}} = P \times \frac{2}{\sqrt{2}} > P \tag{6.3}$$

The two-option hedge is more expensive if the volatility term structure is flat. However, if the volatility term structure is steep enough, it could be worth rolling the shorter term hedges. Specifically if the volatilities for each subperiod, σ_1 and σ_2, and the total volatility, σ, are related by

$$\frac{\sigma_1 + \sigma_2}{\sqrt{2}} < \sigma \tag{6.4}$$

then it is worthwhile rolling shorter-dated hedges.

As an example, consider the situation in which we can either buy a 2-year option or a 1-year option and then another 1-year option when the first one expires. The stock is $100, the strike is 100, volatility is 30%, and rates are zero. The 2-year put is worth 16.8. Each single 1-year put is worth 11.9. So here the 2-year option is the cheaper alternative.

But consider the case where the 2-year option has an implied volatility of 40% and the 1-year option has an implied volatility of 20%. Also assume that this term structure is constant in time, so when the first 1-year option expires we can buy the next one at 20% as well. Now each 1-year option is worth 8.0 and the 2-year option is worth 22.3. Here, we are better off buying a 1-year option and rolling it later.

Expiration Choice

As I've emphasized throughout this chapter, there is more to choosing an option to trade than simply finding the one that has the highest expected return. Many other metrics are also relevant. Median return, drawdown exposure, and percent of trades that are profitable are also statistics to consider. However, it also seems reasonable that our search can start with expected value and expand from there. The purpose of trading is to make money. Risk management should aim to protect expected value rather than minimize risk. The safest position is no position. That will also make no money.

The profitability of option trading is driven by the variance premium. Many other effects are important but the variance premium will always be dominant. It is to options what evolution is to biology or what gravity is to physics. So when choosing an expiration our prime consideration is which one has the most variance premium. If we are selling, we want the expiration with the highest premium. If we are buying, we want the expiration with the lowest premium.

Israelov and Tummala (2017) studied this problem and wrote a paper whose title is the perfect statement of our issue: “Which Index Options Should You Sell?” By looking at S&P 500 option performances from 1996 to 2015, they showed that short-dated options had the highest variance premia. Their explanation for this was this:

> Option buyers seek to purchase insurance for their portfolio and are typically concerned about monthly or quarterly returns.… It is intuitive that the options which most directly match these preferences are the most attractively compensated for option sellers. (p. 14)

This is quite possibly true, but I think the more important reason is compensation for risk. Short-dated option risk is dominated by gamma, and long-dated option risk is due to vega. The old story is "vega wounds but gamma kills." The sellers of short-dated options are taking the most risk and they should be most compensated. A good rule when looking for a variance premium is to look for situations with the most risk. The variance premium is (a mispriced) compensation for risk. The higher the risk, the higher the mispricing.

This hypothesis is consistent with the results of Tosi and Ziegler (2017). Using S&P 500 option data from 1996 to 2015 they showed that the returns from shorting out-of-the-money put options were concentrated in the few days preceding their expiration. Back-month options generated almost no returns.

Their proffered reason for this was:

> The concentration of the option premium at the end of the cycle reflects changes in options' risk characteristics. Specifically, options' convexity risk increases sharply close to maturity, making them more sensitive to jumps in the underlying price. By contrast, volatility risk plays a smaller role close to maturity. (abstract)

And conclude:

> Our results imply that speculators wishing to harvest the put option premium should short front-month options only during the last days of the cycle, while investors wishing to protect against downside risk should use back-month options to reduce hedging costs. (abstract)

Other studies that reach the same conclusion are by Andries et al. (2015), Dew-Becker et al. (2014), and van Binsbergen and Koijen (2015).

Conclusion

There is no "best" strategy. The choice of what to select is a matter of personal risk preferences. Strangles win more often than straddles but have less upside and more downside. Butterflies and condors are more expensive than straddles and strangles in terms of transaction costs. They will also realize their maximum possible loss a significant amount of the time.

When choosing a short strike, the trader needs to balance receiving the most edge by selling the options with the highest implied volatility and the amount of risk that this produces. Similarly, shorter-dated options will have more variance premium than longer-dated ones but they also have more potential for catastrophe.

Summary

- Selling OTM option structures (strangles or condors) will give higher median returns and a higher win percentage but this can make it more difficult to distinguish between good trades with expected value and good luck.
- The highest volatility premium is in short-dated options. Long-dated options have very little volatility premium.
- The highest volatility premium is in far-out-of-the-money puts. Out-of-the-money calls have very little volatility premium.

CHAPTER 7
Directional Option Trading

The genius of the BSM model is the idea that the direction of the underlying doesn't matter when pricing an option. But although this methodology leads to an arbitrage-free replication value, it is still possible to trade options to make bets on the underlying direction. If one believes the story of Thales and the olive presses, this was the original point of options. Even now most traders use options directionally. Indeed, many retail traders can only buy options, and directional trading is essentially their only available tactic.

In this chapter, I will discuss directional option speculation, starting with the theory of pricing with a directional view and then discussing the choice of strikes, structures, and expirations. All examples will be given in terms of long calls, assuming a bullish bias, but the ideas are trivially generalizable to both puts and short option positions. I'm also going to assume each option is on one share.

Subjective Option Pricing

Options offer many advantages over trading the underlying. The main advantage is the ability to speculate on a more nuanced view than just "up or down." Also, leverage and the possibility of highly skewed payoffs can be useful. However, these all add considerable complexity.

Here I'm going to (possibly optimistically) assume that the trader has a valid prediction method for the underlying and show how she should monetize this view. It is never easy to predict the direction of the underlying. But when trading options, it is easy to be right in your prediction of the underlying and still lose money. It is never good to solve the hardest part of a problem and still fail.

The simplest directional option trading strategy is to buy a call if you think the underlying will expire above the strike by more than the option premium. This can be kindly characterized as "model-free directional trading" and more realistically as "guessing."

Consider this set of call prices on a $100 stock:

Strike	Price
95	8.7
100	5.6
105	3.2

Strike	Price
110	1.5

If we know the stock will expire at $120, our strike choice is trivial. Investing $100 in each option would give the following profits:

Strike	Profit
95	$187
100	$257
105	$369
110	$567

But the problems with trading are never about optimizing results when predictions are correct. The real issue is how to control risk when we are wrong. If the stock only goes to $106, our profits will be completely different:

Strike	Profit
95	$26
100	$7
105	−$68
110	−$100

These extreme differences illustrate the need for a better plan.

A Theory of Subjective Option Pricing

The dynamic hedging strategy used in the BSM model removes the need to use a drift parameter. But that isn't to say that we *can't* include drift in our personal pricing. We won't be in the risk-neutral paradigm anymore and our theoretical values will disagree with market prices, but that is to be expected. If we agreed with market prices, we wouldn't be speculating.

BSM showed that the rate of return of the underlying is not relevant for *pricing* options. But the underlying return obviously does affect the *return* of the unhedged option.

If we have a valuation model that explicitly includes drift, we can use it to compare the theoretical values to market prices and find the most attractive opportunities.

Luckily, several pre-BSM pricing models did include the return on the stock. Boness (1962, 1964) found an option pricing model that is functionally the same as BSM but is based in the real, rather than the risk-neutral, world. It isn't arbitrage free, but it answers the most important question a directional speculator has: If she thinks the stock is going up, what option should she buy?

Instead of calculating the call value by taking the expectation of the payoff in the risk-neutral world, Boness's result is the expectation in the physical world.

With the normal notation:

$$C = E_p[max(0, S - X)] \tag{7.1}$$

$$= Sexp(\mu T)N(d_3) - XN(d_4) \tag{7.2}$$

where

$$d_3 = \frac{ln\left(\frac{S}{X}\right) + \left(\mu + \frac{\sigma^2}{2}T\right)}{\sigma\sqrt{T}} \tag{7.3}$$

$$d_4 = \frac{ln\left(\frac{S}{X}\right) + \left(\mu - \frac{\sigma^2}{2}T\right)}{\sigma\sqrt{T}} \tag{7.4}$$

To arrive at this formula, Boness had to make an assumption about returns.

He needed to say that the stock return is the rate used to discount the strike in the put-call relationship, that is, all cash flows would be invested in the stock.

In the normal derivation of put-call parity, we form a portfolio that is long a put, short a call, and long a share. At expiration this portfolio is worth the strike price, which means the portfolio is currently worth the discounted strike value. Since the middle ages (where the idea was used as the basis for mortgage lending) it has been known that the correct discount factor is the interest rate. Drift is irrelevant. If anything other than the interest rate is used as a discount factor, an arbitrage opportunity exists.

Even if we accept that drift is a real phenomenon, it is also reasonable to include an interest rate as an alternative investment opportunity. The stock appreciates at μ and cash is stored at r. It might appear that no investor would operate like this. If $\mu > r$, why would he not invest all money in the stock and ignore the interest rate completely (Boness's model does this)? In reality, people generally do split investments between assets with different returns and risks. The stock has a higher return but also higher risk, which is reflected in the volatility parameter.

It is possible to construct a pricing model that does this with a formal argument from a modified BSM PDE. But this isn't necessary. Our model needs to assume cash is invested at a risk-free interest rate and the forward price of the stock is driven by a (physical world) drift.

If we slightly reinterpret some parameters, this model already exists: the generalized BSM prices (European) options when the underlying pays a continuous dividend yield. We use this, and the interest rate, to price options of the forward value of the stock while also assuming that cash flows are discounted at the risk-free rate. Traders use this model now and interpret the dividend yield and interest rate to be (imperfectly known) objective variables. (Note that in practice different traders will have different dividend estimates and marginal rates. This could, and in rare cases does, allow arbitrage.)

We turn this into a model that incorporates drift by reinterpreting the dividend yield as a subjective drift estimate. We use our drift estimate to give a subjective estimate of the underlying's forward price. This model will also allow arbitrage, but it will be consistent with our opinions of the real world.

The prices of calls in this model are

$$C = Sexp(\mu T)N(d_3) - exp(-rT)XN(d_4) \tag{7.5}$$

where

$$d_3 = \frac{ln\left(\frac{S}{X}\right) + \left(r + \mu + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \tag{7.6}$$

$$d_4 = \frac{ln\left(\frac{S}{X}\right) + \left(r + \mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \tag{7.7}$$

As with any other pricing variable or parameter, it helps to have a greek to measure the impact of an incorrect estimate. The partial derivative of the subjective option price with respect to the drift is given by

$$\frac{\partial C}{\partial \mu} = TSexp(\mu T)N(d_3) \tag{7.8}$$

Now we can derive subjective values for options and see which are most mispriced. In this example, we consider 1-year options on a $100 stock, with a volatility of 30%, a drift of 10%, and zero interest rates. The risk-neutral BSM prices and the subjective prices are shown in Table 7.1.

As the drift is primarily going to be a delta effect, it should be no surprise that the greatest absolute difference in values is in the lower strikes. If an investor wants to buy a fixed number of options, these might be the best choice, but in this case, it is probably better to just buy the underlying because the discrepancy would be greatest there. The largest percentage edge ("bang for the buck") is in the highest strikes, so these appear best for a trader who wants to invest a certain dollar amount.

TABLE 7.1 A Comparison of Risk-Neutral and Subjective Option Prices

Call Strike	BSM Price	Subjective Price	Difference in $ from BSM Value	Difference as % of BSM Value
80	22.53	32.53	8.99	0.38
85	20.09	28.55	8.46	0.42
90	17.01	24.88	7.86	0.46
95	14.29	21.52	7.23	0.51
100	11.92	18.49	6.57	0.55
105	9.88	15.79	5.91	0.60
110	8.14	13.41	5.27	0.65
115	6.67	11.32	4.65	0.70
120	5.44	9.51	4.07	0.75

However, this analysis doesn't consider the different risk characteristics of options with different strikes.

Now that we have a theoretical model, we can talk about the distribution of subjective option returns (assuming our drift and volatility estimates are correct).

Distribution of Option Returns: Summary Statistics

$$\text{Average Profit of a Call} = \textit{Subjective Value} - \textit{Cost} \tag{7.9}$$

$$\text{Chance of Expiring in the Money} = N(d_4) \tag{7.10}$$

$N(d_4)$ is the true probability of expiring in the money, as opposed to the incorrect but frequently stated number, $N(d_2)$, which is the risk-neutral probability. And the drift is a large determinant of whether an option expires in the money. As an example of the size of the difference, consider a 1-year 110 strike call on a $100 stock, with volatility of 30%, a drift of 10%, and an interest rate of zero. The risk-neutral probability of exercise is 32%, whereas the subjective probability is 45%. At very short timescales the volatility will overwhelm the drift, but in general it is bad to assume that the risk-neutral probabilities are indicative of anything in the real world. Figures 7.1 and 7.2 show how $N(d_4)$ depends on the drift and how it varies with strike.

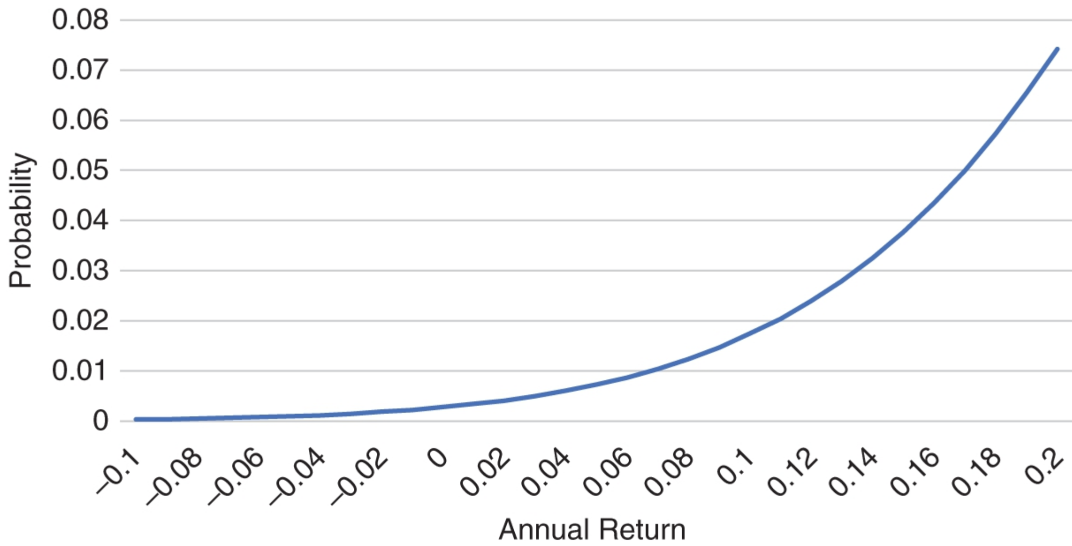


FIGURE 7.1 Probability of the 3-month 150 strike call expiring in the money. Stock is $100, volatility is 30%, and rates are zero.

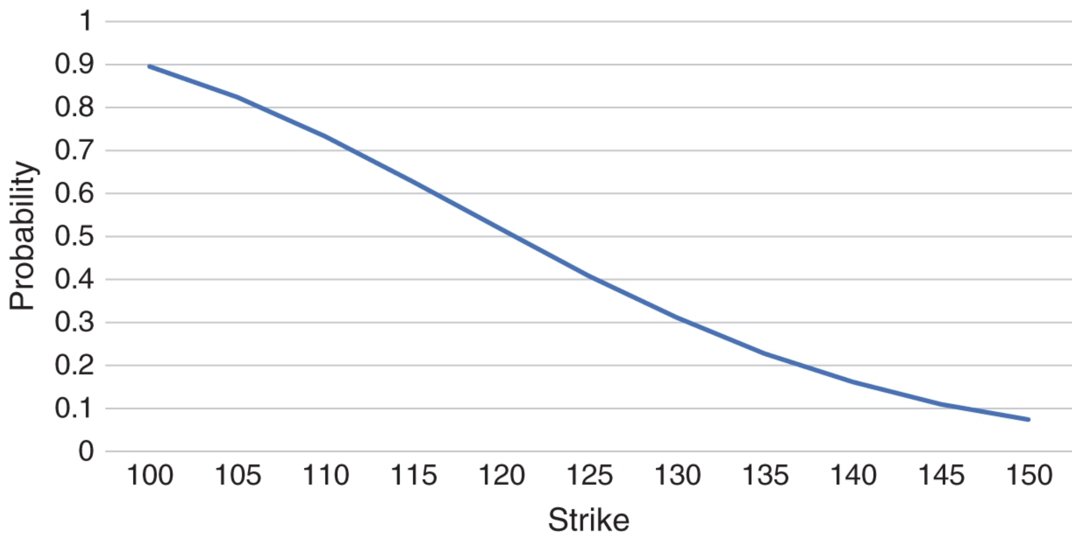


FIGURE 7.2 Probability of the 3-month calls expiring in the money when the return is 20%. Stock is $100, volatility is 30%, and rates are zero.

Chance of Profit = $N(d_5)$

$$d_5 = \frac{ln\left(\frac{S}{X+cost}\right) + \left(r + \mu - \frac{\sigma^2}{2}T\right)}{\sigma\sqrt{T}} \tag{7.11}$$

Also, by assuming the underlying price is lognormally distributed, we can easily calculate the median intrinsic value and hence profit.

$$\text{Median Final Underlying Price} = S_0 exp(\mu T) \tag{7.12}$$

$$\text{Median Option Profit} = max(S_0 exp(\mu T) - X, 0) - Cost \tag{7.13}$$

Further,

$$\text{Percentile, } p\text{, of Final Underlying Price} = S_0 exp\left(\mu T + \sigma\sqrt{T}N^{-1}(p)\right) \tag{7.14}$$

So, Percentile of Option Profit

$$= max\left(S_0 exp\left(\mu T + \sigma\sqrt{T}N^{-1}(p)\right) - X, 0\right) - Cost \tag{7.15}$$

An example is shown in Figure 7.3.

It is possible to calculate the moments of the options returns (Ben-Meir and Schiff, 2012; Boyer and Vorkink, 2014; Sinclair and Brooks, 2017), but the equations are complicated and give no real insight. The most important fact is that options have significantly positive skew. It is also important to note that this extreme skewness occurs even when the underlying has normally distributed returns. This skewness is intrinsic to options and is not inherited from skewness of the underlying. If the underlying has non-normal returns, the effects on the options will be magnified. In this case, there won't be analytic expressions for the option moments and simulations will be necessary to understand the option return moments.

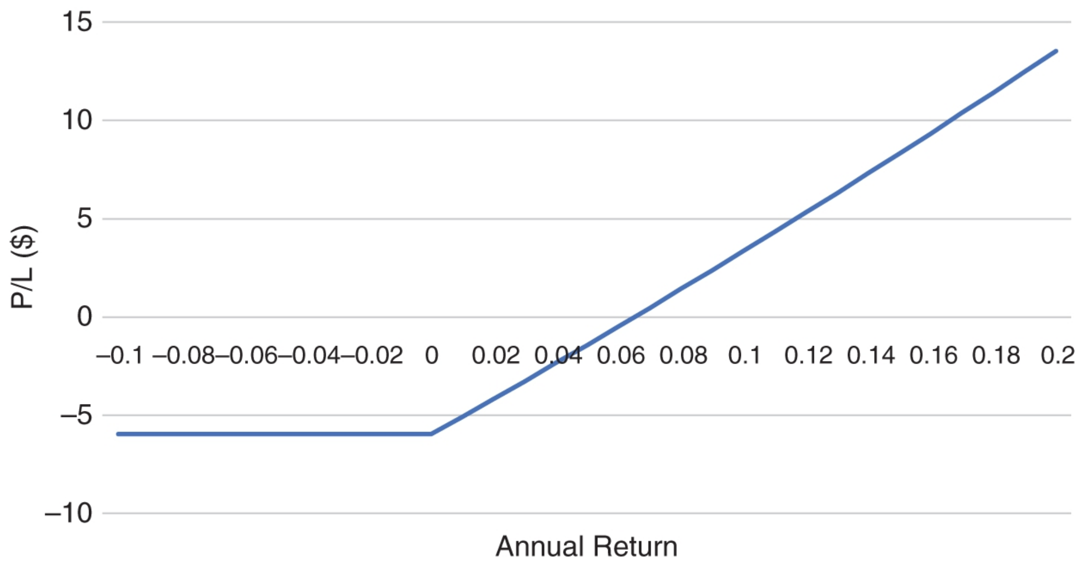


FIGURE 7.3 Ninetieth percentile of the profit of the 3-month 100-strike call. Stock is $100, volatility is 30%, and rates are zero (the risk-neutral call value is $5.98).

Strike Choice

All of the individual risk measurements given above need to be considered. None is sufficient on its own to determine that a particular option is "best." And this is probably a good thing, because "best" or "optimal" is only optimal with respect to a given criterion, and trading decisions need to be based on more than just one criterion.

In the specific case of strike selection (or investment selection in general) it will be impossible to give a single optimal solution. Some problems can't sensibly be solved this way. Think about the question, "What is the best car in the world?" Here, there are many plausible definitions of *best*. Does best mean fastest? Most luxurious? Safest? Greenest? Cheapest to buy? Cheapest to run? Most reliable? Within each category it is possible to make valid comparisons. A Ferrari is better than a Lamborghini. A Toyota is better than a Geo. But the argument over whether a Ferrari is better than a Toyota is unresolvable, because different car users have different goals and preferences.

In some other situations, it *is* possible to define a clear and unambiguous definition of *best*. Consider baseball. The goal of a baseball team is to win games. The best player is the one who most helps his team to do this. The individual skills of hitting, throwing, catching, and running are now seen as components of the ability to create wins, rather than unrelated goals in themselves. So here, a statistic that aggregates the component skills by putting them on a consistent scale and converting them into a measure of wins created is very useful. You can then meaningfully compare a power-hitting catcher to a fast, agile shortstop. But this introduces the danger of overreliance on the power of this single statistic. This number will still have methodological issues, and there will always be sampling problems with the individual component measurements. Even the best composite statistic should be a starting point rather than a definitive answer.

Trading is somewhere in between these situations. The goal of making money is absolute but risk is personal, both in terms of how it relates to a given trader's edges and abilities but also in terms of risk tolerances and aversions. It should be obvious that different people have different levels of risk aversion. This could be personal or it could be because of an external mandate. Anyone trading someone else's money will have to conform to the risk preferences of the capital provider. But it is also important to remember that risk depends on the skills of a specific trader. Risk is all of the things outside our control. So, traders with different sources of edge will have different remaining risk factors. If one trader has an edge in volatility prediction and another doesn't, volatility is an edge for the first and a risk to the second. This is true in most of life. For a heart surgeon, doing a bypass is a low-risk operation. For a random person, it would be murder.

So, a composite statistic for comparing risk and reward will be useful, but we should also not expect too much of it.

The most well-known of these statistics is the Sharpe ratio, the ratio of (excess) return to the volatility of the return. It is also well-known that the Sharpe ratio is not perfect. It has a large sampling error, generally comparable in size to the estimate. It doesn't distinguish between downside and upside volatility. It doesn't take higher-order moments into account at all. These are all problems, but the specific option-related issue is that we will be dealing with heavily skewed returns. (Skew can be a feature, not a bug. Positive skew is a good reason to include long options in a portfolio.)

The first work to address this failing was done by Hodges (1998), who showed the nature of the problem with a very simple example. We have two probability distributions, A and B, of excess returns.

Distribution A

Return	**−25%**	**−15%**	**−5%**	**5%**	**15%**	**25%**	**35%**
Probability	0.01	0.04	0.25	0.40	0.25	0.04	0.01

Distribution B

Return	**−25%**	**−15%**	**−5%**	**5%**	**15%**	**25%**	**45%**
Probability	0.01	0.04	0.25	0.40	0.25	0.04	0.01

Summary Statistics

Distribution	A	B
Mean return	5.00%	5.10%
Standard deviation	10.00 %	10.34 %
Sharpe ratio	0.50	0.493

Clearly distribution B is better than distribution A. The only difference is that the outcome of 35% has been increased to 45%. But this (good) change has increased the standard deviation more than the return, so the Sharpe ratio of distribution B is lower than that of distribution A.

Hodges derived a generalized Sharpe ratio (GSR) for an investor with exponential utility, but it was necessary to know the complete distribution of payoffs to make the calculation.

Pézier (2004) applied similar reasoning to create a GSR that only requires the moments of the distribution (see Maillard, 2018, for a full derivation and discussion). His GSR is

$$GSR = SR\left(1 + \frac{\lambda_3}{6}SR - \frac{\lambda_4 - 3}{24}SR^2\right) \tag{7.16}$$

TABLE 7.2 Projected Performance Numbers for Long Positions in Different Strike 3-Month Calls on a $100 Stock with a Drift of 10%, Volatility of 30%, and Zero Interest Rates

Strike	Average Dollar Profit	Average Percentage Return	Median Percentage Profit	90th Percentile Percentage Profit	Probability of Profit	GSR
80	$2.41	11.8%	10.4%	116.9%	52%	0.34
85	$2.26	14.1%	9.7%	145.7%	50%	0.33
90	$2.04	16.9%	4.2%	185.0%	48%	0.32
95	$1.75	20.2%	−13.1%	237.6%	44%	0.31
100	$1.43	23.8%	−57.6%	305.8%	37%	0.27
105	$1.10	27.8%	−100%	387.8%	31%	0.25
110	$0.80	32.0%	−100%	470.4%	24%	0.23
115	$0.56	36.6%	−100%	509.0%	17%	0.20
120	$0.37	41.3%	−100%	378.3%	12%	0.16

where *SR* is the standard Sharpe ratio, λ_3 is the skewness of returns, and λ_4 is the kurtosis. For normal returns, the GSR reduces to the Sharpe ratio. Positive skewness increases the GSR. Negative skewness lowers the GSR. Any kurtosis lowers the GSR.

Table 7.2 gives the various statistics for different 3-month call options on a $100 stock with a return of 10%. Both realized and implied volatilities are 30% and rates are zero. Each trader needs to choose the strike that most closely matches what they are looking for.

This analysis assumes we have paid the correct volatility level for the options. If we pay too much, our results look much worse. Even when trading purely directionally, implied volatility is very important. This is shown in Table 7.3 where we assume that the implied volatility was 30% but realized volatility was only 22% (this roughly corresponds to the typical variance premium). This effect needs to be considered if the implied volatility of the strike under consideration is very different from the ATM volatility.

(Interestingly, the GSR of the 120 strike is better than that of the 105, 110, and 115 strikes. This is because of the extreme skew of the results.)

TABLE 7.3 Projected Performance Numbers for Long Positions in Different Strike 3-Month Calls on a $100 Stock with a Drift of 10%, Implied Volatility of 30%, Realized Volatility of 22%, and Zero Interest Rates

Strike	Average Dollar Profit	Average Percentage Return	Median Percentage Profit	90th Percentile Percentage Profit	Probability of Profit	GSR
80	$2.16	10.6%	10.4%	86.5%	54%	0.37
85	$1.74	10.8%	9.7%	106.8%	53%	0.33
90	$1.12	9.3%	4.2%	133.4%	50%	0.24
95	$0.43	4.9%	−13.1%	166.3%	43%	0.10
100	−$0.17	−2.6%	−57.6%	201.9%	36%	−0.04
105	−$0.53	−13.3%	−100%	230.6%	26%	−0.19
110	−$0.65	−26.0%	−100%	222.1%	18%	−0.21
115	−$0.60	−39.8%	−100%	100.7%	11%	−0.21
120	−$0.47	−53.1%	−100%	−100%	6%	−0.06

Fundamental Considerations

So far, we have assumed our forecast was only of the mean and variance. Sometimes we may have a more complex view. For example, this is common in the Eurodollar market. Traders tend to forecast in discrete increments; for example, a 25 bp cut has a 40% chance of occurrence, instead of continuous outcomes, that is, a mean return of 5%.

In these cases, each strike should be evaluated with a different subjective drift parameter. Although, given the trader's bias toward a certain probability distribution, the analysis will probably confirm only preexisting opinions (opinions in, opinions out).

Traders in most other products should be careful to ask themselves if their forecasts of the distribution lead to expected value. It is hard to predict volatility. It is harder to predict return. Predicting the full distribution is probably a manifestation of overconfidence.

Conclusion

There is no simple answer to the question, “What strike should I buy?” Basing the decision on maximizing average return, median return, or probability of profit will lead to different answers. And there are many other statistics that could be sensibly considered. It is also quite likely that a trader's criteria will change based on the rest of her portfolio. The decision needs to be made based on personal utility and on a case-by-case basis.

Summary

- By interpreting the dividend yield term in the generalized BSM model to be a drift parameter we can get subjective option values that depend on return.
- These prices are not arbitrage free but can be used to derive real-world statistics (as opposed to risk-neutral statistics).
- Different evaluation criteria will suggest very different “optimal” strikes.

CHAPTER 8
Directional Option Strategy Selection

In addition to choosing a strike and expiration, the trader needs to decide what strategy to employ. There are many option structures that could be used to speculate directionally, but here we will confine ourselves to those that could be considered the fundamental building blocks of the others: the long call, the long call spread, the short put, the short put spread, and the risk reversal.

