For our example in Table 6-1 we derive $Cov_{XY} = -0.1567$. Equation (6.3) is = Covariance.S in Excel and cov in MATLAB. The covariance is not easy to interpret, since it takes values between $-\infty$ and $+\infty$. Therefore, it is more convenient to use the Pearson correlation coefficient ρ_{XY}, which is a standardized covariance; that is, it takes values between −1 and +1. The Pearson correlation coefficient is:

$$\rho_{XY} = \frac{Cov_{XY}}{\sigma_X \sigma_Y} \tag{6.4}$$

For our example in Table 6-1, $\rho_{XY} = -0.7403$, showing that the returns of assets *X* and *Y* are highly negatively correlated. Equation (6.4) is 'correl' in Excel and 'corrcoef' in MATLAB. For the derivation of the numerical examples of Equations (6.2) to (6.4) and more information on the covariances, see Appendix A.

We can calculate the standard deviation for our two-asset portfolio *P* as

$$\sigma_P = \sqrt{w_X^2\sigma_X^2 + w_Y^2\sigma_Y^2 + 2w_X w_Y Cov_{XY}} \tag{6.5}$$

With equal weights, i.e., $w_X = w_Y = 0.5$, the example in Table 6-1 results in $\sigma_P = 16.66\%$.

Importantly, the standard deviation (or its square, the variance) is interpreted in finance as risk. The higher the standard deviation, the higher the risk of an asset or a portfolio. Is standard deviation a good measure of risk? The answer is: It's not great, but it's one of the best we have. A high standard deviation may mean high upside potential, so it penalizes possible profits! But a high standard deviation naturally also means high downside risk. In particular, risk-averse investors will not like a high standard deviation, i.e., high fluctuation of their returns.

An informative performance measure of an asset or a portfolio is the risk-adjusted return, i.e., the return/risk ratio. For a portfolio it is μ_P/σ_P, which we derived in Equations (6.1) and (6.5). In Figure 6-3 we observe one of the few free lunches in finance: the lower (preferably negative) the correlation of the assets in a portfolio, the higher the return/risk ratio. For a rigorous proof, see Markowitz (1952) and Sharpe (1964).

Figure 6-3 shows the high impact of correlation on the portfolio return/risk ratio. A high negative correlation results in a return/risk ratio of close to 250%, whereas a high positive correlation results in a 50% ratio. The Equations (6.1) to (6.5) are derived within the framework of the Pearson correlation approach. We will discuss the limitations of this approach in Chapter 8.

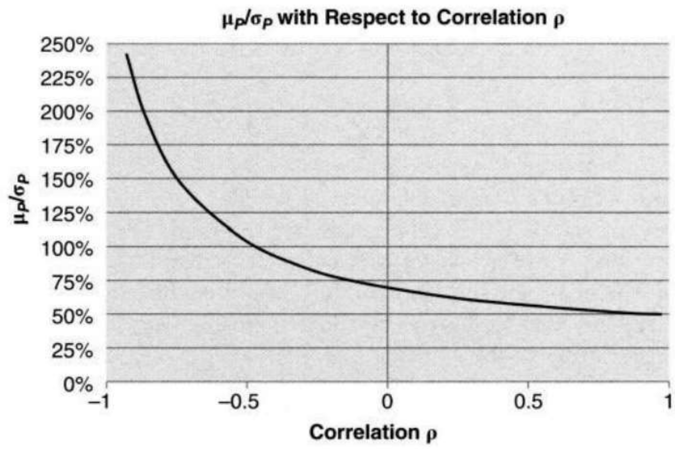


FIGURE 6-3 The negative relationship of the portfolio return/risk ratio μ_P/σ_P with respect to the correlation ρ of the assets in the portfolio (input data are from Table 6-1).

Only by great risks can great results be achieved.

—Xeres

Trading and Correlation

In finance, every risk is also an opportunity. Therefore, at every major investment bank and hedge fund *correlation desks* exist. The traders try to forecast changes in correlation and attempt to financially gain from these changes in correlation. We already mentioned the correlation strategy "pairs trading." Generally, *correlation trading* means trading assets whose prices are determined at least in part by the comovement of one or more asset in time. Many types of correlation assets exist.

Multi-Asset Options

A popular group of correlation options are multi-asset options, also termed rainbow options or mountain range options. Many different types are traded. The most popular ones are listed here. S_1 is the price of asset 1 and S_2 is the price of asset 2 at option maturity. *K* is the strike price, i.e., the price determined at option start, at which the underlying asset can be bought in the case of a call, and the price at which the underlying asset can be sold in the case of a put.

- Option on the better of two. Payoff = max(S_1, S_2).
- Option on the worse of two. Payoff = min(S_1, S_2).
- Call on the maximum of two. Payoff = max[0, max(S_1, S_2) − K].
- Exchange option (as a convertible bond). Payoff = max(0, $S_2 - S_1$).
- Spread call option. Payoff = max[0, ($S_2 - S_1$) − K].
- Option on the better of two or cash. Payoff = max(S_1, S_2, cash).
- Dual-strike call option. Payoff = max(0, $S_1 - K_1$, $S_2 - K_2$).
- Portfolio of basket options. Payoff = $\left[\sum_{i=1}^{n} n_i S_i - K, 0\right]$ where n_i is the weight of assets i.

Importantly, the prices of these correlation options are highly sensitive to the correlation between the asset prices S_1 and S_2. In the list above, except for the option on the worse of two, the lower the correlation, the higher the option price. This makes sense since a low, preferable negative correlation means that if one asset decreases, on average the other increases. So one of the two assets is likely to result in a high price and a high payoff. Multi-asset options can be conveniently priced using closed form extensions of the Black-Scholes-Merton 1973 option model.

Let's look at the evaluation of an exchange option with a payoff of max(0, $S_2 - S_1$). The payoff shows that the option buyer has the right to give away asset 1 and receive asset 2 at option maturity. Hence, the option buyer will exercise her right if $S_2 > S_1$. The price of the exchange option can be derived easily. We first rewrite the payoff equation max(0, $S_2 - S_1$) = S_1 max[0, (S_2/S_1) − 1]. We then input the covariance between asset S_1 and S_2 into the implied volatility function of the exchange option using a variation of Equation (6.5):

$$\sigma_E = \sqrt{\sigma_A^2 + \sigma_B^2 - 2\mathrm{Cov}_{AB}} \qquad \textbf{(6.5a)}$$

where σ_E is the implied volatility of S_2/S_1 which is input into the standard Black-Scholes-Merton 1973 option pricing model.

Importantly, the exchange option price is highly sensitive to the correlation between the asset prices S_1 and S_2, as seen in Figure 6-4.

From Figure 6-4 we observe the strong impact of the correlation on the exchange option price. The price is close to 0 for high correlation and $15.08 for a negative correlation

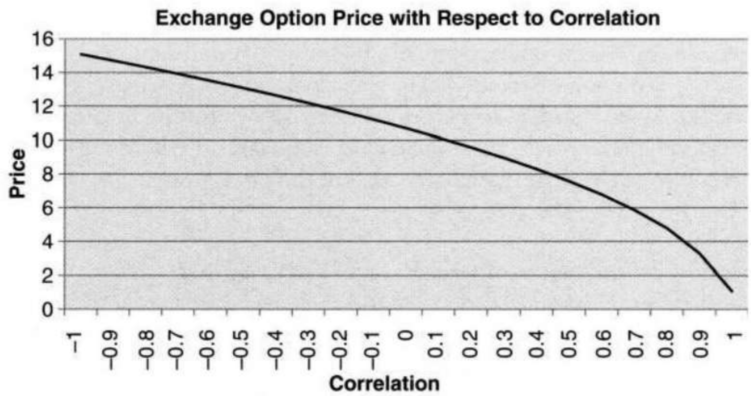


FIGURE 6-4 Exchange option price with respect to correlation of the assets in the portfolio.

of −1. As in Figures 6-2 and 6-3, the correlation approach underlying Figure 6-4 is the Pearson correlation model. We will discuss the limitations of the Pearson correlation model in Chapter 8.

Quanto Option

Another interesting correlation option is the quanto option. This is an option that allows a domestic investor to exchange his potential option payoff in a foreign currency back into his home currency at a fixed exchange rate. A quanto option therefore protects an investor against currency risk. For example, an American believes the Nikkei will increase, but she is worried about a decreasing yen, which would reduce or eliminate her profits from the Nikkei call option. The investor can buy a quanto call on the Nikkei, with the yen payoff being converted into dollars at a fixed (usually the spot) exchange rate.

Originally, the term *quanto* comes from the word *quantity,* meaning that the amount that is reexchanged to the home currency is unknown, because it depends on the future payoff of the option. Therefore the financial institution that sells a quanto call does not know two things:

1. How deep in the money the call will be, i.e., which yen amount has to be converted into dollars.
2. The exchange rate at option maturity at which the stochastic yen payoff will be converted into dollars.

The correlation between (1) and (2) i.e., the price of the underlying S' and the exchange rate X, significantly influences the quanto call option price. Let's consider a call on the Nikkei S' and an exchange rate X defined as domestic currency per unit of foreign currency (so \$/1 yen for a domestic American) at maturity.

If the correlation is positive, an increasing Nikkei will also mean an increasing yen. That is favorable for the call seller. She has to settle the payoff, but only needs a small yen amount to achieve the dollar payment. Therefore, the more positive the correlation coefficient, the lower the price for the quanto option. If the correlation coefficient is negative, the opposite applies: If the Nikkei increases, the yen decreases in value. Therefore, more yen are needed to meet the dollar payment. As a consequence, the lower the correlation coefficient, the more expensive the quanto option. Hence we have a similar negative relationship between the option price and correlation as in Figure 6-2.

Quanto options can be conveniently priced closed form applying an extension of the Black-Scholes-Merton 1973 model.

Correlation Swap

The correlation between assets can also be traded directly with a correlation swap. In a correlation swap a fixed (i.e., known) correlation is exchanged with the correlation that will actually occur, called realized or stochastic (i.e., unknown) correlation, as seen in Figure 6-5.

Paying a fixed rate in a correlation swap is also called *buying correlation*. This is because the present value of the correlation swap will increase for the correlation buyer if the realized correlation increases. Naturally the fixed rate receiver is *selling correlation*.

The realized correlation ρ in Figure 6-5 is the correlation between the assets that actually occurs during the time of the swap. It is calculated as:

$$\rho_{\text{realized}} = \frac{2}{n^2 - n} \sum_{i>j} \rho_{i,j} \qquad \textbf{(6.6)}$$

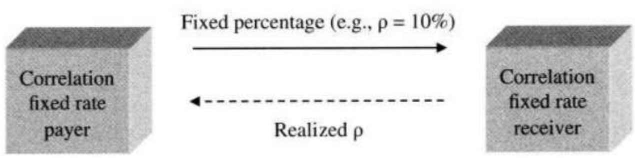


FIGURE 6-5 A correlation swap with a fixed 10% correlation rate.

where $\rho_{i,j}$ is the Pearson correlation between asset i and j, and n is the number of assets in the portfolio.

The payoff of a correlation swap for the correlation fixed rate payer at maturity is:

$$N(\rho_{\text{realized}} - \rho_{\text{fixed}}) \qquad \textbf{(6.7)}$$

where N is the notional amount. Let's look at an example of a correlation swap.

Correlation swaps can indirectly protect against decreasing stock prices. As we will see in this chapter, as well as in Chapter 7, when stocks decrease, typically the correlation between the stocks increases. Hence a fixed correlation payer protects himself indirectly against a stock market decline.

Example 6.1 Payoff of a Correlation Swap

What is the payoff of a correlation swap with three assets, a fixed rate of 10%, a notional amount of \$1,000,000, and a 1-year maturity?

First, the daily log returns $\ln(S_t/S_{t-1})$ of the three assets are calculated for 1 year.[1] Let's assume the realized pairwise correlations of the log returns at maturity are as displayed in Table 6-2.

The average correlation between the three assets is derived by Equation (6.6). We apply the correlations only in the white area from Table 6-2, since these satisfy $i > j$.

Hence we have $\rho_{\text{realized}} = \frac{2}{3^2 - 3}(0.5 + 0.3 + 0.1) = 0.3$.

Following Equation (6.7), the payoff for the correlation fixed rate payer at swap maturity is \$1,000,000 × (0.3 − 0.1) = \$200,000.

TABLE 6-2 Pairwise Pearson Correlation Coefficient at Swap Maturity

	$S_{j=1}$	$S_{j=2}$	$S_{j=3}$
$S_{i=1}$	1	0.5	0.1
$S_{i=2}$	0.5	1	0.3
$S_{i=3}$	0.1	0.3	1

[1] Log returns $\ln(S_1/S_0)$ are an approximation of percentage returns $(S_1 - S_0)/S_0$. We typically use log returns in finance since they are additive in time, whereas percentage returns are not. For details see Appendix B.

Currently, year 2013, there is no industry-standard valuation model for correlation swaps. Traders often use historical data to anticipate $\rho_{realized}$. In order to apply swap valuation techniques, we require a term structure of correlation in time. However, no correlation term structure currently exists. We can also apply stochastic correlation models to value a correlation swap. Stochastic correlation models are currently emerging.

Buying Call Options on an Index and Selling Call Options on Individual Components

Another way of buying correlation (i.e., benefiting from an increase in correlation) is to buy call options on an index such as the Dow Jones Industrial Average (the Dow) and sell call options on individual stocks of the Dow. As we will see in Chapter 7, there is a positive relationship between correlation and volatility. Therefore, if correlation between the stocks of the Dow increases, so will the implied volatility[2] of the call on the Dow. This increase is expected to outperform the potential loss from the increase in the short call positions on the individual stocks.

Creating exposure on an index and hedging with exposure on individual components is exactly what the "London whale," JPMorgan's London trader Bruno Iksil, did in 2012. Iksil was called the London whale because of his enormous positions in credit default swaps (CDSs).[3] He had sold CDSs on an index of bonds, the CDX.NA.IG.9, and hedged them by buying CDSs on individual bonds. In a recovering economy this is a promising trade: Volatility and correlation typically decrease in a recovering economy. Therefore, the sold CDSs on the index should outperform (decrease more than) the losses on the CDSs of the individual bonds.

But what can be a good trade in the medium and long term can be disastrous in the short term. The positions of the London whale were so large that hedge funds short-squeezed him: They started to aggressively buy the CDS index CDX.NA.IG.9. This increased the CDS values in the index and created a huge (paper) loss for the whale. JPMorgan was forced to buy back the CDS index positions at a loss of over $2 billion.

Paying Fixed in a Variance Swap on an Index and Receiving Fixed on Individual Components

A further way to buy correlation is to pay fixed in a variance swap on an index and to receive fixed in variance swaps on individual components of the index. The idea is the same as the idea with respect to buying a call on an index and selling a call on the individual components: If correlation increases, so will the variance. As a consequence, the present value for the variance swap buyer, the fixed variance swap payer, will increase. This increase is expected to outperform the potential losses from the short variance swap positions on the individual components.

In the preceding trading strategies, the correlation between the assets was assessed with the Pearson correlation approach. As mentioned, we will discuss the limitations of this model in Chapter 8.

Risk Management and Correlation

After the global financial crisis from 2007 to 2009, financial markets have become more risk averse. Commercial banks, investment banks, as well as nonfinancial institutions have increased their risk management efforts. As in the investment and trading environment, correlation plays a vital part in risk management. Let's first clarify what risk management means in finance.

Financial risk management is the process of identifying, quantifying, and, if desired, reducing financial risk. The three main types of financial risk are:

1. Market risk.
2. Credit risk.
3. Operational risk.

Additional types of risk may include systemic risk, liquidity risk, volatility risk, and correlation risk. We will concentrate in this chapter on market risk. Market risk consists of four types of risk: (1) equity risk, (2) interest rate risk, (3) currency risk, and (4) commodity risk.

There are several concepts to measure the market risk of a portfolio, such as value-at-risk (VaR), expected shortfall

[2] Implied volatility is volatility derived (implied) by option prices. The higher the implied volatility, the higher the option price.

[3] Simply put, a credit default swap (CDS) is an insurance against default of an underlying (e.g., a bond). However, if the underlying is not owned, a long CDS is a speculative instrument on the default of the underlying (just like a naked put on a stock is a speculative position on the stock going down). See Meissner (2005) for more.

(ES), enterprise risk management (ERM), and more. VaR is currently (year 2013) the most widely applied risk management measure. Let's show the impact of asset correlation on VaR.[4]

First, what is value-at-risk (VaR)? VaR measures the maximum loss of a portfolio with respect to a certain probability for a certain time frame. The equation for VaR is:

$$\mathrm{VaR}_P = \sigma_P \alpha \sqrt{x} \tag{6.8}$$

where VaR_P is the value-at-risk for portfolio P, and α is the abscise value of a standard normal distribution corresponding to a certain confidence level. It can be derived as =normsinv(confidence level) in Excel or norminv (confidence level) in MATLAB. α takes the values $-\infty < \alpha < +\infty$. x is the time horizon for the VaR, typically measured in days; σ_P is the volatility of the portfolio P, which includes the correlation between the assets in the portfolio. We calculate σ_P via

$$\sigma_P = \sqrt{\beta_h C \beta_v} \tag{6.9}$$

where β_h is the horizontal β vector of invested amounts (price time quantity), β_v is the vertical β vector of invested amounts (also price time quantity),[5] and C is the covariance matrix of the returns of the assets.

Let's calculate VaR for a two-asset portfolio and then analyze the impact of different correlations between the two assets on VaR.

Example 6.2 Deriving VaR of a Two-Asset Portfolio

What is the 10-day VaR for a two-asset portfolio with a correlation coefficient of 0.7, daily standard deviation of returns of asset 1 of 2%, of asset 2 of 1%, and $10 million invested in asset 1 and $5 million invested in asset 2, on a 99% confidence level?

First, we derive the covariances (Cov):

$$\mathrm{Cov}_{11} = \rho_{11}\,\sigma_1\,\sigma_1 = 1 \times 0.02 \times 0.02 = 0.0004^{6}$$

$$\mathrm{Cov}_{12} = \rho_{12}\,\sigma_1\,\sigma_2 = 0.7 \times 0.02 \times 0.01 = 0.00014$$

$$\mathrm{Cov}_{21} = \rho_{21}\,\sigma_2\,\sigma_1 = 0.7 \times 0.01 \times 0.02 = 0.00014$$

$$\mathrm{Cov}_{22} = \rho_{22}\,\sigma_2\,\sigma_2 = 1 \times 0.01 \times 0.01 = 0.0001 \tag{6.10}$$

Hence our covariance matrix is $C = \begin{pmatrix} 0.0004 & 0.00014 \\ 0.00014 & 0.0001 \end{pmatrix}$

Let's calculate σ_P following Equation (6.9). We first derive $\beta_h C$

$$(10\ 5)\begin{pmatrix} 0.0004 & 0.00014 \\ 0.00014 & 0.0001 \end{pmatrix} = (10 \times 0.0004 + 5 \times 0.00014\ \ 10 \times 0.00014 + 5 \times 0.0001) = (0.0047\ 0.0019)$$

and then $(\beta_h C)\beta_v = (0.0047\ 0.0019)\begin{pmatrix} 10 \\ 5 \end{pmatrix} = 10 \times 0.0047 + 5 \times 0.0019 = 5.65\%$

Hence we have $\sigma_P = \sqrt{\beta_h C \beta_v} = \sqrt{5.65\%} = 23.77\%$.

We find the value for α in Equation (6.8) from Excel as =normsinv (0.99) = 2.3264, or from MATLAB as norminv(0.99) = 2.3264.

Following Equation (6.8), we now calculate the VaR_P as $0.2377 \times 2.3264 \times \sqrt{10} = 1.7486$.

Interpretation: We are 99% certain that we will not lose more than $1.75486 million in the next 10 days due to market price changes of assets 1 and 2.

The number $1.7486 million is the 10-day VaR on a 99% confidence level. This means that on average once in a hundred 10-day periods (so once every 1,000 days) this VaR number of $1. 7486 million will be exceeded. If we have roughly 250 trading days in a year, the company is expected to exceed the VaR about once every four years. The Basel Committee for Banking Supervision (BCBS) considers this to be too often. Hence, it requires banks, which are allowed to use their own models (called internal model-based approach), to hold capital for assets in the

[4] We use a variance-covariance VaR approach in this book to derive VaR. Another way to derive VaR is the nonparametric VaR. This approach derives VaR from simulated historical data. See Markovich (2007) for details.

[5] More mathematically, the vector β_h is the transpose of the vector β_v, and vice versa: $\beta_h^T = \beta_v$ and $\beta_v^T = \beta_h$. Hence we can also write Equation (6.9) as $\sigma_P = \sqrt{\beta_h C \beta_h^T}$.

[6] The attentive reader realizes that we calculated the covariance differently in Equation (6.3). In Equation (6.3) we derived the covariance from scratch, inputting the return values and means. In Equation (6.10) we are assuming that we already know the correlation coefficient ρ and the standard deviation σ.

trading book[7] in the amount of at least 3 times the 10-day VaR (plus a specific risk charge for credit risk). In Example 6.2, if a bank is granted the minimum of 3 times the VaR, a VaR capital charge for assets in the trading book of $1,7486 million × 3 = $5.2539 million is required by the Basel Committee.[8]

Let's now analyze the impact of different correlations between the asset 1 and asset 2 on VaR. Figure 6-6 shows the impact.

As expected, we observe from Figure 6-6 that the lower the correlation, the lower the risk, measured by VaR. Preferably the correlation is negative. In this case, if one asset decreases, the other asset on average increases, hence reducing the overall risk. The impact of correlation on VaR is strong. For a perfect negative correlation of −1, VaR is $1.1 million; for a perfect positive correlation, VaR is close to $1.9 million.

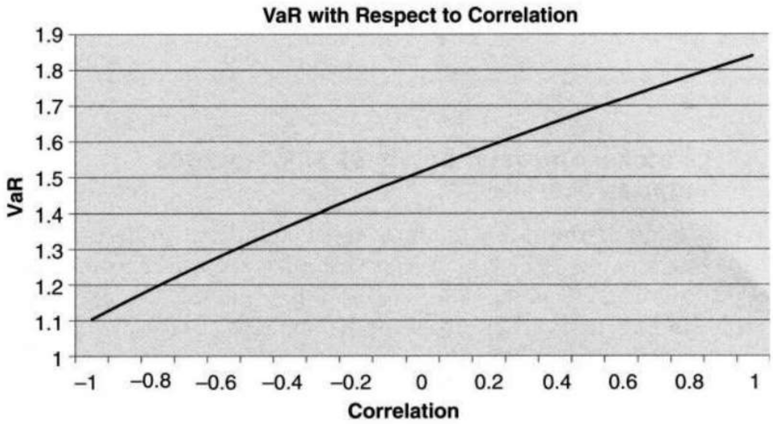


FIGURE 6-6 VaR of the two-asset portfolio of Example 6.2 with respect to correlation ρ between asset 1 and asset 2.

There are no toxic assets, just toxic people.

The Global Financial Crisis of 2007 to 2009 and Correlation

Currently, in 2013, the global financial crisis of 2007 to 2009 seems almost like a distant memory. The U.S. stock market has recovered from its low in March 2009 of 6,547 points and has more than doubled to over 15,000. World economic growth is at a moderate 2.5%. However, the U.S. unemployment rate is stubbornly high at around 8% and has not decreased to pre-crisis levels of about 5%. Most important, to fight the crisis, countries engaged in huge stimulus packages to revive their faltering economies. As a result, enormous sovereign deficits are plaguing the world economy. The European debt crisis, with Greece, Cyprus, and other European nations virtually in default, is a major global economic threat. The U.S. debt is also far from benign with a debt-to-GDP ratio of over 80%. One of the few nations that is enjoying these enormous debt levels is China, which is happy buying the debt and taking in the proceeds.

A crisis that brought the financial and economic system worldwide to a standstill is naturally not monocausal, but has many reasons. Here are the main ones:

- An extremely benign economic and risk environment from 2003 to 2006 with record low credit spreads, low volatility, and low interest rates.
- Increasing risk taking and speculation of traders and investors who tried to benefit in these presumably calm times. This led to a bubble in virtually every market segment, such as the housing market, mortgage market (especially the subprime mortgage market), stock market, and commodity market. In 2007, U.S. investors had borrowed 470% of the U.S. national income to invest and speculate in the real estate, financial, and commodity markets.
- A new class of structured investment products, such as collateralized debt obligations (CDOs), CDO-squareds, constant-proportion debt obligations (CPDOs), constant-proportion portfolio insurance (CPPI), as well as new products like options on credit default swaps (CDSs), credit indexes, and the like.
- The new copula correlation model, which was trusted naively by many investors and which could presumably correlate the $n(n - 1)/2$ assets in a structured product. Most CDOs contained 125 assets. Hence there

[7] Assets that are marked-to-market, such as stocks, futures, options, and swaps, are in the trading book. Some assets, such as loans and certain bonds, which are not marked-to-market, are in the banking book.