It is possible to construct a matrix that constructs optimal positions using various risk measures such as the Sharpe ratio, the generalized Sharpe ratio (GSR), or the Kelly ratio. Doing this isn't stupid but here are several problems with the approach:

- Different criteria will recommend different structures, and none will express the investor's real utility. All the criteria are useful guides, but none are definitive.
- Often the difference between strategies will be minimal using this method.
- The various risk ratios assume that our forecasts are correct. It is more important to understand what happens if we are wrong. Although it is possible to calculate derivatives of the ratios (e.g., the derivative of GSR with respect to return), it is more instructive to again run simulations.

Long Stock

This is our baseline. We buy 100 shares of a $100 stock. We expect a 20% return and realized volatility is 30%. If our return estimate is correct, the (lognormal) PL distribution after 1-year is shown in Figure 8.1 and summary statistics are shown in Table 8.1.

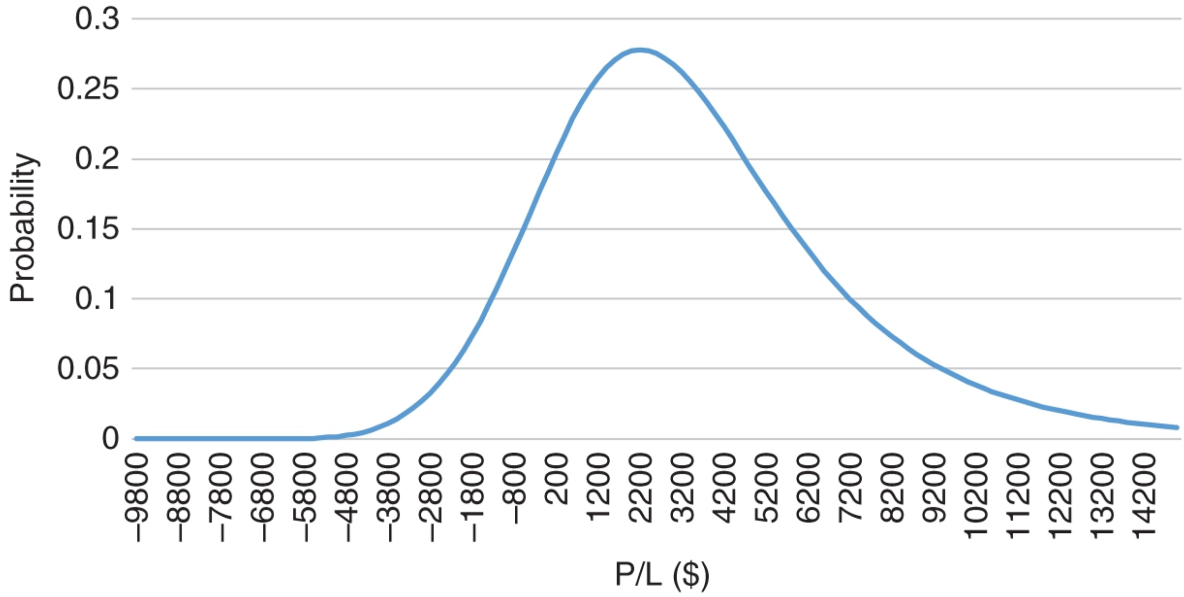


FIGURE 8.1 The PL distribution for 100 shares of a $100 stock with a 20% return; volatility is 30%.

TABLE 8.1 Summary Statistics for 100 Shares of a $100 Stock with a 20% Return (Volatility is 30%.)

Average	$2,640
Standard deviation	$3,736
Skewness	1.12
Excess kurtosis	2.73
Median	$2,214
90th percentile	$7,940
Maximum (in a 10,000-path simulation)	$27,220
10th percentile	−$1,660
Minimum (in a 10,000-path simulation)	−$6,260
Percent profitable	75%

Long Call

Consider a long ATM 1-year call on a $100 stock when rates are zero. We expect a 20% return and both implied and realized volatility are 30%. If our return estimate is correct, the PL distribution from a simulation of 10,000 paths is shown in Figure

8.2 and summary statistics are shown in Table 8.2. The initial value of the call is $11.92.

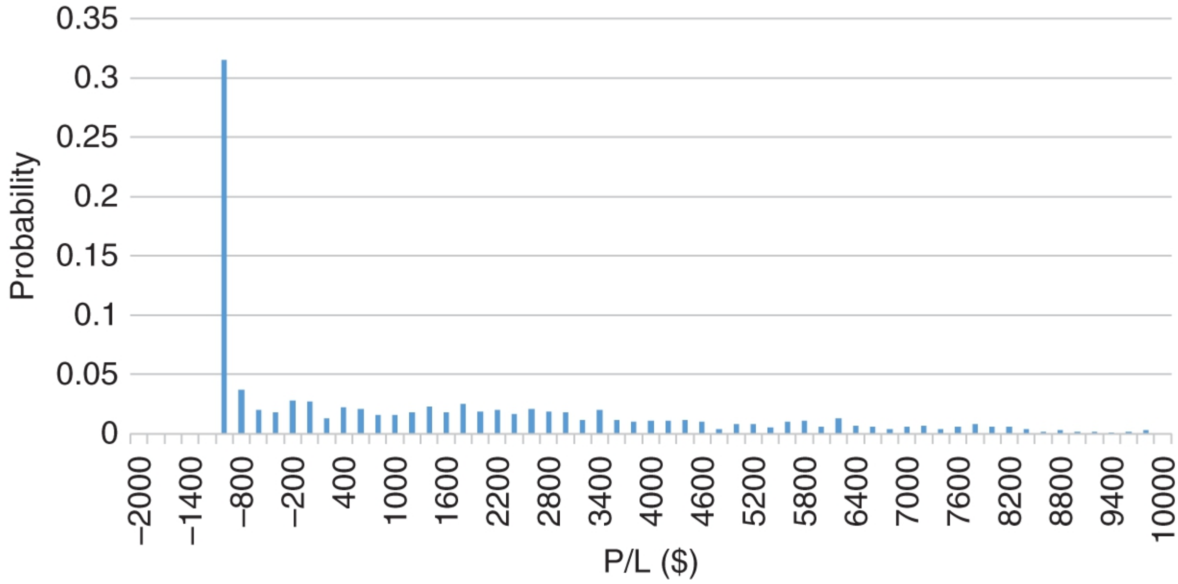


FIGURE 8.2 The PL distribution for a 1-year ATM call option on a $100 stock with a 20% return. Both implied and realized volatilities are 30% and rates are zero.

TABLE 8.2 Summary Statistics of the PL Distribution for a 1-Year ATM Call Option on a $100 Stock with a 20% Return (Both implied and realized volatilities are 30% and rates are zero.)

Average	$1,516
Standard deviation	$3,198
Skewness	1.66
Excess kurtosis	4.12
Median	$538
90th percentile	$5,843
Maximum	$23,068
10th percentile	−$1,192
Minimum	−$1,192
Percent profitable	58%

If we had bought 100 shares of stock, our average median profit would have been $2,200. The option premium pays for the leverage relative to the shares and the limited downside. A call option is similar (but not identical) to a long stock position and a stop-loss order at the strike. However, stops will kill some trades that would eventually have recovered. Options won't do this, and this is the benefit of paying the premium.

Long Call Spread

We buy the 1-year ATM call and sell the 20-delta call (the 135 strike). The PL distribution from a simulation of 10,000 paths is shown in Figure 8.3 and summary statistics are shown in Table 8.3. The initial value of the spread is $9.04.

This P/L distribution is similar to a long position with a stop and a profit target. Although returns are far from normal, the extreme values have been eliminated along with the skewness.

In indices, it is quite possible that the implied skew means you will be selling the short strike at a discount to the ATM volatility. In many other products, you will receive a premium for the short strike. This changes the initial premium and hence the profits but won't change the shape of the terminal distribution.

One benefit that the call spread offers over the call is related to psychology. When holding a call, particularly an OTM call, you are paying for the extreme upside. This means you need to continue to hold the option. A lot of traders have trouble with this (in my experience, amateurs can't take losses and professionals are too inclined to take profits). Instead of fighting this tendency, it may be better to buy a call spread instead of a call. The short strike will be the profit target and the position won't have cost as much to initiate.

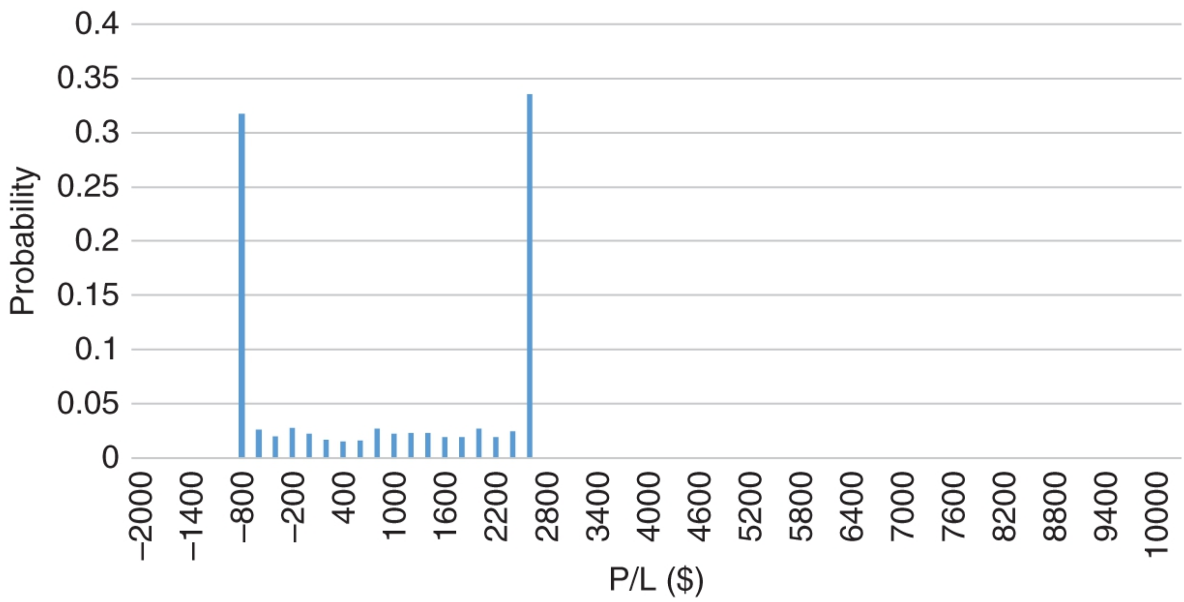


FIGURE 8.3 The PL distribution for a 1-year ATM/20-delta call spread on a $100 stock with a 20% return. Both implied and realized volatilities are 30% and rates are zero.

TABLE 8.3 Summary Statistics of the PL Distribution for a 1-Year ATM/20-Delta Call Spread on a $100 Stock with a 20% Return (Both implied and realized volatilities are 30% and rates are zero.)

Average	$819
Standard deviation	$1,502
Skewness	0.01
Excess kurtosis	−1.76
Median	$759
90th percentile	$2,596
Maximum	$2,596
10th percentile	−$904
Minimum	−$904
Percent profitable	58%

Short Put

We sell the 1-year ATM put. The PL distribution from a simulation of 10,000 paths is shown in Figure 8.4 and summary statistics are shown in Table 8.4. The initial value of the put is $11.92.

Choosing a short put instead of a long call is really about preferring a high probability of a smaller profit to a larger average profit and positive skewness.

Covered Calls

A covered call consists of a long position in a stock and short position in a call on that stock. In exchange for receiving the option premium, the investor has her upside capped. Figure 8.5 shows the covered call payoff at expiration when a 100-strike call option is sold for $5.

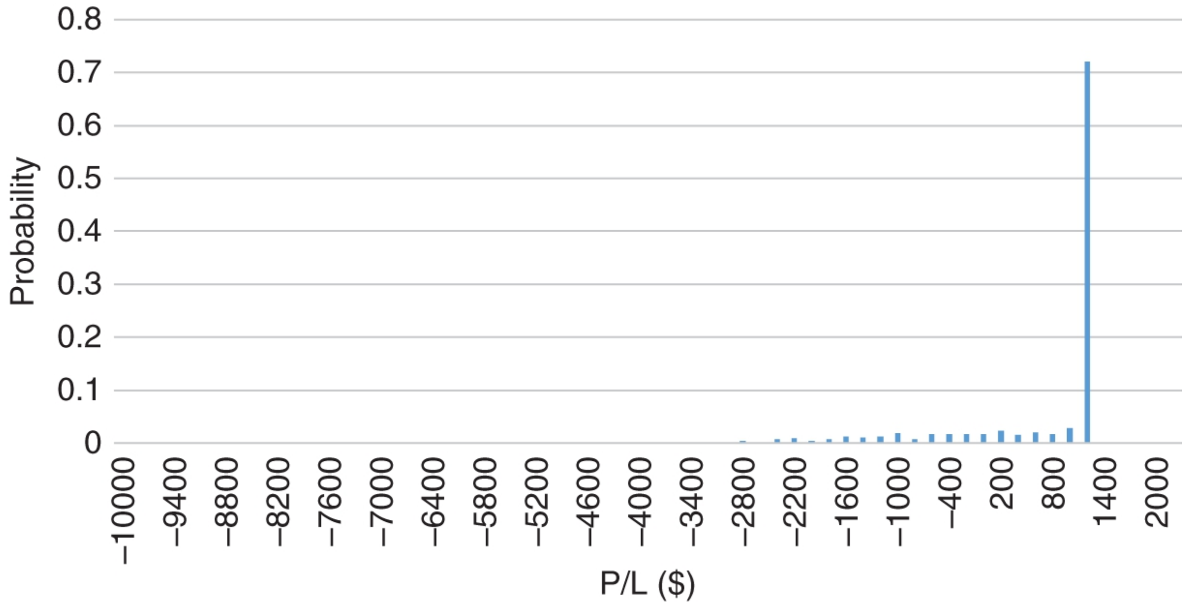


FIGURE 8.4 The PL distribution for a short 1-year ATM put option on a $100 stock with a 20% return. Both implied and realized volatilities are 30% and rates are zero.

TABLE 8.4 Summary Statistics of the PL Distribution for a Short 1-Year ATM Put Option on a $100 Stock with a 20% Return (Both implied and realized volatilities are 30% and rates are zero.)

Average	$706
Standard deviation	$986
Skewness	−2.22

Excess kurtosis	4.53
Median	$1,192
90th percentile	$1,192
Maximum	$1,192
10th percentile	−$8,43
Minimum	−$5,172
Percent profitable	78%

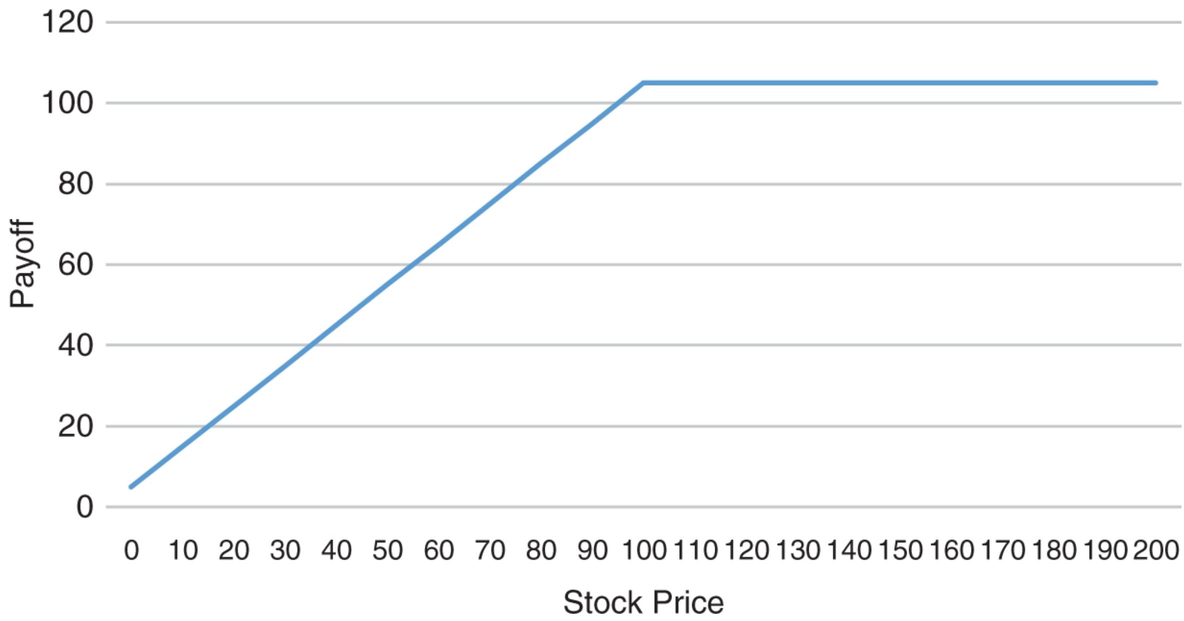


FIGURE 8.5 The payoff of the covered call as a function of stock price at expiration.

Synthetically a covered call is the same as a short put. Instead of selling a call against an established long stock position, investors sometimes sell a put and hold enough cash to be able to purchase the stock if they are assigned. Synthetically, this position is the same as a covered call with the same strike. However, there are some differences in how and why these strategies are used:

- Some investors are prohibited from put selling, but they can write covered calls.
- The chosen strikes tend to be different, with both strategies generally implemented with out-of-the-money options.
- Selling out-of-the-money put options means the trader usually benefits from selling at an implied volatility premium.

The psychological effects on the trader are also somewhat different. The holder of a covered call tends to be happy with rallies, whereas the seller of the put often feels she has missed out in the case of a large rally. This is because we frame the situations differently. Covered calls are framed as a situation in which we are long and are prepared to sell, whereas short puts are seen as a situation in which we are waiting to get long at a certain price. Part of this is due to the different strike choices but the reasoning is still specious. This shouldn't be a relevant consideration in strategy selection, but in practice it is.

Covered calls have been popular with retail traders due to the argument that "I would sell the stock if it went to the strike price, so why not get paid to do that?" This reasoning is poor, but covered calls are also the rare example of a popular retail strategy that works well and makes good sense. Over time they have delivered equity-like returns with lower risk. For example, consider the CBOE BuyWrite Index. This consists of holding the SPX portfolio and selling slightly out-of-the-money 1-month calls that are held until expiration. The performance is shown in Figure 8.6 and summarized in Table 8.5, together with the S&P 500 (including dividends).

We can see that the outperformance of the covered call strategy is robust with respect to the exact implementation by looking at the results of BXY (which sells 2% out of the money calls) and BXMD (which sells 30-delta calls). These results are summarized in Table 8.6.

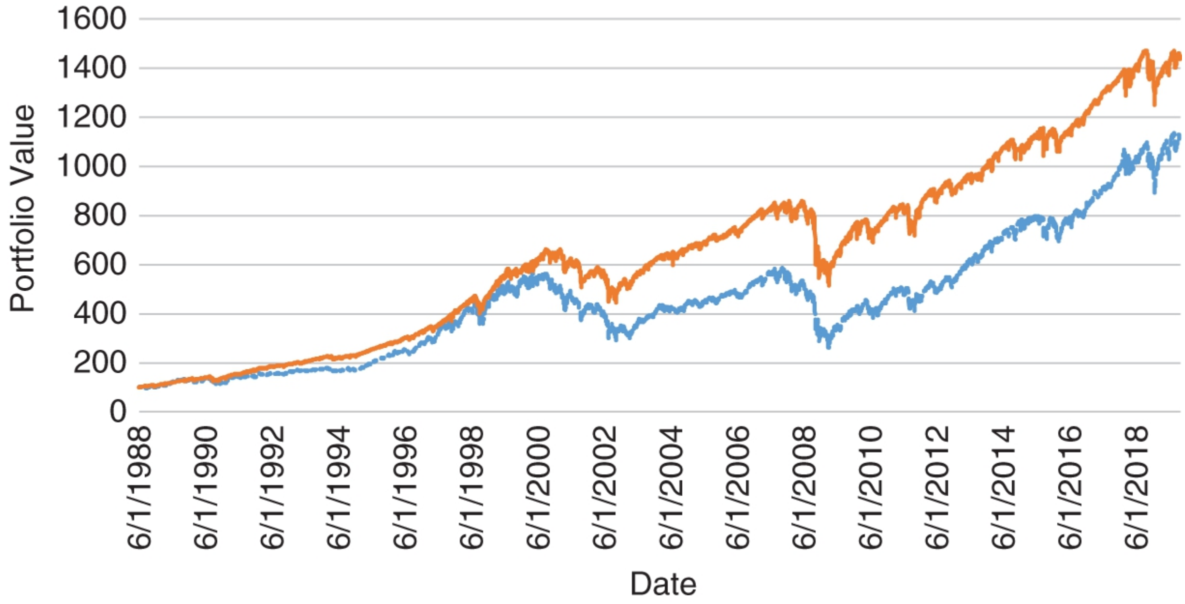


FIGURE 8.6 The performance of the CBOE BuyWrite Index compared to that of the S&P 500 index from June 1988 to September 2019.

TABLE 8.5 Summary Statistics for BXM and the S&P 500

Statistic	BXM	S&P 500
Annual return	8.5%	7.7%
Volatility	12.6%	17.3%
Max drawdown	40.1%	56.8%
Skew	−0.67	−0.29

TABLE 8.6 Summary Statistics for BXY, BXMD, and the S&P 500 from June 1988 to July 2019

Statistic	BXY	BXMD	S&P 500
Annual return	8.6%	10.3%	7.7%
Volatility	12.6%	14.7%	17.3%
Max drawdown	40.1%	46.9%	56.8%
Skew	−0.67	−0.46	−0.29

Components of Covered Call Profits

The discussion of strike choice up until now has only focused on the risk characteristics of the options. We also need to consider the factors that drive option returns when making the choice. As an example, we consider a covered call. This is the simplest possible option position, but this analysis is quite general.

The reason that covered calls can provide equity-like returns with lower volatility is that they are exposed to two profitable factors: the equity market risk factor and the volatility premium. Selling a call against an existing position reduces the portfolio's exposure to the stock while adding a short volatility exposure. The lower volatility of a covered call position is due to the diversification that two factor exposures provide.

Consider a stock with a current price of $100. We assume that the stock increases by 10% a year and has a volatility of 15%. We also assume that dividends and rates are zero. We sell a 1-year call option with an implied volatility of 20% against this position.

This at-the-money covered call has a delta of 0.47 when evaluated at the realized volatility, so it will earn 4.7% from its exposure to the stock's appreciation (to a first approximation). Also, this option has a premium of $7.97 ($2.00 more than it would have been worth at the true realized volatility). So, about 25% of the option premium that the seller collected is harvesting the volatility premium. That is, we expect to gain 2.0% a year from harvesting the short volatility premium. Here the total expected return of the covered call is 6.7%. Although this example has a lower expected return than the stock, it also has lower volatility.

Obviously, this theoretical decomposition varies with respect to the volatility premium, and the ex-post return is affected by both realized volatility and return. Also, the strike choice and expiration of the short call determines how much of the profit comes from directional exposure and how much comes from volatility harvesting.

First, we look at the decomposition as a function of a strike assuming all strikes have the same variance premium. This is shown in Figure 8.7.

We can see that although the equity premium is an increasing function of strike, the volatility premium is peaked at strikes just above the current stock price. This is where the option's sensitivity to volatility (vega) is greatest, so it is also where exposure to the volatility premium is maximized.

Note that if there is no volatility premium, the covered call earns just the stock return multiplied by its delta (the exposure to its underlying equity). This is the most important concept. Selling options only makes sense if a volatility premium exists. Further, if an investor is more confident in the existence of the volatility premium, he should sell options just above the ATM. Conversely, if he is more confident in the equity return, he should sell options that are further out of the money. The fact that different investors have different forecasting abilities means that they will also choose different option structures.

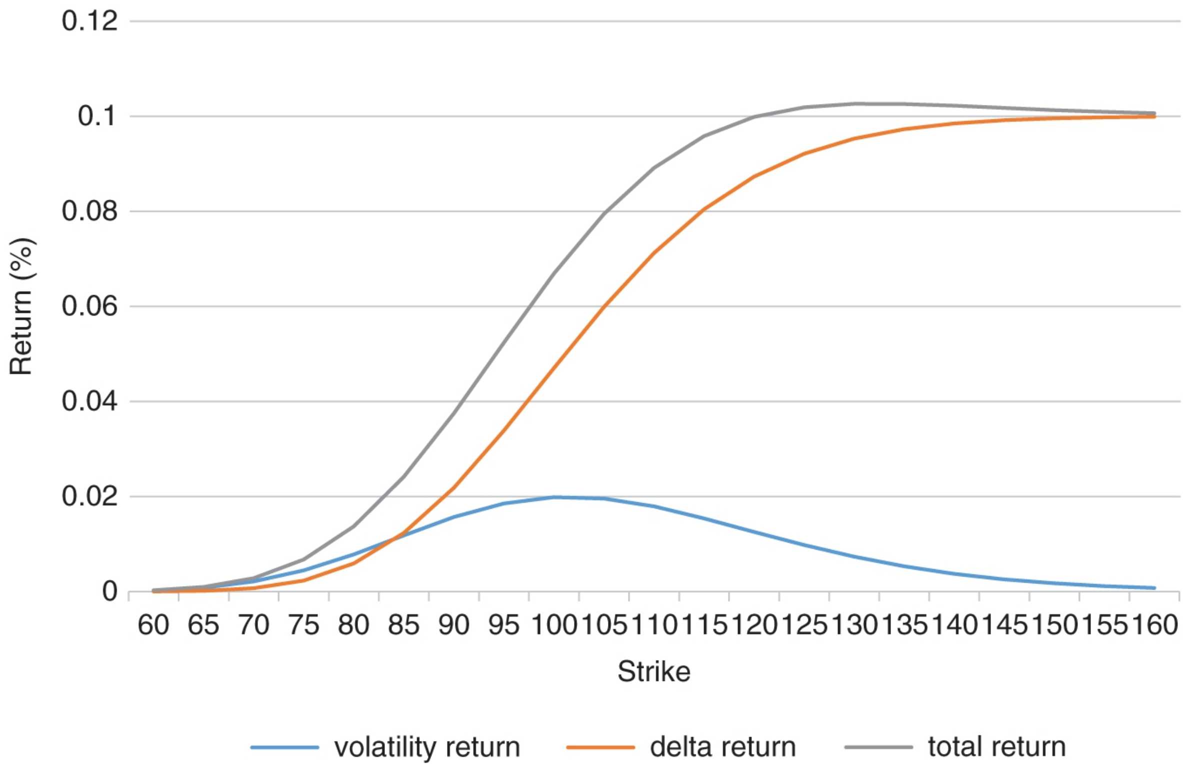


FIGURE 8.7 The total profit of the covered call and how much comes from equity return and volatility premium.

Covered Calls and Fundamentals

It is well-known that value stocks, momentum stocks, low-beta stocks, and small-cap stocks tend to be the best performers. So, investors should preferentially own these. Sadly, with the exception of momentum stocks, these classes of equities tend to have the lowest variance premia. This means that the investor has to choose between better delta performance or better variance performance and make the strike choice accordingly. Again, this choice will depend on the ability of the individual trader.

It is also important to note that the variance premium and subsequent stock returns (the two sources of edge) are not completely independent. High-variance premium predicts high future stock returns. This effect has been extensively studied (see, for example, Bollerslev and Zhou, 2007; Bollerslev and Todorov, 2011; Kelly and Jiang, 2014; Bollerslev et al., 2014) and exists for both single stocks and at the index level. This effect is strong enough that it can be used as a timing signal: when the variance premium is high is a good time to enter covered call positions.

Some explanations for the effect are very complex but, very simply, stock markets have tended to go up and volatility rises during drops, so, a high-variance premium is correlated with temporary dips.

Short Put Spread

We sell the 1-year ATM put and buy the 20-delta put (the 81 strike). The PL distribution from a simulation of 10,000 paths is shown in Figure 8.8 and summary statistics are shown in Table 8.7. The initial value of the spread is $8.11.

Unfortunately, in most products the long put will have a significantly higher implied volatility than the short ATM option. For example, the S&P 500 20-delta put currently has an implied volatility of about 1.36 times that of the ATM. So, if the ATM volatility is 30%, we would be paying 40.8% for the long 20-delta put. This doesn't greatly change the *shape* of the distribution but will be a significant drag on profits. This is shown in Table 8.8.

For many definitions of *conservative,* the short spread is the most conservative directional strategy. It has a high winning percentage, a high median, and a capped downside.

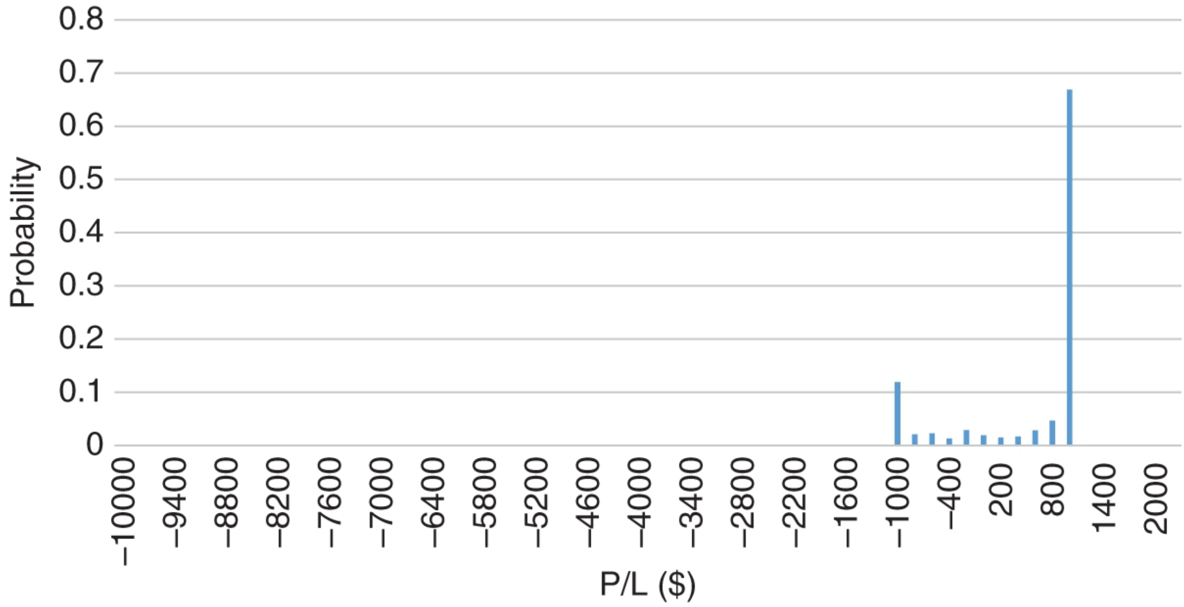


FIGURE 8.8 The PL distribution for a short 1-year ATM/20-delta put spread on a $100 stock with a 20% return. Both implied and realized volatilities are 30% and rates are zero.

TABLE 8.7 Summary Statistics of the PL Distribution for a Short 1-Year ATM/20-Delta Put Spread on a $100 Stock with a 20% Return (Both implied and realized volatilities are 30% and rates are zero.)

Average	$422
Standard deviation	$680
Skewness	−1.42
Excess kurtosis	0.43
Median	$811
90th percentile	$811
Maximum	$811
10th percentile	−$1,072
Minimum	−$1,089
Percent profitable	78%

TABLE 8.8 Summary Statistics of the PL Distribution for a Short 1-Year ATM/20-Delta Put Spread on a $100 Stock with a 20% Return (The ATM-implied volatility is 30% and the implied volatility of the 20-delta put is 40.8%. Realized volatilities are 30% and rates are zero.)

Average	$216
Standard deviation	$720
Skewness	−1.52
Excess kurtosis	0.62
Median	$634
90th percentile	$634
Maximum	$634
10th percentile	−$1,404
Minimum	−$1,666
Percent profitable	76%

Risk Reversal

We sell the 1-year 20-delta put and buy the 20-delta call. The PL distribution from a simulation of 10,000 paths is shown in Figure 8.9 and summary statistics are shown in Table 8.9. The initial value of the position is a credit of $93.

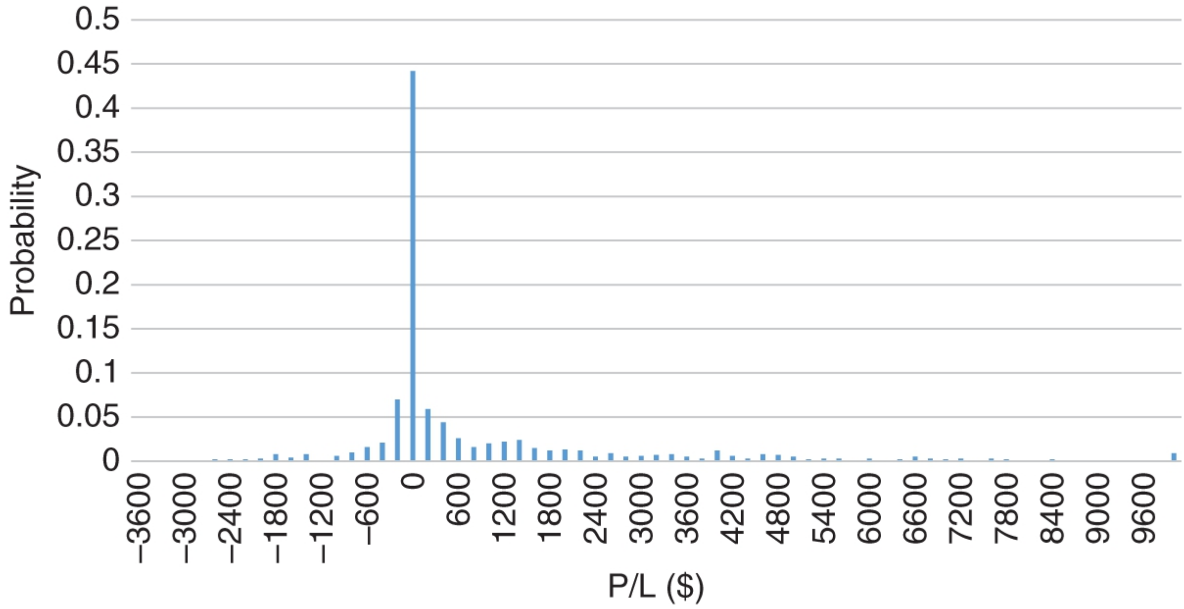


FIGURE 8.9 The PL distribution for a 1-year 20-delta risk reversal on a $100 stock with a 20% return. Both implied and realized volatilities are 30% and rates are zero.

TABLE 8.9 Summary Statistics of the PL Distribution for a 1-Year 20-Delta Risk Reversal on a $100 Stock with a 20% Return (Both implied and realized volatilities are 30% and rates are zero.)

Average	$948
Standard deviation	$2,140
Skewness	2.69
Excess kurtosis	9.00
Median	$79
90th percentile	$3,666
Maximum	$22,680
10th percentile	−$251
Minimum	−$3,802
Percent profitable	85%

Although this position has an initial delta of 40, as the stock moves the delta changes. This is reflected in the fact that the maximum profit and losses are practically the same as for a long stock position. In exceptionally large moves the 20-delta options become 100 delta and the risk reversal mimics a stock position. However, the risk reversal has lower average and median than the stock position. There will be times when the long calls expire worthless even when the stock rallies. The reason the median value is positive is that the position was entered into at a credit.

In most markets, we will be benefiting from selling the 20-delta put at an inflated volatility and, at least for indices, we will be buying the 20-delta call at a volatility under that of the ATM. For the S&P 500, the 20-delta call currently has a volatility of about 0.77 of the ATM volatility. Assuming this, the effect on summary statistics is shown in Table 8.10. The fact the position performs better when there is a skew is entirely due to the extra premium we collect. The initial value of this position is a credit of $330.

Because we can collect a reasonable premium from this position, we can still buy a teeny put to hedge the downside risk. For example, the implied volatility of a 5-delta index put is usually about 1.7 times the ATM volatility. In this case that gives an implied volatility of 51% and a premium of $1.28 for this teeny put. That reduces the initial credit to $202 and lowers the average and percentile numbers by the same amount.

This downside hedged risk reversal is my personal favorite bullish directional position.

- It has limited downside risk.

TABLE 8.10 Summary Statistics of the PL Distribution for a 1-Year 20-Delta Risk Reversal on a $100 Stock with a 20% Return (The call implied volatility is 23.1% and the implied volatility of the 20-delta put is 40.8%. Realized volatility is 30% and rates are zero.)

Average	$1,430
Standard deviation	$2,320
Skewness	2.69
Excess kurtosis	10.2
Median	$366
90th percentile	$4,448
Maximum	$27,262

10th percentile	$85
Minimum	–$3,205
Percent profitable	91%

- It has the potential for large wins.
- It takes advantage of the skewness premium.

Aside: The Risk Reversal as a Skew Trade

As demonstrated, the risk reversal is an effective way to profit from the implied skewness premium. However, despite many views to the contrary, it isn't particularly useful for speculating on the movement of the implied skew itself. Although the implied skew does fluctuate, the size of its moves is dwarfed by the effects of the stock movement and the *level* of implied volatility.

Consider the risk-reversal just discussed. Imagine we are selling the put and buying the call because we think the slope of the skew will flatten. If we think the put volatility ratio to the ATM volatility will drop and the call ratio will increase, we will make a profit of

$$Profit = Vega_{call}(\sigma_{callnew} - \sigma_{callold}) + Vega_{put}(\sigma_{putold} - \sigma_{putnew}) \tag{8.1}$$

Consider a 1-month risk reversal on a $100 stock. The 20-delta put (91 strike) has an implied volatility of 40.8% and the 20-delta call has an implied volatility of 23.1%. We sell the put and buy the call because we expect the skew to flatten. Table 8.11 shows the profits we make on the position for various degrees of flattening.

However, the expected daily move of a $100 stock with a volatility of 30% is $1.50. If the stock drops to $98.5, the risk reversal loses $94, and if the stock rallies to $101.5, the risk reversal will make $16. So, an average daily P/L due to the stock's random fluctuations is $55. Even if the implied curve flattens by three volatility points on both the calls and the puts on one day, the skew-related profit will still only be of the order of the delta/gamma P/L. It is exceptionally unlikely that a move will occur on the day after the trade is initiated. This analysis also ignores changes in the level of the volatility curve and the effect of correlations between stock returns and implied skewness. Long-dated options will have less gamma to cause problems and more vega to make money off implied volatility changes. However, the long-dated implied volatility curves are much more stable than short-dated ones.

TABLE 8.11 Results for a Short Put–Long Call 20-Delta Risk Reversal for Various Amounts of Implied Volatility Curve Flattening

Put volatility	0.408	0.398	0.388	0.378
Call volatility	0.231	0.241	0.251	0.261
Risk reversal value	0.43	0.25	0.08	−0.09
Profit ($)	0	18	35	52

It is *possible* to make money with this trade. The idea of taking advantage of reversion of the implied skew is a sensible one. But the edge from this prediction is likely to be overwhelmed by noise.

Ratio Spreads

Although all directional option positions are dependent on volatility, some have a higher dependence than others. Ratio spreads are an extreme example. Although they are often used to speculate on direction, their primary exposure is to volatility. They have the payoff of a broken wing butterfly, something we always think of as primarily a volatility position.

We buy the 1-year ATM call and sell two of the 20-delta calls. The PL distribution from a simulation of 10,000 paths is shown in Figure 8.10 and summary statistics are shown in Table 8.12. The initial value of the position is a debit of $749.

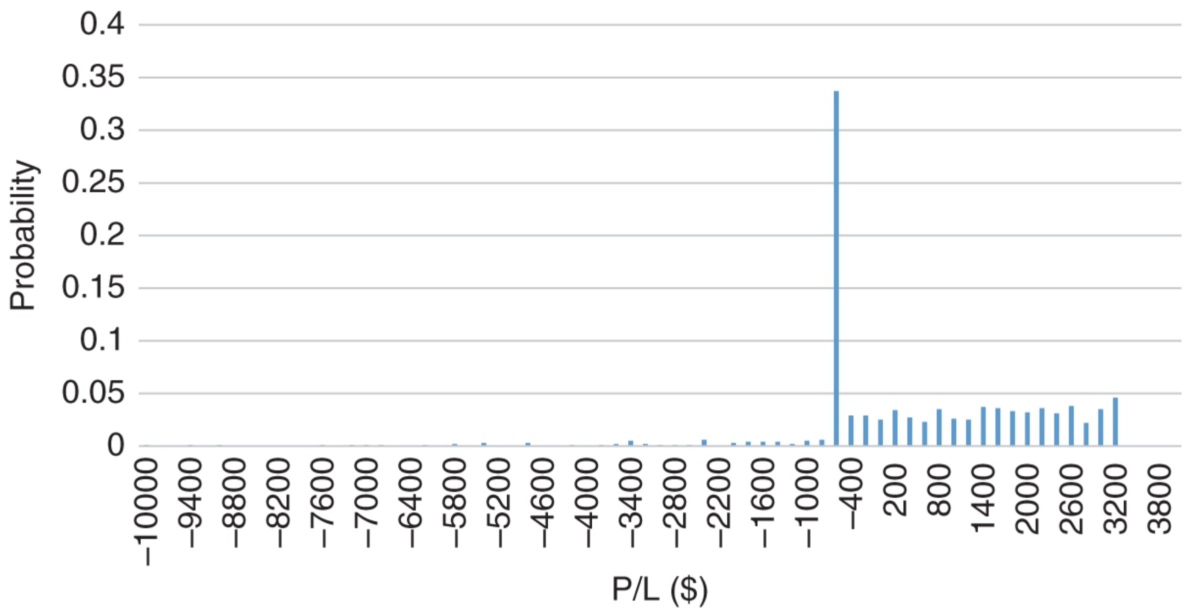


FIGURE 8.10 The PL distribution for a 1-year ATM/20-delta risk one-by-two call spread on a $100 stock with a 20% return. Both implied and realized volatilities are 30% and rates are zero.

TABLE 8.12 Summary Statistics of the PL Distribution for a 1-Year 20-Delta Risk Reversal on a $100 Stock with a 20% Return (Both implied and realized volatilities are 30% and rates are zero.)

Average	$381
Standard deviation	$1780
Skewness	−1.10
Excess kurtosis	5.20
Median	$190
90th percentile	$2,670
Maximum	$3,140
10th percentile	−$749
Minimum	−$15,230
Percent profitable	52%

As I have emphasized, there are few hard-and-fast rules in option trading, but the version of the ratio spread that is long the one option and short the second is generally best avoided. In terms of finding edge, the trader needs a good prediction of both volatility and direction. It is difficult to predict either, let alone both. There are better alternatives for both directional and volatility speculation. It is tempting to use the sale of the two options to "finance" the purchase of the one, but this is only done by taking on unlimited risk. If a trader has enough edge (either in volatility or direction) to do a trade at all, she shouldn't be afraid to pay the option premium. There are no free lunches and there are no free option positions.

Another reason that is often given for trading a ratio is to short a high implied skew. This is usually done with puts because puts usually have a more pronounced implied skew. By selling the two farther-out-of-the-money options, the trader can short volatility at the higher implied volatility and mitigate the risk with the single long option. This trade has all of the same problems mentioned in the section on trading risk reversals to capture skew, but ratio spreads are an even worse vehicle for trading an idea that isn't very good to start with.

If a trader wants to collect the skew premium, the safest way is to sell a put spread rather than possibly offset this by purchasing a call spread. The long put will almost certainly be the option with the highest implied volatility in the structure, but because the short put will still be trading at a volatility premium to the ATM this position is still short implied skew.

An effective way to use a ratio spread is to buy the two options and sell the one as a relatively cheap catastrophe hedge. If the options are struck far enough out of the money, the hope is they will only come into play in a huge crash and we will then end up being long vega and net options (the most robust risk control number). This trade is still not a free lunch. If instead of a crash we have a slow move downwards, we can end up short gamma and paying theta, but the bad scenarios are ones that occur slowly so at least we can rebalance. This incurs transaction costs but at least we have a chance to trade.

As an example, with SPY 299 on October 23, 2019, we can sell the November 15 266 put for 0.16 and buy two of the 258 puts for 0.10 each. For the same outlay of 0.04 we could have bought the 241 put. Obviously, in almost all situations we will lose the 0.04 premium. The important thing in evaluating these positions isn't the overall probability distribution; it is how we look in the event of a crash. The risk slides for these two positions are shown in

Tables 8.13 and 8.14. On October 23 the ATM volatility was 11.4%. To estimate the relevant implied volatilities for each underlying price level I could use a regression model to find the historical relationship between price and volatility moves. Such a model is useful for normal trading purposes but for estimating tail event parameters it is at best useless and possibly dangerous if it gives a false sense of certainty. Instead I'm going to assign what I think are possible and I hope overly pessimistic volatility numbers.

TABLE 8.13 The Risk Slide for the Single 241 Put

SPY price change	−30%	−20%	−10%
Postulated IV	120%	80%	30%
Delta	−0.61	−0.47	−0.07
Vega	$2120	$2515	$965
P/L	$4650	$2070	$65

TABLE 8.14 The Risk Slide for the 258/266 One-By-Two Put Spread

SPY price change	−30%	−20%	−10%
Postulated IV	120%	80%	30%
Delta	−0.68	−0.53	0.14
Vega	$2050	$2556	−$410
P/L	$5300	$2560	$42

One could argue than in a “small” crash the single teeny put behaves better because it leaves us short delta and long vega, although the P/L superiority is small. However, in severe crashes the ratio spread gives significantly better protection.

Conclusion

As with volatility position selection, there is no one “best” strategy when using options to speculate directionally. The most important consideration is probably whether the variance premium is low or high. This, more than risk preferences, is paramount in deciding to be long or short options. Then the trader can decide on the structure based on preference for winning percentage, maximum profit, and maximum loss, and so on. Finally, strikes can be chosen by considering the risk characteristics discussed in Chapter Seven.

Summary

- Variance and skew premia are the most important factors even when trading options directionally.
- Single options have the best correlation between a profit and a successful prediction.
- Spreads are useful for mitigating the dependency of the trade on the variance premium. The fact they also create a stop (or profit target) is useful as a risk management tool but predicting an underlying's range is probably too hard to do consistently.

CHAPTER 9
Trade Sizing

It certainly is not true that good risk management can turn a strategy with no positive expectation into a winner. Risk management can change the return distribution of a strategy with dynamic trade sizing and the use of stops and profit targets, but unless there is real edge somewhere the idea will eventually lose money. However, this doesn't make risk control irrelevant. Bad risk management can turn a potentially profitable idea into a loser. And, because risk management is the only part of the trading process that is completely under the control of the trader, there is no excuse for not doing it as well as possible.

In this chapter, I restate why the Kelly criterion should form the basis of a trade-sizing scheme and look at two extensions that are particularly important when trading options: non-normal trade returns and uncertainty about the return distribution.

The Kelly Criterion

It is well-known that investing according to the Kelly criterion (Kelly, 1956) will theoretically outperform any other sizing strategy. No other sizing scheme will produce greater long-term growth. This doesn't mean that everyone tries or even wants to invest like this. There are three kinds of reasons for this:

- The Kelly criterion maximizes the long-term growth rate of the bankroll. But it is completely legitimate to have other goals. For example, many traders are more interested in maximizing the probability of hitting a goal in a given time (see Browne, 1999, 2000a, 2000b). No one has a utility function that is as simple as the log function that matches the Kelly scheme. This is fair.
- Some traders try to deny the math. They don't like the volatile return stream and decide that this is a flaw of the system. It isn't. Whether or not you like what Kelly says, it is a mathematical fact. It's objectively true. It is common for people to conflate their dislike of a situation with its truth.

Examples of this are evolution, climate change, and the Kelly criterion.

- Some people create strawman arguments about the mathematics, claiming that Kelly only applies to simplified, unrealistic situations. This isn't true at all. The mathematics of maximizing growth rate are quite general.

A derivation of the Kelly criterion for both discrete and continuous outcomes is given in Sinclair (2013), together with a discussion of the distribution of results we can expect when investing this way.

The important results can be summarized as follows.

Good

- Kelly maximizes growth rate.
- The expected time to reach any goal is minimized.
- It is impossible to go bankrupt.
- The strategy depends only on the current bankroll, not the specific trade results that led to it.
- It is essentially unbeatable.

Bad

- The best bets can be uncomfortably large.
- Portfolio volatility and drawdowns are large.
- Because of compounding, it is reasonably common to find that an equal number of wins and losses leaves you with a net loss.

Here I want to look at two slight extensions that are very important to traders, particularly option traders. What happens when outcomes are highly non-normal? And what happens when we are uncertain of outcomes and probabilities?

These are two aspects of the same general problem. Trading success is largely dependent on how robust our ideas are. At best our knowledge is uncertain. Probably our knowledge is incomplete and only partially correct. In particular, we will be ignorant of the true probabilities of rare events. These will be the events that drive non-normality, and, because they appear only rarely, they will be those that we are most uncertain of.

To introduce the ideas, we first look at the fairly impractical case of discrete trade results.

Non-normal Discrete Outcomes

Imagine we have a discrete set of outcomes W_i, each with probability p_i. We bet a fraction, f, of our bankroll on each opportunity. So, the gain factor per trade is

$$G = \prod_i (1 + f W_i)^{p_i} \tag{9.1}$$

Alternatively, the exponential growth per unit bet is

$$\sum_i p_i \log(1 + f W_i) \tag{9.2}$$

To find the value of f that maximizes the exponential growth rate, we differentiate with respect to f and set the derivative to zero. If i is greater than 2, this is unwieldy or impossible and we need to use numerical methods. A numerical solution of a simple example is illustrative.

Case One

$p_1 = 0.55$

$p_2 = 0.45$

$W_1 = 1$

$W_2 = -1$

(55% chance of winning a dollar and 45% chance of losing a dollar)

This implies $f_{\max} = 0.1$. The dependence of the exponential growth rate on f is shown in Figure 9.1.

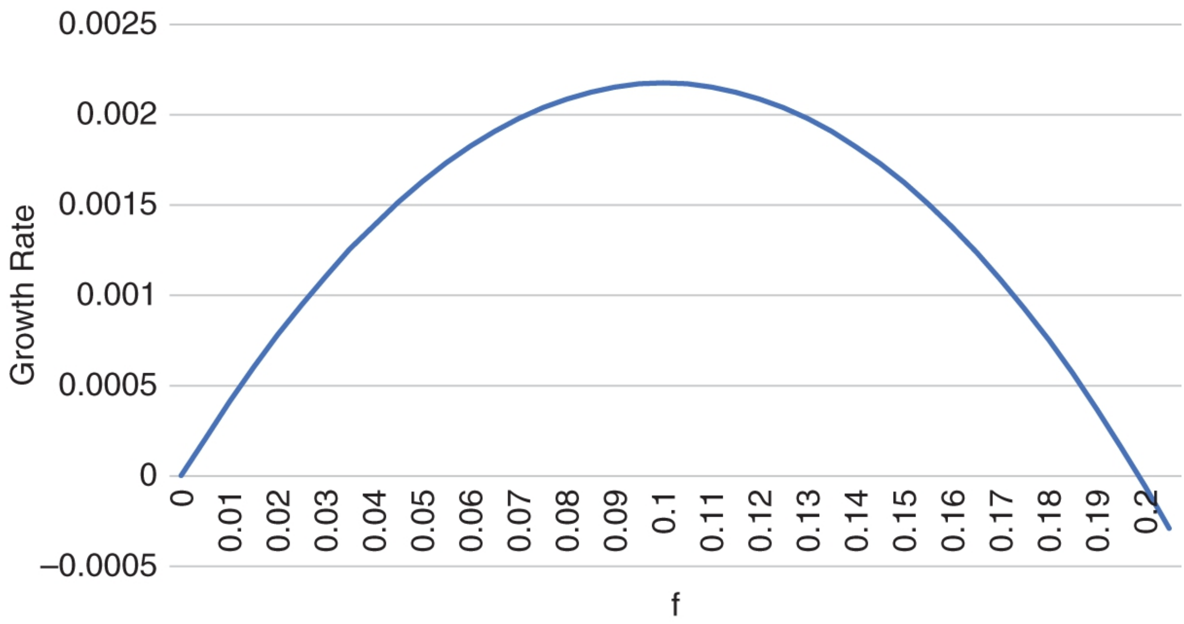


FIGURE 9.1 Growth rate as a function of f (p1=0.55,p2=0.45,W1=1, W2=−1).

Reconsidered Case

$p_1 = 0.55$

$p_2 = 0.43$

$p_3 = 0.02$

$W_1 = 1$

$W_2 = -1$

$W_3 = -3$

Here the probability, p_3, of the extreme event is low enough that we could easily misestimate it from historical data. But the implications of including this small probability are far from negligible. The growth rate is illustrated in Figure 9.2.

In this case $f_{\max}$ is 0.5, half of that in the two-outcome case. And growth rate becomes negative for $f > 0.1$. An event with only a 2% chance of occurrence could easily be missed when we estimate parameters, and if we erroneously think $p_3 = 0$, we will bet at a size that gives a negative growth rate.

This phenomenon is qualitatively similar when returns are continuous. Because this is the more relevant situation for trading, that is what we will assume when deriving methods to live with these issues.

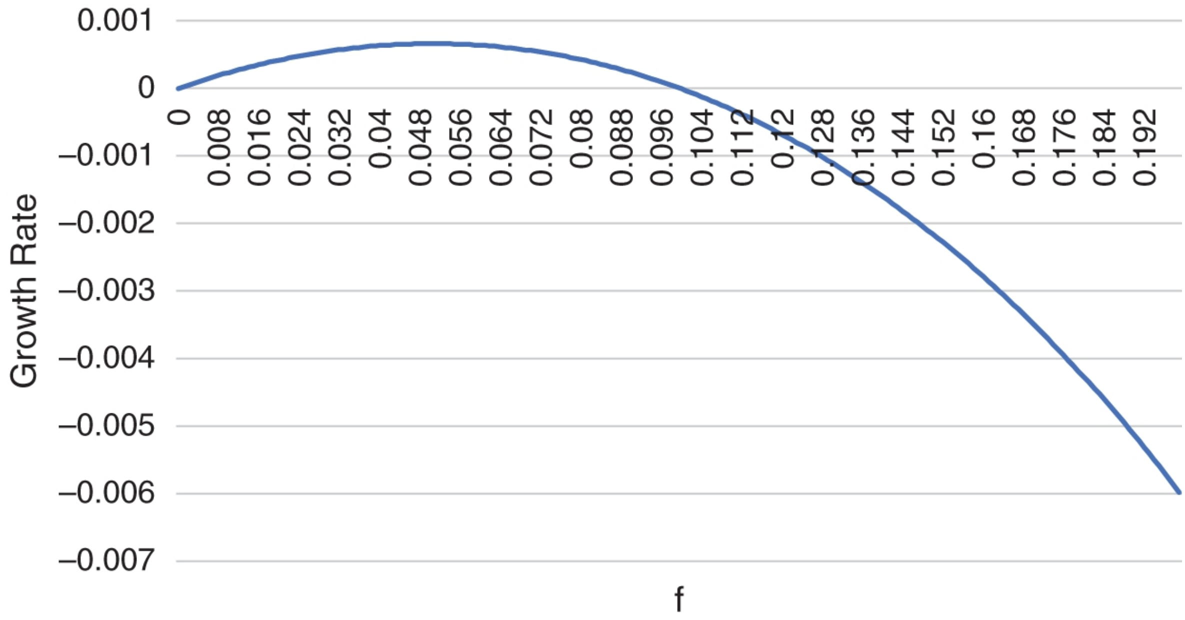


FIGURE 9.2 Growth rate as a function of f (P1 = 0.55, P2 = 0.44, P3 = 0.01, W1 = 1, W2 = −1, W3 = −3).

Non-normal Continuous Outcomes

We are interested in the case where the outcome of a trade is known to have a certain continuous distribution. We bet a fraction, f, of our wealth at the start of each period so that

$$B_n = B_{n-1} + fB_{n-1}g(X_n) \tag{9.3}$$

where B_n is the random variable giving the result of the nth trade and it has the payoff $g(X_n)$. After a sequence of n trades our bankroll will be

$$B_n = B_0 \prod_{i=1}^{n} (1 + fg(X_i)) \tag{9.4}$$

Now we take logarithms:

$$\ln\left(\frac{B_n}{B_0}\right) = \sum_{i=1}^{n} \ln(1 + fg(X_i)) \tag{9.5}$$

so

$$E\left[\ln\left(\frac{B_n}{B_0}\right)\right] = nE[\ln(1 + fg(X_n))] \tag{9.6}$$

$$= n\int \ln(1 + fg(x))\Phi(x)dx \tag{9.7}$$

where $\Phi(x)$ is the distribution function that describes the results of the trades. If we maximize over the bankroll fraction, f, we find that the optimal value is the one that satisfies

$$\int \frac{g(x)\Phi(x)dx}{1 + fg(x)} = E\left(\frac{g(x)}{1 + fg(x)}\right) = 0 \tag{9.8}$$

Applying a Taylor expansion to this equation gives

$$0 = \int g\Phi(1 + fg)^{-1} \tag{9.9}$$

$$= \int g\Phi(1 - fg + f^2g^2 + \ldots) \tag{9.10}$$

$$= \int g\Phi - f\int g^2\Phi + f^2\int g^3\Phi + \ldots \tag{9.11}$$

This can be further simplified if we note that

$$\int g(x)\Phi(x)dx \equiv \mu \tag{9.12}$$

is the payoff to a unit bet.

Further

$$\int g^2(x)\ \Phi(x)dx \equiv \mu^2 + \sigma^2 \tag{9.13}$$

$$\int g^3(x)\ \Phi(x)dx \equiv \lambda_3 \tag{9.14}$$

$$\int g^4(x)\ \Phi(x)dx \equiv \lambda_4 \tag{9.15}$$

where λ_3 and λ_4 are the third and fourth raw moments of Φ.