[8] In a recent Consultative Document (May 2012), the Basel Committee has indicated that it is considering replacing VaR with expected shortfall (ES). Expected shortfall measures tail risk (i.e., the size and probability of losses beyond a certain threshold). See www.bis.org/publ/bcbs219.pdf for details. Loosely speaking, VaR answers the question: "What is the maximum loss in good times?" Expected shortfall answers the question: "What is the loss in bad times?"

are 125(125 − 1)/2 = 7,750 asset correlation pairs to be quantified and managed.

- A moral hazard of rating agencies, which were paid by the same companies whose assets they rated. As a consequence, many structured products received AAA ratings and gave the illusion of little price and default risk.
- Risk managers and regulators who lowered their standards in light of the greed and profit frenzy. We recommend an excellent (anonymous) paper in the *Economist:* "A Personal View of the Crisis, Confessions of a Risk Manager."

Let's concentrate on the correlation aspect of the crisis. Around 2003, two years after the Internet bubble burst, the risk appetite of the financial markets increased, and investment banks, hedge funds, and private investors began to speculate and invest in the stock markets, commodity markets, and especially the real estate market.

In particular, residential mortgages became an investment object. The mortgages were packaged in collateralized debt obligations (CDOs), and then sold off to investors nationally and internationally. The CDOs typically consist of several tranches; that is, the investor can choose a particular degree of default risk. The equity tranche holder is exposed to the first 3% of mortgage defaults, the mezzanine tranche holder is exposed to the 3% to 7% of defaults, and so on. The new copula correlation model derived by Abe Sklar in 1959 and transferred to finance by David Li in 2000 could presumably manage the default correlations in the CDOs.

The first correlation-related crisis, which was a forerunner of the major one to come in 2007 to 2009, occurred in May 2005. General Motors was downgraded to BB and Ford was downgraded to BB+, so both companies were now in junk status. A downgrade to junk status typically leads to a sharp bond price decline, since many mutual funds and pension funds are not allowed to hold junk bonds.

Importantly, the correlation of the bonds in CDOs that referenced investment grade bonds decreased, since bonds of different credit qualities are typically lower correlated. This led to huge losses of hedge funds, which had put on a strategy where they were short the equity tranche of the CDO and long the mezzanine tranche of the CDO.

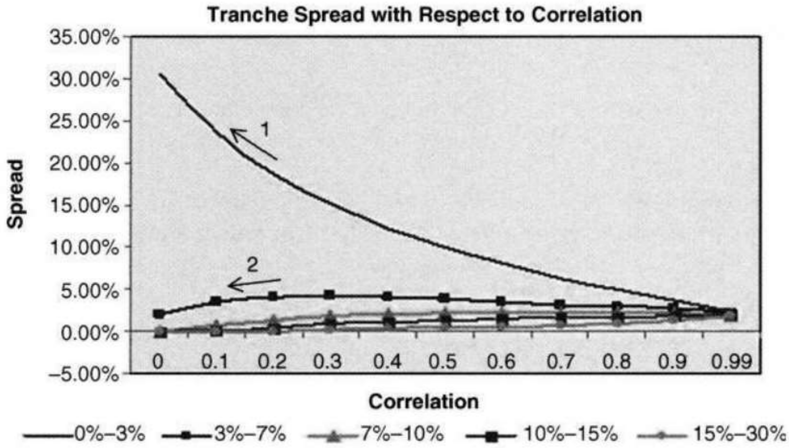


FIGURE 6-7 CDO tranche spread with respect to correlation.

Figure 6-7 shows the dilemma. Hedge funds had shorted the equity tranche[9] (0% to 3 % in Figure 6-7) to collect the high equity tranche spread. They had then presumably hedged[10] the risk by going long the mezzanine tranche[11] (3% to 7% in Figure 6-7). However, as we can see from Figure 6-7, this hedge is flawed.

When the correlations of the assets in the CDO decreased, the hedge funds lost on both positions.

1. The equity tranche spread increased sharply; see arrow 1. Hence the fixed spread that the hedge funds received in the original transaction was now significantly lower than the current market spread, resulting in a paper loss.
2. In addition, the hedge funds lost on their long mezzanine tranche positions, since a lower correlation lowers the mezzanine tranche spread; see arrow 2. Hence the spread that the hedge funds paid in the original transactions was now higher than the market spread, resulting in another paper loss.

As a result of the huge losses, several hedge funds, such as Marin Capital, Aman Capital, and Baily Coates

[9] Shorting the equity tranche means being short credit protection or selling credit protection, which means receiving the (high) equity tranche contract spread.

[10] To hedge means to protect or to reduce risk.

[11] Going long the mezzanine tranche means being long credit protection or buying credit protection, which means paying the (fairly low) mezzanine tranche contract spread.

Cromwell, filed for bankruptcy. It is important to point out that the losses resulted from a lack of understanding of the correlation properties of the tranches in the CDO. The CDOs themselves can hardly be blamed or be called toxic for their correlation properties.

From 2003 to 2006 the CDO market, mainly referencing residential mortgages, had exploded, increasing from $64 billion to $455 billion. To fuel the CDOs, more and more questionable subprime mortgages were given, named NINJA loans, standing for "no income, no job or assets." When housing prices started leveling off in 2006, the first mortgages started to default. In 2007 more and more mortgages defaulted, finally leading to a real estate market collapse. With it the huge CDO market collapsed, leading to the stock market and commodity market crash and a freeze in the credit markets. The financial crisis spread to the world economies, creating a global severe recession, now called the Great Recession.

In a systemic crash like this, naturally many types of correlations increase (see also Figure 6-8). From 2007 to 2009, default correlations of the mortgages in the CDOs increased. This actually helped equity tranche investors, as we can see from Figure 6-7: If default correlations increase, the equity tranche spread decreases, leading to an increase in the value of the equity tranche. However, this increase was overcompensated by a strong increase in default probability of the mortgages. As a consequence, tranche spreads increased sharply, resulting in huge losses for the equity tranche investors as well as investors in the other tranches.

Correlations between the tranches of the CDOs also increased during the crisis. This had a devastating effect on the super-senior tranches. In normal times, these tranches were considered extremely safe since (1) they were AAA rated and (2) they were protected by the lower tranches. But with the increased tranche correlation and the generally deteriorating credit market, these super-senior tranches were suddenly considered risky and lost up to 20% of their value.

To make things worse, many investors had leveraged the super-senior tranches, termed leveraged super-senior (LSS) tranches, to receive a higher spread. This leverage was typically 10 to 20 times, meaning an investor paid $10,000,000 but had risk exposure of $100,000,000 to $200,000,000. What made things technically even worse was that these LSSs came with an option for the investors to unwind the super-senior tranche if the spread had widened (increased). Many investors started to purchase the LSS spread at very high levels, realizing a loss and increasing the LSS tranche spread even further.

In addition to the overinvestment in CDOs, the credit default swap (CDS) market also exploded from its beginnings in the mid-1990s from about $8 trillion in 2004 to almost $60 trillion in 2007. CDSs are typically used as insurance to protect against default of a debtor, as we explained in Figure 6-1. No one will argue that an insurance contract is toxic. On the contrary, it is the principle of an insurance contract to spread the risk to a wider audience and hence reduce individual risk, as we can see from health insurance or life insurance contracts.

CDSs, though, can also be used as speculative instruments. For example, the CDS seller (i.e., the insurance seller) hopes that the insured event (e.g., default of a company or credit deterioration of the company) will not occur. In this case the CDS seller keeps the CDS spread (i.e., the insurance premium) as income, as American International Group (AIG) tried to do in the crisis. A CDS buyer who does not own the underlying asset is speculating on the credit deterioration of the underlying asset, just like a naked put option holder speculates on the decline of the underlying asset.

So who is to blame for the 2007–2009 global financial crisis? The quants, who created the new products such as CDSs and CDOs and the models to value them? The upper management and the traders, who authorized and conducted the overinvesting and extreme risk taking? The rating agencies, who gave an AAA rating to many of the CDOs? The regulators, who approved the overinvestments? The risk managers, who allowed the excessive risk taking?

The entire global financial crisis can be summed up in one word: Greed! It was the upper management, the traders, and the investors who engaged in excessive trading and irresponsible risk taking to receive high returns, huge salaries, and generous bonuses. For example, the London unit of AIG had sold close to $500 billion in CDSs without much reinsurance! Their main hedging strategy seemed to have been: Pray that the insured contracts don't deteriorate. The investment banks of the small Northern European country of Iceland had borrowed 10 times Iceland's national GDP and invested it. With this leverage, Iceland naturally went de facto into bankruptcy in 2008, when the credit markets deteriorated. Lehman Brothers, before filing for bankruptcy in September 2008, reported a leverage of 30.7 (i.e., $691 billion in assets and only $22 billion in stockholders' equity). The true leverage was even higher, since Lehman tried to hide the leverage with materially misleading repo transactions.[12] In addition, Lehman had 1.5 million derivatives transactions with 8,000 different counterparties on its books.

Did the upper management and traders of hedge funds and investment banks admit to their irresponsible leverage, excessive trading, and risk taking? No. Instead they created the myth of the toxic asset, which is absurd. It is like a murderer saying, "I did not shoot that person. It was my gun!" Toxic are not the financial products, but humans and their greed.

Most traders were well aware of the risks that they were taking. In the few cases where traders did not understand the risks, the asset itself cannot be blamed; rather, the incompetence of the trader is the reason for the loss. While it is ethically disappointing that the investors and traders did not admit to their wrongdoing, at the same time it is understandable. If they admitted to irresponsible trading and risk taking, they would immediately be prosecuted.

Naturally, risk managers and regulators have to take part of the blame for allowing the irresponsible risk taking. The moral hazard of the rating agencies, being paid by the same companies whose assets they rate, also needs to be addressed.

We will discuss the role of financial models, their benefits, and their limitations at the beginning of Chapter 8.

Regulation and Correlation

Correlations are critical inputs in regulatory frameworks such as the Basel accords, especially in regulations for market risk and credit risk. Let's clarify.

What Are Basel I, II, and III?

Basel I, implemented in 1988; Basel II, implemented in 2006; and Basel III, which is currently being developed and implemented until 2018, are regulatory guidelines to ensure the stability of the banking system.

The term *Basel* comes from the beautiful city of Basel in Switzerland, where the honorable regulators meet. None of the Basel accords has legal authority. However, most countries (about 100 for Basel II) have created legislation to enforce the Basel accords for their banks.

Why Basel I, II, and III?

The objective of the Basel accords is to "provide incentives for banks to enhance their risk measurement and management systems" and "to contribute to a higher level of safety and soundness in the banking system." In particular, Basel III is being developed to address the deficiencies of the banking system during the financial crisis of 2007 to 2009. Basel III introduces many new ratios to ensure liquidity and adequate leverage of banks. In addition, new correlation models will be implemented that deal with double defaults in insured risk transactions as displayed in Figure 6-1. Correlated defaults in a multi-asset portfolio quantified with the Gaussian copula, correlations in derivatives transactions termed credit value adjustment (CVA), and correlations in what is called wrong-way risk (WWR) are currently being discussed.

HOW DOES CORRELATION RISK FIT INTO THE BROADER PICTURE OF RISKS IN FINANCE?

As already mentioned, we differentiate three main types of risks in finance:

1. Market risk
2. Credit risk
3. Operational risk

Additional types of risk may include systemic risk, concentration risk, liquidity risk, volatility risk, legal risk, reputational risk, and more. Correlation risk plays an important part in market risk and credit risk, and is closely related to systemic risk and concentration risk. Let's discuss it.

Correlation Risk and Market Risk

Correlation risk is an integral part of market risk. Market risk is comprised of equity risk, interest rate risk, currency risk, and commodity risk. Market risk is typically measured

[12] *Repo* stands for repurchase transaction. It can be viewed as a short-term collateralized loan.

with the value-at-risk (VaR) concept. VaR has a covariance matrix of the assets in the portfolio as an input. So market risk implicitly incorporates correlation risk, i.e., the risk that the correlations in the covariance matrix change. We have already studied the impact of different correlations on VaR in the section, "Risk Management and Correlation."

Market risk is also quantified with expected shortfall (ES), also termed conditional VaR or tail risk. Expected shortfall measures market risk for extreme events, typically for the worst 0.1 %, 1%, or 5% of possible future scenarios. A rigorous valuation of expected shortfall naturally includes the correlation between the asset returns in the portfolio, as VaR does.[13]

Correlation Risk and Credit Risk

Correlation risk is also a critical part of credit risk. Credit risk is comprised of (1) migration risk and (2) default risk. Migration risk is the risk that the credit quality of a debtor decreases, i.e., migrates to a lower credit state. A lower credit state typically results in a lower asset price, so a paper loss for the creditor. We already studied the effect of correlation risk of an investor who has hedged his bond exposure with a CDS. We derived that the investor is exposed to the correlation between the reference asset and the counterparty, the CDS seller. The higher the correlation, the higher the CDS paper loss for the investor and, importantly, the higher the probability of a total loss of the investment.

The degree to which defaults occur together (i.e., default correlation) is critical for financial lenders such as commercial banks, credit unions, mortgage lenders, and trusts, which give many types of loans to companies and individuals. Default correlations are also critical for insurance companies, which are exposed to credit risk of numerous debtors. Naturally, a low default correlation of debtors is desired to diversify the credit risk. Table 6-3 shows the default correlation from 1981 to 2001 of 6,907 companies, of which 674 defaulted.

The default correlations in Table 6-3 are one-year default correlations averaged over the time period 1981 to 2001.

[13] Unfortunately, different authors use different definitions (and notation) for ES. To study ES, we recommend the original ES paper by Artzner et. al. (1997), an educational paper by Yamai and Yoshiba (2002), as well as Acerbi and Tasche (2001) and McNeil, Frey, and Embrechts (2005).

We will see how to calculate default correlations in Chapter 9, especially in the section, "The Binomial Correlation Measure" (Lucas 1995).

From Table 6-3, we observe that default correlations between industries are mostly positive with the exception of the energy sector. This sector is typically viewed as a recession-resistant, stable sector with little or no correlation to other sectors. We also observe that the default correlation within sectors is higher than between sectors. This suggests that systematic factors (e.g., a recession or structural weakness such as the general decline of a sector) have a greater impact on defaults than do idiosyncratic factors. Hence if General Motors defaults, it is more likely that Ford will default, rather than Ford benefiting from the default of its rival.

Since the intrasector default correlations are higher than intersector default correlations, a lender is advised to have a sector-diversified loan portfolio to reduce default correlation risk.

Defaults are binomial events, either default or no default. So principally we can use a simple correlation model such as the binomial model of Lucas (1995) to analyze them, which we will do in Chapter 9. However, we can also analyze defaults in more detail and look at term structure of defaults. Let's assume a creditor has given loans to two debtors. One debtor is A rated, and one is CC rated. A historical default term structure of these bonds is displayed in Table 6-4.

For most investment grade bonds, the term structure of default probabilities increases in time, as we see from Table 6-4 for the A-rated bond. This is because the longer the time horizon, the higher the probability of adverse internal events such as mismanagement, or adverse external events such as increased competition or a recession. For bonds in distress, however, the default term structure is typically inverse, as seen for the CC-rated bond in Table 6-4. This is because for a distressed company the immediate future is critical. If the company survives the coming problematic years, the probability of default decreases.

For a creditor, the default correlation of her debtors is critical. As mentioned, a creditor will benefit from a low default correlation of her debtors, which spreads the default correlation risk. We can correlate the default term structures in Table 6-4 with the famous (now infamous) copula model, which will be discussed in Chapter 9.

TABLE 6-3 Default Correlation of 674 Defaulted Companies by Industry

One-Year U.S. Default Correlations—Non-Investment-Grade Bonds 1981–2001

	Auto	Cons	Ener	Fin	Build	Chem	HiTech	Insur	Leis	Tele	Trans	Util
Auto	3.80%	1.30%	1.20%	0.40%	1.10%	1.60%	2.80%	−0.50%	1.00%	3.90%	1.30%	0.50%
Cons	1.30%	2.80%	−1.40%	1.20%	2.80%	1.60%	1.80%	1.10%	1.30%	3.20%	1.30%	1.90%
Ener	1.20%	−1.40%	6.40%	−2.50%	−0.50%	0.40%	−0.10%	−1.60%	−1.00%	−1.40%	−0.10%	0.70%
Fin	0.40%	1.20%	−2.50%	**5.20%**	2.60%	0.10%	2.30%	3.00%	1.60%	3.70%	1.50%	4.50%
Build	1.10%	2.80%	−0.50%	2.60%	**6.10%**	1.20%	2.30%	1.80%	2.30%	**6.50%**	4.20%	1.30%
Chem	1.60%	1.60%	0.40%	0.10%	1.20%	3.20%	1.40%	−1.10%	1.10%	2.80%	1.10%	1.00%
HiTech	2.80%	1.80%	−0.10%	0.40%	2.30%	1.40%	3.30%	0.00%	1.10%	2.80%	1.10%	1.00%
Insur	−0.50%	1.10%	−1.60%	3.00%	1.80%	−1.10%	0.00%	**5.60%**	1.20%	−2.60%	2.30%	1.40%
Leis	1.00%	1.30%	−1.00%	1.60%	2.30%	1.10%	1.40%	1.20%	2.30%	4.00%	2.30%	0.60%
Tele	3.90%	3.20%	−1.40%	3.70%	**6.50%**	2.80%	4.70%	−2.60%	4.00%	**10.70%**	3.20%	−0.80%
Trans	1.30%	2.70%	−0.10%	1.50%	4.20%	1.10%	1.90%	2.30%	2.30%	3.20%	4.30%	−0.20%
Util	.50%	1.90%	0.70%	4.50%	1.30%	1.00%	1.00%	1.40%	0.60%	−0.80%	−0.20%	**9.40%**

Correlations above 5% are bold.

Source: Standard & Poor's (S&P) 500.

TABLE 6-4 Term Structure of Default Probabilities for an A-Rated Bond and a CC-Rated Bond in 2002

	Year									
	1	2	3	4	5	6	7	8	9	10
A	0.02%	0.07%	0.13%	0.14%	0.15%	0.17%	0.18%	0.21%	0.24%	0.25%
CC	23.83%	13.29%	10.31%	7.62%	5.04%	5.13%	4.04%	4.62%	2.62%	2.04%

Source: Moody's.

This will allow us to answer such questions as: "What is the joint probability of debtor 1 defaulting in year 3 and debtor 2 defaulting in year 5?"

Correlations always increase in stressed markets.

—John Hull

Correlation Risk and Systemic Risk

So far, we have analyzed correlation risk with respect to market risk and credit risk and have concluded that correlations are a critical input when quantifying market risk and credit risk. Correlations are also closely related to systemic risk, which we define here.

Systemic Risk The risk of a financial market or an entire financial system collapsing.

An example of systemic risk is the collapse of the entire credit market in 2008. At the height of the crisis in September 2008, when Lehman Brothers filed for bankruptcy, the credit markets were virtually frozen with essentially no lending activities. Even as the Federal Reserve guaranteed interbank loans, lending resumed only very gradually and slowly.

The stock market crash starting in October 2007 with the Dow Jones Industrial Average at 14,093 points and then falling by 53.54% to 6,547 points by March 2009 is also a systemic market collapse. All but one of the Dow 30 stocks had declined. Walmart was the lone Dow stock that was up during the crisis. Of the S&P 500 stocks, 489 declined during this time frame. The 11 stocks that were up were:

1. Apollo Group (APOL), educational sector; provides educational programs for working adults and is a subsidiary of the University of Phoenix.
2. AutoZone (AZO), auto industry; provides auto replacement parts.
3. CF Industries (CF), agricultural industry; provides fertilizer.
4. DeVry Inc. (DV), educational sector; holding company of several universities.
5. Edward Lifesciences (EW), pharmaceutical industry; provides products to treat cardiovascular diseases.
6. Family Dollar (FDO), consumer staples.
7. Gilead Pharmaceuticals (GILD), pharmaceutical industry; provides HIV, hepatitis medications.
8. Netflix (NFLX), entertainment industry; provides Internet subscription service.
9. Ross Stores (ROST), consumer staples.
10. Southwestern Energy (SWN), energy sector.
11. Walmart (WMT), consumer staples.

From this list we can see that the consumer staples sector (which provides such basic necessities as food and household items) fared well during the crisis. The educational sector also typically thrives in a crisis, since many unemployed seek to further their education.

Importantly, systemic financial failures such as the one from 2007 to 2009 typically spread to the economy, with a decreasing GDP, increasing unemployment, and therefore a decrease in the standard of living.

Systemic risk and correlation risk are highly dependent. Since a systemic decline in stocks involves almost the entire stock market, correlations between the stocks increase sharply. Figure 6-8 shows the relationship between the percentage change of the Dow Jones Industrial Average, short "Dow," and the correlation between the stocks in the Dow before the crisis from May 2004 to October 2007 and during the crisis from October 2007 to March 2009.

In Figure 6-8 we downloaded daily closing prices of all 30 stocks in the Dow and put them into monthly bins. We then derived monthly 30 x 30 correlation matrices using

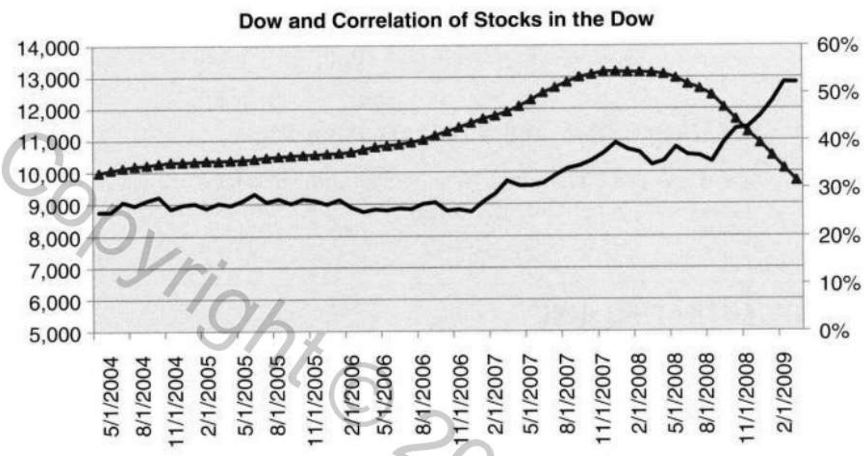


FIGURE 6-8 Relationship between the Dow (graph with triangles, numerical values on left axis) and correlation between the stocks in the Dow (numerical values on right axis).

the Pearson correlation measure and averaged the matrices. We then smoothed the graph by taking the one-year moving average.

From Figure 6-8 we can observe a somewhat stable correlation from 2004 to 2006, when the Dow increased moderately. In the time period from January 2007 to February 2008 we observe that the correlation in the Dow increases when the Dow increases more strongly. Importantly, in the time of the severe decline of the Dow from August 2008 to March 2009, we observe a sharp increase in the correlation from noncrisis levels of on average 27% to over 50%. In Chapter 7, we will observe empirical correlations in detail, and we will find that at the height of the crisis in February 2009 the correlation of the stocks in the Dow reached a high of 96.97%. Hence, portfolios that were considered well diversified in benign times experienced a sharp increase in correlation and hence unexpected losses due to the combined, highly correlated decline of many stocks during the crisis.

Correlation Risk and Concentration Risk

Concentration risk is a fairly new risk category and therefore not yet uniquely defined. We provide a sensible definition.

> **Concentration Risk** The risk of financial loss due to a concentrated exposure to a particular group of counterparties.

Concentration risk can be quantified with the concentration ratio. For example, if a creditor has 10 loans of equal size, the concentration ratio would be 1/10 = 0.1. If a creditor has only one loan to one counterparty, the concentration ratio would be 1. Naturally, the lower the concentration ratio, the more diversified is the default risk of the creditor, assuming the default correlation between the counterparties is smaller than 1.

We can also categorize counterparties into groups, for example sectors. We can then analyze sector concentration risk. The higher the number of different sectors a creditor has lent to, the higher is the sector diversification. High sector diversification reduces default risk, since intrasector defaults are more highly correlated than counterparties in different sectors, as seen in Table 6-3.

Naturally, concentration risk and correlation risk are closely related. Let's verify this in an example.

Example 6.3 Concentration Ratio and Correlation

CASE A

The commercial bank *C* has lent $10,000,000 to a single company, *W*. So *C*'s concentration ratio is 1. Let's assume company W has a default probability (P_W) of 10%. Hence the expected loss (EL) for bank *C* is $10,000,000 × 0.1 = $1,000,000 (see Figure 6-9).

CASE B

The commercial bank *C* has lent $5,000,000 to company *X* and $5,000,000 to company *Y*. Let's assume both *X* and *Y* have a 10% default probability. So *C*'s concentration ratio is reduced to ½.