So, if f is small, we can truncate the series after the first term to get

$$f_{max} \approx \frac{\mu}{\mu^2 + \sigma^2} \tag{9.16}$$

And further, if μ is small, we can further approximate by

$$f_{max} \approx \frac{\mu}{\sigma^2} \tag{9.17}$$

This is the usual expression for the Kelly ratio of a trade with a continuum of outcomes, but it is only an approximation and if we are in a situation where skewness is important, a better approximation can be obtained if we keep the third term, so that equation 9.13 becomes

$$0 = \mu - f(\mu^2 + \sigma^2) + f^2\lambda_3 + \ \ldots \tag{9.18}$$

We can solve this equation to get

$$f_{max} = \frac{\mu^2 + \sigma^2 \pm \sqrt{(\mu^2 + \sigma^2)^2 - 4\lambda_3\mu}}{2\lambda_3} \tag{9.19}$$

Equation 9.19 only has real solutions if

$$(\mu^2 + \sigma^2)^2 > 4\lambda_3\mu \tag{9.20}$$

(which is a limitation of our sloppy use of asymptotics).

Further, it isn't immediately obvious which root is the correct one. Also, the case where skewness is zero leads to a singularity. We can address these issues by taking the limit as skewness approaches zero.

To do this note that

$$\sqrt{a-b} = \sqrt{a\left(1-\frac{b}{a}\right)} \approx \sqrt{a}\left(1-\frac{b}{2a}-\frac{b^2}{8a^2}\right) \tag{9.21}$$

(if b is small relative to a).

So if

$$(\mu^2+\sigma^2)^2 \gg 4\lambda_3\mu \tag{9.22}$$

we can write

$$\sqrt{(\mu^2+\sigma^2)^2-4\lambda_3\mu} \approx (\mu^2+\sigma^2)\left(1-\frac{2\lambda_3\mu}{(\mu^2+\sigma^2)^2}-\frac{2\lambda_3{}^2\mu^2}{(\mu^2+\sigma^2)^4}\right) \tag{9.23}$$

And so the negative root of equation 9.19 is approximately

$$f_{max} \approx \frac{\mu^2+\sigma^2-(\mu^2+\sigma^2)+(\mu^2+\sigma^2)\frac{2\lambda_3\mu}{(\mu^2+\sigma^2)^2}+(\mu^2+\sigma^2)\frac{2\lambda_3{}^2\mu^2}{(\mu^2+\sigma^2)^4}}{2\lambda_3} \tag{9.24}$$

which simplifies to

$$f_{max} \approx \frac{\mu}{\mu^2+\sigma^2}+\frac{\lambda_3\mu^2}{(\mu^2+\sigma^2)^3} \tag{9.25}$$

So, in order for the limiting case to agree with the Kelly fraction when trades are normally distributed (equations 9.16 and 9.17), we need to take the negative root.

From a practitioner's perspective, the important thing is to note that negative skewness decreases the optimal investment fraction and positive skewness increases the optimal investment fraction. This effect is shown in Figure 9.3.

Figure 9.4 shows the approximation of equation 9.25.

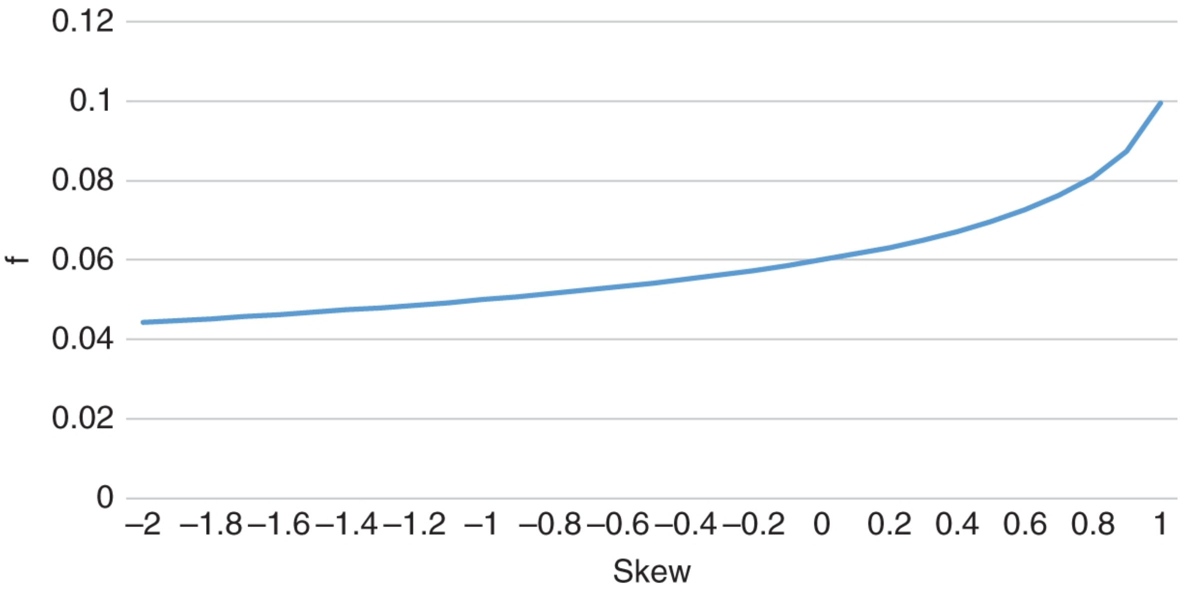


FIGURE 9.3 The optimal investment fraction as a function of skewness (return is 0.015, volatility is 0.5).

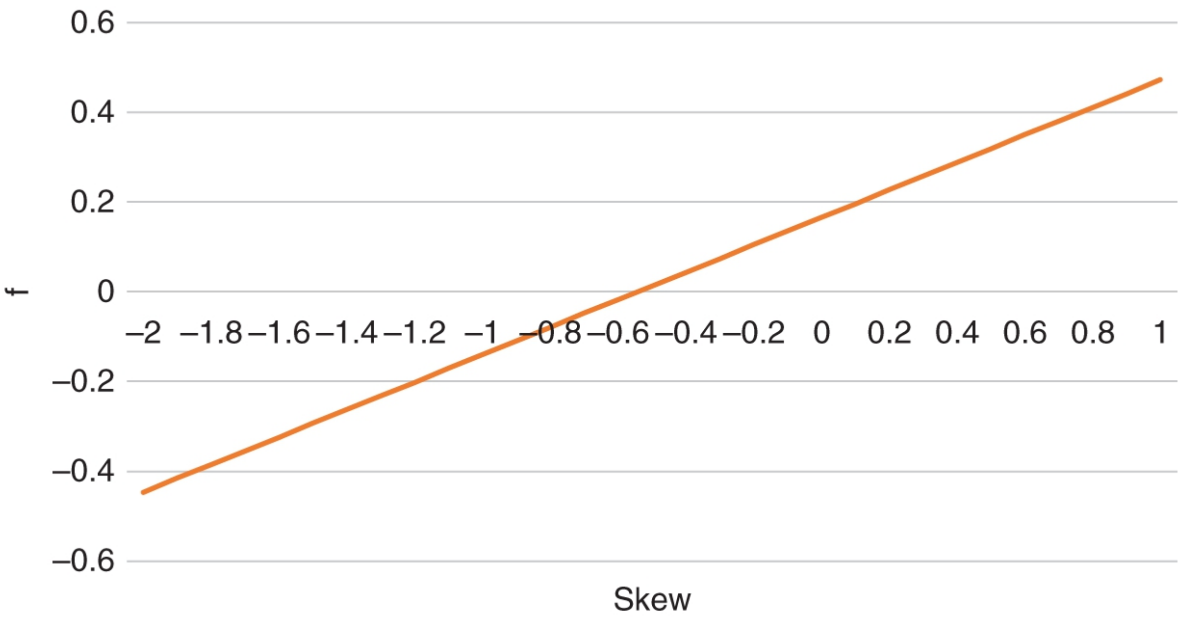


FIGURE 9.4 The approximate investment fraction as a function of skewness (return is 0.015, volatility is 0.5).

Uncertain Parameters

The value of the optimal sizing fraction will generally need to be estimated from empirical data. Because empirical data will always have sampling errors and uncertainty, the estimate of the sizing parameter will also have a degree of uncertainty attached to it.

This is well-known by professional gamblers. And to mitigate the risk of over-betting, bettors following a Kelly scheme often modify the Kelly criterion by investing only a fraction of the optimal amount. These schemes are known as "fractional Kelly" sizing. By doing this, traders accept that they will be reducing growth but will also more drastically reduce variance.

However, simply scaling the investment fraction doesn't protect against a bigger problem: the case in which the investment fraction is estimated to be positive, but the true value is negative. In this case, investing any positive fraction of the bankroll will be over-betting.

In order to estimate the chances of this happening we need the variance of the estimated Kelly criterion ratio (Sinclair, 2014). f_{max} (approximated in equation 9.17) is a statistical estimator and has an associated probability distribution.

First consider the case of normal trade results. Here the central limit theorem says that the estimators of the mean, $\hat{\mu}$, and variance, $\hat{\sigma}^2$, asymptotically have the following normal distributions, where μ and σ^2 are the population mean and variances respectively.

$$\sqrt{n}(\hat{\mu} - \mu) \rightarrow N(0, \sigma^2) \quad (9.26)$$

$$\sqrt{n}(\hat{\sigma^2} - \sigma^2) \rightarrow N(0, 2\sigma^4) \quad (9.27)$$

Alternatively, the estimation errors of mean, $\hat{\mu}$, and variance, $\hat{\sigma}^2$, can be approximated by

$$Var(\hat{\mu}) = \frac{\sigma^2}{n} \quad (9.28)$$

$$Var(\hat{\sigma}^2) = \frac{2\sigma^4}{n} \tag{9.29}$$

denoted by $f(\mu,\sigma^2)$, the Kelly ratio of equation 9.17. So the estimator is just $f(\hat{\mu}, \hat{\theta})$. The estimation errors in the mean and variance will lead to estimation errors in f.

If we define theta to be the column vector of the normal distribution's parameters, this has an estimate of $\theta = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix}$.

For IID returns, $\sqrt{n}(\hat{\theta} - \theta) \rightarrow N(0, Var(\theta))$ where $Var(\theta)$ is the variance of the estimation error of θ .

Denoting the estimator of the Kelly ratio to be $f(\theta)$ where $f()$ is now a function that estimates the Kelly ratio, we next apply the delta method (see, for example, Oehlert, 1992).

This states that the variance of a function $f(\theta)$ is

$$Var(f) = \frac{\partial f}{\partial \theta} Var(\theta) \frac{\partial f}{\partial \theta\prime} \tag{9.30}$$

$$\text{But } n\ Var(\theta) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \tag{9.31}$$

and

$$\frac{\partial f}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma^2} & \frac{-\hat{\mu}}{\sigma^4} \end{pmatrix} \tag{9.32}$$

so, evaluating equation 9.30 gives the asymptotic variance of our estimate of the Kelly ratio as

$$n\ Var(f) = \frac{1}{\sigma^2} + \frac{2\hat{\mu}^2}{\sigma^4} \tag{9.33}$$

If the trade returns are not normally distributed, we need to make use of the result (Zhang, 2007) that

$$n\ Cov(\mu, \sigma^2) = \lambda_3 \tag{9.34}$$

where λ_3 is the third central moment of the population distribution. Now equation 9.30 gives

$$n\ Var(f) = \begin{pmatrix} \frac{1}{\sigma^2} & \frac{-\hat{\mu}}{\sigma^4} \end{pmatrix} \begin{bmatrix} \sigma^2 & \lambda_3 \\ \lambda_3 & 2\sigma^4 \end{bmatrix} \begin{pmatrix} \frac{1}{\sigma^2} \\ \frac{-\hat{\mu}}{\sigma^4} \end{pmatrix} \tag{9.35}$$

$$n\ Var(f) = \frac{1}{\sigma^2} + \frac{2\hat{\mu}^2}{\sigma^4} - \frac{2\hat{\mu}\lambda_3}{\sigma^6} \tag{9.36}$$

It isn't possible to find the variance of the sizing fraction given by equation 9.19, because the variance of the skewness would need to be evaluated for the particular distribution the results were drawn from. The best we can do in general is to measure the empirical skewness, calculate the sizing ratio using equation 9.19, then estimate the variance around that value by using equation 9.36.

We now use an example of real trade results to show the importance of including estimation error in trade sizing. The trade results are from a proprietary short volatility strategy. It is somewhat typical of many such strategies in that it has a positive expected value but a large negative kurtosis. The summary statistics for these trade results are given in Table 9.1 and the distribution of results is shown in Figure 9.5.

We can rearrange (and slightly modify) equation 9.36 to give an explicit expression for the estimated standard deviation of the Kelly ratio.

$$sd(f) = \sqrt{\frac{\frac{1}{\sigma^2} + \frac{2\hat{\mu}^2}{\sigma^4} - \frac{2\hat{\mu}\lambda_3}{\sigma^6}}{n-1}} \tag{9.37}$$

where the denominator of $n - 1$ is due to applying Bessel's correction.

TABLE 9.1 Summary Statistics for the Option Trade

Sample size	1000
Mean	$0.059
Standard deviation	$1.137
Skewness	($6.199)

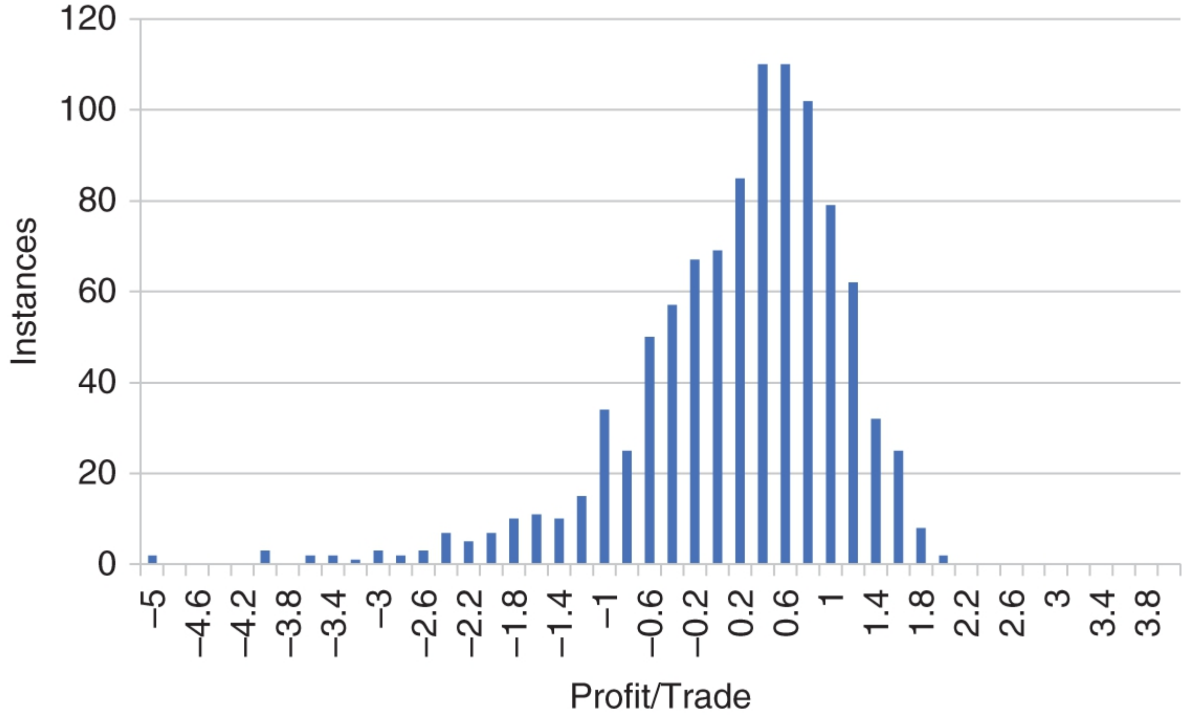


FIGURE 9.5 The distribution of the option trade results.

Because of the central limit theorem, we know that the distribution of f is normal so we can calculate the probability that f is actually below any critical value f^*.

$$Prob(f < f^*) = Z\left(\frac{(f - f^*)\sqrt{n-1}}{\sqrt{\frac{1}{\sigma^2} + \frac{2\hat{\mu}^2}{\sigma^4} - \frac{2\hat{\mu}\lambda_3}{\sigma^6}}}\right) \quad (9.38)$$

where *Z* is the cumulative distribution function of the normal distribution with mean of *f* and a standard deviation calculated from equation 9.38.

Equation 9.17 gives the Kelly ratio as 0.046, but equation 9.37 tells us that the standard deviation of this point estimate is 0.031, so our point estimate is only 1.4 standard deviations above zero. There is an 7% chance that the true Kelly ratio of the population is less than zero.

Having an expression for the sampling distribution also enables us to estimate the chance that we are over-betting so much that our growth rate is negative. This case corresponds to the true value of *f* being roughly less than half the estimated value. Equation 9.38 tells us this is 25%.

TABLE 9.2 Fractional Schemes Corresponding to Various Probabilities of Over-Betting

Chance of Over-betting	Corresponding Benchmark	Kelly Scale Factor
0.1	0.0022	0.0480
0.15	0.0104	0.2301
0.2	0.0169	0.3748

TABLE 9.3 Fractional Schemes Corresponding to Various Probabilities of Over-Betting When Setting Skewness of the Trading Results to Zero

Chance of Over-betting	Corresponding Benchmark	Kelly Scale Factor
0.1	0.0092	0.2054
0.15	0.0161	0.3574
0.2	0.0215	0.4782

This leads us to a complimentary way to use the information. We can use equation 9.38 to solve for a benchmark given that we want a certain chance of over-betting. For example, we have just seen that using a benchmark of half the measured Kelly fraction (i.e., betting at "half-Kelly") still implies a 25% chance that we will be over-betting. Table 9.2 shows the probabilities of over-betting for various fractional Kelly schemes.

So, in order to introduce a margin of safety we would need to scale the measured Kelly ratio by a considerable amount. This is in line with the practice of professional gamblers. Much of this need for scaling is due to the presence of negative skewness. If the returns were normally distributed, the scaling could be reduced. This is shown in Table 9.3.

Kelly and Drawdown Control

Even after calculating and allowing for our measurement uncertainty, it is likely that investors will find that investing the full Kelly fraction leads to results that are unpalatably volatile. And the more edge there is, the higher the Kelly ratio will be and so the higher the volatility will be. Good trades are the most volatile.

The standard way to mitigate drawdowns is to trade using a fraction of the Kelly ratio. In this case, both growth rate and volatility will drop. If we trade at a fraction, f, of the Kelly ratio, the growth rate is

$$GR(f) = \left(f - \frac{f^2}{2}\right)\frac{\mu^2}{\sigma^2} \tag{9.39}$$

This is maximized for $f = 1$ and drops as the investment fraction is reduced from this (Figure 9.6).

Measures of risk and drawdown also decrease as the investment fraction drops. The ratio of growth rate to drawdown is shown in Figure 9.7 for the case where $\mu = 0.05$ and $\sigma = 0.3$.

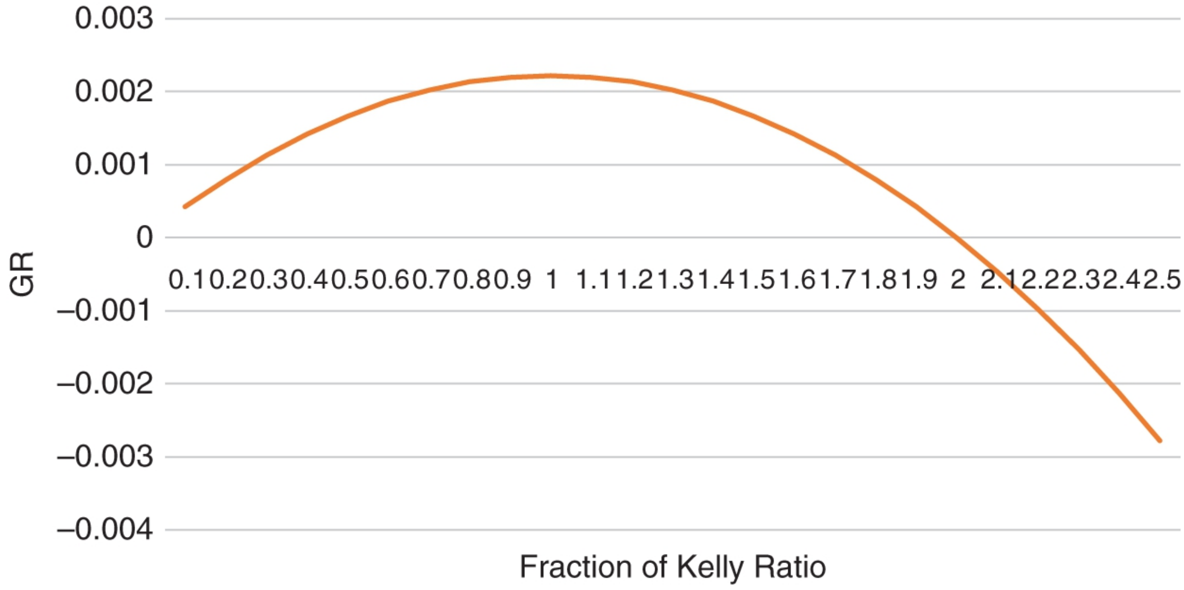


FIGURE 9.6 The dependence of growth rate on the fractional Kelly ratio.

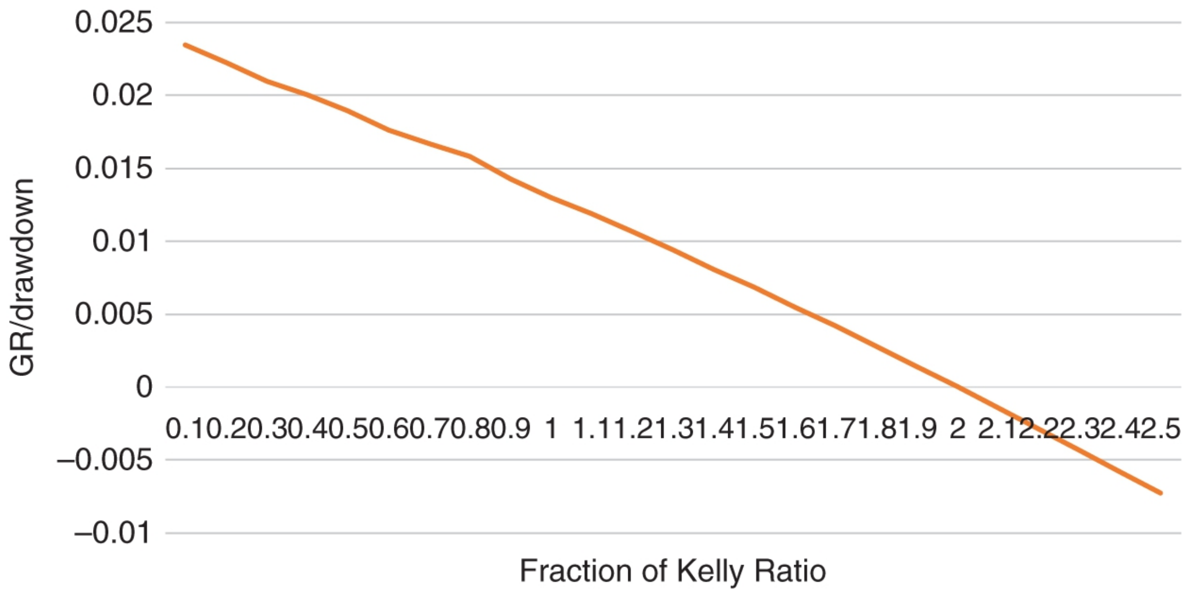


FIGURE 9.7 The growth rate to drawdown ratio as a function of the scaling factor.

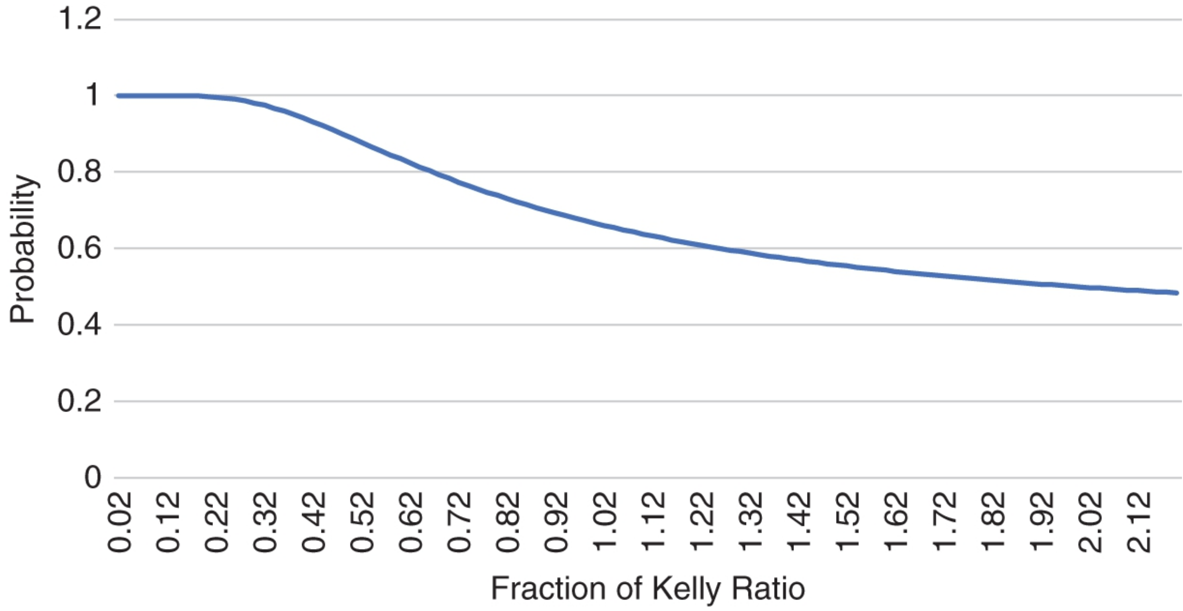


FIGURE 9.8 The probability of reaching 200% before being stopped out at 0% when trading at fractions of the Kelly ratio.

So, if drawdown is the primary risk factor, you should be very cautious indeed.

Another way to quantify the effect of changing the investment ratio is by looking at the chance of hitting a lower barrier (possibly the stop level) before reaching a target. If the stop percentage level is A, and the target is B, the probability is given by

$$P(A, B) = \frac{1 - A^{1-\frac{2}{f}}}{B^{1-\frac{2}{f}} - A^{1-\frac{2}{f}}} \tag{9.40}$$

(Interestingly, neither the return nor volatility of the return stream is relevant. A "good" trade means the level will be hit earlier but doesn't change the relative probability.) The dependence of the probability on the scaling fraction, f, is shown in Figure 9.8 for the case A = 50%, B = 200%.

A very conservative trading size will raise the probability of a good outcome because the volatility associated with Kelly betting is dampened to the point that hitting a stop is unlikely.

However, a very small scaling factor also means that the expected time to hit the goal increases. The expected exit time is given by

$$E[T] = \frac{1}{GR} ln\left(\frac{B^{P(A,B)}}{A^{P(A,B)-1}}\right) \quad (9.41)$$

Expected exit time as a function of f, for the case A = 50% and B = 200%, and μ = 0.05 and σ = 0.3. is shown in Figure 9.9.

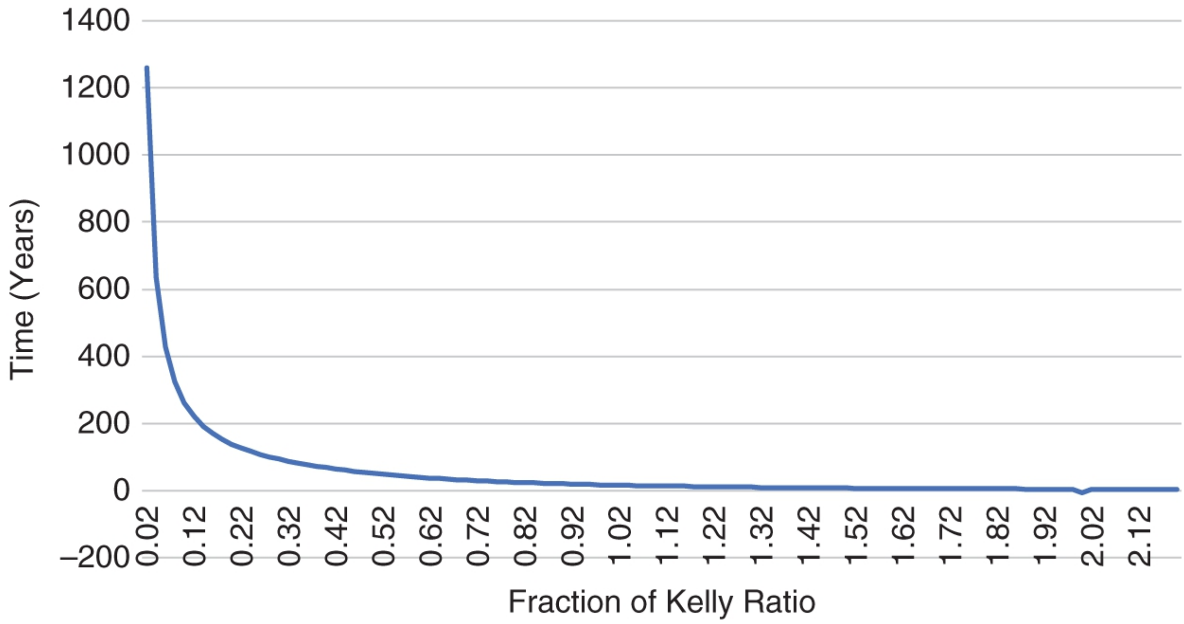


FIGURE 9.9 The expected time to reach 200% before being stopped out at 50% when trading at fractions of the Kelly ratio for the case where μ = 0.05 and σ = 0.3.