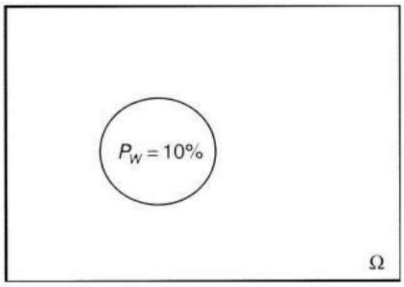


FIGURE 6-9 Probability space for the default probability of a single loan to *W*.

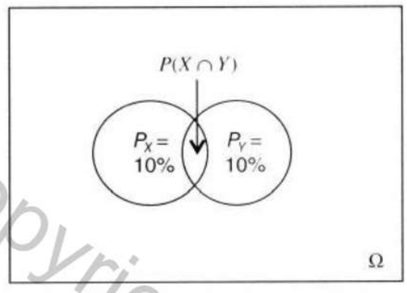


FIGURE 6-10 Probability space for loans to companies *X* and *Y*.

If the default correlation between *X* and *Y* is bigger than 0 and smaller than 1, we derive that the worst-case scenario [i.e., the default of *X* and *Y*, $P(X \cap Y)$, with a loss of $1,000,000] is reduced, as seen in Figure 6-10.

The exact joint default probability $P(X \cap Y)$ depends on the correlation model and correlation parameter values, which will be discussed in Chapters 8 and 9. For any model, though, if default correlation between *X* and *Y* is 1, then there is no benefit from the lower concentration ratio. The probability space would have the form as in Figure 6-9.

CASE C

If we further decrease the concentration ratio, the worst-case scenario (i.e., the expected loss of 10%) decreases further. Let's assume the lender *C* gives loans to three companies, *X*, *Y*, and *Z*, of $3.33 million each. Let's assume that the default probabilities of *X*, *Y*, and *Z* are 10% each. Therefore the concentration ratio decreases to ⅓. The probabilities are displayed in Figure 6-11.

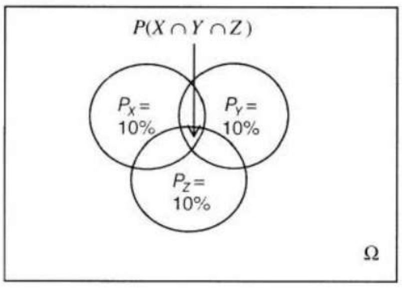


FIGURE 6-11 Probability space for loans to companies *X*, *Y*, and *Z*.

Hence from Figures 6-9 to 6-11 we observe the benefits of a lower concentration ratio. The worst-case scenario, an expected loss of $1,000,000, reduces with a decreasing concentration ratio.

A decreasing concentration ratio is closely related to a decreasing correlation coefficient. Let's show this. The defaults of companies *X* and *Y* are expressed as two binomial variables, which take the value 1 if in default, and 0 otherwise. Equation (6.11) gives the joint probability of default for the two binomial events:

$$P(X \cap Y) = \rho_{XY}\sqrt{P_X(1-P_X)P_Y(1-P_Y)} + P_X P_Y \quad \textbf{(6.11)}$$

where ρ_{XY} is the correlation coefficient and

$$\sqrt{P_X(1-P_X)} \quad \textbf{(6.12)}$$

is the standard deviation of the binomially distributed variable *X*.

Let's assume again that the lender C has given loans to *X* and *Y* of $5,000,000 each. Both *X* and *Y* have a default probability of 10%. Following Equation (6.12), this means that the standard deviation for *X* and *Y* is $\sqrt{0.1 \times (1-0.1)} = 0.3$.

Let's first look at the case where the default correlation is $\rho_{XY} = 1$. This means that *X* and *Y* cannot default individually. They can only default together or survive together. The probability that they default together is 10%. Hence the expected loss is the same as in case A: EL = ($5,000,000 + $5,000,000) × 0.1 = $1,000,000. We can verify this with Equation (6.11) for the joint probability of two binomial events, $P(X \cap Y) = 1 \times \sqrt{0.1(1-0.1) \times 0.1(1-0.1)} + 0.1 \times 0.1 = 10\%$. The probability space is graphically the same as Figure 6-9 with $P_X = P_Y =$ 10% as the probability event.

If we now decrease the correlation coefficient, we can see from Equation (6.11) that the worst-case scenario, the joint default probability of *X* and *Y*, $P(X \cap Y)$, will decrease. For example, $\rho_{XY} = 0.5$ results in $P(X \cap Y) = 5.5\%$, and $\rho_{XY} =$ 0 results in $P(X \cap Y) = 1\%$. Interestingly, even a slightly negative correlation coefficient can result in a positive joint default probability if the standard deviation of the binomial events is fairly low and the default probabilities are high. In our example, the standard deviation of both entities is 30% and a default probability of both entities is 10%. Together with a negative correlation coefficient of −0.1, following Equation (6.11), this leads to a joint default probability of 0.1 %. We will discuss the binomial correlation model in more detail in Chapter 9.

In conclusion, we have shown the beneficial aspect of a lower concentration ratio, which is closely related to a lower correlation coefficient. In particular, both a lower concentration ratio and a lower correlation coefficient reduce the worst-case scenario for a creditor, the joint probability of default of his debtors.

A higher (copula) correlation between assets results in a higher credit value-at-risk (CVaR). CVaR measures the maximum loss of a portfolio of correlated debt with a certain probability for a certain time frame. Hence CVaR measures correlated default risk and is analogous to the VaR concept for correlated market risk, which we discussed earlier.

A WORD ON TERMINOLOGY

As mentioned in the section, "Trading and Correlation," we find the terms *correlation desks* or *correlation trading* in trading practice. Correlation trading means that traders trade assets or execute trading strategies whose value is at least in part determined by the comovement of two or more assets in time. We already mentioned pairs trading, the exchange option, and the quanto option as examples of correlation trading. In trading practice, the term *correlation* is typically applied quite broadly, referring to any comovement of asset prices in time.

However, in financial theory, especially in recent publications, the term *correlation* is often defined more narrowly, referring only to the linear Pearson correlation model, as in Cherubini, Luciano, and Vecchiato (2004), Nelsen (2006), or Gregory (2010). These authors refer to other than Pearson correlation coefficients as dependence measures or measures of association. However, in financial theory the term *correlation* is also often applied generally to describe dependencies, as in the terms *credit correlation*, *default correlation*, or *volatility-asset return correlation*, which are quantified by non-Pearson models as in Heston (1993), Lucas (1995), or Li (2000).

In this book, we will refer to the Pearson coefficient, discussed in Chapter 8, as correlation coefficient and the coefficients derived by non-Pearson models as dependency coefficients. In accordance with most literature, we will refer to all methodologies that measure some form of dependency as correlation models or dependency models. Ordinal dependence measures, discussed in Chapter 8, which are related to the Pearson correlation approach, will be termed rank correlation measures.

SUMMARY

There are two types of financial correlations: (1) *Static* correlations measure how two or more financial assets are associated within a certain time period, for example a year. (2) *Dynamic* financial correlations measure how two or more financial assets move together in time.

Correlation risk can be defined as the risk of financial loss due to adverse movements in correlation between two or more variables. These variables can be financial variables such as correlated defaults of two debtors or non-financial such as the correlation between political tensions and an exchange rate. Correlation risk can be nonmonotonous, meaning that the dependent variable, for example the CDS spread, can sometimes increase and sometimes decrease when the correlation parameter value increases.

Correlations and correlation risk are critical in many areas in finance, such as investments, trading, and especially in risk management, where different correlations result in very different degrees of risk. Correlations also play a key role in a systemic crisis, where correlations typically increase and can lead to high unexpected losses. As a result, the Basel III accord has introduced several correlation concepts and measures to reduce correlation risk.

Correlation risk can be categorized as its own type of risk. However, correlation parameters and correlation matrices are critical inputs and hence a part of market risk and credit risk. Market risk and credit risk are highly sensitive to changing correlations. Correlation risk is also closely related to concentration risk, as well as systemic risk, since correlations typically increase in a systemic crisis.

The term *correlation* is not uniquely defined. In trading practice, *correlation* is applied quite broadly and refers to the comovements of assets in time, which may be measured by different correlation concepts. In financial theory, the term *correlation* is often defined more narrowly, referring only to the linear Pearson correlation coefficient. Non-Pearson correlation measures are termed *dependence measures* or *measures of association.*

APPENDIX A

Dependence and Correlation

Dependence

In statistics, two events are considered dependent if the occurrence of one affects the probability of another. Conversely, two events are considered independent if the occurrence of one does not affect the probability of another. Formally, two events, A and B, are independent if and only if the joint probability equals the product of the individual probabilities:

$$P(A \cap B) = P(A)P(B) \tag{6.13}$$

Solving Equation (6.13) for $P(A)$, we get

$$P(A) = \frac{P(A \cap B)}{P(B)}$$

Following the Kolmogorov definition $\frac{P(A \cap B)}{P(B)} \equiv P(A \mid B)$, we derive

$$P(A) = \frac{P(A \cap B)}{P(B)} = P(A \mid B) \tag{6.14}$$

where $P(A \mid B)$ is the conditional probability of A with respect to B. $P(A \mid B)$ reads "probability of A given B." In Equation (6.14), the probability of A, $P(A)$, is not affected by B, since $P(A) = P(A \mid B)$; hence the event A is independent from B.

From Equation (6.14) we also derive

$$P(B) = \frac{P(A \cap B)}{P(A)} = P(B \mid A) \tag{6.15}$$

Hence from Equation (6.13) it follows that A is independent from B and B is independent from A.

Example of Statistical Independence

The historical default probability of company A is $P(A) = 3\%$, the historical default probability of company B is $P(B) = 4\ \%$, and the historical joint probability of default is $3\% \times 4\% = 0.12\%$. In this case $P(A)$ and $P(B)$ are independent. This is because from Equation (6.14), we have

$$P(A) = \frac{P(A \cap B)}{P(B)} = P(A \mid B) = 3\% = \frac{3\% \times 4\%}{4\%} = 3\%$$

Since $P(A) = P(A \mid B)$, the default probability of company A is independent from the default probability of company B. Using Equation (6.15), we can do the same exercise for the default probability of company B, which is independent from the default probability of company A.

Correlation

As mentioned earlier, the term *correlation* is not uniquely defined. In trading practice, the term *correlation* is used quite broadly, referring to any comovement of asset prices in time. In statistics, correlation is typically defined more narrowly and typically refers to the linear dependency derived in the Pearson correlation model. Let's look at the Pearson covariance and relate it to the dependence discussed earlier.

A covariance measures how strong the linear relationship between two variables is. These variables can be deterministic (meaning their outcome is known), as the historical default probabilities in Equation (6.13). For random variables (variables with an unknown outcome such as flipping a coin), the Pearson covariance is derived with expectation values:

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y) \tag{6.16}$$

where $E(X)$ and $E(Y)$ are the expected values of (X) and (Y) respectively, also known as the mean, and $E(XY)$ is the expected value of the product of the random variables X and Y.

The covariance in Equation (6.16) is not easy to interpret. Therefore, often a normalized covariance, the correlation coefficient, is used. The Pearson correlation coefficient $\rho(XY)$ is defined as

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \tag{6.17}$$

where $\sigma(X)$ and $\sigma(Y)$ are the standard deviations of X and Y, respectively. While the covariance takes values between $-\infty$ and $+\infty$, the correlation coefficient conveniently takes values between -1 and $+1$.

Independence and Uncorrelatedness

From Equation (6.13) above we find that the condition for independence of two random variables is $E(XY) = E(X)E(Y)$. From Equation (6.16) we see that $E(XY) = E(X)E(Y)$ is equal to a covariance of zero. Therefore, if two variables are independent, their covariance is zero.

Is the reverse also true? Does a zero covariance mean independence? The answer is no. Two variables can have a zero covariance even when they are dependent! Let's show this with an example. For the parabola $Y = X^2$, Y is clearly dependent on X, since Y changes when X changes. However, the correlation of the function $Y = X^2$ derived by

Equations (6.16) or (6.17) is zero! This can be shown numerically and algebraically.

Algebraically, we have from Equation (6.16):

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

Inputting $Y = X^2$, we derive

$$Cov(X, Y) = E(X\,X^2) - E(X)E(X^2)$$
$$= E(X^3) - E(X)E(X^2)$$

Let X be a uniform variable bounded in $[-1, +1]$. Then the mean $E(X)$ is zero and we have

$$Cov(X, Y) = 0 - 0E(X^2)$$
$$= 0$$

In conclusion, the Pearson covariance or correlation coefficient can give values of zero; that is, it tells us that the variables are uncorrelated, even if the variables are dependent! This is because the Pearson correlation concept measures only linear dependence. It fails to capture nonlinear relationships. This shows the limitation of the Pearson correlation concept for finance, since most financial relationships are nonlinear. See Chapter 8 for a more detailed discussion on the Pearson correlation model.

APPENDIX B

On Percentage and Logarithmic Changes

In finance, growth rates are expressed as relative changes, $(S_t - S_{t-1})/S_{t-1}$, where S_t and S_{t-1} are the prices of an asset at time t and $t - 1$, respectively. For example, if $S_t = 110$, and $S_{t-1} = 100$, the relative change is $(110 - 100)/100 = 0.1 = 10\%$.

We often approximate relative changes with the help of the natural logarithm:

$$(S_t - S_{t-1})/S_{t-1} \approx \ln(S_t/S_{t-1}) \qquad \textbf{(6.18)}$$

This is a good approximation for small differences between S_t and S_{t-1}. $\ln(S_t/S_{t-1})$ is called a log return. The advantage of using log returns is that they can be added over time. Relative changes are not additive over time. Let's show this in two examples.

Example 1: A stock price at t_0 is \$100. From t_0 to t_1, the stock increases by 10%. Hence the stock increases to \$110. From t_1 to t_2 the stock increases again by 10%. So the stock price increases to \$110 × 0.1 = \$121. This increase of 21% is higher than adding the percentage increases of 10% + 10% = 20%. Hence percentage changes are not additive over time.

Let's look at the log returns. The log return from t_0 to t_1 is ln(110/100) = 9.531%. From t_1 to t_2 the log return is ln(121/110) = 9.531%. When adding these returns, we get 9.531% + 9.531% = 19.062%. This is the same as the log return from t_0 to t_2; that is, ln(121/100) = 19.062%. Hence log returns are additive in time.[14]

Let's now look at another, more extreme example.

Example 2: A stock price in t_0 is \$100. It moves to \$200 in t_1 and back to \$100 in t_2. The percentage change from t_0 to t_1 is (\$200 − \$100)/\$100 = 100%. The percentage change from t_1 to t_2 is (\$100 − \$200)/(200) = −50%. Adding the percentage changes, we derive +100% − 50% = +50%, although the stock has not increased from t_0 to t_2! Naturally, this type of performance measure is incorrect and not allowed in accounting.

Log returns give the correct answer: The log return from t_0 to t_1 is ln(200/100) = 69.31%. The log return from t_1 to t_2 is ln(100/200) = −69.31%. Adding these log returns in time, we get the correct return of the stock price from t_0 to t_2 of 69.31% − 69.31% = 0%.

References and Suggested Readings

Acerbi, D., and D. Tasche. 2001. "Expected Shortfall: A Natural Coherent Alternative to Value at Risk." www.bis.org/bcbs/ca/acertasc.pdf.

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath. 1997. "Thinking Coherently." *Risk* 10(11):68–71.

Brownless, C., and R. Engle. 2012. "Volatility, Correlation and Tails in Systemic Risk Measurement." Working paper.

Cherubini, U., E. Luciano, and W. Vecchiato. 2004. *Copula Methods in Finance.* Hoboken, NJ: John Wiley & Sons.

Economist, August 2008. "A Personal View of the Crisis, Confessions of a Risk Manager."

Gregory, J. 2010. *Counterparty Credit Risk.* Hoboken, NJ: John Wiley & Sons.

[14] We could also have solved for the absolute value 121, which matches a logarithmic growth rate of 9.531%: $\ln(x/110) = 9.531\%$, or $\ln(x) - \ln(110) = 9.531\%$, or $\ln(x) = \ln(110) + 9.531\%$. Taking the power of e, we get $e^{(\ln(x))} = x = e^{(\ln(110)+0.09531)} = 121$.

Heston, S. (1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *The Review of Financial Studies* 6:327–343.

Hull, J. 2012. *Risk Management and Financial Institutions.* 3rd ed. Hoboken, NJ: John Wiley & Sons.

Jones, S. 2009. "The Formula That Felled Wall St." *Financial Times,* April 24, 2009.

Jorion, P. 2006. *Value at Risk: The New Benchmark for Managing Financial Risk.* 3rd ed. Hoboken, NJ: John Wiley & Sons.

Kettunen, J., and G. Meissner. 2006. "Valuing Default Swaps on Correlated LMM Processes." *Journal of Alternative Investments,* Summer.

Li, D. 2000. "On Default Correlation: A Copula Approach," *Journal of Fixed Income* 9:119–149.

Lucas, D. 1995. "Default Correlation and Credit Analysis." *Journal of Fixed Income* 4:76–87.

Markovich, N. 2007. *Nonparametric Analysis of Univariate Heavy-Tailed Data.* Hoboken, NJ: John Wiley & Sons.

Markowitz, H. M. 1952. "Portfolio Selection." *Journal of Finance* 7:77–91.

McNeil, A., R. Frey, and P. Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton, NJ: Princeton University Press.

Meissner, G. 2005. *"Credit Derivatives—Application, Pricing, and Risk Management,"* Wiley-Blackwell.

Nelsen, R. 2006. *An Introduction to Copulas.* 2nd ed. Springer, New York.

Sharpe, W. F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance* 19:425–442.

Sklar, A. 1959. "Fonctions de Répartition à n Dimensions et Leurs Marges," *Publications de l'Institut de Statistique de L'Université de Paris* 8:229–231.

Yamai, Y., and T. Yoshiba. 2002. "On the Validity of Value-at-Risk: Comparative Analyses with Expected Shortfall." *Monetary and Economic Studies,* January: 57–86.

Empirical Properties of Correlation: How Do Correlations Behave in the Real World?

7

Learning Objectives

After completing this reading you should be able to:

- Describe how equity correlations and correlation volatilities behave throughout various economic states.
- Calculate a mean reversion rate using standard regression and calculate the corresponding autocorrelation.
- Identify the best-fit distribution for equity, bond, and default correlations.

Excerpt is Chapter 2 of Correlation Risk Modeling and Management *by Gunter Meissner.*

Anything that relies on correlation is charlatanism.

—Nassim Taleb

In this chapter we show that, contrary to common beliefs, financial correlations display statistically significant and expected properties. We show that correlation levels as well as correlation volatility are generally higher in economic crises, which should be taken into consideration by traders and risk managers. We also find strong mean reversion in correlations as well as expected behavior of autocorrelation. The distribution of correlations is typically not normal or lognormal.

HOW DO EQUITY CORRELATIONS BEHAVE IN A RECESSION, NORMAL ECONOMIC PERIOD, OR STRONG EXPANSION?

In our study, we observed daily closing prices of the 30 stocks in the Dow Jones Industrial Average (Dow) from January 1972 to October 2012. This resulted in 10,303 daily observations of the Dow stocks and hence 10,303 × 30 = 309,090 closing prices. We built monthly bins and derived 900 correlation values (30 × 30) for each month, applying the Pearson correlation approach. Since we had 490 months in the study, all together we derived 490 × 900 = 441,000 correlation values. We eliminated the unity correlation values on the diagonal of each correlation matrix and derived 441,000 − (30 × 490) = 426,300 correlation values as inputs.

The composition of the Dow is changing in time, with successful stocks being put into the Dow and unsuccessful stocks being removed. Our study is comprised of the Dow stocks that represent the Dow at each particular point in time.

Figure 7-1 shows the 490 monthly averaged correlation levels from 1972 to 2012 with respect to the state of the economy. We differentiate three states: an *expansionary period* with gross domestic product (GDP) growth rates of 3.5% or higher, a *normal economic period* with growth rates between 0% and 3.49%, and a *recession* with two consecutive quarters of negative growth rates.

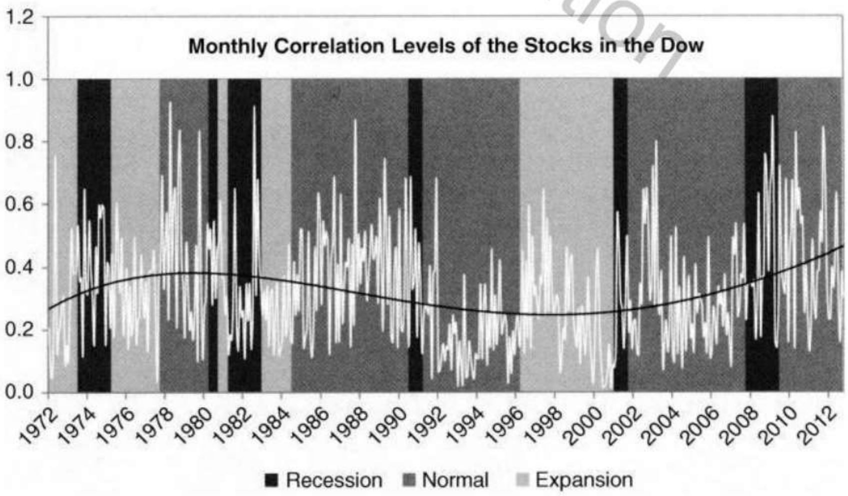


FIGURE 7-1 Average correlation of monthly 30 × 30 Dow stock bins. The light gray background represents an expansionary economic period, the medium gray background a normal economic period, and the dark gray background a recession. The horizontal line shows the polynomial trend line of order 4.

Figure 7-2 shows the volatility of the averaged monthly correlations.

From Figures 7-1 and Figures 7-2 we observe the somewhat erratic behavior of Dow correlation levels and volatility. However, Table 7-1 reveals some expected results.

From Table 7-1 we observe that correlation levels are lowest in strong economic growth times. The reason may be that in strong growth periods equity prices react primarily to idiosyncratic, not macroeconomic factors. In recessions, correlation levels typically increase, as shown in Table 7-1. In addition, Figure 7-8, that correlation levels increased sharply in the Great Recession from 2007 to 2009. In a recession, macroeconomic factors seem to dominate idiosyncratic factors, leading to a downturn of multiple stocks.

A further expected result in Table 7-1 is that correlation volatility is lowest in an economic expansion and highest in worse economic states. We did expect a higher correlation volatility in a recession compared to a normal economic state. However, it seems that high correlation levels in a recession remain high without much additional volatility. Generally, correlation volatility is high, as we can also observe from the somewhat erratic correlation function in Figure 7-1. We will analyze whether the correlation

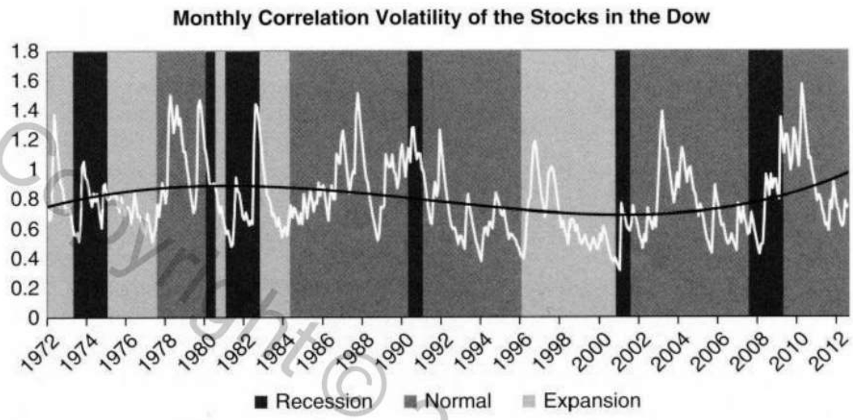


FIGURE 7-2 Correlation volatility of the average correlation of monthly 30 × 30 Dow stock bins with respect to the state of the economy. The horizontal line shows the polynomial trend line of order 4.