(Note that the average exit time is monotonically decreasing in f, even decreasing when expected growth rate is negative. At this point, you will probably get stopped out, but you also might just get lucky.)

In terms of risk mitigation, it is easy to make a case for trading at only a fraction of Kelly. However, doing this will also affect the good points of the method. Further, due to sampling issues when we estimate the parameters, we may be betting in a negative expectation game. We can mitigate these issues by combining the Kelly concept with a stop.

The Effect of Stops

To many traders, the use of stops is seen as an essential part of risk control and money management. And usually they take the utility of stops to be self-evident. "How can you go broke if you limit your losses?" "Cut your losses and let your profits run."

"Losers add to losers." But the effect of stop use is quite complex. In this section I will explain where stops help, where they don't, their effect on profitability, and, if you are using them, where to set them.

First, we examine what the use of stops does to the distribution of our trading results. A hypothetical trade result example is shown in Figure 9.10. The trade results are normally distributed with a mean return of 10% and a standard deviation of 15%.

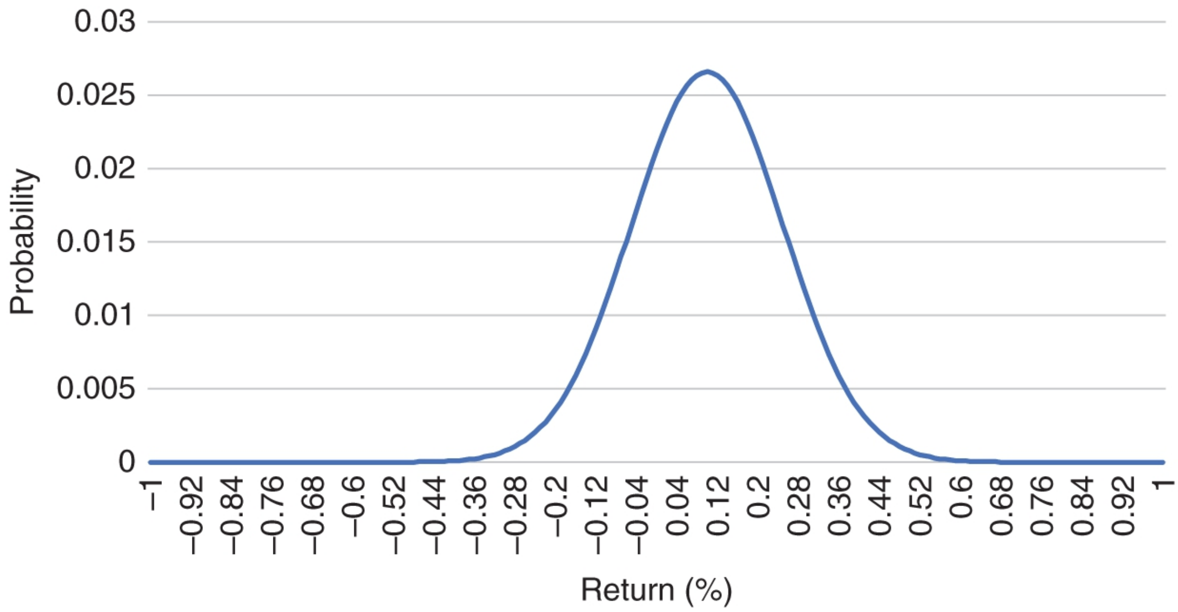


FIGURE 9.10 The return distribution of our trading strategy.

However, we can also see that a significant number of trades were losers (here nearly 11% of trades will lose more than 15%). A natural thought would be to introduce a stop loss to somehow "cut off" the left-hand side of the distribution.

It is tempting to think that use of a stop simply truncates the downside of the distribution by capping losses at a certain level. Redrawing the distribution to reflect this intuition situation gives us Figure 9.11, where a stop has been introduced at the 15% down level, so all loses are capped at 15%.

But a little thought will be enough to see that this distribution isn't possible. The trades that get stopped out don't just disappear. Their results still have to be accounted for (mathematically, the integral of the probability density function must still be 1). So we next surmise that these trades cluster around the level of the stop. But this still misses an important point.

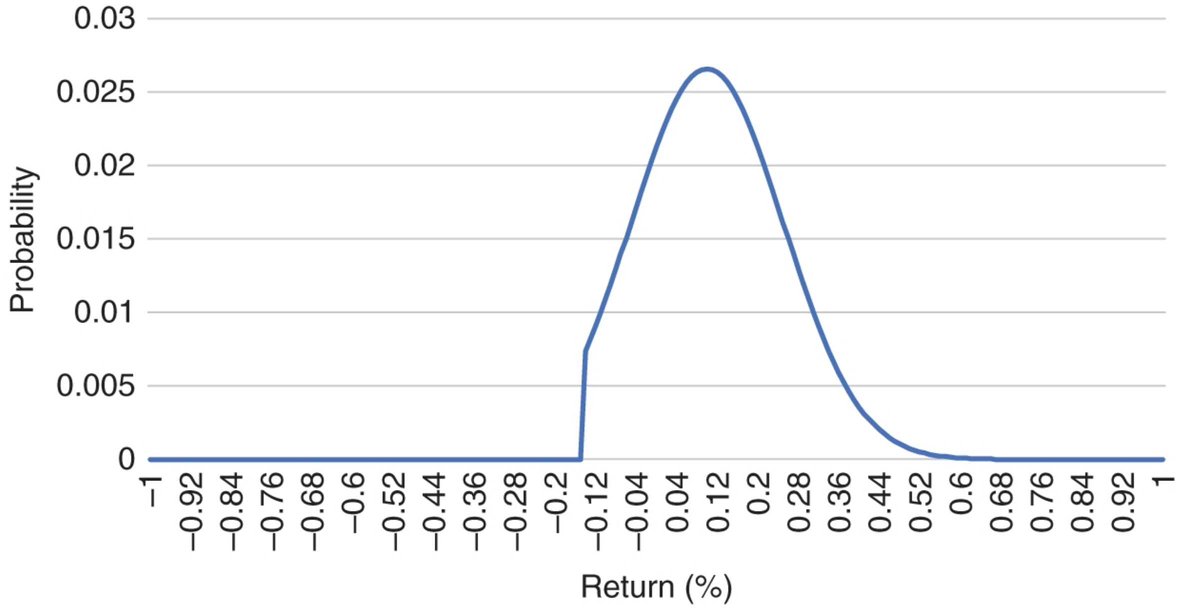


FIGURE 9.11 The hoped-for distribution when a stop has been added.

Many trades that at the end of the period will be small winners and will have been stopped out beforehand. Note that the large winners will still remain because these largely consist of trades that started as winners and never looked back, but the presence of a stop will drastically reduce the number of small winners. This is the hidden cost of using a stop.

To see the exact effect, I simulated 1,000 GBM paths that represented checking the performance of the trade once a day for a year. Again, the expected final return was 10%, the standard deviation was 15%, and a stop was placed to cap losses at 15%. Results are shown in Figure 9.12.

There are several things to note here. First, the average return is negatively affected by using stops. The "unstopped" investment has a mean return of 10% (as we planned for the simulation), but when we add a stop this drops to 9.6%, and the median is now 8.6%. This is the mathematical inevitability of assuming that the (unstopped) results follow the normal distribution. By adding a stop, we eliminate the big losers (in this case of greater than 15%) but we also eliminate trades that would have later recovered above the stop level. And because the normal distribution has more density around the mean than it does in the wings, there are more of these marginal trades than there are big losers.

This analysis has been for a fixed stop, set at a given distance from our entry price. The other type of stop is the trailing stop, which is set so it stays a certain distance from the highest amount the investment has made. A trailing stop is a very comforting strategy. It protects us from the painful experience of seeing our winners turn into losers.

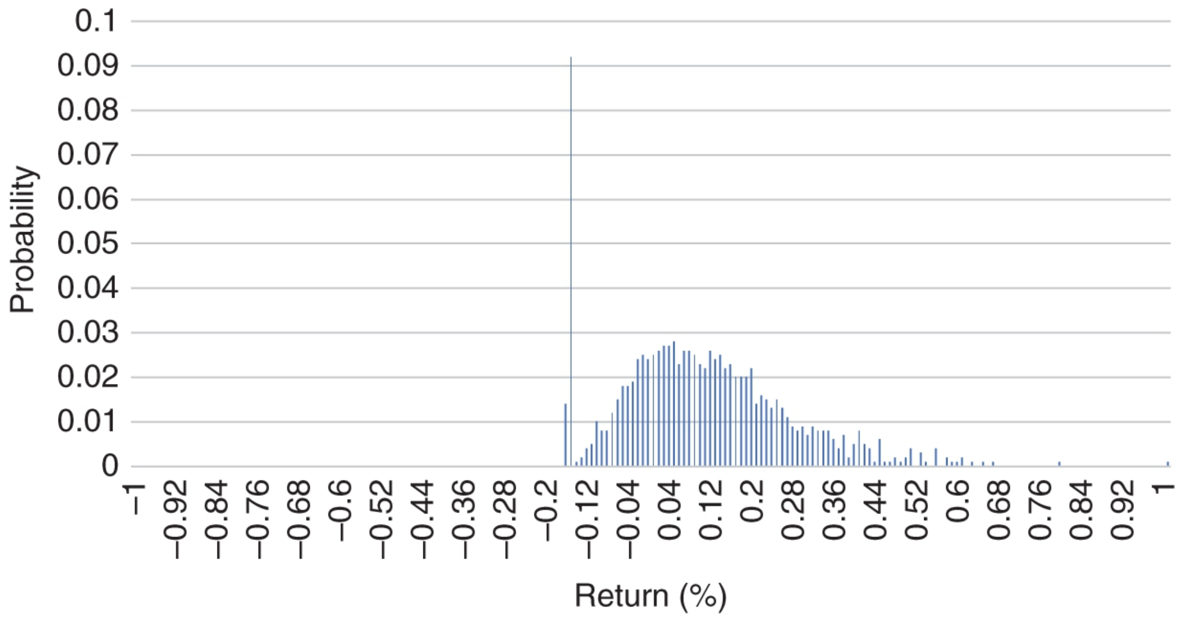


FIGURE 9.12 The true distribution when a stop has been added.

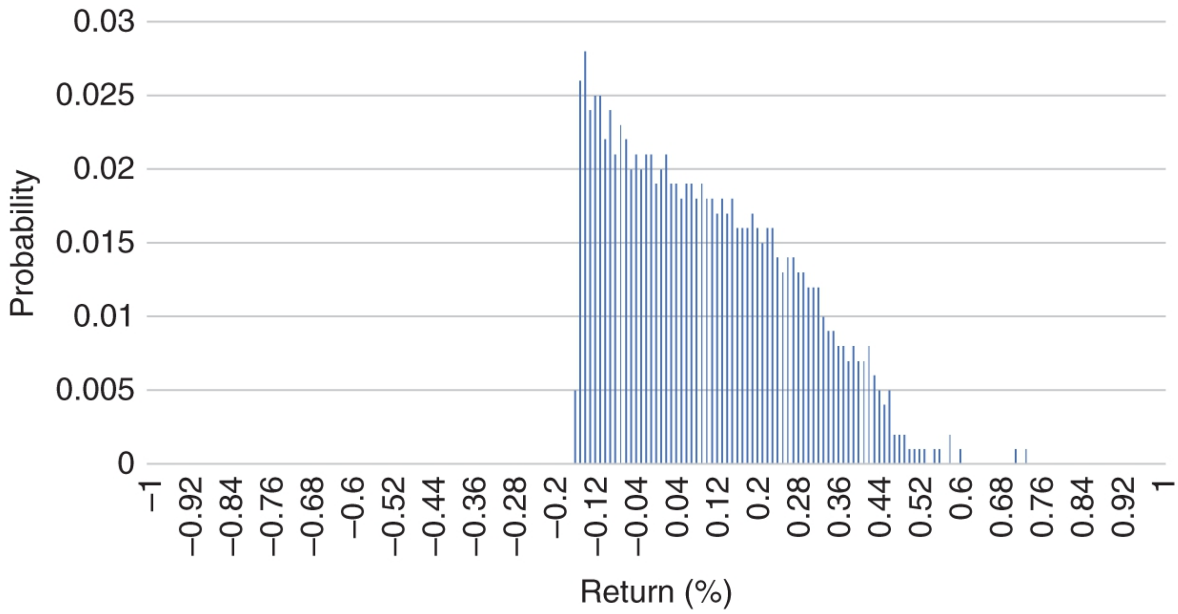


FIGURE 9.13 The return distribution of the simulated trade when using a trailing stop.

However, the trailing stop costs even more than the fixed stop. This is because some positions that would be large winners don't get the chance to fully realize their potential. When using fixed stops, some investments benefit by getting away from the stop and having the chance to develop. Trailing stops are always in play.

Adapting our previous simulation confirms this. The trailing stop is always 15% below the high of the trade. The addition of a trailing stop lowers the mean return of the trade to only 9.4%, and the median is 8.0%. This is lower than the unstopped return and also lower than when employing a fixed stop. Figure 9.13 shows the distribution when we employ a trailing stop. The distribution in this case is now totally different from that of the unstopped investment.

Stops don't just stop losses. They drastically change the shape of the return distribution and can lower the average return. Adding stops won't transform a losing strategy into a winning strategy. The only reason that we would add a stop is that we prefer the shape of the stopped distribution, that is, we prefer to trade lots of small losses *and* fewer small wins for some large losses.

Although this is true, the real world is far more complex. Returns are not normally distributed and the results will change over time. Also, there is no reason a trader need only be interested in maximizing returns. Safety and risk control are important for both financial and psychological reasons.

Obviously, many traders like using stops. Indeed, some insist that stops are absolutely essential and that their appropriate use is a good predictor of overall, long-term success. Given that most tests show that stops cost money, what is it that these traders are thinking?

Let's take a quick look at some of the common arguments given for using stops:

> **Stops limit losses.** For any given trade this is trivially true. Ignoring slippage and trading costs, we can't lose more than a certain predefined amount on any investment when we use a stop. But, as we have seen in our simple simulation, in aggregate this is an illusion. Using stops will lower returns in the long term so this reason at best needs more justification and is probably just incorrect.
>
> **Stops are a form of discipline.** It is true that if you always use stops, then you have displayed discipline. But discipline needs to be about applying a sensible methodology

consistently, not just doing *something* consistently. If your idea has negative return, executing it diligently will result in greater losses than doing it haphazardly.

Using stops means all trades risk the same amount. Some traders have the idea that it is important to risk a certain set amount on each investment (1% or 2% is often the amount given). This is generally false. Different trades, even within the same strategy, will have different projected risks and returns. It is best to take this into account when trade sizing. Not doing so will lead to lower total returns and unnecessary risk. Further, trade sizing and risk control are a different issue from setting a loss for each single position.

A stop is a predefined exit. The trade will be stopped out when we don't want to be in the position anymore. It should be obvious that exiting a position when we don't want to be in it is an excellent idea. If we use stops to formalize this, then they are perfectly sensible. But we need to think more about what this means. If we are exiting a position purely because the price has moved by a given amount, then we are assuming the trade has positive autocorrelation: the move that has already happened is predictive of a future move. Price-based stops are a trend-following system. So they make a good deal of sense if we are explicitly betting on momentum. Conversely, they don't make a lot of sense if we are trading something that we expect to revert. In this case we will be exiting trades at points where we see potential for future profit. We should never exit a position when a trader with the same strategy and no position would want to enter.

Rephrased, the reason we should get out of a position that has moved against us is if, and only if, we expect the move to continue. The loss has already been incurred; we need to think about our current risk, not the sunk cost of the incurred loss. And if we are only basing the stop on the price action, we are saying price direction alone determines future price direction, that is, we are trend following.

A position should be exited when we are wrong. Sometimes this will coincide with losing money. In this case a stop is harmless. But sometimes losing money corresponds to situations for which we have more edge. Here a stop is actively damaging and contrary to the idea behind the strategy.

Stop Placement

Once we have decided we want to use a stop, we still need to choose where to put it. There are two aspects to this: game theory and statistics.

The role of game theory is often emphasized although it is the less important consideration. The idea is to avoid placing stops in “obvious” price levels such as around the highs or lows of the previous session, key technical levels, or round numbers. Other traders could take advantage of this predictability. For example, imagine the market was quoted at 98 bid and 99 offered and a trader knew there were likely to be buy stops set at 100. He could get long at 99, print a trade at 100, and set off the stops. This would drive the price higher, giving him an instant profit.

This sort of tactic was common on trading floors, but the higher liquidity and different sociological structure of electronic markets means the idea is less relevant than it once was. It probably doesn't hurt to avoid placing orders at these levels, but it almost certainly doesn't matter much.

The more important consideration is choosing a stop price that correctly balances risk and costs. If a stop is too far away, it isn't doing much to reduce risk and if it is too close, it will get hit too often, increasing transaction costs and not giving trades enough of a chance to become winners.

There have been many attempts to establish a theoretical basis for stop placement. Unfortunately, the results are highly dependent on both the assumed price process and the trader's utility function. For example, a risk-neutral trader who is long an instrument with a positive drift will never use a stop.

We can never be completely sure about either the price process or parameters and our utility function is likely to be a function of many things other than just wealth. So, these theories aren't helpful.

There have also been a number of empirical tests of the efficacy of stops. The problem is that these are tests of the interplay between a stop and a particular strategy. They can possibly help as guides, but they will never apply particularly well to a different strategy or set of instruments. Some such studies are those by Lei and Li (2009), Clare et al. (2013), and Han et al. (2016).

Instead we need to use a completely empirical method for each strategy we are considering. Here, the only theoretical assumption is the usual one that the past return distribution is predictive of the future. It is never a great idea to base a trading decision only on data analysis but at this point I don't think that there is a better way to choose stop parameters than just testing various ideas.

Incorporating Stops into the Kelly Criterion

Conceptually, we split the trading account into two parts: a risky part that we trade according to the Kelly criterion and the riskless part that we hold in cash. If we never transfer profits from the risky part to the safe part, we are using a fixed stop. For example, we split our $100 account into $80 cash and an active trading subaccount of $20. We will never be able to lose more than $20 (in theory we can't even lose all of that because we will be trading proportionally) but we can suffer larger peak-to-trough drawdowns if we first make money and then lose those profits as well as the original $20. This idea has been studied by Grossman and Zhou (1993), Cvitanic and Karatzas (1995), and Browne (2000b).

Consider trading our $100 in two situations with identical growth rates. Our trading strategy has a mean return of 5% and a volatility of 30%. In one case we trade at one-quarter of the Kelly ratio, which according to equation 9.39 gives a growth rate of 0.0061. So, after one year the expected account value is $100.61, and the volatility is 0.75, or $7,546. The complete distribution of results is shown in Figure 9.14.

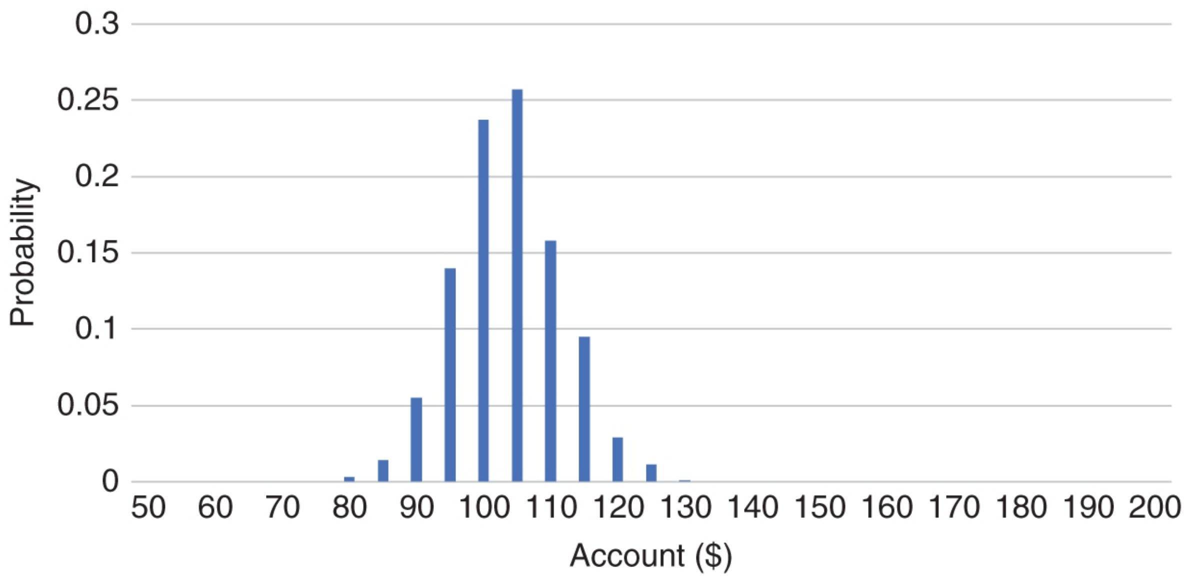


FIGURE 9.14 The distribution of the final account after 10,000 simulations of a GBM where μ = 0.05 and σ = 0.3 when using a quarter of the Kelly ratio.

Full Kelly gives a growth rate of 0.139. To achieve the same expected account value using a smaller subaccount, W_r, which trades at full Kelly, we solve the equation

$$Profit = W_r exp(GR) - W_r \tag{9.42}$$

which gives W_r = \$43.7. Initial volatility is 0.3 or \$13. So, the subaccount method gives the same growth rate but with a higher dollar volatility. This doesn't seem promising. But things are better when we consider the entire distribution. We have given ourselves some downside protection while still retaining the possibility of the extreme growth that Kelly can give. This is shown in Figure 9.15.

Another method is to adjust the amount in our safe subaccount to stay a constant percentage below the peak. We will be using a trailing stop on the account value. This time we use a trailing stop level of 43%. The distribution of results from using this method are shown in Figure 9.16.

From the figures the higher volatility of the subaccount method is largely due to the positive upside. Summary statistics are shown in Table 9.4.

Using a percentage-based trailing stop retains many good features of the aggressiveness of the Kelly criterion. It does better both on average and in the extreme cases. The great results are those that start well and continue well, never being affected by the stop. And when the stop does come into play, the trailing stop idea protects capital nearly as well as the other two, more conservative, methods. Trading according to full Kelly on a percentage-based subaccount seems to be the best idea no matter what criteria are considered.

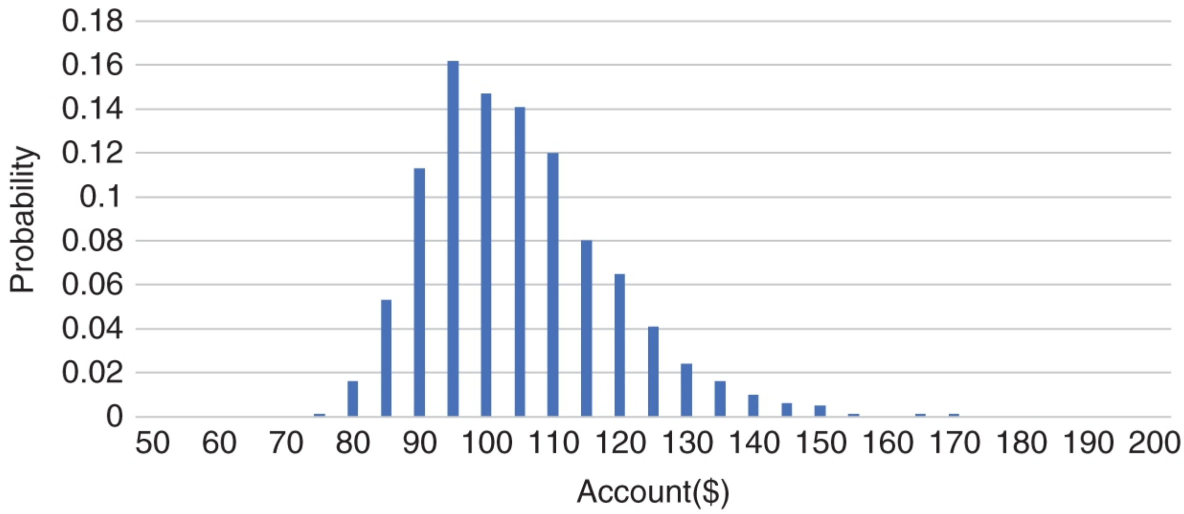


FIGURE 9.15 The distribution of the final account after 10,000 simulations of a GBM where μ=0.05 and σ=0.3 when only trading a subaccount that begins at $43.70 at full Kelly.

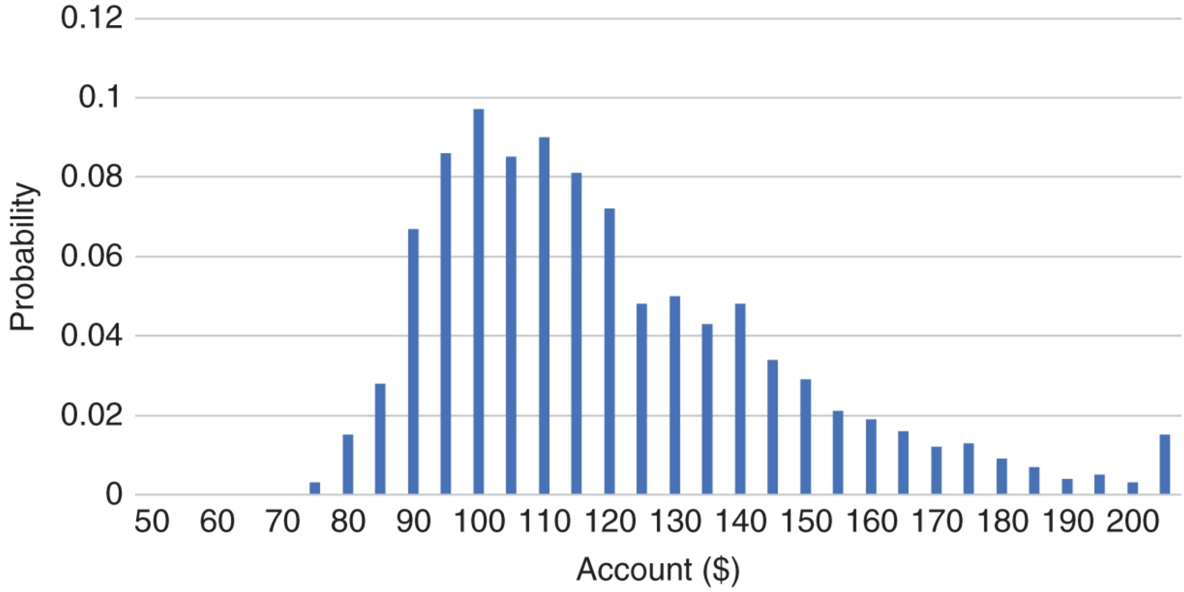


FIGURE 9.16 The distribution of the final account after 10,000 simulations of a GBM where μ = 0.05 and σ = 0.3 when only trading a subaccount of 43.7% of the total at full Kelly.

TABLE 9.4 A Comparison of Trading at Quarter Kelly and Trading Full Kelly in Subaccounts

Statistic	Quarter Kelly	Subaccount with Fixed Maximum Loss	Subaccount with Trailing % Loss
Average	$101.70	$102.20	$117.13
Minimum	$78.30	$68.75	$69.36
Maximum	$129.50	$187.60	$305.60
10th percentile	$91.70	$86.30	$89.30
90th percentile	$111.10	$120.90	$153.40
Average maximum drawdown	9%	12%	12%
Maximum drawdown	26%	38%	30%

Conclusion

The Kelly criterion is only optimal in the sense that it maximizes growth. There are many other things a trader could be interested in maximizing. However, using a modified Kelly scheme that explicitly includes higher-order moments in the return stream, sampling errors and drawdown control is a consistent, sensible method that is vastly preferable to either an ad-hoc strategy or one that is based on mistaken ideas.

Summary

- Adjust the Kelly ratio to account for skew. Trade negatively (positively) skewed strategies and instruments smaller (larger) than suggested by the standard Kelly criterion approximation.
- Quantify the uncertainty in the estimate of the Kelly ratio. Scale the raw ratio to give a prespecified chance of it being non-negative.
- Split the trading account into an untraded percentage and a traded percentage. Apply full Kelly to the risky part of the

portfolio.

CHAPTER 10
Meta Risks

Most discussion of the risks related to options focuses on market risks. Obviously, it is important to mitigate these risks, but they *can* be mitigated. They are entirely under the control of the trader. If a trader blows up because of a market move, it is due to either ignorance or overconfidence.