TABLE 7-1 Correlation Level and Correlation Volatility with Respect to the State of the Economy

	Correlation Level	Correlation Volatility
Expansionary period	27.46%	71.17%
Normal economic period	32.73%	83.40%
Recession	36.96%	80.48%

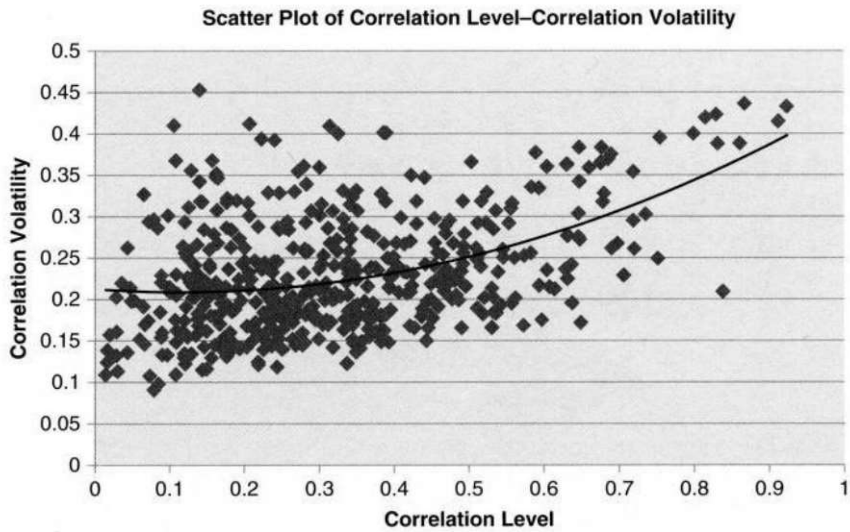


FIGURE 7-3 Positive relationship between correlation level and correlation volatility with a polynomial trend line of order 2.

volatility is an indicator for future recessions later in the chapter. Altogether, Table 7-1 displays the higher correlation risk in bad economic times, which traders and risk managers should consider in their trading and risk management.

From Table 7-1 we observe a generally positive relationship between correlation level and correlation volatility. This is verified in more detail in Figure 7-3.

DO EQUITY CORRELATIONS EXHIBIT MEAN REVERSION?

Mean reversion is the tendency of a variable to be pulled back to its long-term mean. In finance, many variables, such as bonds, interest rates, volatilities, credit spreads, and more, are assumed to exhibit mean reversion. Fixed coupon bonds, which do not default, exhibit strong mean reversion: A bond is typically issued at par, for example at $100. If the bond does not default, at maturity it will revert to exactly that price of $100, which is typically close to its long-term mean.

Interest rates are also assumed to be mean reverting: In an economic expansion, typically demand for capital is high and interest rates rise. These high interest rates will eventually lead to a cooling off of the economy, possibly leading to a recession. In this process, capital demand decreases and interest rates decrease from their high levels towards their long-term mean, eventually falling below their long-term mean. Being in a recession, eventually economic activity increases again, often supported by monetary and fiscal policy. In this reviving economy, demand for capital increases, in turn increasing interest rates to their long-term means.

How Can We Quantify Mean Reversion?

Mean reversion is present if there is a negative relationship between the change of a variable, $S_t - S_{t-1}$, and the variable at $t - 1$, S_{t-1}. Formally, mean reversion exists if

$$\frac{\partial(S_t - S_{t-1})}{\partial S_{t-1}} < 0 \qquad \textbf{(7.1)}$$

where

S_t: price at time t

S_{t-1}: price at the previous point in time $t-1$

∂: partial derivative coefficient

Equation (7.1) tells us: If S_{t-1} increases by a very small amount, $S_t - S_{t-1}$ will decrease by a certain amount, and vice versa. This is intuitive: If S_{t-1} has decreased and is low at $t-1$ (compared to the mean of S, μ_s), then at the next point in time t, mean reversion will pull up S_{t-1} to μ_s and therefore increase $S_t - S_{t-1}$. If S_{t-1} has increased and is high in $t-1$ (compared to the mean of S, μ_s), then at the next point in time t, mean reversion will pull down S_{t-1} to μ_s and therefore decrease $S_t - S_{t-1}$. The degree of the pull is the degree of the mean reversion, also called mean reversion rate, mean reversion speed, or gravity.

Let's quantify the degree of mean reversion. Let's start with the discrete Vasicek 1977 process, which goes back to Ornstein-Uhlenbeck 1930:

$$S_t - S_{t-1} = a(\mu_S - S_{t-1})\Delta t + \sigma_S \varepsilon \sqrt{\Delta t} \qquad \textbf{(7.2)}$$

where

S_t: price at time t

S_{t-1}: price at the previous point in time $t-1$

a: degree of mean reversion, also called mean reversion rate or gravity, $0 \le a \le 1$

μ_s: long-term mean of S

σ_s: volatility of S

ε: random drawing from a standardized normal distribution at time t, $\varepsilon(t)$: $n \sim (0,1)$

We can compute ε as =normsinv(rand()) in Excel/VBA and norminv (rand) in MATLAB.

We are currently interested only in mean reversion, so for now we will ignore the stochasticity part in Equation (7.2), $\sigma_S \varepsilon \sqrt{\Delta t}$.

For ease of explanation, let's assume $\Delta t = 1$. Then, from Equation (7.2) we see that a mean reversion parameter of $a = 1$ will pull S_{t-1} to the long-term mean μ_s completely at every time step. For example, if S_{t-1} is 80 and μ_s is 100, then $a(\mu_s - S_{t-1}) = 1 \times (100 - 80) = 20$, so the S_{t-1} of 80 is mean reverted up to its long-term mean of 100. Naturally, a mean reversion parameter a of 0.5 will lead to a mean reversion of 50% at each time step, and a mean reversion parameter a of 0 will result in no mean reversion.

Let's now quantify mean reversion. Setting Δt to 1, Equation (7.2) without stochasticity reduces to

$$S_t - S_{t-1} = a(\mu_S - S_{t-1}) \qquad \textbf{(7.3)}$$

or

$$S_t - S_{t-1} = a\,\mu_S - a\,S_{t-1} \qquad \textbf{(7.4)}$$

To find the mean reversion rate a, we can run a standard regression analysis of the form

$$Y = \alpha + \beta X$$

Following Equation (7.4), we are regressing $S_t - S_{t-1}$ with respect to S_{t-1}:

$$\underbrace{S_t - S_{t-1}}_{Y} = \underbrace{a\,\mu_S}_{\alpha} - \underbrace{a\,S_{t-1}}_{\beta X} \qquad \textbf{(7.5)}$$

Importantly, from Equation (7.5), we observe that the regression coefficient β is equal to the negative mean reversion parameter a.

We now run a regression of Equation (7.5) to find the empirical mean reversion of our correlation data. Hence S represents the 30 × 30 Dow stock monthly average correlations from 1972 to 2012. The regression analysis is displayed in Figure 7-4.

The regression function in Figure 7-4 displays a strong mean reversion of 77.51%. This means that on average in every month, a deviation from the long-term correlation mean (34.83% in our study) is pulled back to that long-term mean by 77.51%. We can observe this strong mean reversion also by looking at Figure 7-1. An upward spike in correlation is typically followed by a sharp decline in the next time period, and vice versa.

Let's look at an example of modeling correlation with mean reversion.

Example 7.1 Expected Correlation

The long-term mean of the correlation data is 34.83%. In February 2012, the averaged correlation of the 30 × 30 Dow correlation matrices was 26.15%. From the regression function from 1972 to 2012, we find that the average mean reversion is 77.51%. What is the expected correlation for March 2012 following Equation (7.3) or (7.4)?

Solving Equation (7.3) for S_t, we have $S_t = a(\mu_s - S_{t-1}) + S_{t-1}$. Hence the expected correlation in March is

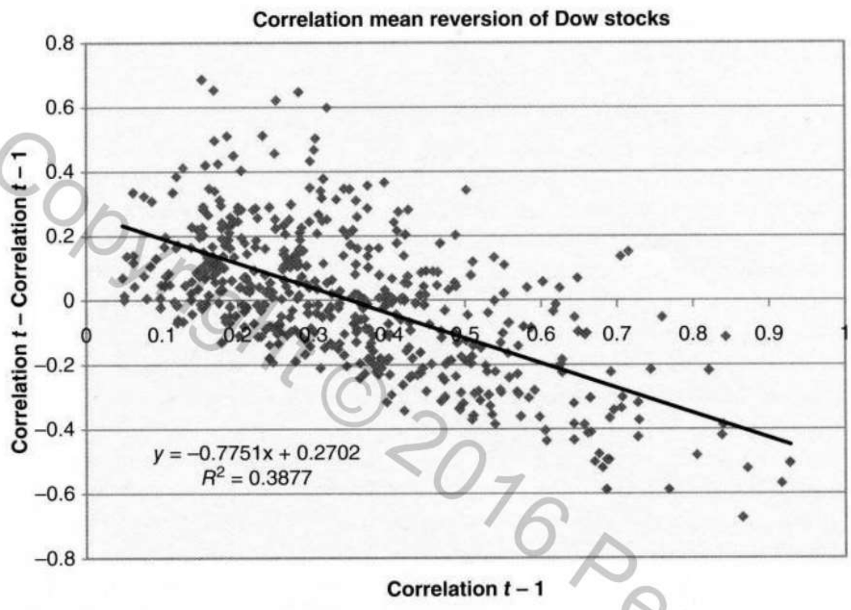


FIGURE 7-4 Regression function (8.5) for 490 monthly average Dow correlations from 1972 to 2012.

$$S_t = 0.7751 \times (0.3483 - 0.2615) + 0.2615 = 0.3288$$

As a result, we find that the mean reversion rate of 77.51% increases the correlation in February 2012 of 26.15% to an expected correlation in March 2012 of 32.88%.[1]

DO EQUITY CORRELATIONS EXHIBIT AUTOCORRELATION?

Autocorrelation is the degree to which a variable is correlated to its past values. Autocorrelation can be quantified with the Nobel Prize–winning autoregressive conditional heteroscedasticity (ARCH) model of Robert Engle (1982) or its extension, generalized autoregressive conditional heteroscedasticity (GARCH) by Tim Bollerslev (1986). However, we can also regress the time series of a variable to its past time series values to derive autocorrelation. This is the approach we will take here.

In finance, positive autocorrelation is also termed *persistence*. In mutual fund or hedge fund performance analysis, an investor typically wants to know if an above-market performance of a fund has persisted for some time (i.e., is positively correlated to its past strong performance).

[1] Note that we have omitted any stochasticity, which is typically included when modeling financial variables, as shown in Equation (7.2).

The question whether autocorrelation exists is an easy one. Autocorrelation is the "reverse property" to mean reversion: The stronger the mean reversion (i.e., the more strongly a variable is pulled back to its mean), the lower the autocorrelation (i.e., the less it is correlated to its past values), and vice versa.

For our empirical correlation analysis, we derive the autocorrelation (AC) for a time lag of one period with Equation (7.6):

$$AC(\rho_t, \rho_{t-1}) = \frac{Cov(\rho_t, \rho_{t-1})}{\sigma(\rho_t)\sigma(\rho_{t-1})} \quad \textbf{(7.6)}$$

where

AC: autocorrelation

ρ_t: correlation values for time period t (in our study the monthly average of the 30 × 30 Dow stock correlation matrices from 2/1/1972 to 12/13/2012, after eliminating the unity correlation on the diagonal)

ρ_{t-1}: correlation values for time period $t - 1$ (i.e., the monthly correlation values starting and ending one month prior than period t)

Cov: covariance; see Equation (7.3) for details.

Equation (7.6) is algebraically identical with the Pearson correlation coefficient Equation (6.4) in Chapter 6. The autocorrelation just uses the correlation values of time period t and time period $t - 1$ as inputs.

Following Equation (7.6), we find the one-period lag autocorrelation of the correlation values from 1972 to 2012 to be 22.49%. As mentioned earlier, autocorrelation is the opposite property of mean reversion. Therefore, not surprisingly, the autocorrelation of 22.49% and the mean reversion in our study of 77.51% (see previous section) add up to 1.

Figure 7-5 shows the autocorrelation with respect to different time lags.

From Figure 7-5 we observe that time lag 2 autocorrelation is highest, so autocorrelation with respect to two months prior produces the highest autocorrelation. Altogether we observe the expected decay in autocorrelation with respect to time lags of earlier periods.

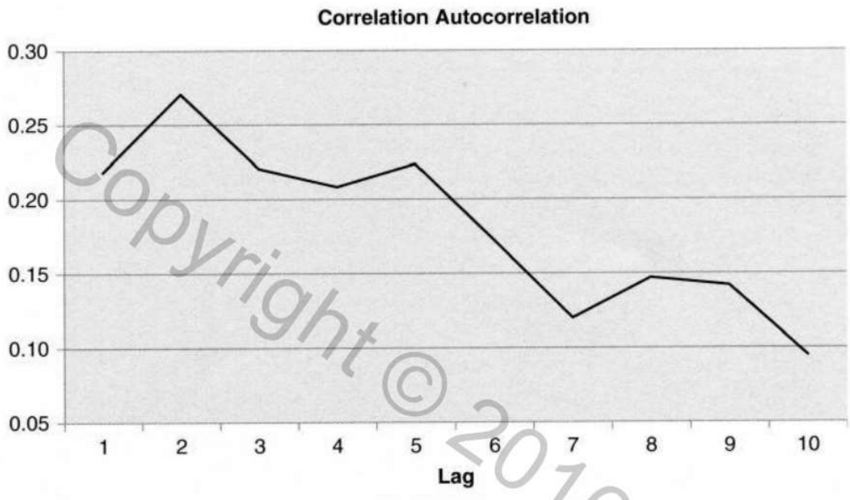


FIGURE 7-5 Autocorrelation of monthly average 30 × 30 Dow stock correlations from 1972 to 2012. The time period of the lags is months.

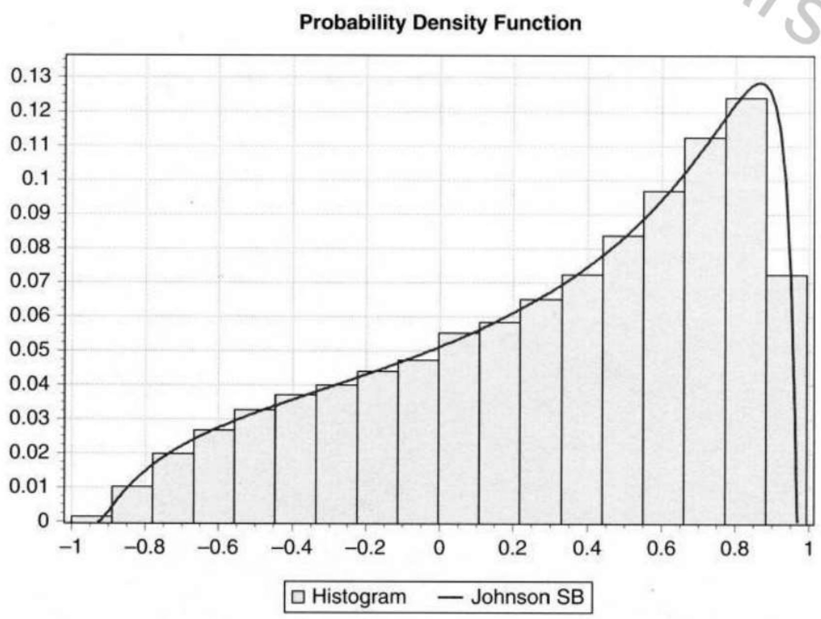


FIGURE 7-6 Histogram of 426,300 correlations between the Dow 30 stocks from 1972 to 2012. The continuous line shows the Johnson SB distribution, which provided the best fit.

HOW ARE EQUITY CORRELATIONS DISTRIBUTED?

The input data of our distribution tests are daily correlation values between all 30 Dow stocks from 1972 to 2012. This resulted in 426,300 correlation values. The distribution is shown in Figure 7-6.

From Figure 7-6, we observe the mostly positive correlations between the stocks in the Dow. In fact, 77.23% of all 426,300 correlation values were positive.

We tested 61 distributions for fitting the histogram in Figure 7-6, applying three standard fitting tests: (1) Kolmogorov-Smirnov, (2) Anderson-Darling, and (3) chi-squared. Not surprisingly, the versatile Johnson SB distribution with four parameters, γ and δ for the shape, μ for location, and σ for scale, provided the best fit. Standard distributions such as normal distribution, lognormal distribution, or beta distribution provided a poor fit.

IS EQUITY CORRELATION VOLATILITY AN INDICATOR FOR FUTURE RECESSIONS?

In our study from 1972 to 2012, six recessions occurred:

1. A severe recession in 1973–1974 following the first oil price shock.
2. A short recession in 1980.
3. A severe recession in 1981–1982 following the second oil price shock.
4. A mild recession in 1990–1991.
5. A mild recession in 2001 after the Internet bubble burst.
6. The Great Recession from 2007 to 2009 following the global financial crisis.

Table 7-2 displays the relationship of a change in the correlation volatility preceding the start of a recession.

From Table 7-2 we observe the severity of the 2007–2009 Great Recession, which exceeded the severity of the oil price shock–induced recessions in 1973–1974 and 1981–1982.

From Table 7-2 we also notice that, except for the mild recession in 1990–1991, before every recession a downturn in correlation volatility occurred. This coincides with the fact that correlation volatility is low in an expansionary

TABLE 7-2 Decrease in Correlation Volatility Preceding a Recession

	% Change in Correlation Volatility before Recession	Severity of Recession (% Change of GDP)
1973–1974	−7.22%	−11.93%
1980	−10.12%	−6.53%
1981–1982	−4.65%	−12.00%
1990–1991	0.06%	−4.05%
2001	−5.55%	−1.80%
2007–2009	−2.64%	−14.75%

The decrease in correlation volatility is measured as a six months change of six-month moving average correlation volatility. The severity of the recession is measured as the total GOP decline during the recession.

period (see Table 7-1), which often precedes a recession. However, the relationship between a decline in volatility and the severity of the recession is statistically nonsignificant. The regression function is almost horizontal and the R^2 is close to zero. Studies with more data, going back to 1920, are currently being conducted.

PROPERTIES OF BOND CORRELATIONS AND DEFAULT PROBABILITY CORRELATIONS

Our preliminary studies of 7,645 bond correlations and 4,655 default probability correlations display properties similar to those of equity correlations. Correlation levels were higher for bonds (41.67%) and slightly lower for default probabilities (30.43%) compared to equity correlation levels (34.83%). Correlation volatility was lower for bonds (63.74%) and slightly higher for default probabilities (87.74%) compared to equity correlation volatility (79.73%).

Mean reversion was present in bond correlations (25.79%) and in default probability correlations (29.97%). These levels were lower than the very high equity correlation mean reversion of 77.51%.

The default probability correlation distribution is similar to the equity correlation distribution (see Figure 7-4) and can be replicated best with the Johnson SB distribution. However, the bond correlation distribution shows a more normal shape and can be best fitted with the generalized extreme value distribution and quite well with the normal distribution. The bond correlation and default probabilities results are currently being verified with a larger sample database.

SUMMARY

The following are the main findings of the empirical correlation analysis.

- Our study confirmed that the worse the state of the economy, the higher are equity correlations. Equity correlations were extremely high in the Great Recession of 2007 to 2009 and reached 96.97% in February 2009.
- Equity correlation volatility is lowest in an expansionary period and higher in normal and recessionary economic periods. Traders and risk managers should take these higher correlation levels and higher correlation volatility that markets exhibit during economic distress into consideration.
- Equity correlation levels and equity correlation volatility are positively related.
- Equity correlations show very strong mean reversion. The Dow correlations from 1972 to 2012 showed a monthly mean reversion of 77.51%. Hence, when modeling correlation, mean reversion should be included in the model.
- Since equity correlations display strong mean reversion, they display low autocorrelation. Autocorrelations show the typical decrease with respect to time lags.
- The equity correlation distribution showed a distribution that can be replicated well with the Johnson SB distribution. Other distributions such as normal, lognormal, and beta distributions did not provide a good fit.
- First results show that bond correlations display properties similar to those of equity correlations. Bond correlation levels and bond correlation volatilities are generally higher in bad economic times. In addition, bond correlations exhibit mean reversion, although lower mean reversion than equity correlations exhibit.
- First results show that default correlations also exhibit properties seen in equity correlations. Default probability correlation levels are slightly lower than equity correlations levels, and default probability correlation

volatilities are slightly higher than equity correlations. Studies with more data are currently being conducted.

References and Suggested Readings

Ang, A., and J. Chen. 2002. "Asymmetric Correlations of Equity Portfolios." *Journal of Financial Economics* 63:443–494.

Barndorff-Nielsen, O. E., and N. Shephard. 2004. "Econometric Analysis of Realized Covariation: High Frequency Covariance, Regression and Correlation in Financial Economics." *Econometrica* 72:885–925.

Bekaert, G., and C. R. Harvey. 1995. "Time-Varying World Market Integration." *Journal of Finance* 50:403–444.

Bollerslev, Tim. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31(3):307–327.

De Santis, G., B. Litterman, A. Vesval, and K. Winkelmann. 2003. "Covariance Matrix Estimation." In *Modern Investment Management: An Equilibrium Approach*, by Bob Litterman and the Quantitative Resources Group, Goldman Sachs Asset Management, 224–248. Hoboken, NJ: John Wiley & Sons.

Engle, R. 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation," *Econometrica* 50:987–1008.

Erb, C., C. Harvey, and T. Viskanta. 1994. "Forecasting International Equity Correlations." *Financial Analysts Journal*, November/December: 32–45.

Goetzmann, W. N., L. Li, and K. G. Rouwenhorst. 2005. "Long-Term Global Market Correlations." *Journal of Business* 78:1–38.

Ledoit, O., P. Santa-Clara, and M. Wolf. 2003. "Flexible Multivariate GARCH Modeling with an Application to International Stock Markets." *Review of Economics and Statistics* 85:735–747.

Longin, F., and B. Solnik. 1995. "Is the Correlation in International Equity Returns Constant: 1960–1990?" *Journal of International Money and Finance* 14(1):3–26.

Longin, F., and B. Solnik. 2001. "Extreme Correlation of International Equity Markets." *Journal of Finance* 56:649–675.

Moskowitz, T. 2003. "An Analysis of Covariance Risk and Pricing Anomalies." *Review of Financial Studies* 16:417–457.

Uhlenbeck, G. E., and L. S. Ornstein. 1930. "On the Theory of Brownian Motion." *Physical Review* 36: 823–841.

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

Statistical Correlation Models—Can We Apply Them to Finance?

8

Learning Objectives

After completing this reading you should be able to:

- Evaluate the limitations of financial modeling with respect to the model itself, calibration of the model, and the model's output.
- Assess the Pearson correlation approach, Spearman's rank correlation, and Kendall's τ, and evaluate their limitations and usefulness in finance.

Excerpt is Chapter 3 of Correlation Risk Modeling and Management *by Gunter Meissner.*

Great achievements involve great risk.

—Dalai Lama

Correlation models measure the degree of association between two or more variables. In this chapter we discuss three popular statistical correlation measures:

1. The Pearson correlation measure.
2. The Spearman rank correlation.
3. The Kendall τ.

We will analyze the properties of these correlation measures and evaluate whether it is appropriate to apply them to financial variables.

Let's first generally assess the role of models in finance.

Models are not perfect. That doesn't mean they are not useful.

—Robert Merton

A WORD ON FINANCIAL MODELS

The financial reality is extremely complex, with thousands of investors, who may behave irrationally, and numerous markets such as equity, fixed income, commodities, foreign exchange, real estate, and more, which are correlated. In addition, numerous financial institutions and a great number of financial products such as stocks, bonds, indexes, exchange-traded funds (ETFs), structured products, and derivatives exist. Naturally, there is no financial model that can replicate the immense complexity of these financial systems and their products. In the 1980s and 1990s, econometricians actually tried to replicate this complexity by models with hundreds of equations and variables. However, these models have failed to produce convincing results.

Does this mean financial modeling is senseless? No. Financial models are useful tools to help us understand the financial system. The value-at-risk (VaR) model, for example (discussed in Chapter 6), can give us a good estimate about our market risk. The copula model (which will be discussed in this chapter and which is applied in the Basel framework) can give us a good estimate about the credit value-at-risk (CVaR) of a portfolio. The Black-Scholes-Merton (BSM) model can give us a good idea about the value of an option.

Importantly, however, we have to be constantly aware of the limitations of any financial model. In this respect, there are three main aspects to consider.

The Financial Model Itself

In physics we have models and relationships that are accurate and constant in time. For example, the relationship $E = mc^2$ is true and will be true in the future in normal physical environments. However, financial models such as VaR, CVaR, and BSM are models that depend on market prices as inputs. These market prices are determined by human beings and can therefore behave randomly and unexpectedly. (That's why we often use random models in financial modeling, since we believe they can better replicate random human behavior.) Therefore, we always have to be aware that any financial model is at best an approximation of reality and should never be trusted uncritically.

We also have to assess whether the model actually has problems with respect to approximating reality. The VaR model, for example, assumes a normal distribution of asset returns. However, in reality we find that asset returns have fat tails, so it would be better to use a model with higher kurtosis. The Black-Scholes-Merton (BSM) option pricing model assumes a constant volatility for all strikes. However, it is well known that traders apply a volatility smile in currencies markets (i.e., higher volatility for out-of-the-money calls and puts) and a volatility skew in equity markets (i.e., higher volatility for out-of-the-money puts). Risk managers and traders have to critically observe whether a model should be applied to price and hedge, or the model risk is too high; that is, the application of the model to replicate reality is not feasible.