There is a far more dangerous type of risk. These are risks from outside the market. Options don't exist in a vacuum. They are traded as part of an economic system that contains moral, legal, and governmental risks. These are the risks that are hardest to anticipate, but it is still important to learn as much as possible about them and to take any precautions we can.

Currency Risk

All traders will have currency risk. You can't execute stock and option trades with anything other than fiat currency. Even if we ignore exchange rate risk (because your currency of determination might be the currency of the trades), fiat currency is risky due to inflation. For example, Zimbabwe has had a stock exchange since 1993. In the same period the inflation rate has reached up to 98% *a day*. It is respectable if a trader averages 0.1% a day. No one is good enough to keep up with hyperinflation.

The worst examples of historical inflation have been in war zones or places experiencing massive political instability: Greece in 1944, Hungary in 1946, Yugoslavia in 1994, Germany in 1923, China in 1949, and Venezuela in 2018. It would be simple to write, "Avoid living in such situations." But this is generally much easier said than done. A lot of people would love to have left these places but didn't have the means to do so. And even in the very stable country of New Zealand, the inflation rate reached 17% during the 1970s. It is tempting to think that these periods are confined to the past but that seems very unlikely. Inflation cannot generally be avoided. The only thing traders can do is be aware of it, and when active trading isn't producing real returns stop and invest in something else.

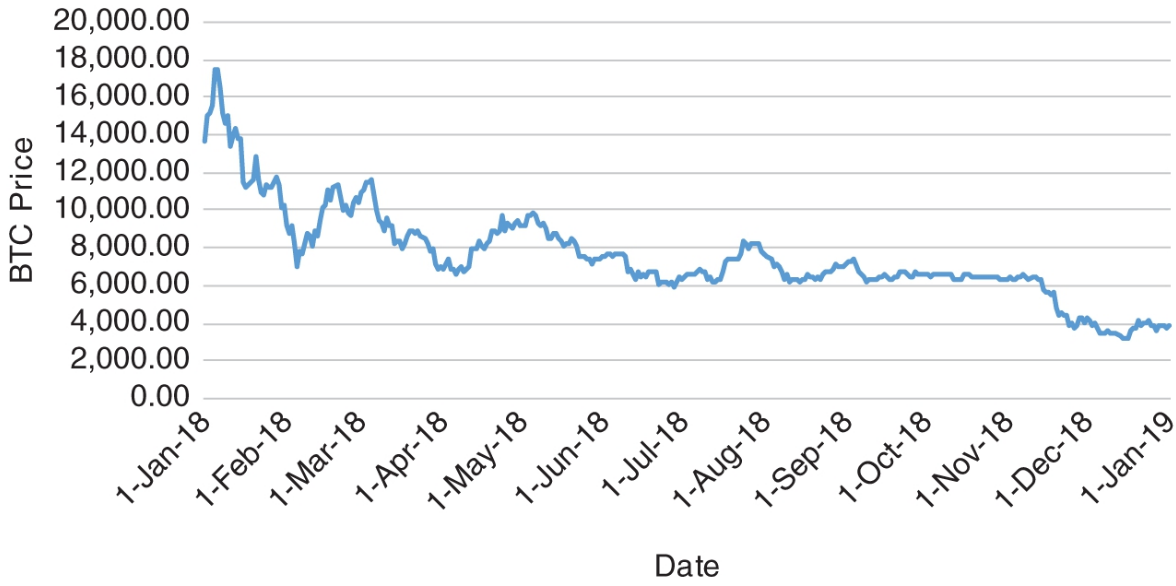


FIGURE 10.1 The price of Bitcoin in USD in 2018.

Cryptocurrencies have their own problems. Although they are not subject to the inflationary spending of a government, they are exceptionally volatile. In 2018, Bitcoin had a low of $3,190 and a high of $17,700 (Figure 10.1). Average annualized volatility was 68%. Typically, major currencies have cross-rate volatilities in the single digits. For example, in 2018 the volatility of euro/USD was 7%. And the volatility of USD/yen was 6%.

And although cryptocurrency exchanges are regulated as businesses in their country of incorporation, they are not run with the same degree of care and oversight that we can expect from a traditional exchange. A good example of this occurred in 2018 when the owner of the Canadian exchange QuadrigaCX died. He was the only person who knew the passwords to the exchange's cold storage. As of February 2019, customers were unable to access US$190 million.

Also, no significant exchange will let you pay for anything in Bitcoin. Almost no one will. Bitcoin fails as a currency both because it is far more volatile than what you will be buying with it and because it is not a widely accepted medium of exchange. As of January 2019, the total value of all the Bitcoin in the world was US$64 billion. In comparison the total amount of USD was 16 trillion. To become a legitimate currency Bitcoin (and cryptocurrencies in general) would need to both vastly increase in value and vastly drop in volatility.

Finally, many financial institutions want nothing to do with any cryptocurrencies. Quoting a senior banker in the clearing division of a major US bank, “If you have anything to do with crypto, this conversation is over.” Part of this reluctance is due to the current uncertainty of the regulatory environment. Some more is because banks are looking to develop private versions of the blockchain. Finally, although it seems that the blockchain could eventually help KYC (“know your customer”) due diligence, banks are currently more worried that the verifiability of each transaction in the blockchain can directly tie them to criminal activity.

Lessons

- Currency risk is unavoidable but can be diversified.
- Cryptocurrencies are speculative instruments and not currency in the normal sense.

Theft and Fraud

Thieves steal things that are worth money. And, as nothing is more clearly worth money than money itself, the financial industry provides good targets for thieves, con artists, and hucksters. It is impossible to be completely protected from this risk. Everyone who trades or invests will rely on other people. Even those who execute their own trades on an exchange will still have risk associated with their custodian and clearinghouse.

The good news is that most criminals and fraudsters are not masterminds. Many simply rely on the fact that if they are charming, no one will do the necessary due diligence.

The first thing is to look at the people involved. Everything a fund does is because of the people running it. You are not investing *in* a fund. You are investing *with* people. It is important to know who the managers are, their backgrounds, their experiences, their levels of involvement, and the amount, both professionally and financially they have invested. Do they have criminal convictions? Have they been censured by regulatory bodies (leaving aside floor traders being fined by the exchange for a fight over a spot or something)? Do they treat your money even more respectfully than they treat their own? For that matter, do they have a significant amount of their net worth invested in the fund? Do only one or two people in the fund have access to the account statements? Often you can get a lot of insight about a person's character just by asking around. The trading world is a small place and someone with a dubious reputation quickly becomes known (even if this isn't discussed on the record).

Any legitimate organization will be pleased an investor asks these things. For the manager, it is both understandable and makes dealing with the customer easier. It is always better to be in business with sensible people.

Further, any legitimate fund should have a third-party administrator. This is a separate firm that has access to bank statements and brokerage reports. They verify positions and returns and generate the monthly reports sent to clients. Don't accept an excuse such as, “We are just a startup and can't afford a third-party administrator.” It is non-negotiable. Had Bernie Madoff's victims insisted on the oversight of an independent body, they would have immediately discovered that Madoff had done no trades and held no positions.

More generally, a lot of financial disasters can be traced back to the victims not knowing about their inventory and hence their risk. Many of the catastrophic and infamous derivatives losses have been caused when an institution or investor did not know what their position was.

Example One: Baring's Bank

Baring's bank was the second oldest merchant bank in the world. It was founded in 1762 by Sir Francis Baring. In 1990, Nick Leeson was employed in the Singapore branch of Baring's. He was authorized to act as a floor broker and to arb the Nikkei futures between the Singapore and Osaka exchanges. All brokers have an “error account.” Any trades that are in some way mistakes get put into the error account. This is meant to be for accounting purposes. As soon as the error is found the broker is supposed to exit the trade by liquidating the position. Leeson had a small error that led to a position of 20 Nikkei futures. Leeson changed the account number so the head office didn't see reports and began to trade in the account to recoup a loss of £20,000, eventually accumulating a large position in short Nikkei straddles.

On January 17, 1995, a major earthquake hit Kobe and, over the next four days, the Nikkei dropped 16%, making the short straddles enormous losers. Leeson started to make even bigger trades trying to recoup losses. This didn't work. He lost £827 million (the equivalent of about US$2.5 billion in 2018). On February 23, he absconded to Malaysia and sent a confession note (consisting only of "I'm sorry") to London at the same time. Barings was insolvent and was eventually sold to ING for one pound.

Lessons

- Never let a trader or fund be both a trader and a risk manager.
- A trader shouldn't have the ability to create accounts.
- Don't invest in products or strategies that you don't understand. The London office believed Leeson's falsified profit numbers while someone familiar with the Nikkei arb would know that the reported numbers were impossible.

Example Two: Yasumo Hamanaka, aka "Mr. Copper"

In 1995 (a big year for rogue traders) the head copper trader at Sumitomo bank, Yasumo Hamanaka, lost about US$1.8 billion (roughly $3 billion in 2018 terms). Over a period of about 10 years, he accumulated about 5% of the world's copper, owning both the actual metal and the futures. Five percent of a commodity might not seem excessive, but copper is illiquid and has relatively low turnover. Using his large futures position, Hamanaka kept the price artificially high. As well as making his position look profitable, this enabled him to earn inflated commissions on physical copper trades, in which commission is calculated as a percentage of the price (which is in itself ridiculous, because brokers do no more work to sell something for $100 than they do when it is $50).

Other traders speculated that Sumitomo was manipulating the price, but that is weak evidence as traders are generally paranoid and think every price is being manipulated. No one could confirm suspicions because the London Metal Exchange had no position limits and didn't publish open interest statistics.

In 1995, the price of copper dropped as Chinese mining increased. This new supply had a greater effect than Hamanaka could counter, and liquidating his huge position only added to the price decline. He was removed from his copper trading job and Sumitomo liquidated the position.

Sumitomo claimed they were totally ignorant of the positions. Although they had to keep pumping huge sums of money to collateralize the trade, ignorance is still possible (as we saw in the Baring's case, treasury and management don't always understand the positions they have). They said Mr. Hamanaka hid trade confirmations. He *was* later convicted of forgery. Nonetheless, Sumitomo was either complicit, ignorant, or negligent.

Lessons

- Treasury and management need to be aware of how much margin it takes to hold a position.
- Beware of instruments that have limited liquidity or transparency.
- Ask, “Is it reasonable for a trader to make this much money trading this product?” If you can't answer this, you shouldn't invest.

Example Three: Bernie Madoff

Bernie Madoff was a very well-respected figure in the investing world. The son of a plumber (later a stockbroker), he was genuinely a self-made man. His early jobs included lifeguard and a sprinkler installer. In 1960 Madoff founded a penny-stock brokerage, which eventually grew into Bernard L. Madoff Investment Securities. He became known as an ethical, philanthropic businessman. He was the chairman of the NASD board of directors and advised the SEC.

His business had two divisions. There was a broker-dealer and there was a money management firm. In 2008, it was discovered that the money management division was running a $50 billion Ponzi scheme in which Madoff paid redemptions from new investor deposits rather than any investment returns.

Madoff claimed consistent returns of 12% a year in all market conditions. Given that from 1950 to 2007, the S&P 500 averaged only about 9% a year, his fund found it easy to raise money. Unfortunately, the S&P 500 dropped over 38% in 2008. Many people asked for their money, and there were no inflows to cover the withdrawals.

(There were rumors that some of these reported returns were due to the investment division front-running client orders from the brokerage. Some investors thought they were the smart insiders in the scam. If this is true, it suggests the phrase, “You can't cheat an honest man.”)

Madoff's story was that he employed a split strike conversion (more commonly called a collar) in which he was long stock, short calls, and long puts. The fact that he told people about his strategy should have shut him down earlier. If you tell people what you are doing, they can more easily see if your returns are plausible. Although it is possible for a good trader to beat a strategy benchmark, the benchmark will still give a good baseline.

And it didn't even take an enormous amount of work to see that the purported returns were implausible. The first skeptic, an investment analyst called Harry Markopolos, claimed he could see within five minutes that the results were suspicious and only another four hours to prove it. He first notified the SEC in 2000 and then again in 2001, 2005, and 2007. He wrote the book *No One Would Listen* (Markopolos, 2010) about the experience. A detailed analysis of the split-strike conversion was published by Bernard and Boyle (2009).

Not only were Madoff's returns not real, most of his trades weren't, either. From at least the 1990s, he would just create fake trade confirmations based on idealized prices.

It is hard to quantify how much money really disappeared, given that most of it was fictional the whole time, but by 2019 only about $11 billion of the initially invested $17.5 billion had been returned to investors.

Lessons

- Check the plausibility of the strategy. Is the return stream consistent with a theoretical analysis? Is there enough liquidity to execute the claimed strategy?

- It is dangerous to rely on a superficial character assessment. Madoff had an exceptional reputation and it still isn't at all clear why someone as legitimately successful as he was would run a fraudulent operation.
- Only invest in funds with a separate TPA and custodian.

Index Restructuring

It is easy to forget that indices aren't real. They are just numbers calculated according to some methodology by a third party. There is nothing to stop the publisher changing the methodology or composition of the index. Generally, these changes are inconsequential (for example, most indices add and delete some components regularly). But not always.

The EuroSTOXX 50 is an index that is composed of the 50 largest capitalized companies in the eurozone, irrespective of what country they are from. At least it is now. Originally it was designed to have companies from each eurozone country. When STOXX made this change, they reweighted the index so its value wouldn't change. But the dividend yield, and hence the implied forward basis to cash, changed enormously (by about 50%).

Any option trader with a forward position lost money (still my largest 1-day loss).

A similar situation occurred in February 2018, when the volatility ETNs spiked (see Sinclair, 2018). UVXY, the ProShares Ultra VIX Short-Term Futures ETF, had its leverage reduced from 2 to 1.5. This meant that the expected future volatility dropped by 25%, which hurt any trader who was long vega. It isn't entirely clear whether this was allowed under the terms of the prospectus, and as of May 2019 lawsuits arguing about this have not been decided. But it seems likely option buyers will not recoup all losses.

Lesson

- Be aware of what contract specifications can be changed.

Arbitrage Counterparty Risk

Often, spread traders or arbitrageurs find they have large profits at one institution and large losses at another. This leads to credit risk.

A LIFFE broker had made a lot of money on spread bet arbitrage, buying low at some bookies and selling high at others. Unfortunately, almost all his winning bets were held by one bookmaker. And that person wasn't capable of paying. The simple thing would have been for him to just default, because gambling debts in England were not legally enforceable. But that wouldn't have been good for business, so he did something else.

The bookie's shop was between the train station and the broker's house. So, every evening, there were a couple of attractive, morally malleable young women waiting at the station when the broker disembarked. They escorted (in both senses) him to the shop where he was entertained with alcohol and cocaine while watching and gambling on greyhound races. Three months later the bookie had his money back (and the broker had STDs).

Lessons

- Try to stay flat counterparty risk.
- Don't do anything that would make your mother cry.

Conclusion

Amateur option traders lose money due to practically every decision they make. Professionals should be able to manage their market risks so that no single loss will be catastrophic. The risks that a professional should be most concerned about are those created by political instability, contract specification changes, the stability of financial institutions, and fraud. These can never be totally avoided. All a trader can do is to check everything that can be checked and avoid being completely exposed to any single country, currency, or institution.

Summary

- As much as possible, separate compliance, trading, and risk management.
- Never invest in a strategy you don't understand.
- Avoid illiquid products and situations.

- Try to diversify across institutions, currencies, managers, and countries.

CONCLUSION

Football is like chess, but with dice.

—Peter Krawietz, assistant manager of Liverpool Football Club, from Biermann (2019)

The same is true of option trading. Skill is essential, and the more knowledge and experience a trader accumulates the greater her chance of success. But there is also a tremendous amount of randomness in any individual trade. It is quite possible to predict realized volatility and direction correctly and still lose money as options also depend on implied volatility, interest rates, borrow rates, and dividends. The situation is even more complex if options are hedged as then path dependency is introduced.

The most important concept in trading is accepting that we will be making decisions in situations of great uncertainty. And this is not even the comparatively tame uncertainty of Knight (1921), in which the probabilities are unknown but are at least well defined. Traders operate in a realm of ignorance and unknowability where probabilities are changing, poorly defined, and the events they measure change. We will never know more than a tiny fraction of what can be known. And what can be known is a tiny fraction of all that there is.

This is not a reason to stop looking for trades with edge. It is a reason to look very hard and test ideas as rigorously as possible. Edges exist, but they need to be very robust to withstand the enormous amount of noise in the world. A trading strategy needs to have valid test statistics when applied to several markets. Ideally, it will also have a clear reason for existence. And all strategies should be robust with respect to the details of implementation.

When looking for ideas it is important to focus on phenomena rather than parametrizations or models. For example, volatility is important because it measures uncertainty and variability, not because it is the standard deviation of returns. That is just a mathematical expression of the core idea, chosen largely due to its mathematical tractability. Many other statistics could express the idea of variability. A good trading phenomenon is one that can be measured, modeled, and traded in many ways. For example, momentum investing is merely the observation that stocks tend to continue in the direction they have been moving. This can be studied at daily, weekly, or monthly time scales. It can be quantified by moving average rules, returns over previous periods, or numerous signal processing methods. The general observation is robust. The details of the trading model are of course important. It is possible to lose money trading a strong phenomenon by using a poor model. But the phenomenon itself is the most important thing.

It is important to continually search for new ideas. Max Planck said that science progresses one funeral at a time, and trading methods seem to as well. Although the markets are always changing, individual traders tend not to. When the trading floors closed, a lot of floor traders tried to apply the same techniques to trading on the screens. They didn't adapt. They kept trying to apply an obsolete set of methods until they retired. The most important thing for traders is that they are in a position to trade. So, we need to keep adapting to stay in the game.

Remember to be primarily a *trader,* not an *option* trader. Options are just a tool to express opinions. They are useful because of the various characteristics of volatility, but they won't be the best tool in every situation. Nietzsche (1878) (it is no coincidence that the favorite philosopher of many traders was a syphilitic maniac …) said, “Many people are obstinate about the path once it is taken, few people about the destination.” Remember this and don't fall into the trap of thinking more about options than trading.

Trading will always involve uncertainty. No matter how hard we work, we will still need luck. Both Napoleon and Eisenhower expressed their preference for lucky generals over talented ones. But it is important to remember that all their generals had reached that rank because they were talented. Talent was a given. It is the same with trading. Luck will play a part, but over a long career it will generally separate those of equal talent and knowledge rather than elevate the merely lucky. Learn all that you can but be sanguine about randomness.

Good luck.

APPENDIX 1
Traders' Adjustments to the BSM Assumptions

The Existence of a Single, Constant Interest Rate

The BSM model assumes a constant risk-free interest rate. There is no such thing as *an* interest rate. Interest rates have a bid-ask spread. We borrow and lend at different rates. Further, there is a different interest rate for each maturity: the yield curve. All these rates change over time. And no interest rate is risk free.

Before we even discuss the effects of any mispricings due to interest rates, it is important to note that very few traders hedge their own interest rate risk. At a large firm, this will be handled by the risk management group, which will hedge the firm's net exposure by aggregating the exposures of all positions denominated in each currency. Independent traders generally don't hedge interest rate risk at all. It is too expensive in terms of transaction costs and ties up margin. If independent traders start to accumulate too much rho, they will reduce it by trading options. Market-makers will shade their rate input so they trade out of their rho position in the same way that they shade their volatility inputs if they want to reduce volatility risks. Positional traders will either trade a reversal, conversion, or a box. This somewhat lax attitude toward rho is an indication of how robust the BSM model is with respect to interest rate inaccuracy.

Different maturity loans and bonds have different interest rates. This forms the yield curve. In theory this is no problem at all. We just hedge with the bond (or in practice the Eurodollar strip) corresponding to the lifetime of the option. However, the steeper this curve is the higher the chances are that we will be using an incorrect interest rate. How incorrect does the rate have to be before we develop significant price errors or, more important, delta errors? And what size is the likely input error?

A 1-year European call on a $100 stock that pays no dividends, struck at 130, priced with a volatility of 30% and 5% interest rate, has a value of $4.67 and a delta of 0.288. If we incorrectly used a rate of 4%, we would get a value of $4.44 and a delta of 0.277. A delta difference of 0.011 isn't totally insignificant, but the same size error would result if we used an incorrect implied volatility of 29%, an input error that is far more likely.

Further, in the current environment a trader would need to be extremely inattentive to have an interest rate input that is incorrect by 1%. On March 18, 2019, the US 1-month zero coupon rate was 2.47% and the 1-year rate was 2.52%. The effect of pricing options off the wrong spot in the yield curve was practically zero.

Interest rates are also volatile. Although the BSM model can handle a static yield curve, one where different maturities have different rates but they are unchanging, stochastic interest rates are more of an issue. When the underlying's volatility is constant, Merton (1973) showed that the current zero-coupon bond yield will still work, even when rates are stochastic. However, this does not work if volatility is also stochastic. But the effects of rate volatility are also negligible.

Just as the absolute size of rate errors is small, so is the volatility of rates. In particular, the volatility of rates is much lower than the volatility of volatility. The absolute daily changes of the VIX, and the 1-year rate, are shown in Figures A1.1 and A1.2.

The standard deviation of daily VIX changes has been 1.7 points. The standard deviation of daily changes in the 1-year rate was 0.037%.

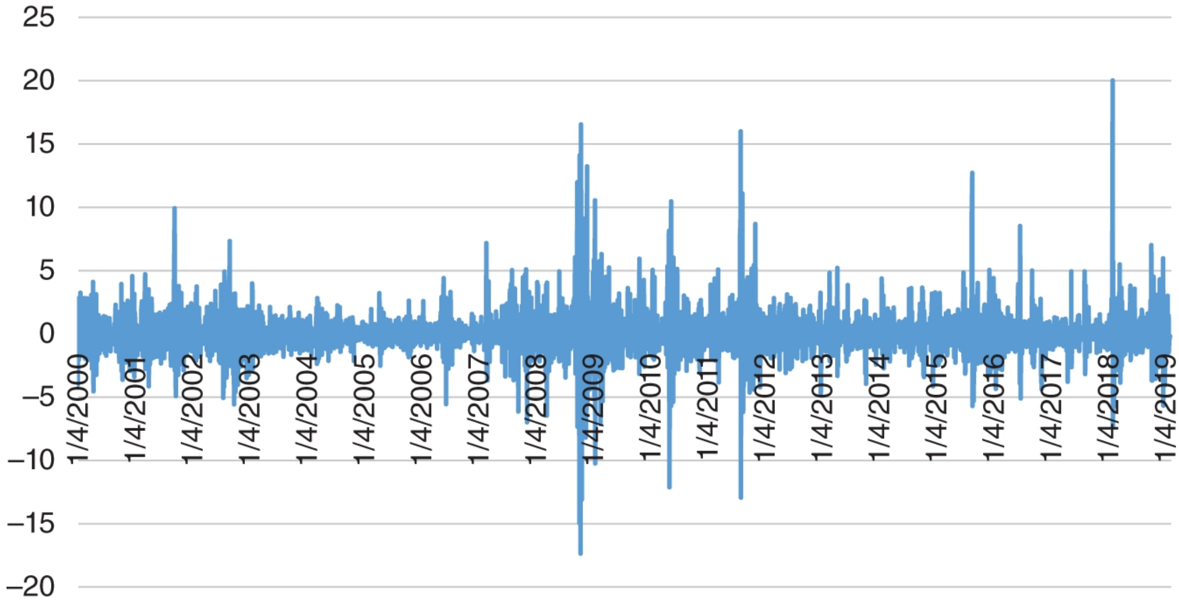


FIGURE A1.1 The daily VIX changes from 2000 to 2018.

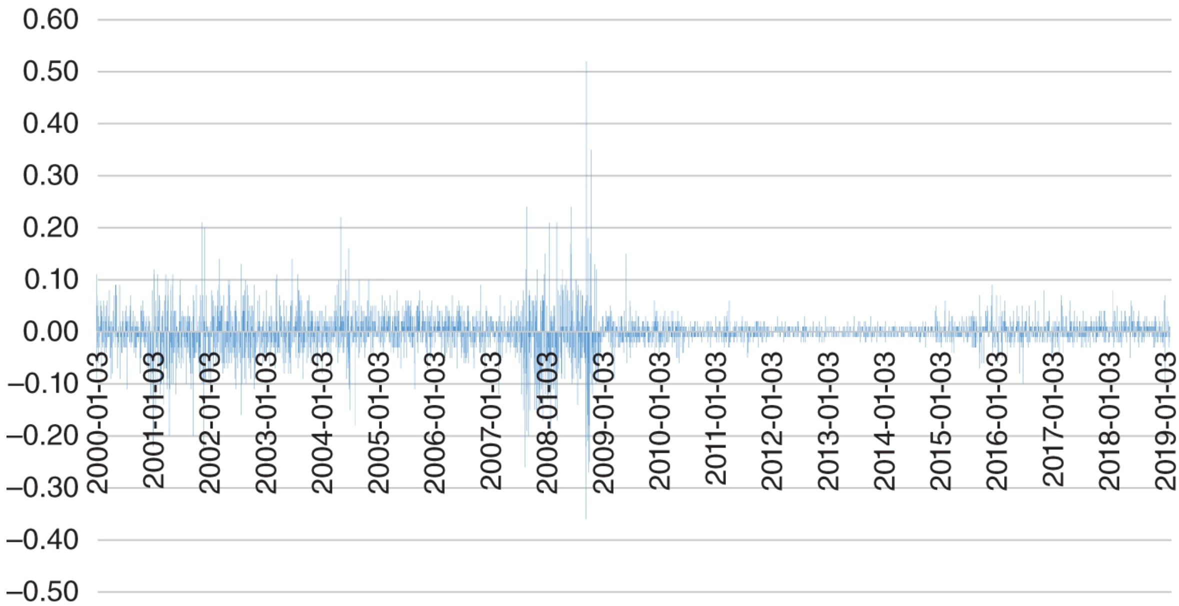


FIGURE A1.2 The daily 1-year rate changes from 2000 to 2018.

Because the volatility of rates is so comparatively low, it isn't necessary to use a model that incorporates stochastic interest rates. This has been confirmed by several empirical studies. Bakshi et al. (1997, 2000) showed that, after accounting for stochastic volatility, adding stochastic rates did little to improve pricing and hedging for options, even LEAPs with up to three years to expiration. Kim (2002) found an even stronger result: incorporating stochastic rates into an equity option pricing model offered no improvement over the BSM model.

One situation in which an incorrect rate can cause problems is when making early exercise decisions. When exercising a put on a stock we need to decide if interest income on the proceeds from a short stock position is greater than the amount of optionality value we are losing. If rates are stochastic, we might get this calculation wrong. There isn't a lot we can do about this. Certainly, no model can help.

Finally, interest rates have a bid-ask spread. It is possible to modify the BSM model to take this into account (Bergman, 1995). The analysis is similar to the modification necessary when the underlying has a bid-ask spread. And, as in that case, differential interest rates imply that the option has a band of values rather than a single price. But again, the effects are very, very small in practice.

The Stock Pays No Dividends

The BSM model assumes the underlying stock pays no dividends. Correcting this is trivial. We simply price the option off the stock minus the discounted value of the dividend. So, in the case of a single discrete dividend, *D,*

$$S \rightarrow S - Dexp(-rt) \quad \text{(A1.1)}$$

In some cases, a continuous dividend yield, q, is a fair approximation. In this case,

$$S \rightarrow Sexp(-qt) \quad \text{(A1.2)}$$

A similar adjustment is needed if a stock becomes hard to borrow. The BSM assumes that sale proceeds can be invested at the risk-free rate but if a stock is hard to borrow, the trader receives a lower rate, r-λ, where λ is the borrowing penalty.

Absence of Taxes

The BSM model ignores taxes. Some traders are taxed as individuals and some as corporations. Sometimes profits will be taxed at the short-term capital gains rate, sometimes at the long-term capital gains rate, and sometimes at a mix of the two rates.

Tax cheats pay no taxes. Foreign investors may have other tax complications.

If all investors had the same tax rate, BSM could be adjusted by using a modified interest rate. The problem isn't the difficulty of including taxes in a pricing model; it is that different people have different taxes. An investor's tax situation will affect his trading strategies (Scholes, 1976), but it is impossible to construct a pricing model that considers different, unknown tax obligations. Options will be worth different amounts to different people, but we can't value the effect.

Probably due to the difficulty of estimating the aggregate marginal tax rate, little empirical work has been done on this issue, either. Mason and Utke (2019) compared SPX and SPY options. ASPX option profits are taxed 60% at the long-term capital gains rate and 40% at the short-term rate. SPY options are taxed at the higher short-term rate. After controlling for dividends and American/European exercise features they concluded that there was a persistent price difference where SPY options were cheaper. Their conclusion was that this could be attributed to people being less inclined to buy options with higher tax rates. However, the effect was so small that it could be explained by dirty data or data processing procedures (H. Contini, personal communication, 2019).

The Ability to Trade and Short the Underlying

The BSM formalism relies on the ability to hedge in the underlying. Sometimes we will need to be able to short the underlying. Sometimes this isn't possible. If we cannot short, then the BSM hedging strategy clearly won't work. To see how pricing is affected, think about pricing a forward, *F*. The current price of the underlying is *S*, time until delivery is *T*, and interest rates are *r*.