In rare cases, a financial model has mathematical inconsistencies. For example, this is the case when pricing up-and-out calls and puts and down-and-out calls and puts on the BSM model. If the knock-out strike KO is equal to the strike K and the interest rate r equals the underlying asset return q, the model is insensitive to changes in implied volatility. In the case of $KO = K$ and $r = q = 0$, the model is insensitive to changes in volatility and option maturity. Similarly, lookback options cannot be valued on a standard extension of the BSM model if the interest rate is equal to the return (i.e., $r = q$). In this case, a new algorithm must and can be found. Naturally, traders and risk managers have to be aware of mathematical inconsistencies of their models to avoid incorrect pricing and hedging.

The Calibration of the Model

Calibrating a model means finding the values for the parameters of the model, so that the model can produce the prices that are found in the market. Once we find those parameter values, the model can then be applied to value products for which few or no market prices are available. A critical issue is what time frame should be observed when calibrating the parameter values. This was a significant problem in the 2007–2009 global financial crisis. Risk managers fed their VaR, CVaR, and collateralized debt obligation (CDO) models the benign volatility and correlation data from noncrisis years, especially from 2003 to 2006. Hence the risk numbers that came out of the models significantly underestimated the catastrophic events from 2007 to 2009. Naturally, no model can produce realistic outputs when it is fed unrealistic inputs. In programming terminology: *Garbage in, garbage out!*

Models also need to be stress-tested. This means that extreme scenarios such as economic recessions and systemic market crashes are simulated. This can give risk managers and traders a good estimate of the risks of their portfolios in distressed times. Not surprisingly, Basel III and the U.S. Federal Reserve are requiring financial institutions to perform stress tests. In 2012, 15 of 19 financial institutions passed the Fed's required stress tests, i.e., "had enough capital to withstand a severe recession." See www.nytimes.com/2012/03/14/business/jpmorgan-passes-stress-test-raises-dividend.html for more details.

Mindfulness about Models

As mentioned earlier, no financial model is or will ever be able to replicate exactly the complexity of the financial system. Therefore we have to constantly be aware of the limitations of financial models. These limitations were ignored in the crisis of 2007 to 2009, when many traders and risk managers blindly trusted the new copula correlation model. When real estate prices declined sharply in 2007 to 2009, and structured products such as CDOs, which referenced mortgages, declined by 50% or more, the losses were not anticipated by the copula model for two reasons:

1. The correlation assumptions of the copula model were violated in the systemic crash. The copula model assumes a negative correlation between the equity tranche and senior tranches, as we already saw in Figure 6-3. However, correlation increased sharply during the crisis, and equity tranche values *and* senior tranche values both declined.
2. In addition, the copula models were calibrated with the benign data from low-risk periods, as mentioned previously.

In conclusion, there needs to be human judgment when the outputs of models are assessed. The outputs have to be viewed in consideration of the limitations of any financial model. As David Li, who transferred the copula model to finance, put it: "The most dangerous thing is when people believe everything that comes out of it [the copula model]."

We will now address correlation models used in statistics, which measure associations between two or more variables, and discuss their usefulness in finance.

STATISTICAL CORRELATION MEASURES

In the following section, we analyze the most widely applied correlation concept in science, the Pearson correlation model. We find that the Pearson correlation model, despite its popularity, has severe limitations when applied in financial analysis.

The Pearson Correlation Approach and Its Limitations for Finance

Most introductory statistics courses cover the Pearson product moment correlation coefficient or Pearson correlation coefficient ρ. The Pearson approach measures the strength of the linear association between two variables. In fact, we have already applied the Pearson approach numerous times in this book, for example in Chapter 6 when discussing correlations in investments, trading, and market risk management. Let's now look at the Pearson correlation model in detail. The Pearson correlation coefficient ρ is defined:

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \tag{8.1}$$

where X and Y are sets $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_n\}$ with the elements $x_1, \ldots, x_n$ and $y_1, \ldots, y_n \in R$.

We already used this equation with x_t being the asset returns of asset X at time t and y_t being the asset returns of asset Y at time t. $\sigma(X)$ and $\sigma(Y)$ in Equation (8.1) are the standard deviation of X and Y, respectively. The covariance in Equation (8.1) is defined as in Equation (8.2):

$$\mathrm{Cov}(X, Y) = \frac{1}{n-1}\sum_{t=1}^{n}(X_t - \mu_X)(Y_t - \mu_Y) \tag{8.2}$$

If we deal with random sets (whose outcome is unknown, such as rolling a die), the covariance is quantified with expectation values. Hence the covariance is $E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$, where $E(X)$ is the expected value of X and $E(Y)$ is the expected value of Y. $E(XY)$ is the expected value of the product of the random variables X and Y. Also, the variances of X and Y are $\sigma_X^2 = E(X^2) - E(X)^2$ and $\sigma_Y^2 = E(Y^2) - E(Y)^2$, respectively. Hence for random sets Equation (8.1) assumes the equivalent form

$$\rho_1(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - (E(X))^2}\sqrt{E(Y^2) - (E(Y))^2}} \tag{8.3}$$

The application of the Pearson correlation coefficient and the related least squares linear regression analysis is a standard statistical tool in finance; see for example Fitch (2006), who regresses correlations between asset returns with sector specific regional factor loadings. Das et. al. 2006 linearly regress the mean probability of default with market volatility and debt to asset ratios. Altman et. al. (2005) apply the Pearson correlation approach and its extensions to verify the negative correlation between default rates and recovery rates.

However, the limitations of the Pearson correlation approach in finance are evident for five reasons:

1. Linear dependencies, which are evaluated in Equations (8.1) and (8.3), do not appear often in finance. We have already seen in Figures 6-2 to 6-4 and 6-6 to 7-8 that financial relationships are typically nonlinear.
2. Zero correlation derived in Equations (8.1) and (8.3) does not necessarily mean independence. This is because only the first two moments, mean and standard deviation, are considered in Equations (8.1) and (8.3). For example, the parabola $Y = X^2$ will lead to $\rho = 0$, which is arguably misleading. See Appendix A of Chapter 6 for details.
3. Linear correlation measures are natural dependence measures only if the joint distribution of the variables is elliptical.[1] However, only a few distributions such as the multivariate normal distribution and the multivariate Student's t distribution are special cases of elliptical distributions, for which linear correlation measure can be meaningfully interpreted.[2]
4. The variances of the sets X and Y have to be finite. However, for distributions with strong kurtosis, for example the Student's t distribution with $v \leq 2$, the variance is infinite.
5. In contrast to the copula approach, which is invariant to strictly increasing transformations, the Pearson correlation approach is typically not invariant to transformations. For example, the Pearson correlation between pairs X and Y is in general different from the Pearson correlation between the pairs $\ln(X)$ and $\ln(Y)$. Hence the information value of the Pearson correlation coefficient after data transformation is limited.

For these reasons, the application of the Pearson correlation concept in finance is questionable. The linear Pearson correlation coefficient can at best serve as an approximation for the typically nonlinear relationship between financial variables.

Let's now discuss two ordinal correlation measures and evaluate their usefulness for financial applications.

Spearman's Rank Correlation

Ordinal correlation measures such as Spearman's rank correlation and Kendall's τ have gained popularity in finance in the recent past. Let's discuss them both and then assess whether they are applicable to finance.

The Spearman's rank correlation concept is an ordinal correlation measure. This means that the numerical values of the elements in a set are not relevant for deriving the correlation, just the order of the elements. The Spearman's correlation coefficient is sometimes referred to as the Pearson correlation coefficient for ranked variables. It will result in a perfect correlation coefficient of 1 if an increase in the elements x_i is always accompanied by an increase in y_i, regardless of the numerical increase, and vice versa. The Spearman correlation approach is nonparametric in the sense that it can be applied without requiring knowledge of the joint distribution of the variables.

Let's look at the example in Chapter 6. We have two assets, which have performed as in Table 8-1.

[1] An elliptical distribution is a generalization of multivariate normal distributions.

[2] See Embrechts, McNeil, and Straumann (1999) and Bingham and Kiesel (2002) for details.

We had derived the Pearson correlation coefficient for the assets' returns in Table 6-1 as −0.7403. Let's now derive the Spearman rank correlation coefficient.

1. We first have to order the return set pairs of X and Y with respect to the set X. This is done in columns 2 and 3 of Table 8-2.
2. We then derive the ranks of X_i and Y_i. This is done in columns 4 and 5.
3. We now derive the difference of the ranks in column 6 and square the difference in column 7.

The Spearman rank correlation coefficient ρ_s is defined as

$$\rho_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \quad \textbf{(8.4)}$$

For our example in Table 8-2, we derive $\rho_S = 1 - \frac{6 \times 36}{5(5^2 - 1)} = -0.8$. Since the Spearman correlation coefficient is defined between −1 and +1, we find that the returns of assets X and Y are highly negatively correlated according to the Spearman rank correlation concept. The −0.8 Spearman correlation is similar to the Pearson correlation coefficient of −0.7403, which we had derived in Chapter 6. Before we evaluate the usefulness of the Spearman correlation coefficient for finance, let's discuss another rank correlation measure.

Kendall's τ

Kendall's τ is a further, fairly popular ordinal correlation measure applied in finance. As with the Spearman's correlation coefficient, the Kendall τ is nonparametric and will result in a perfect correlation coefficient of 1 if an increase in the variable x is always accompanied by an increase in y, regardless of the numerical increase, and vice versa. In most other cases, the two rank correlation measures are not equal.

The Kendall τ is defined as

$$\tau = \frac{n_c - n_d}{n(n-1)/2} \quad \textbf{(8.5)}$$

where n_c is the number of concordant data pairs and n_d is the number of discordant pairs.

A concordant pair is defined as any pair of observations where $x_t > y_t$ and $x_{t^*} > y_{t^*}$ or $x_t < y_t$ and $x_{t^*} < y_{t^*}$, where $t \neq t^*$.

TABLE 8-1 Performance of a Portfolio with Two Assets

	Asset X	Asset Y	Return of Asset X	Return of Asset Y
2008	100	200		
2009	120	230	20.00%	15.00%
2010	108	460	−10.00%	100.00%
2011	190	410	75.93%	−10.87%
2012	160	480	−15.79%	17.07%
2013	280	380	75.00%	−20.83%
		Average	29.03%	20.07%

TABLE 8-2 Ranked Asset Returns to Derive the Spearman Correlation Coefficient

	Ranked Return of X_i	Assigned (same year) Return of Y_i	Rank of X_i	Rank of Y_i	d_i	d_i^2
2012	−15.79%	17.07%	1	4	−3	9
2010	−10.00%	100.00%	2	5	−3	9
2009	20.00%	15.00%	3	3	0	0
2013	75.00%	−20.83%	4	1	3	9
2011	75.93%	−10.87%	5	2	3	9
						Sum = 36

A discordant data pair is where $x_t > y_t$ and $x_{t^*} < y_{t^*}$ or $x_t > y_t$ and $x_{t^*} < y_{t^*}$, where $t \neq t^*$.

A pair is neither concordant nor discordant if $x_t = y_t$ or $x_{t^*} = y_{t^*}$.

Let's calculate the Kendall τ for our example in Table 8-2. We have five observation pairs and therefore $5 \times (5 - 1)/2 = 10$ combinations of pairs to evaluate. We have the two concordant pairs {(1,4),(2,5)}, {(4,1),(5,2)} and the four discordant pairs {(1,4),(4,1)}, {(1,4),(5,2)}, {(2,5),(4,1)}, ((2,5),(5,2)}. The pairs {(1,4),(3,3)}, ((2,5),(3,3)}, {(3,3),(4,1)}, and ((3,3),(5,2)} are neither concordant nor discordant. From Equation (8.5), we derived the Kendall $\tau = \frac{2-4}{5(5-1)/2} = -0.2$. Since Kendall's τ is defined between -1 and $+1$, we can interpret the -0.2 as: The association between the returns of assets X and Y is slightly negative when calculated by the Kendall τ concept.

SHOULD WE APPLY SPEARMAN'S RANK CORRELATION AND KENDALL'S τ IN FINANCE?

Rank correlation measures have been popular in analyzing rating categories (i.e., the categories AAA, AA, A, . . . , to D), since these are ordinal. Cherubini and Luciano (2002) apply Spearman's rank correlation and Kendall's τ to analyze the dependence of market prices and counterparty risk measured by rating categories in a copula setting. Burtschell, Gregory and Laurent (2008) compare Kendall's τ to various copulas and find significant difference in the correlation approaches when inferring CDO tranche spreads. Anderson (2010) analyzes CDS correlations and finds that Spearman's rank correlations for CDS spreads more than doubled during the financial crisis from July 2007 to March 2009.

Ordinal rank correlation measures are an appropriate tool if the observations are ordinal. The problem with applying ordinal rank correlations to cardinal observations is that ordinal correlations are less sensitive to outliers. To show this, let's double the outliers of the returns of asset X in Table 8-2. We derive Table 8-3.

The values in Table 8-3 result in an increase of the Pearson correlation coefficient from -0.7402 to -0.6108 in Table 8-2, which will increase risk when input into VaR. However, since the numerical value of outliers in the rank correlations Spearman and Kendall are irrelevant, the correlations in the rank correlation measures do not change. This is an unwelcome property, especially in risk management. For example, a severe loss that may have occurred in the past is not numerically assessed. This can lead to the illusion of less risk than is actually present!

A special problem with the Kendall τ is when many non-concordant and many nondiscordant pairs occur, which are omitted in the calculation. This may lead to only a few concordant and discordant pairs, which can distort the Kendall τ coefficient. To a certain degree this is the case in our example of Table 8-2. Of the 10 observation pairs, four are neither concordant nor discordant, leaving just six pairs to be evaluated.

We can conclude that the application of statistical correlation measures to assess financial correlations is limited. The main concern with the Pearson correlation coefficient

TABLE 8-3 Table 8-2 but with Increased Outliers for Asset X

	Ranked Return of X_i	Assigned (same year) Return of Y_i	Rank of X_i	Rank of Y_i	d_i	d_i^2
2012	−31.58%	17.07%	1	4	−3	9
2010	−10.00%	100.00%	2	5	−3	9
2009	20.00%	15.00%	3	3	0	0
2013	75.00%	−20.83%	4	1	3	9
2011	151.86%	−10.87%	5	2	3	9
						Sum = 36

is that it evaluates linear relationships. However, financial variables are mostly nonlinear. In addition, the limited interpretation for nonelliptical data is problematic. Statistical rank correlation measures should not be applied to cardinal financial variables, especially since the sensitivity to outliers is low. These outliers, for example high losses, are critical when evaluating correlations and risk. Statistical rank correlation measures are appropriate only if the financial variables are ordinal as, for example, rating categories.

Since the application of the statistical correlation concepts is limited in finance, quants have developed specific financial correlation measures, which we will discuss in Chapter 9.

SUMMARY

In this chapter, we first generally assessed the value of financial modeling. The financial reality is extremely complex, with numerous markets, complex products, and—most critically—investors who can behave irrationally. No financial model will ever be able to replicate this complex financial reality perfectly. However, this does not mean financial models are useless. Financial models can give a good approximation of the reality and help us better understand the behavior of financial processes. They can further help us forecast future crises and help us understand and manage financial risk.

In this chapter we also discussed statistical correlation approaches and investigated whether they are appropriate for financial modeling. By far the most widely applied correlation concept in statistics is the Pearson correlation model. The reason for the popularity of the Pearson model is its mathematical simplicity and high intuition. The Pearson correlation model is widely applied in finance. But should we actually apply it to financial modeling? The answer is "not really," especially not for complex financial correlations, as, for example, correlations in a CDO; see Chapter 9.

The Pearson approach suffers from a variety of problems: most importantly, it measures only *linear* relationships. However, most financial correlations are nonlinear. As a result, zero correlation derived by the Pearson approach does not necessarily mean independence (see also Appendix A of Chapter 6), so the Pearson correlation outcome can be quite misleading. The Pearson correlation approach can at best serve as a good approximation of the mostly nonlinear financial correlations found in practice. When applying the simple, linear Pearson correlation model to financial correlations, we should constantly be aware of its severe limitations.

Ordinal or rank correlations measures such as Spearman's rank correlation and Kendall's τ do not consider numerical values but just the order of the elements (i.e., higher or lower) when deriving correlations. For financial variables that are ordinal, such as rating categories, ordinal correlation measures are appropriate. However, the application of ordinal correlation measures to cardinal data is not appropriate, since ordinal correlation measures ignore the extreme values of outliers. This can give the illusion of less risk than is present.

References and Suggested Readings

Altman, E., B. Brooks, A. Resti, and A. Sironi. 2005. "The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications." *Journal of Business* 78(6): 2203–2227.

Anderson, M. 2010. "Contagion and Excess Correlation in Credit Default Swaps," Working paper.

Bingham, N., and R. Kiesel. 2001. "Semi-Parametric Modelling in Finance: Theoretical Foundations." *Quantitative Finance* 1:1–10.

Burtschell, X., J. Gregory, and J-P. Laurent. 2008. "A Comparative Analysis of CDO Pricing Models" in *The Definitive Guide to CDOs—Market, Application, Valuation and Hedging,* London: Risk Books.

Cherubini, U., and E. Luciano. 2002. "Copula Vulnerability," *RISK.*

Das, S., L. Freed, G. Geng, and N. Kapadia. 2006. "Correlated Default Risk," *The Journal of Fixed Income*, Fall.

Embrechts, A., A. McNeil, and D. Straumann. 1999. "Correlations and Dependence in Risk Management: Properties and Pitfalls." *Mimeo ETHZ Zentrum.*

Fitch. 2006. "Global Rating Criteria for Collateralized Debt Obligations," from www.fitchratings.com.

Pearson, K. 1900. "On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling." *Philosophical Magazine Series 5* 50(302):157–175.

Soper, H. E., A. W. Young, B. M. Cave, A. Lee, and K. Pearson. 1917. "On the Distribution of the Correlation Coefficient in Small Samples: Appendix II to the Papers of 'Student' and R. A. Fisher; A Co-operative Study." *Biometrika* 11:328–413.

Spearman, Charles B. 2005. *The Abilities of Man: Their Nature and Measurement.* New York: Blackburn Press.

Financial Correlation Modeling—Bottom-Up Approaches

9

Learning Objectives

After completing this reading you should be able to:

- Explain the purpose of copula functions and the translation of the copula equation.
- Describe the Gaussian copula and explain how to use it to derive the joint probability of default of two assets.
- Summarize the process of finding the default time of an asset correlated to all other assets in a portfolio using the Gaussian copula.

Excerpt is from Chapter 4 of Correlation Risk Modeling and Management *by Gunter Meissner.*

COPULA CORRELATIONS

A fairly recent and famous as well as infamous correlation approach applied in finance is the copula approach. Copulas go back to Abe Sklar (1959). Extensions are provided by Schweizer and Wolff (1981) and Schweizer and Sklar (1983). One-factor copulas were introduced to finance by Oldrich Vasicek in 1987. More versatile, multivariate copulas were applied to finance by David Li in 2000.

When flexible copula functions were introduced to finance in 2000, they were enthusiastically embraced but then fell into disgrace when the global financial crisis hit in 2007. Copulas became popular because they could presumably solve a complex problem in an easy way: It was assumed that copulas could correlate multiple assets, for example the 125 assets in a CDO, with a single (although multidimensional) function. There are benefits and limitations of the Gaussian copula for valuing CDOs. Let's first look at the math of the copula correlation concept.

Copula functions are designed to simplify statistical problems. They allow the joining of multiple univariate distributions to a single multivariate distribution. Formally, a copula function C transforms an n-dimensional function on the interval [0, 1] into a unit-dimensional one:

$$C: [0, 1]^n \rightarrow [0, 1] \tag{9.1}$$

More explicitly, let $G_i(u_i) \in [0,1]$ be a univariate, uniform distribution with $u_i = u_1, \ldots, u_n$, and $i \in N$. Then there exists a copula function C such that

$$C[G_1(u_1), \ldots, G_n(u_n)] = F_n[F_1^{-1}(G_1(u_1)), \ldots, F_n^{-1}(G_n(u_n)); \rho_F] \tag{9.2}$$

where $G_i(u_i)$ are called marginal distributions, F_n is the joint cumulative distribution function, F_i^{-1} is the inverse of F_i, and ρ_F is the correlation structure of F_n.

Equation (9.2) reads: Given are the marginal distributions $G_1(u_1)$ to $G_n(u_n)$. There exists a copula function that allows the mapping of the marginal distributions $G_1(u_1)$ to $G_n(u_n)$ via F^{-1} and the joining of the (abscise values) $F^{-1}(G_i(u_i))$ to a single, n-variate function $F_n[F_1^{-1}(G_1(u_1)), \ldots, F_n^{-1}(G_n(u_n))]$ with correlation structure of ρ_F.

If the mapped values $F_i^{-1}(G_i(u_i))$ are continuous, it follows that C is unique. For detailed properties and proofs of Equation (9.2), see Sklar (1959) and Nelsen (2006).

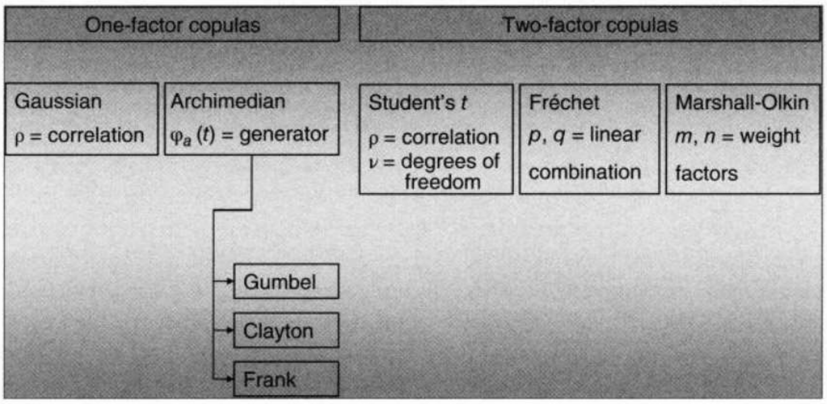


FIGURE 9-1 Popular copula functions in finance.

Numerous types of copula functions exist. They can be broadly categorized in one-parameter copulas as the Gaussian copula[1] and the Archimedean copula family, the most popular being Gumbel, Clayton, and Frank copulas. Often cited two-parameter copulas are Student's *t*, Fréchet, and Marshall-Olkin. Figure 9-1 shows an overview of popular copula functions.

The Gaussian Copula

Due to its convenient properties, the Gaussian copula C_G is among the most applied copulas in finance. In the n-variate case, it is defined

$$C_G[G_1(u_1), \ldots, G_n(u_n)] = M_n[N^{-1}(G_1(u_1)), \ldots, N^{-1}(G_n(u_n)); \rho_M] \tag{9.3}$$

where M_n is the joint, n-variate cumulative standard normal distribution with ρ_M, the $n \times n$ symmetric, positive-definite correlation matrix of the n-variate normal distribution M_n. N^{-1} is the inverse of a univariate standard normal distribution.

If the $G_x(u_x)$ are uniform, then the $N^{-1}(G_x(u_x))$ are standard normal and M_n is standard multivariate normal. For a proof, see Cherubini et. al. 2004.

It was David Li (2000) who transferred the copula approach of Equation (9.3) to finance. He defined the cumulative default probabilities Q for entity i at a fixed

[1] Strictly speaking, only the *bivariate* Gaussian copula is a one-parameter copula, the parameter being the copula correlation coefficient. A multivariate Gaussian copula may incorporate a correlation *matrix*, containing various correlation coefficients.

time t, $Q_i(t)$ as marginal distributions. Hence we derive the Gaussian default time copula C_{GD},

$$C_{GD}[Q_i(t), \ldots, Q_n(t)] = M_n[N^{-1}(Q_1(t)), \ldots, N^{-1}(Q_n(t)); \rho_M] \quad \textbf{(9.4)}$$

Equation (9.4) reads: Given are the marginal distributions, that is, the cumulative default probabilities Q of entities i = 1 to n at times t, $Q_i(t)$. There exists a Gaussian copula function C_{GD}, which allows the mapping of the marginal distributions $Q_i(t)$ via N^{-1} to standard normal and the joining of the (abscise values) $N^{-1}Q_i(t)$ to a single n-variate standard normal distribution M_n with the correlation structure ρ_M.

More precisely, in Equation (9.4) the term N^{-1} maps the cumulative default probabilities Q of asset i for time t, $Q_i(t)$, percentile to percentile a univariate standard normal distribution. So the 5th percentile of $Q_i(t)$ is mapped to the 5th percentile of the standard normal distribution, the 10th percentile of $Q_i(t)$ is mapped to the 10th percentile of the standard normal distribution, and so forth. As a result, the $N^{-1}(Q_i(t))$ in Equation (9.4) are abscise (x-axis) values of the standard normal distribution. For a numerical example, see Example 9.1 and Figure 9-2. The $N_i^{-1}(Q_i(t))$ are then joined to a single n-variate distribution M_n by applying the correlation structure of the multivariate normal distribution with correlation matrix ρ_M. The probability of n correlated defaults at time t is given by M_n.