If $F > S\exp(rT)$, sell the forward contract and buy the underlying for *S* with borrowed money.

On the expiration date, we deliver the underlying, and receive the agreed forward price, *F*.

We then repay the lender the borrowed amount plus interest. This amount totals $S\,exp(rT)$. The difference between the two amounts is the arbitrage profit. The principle of no-arbitrage means this can't happen so

$$F \leq S\,exp(rT) \tag{A1.3}$$

But if $F < S\,exp(rT)$, we can't form the arbitrage portfolio. In this case we would need to buy a forward and short the underlying, so we cannot rule out the inequality.

If we can't short the underlying, the forward price can be less than the no-arbitrage value. Consequently, call values can be less than the BSM value and put values can be more than the BSM values. The put-call parity relationship is "shifted" as well. This means that pricing in the risk-neutral world isn't possible. If we can't trade the underlying at all, then we are even more lost.

It is unlikely that most traders will trade options on completely untradeable options, but there are cases where the underlying is, or becomes, illiquid, so it is worth knowing how to deal with an untradeable underlying as a limiting case.

Before an exchange lists options, the underlying has to have a certain amount of liquidity. For example, to have options on a stock in the United States the company must be listed on either the NYSE, AMEX, or NASDAQ. The company must have at least 7 million shares and the company must have more than 2,000 shareholders. But after the listing it is quite possible for the underlying to become less liquid.

There are also instances where we want to price employee options on a non-traded company. This stock is impossible to trade.

Finally, there are cases where the underlying becomes impossible to short either through lack of a borrow or through a regulatory change. Although this illiquidity is only on one side of the market it is still enough to create problems.

One way to approach this problem is to use the "real option" approach. Although the term *real option* is new, the idea isn't. In 1930, Irving Fisher explicitly wrote about the options that a business owner had (although his credibility may have been lessened due to his famous 1929 statement that "stocks have reached what looks like a permanently high plateau”). However, the publication of the BSM model has enabled the concept to be quantified.

An example of a real option is the decision to invest in a project. Preliminary research costs are the price of the option. The exercise price is the future profits. The payoff is the difference between these two.

There are some important distinctions between real options and financial options:

- They are untraded in a conventional sense although they can be “bought” or “sold” by buying or selling the business units that create them.
- The value of the option is directly dependent on the actions of management because their actions control the initial premium.
- These options are often more dependent on uncertainty than on the measurable volatility of the underlying.

Some real options can be priced by using the BSM framework because they involve an underlying that is a traded asset. An example of this situation would be the option to start producing oil. The premium would be the cost of equipment and exploration. The underlying asset is oil. Because oil is traded, this option can be hedged and the risk-neutrality arguments of BSM can be applied.

But many real options are not on a traded asset. For example, if the underlying is some sort of intellectual property, it will be impossible to hedge, and hence the BSM formalism will be inapplicable. The principle of no-arbitrage cannot be used. It is common in these circumstances to assume the existence of a traded security that is perfectly correlated with the underlying. This is almost always a very unrealistic assumption.

However, if there are traded securities with imperfect correlation to the underlying (obviously the higher the better), we can find bounds on our real option prices. This was done by Capinski and Patena (2003) (an application of the no-good-deal theory of Cerny and Hodges, 2000), and, in the case of a perfectly correlated asset, their model gives the BSM option price.

Their main assumption is that adding an option to a portfolio won't change the Sharpe ratio. This isn't always true, but it is generally a good approximation (and is the same assumption that Black and Scholes used in their derivation). Unfortunately, the bounds produced are too large for any practical valuation processes (the range is generally from zero to the BSM price assuming perfect correlation).

Hedging options with a correlated underlying produces a large dispersion of P/L. In Figure A1.3 we show the effect on the P/L of an option that is hedged daily with the underlying. The duration is one year. Both realized and implied volatility are 30%, rates a zero, and total vega is $1.000. One thousand paths of GBM were used to find the dispersion.

In conclusion, when the underlying is untraded the best you can do is find another product that should be similar (in the co-integrated sense). This will let you value the option using the BSM concept because you can now hedge. It probably won't be a good hedge, but you need to do the best you can, accept the variance, and ask for enough edge. You can't create information from nothing, but you can get someone else to accept some of the uncertainty through the hedge.

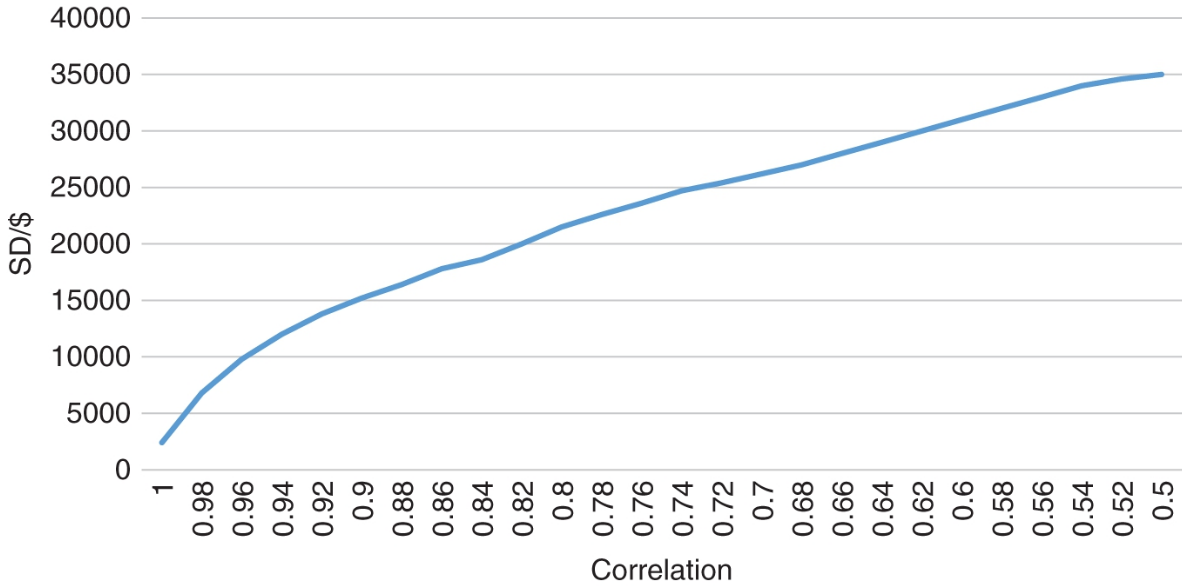


FIGURE A1.3 The standard deviation *pf* the P/L for an option hedged with an imperfectly correlated underlying.

Nonconstant Volatility

The volatility of the underlying is not constant. The rolling 30-day close-to-close volatility of the S&P 500 is shown in Figure A1.4.

Any nonconstancy of volatility creates a non-normal return distribution. The simplest example is if the true return process is a mixture of normal distributions with different volatilities. The kurtosis of a mixture of normal distributions, with zero means, is given by

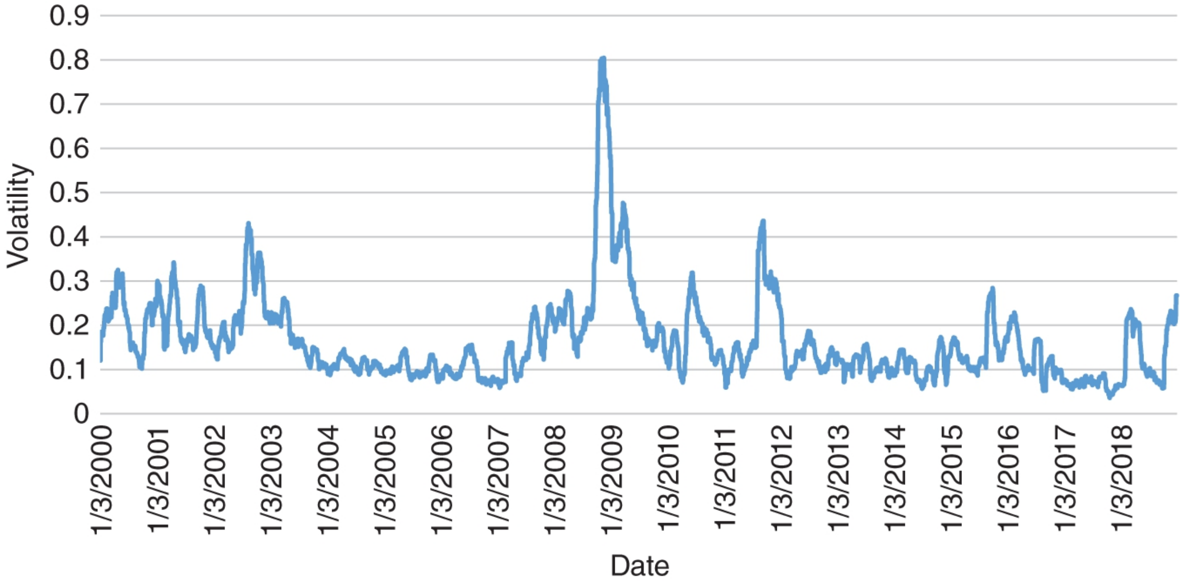


FIGURE A1.4 S&P 500 30-day volatility from January 2000 through to the end of 2018.

$$\frac{3(p_1\sigma_1^4 + p_2\sigma_2^4)}{(p_1\sigma_1^2 + p_2\sigma_2^2)^2} \tag{A1.4}$$

where p_x is the probability of being in the state with volatility, σ_x. So, if there is a 50% chance of being in a state with a volatility of 20%, and a 50% chance of being in a state with a volatility of 80%, the distribution will have a kurtosis of 5.3.

We could price options off the resulting distribution and then back out the equivalent BSM implied volatilities to see how stochasticity leads to a volatility smile, but we can also show this by directly thinking about implied volatility as the stochastic variable.

The implied volatility of an option is the forward-looking estimate of the average underlying volatility over the lifetime of the option. If volatility is not constant, an implied volatility smile will appear.

Consider a case in which a $100 stock has a volatility with a 50% chance of being 20% and a 50% chance of being 80%. So, half the time the 30-day, 100 call will be worth $2.29 and half the time it will be worth $9.13. The average value is $5.71. This corresponds to an implied volatility of 50%, the average of the two possible volatility states. Now consider a 120 call. In the low volatility state, it is worth nothing and in the high volatility state it is worth $3.04. In this case the average value is $1.52, which corresponds to an implied volatility of about 62%.

This situation would create a symmetric smile, but extending the idea so that underlying moves are negatively correlated to volatility moves gives the asymmetric implied volatility skew found in most products. Imagine we have the same two volatility states (20% and 80%) but now the low volatility is associated with a stock price of $102, and the high volatility state corresponds to a stock price of $98. Going through the same exercise as just described we find the 80 strike has an implied volatility of 68%, the 100 strike has an implied volatility of slightly over 50%, and the 120 strike has an implied volatility of 59%.

Nonconstant volatility creates the convexity of the smile. The correlation between volatility moves and stock moves creates the slope of the smile. Other causes also exist, but stochastic volatility is clearly a contributor to the implied volatility structure.

Many pricing models that address this effect have been developed. For example, as of January 4, 2019, ssrn.com listed 717 papers with *stochastic volatility* in their title and 3,287 with those words in their title, abstract, or keywords.

Market-makers also know about this effect. They typically use a modified BSM model that has a different volatility for each strike, which they periodically update. Is using a deterministic volatility model and sporadically changing the parameter inferior to using an explicitly stochastic volatility model?

A number of studies that compare the performance of the modified BSM model to different correctly specified models have reached the same conclusion, including Jung (2000), Engle and Rosenberg (2002), Bollen and Raisel (2003), Yung and Zhang (2003), Li and Pearson (2007), Jung and Corrado (2009), and Hull and White, 2017. A stochastic volatility model that captures the true stochastic process offers practically no improvement over the tweaked BSM model used by traders. In each case, the hedging errors due to using BSM are economically insignificant.

A further consideration is that even a stochastic process that is fairly effective at modeling the dynamics of the underlying will not be able to exactly match all of the implied volatilities. The variance-reducing effect of hedging at the implied volatilities (Ahmad and Wilmott, 2005) will be reduced. Even if a stochastic volatility model gives a higher terminal profit, it will always lead to a more variable day-to-day PL.

Given that we won't ever know the true volatility generating process, using the simpler, better understood, and non-parametric-modified BSM model is probably the best solution.

Conclusion

Although the BSM model is based on many clearly incorrect assumptions, most of these can be corrected for with simple adjustments (some more ad-hoc than others). BSM has three large advantages over more complex models. It is an industry-wide standard that forms the basis for trader communication. Most traders have based all of their learned intuition on the BSM model. It is simple and implemented in all trading programs. Given these advantages, there is no compelling reason to use a more “correct” model over the BSM with the various adjustments.

Summary

- BSM doesn't handle stochastic interest rates or interest rates with transaction costs, but in practice this doesn't really matter because the effect is typically small.
- Adding dividends and carry rates to BSM is easy.
- Very little research has been done on taxes and BSM, but this is true of all other models as well.
- It is possible to trade options on illiquid underlyings if there is a correlated hedging instrument.
- Volatility is stochastic, but BSM with periodically adjusted volatility inputs works as well as stochastic volatility pricing models.

APPENDIX 2
Statistical Rules of Thumb

Just as a lot of popular trading rules have no basis in fact, the generally accepted origin of the phrase “rule of thumb” is no more than a myth. The story is that there was a 17th-century English law that specified that the maximum width of a stick for wife beating was the width of a man's thumb. No such law ever existed.

Nonetheless, these simple heuristics can be useful for quick estimates or establishing a baseline.

Converting Range Estimates to Option Pricing Inputs

In order to either subjectively price options or simulate outcomes, you will need estimates of mean return and the variance of the returns. Unfortunately, it is quite common for fundamental analysts to give only a most-likely case and a range of possible outcomes. It is probably best to use the common time series methods to estimate volatility (refer to Sinclair, 2013), but the analysts' return information could be useful.

The approximations for doing this conversion are part of a class called *three-point-estimators*. There are many such methods that differ due to the assumed shape of the approximating distribution.

First, convert the analyst's numbers to annualized percentages. So, if she has a low stock price estimate of $90 on a $100 stock in the next three months, that would convert to a negative 40% return. If we assume the data is drawn from a triangular distribution, and our low estimate is l, the median estimate is m and the high estimate is h, the mean estimate is

$$mean = \frac{(l + m + h)}{3} \tag{A2.1}$$

This is just an average of our data. This makes sense if we have no idea of the historical accuracy of the analyst. But sometimes we want to more heavily weight the median. This would be the case if we have higher confidence in the analyst. If we don't have much of an idea how good the analyst is, we will have no reason to more heavily weight the median (we should also lower our resulting trade sizes in this situation).

A weighted three-point estimate of the mean is

$$mean = \frac{(l + 4m + h)}{6} \tag{A2.2}$$

(This assumes a beta distribution for the true process.)

If we had confidence in the analyst, we would just use her median value as the mean.

This method is far from ideal, but it could also be the only way to use the information from an otherwise valuable source.

Rule of Five

This is an exceptionally simple heuristic that I read about in *How to Measure Anything: Finding the Value of Intangibles in Business* by Douglas Hubbard (2007). He gives no references and I've never seen it mentioned anywhere else.

If we are randomly sampling from a population, we can be 93% certain that the population median lies within the range of five measurements. For example, if I have a GBM with unchanging parameters and I measure the volatility over five different periods and get 30%, 20%, 26%, 18%, and 23%, I can be 93% certain that the true median volatility is between 18% and 30%. The usual caveats about what a random sample is apply. And often the range, and hence confidence interval, will be wide but it is still valuable to use even a very small sample to establish a base rate.

The only way the median could be outside the range is if all measurements were either above the median or below it. Because the chance of any observation being above the median is 50%, the probability of five measurements all being above the median is 5^5 = 0.03125. Because there is the same chance of five measurements being below the median, the total probability of being outside the range is 0.065. We have a 93.5% chance of the median being within the range. This idea is generalizable to different confidence levels and by considering only certain subranges.

As with all heuristics, the Rule of Five is only approximate, but if it increases your knowledge by any amount, it is worth using.

Rule of Three

This is a method to quickly estimate the probability of something that has never happened before. Obviously, in some situations you will have prior knowledge and won't need to rely completely on a purely mathematical bound, but this is a useful method to at least establish a base rate. For more discussion refer to Hanley and Lippman-Hand (1983) and Louis (1981).

A 95% upper bound of the occurrence rate is given by

$$\frac{3}{n} \tag{A2.3}$$

where n is the number of observations.

So if we haven't seen an event in 30 observations, a 95% bound on the chance of the event happening in the next period is 3/30, or 10%. This might seem high but remember this is an upper bound and we are using no specific information about the particular situation.

To derive the result, take the chance of the event to be p, which is what we want to estimate. In each separate time period the chance of the event not happening is $1-p$. So, after n periods the chance of there being no events is

$$(1-p)^n \tag{A2.4}$$

We want to find p such that this probability is less than 5%. This gives the bound on the event not happening.

So we solve the equation

$$(1-p)^n < 0.05 \tag{A2.5}$$

for p.

Taking logs

$$n \ln(1 - p) < \ln 0.5 \qquad \text{(A2.6)}$$

The logarithm of 0.5 is about −3. If *p* is small, which it has to be if the event hasn't been observed, ln(1−*p*) is roughly -*p* (the first term of a Taylor series). So we get

$$-np < -3 \qquad \text{(A2.7)}$$

which gives the critical value of *p*.

Clearly, this idea can also be used with other confidence intervals and, less clearly, can be generalized to the case when one observation of the event has occurred.

We have made two very important assumptions in our derivation: *p* needs to be constant and the successive observations need to be independent. So this heuristic works well for a question such as, "Given that I've run through the fireworks factory with a lighted candle 20 times with no explosion, what is the chance I can do it again?" But it couldn't be applied to the question, "Given that I've been alive for 70 years, what is the chance of me surviving another year?" Here the normal process of aging means the probability of death increases with age, so *p* isn't constant, and any health issues mean the observations aren't independent.

In questions a trader might be interested in, things probably won't be this clear. For example, think about the case of a company going bankrupt next year. Is the probability of bankruptcy independent of what has happened in the past? In some cases, this is a decent guess. It could be valid in the case of a company that is very dependent on one product and is most vulnerable to a disruptive technology. But in the case of a company like Uber, where the current business model clearly needs to be changed if it is going to survive, the assumption probably isn't valid. Similarly, a company with an established business model might have a relatively constant probability of bankruptcy, whereas a startup won't.

APPENDIX 3
Execution

After deciding what trade to do, we still need to do it. Trade execution is a complex subject. Many open-outcry traders made a lot of money purely because of their ability to execute well. Execution was such a valuable skill that being good at it could mask a lot of other trading weaknesses. It was possible to be a good floor trader if your *only* skill was execution. Similarly, on the electronic exchanges many firms have done very well with algorithmic trading, order-type arbitrage, and latency advantages.

Even if execution ability isn't going to be a source of alpha (which it absolutely can be), a large enough trader may need to seriously consider optimizing execution. But most traders won't be in this situation. They will obviously want to minimize trading costs but won't trade enough to need to invest in an algorithmic system. For these people, using the built-in execution algorithms in the widely available (and free) brokerage-provided trading systems should be fine. Twenty years ago, it was reasonably easy to beat most VWAP systems (volume-weighted average price). Now it isn't.

However, it is still important to understand how to think about transaction costs generally. All financial decision-making is about balancing risk and reward. In the case of trade execution, the issue is how much we should pay to do a trade. If we are too aggressive, we will kill our returns by paying too much, but if we are too passive, we won't ever make any trades at all. This is true no matter what size we are trading or how liquid the product is. The situation may differ by degree, but the principles will be the same. The mathematics of balancing expected return and trading cost can become complex but there are also some broadly applicable rules of thumb that we can use.

The decision to make a trade is contingent on the instrument being at some particular price, the “decision price.” This is often the price in the middle of the bid-ask spread but it could be any price at all. A transaction cost is the price premium paid above the decision price for buyers and the discount below the decision price for sellers. The total cost has several components. Here they are in rough order of how evident they are:

- **Commissions and fees.** Commissions are paid to brokers and fees are charged by exchanges and regulatory bodies, but they are both fixed and visible. Who they go to doesn't matter to us as traders, so we will consider them to be the same.
- **Bid-ask spreads.** The bid-ask spread is the difference between the highest bid price and the lowest ask price. It exists to compensate the market-makers for providing liquidity. Because market-makers have their own set of problems, the bid-ask spread is highly variable, both during the day and in different market regimes. Further, the displayed bid-ask spread will often not be the true spread. The visibly quoted spread will be for a given size. Orders smaller than this can often be filled inside the spread and larger orders will usually pay a wider spread.
- **Price change.** Price change is the change in value of the instrument between our decision to trade and the execution. This can be positive or negative depending on whether we are buying or selling and whether the market is rallying or dropping. Generally, when we are entering a trade the cost will be negative because we will be selling something we expect to drop or buying something we expect to increase in value. But the cost will be random when we exit, because if we had a price view, we wouldn't be exiting. This is visible and variable.
- **Market impact.** Market impact is the price change due to our order. Because the price will also be changing of its own accord, market impact is invisible. It is also variable and depends on what other traders are doing at the same time. Some of the market impact will be temporary and lasts only until the market has absorbed the new trade, but some is permanent as the market processes the new information from the more aggressive price takers.
- **Opportunity cost.** Opportunity cost is the lost profit when we do not enter a trade. This is generally due to insufficient liquidity. This cost is invisible and variable.

The trader's execution dilemma is to maximize return given these costs. If he is too aggressive, the realized costs will be high. If he is too passive, there will be significant timing risk and opportunity cost because many trades will be missed.

Commissions, fees, and taxes are inescapable by-products of our trading strategy. A given strategy can't just be made to trade less often or to take a different holding period. Then it is a different strategy. These costs should be considered when the strategy is being planned and tested but once we are trading, they are what they are. Changing these would require changing the strategy.

The bid-ask spread is also in some ways a fact of nature. If you demand to be filled, you will pay the spread. Many traders have a hard time even accepting this, thinking that they can buy on the bid and sell on the offer. Of course, it is possible to try to do this, but if you want a guaranteed fill, you will need to pay the spread.

But the effective spread will probably not be the difference between the low offer and the high bid. To see this, look at the order book in Table A3.1.

TABLE A3.1 The Order Book of All Bids and Offers for UVXY (ProShares Ultra VIX Short-Term Futures ETF) on the Morning of August 10, 2016

Bid Size	Price	Ask Size
	20.71	1,200
	20.70	1,200
	20.69	1,100
	20.68	1,800
	20.67	600
	20.66	1,400
	20.65	800
200	20.64	
1,000	20.63	
900	20.62	
500	20.61	
700	20.60	
1,300	20.59	

The best bid is 20.64 and the best offer is 20.65, but the 20.65 offer contains only 800 shares for sale. So, if we wanted to buy up to 800 shares, we could say that the spread was $.01 (the difference between 20.64 and 20.65). But if we wanted to buy 5,000 shares, we would need to pay up to 20.69, with an average purchase price of 20.6692. Similarly, to sell 5,000 shares would give an average fill price of 20.608. So, for an order of 5,000 shares the effective bid-ask spread is $0612. The bid-ask spread is contingent on the size of the order. In some cases, such as where we have a populated order book, this is self-evident. But it is always the case.

The example of the UVXY order book that we just gave showed the case in which the actual, effective spread is wider than the difference between the best bid and best offer. But there are also cases in which the actual spread is narrower than the currently posted spread. This can often happen in option markets. The market-makers don't want to run the risk of showing tight prices for large size, but they will probably trade tighter for smaller size. For example, the indicated spread might be a bid of 9.0 for 100 options and an offer of 9.5 for 100 options but the market-maker might be prepared to pay 9.2 for 5 and sell 5 at 9.3.

Sometimes you can see the actual spread and sometimes you can't. Sometimes you have to "fish" by placing a small order and seeing where you get filled. But remember that the spread is a fact of life. If the market is showing 9.0 bid for 100 and 100 offered at 9.5, it is highly unlikely you are going to get filled if you try to sell 1,000 at 9.2.

Generally speaking, the following are true:

- The fill-price for an order of infinitesimal size will be the mid-point of what you see.
- You need to pay a spread if you demand to be filled.
- The bigger your order the further away from the "zero size" price you will need to go.

Somewhat related to the bid-ask spread is the idea of market impact. This is the amount that a given order changes the market price. It is useful to further split market impact into temporary and permanent impacts.

First, let's think about temporary market impact. This is just the fact that by doing a trade we will take out some of the orders in the book. Most of this is temporary because we would expect the book to fill back in again. But it won't completely, and this difference is permanent impact. There are a couple of ways to see that permanent impact has to exist. The first is that all trades convey information. If someone is buying, they must think the price is going up. The price of a security is just the aggregation of all of this information and every new order will cause some adjustment. The second reason is to prevent arbitrage. If all of the market impact was temporary, we could split our order into smaller pieces and always be guaranteed to pay a smaller spread than by doing a single large trade. We would just wait for the market to repopulate. In the UVXY example, the impact when buying 500 shares is only $.01. Why not just buy this many, wait for the book to fill back in, then do it 10 more times? Even if we ignore fixed fees and price appreciation, the effect of permanent market impact means doing this is not automatically advantageous.

If we are to make a sensible decision about whether to execute in slices or all at once, we will need to have a model of the order book dynamics. At any random time, what is the spread as a function of order size?

If we had such a model we could decide on the optimal “slicing” procedure, dividing the total order into suborders that minimize impact. But although there are many such models, most are far too complex to be of any use to a non-quantitative trader (and it also isn't clear to me that they add enough value to make all the work worthwhile anyway). If you are trading large enough size for such a model to be useful (a decent guess for “large enough” would be an order about 1% of the volume in a given time period), you should probably use one of the execution algorithms provided by most professional-level brokerage firms. If you aren't trading this big, it probably isn't too important exactly how to split an order.

The class of algorithms that is designed to minimize market impact is VWAP trading, which aims to match the volume weighted average price. The VWAP itself is a measure of the average transaction price for all market activity in a given amount of time. It is often considered a sort of fair execution price and many traders have their own execution benchmarked to the VWAP price.

Achieving VWAP is theoretically easy, but obviously is harder in real life. By definition, to perfectly match VWAP a trader would need to participate in every single trade. For example, if a trader's order was 10% of the volume in a given time, he would need to participate by having 10% of every trade. This is impossible for several reasons:

- Most exchange filling algorithms won't just let you get on any trade you want. A lot operate on a time-priority basis.
- You can't know what volume will trade until after the event.
- It is impossible to enter enough individual orders to participate on every trade.

However, the volume profile is stable enough that we can get a good idea of volume per unit time by looking at historical numbers. The numbers are so stable that many brokers offer a guaranteed VWAP execution where they promise your fill price will be the actual realized VWAP. They will never actually execute at this price, but their tracking errors are small and bias free.

VWAP strategies are good at lowering market impact because they are trading proportionally to the amount of volume in the market. They don't try to push large volume into thin markets, which would move the price, the definition of market impact. But market impact is just one trading cost and VWAP strategies are not the best at managing the most important cost for the active trader: timing cost.

The timing cost is the amount the market moves in the time between the trading decision and the end of the execution. This can either be positive or negative. As an example, consider the VIX Fed trade. Here my thesis is that the VIX futures rally in the 15 minutes before the Fed announcement and crash immediately afterwards. This means that I can be somewhat passive when selling before the announcement, sitting on the offer as the price should come to me. The same idea is true when I'm buying back the position after the announcement. I can afford to wait on the bids.

Conversely if I was buying futures, say 30 minutes before the announcement, in the anticipation that they would rally right up until the release I would need to be aggressive and lift the offer. I can't sit on the bid because the price will be naturally going away from the bid.

Understanding this idea is crucial for the active trader because we will generally be trading in the expectation of making a quick profit on a market move. If we dither around and refuse to pay the spread, we may never get filled and miss all of the profit.

Example

Let's say we want to sell the straddle on a stock because we expect implied volatility to collapse after the earnings are announced. On the open of the day before the announcement, the straddle is 2.1 bid and offered at 2.5. We want to sell 20 straddles and the bid and offer are both for 100.

Because I know that the implied volatility will generally tend to rise until the announcement, I can afford to let the market come to me. So, I would put in a mid-market order to sell 20 straddles at 2.3. If this doesn't get filled, then we need to try something else. I know from back testing and previous trade results that I expect the straddle to drop to 1.0 after the announcement. From the mid-price this would give a profit of 1.3, but even from the bid the profit would be 1.1. This profit projection has some uncertainty to it but according to my own risk-reward calculus I would never miss out on a profit of 1.1 trying to get a better fill by 0.2. I might be able to get filled at a better price so I will do a little fishing. I will put an offer to sell one lot and gradually walk it down, lowering the offer until I get filled. Then I will offer the remaining 19 straddles at slightly better than that. If I don't get filled, I will just hit the bid. The trade is expected to be profitable, so I don't want to miss it.

By far the most important aspect of execution to an active trader is not missing the trade. Psychologically, an easy way to avoid missing trades is to value the trade against the bid if selling and the offer if buying. This stops you being intimidated by the size of the spread.

Keep in mind the expected direction of the instrument you are trading. If you are selling into a rising market, you can be patient, but if you are buying into a rising market, you will need to be aggressive.

REFERENCES

Abreu, D., and M. Brunnermeier. 2003. "Bubbles and Crashes." *Econometrica* 71: 173–204.

Ahmad, R., and P. Wilmott. 2005. "Which Free Lunch Would You Like Today, Sir?" *Wilmott Magazine* (November): 64–79.

Andersson, M., L. Overby, and S. Sebestyen. 2009. "Which News Moves the Euro Area Bond Market." *German Economic Review* 10: 1–31.

Andries, M., T. Eisenbach, M. Schmalz, and Y. Wang. 2015. "The Term Structure of the Price of Volatility Risk." Working paper, Toulouse School of Economics.

Aronson, D. 2007. *Evidence-Based Technical Analysis*. New York: Wiley.

Bachelier, L. 1900. *Theorie de la Speculation*. Paris: Gauthier-Villars. (Reprinted in English in Cootner, 1964).

Bakshi, Z., C. Cao, and Z. Chen. 1997. "Empirical Performance of Alternative Option Pricing Models." *The Journal of Finance* 52: 2003–2049.

Bakshi, Z., C. Cao, and Z. Chen. 2000. "Pricing and Hedging Long-term Options." *Journal of Econometrics* 94: 277–318.

Ball, R. J., and P. Brown. 1968. "An Empirical Evaluation of Accounting Income Numbers." *Journal of Accounting Research* 6: 159–178.

Baltas, N. 2019. "The Impact of Crowding in Alternative Risk Premia Investing." *Financial Analysts Journal*. SSRN: 33603350.

Barber, B., and T. Odean. 2008. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies* 21: 785–818.

Bartov, E., S. Radhakrishnan, and I. Krinsky. 2000. "Investor Sophistication and Patterns in Stock Returns after Earnings Announcements." *The Accounting Review* 75: 43–63.

Beaver, W. 1968. "The Information Content of Annual Earnings Announcements." *Journal of Accounting Research, Empirical Research in Accounting: Selected Studies* 6: 67–92.

Ben-Meir, A., and J. Schiff. 2012. "The Variance of Standard Option Returns." http://arxiv.org/abs/1204.3452

Bergman, Y. 1995. "Option Pricing with Differential Interest Rates." *Review of Financial Studies* 8: 475–500.

Bernard, C., and P. Boyle. 2009. "Mr. Madoff's Amazing Returns: An Analysis of the Split-Strike Conversion Strategy." *Journal of Derivatives* 17: 62–76.

Bernard, V., and J. K. Thomas. 1989. "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?" *Journal of Accounting Research* 27: 1–36.

Bernard, V. L., and J. K. Thomas. 1990. "Evidence That Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings." *Journal of Accounting and Economics* 13: 305–340.

Beunza, D., and D. Stark. 2012. "From Dissonance to Resonance: Cognitive Interdependence in Quantitative Finance." *Economy and Society* 41: 383–417.

Biermann, C. 2019. *Football Hackers: The Science and Art of a Data Revolution*. London: Blink Publishing.

Black, F. 1986. "Noise." *Journal of Finance* 41: 528–543.

Bollen, N., and E. Raisel. 2003. "The Performance of Alternative Valuation Models in the OTC Currency Option Market." *Journal of International Money and Finance* 22: 33–64.

Bollerslev, T., J. Marrone, L. Xu, and H. Zhou. 2014. "Stock Return Predictability and Variance Risk Premia, *Statistical Inference and International Evidence* 49: 633–661.

Bollerslev, T., and V. Todorov. 2011. "Tails, Fears, and Risk Premia." *The Journal of Finance* 66: 2165–2211.

Bollerslev, T., and H. Zhou. 2007. “Expected Stock Returns and Variance Risk Premia.” Working paper, Federal Reserve Board.

Bondarenko, O. 2003. “Why Are Put Options So Expensive?” Working paper, University of Illinois-Chicago.

Boness, J. 1962. “A Theory and Measurement of Stock Option Value.” PhD Dissertation, University of Chicago.

Boness, J. 1964. “Elements of a Theory of Stock-Option Value.” *Journal of Political Economy* 72: 163–175.

Box, G. 1976. “Science and Statistics.” *Journal of the American Statistical Association* 71: 791–799.

Boyer, B., and K. Vorkink. 2014. “Stock Options as Lotteries.” *The Journal of Finance* 69: 1485–1527.

Brenner, M., and M. Subrahmanyam. 1988. “A Simple Formula to Compute the Implied Standard Deviation.” *Financial Analysts Journal* 5: 80–83.

Bronzin, V. 1906. *Lehrbuch der Politischen Arithmetik*. Leipzig: Franz Deuticke.

Browne, S. 1999. “Reaching Goals by a Deadline: Digital Options and Continuous-Time Active Portfolio Management.” *Advances in Applied Probability* 31: 551–577.

Browne, S. 2000a. “Can You Do Better Than Kelly in the Short Run?” In *Finding the Edge: Mathematical Analysis of Casino Games,* ed. O. Vancura, J. Cornelius, and W. Eadington, 215–231. Reno: University of Nevada, Reno Bureau of Business.

Browne, S. 2000b. “Risk Constrained Dynamic Active Portfolio Management.” *Management Science* 46: 1188–1199.

Cahan, R., and Y. Luo. 2013. “Standing Out from the Crowd: Measuring Crowding in Quantitative Strategies.” *Journal of Portfolio Management* 39: 14–23.

Cao, J., A. Vasquez, X. Xiao, and X. Zhan. 2018. “Volatility Uncertainty and the Cross Section of Option Returns.” SSRN: 3178263.

Cao, J., B. Han, Q. Tong, and X. Zhan. 2015. "Option Return Predictability," 27th Annual Conference on Financial Economics and Accounting Paper, Rotman School of Management Working Paper No. 2698267.

Capinski, M., and W. Patena. 2003. "Real Options-Realistic Valuation." SSRN: 476721.

Cerny, A., and S. Hodges. 2000. "The Theory of Good-Deal Pricing in Financial Markets." Imperial College Management School, London.

Chabot, B., E. Ghysels, and R. Jagannathan. 2009. "Price Momentum in Stocks: Insights from Victorian Age Data." NBER Working Paper No. w14500.

Choi, H., P. Mueller, and A. Vedolin. 2017. "Bond Variance Risk Premiums." *Review of Finance* 21: 987–1022.

Chordia, T., A. Subrahmanyam, and Q. Tong. 2014. "Have Capital Market Anomalies Attenuated in the Recent Era of High Liquidity and Trading Activity?" *Journal of Accounting and Economics* 58: 41–58.

Chen, E., and A. Clements. 2007. "S&P 500 Implied Volatility and Monetary Policy Announcements." *Finance Research Letters* 4: 227–232.

Chung, S., and H. Lewis. 2017. "Earnings Announcements and Option Returns." *Journal of Empirical Finance* 40: 220–235.

Clare, A., J. Seaton, P. Smith, and S. Thomas. 2013. "Breaking into the Blackbox: Trend Following, Stop Losses and the Frequency of Trading–The Case of the S&P500." *Journal of Asset Management* 14: 182–194.

Cootner, P. 1964. *The Random Character of Stock Market Prices*. Cambridge, MA: MIT Press.

Cvitanic, J., and I. Karatzas. 1995. "On Portfolio Optimization under 'Drawdown' Constraints." *IMA Volumes in Mathematics and Its Applications* 65: 35–46.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam. 1998. "Investor Psychology and Security Market Under- and Overreactions."

Journal of Finance 53: 1839–1885.

De Grauwe, P., and M. Grimaldi. 2004. "Bubbles and Crashes in a Behavioural Finance Model." Sveriges Riksbank Working Paper No. 164, Stockholm, May.

DellaVigna, S., and J. Pollet. 2009. "Investor Inattention and Friday Earnings Announcements." *Journal of Finance* 64: 709–749.

Dew-Becker, I., S. Giglio, A. Le, and M. Rodriguez. 2014. "The Price of Variance Risk." Working paper, Kellogg.

Diba, B., and H. Grossman. 1988. "The Theory of Rational Bubbles in Stock Prices." *The Economic Journal* 98: 746–754.

Di Pietro, V., and G. Vainberg. 2006. "Systematic Variance Risk and Firm Characteristics in the Equity Options Market." SSRN: 858304.

do Vale, R., R. Pieters, and M. Zeelenberg. 2016. "The Benefits of Behaving Badly on Occasion: Successful Regulation by Planned Hedonic Deviations." *Journal of Consumer Psychology* 26: 17–28.

Driessen, J., and P. Maenhout. 2006. "The World Price of Jump and Volatility Risk." Working paper, Insead.

Driessen, J., P. Maenhout, and G. Vilkov. 2009. "The Price of Correlation Risk: Evidence from Equity Options." *Journal of Finance* 64: 1377–1406.

Dyson, F. 2004. "A Meeting with Enrico Fermi." *Nature* 427 (January 22): 297.

Easton, P., G. Gao, and P. Gao. 2010. "Pre-earnings Announcement Drift." SSRN: 1786697.

Ederington, L., and J. Lee. 1996. "The Creation and Resolution of Market Uncertainty: The Impact of Information Releases on Implied Volatility." *Journal of Financial and Quantitative Analysis* 31: 513–539.

Engle, R., and J. Rosenberg. 2002. "Empirical Pricing Kernels." *Journal of Financial Economics* 64: 341–372.

Erb, C., and C. Harvey. 2006. "The Strategic and Tactical Value of Commodity Futures." *Financial Analysts Journal* 62: 69–97.

Fama, E. F. 1965a. "The Behavior of Stock-Market Prices." *The Journal of Business* 38: 34–105.

Fama, E. F. 1965b. "Random Walks in Security Prices." *Financial Analysts Journal* 41: 55–59.

Fama, E. F. 1998. "Market Efficiency, Long-Term Returns, and Behavioral Finance." *Journal of Financial Economics* 49: 283–306.

Fama, E., and K. French. 1997. "Industry Costs of Equity." *Journal of Financial Economics* 43: 153–193.

Fama, E., and K. French. 2010. "Size, Value and Momentum in International Stock Returns." Fama-Miller Working Paper, Tuck School of Business Working Paper No. 2011–85, Chicago Booth Research Paper No. 11–10.

Fernandez-Perez, A., B. Frijns, and A. Tourani-Rad. 2017. "When No News Is Good News–The Decrease in Investor Fear after the FOMC Announcement." *Journal of Empirical Finance* 41: 187–199.

Fisher, I. 1930. *The Theory of Interest, as Determined by Impatience to Spend Income and Opportunity to Invest It*. New York: McMillan.

Fligstein, N., and L. Dauter. 2007. "The Sociology of Markets." *The Annual Review of Sociology* 33: 105–128.

Foster, G., C. Olsen, and T. Shelvin. 1984. "Earnings Releases, Anomalies and the Behavior of Security Returns." *The Accounting Review* 59: 574–603.

Fuss, R., F. Mager, H. Wohlenberg, and L. Zhao. 2011. "The Impact of Macroeconomic Announcements on Implied Volatility." *Applied Financial Economics* 21: 1571–1580.

Gao, C., Y. Xing, and X. Zhang. 2017. "Anticipating Uncertainty: Straddles around Earnings Announcements." SSRN: 2204549.

Geczy, C., and M. Samnov. 2016. "Two Centuries of Price Return Momentum." *Financial Analysts Journal* 72: 32–56.

Gorton, G., F. Hayashi, and K. Rouwenhorst. 2013. "The Fundamentals of Commodity Futures Returns." *Review of Finance* 17: 35–105.

Gorton, G., and K. Rouwenhorst. 2006. "Facts and Fantasies about Commodity Futures." *Financial Analysts Journal* 62: 47–68.

Gospodinov, N., and I. Jamali. 2012. "The Effects of Federal Funds Rate Surprises on S&P500 Volatility and Volatility Risk Premium." *Journal of Empirical Finance* 19: 497–510

Gray, W. 2014. "Are You Trying Too Hard?" SSRN: 2481675.

Grosshans, D., and S. Zeisberger. 2018. "All's Well That Ends Well? On the Importance of How Returns Are Achieved." *Journal of Banking and Finance* 87: 397–410.

Grossman, S., and J. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review* 70: 393–408.

Grossman, S., and Z. Zhou., 1993. "Optimal Investment Strategies for Controlling Drawdowns." *Mathematical Finance* 3: 241–276.

Han, Y., G. Zhou, and Y. Zhu. 2016. "Taming Momentum Crashes: A Simple Stop-Loss Strategy." SSRN: 2407199.

Hanley, J. A., and A. Lippman-Hand. 1983. "If Nothing Goes Wrong Is Everything All Right?" *Journal of the American Medical Association* 249: 1743–1745.

Harvey, C., Y. Liu, and Z. Hequing. 2016. "… and the Cross-Section of Expected Returns." *The Review of Financial Studies* 29: 5–68.

He, W., Y. Lee, and P. Wei. 2010. "Do Option Traders on Value and Growth Stocks React Differently to New Information?" *Review of Quantitative Finance and Accounting* 34: 371–381.

Higgins, L. 1906. *The Put-and-Call*. London: Effingham Wilson.

Hirshleifer, D., S. Lim, and S. Teoh. 2009. "Driven to Distraction: Extraneous Events and Underreaction to Earnings News." *Journal of Finance* 64: 2289–2325.

Hodges, S. 1998. “A Generalization of the Sharpe Ratio and Its Applications to Valuation Bounds and Risk Measures.” Working paper, Financial Options Research Centre, University of Warwick.

Hodges, S., R. Tompkins, and W. Ziemba. 2003. “The Favorite/Long-Shot Bias in S&P 500 and FTSE 100 Index Futures Options: The Return to Bets and the Cost of Insurance.” EFA 2003 Annual Conference Paper No. 135.

Hogan, R. 2011. “Quantifying the Variance Risk Premium in VIX Options.” CMC Senior Thesis Paper 147. http://scholarship.claremont.edu/cmc_theses/147.

Hou, K., C. Xue, and L. Zhang. 2017. “Replicating Anomalies.” NBER Working Papers 23394, National Bureau of Economic Research.

Huang, D., C. Schlag, I. Shaliastovich, and J. Thime. 2018. “Volatility of Volatility Risk.” *Journal of Financial and Quantitative Analysis* 53: 1–63.

Hubbard, D. W. 2007. *How to Measure Anything: Finding the Value of Intangibles in Business*. 3rd ed. New York: Wiley.

Hull, J., and A. White. 2017. “Optimal Delta Hedging for Options.” *Journal of Banking and Finance* 82: 180–190.

Ioannidis, J. 2005. “Why Most Published Research Findings Are False.” *PLoS Medicine* 2: e124.

Israelov, R., and H. Tummala. 2017. “Which Index Options Should You Sell.” SSRN: 2990542.

Jansen, I., and A. Nikiforov. 2016. “Fear and Greed: A Returns-Based Trading Strategy around Earnings Announcements.” *Journal of Portfolio Management* 42: 88–95.

Jenkinson, T., H. Jones, and J. Martinez. 2015. “Picking Winners? Investment Consultants' Recommendations of Fund Managers.” *The Journal of Finance* 71: 2333–2370.

Jones, C., O. Lamont, and R. Lumsdaine. 1998. “Macroeconomic News and Bond Market Volatility.” *Journal of Financial Economics* 47: 315–337.

Jones, C., and J. Shemesh. 2017. "Option Mispricing around Nontrading Periods." *The Journal of Finance* 73: 861–900.

Jung, D. 2000. "The Model Risk of Option Pricing Models When Volatility Is Stochastic: A Monte Carlo Simulation Approach." PhD dissertation, University of Missouri–Columbia.

Jung, D., and C. Corrado. 2009. "Tweaking Black Scholes." SSRN: 1330248.

Kaeck, A. 2017. "Variance-of-Variance Risk Premium." *Review of Finance* 22: 1549–1579.

Katona, G. 1975. *Psychological Economics*. New York: Elsevier.

Kelly, B., and H. Jiang. 2014. "Tail Risk and Asset Prices." *Review of Financial Studies* 27: 2841–2871.

Kelly, J. 1956. "A New Interpretation of Information Rate." *Bell System Technical Journal* 35: 917–926.

Keynes, J. M. 1930. *A Treatise on Money*. London: MacMillan.

Kim, Y.-J. 2002. "Option Pricing under Stochastic Interest Rates: An Empirical Investigation." *Asia-Pacific Financial Markets* 9: 23–44.

Kipling, R. 1910. *"If"* in *Rewards and Fairies*. London: Doubleday.

Knight, F. 1921. *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin.

Koijen, R., M. Schmeling, and E. Vrugt. 2015. "Survey Expectations of Returns and Asset Pricing Puzzles." London Business School IFA working paper.

Kozhan, R., A. Neuberger, and P. Schneider. 2011. "The Skew Risk Premium in Index Option Prices." AFA 2011 Denver Meetings Paper. SSRN: 1571700.

Lamont, O., and A. Frazzini. 2007. "The Earnings Announcement Premium and Trading Volume." NBER Working Papers 13090, National Bureau of Economic Research.

Lee, Y. 2017. "Market Reactions to Unexpected Relative Earnings Performance." *Asia-Pacific Journal of Accounting and*

Economics 24: 339–357.

Lei, A., and H. Li. 2009. “The Value of Stop Loss Strategies.” *Financial Services Review* 18: 23–51.

Lempérière, Y., C. Deremble, P. Seager, M. Potters, and J. P. Bouchaud. 2014. “Two Centuries of Trend Following.” *Journal of Investment Strategies* 3: 41–61.

Li, M., and N. Pearson. 2007. “A ‘Horse Race’ Among Competing Option Pricing Models Using S&P 500 Index Options.” SSRN: 952770.

Lintner, J. 1965. “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets.” *The Review of Economics and Statistics* 47: 13–37.

Lo, B., and S. Zhang. 2005. “The Volatility Risk Premium Embedded in Currency Options.” *Journal of Financial and Quantitative Analysis* 40: 803–832.

Londono, J. 2011. “The Variance Risk Premium around the World.” Federal Reserve International Discussion Papers, Number 1035.

Londono, J., and H. Zhou. 2017. “Variance Risk Premiums and the Forward Premium Puzzle.” *Journal of Financial Economics* 124: 415–440.

Louis, T. 1981. “Confidence Intervals for a Binomial Parameter after Observing No Successes.” *The American Statistician* 35: 154.

Lucca, D., and E. Moench. 2011. “The pre-FOMC Announcement Drift.” Staff Report, No. 512, Federal Reserve Bank of New York.

Maillard, D. 2018. “Adjusted Sharpe Ratio: Some Caveats.” SSRN: 3284396.

Mandelbrot, M. 1963. “The Variation of Certain Speculative Prices.” *The Journal of Business* 36: 394–419.

Markopolos, H. 2010. *No One Would Listen: A True Financial Thriller*. New York: Wiley.

Mason, P., and S. Utke. 2019. “Investor Taxes and Option Prices.” SSRN: 332680.

Massey, A., and A. Hill. 2012. “Dieting and Food Craving: A Descriptive, Quasi-Prospective Study.” *Appetite* 58: 781–785.

McLean, R., and J. Pontiff. 2016. “Does Academic Research Destroy Stock Return Predictability?” *Journal of Finance* 71: 5–32.

Merton, R. 1973. “Theory of Rational Option Pricing.” *The Bell Journal of Economics and Management Science* 4: 141–183.

Mills, F. 1927. “The Behavior of Prices.” National Bureau of Economic Research, New York.

Mitchell, W. 1915. “The Making and Using of Index Numbers.” Bulletin 1973 of the U.S. Bureau of Labor Statistics.

Mossin, J. 1966, “Equilibrium in a Capital Asset Market.” *Econometrica* 34: 768–783.

Muravyev, D., and X. Ni. 2018. “Why Do Option Returns Change Sign from Day to Night?” SSRN: 2820264.

Nietzsche, F. 1878, “Human, All Too Human,” Section Nine. *Man Alone with Himself*, aphorism 494.

Nikkinen, J., and P. Sahlström. 2004. “Impact of the Federal Open Market Committee's Meetings and Scheduled Macroeconomic News on Stock Market Uncertainty.” *International Review of Financial Analysis* 13: 1–12.

Oehlert, G. 1992. “A Note on the Delta Method.” *American Statistician* 46: 27–29.

Olivier, M. 1926. “Les Nombres Indices de la Variation des Prix.” Doctoral Dissertation: Paris

Osborne, M. 1959. “Brownian Motion in the Stock Market.” *Operations Research* 7: 145–173, 807–811.

Patell, J., and M. Wolfson. 1979. “Anticipated Information Releases Reflected in Call Option Prices.” *Journal of Accounting and Economics* 1: 117–140.

Park, Y. 2015. “Volatility-of-Volatility and Tail Risk Hedging Returns.” *Journal of Financial Markets* 26: 38–63.

Pézier, J. 2004. “Risk and Risk Aversion. In *The Professional Risk Managers' Handbook,* ed. C. Alexander and E. Sheedy. New York: PRMIA Publications.

Philippon, T. 2012. “Finance versus Wal-Mart: Why Are Financial Services So Expensive.” In *Rethinking the Financial Crisis,* ed. A. Blinder, A. Lo, and R. Solow. New York: Russell Sage Foundation.

Poon, S. 2005. *A Practical Guide to Forecasting Financial Market Volatility*. London: Wiley.

Poon, S., and C.W.J. Granger. 2003. “Forecasting Volatility in Financial Markets: A Review.” *Journal of Economic Literature* 41 (2): 478–539.

Prokopczuk, M., and C. Simen. 2014. “Variance Risk Premia in Commodity Markets.” SSRN: 2195691.

Ramnath, S. 2002. “Investor and Analyst Reactions to Earnings Announcements of Related Firms: An Empirical Analysis.” *Journal of Accounting Research* 40: 1351–1376.

Ross, S. 1976. “The Arbitrage Theory of Capital Asset Pricing.” *Journal of Economic Theory* 13: 341–360.

Ruan, X. 2017. “Cross Section of Option Returns and Volatility-of-Volatility.” SSRN: 3055982.

Samuelson, P. 1965. “Proof That Properly Anticipated Prices Fluctuate Randomly.” *Industrial Management Review* 6: 41–49.

Scholes, M. 1976. “Taxes and the Pricing of Options.” *The Journal of Finance* 31: 319–332.

Shaikh, I. and P. Padhi. 2013. “Macroeconomic Announcements and Implied Volatility Index: Evidence from India VIX.” *Margin: The Journal of Applied Economic Research* 7: 417–442.

Sharpe, W. 1964. “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.” *The Journal of Finance*

19: 425–442.

Simon, D., and J. Campasano. 2014. “The VIX Futures Basis: Evidence and Trading Strategies.” *The Journal of Derivatives* 21: 54–69.

Sinclair, E. 2010. *Option Trading: Pricing and Volatility Strategies and Techniques*. New York: Wiley.

Sinclair, E. 2013. *Volatility Trading*. 2nd ed. New York: Wiley.

Sinclair, E. 2014. “Confidence Intervals for the Kelly Criterion.” *Journal of Investment Strategies* 3: 65–74.

Sinclair, E. 2018. “Reflections on Recent Volatility.” *Journal of Investment Strategies* 7: 77–82.

Sinclair, E., and R. Brooks. 2017 “The Skewness and Kurtosis of European Options and the Implications for Trade Sizing.” SSRN: 2956161.

So, E., and S. Wang. 2014. “News-Driven Return Reversals: Liquidity Provision ahead of Earnings Announcements.” *Journal of Financial Economics* 114: 20–35.

Taylor, D. 2011. “Retail Investors and the Adjustment of Stock Prices to Earnings Information.” Wharton School Working Papers, University of Pennsylvania.

Tosi, A., and A. Ziegler. 2017. “The Timing of Option Returns.” SSRN: 2909163.

Treynor, J. 1962. “Toward a Theory of Market Value of Risky Assets.” Unpublished manuscript.

Vähämaa, S., and J. Äijö. 2011. “The Fed's Policy Decisions and Implied Volatility.” *The Journal of Futures Markets* 31: 995–1010.

van Binsbergen, J., and R. Koijen. 2015. “The Term Structure of Returns: Facts and Theory.” NBER Working Paper 21234.

Vasquez, A. 2017. “Equity Volatility Term Structures and the Cross Section of Option Returns.” *Journal of Financial and Quantitative Analysis* 52: 1–28.

Vilkov, G. 2008. "Variance Risk Premium Demystified." SSRN: 891360.

Watts, R. L. 1978. "Systematic 'Abnormal' Returns After Quarterly Earnings Announcements." *Journal of Financial Economics* 6: 127–150.

Yung, H., and H. Zhang. 2003. "An Empirical Investigation of the GARCH Option Pricing Model: Hedging Performance." *Journal of Futures Markets* 23: 1191–1207.

Zhang, L. 2007. "Sample Mean and Sample Variance: Their Covariance and Their (In)Dependence." *American Statistician* 61: 159–160.

Index

J

K

L

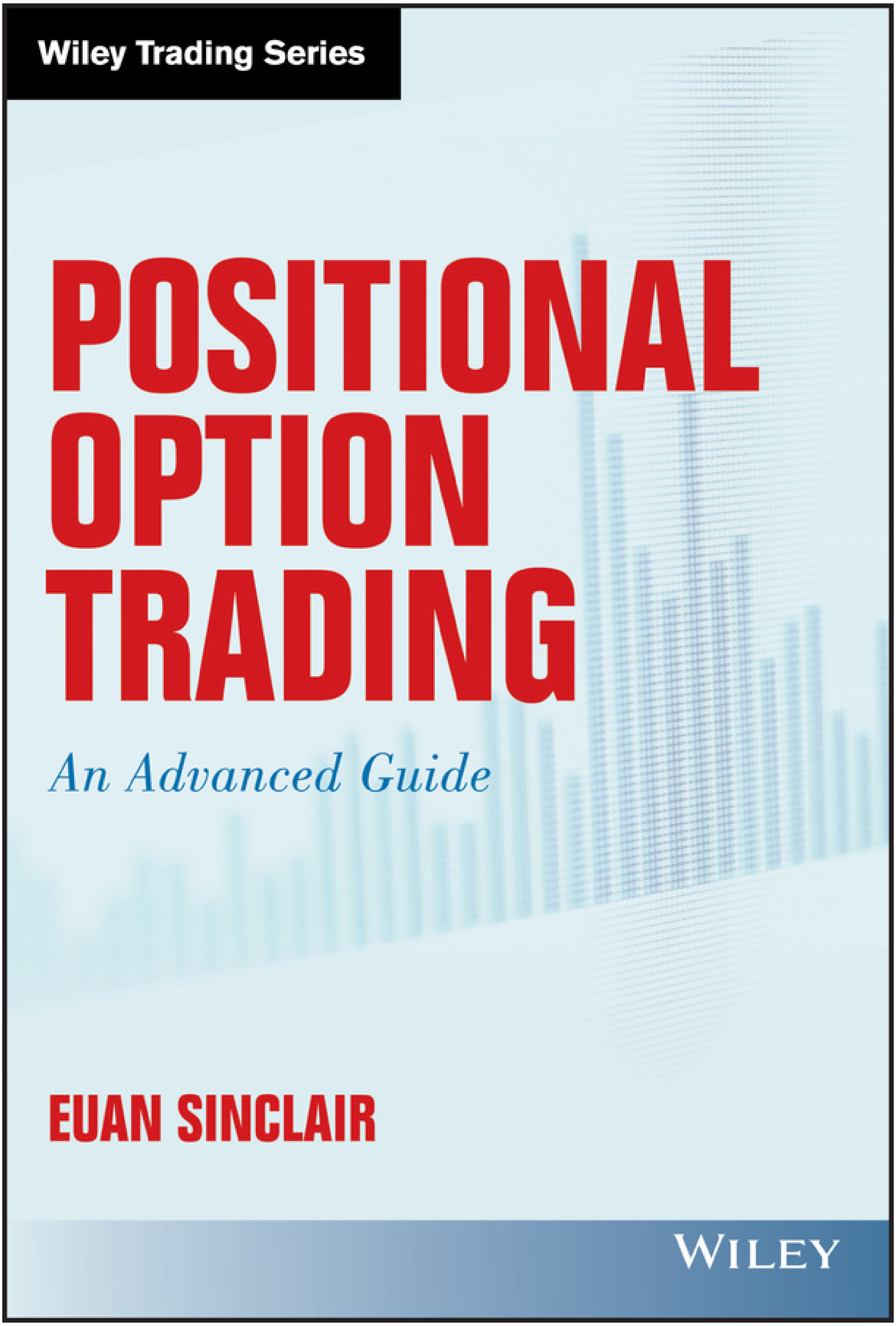
M

Q

R

S

WILEY END USER LICENSE AGREEMENT


Wiley Trading Series
POSITIONAL OPTION TRADING
An Advanced Guide
EUAN SINCLAIR
WILEY