We will now look at the Gaussian copula in an example.

Example 9.1 Deriving the Joint Probability of Default of Two Entities with the Gaussian Copula

Let's assume we have two companies, B and Caa, with their estimated default probabilities for year 1 to 10 as displayed in Table 9-1.

Default probabilities for investment grade companies typically increase in time, since uncertainty increases with time. However, in Table 9-1 we have two companies currently in distress. For these companies the next few years will be the most difficult. If they survive these next years, their default probability will decrease.

Let's now find the joint default probabilities of the companies B and Caa for any time t with the Gaussian copula function (9.4). First we map the cumulative default probabilities $Q(t)$, which are in columns 3 and 5 in Table 9-1, to the standard normal distribution via $N^{-1}(Q(t))$. Computationally this can be done with = normsinv(Q(t)) in Excel or norminv(Q(t)) in MATLAB. Graphically the mapping $N^{-1}(Q(t))$ can be represented in two steps, which are displayed in Figure 9-2. In the lower graph of Figure 9-2, the cumulative default probability of asset B, $Q_B(t)$, is

TABLE 9-1 Default Probability and Cumulative Default Probability of Companies *B* and *Caa*

Default Time t	Company *B* Default Probability	Company *B* Cumulative Default Probability $Q_B(t)$	Company *Caa* Default Probability	Company *Caa* Cumulative Default Probability $Q_{Caa}(t)$
1	6.51%	6.51%	23.83%	23.83%
2	7.65%	14.16%	13.29%	37.12%
3	6.87%	21.03%	10.31%	47.43%
4	6.01%	27.04%	7.62%	55.05%
5	5.27%	32.31%	5.04%	60.09%
6	4.42%	36.73%	5.13%	65.22%
7	4.24%	40.97%	4.04%	69.26%
8	3.36%	44.33%	4.62%	73.88%
9	2.84%	47.17%	2.62%	76.50%
10	2.84%	50.01%	2.04%	78.54%

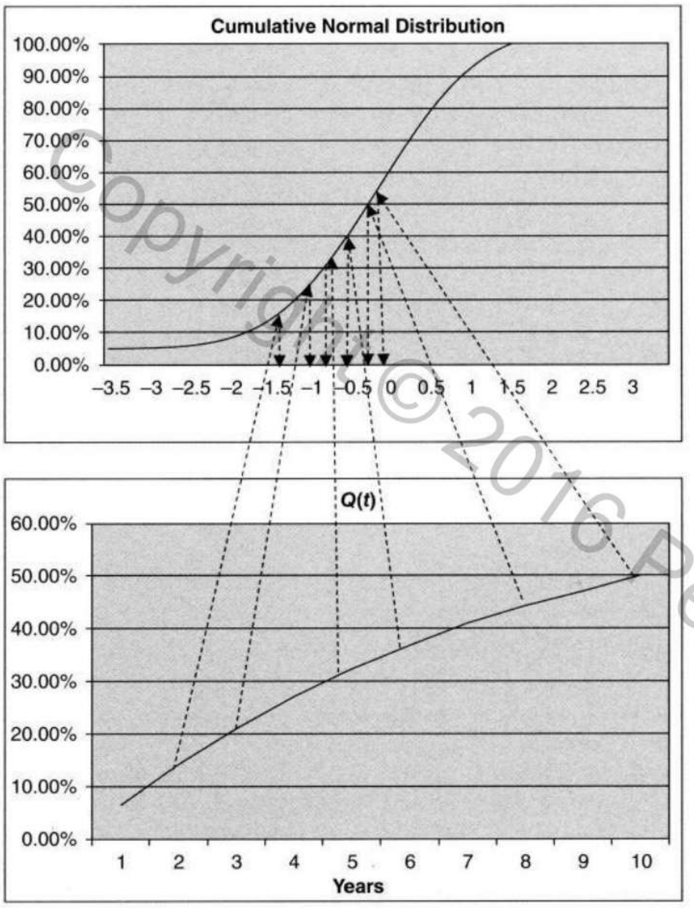


FIGURE 9-2 Graphical representation of the copula mapping $N^{-1}(Q(t))$.

displayed. We first map these cumulative probabilities percentile to percentile to a cumulative standard normal distribution in the upper graph of Figure 9-2 (up arrows). In a second step the abscise (x-axis) values of the cumulative normal distribution are found (down arrows).

The same mapping procedure is done for company *Caa*; the cumulative default probabilities of company *Caa*, which are displayed in Table 9-1 in column 5, are mapped percentile to percentile to a cumulative standard normal distribution via $N^{-1}(Q_{Caa}(t))$.

We have now derived the percentile to percentile mapped cumulative default probability values of our companies to a cumulative standard normal distribution. These values are displayed in Table 9-2, columns 3 and 5.

We can now use the derived $N^{-1}(Q_B(t))$ and $N^{-1}(Q_{Caa}(t))$ and apply them to Equation (9.4). Since we have only $n = 2$ companies B and Caa in our example, Equation (9.4) reduces to

$$M_2[N^{-1}(Q_B(t)), N^{-1}(Q_{Caa}(t)); \rho] \quad \textbf{(9.5)}$$

From Equation (9.5) we see that since we have only two assets in our example, we have only one correlation coefficient ρ, not a correlation matrix ρ_M.

Importantly, the copula model now assumes that we can apply the correlation structure ρ_M or a single ρ of the multivariate distribution (in our case the Gaussian multivariate distribution M) to the transformed marginal distributions $N^{-1}(Q_B(t))$ and $N^{-1}(Q_{Caa}(t))$. This is done for mathematical and computational convenience.

The bivariate normal distribution M_2 is displayed in Figure 9-3.

The code for the bivariate cumulative normal distribution M can be found on the Internet.

We now have all necessary ingredients to find the joint default probabilities of our companies B and *Caa*. For example, we can answer the question: What is the joint default probability Q of companies B and *Caa* in the next year assuming a one-year Gaussian default correlation of 0.4? The solution is:

$$Q(t_B \le 1 \cap t_{Caa} \le 1) \equiv M(x_B \le -1.5133 \cap x_{Caa} \le -0.7118, \rho = 0.4) = 3.44\% \quad \textbf{(9.6)}$$

where t_B is the default time of company B and t_{Caa} is the default time of company *Caa*. x_B and x_{Caa} are the mapped abscise values of the bivariate normal distribution, which are derived from Table 9-2.

In another example, we can answer the question: What is the joint probability of company B defaulting in year 3 and company *Caa* defaulting in year 5? It is

$$Q(t_B \le 3 \cap t_{Caa} \le 5) \equiv M(x_B \le -0.8054 \cap x_{Caa} \le 0.2557, \rho = 0.4) = 16.93\% \quad \textbf{(9.7)}$$

Equations (9.6) and (9.7) show why this type of copula is also called default time copula. We are correlating the default times of two or more assets t_i.

TABLE 9-2 Cumulative Default Probabilities and Corresponding Standard Normal Percentiles of Companies *B* and *Caa*

Default Time *t*	Company *B* Cumulative Default Probability $Q_B(t)$	Company *B* Cumulative Standard Normal Percentiles $N^{-1}(Q_B(t))$	Company *Caa* Cumulative Default Probability $Q_{Caa}(t)$	Company *Caa* Cumulative Standard Normal Percentiles $N^{-1}(Q_{Caa}(t))$
1	6.51%	−1.5133	23.83%	−0.7118
2	14.16%	−1.0732	37.12%	−0.3287
3	21.03%	−0.8054	47.43%	−0.0645
4	27.04%	−0.6116	55.05%	0.1269
5	32.31%	−0.4590	60.09%	0.2557
6	36.73%	−0.3390	65.22%	0.3913
7	40.97%	−0.2283	69.26%	0.5032
8	44.33%	−0.1426	73.88%	0.6397
9	47.17%	−0.0710	76.50%	0.7225
10	50.01%	0.0003	78.54%	0.7906

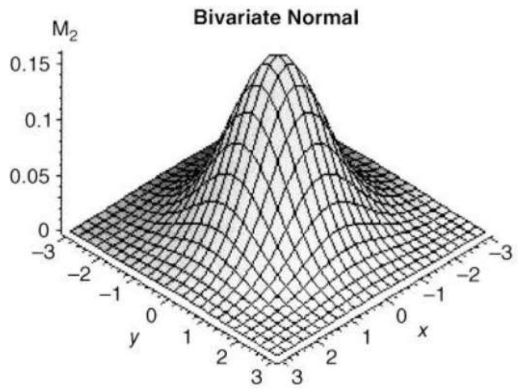


FIGURE 9-3 Bivariate (noncumulative) normal distribution M_2.

Simulating the Correlated Default Time for Multiple Assets

The preceding example considers only two assets. We will now find the default time for an asset that is correlated to the default times of all other assets in a portfolio using the Gaussian copula.

To derive the default time τ of asset *i*, τ_i, which is correlated to the default times of all other assets $i = 1, \ldots, n$, we first derive a sample $M_n(\cdot)$ from a multivariate copula [r.h.s. of Equation (9.5) in the Gaussian case], $M_n(\cdot) \in [0, 1]$. This is done via Cholesky decomposition. The sample includes the default correlation via the default correlation matrix ρ_M of the *n*-variate standard normal distribution M_n. An example of a default correlation matrix was displayed in Chapter 6 in Table 6-3. We equate the sample $(\cdot)$ from M_n, $M_n(\cdot)$ with the cumulative individual default probability Q of asset i at time τ, $Q_i(\tau_i)$. Therefore,

$$M_n(\cdot) = Q_i(\tau_i) \tag{9.8}$$

or

$$\tau_i = Q_i^{-1}(M_n(\cdot)) \tag{9.9}$$

There is no closed-form solution for Equations (9.8) or (9.9). To find the solution, we first take the sample $M_n(\cdot)$ and use Equation (9.8) to equate it to $Q_i(\tau_i)$. This can be done with a search procedure such as Newton-Raphson. We can also use a simple lookup function in Excel.

Let's assume the random drawing from $M_n(\cdot)$ was 35%. We now equate 35% with the market given function $Q_i(\tau_i)$ and find the expected default time of asset i,τ_i. This is displayed in Figure 9-4, where τ_i = 5.5 years. We repeat this procedure numerous times, for example 100,000 times, and average each τ_i of every simulation to find our estimate for τ_i. Importantly, the estimated default time of asset i, τ_i, includes the default correlation with the other assets in the portfolio, since the correlation matrix is an input of the n-variate standard normal distribution M_n.

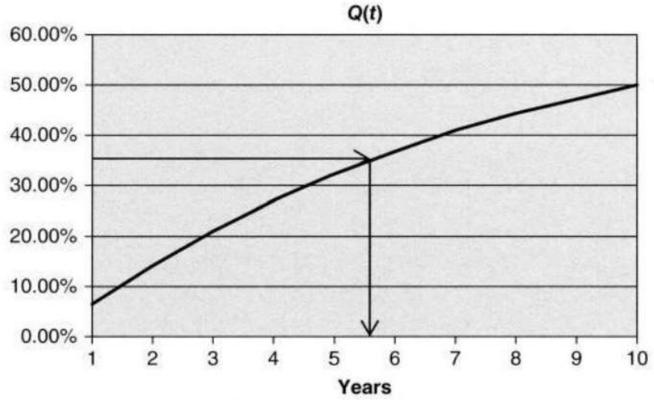


FIGURE 9-4 Finding the default time τ of 5.5 years from Equation (9.8) for a random sample of the n-variate normal distribution $M_n(\cdot)$ of 35%.

Source: CDO mapping explained.xls.

Empirical Approaches to Risk Metrics and Hedging

10

Learning Objectives

After completing this reading you should be able to:

- Explain the drawbacks to using a *DV01*-neutral hedge for a bond position.
- Describe a regression hedge and explain how it can improve a standard *DV01*-neutral hedge.
- Calculate the regression hedge adjustment factor, beta.
- Calculate the face value of an offsetting position needed to carry out a regression hedge.
- Calculate the face value of multiple offsetting swap positions needed to carry out a two-variable regression hedge.
- Compare and contrast level and change regressions.
- Describe principal component analysis and explain how it is applied to constructing a hedging portfolio.

Excerpt is Chapter 6 of Fixed Income Securities, *Third Edition, by Bruce Tuckman.*

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

Central to the *DV*01-style metrics and hedges and the multifactor metrics and hedges are implicit assumptions about how rates of different term structures change relative to one another. In this chapter, the necessary assumptions are derived directly from data on rate changes.

The chapter begins with single-variable hedging based on regression analysis. In the example of the section, a trader tries to hedge the interest rate risk of U.S. nominal *versus* real rates. This example shows that empirical models do not always describe the data very precisely and that this imprecision expresses itself in the volatility of the profit and loss of trades that depend on the empirical analysis.

The chapter continues with two-factor hedging based on multiple regression. The example for this section is that of an EUR swap market maker who hedges a customer trade of 20-year swaps with 10- and 30-year swaps. The quality of this hedge is shown to be quite a bit better than that of nominal *versus* real rates. Before concluding the discussion of regression techniques, the chapter comments on level *versus* change regressions.

The final section of the chapter introduces principal component analysis, which is an empirical description of how rates move together across the curve. In addition to its use as a hedging tool, the analysis provides an intuitive description of the empirical behavior of the term structure. The data illustrations for this section are taken from USD, EUR, GBP, and JPY swap markets. Considerable effort has been made to present this material at as low a level of mathematics as possible.

A theme across the illustrations of the chapter is that empirical relationships are far from static and that hedges estimated over one period of time may not work very well over subsequent periods.

SINGLE-VARIABLE REGRESSION-BASED HEDGING

This section considers the construction of a relative value trade in which a trader sells a U.S. Treasury bond and buys a U.S. Treasury TIPS (Treasury Inflation Protected Securities). As mentioned in the Overview, TIPS make real or inflation-adjusted payments by regularly indexing their principal amount outstanding for inflation. Investors in TIPS, therefore, require a relatively low real rate of return. By contrast, investors in U.S. Treasury bonds—called nominal bonds when distinguishing them from TIPS—require a real rate of return plus compensation for expected inflation plus, perhaps, an inflation risk premium. Thus the spread between rates of nominal bonds and TIPS reflects market views about inflation. In the relative value trade of this section, a trader bets that this inflation-induced spread will increase.

The trader plans to short \$100 million of the (nominal) 3⅝s of August 15, 2019, and, against that, to buy some amount of the TIPS 1⅞s of July 15, 2019. Table 10-1 shows representative yields and *DV*01s of the two bonds. The TIPS sells at a relatively low yield, or high price, because its cash flows are protected from inflation while the *DV*01 of the TIPS is relatively high because its yield is low. In any case, what face amount of the TIPS should be bought so that the trade is hedged against the level of interest rates, i.e., to both rates moving up or down together, and exposed only to the spread between nominal and real rates?

One choice is to make the trade *DV*01-neutral, i.e., to buy F^R face amount of TIPS such that

$$F^R \times \frac{.081}{100} = 100\text{mm} \times \frac{.067}{100}$$

$$F^R = 100\text{mm} \times \frac{.067}{.081} = \$82.7\text{mm} \tag{10.1}$$

This hedge ensures that if the yield on the TIPS and the nominal bond both increase or decrease by the same number of basis points, the trade will neither make nor lose money. But the trader has doubts about this choice because changes in yields on TIPS and nominal bonds may very well not be one-for-one. To investigate, the trader collects data on daily changes in yield of these two bonds from August 17, 2009, to July 2, 2010, which are then graphed in Figure 10-1, along with a regression line, to be discussed shortly. It is immediately apparent from the graph that, for example, a five basis-point change in the yield of the TIPS does not imply, with very

TABLE 10-1 Yields and *DV*01s of a TIPS and a Nominal U.S. Treasury as of May 28, 2010

Bond	Yield (%)	*DV*01
TIPS 1⅞s of 7/15/19	1.237	.081
3⅝s of 8/15/19	3.275	.067

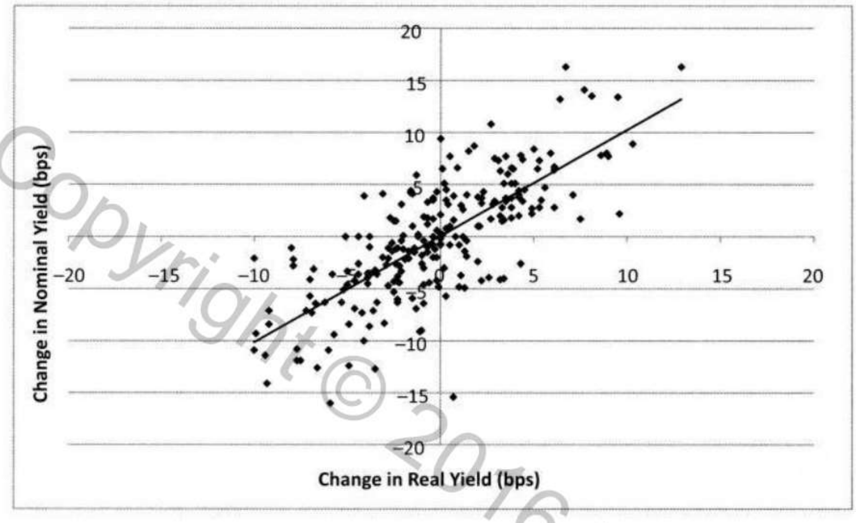


FIGURE 10-1 Regression of changes in the yield of the Treasury $3\frac{5}{8}$s of August 15, 2019, on changes in the yield of the TIPS 1.875s of July 15, 2019, from August 17, 2009, to July 2, 2010.

Relative to the *DV*01 hedge of $82.7 million in (10.1), the hedge in (10.2) increases the amount of TIPS to compensate for the empirical fact that, on average, the nominal yield changes by more than one basis point for every basis-point change in the real yield.

The next subsection introduces *regression analysis*, which is used both to estimate the coefficient 1.0189, used in Equation (10.2), and to assess the properties of the resulting hedge.

high confidence, a unique change in the nominal yield, nor even an average change of five basis points. In fact, while the daily change in the real yield was about five basis points several times over the study period, the change in the nominal yield over those particular days ranged from 2.2 to 8.4 basis points. This lack of a one-to-one yield relationship calls the *DV*01 hedge into question. For context, by the way, it should be noted that graphing the changes in the yield of one nominal Treasury against changes in the yield of another, of similar maturity, would result in data points much more tightly surrounding the regression line.

With respect to improving on the *DV*01 hedge, there is not much the trader can do about the dispersion of the change in the nominal yield for a given change in the real yield. That is part of the risk of the trade and will be discussed later. But the trader can estimate the average change in the nominal yield for a given change in the real yield and adjust the *DV*01 hedge accordingly. For example, were it to turn out—as it will—that the nominal yield in the data changes by 1.0189 basis points per basis-point change in the real yield, the trader could adjust the hedge such that

$$F^R \times \frac{.081}{100} = 100\text{mm} \times \frac{.067}{100} \times 1.0189$$

$$F^R = \$100\text{mm} \times \frac{.067}{.081} \times 1.0189 = \$84.3\text{mm} \qquad \textbf{(10.2)}$$

Least-Squares Regression Analysis

Let Δy_t^N and Δy_t^R be the changes in the yields of the nominal and real bonds, respectively, and assume that

$$\Delta y_t^N = \alpha + \beta \Delta y_t^R + \epsilon_t \qquad \textbf{(10.3)}$$

According to Equation (10.3), changes in the real yield, the *independent variable*, are used to predict changes in the nominal yield, the *dependent variable*. The intercept, α, and the slope, β, need to be estimated from the data. The error term ϵ_t is the deviation of the nominal yield change on a particular day from the change predicted by the model. *Least-squares estimation* of (10.3), to be discussed presently, requires that the model be a true description of the dynamics in question and that the errors have the same probability distribution, are independent of each other, and are uncorrelated with the independent variable.[1]

As an example of the relationship between the nominal and real yields in (10.3), say that the parameters estimated

[1] Since the nominal rate is the real rate plus the inflation rate, the error term in Equation (10.3) contains the change in the inflation rate. Therefore, the assumption that the independent variable be uncorrelated with the error term requires here that the real rate be uncorrelated with the inflation rate. This is a tolerable, though far from ideal, assumption: the inflation rate can have effects on the real economy and, consequently, on the real rate.

If the regression were specified such that the real rate were the dependent variable and the nominal rate the independent variable, the requirement that the error and the dependent variable be uncorrelated would certainly not be met. In that case, the error term contains the inflation rate and there is no credible argument that the nominal rate is even approximately uncorrelated with the inflation rate. Consequently, a more advanced estimation procedure would be required, like that of *instrumental variables*.

with the data, denoted $\hat{\alpha}$ and $\hat{\beta}$, are 0 and 1.02 respectively. Then, if Δy_t^R is 5 basis points on a particular day, the predicted change in the nominal yield, written $\Delta \hat{y}_t^N$, is

$$\begin{aligned}\Delta \hat{y}_t^N &= \hat{\alpha} + \hat{\beta}\Delta y_t^R \\ &= 0 + 1.02 \times 5 = 5.1\end{aligned} \tag{10.4}$$

Furthermore, should it turn out that the nominal yield changes by 5.5 basis points on that day, then the realized error that day, written $\hat{\epsilon}_t$, following Equation (10.3), is defined as

$$\begin{aligned}\hat{\epsilon}_t &= \Delta y_t^N - \hat{\alpha} - \hat{\beta}\Delta y_t^R \\ &= \Delta y_t^N - \Delta \hat{y}_t^N\end{aligned} \tag{10.5}$$

In this example,

$$\hat{\epsilon}_t = 5.5 - 5.1 = .4 \tag{10.6}$$

Least-squares estimation of α and β finds the estimates $\hat{\alpha}$ and $\hat{\beta}$ that minimize the sum of the squares of the realized error terms over the observation period,

$$\sum_t \hat{\epsilon}_t^2 = \sum_t (\Delta y_t^N - \hat{\alpha} - \hat{\beta}\Delta y_t^R)^2 \tag{10.7}$$

where the equality follows from (10.5). The squaring of the errors ensures that offsetting positive and negative errors are not considered as acceptable as zero errors and that large errors in absolute values are penalized substantially more than smaller errors.

Least-squares estimation is available through many statistical packages and spreadsheet add-ins. A typical summary of the regression output from estimating Equation (10.3) using the data in Figure 10-1 is given in Table 10-2. The $\hat{\beta}$ reported in the table is 1.0189, which says that, over the sample period, the nominal yield increases by 1.0189 basis points per basis-point increase in real yields. The constant term of the regression, $\hat{\alpha}$, is not very different from zero, which is typically the case in regressions of changes in a yield on changes in a comparable yield. The economic interpretation of this regularity is that a yield does not usually trend up or down while a comparable yield is not changing.

Table 10-2 reports standard errors of $\hat{\alpha}$ and $\hat{\beta}$ of .2529 and .0525, respectively. Under the assumptions of least squares and the availability of sufficient data, the parameters $\hat{\alpha}$ and $\hat{\beta}$ are normally distributed with means equal to the true model values, α and β, respectively, and with standard deviations that can be estimated as the standard errors given in the table. Therefore, relying on the properties of the normal distribution, the confidence interval $.0503 \pm 2 \times .2529$ or (−.4555, .5561) has a 95% chance of

TABLE 10-2 Regression Analysis of Changes in the Yield of the 3⅝s of August 15, 2019, on the Changes in Yield of the TIPS 1⅞s of July 15, 2019, from August 17, 2009, to July 2, 2010

No. of Observations	**229**	
R-Squared	**56.3%**	
Standard Error	**3.82**	
Regression Coefficients	**Value**	**Std. Error**
Constant ($\hat{\alpha}$)	0.0503	.2529
Change in Real Yield ($\hat{\beta}$)	1.0189	.0595

falling around the true value α. And since this confidence interval does include the value zero, one cannot reject the statistical hypothesis that $\alpha = 0$. Similarly, the 95% confidence interval with respect to β is $1.0189 \pm 2 \times .0595$, or (.8999, 1.1379). So, while regression hedging makes heavy use of the point estimate $\hat{\beta} = 1.0189$, the true value of β may very well be somewhat higher or lower.

Substituting the estimated coefficients from Table 10-2 into the predicted regression equation in the first line of (10.4),

$$\begin{aligned}\Delta \hat{y}_t^N &= \hat{\alpha} + \hat{\beta}\Delta y_t^R \\ \Delta \hat{y}_t^N &= .0503 + 1.0189 \times \Delta y_t^R\end{aligned} \tag{10.8}$$

This relationship is known as the *fitted regression line* and is the straight line through the data that appears in Figure 10-1.

Table 10-2 reports two other useful statistics, the R-squared and the standard error of the regression. The R-squared in this case is 56.3%, which means that 56.3% of the variance of changes in the nominal yield can be explained by the model. In a one-variable regression, the R-squared is just the square of the correlation of the two changes, so the correlation between changes in the nominal and real yields is the square root of 56.3% or about 7.5%. This is a relatively low number compared with typical correlations between changes in two nominal yields, echoing the comment made in reference to the relatively wide dispersion of the points around the regression line in Figure 10-1.

The second useful statistic reported in Table 10-2 is the standard error of the regression, denoted here by $\hat{\sigma}$ and given as 3.82 basis points. Algebraically, $\hat{\sigma}$ is essentially the standard deviation of the realized error terms

$\hat{\epsilon}_t^2$ defined in Equation (10.5). Graphically, each $\hat{\epsilon}_t$ is the vertical line from a data point directly down or up to the regression line and $\hat{\sigma}$ is essentially the standard deviation of these distances. Either way, $\hat{\sigma}$ measures how well the model fits the data in the same units as the dependent variable, which, in this case, are basis points.

The Regression Hedge

The use of the regression coefficient in the hedging example of this section was discussed in the development of Equation (10.2). More formally, denoting the face amounts of the real and nominal bonds by F^R and F^N and their $DV01$s by $DV01^R$ and $DV01^N$, the regression-based hedge, characterized earlier as the $DV01$ hedge adjusted for the average change of nominal yields relative to real yields, can be written as follows:

$$F^R = -F^N \times \frac{DV01^N}{DV01^R} \times \hat{\beta} \qquad \textbf{(10.9)}$$

It turns out, however, that this regression hedge has an even stronger justification. The profit and loss (P&L) of the hedged position over a day is

$$-F^R \times \frac{DV01^R}{100} \Delta y_t^R - F^N \times \frac{DV01^N}{100} \Delta y_t^N \qquad \textbf{(10.10)}$$

Appendix A in this chapter shows that the hedge of Equation (10.9) minimizes the variance of the P&L in (10.10) over the data set shown in Figure 10-1 and used to estimate the regression parameters of Table 10-2.

In the example of this section, $F^N = -\$100$mm, $\hat{\beta} = 1.0189$, $DV01^N = .067$, and $DV01^R = .081$, so, from (10.9), as derived before, $F^R = \$84.279$mm. Because the estimated β happens to be close to one, the regression hedge of about \$84.3 million is not very different from the $DV01$ hedge of \$82.7 million calculated earlier. In fact, some practitioners would describe this hedge in terms of the $DV01$ hedge. Rearranging the terms of (10.9),

$$\frac{-F^R \times DV01^R}{F^N \times DV01^N} = \hat{\beta} = 101.89\% \qquad \textbf{(10.11)}$$

In words, the risk of the (TIPS) hedging portfolio, measured by $DV01$, is 101.89% of the risk of the underlying (nominal) position, measured by $DV01$. Alternatively, the *risk weight* of the hedge portfolio is 101.89%. This terminology does connect the hedge to the common $DV01$ benchmark but is somewhat misleading because the whole point of the regression-based hedge is that the risks of the two securities cannot properly be measured by the $DV01$ alone. It should also be noted at this point that the regression-based and $DV01$ hedges are certainly not always this close in magnitude, even in other cases of hedging TIPS *versus* nominals, as will be illustrated in the next subsection.

An advantage of the regression framework for hedging is that it automatically provides an estimate of the volatility of the hedged portfolio. To see this, substitute F^R from (10.9) into the P&L expression (10.10) and rearrange terms to get the following expression for the P&L of the hedged position:

$$-F^N \times \frac{DV01^N}{100} (\Delta y_t^N - \hat{\beta}\Delta y_t^R) \qquad \textbf{(10.12)}$$

From the definition of $\hat{\epsilon}_t$ in (10.5), the term in parentheses equals $\hat{\epsilon}_t + \hat{\alpha}$. But since $\hat{\alpha}$ is typically not very important, the standard error of the regression $\hat{\sigma}$ can be used to approximate the standard deviation of $\Delta y_t^N - \hat{\beta}\Delta y_t^R$. Hence, the standard deviation of the P&L in (10.12) is approximately

$$F^N \times \frac{DV01^N}{100} \times \hat{\sigma} \qquad \textbf{(10.13)}$$

In the present example, recalling that the standard error of the regression can be found in Table 10-2, the daily volatility of the P&L of the hedged portfolio is approximately

$$\$100\text{mm} \times \frac{.067}{100} \times 3.82 = \$255{,}940 \qquad \textbf{(10.14)}$$

The trader would have to compare this volatility with an expected gain to decide whether or not the risk-return profile of the trade is attractive.

The Stability of Regression Coefficients over Time

An important difficulty in using regression-based hedging in practice is that the hedger can never be sure that the hedge coefficient, β, is constant over time. Put another way, the errors around the regression line might be random outcomes around a stable relationship, as described

[2] If the number of observations is n, the standard error of the regression is actually defined as the square root of $\frac{\sum_t \hat{\epsilon}_t^2}{(n-2)}$. The average of the $\hat{\epsilon}_t$ in a regression with a constant is zero by construction, so the standard error of the regression differs from the standard deviation of the errors only because of the division by $n - 2$ instead of $n - 1$.

by Equation (10.3), or they might be manifestations of a changing relationship. In the former situation a hedger can safely continue to use a previously estimated $\hat{\beta}$ for hedging while, in the latter situation, the hedger should re-estimate the hedge coefficient with more recent data, if available, or with data from a past, more relevant time period. But how can the hedger know which situation prevails?

A useful start for thinking about the stability of an estimated regression coefficient is to estimate that coefficient over different periods of time and then observe if the result is stable or not. To this end, with the same data as before, Figure 10-2 graphs $\hat{\beta}$ for regressions over rolling 30-day windows. This means that the full data set of changes from August 18, 2009, to July 2, 2010, is used in 30-day increments, as follows: the first $\hat{\beta}$ comes from a regression of changes from August 18, 2009, to September 28, 2009; the second $\hat{\beta}$ from that regression from August 19, 2000, to September 29, 2009, etc.; and the last $\hat{\beta}$ from May 24, 2010, to July 2, 2010. The estimates of β in the figure certainly do vary over time, but the range of .75 to 1.29 is not extremely surprising given the previously computed 95% confidence interval with respect to β of (.8999, 1.1379). More troublesome, perhaps, is the fact that the most recent values of $\hat{\beta}$ have been trending up, which may indicate a change in regime in which even higher values of β characterize the relationship between nominal and real rates.

For a bit more perspective before closing this subsection, the period February 15, 2000, to February 15, 2002, when rates were substantially higher, was characterized by significantly higher levels of $\hat{\beta}$ and higher levels of uncertainty with respect to the regression relationship. The two bonds used in this analysis are the TIPS 4¼s of January 15, 2010, and the Treasury 6½s of February 15, 2010. Summary statistics for the regression of changes in yields of the nominal 6½s on the real 4¼s are given in Table 10-3.

Compared with Table 10-2, the estimated β here is 50% larger and the precision of this regression, measured by the R-squared or the standard error of the regression, is substantially worse. The contrast across periods again emphasizes the potential pitfalls of relying on estimated relationships persisting over time. This does not imply, of course, that blindly assuming a β of one, as in *DV*01 hedging, is a generally superior approach.

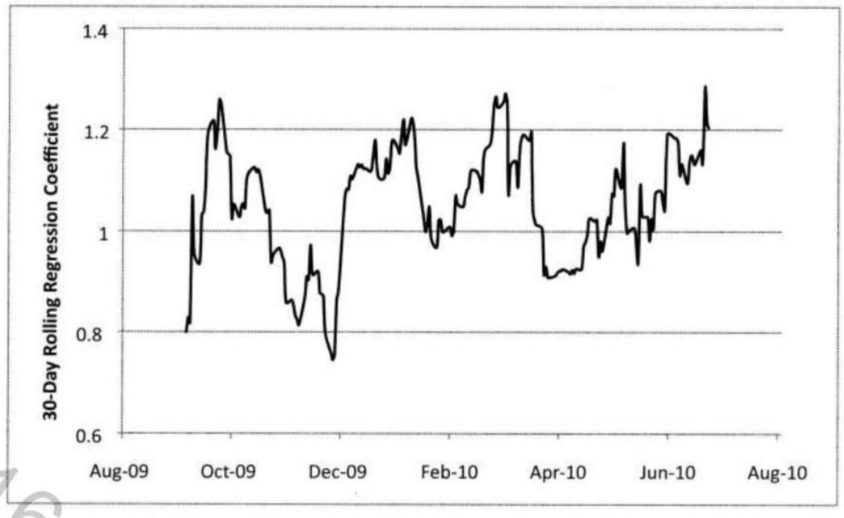


FIGURE 10-2 Rolling 30-day regression coefficient for the change in yield of the Treasury 3⅝s of August 15, 2019, on the change in yield of the TIPS 1⅞s of July 15, 2019.

TABLE 10-3 Regression Analysis of Changes in the Yield of the 6½s of February 15, 2010, on the Changes in Yield of the TIPS 4¼s of January 15, 2010, from February 15, 2000, to February 15, 2002

No. of Observations	**519**	
R-Squared	**43.0%**	
Standard Error	**4.70**	
Regression Coefficients	**Value**	**Std. Error**
Constant ($\hat{\alpha}$)	−.0267	.2067
Change in Real Yield ($\hat{\beta}$)	1.5618	.0790

TWO-VARIABLE REGRESSION-BASED HEDGING

To illustrate regression hedging with two independent variables, this section considers the case of a market maker in EUR interest rate swaps. An algebraic introduction is followed by an empirical analysis.

The market maker in question has bought or received fixed in relatively illiquid 20-year swaps from a customer and needs to hedge the resulting interest rate exposure. Immediately paying fixed or selling 20-year swaps would sacrifice too much if not all of the spread paid by the customer, so the market maker chooses instead to sell a combination of 10- and 30-year swaps. Furthermore, the market maker is willing to rely on a two-variable regression model to describe the relationship between changes in 20-year swap rates and changes in 10- and 30-year swap rates:

$$\Delta y_t^{20} = \alpha + \beta^{10}\Delta y_t^{10} + \beta^{30}\Delta y_t^{30} + \epsilon_t \quad \textbf{(10.15)}$$

Equation (10.15) can be estimated by least squares, analogously to the single-variable case, by minimizing

$$\sum_t (\Delta y_t^{20} - \hat{\alpha} - \hat{\beta}^{10}\Delta y_t^{10} - \hat{\beta}^{30}\Delta y_t^{30})^2 \quad \textbf{(10.16)}$$

with respect to the parameters $\hat{\alpha}$, $\hat{\beta}^{10}$ and $\hat{\beta}^{30}$. The estimation of these parameters then provides a predicted change for the 20-year swap rate:

$$\Delta \hat{y}_t^{20} = \hat{\alpha} + \hat{\beta}^{10}\Delta y_t^{10} + \hat{\beta}^{30}\Delta y_t^{30} \quad \textbf{(10.17)}$$

To derive the notional face amount of the 10- and 30-year swaps, F^{10} and F^{30}, respectively, required to hedge F^{20} face amount of the 20-year swaps, generalize the reasoning given in the single-variable case as follows. Write the P&L of the hedged position as

$$-F^{20}\frac{DV01^{20}}{100}\Delta y_t^{20} - F^{10}\frac{DV01^{10}}{100}\Delta y_t^{10} - F^{30}\frac{DV01^{30}}{100}\Delta y_t^{30} \quad \textbf{(10.18)}$$

Then substitute the predicted change in the 20-year rate from (10.17) into (10.18), retaining only the terms depending on Δy_t^{10} and Δy_t^{30}, to obtain

$$\left[-F^{20}\frac{DV01^{20}}{100}\hat{\beta}^{10} - F^{10}\frac{DV01^{10}}{100}\right]\Delta y_t^{10} + \left[-F^{20}\frac{DV01^{20}}{100}\hat{\beta}^{30} - F^{30}\frac{DV01^{30}}{100}\right]\Delta y_t^{30} \quad \textbf{(10.19)}$$

Finally, choose F^{10} and F^{30} to set the terms in brackets equal to zero, i.e., to eliminate the dependence of the predicted P&L on changes in the 10- and 30-year rates. This leads to two equations with the following solutions:

$$F^{10} = -F^{20}\frac{DV01^{20}}{DV01^{10}}\hat{\beta}^{10} \quad \textbf{(10.20)}$$

$$F^{30} = -F^{20}\frac{DV01^{20}}{DV01^{30}}\hat{\beta}^{30} \quad \textbf{(10.21)}$$

As in the single-variable case, this 10s-30s hedge of the 20-year can be expressed in terms of risk weights. More specifically, the *DV*01 risk in the 10-year part of the hedge and the *DV*01 risk in the 30-year part of the hedge can both be expressed as a fraction of the *DV*01 risk of the 20-year. Mathematically, these risk weights can be found by rearranging (10.20) and (10.21):

$$\frac{-F^{10}\times DV01^{10}}{F^{20}\times DV01^{20}} = \hat{\beta}^{10} \quad \textbf{(10.22)}$$

$$\frac{-F^{30}\times DV01^{30}}{F^{20}\times DV01^{20}} = \hat{\beta}^{30} \quad \textbf{(10.23)}$$

Proceeding now to the empirical analysis, the market maker, as of July 2006, performs an initial regression analysis using data on changes in the 10-, 20-, and 30-year EUR swap rates from July 2, 2001, to July 3, 2006. Summary statistics for the regression of changes in the 20-year EUR swap rate on changes in the 10- and 30-year EUR swap rates are given in Table 10-4. The statistical quality of these results, characteristic of all regressions of like rates, are far superior to those of the nominal against real yields of the previous section: the R-squared or percent variance explained by the regression is 99.8%; the standard error of the regression is only .14 basis points; and the 95% confidence intervals with respect to the two coefficients are extremely narrow, i.e., (.2153, 2289) for the 10-year and (.7691, .7839) for the 30-year. Lastly, in a result similar to those of the regressions of the previous section, the constant is insignificantly different from zero.

Applying the risk-weight interpretation of the regression coefficients given in Equations (10.22) and (10.23), the results in Table 10-4 say that 22.21% of the *DV*01 of the 20-year swap should be hedged with a 10-year swap and 77.65% with a 30-year swap. The sum of these weights,

TABLE 10-4 Regression Analysis of Changes in the Yield of the 20-Year EUR Swap Rate on Changes in the 10- and 30-Year EUR Swap Rates from July 2, 2001, to July 3, 2006

No. of Observations	**1281**	
R-Squared	**99.8%**	
Standard Error	**.14**	
Regression Coefficients	**Value**	**Std. Error**
Constant ($\hat{\alpha}$)	−.0014	.0040
Change in 10-Year Swap Rate ($\hat{\beta}^{10}$)	.2221	.0034
Change in 30-Year Swap Rate ($\hat{\beta}^{30}$)	.7765	.0037

99.86%, happens to be very close to one, meaning that the *DV*01 of the regression hedge very nearly matches the *DV*01 of the 20-year swap, although this certainly need not be the case: minimizing the variance of the P&L of a hedged position, when rates are not assumed to move in parallel, need not result in a *DV*01-neutral portfolio.

Tight as the in-sample regression relationship seems to be, the real test of the hedge is whether it works out-of-sample.[3] To this end, Figure 10-3 tracks the errors of the hedge over time. All of these errors are computed as the realized change in the 20-year yield minus the predicted change for that yield based on the estimated regression in Table 10-4:

$$\hat{\epsilon}_t = \Delta y_t^{20} - (-.0014 + .2221\Delta y_t^{10} + .7765\Delta y_t^{30}) \quad \textbf{(10.24)}$$

The errors to the left of the vertical dotted line are in-sample in that the same Δy_t^{20} used to compute $\hat{\epsilon}_t$ in (10.24) were also used to compute the coefficient estimates −.0014, .2221, and .7765. In other words, it is not that surprising that the $\hat{\epsilon}_t$ to the left of the dotted line are small because the regression coefficients were estimated to minimize the sum of squares of these errors. By contrast, the errors to the right of the dotted line are out-of-sample: these $\hat{\epsilon}_t$ are computed from realizations of Δy_t^{20} after July 3, 2006, but using the regression coefficients estimated over the period from July 2, 2001, to July 3, 2006. It is, therefore, the size and behavior of these out-of-sample errors that provide evidence as to the stability of the estimated coefficients over time.

From inspection of Figure 10-3 the out-of-sample errors are indeed small, for the most part, until August and September 2008, a peak in the financial crisis of 2007–2009. After then the daily errors ran as high as about four basis points and as low as about −5.3 basis points. And while the accuracy of the relationship seems to have recovered somewhat to the far right-end of the graph, by the summer of 2009, the errors there are not nearly so well behaved as at the start of the out-of-sample period.

It is obvious and easy to say that the market maker, during the turbulence of a financial crisis, should have replaced the regression of Table 10-4 and the resulting hedging rule. But replace these with what? What does the market maker do at that time, before there exist sufficient post-crisis data points? And what does the market maker do after the worst of the crisis: estimate a regression from data during the crisis or revert to some earlier, more stable period? These are the kinds of issues that make regression hedging an art rather than a science. In any case, it should again be emphasized that avoiding these issues by blindly resorting to a one-security *DV*01 hedge, or a two-security *DV*01 hedge with arbitrarily assigned risk weights, like 50%-50%, is even less satisfying.

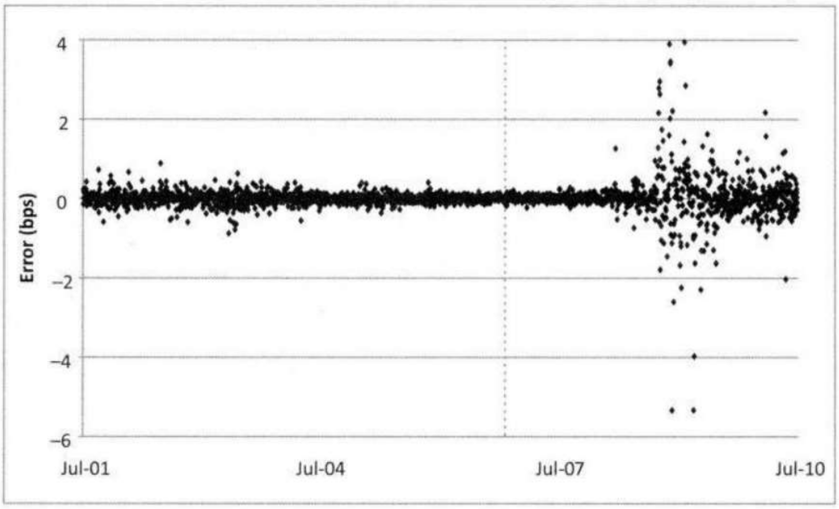


FIGURE 10-3 In- and out-of-sample errors for a regression of changes of 20-year and 10- and 30-year EUR swap rates with estimation period July 2, 2001, to July 3, 2006.

LEVEL VERSUS CHANGE REGRESSIONS

When estimating regression-based hedges, some practitioners regress changes in yields on changes in yields, as in the previous sections, while others prefer to regress yields on yields. Mathematically, in the single-variable case, the level-on-level regression with dependent variable *y* and independent variable *x* is

$$y_t = \alpha + \beta x_t + \epsilon_t \quad \textbf{(10.25)}$$

while the change-on-change regression is[4]

$$y_t - y_{t-1} = \Delta y_t = \beta \Delta x_t + \Delta\epsilon_t \quad \textbf{(10.26)}$$

[3] The phrase *in-sample* refers to behavior within the period of estimation, in this case July 2, 2001, to July 3, 2006. The phrase *out-of-sample* refers to behavior outside the period of estimation, usually after but possibly before that period as well.

[4] It is usual to include a constant term in the change-on-change regression, but for the purposes of this section, to maintain consistency across the two specifications, this constant term is omitted.

By theory that is beyond the scope of this book, if the error terms ϵ_t are independently and identically distributed random variables with mean zero and are uncorrelated with the independent variable, then so are the $\Delta\epsilon_t$, and least squares on either (10.25) or (10.26) will result in coefficient estimators that are *unbiased*,[5] *consistent*,[6] and *efficient*, i.e., of *minimum variance*, in the class of linear estimators. If the error terms of either specification are not independent of each other, however, then the least-squares coefficients of that specification are not necessarily efficient, but retain their unbiasedness and consistency.

To illustrate the economics behind the assumption that error terms are independent of each other, say that $\alpha = 0$, that $\beta = 1$, that y is the yield on a coupon bond, and that x is the yield on another, near-maturity coupon bond. Say further that the yield on the x-bond was 5% yesterday and 5% again today while the yield on the y-bond was 1% yesterday. Because the yield on the x-bond is 5% today, the level Equation (10.25) predicts that the yield on the y-bond will be 5% today, despite its being 1% yesterday. But if the market yield was so far off yesterday's prediction, with a realized error of −4%, then it is more likely that the error today will be not far from −4% and that the yield of the y-bond yield will be closer to 1% than the 5% predicted by (10.25). Put another way, the errors in (10.25) are not likely to be independent of each other, as assumed, but rather persistent, or correlated over time.

The change regression (10.26) assumes the opposite extreme with respect to the errors, i.e., that they are completely persistent. Continuing with the example of the previous paragraph, with the yield on the y-bond at 1% yesterday and the yield on the x-bond unchanged from yesterday, the change regression predicts that y-bond will remain at 1%. But, as reasoned above, it is more likely that the y-bond yield will move some of the way back from 1% to 5%. Hence, the error terms in (10.26) are also unlikely to be independent of each other.

The first lesson to be drawn from this discussion is that because the error terms in both (10.26) and (10.25) are likely to be correlated over time, i.e., *serially correlated*, their estimated coefficients are not efficient. But, with nothing to gainsay the validity of the other assumptions concerning the error terms, the estimated coefficients of both the level and change specifications are still unbiased and consistent.

The second lesson to be drawn from the discussion of this section is that there is a more sensible way to model the relationship between two bond yields than either (10.26) or (10.25). In particular, model the behavior that the y-bond's yield will, on average, move somewhat closer from 1% to 5%. Mathematically, assume (10.25) with the error dynamics

$$\epsilon_t = \rho\epsilon_{t-1} + \nu_t \quad \textbf{(10.27)}$$

for some constant $\rho < 1$. Assumption (10.27) says that today's error consists of some portion of yesterday's error plus a new random fluctuation. In terms of the numerical example, if $\rho = 75\%$, then yesterday's error of −4% would generate an average error today of 75% × −4% or −3% and, therefore, an expected y-bond yield of 5% − 3% or 2%. In this way the error structure (10.27) has the yield of the y-bond converging to its predicted value of 5% given the yield of the x-bond at 5%. While beyond the scope of this book, the procedure for estimating (10.25) with the error structure (10.27) is presented in many statistical texts.

PRINCIPAL COMPONENTS ANALYSIS

Overview

Regression analysis tries to explain the changes in the yield of one bond relative to changes in the yields of a small number of other bonds. It is often useful, however, to have a single, empirical description of the behavior of the term structure that can be applied across all bonds. *Principal Components* (PCs) provide such an empirical description.

To fix ideas, consider the set of swap rates from 1 to 30 years at annual maturities. One way to describe the time series fluctuations of these rates is through the variances of the rates and their pairwise covariances or correlations. Another way to describe the data, however, is to create 30 interest rate factors or components, where each factor describes a change in each of the 30 rates. So, for example, one factor might be a simultaneous change of 5 basis points in the l-year rate, 4.9 basis points in the 2-year rate, 4.8 basis points in the 3-year rate, etc. Principal Components Analysis (PCA) sets up these 30 such factors with the following properties:

1. The sum of the variances of the PCs equals the sum of the variances of the individual rates. In this sense the PCs capture the volatility of this set of interest rates.

[5] An unbiased estimator of a parameter is such that its expectation equals the true value of that parameter.

[6] A consistent estimator of a parameter, with enough data, becomes arbitrarily close to the true value of the parameter.

2. The PCs are uncorrelated with each other. While changes in the individual rates are, of course, highly correlated with each other, the PCs are constructed so that they are uncorrelated.
3. Subject to these two properties or constraints, each PC is chosen to have the maximum possible variance given all earlier PCs. In other words, the first PC explains the largest fraction of the sum of the variances of the rates; the second PC explains the next largest fraction, etc.

PCs of rates are particularly useful because of an empirical regularity: the sum of the variances of the first three PCs is usually quite close to the sum of variances of all the rates. Hence, rather than describing movements in the term structure by describing the variance of each rate and all pairs of correlation, one can simply describe the structure and volatility of each of only three PCs.

The next subsections illustrate PCs and their uses in the context of USD and then global swap markets. For interested readers, Appendix B in this chapter describes the construction of PCs with slightly more mathematical detail, using the simpler context of three interest rates and three PCs. Fully general and more mathematical descriptions are available in numerous other books and articles.

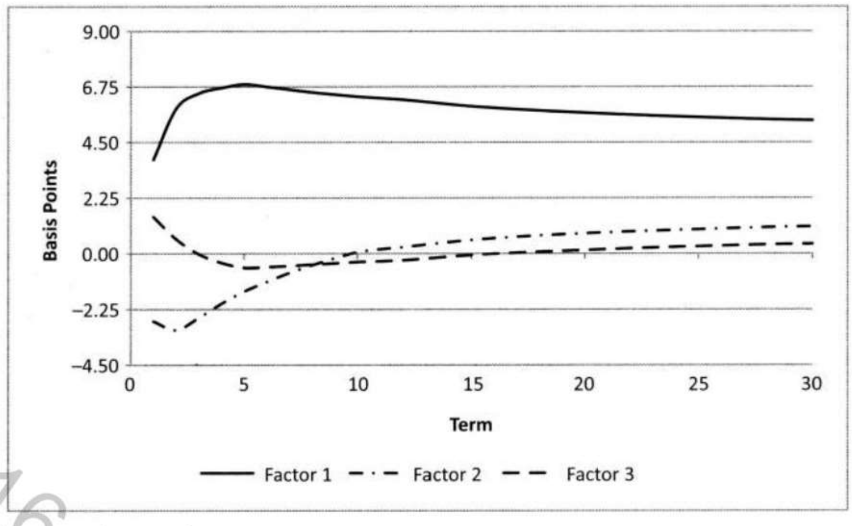


FIGURE 10-4 The first three principal components from USD swap rates from October 2001 to October 2008.

PCAs for USD Swap Rates

Figure 10-4 graphs the first three principal components from daily data on USD swap rates while Table 10-5 provides a selection of the same information in tabular form. Thirty different data series are used, one series for each

TABLE 10-5 Selected Results of Principal Components for the USD Swap Curve from October 1, 2001, to October 2, 2008. Units are basis points or percentages.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	PCs					% of PC Variances			
Term	Level	Slope	Short Rate	PC Vol	Total Vol	Level	Slope	Short Rate	PC Vol/ Total Vol (%)
1	3.80	−2.74	1.48	4.91	4.96	59.8	31.0	9.1	99.05
2	5.86	−3.09	0.59	6.65	6.67	77.7	21.5	0.8	99.74
5	6.85	−1.53	−0.57	7.04	7.06	94.7	4.7	0.7	99.85
10	6.35	0.06	−0.34	6.36	6.37	99.7	0.0	0.3	99.83
20	5.69	0.82	0.14	5.75	5.75	97.9	2.0	0.1	99.95
30	5.38	1.09	0.39	5.51	5.52	95.6	3.9	0.5	99.79
Total	32.47	6.74	2.28	33.25	33.29	95.4	4.1	0.5	99.87

annual maturity from one to 30 years, and the observation period spans from October 2001 to October 2008. (Data from more recent dates will be presented and discussed later in this section.)

Columns (2) to (4) in Table 10-5 correspond to the three PC curves in Figure 10-4. These components can be interpreted as follows. A one standard-deviation increase in the "Level" PC, given in column (2), is a simultaneous 3.80 basis-point increase in the one-year swap rate, a 5.86 basis-point increase in the 2-year, etc., and a 5.38 basis-point increase in the 30-year. This PC is said to represent a "level" change in rates because rates of all maturities move up or down together by, very roughly, the same amount. A one standard-deviation increase in the "Slope" PC, given in column (3), is a simultaneous 2.74 basis-point *drop* in the 1-year rate, a 3.09 basis-point drop in the 2-year rate, etc., and a 6.74 basis-point *increase* in the 30-year rate. This PC is said to represent a "slope" change in rates because short-term rates fall while longer-term rates increase, or *vice versa*. Finally, a one standard-deviation increase in the "Short Rate" PC, given in column (4), is made up of simultaneous increases in short-term rates (e.g., one- and two-year terms), small decreases in intermediate-term rates (e.g., 5- and 10-year terms), and small increases in long-term rates (e.g., 20- and 30-year terms). While this PC is often called a "curvature" change, because intermediate-term rates move in the opposite direction from short- and long-term rates, the short-term rates moves dominate. Hence, the third PC is interpreted here as an additional factor to describe movements in short-term rates.

One feature of the shape of the level PC warrants additional discussion. Short-term rates might be expected to be more volatile than longer-term rates because changes in short-term rates are determined by current economic conditions, which are relatively volatile, while longer-term rates are determined mostly by expectations of future economic conditions, which are relatively less volatile. But since the Board of Governors of the Federal Reserve System, like many other central banks, anchors the very short-term rate at some desired level, the volatility of very short-term rates is significantly dampened. The level factor, which, as will be discussed shortly, explains the vast majority of term structure movements, and reflects this behavior on the part of central banks: very short-term rates move relatively little. Then, at longer maturities, the original effect prevails and longer-term rates move less than intermediate and shorter-term rates.

Column (5) of Table 10-5 gives the combined standard deviation or volatility of the three principal components for a given rate, and column (6) gives the total or empirical volatility of that rate. For the one-year rate, for example, recalling that the principal components are uncorrelated, the combined volatility, in basis points, from the three components is

$$\sqrt{3.80^2 + (-2.74)^2 + 1.48^2} = 4.91 \qquad \textbf{(10.28)}$$

The total or empirical volatility of the one-year rate, however, computed directly from the time series data, is 4.96 basis points. Column (10) of the table gives the ratio of columns (5) and (6), which, for the 1-year rate is 4.9099/4.9572 or 99.05%. (For readability, many of the entries of Table 10-5 are rounded although calculations are carried out to higher precision.)

Columns (7) through (9) of Table 10-5 give the ratios of the variance of each PC component to the total PC variance. For the 1-year rate, these ratios are $3.80^2/4.91^2 = 59.9\%$; $(-2.74)^2/4.91^2 = 31.1\%$; and $1.48^2/4.91^2 = 9.1\%$.

Finally, the last row of the table gives statistics on the square root of the sum of the variances across rates of different maturities. The sum of the variances is not a particularly interesting economic quantity—it does not, for example, represent the variance of any interesting portfolio—but, as mentioned in the overview of PCA, this sum is used to ensure that the PCs capture all of the volatility of the underlying interest rate series.

Having explained the calculations of Figure 10-4 and Table 10-5, the text can turn to interpretation. First and foremost, column (10) of Table 10-5 shows that, for rates of all maturities, the three principal components explain over 99% of rate volatility. And, across all rates, the three PCs explain 99.87% of the sum of the variability of these rates. While these findings represent relatively recent data on U.S. swap rates, similarly high explanatory powers characterize the first three components of other kinds of rates, like U.S. government bond yields and rates in fixed income markets in other countries. These results provide a great deal of comfort to hedgers: while in theory many factors (and, therefore, securities) might be required to hedge the interest rate risk of a particular portfolio, in practice, three factors cover the vast majority of the risk.

Columns (7) through (9) of Table 10-5 show that the level component is far and away the most important in explaining the volatility of the term structure. The construction of principal components, described in the overview, does ensure that the first component is the most important component, but the extreme dominance of this component is a feature of the data. This finding is useful for thinking about the costs and benefits of adding a second or third factor to a one-factor hedging framework. Interestingly too, the dominance of the first factor is significantly muted in the very short end of the curve. This implies that hedging one short-term bond with another will not be so effective as hedging one longer-term bond with another. Or, put another way, relatively more factors or hedging securities are needed to hedge portfolios that are concentrated at the short end of the curve. This makes intuitive sense in the context of the extensive information market participants have about near-term events and their effects on rates relative to the information they have on events further into the future.

Hedging with PCA and an Application to Butterfly Weights

A PCA-based hedge for a portfolio would proceed along the lines of the multi-factor approaches. Start with the current price of the portfolio under the current term structure. Then, shift each principal component in turn to obtain new term structures and new portfolio prices. Next, calculate an '01 with respect to each principal component using the difference between the respective shifted price and the original price. Finally, using these portfolio '01s and analogously constructed '01s for a chosen set of hedging securities, find the portfolio of hedging securities that neutralizes the risk of the portfolio to the movement of each PC.

PCA is particularly useful for constructing empirically-based hedges for large portfolios; it is impractical to perform and assess individual regressions for every security in a large portfolio. For illustration purposes, however, this subsection will illustrate how PCA is used, in practice, to hedge a *butterfly* trade. Most typically, butterfly trades use three securities and either buy the security of intermediate maturity and short the *wings* or short the intermediate security and buy the wings.

To take a relatively common butterfly, consider a trader who believes that the 5-year swap rate is too high relative to the 2- and 10-year swap rates and is, therefore, planning to receive in the 5-year and pay in the 2- and 10-year. As of May 28, 2010, the par swap rates and *DV*01s of the swaps of relevant terms are listed in Table 10-6. (The 30-year data will be used shortly.) To calculate the PCA

TABLE 10-6 Par Swap Rates and *DV*01s as of May 28, 2010

Term	Rate	*DV01*
2	1.235%	.0197
5	2.427%	.0468
10	3.388%	.0842
30	4.032%	.1731

hedge ratios, assume that the trader will receive on 100 notional amount of 5-year swaps and will trade F^2 and F^{10} notional amount of 2- and 10-year swaps. Using the data from Tables 10-5 and 10-6, the equation that neutralizes the overall portfolio's exposure to the level PC is

$$-F^2\frac{.0197}{100}\times 5.86 - F^{10}\frac{.0842}{100}\times 6.35 - 100\times\frac{.0468}{100}\times 6.85 = 0 \quad \textbf{(10.29)}$$

Similarly, the equation that neutralizes the overall exposure to the slope PC is

$$-F^2\frac{.0197}{100}\times(-3.09) - F^{10}\frac{.0842}{100}\times .06 - 100 \times\frac{.0468}{100}\times(-1.53) = 0 \quad \textbf{(10.30)}$$

Solving, $F^2 = -120.26$ and $F^{10} = -34.06$ or, in terms of risk weights relative to the *DV*01 of the five-year swap,

$$\frac{120.26\times\frac{.0197}{100}}{.0468} = 50.6\% \quad \textbf{(10.31)}$$

$$\frac{34.06\times\frac{.0842}{100}}{.0468} = 61.3\% \quad \textbf{(10.32)}$$

In words, the *DV*01 of the five-year swap is hedged 50.6% by the two-year swap and 61.3% by the 10-year swap. Note that the sum of the risk weights is not 100%: the hedge neutralizes exposures to the level and slope PCs, not exposures to parallel shifts. To the extent that the term structure changes as assumed, i.e., as some combination of the first two PCs, then the hedge will work exactly. On the other hand, to the extent that the actual change deviates from a combination of these two PCs, the hedge will not, *ex post*, have fully hedged interest rate risk.

Hedging the interest rate risk of the five-year swap with two other swaps in not uncommon, a practice supported by the large fraction of rate variance explained by the first two PCs. A trader might also decide, however, to hedge the third PC as well. A hedge against the first three PCs, found by generalizing the two-security hedge just discussed, gives rise to risk weights of 28.1%, 139.1%, and −67.4% in the 2-, 10-, and 30-year swaps, respectively, i.e., pay in the 2- and 10-year, but receive in the 30-year.

Is hedging the third PC worthwhile? The answer depends on the trader's risk preferences, but the following analysis is useful. Say that the trader hedges the first two components alone and then the third component experiences a one standard-deviation decrease. The P&L of the trade, per 100 face amount of the 5-year swap, would be

$$\left[-120.26\times\frac{.0197}{100}\times.59+100\times\frac{.0468}{100}\times(-.57)\right.$$
$$\left.-34.06\times\frac{.0842}{100}\times(-.34)\right]=-.031 \quad \textbf{(10.33)}$$

or, for a two standard-deviation move, a loss of a bit more than 6 cents per 100 face amount of the 5-year swap. As these two standard deviations of short rate risk equates to not even 1.5 basis points of convergence of the 5-year swap, a trader might very well not bother with this third leg of the hedge.

Principal Component Analysis of EUR, GBP, and JPY Swap Rates

Figures 10-5 to 10-7 show the first three PCs for the EUR, GBP, and JPY swap rate curves over the same sample period as the USD PCs in Figure 10-4. The striking fact about these graphs is that the shape of the PCs are very much the same across USD, EUR, and GBP. The only significant difference is in magnitudes, with the USD level component entailing larger-sized moves than the level components of EUR and GBP. The PCs of the JPY curve are certainly similar to those of these other countries, but the level component in JPY does not have the same hump: in JPY the first PC does not peak at the five-year maturity point as do the other curves, but increases monotonically with maturity before ultimately leveling off. The significance of this difference in shape will be discussed in the next subsection.

The Shape of PCs over Time

As with any empirically based hedging methodology, a decision has to be made about the relevant time period over which to estimate parameters. This is an issue for regression-based methods, as discussed in this chapter, and it is no less an issue for PCA. As will be discussed in this subsection, the qualitative shapes of PCs have, until very recently, remained remarkably stable. This does not

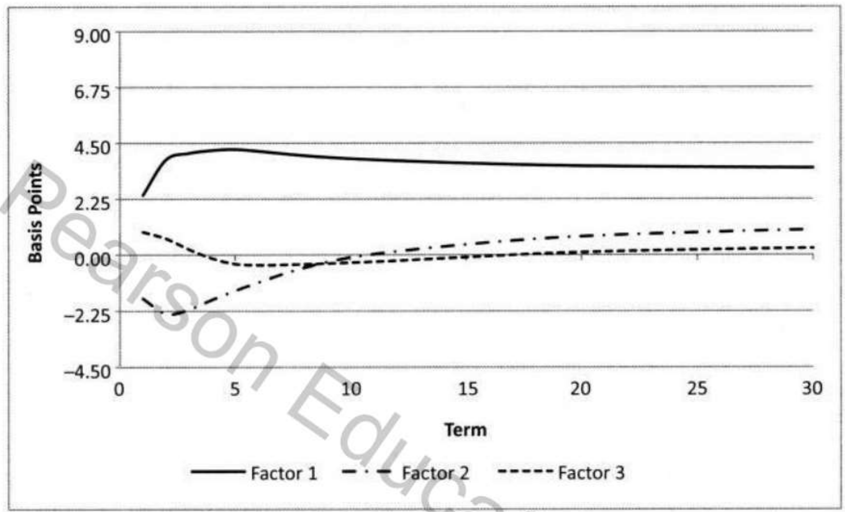


FIGURE 10-5 The first three principal components from EUR swap rates from October 2001 to October 2008.

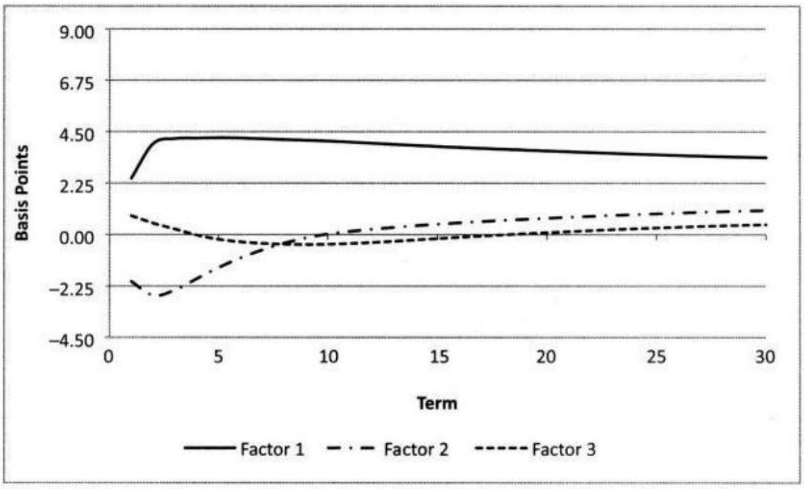


FIGURE 10-6 The first three principal components from GBP swap rates from October 2001 to October 2008.

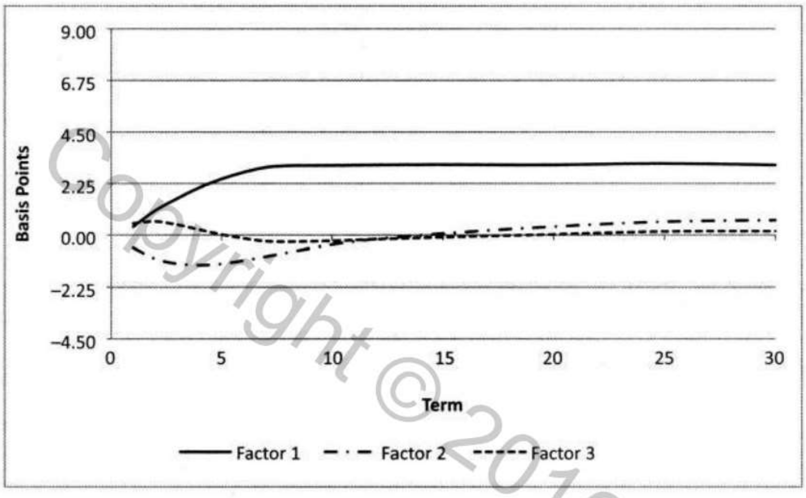


FIGURE 10-7 The first three principal components from JPY swap rates from October 2001 to October 2008.

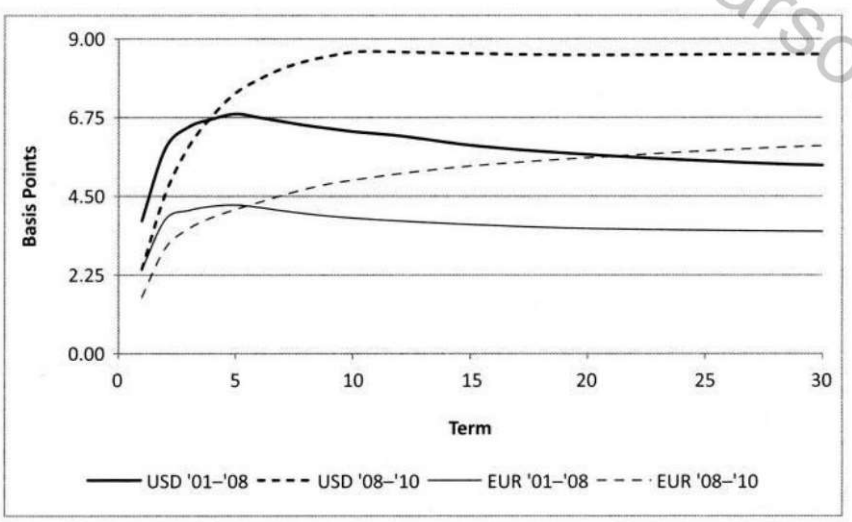


FIGURE 10-8 The first principal component in USD and EUR swap rates estimated from October 2001 to October 2008 and from October 2008 to October 2010.

imply, however, that differences in PCs estimated over different time periods can be ignored in the sense that they have no important effects on the quality of hedges. But having made this point, the text focuses on the relatively recent changes in the shapes of PCs around the world.

Figure 10-4 showed the first three USD PCs computed over the period 2001 to 2008, but, for quite some time, the qualitative shapes of these PCs was pretty much the same.[7] The volatility of rates has changed over time, and with it the magnitude or height of the PC curves, but the qualitative shapes have not changed much. Most recently, however, there has been a qualitative change to the shape of the first PC in USD, EUR, and GBP. In fact, these shapes have become more like the past shape of the first PC in JPY!

Figures 10-8 and 10-9 contrast the level PC over the historical period October 2001 to October 2008 with that of the post-crisis period, October 2008 to October 2010. Figure 10-8 makes the comparison for USD and EUR while Figure 10-9 does the same for GBP and JPY. The historical maximum of the level PC at a term of about five years in USD, EUR, and GBP has been pushed out dramatically to 10 years and beyond. In fact, these shapes now more closely resemble the level PC of JPY over the earlier estimation period. One explanation for this is the increasing certainty that central banks will maintain easy monetary conditions and low rates for an extended period of time. This dampens the volatility of short- and intermediate-term rates relative to that of longer-term rates, lowers the absolute volatility of short-term rates, and increases the volatility of long-term rates, reflecting the uncertainty of the ultimate results of central bank policy. Meanwhile, the level PC for JPY in the most recent period has become even more pronouncedly upward-sloping, consistent with an even longer period of central-bank control over the short-term rate.

[7] See, for example, Figure 2 of Bulent Baygun, Janet Showers, and George Cherpelis, Salomon Smith Barney, "Principles of Principal Components," January 31, 2000. The shapes of the three PCs in that graph, covering the period from January 1989 to February 1998, are qualitatively extremely similar to those of Figure 10-4 in this chapter.

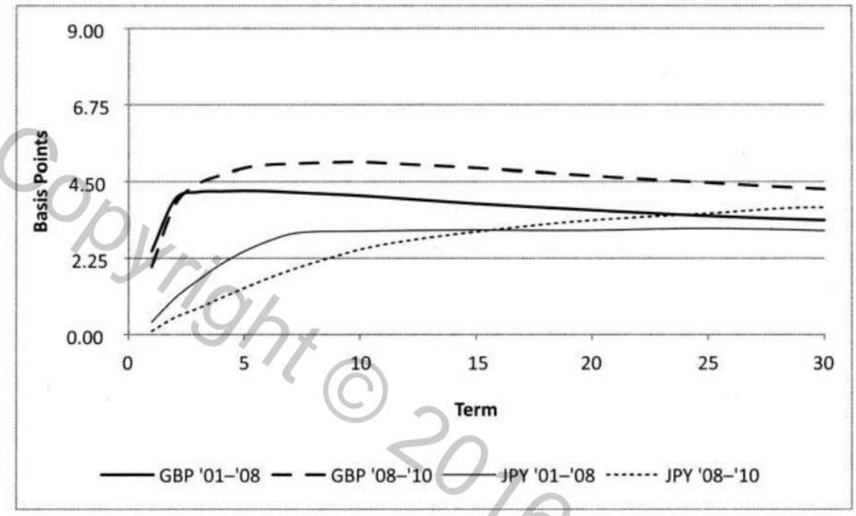


FIGURE 10-9 The first principal component in GBP and JPY swap rates estimated from October 2001 to October 2008 and from October 2008 to October 2010.

APPENDIX A

The Least-Squares Hedge Minimizes the Variance of the P&L of the Hedged Position

The P&L of the hedged position, given in (10.10) and repeated here, is

$$-F^R \times \frac{DV01^R}{100}\Delta y_t^R - F^N \times \frac{DV01^N}{100}\Delta y_t^N \quad \textbf{(10.34)}$$

Let $V(\cdot)$ and $Cov(\cdot,\cdot)$ denote the variance and covariance functions. The variance of the P&L expression in (10.34) is

$$\left(F^R \times \frac{DV01^R}{100}\right)^2 V(\Delta y_t^R) + \left(F^N \times \frac{DV01^N}{100}\right)^2 V(\Delta y_t^N)$$
$$+2\left(F^R \times \frac{DV01^R}{100}\right)\left(F^N \times \frac{DV01^N}{100}\right) Cov(\Delta y_t^R, \Delta y_t^N) \quad \textbf{(10.35)}$$

To find the face amount F^R that minimizes this variance, differentiate (10.35) with respect to F^R and set the result to zero:

$$2F^R\left(\frac{DV01^R}{100}\right)^2 V(\Delta y_t^R) + 2F^N \frac{DV01^R}{100}\frac{DV01^N}{100} Cov(\Delta y_t^R, \Delta y_t^N) = 0 \quad \textbf{(10.36)}$$

Then, rearranging terms,

$$F^N \times DV01^N \times \frac{Cov(\Delta y_t^R, \Delta y_t^N)}{V(\Delta y_t^R)} = -F^R \times DV01^R \quad \textbf{(10.37)}$$

But, by the properties of least squares, not derived in this text,

$$\hat{\beta} \equiv \frac{Cov(\Delta y_t^R, \Delta y_t^N)}{V(\Delta y_t^R)} \quad \textbf{(10.38)}$$

Therefore, substituting (10.38) into (10.37) gives the regression hedging rule (10.9) of the text.

APPENDIX B

Constructing Principal Components from Three Rates

The goal of this appendix is to demonstrate the construction and properties of PCs with a minimum of mathematics. To this end, consider three swap rates, the 10-year, 20-year, and 30-year. Over some sample period, the volatilities of these rates, in basis points per day, are 4.25, 4.20, and 4.15. Furthermore, the correlations among these rates are given in the correlation matrix of Table 10-7.

The combination of data on volatilities and correlations are usefully combined into a *variance-covariance* matrix, denoted by V, where the element in the i^{th} row and j^{th} column gives the covariance of the rate of term i with the rate of term j, or, the correlation of i and j times the standard deviation of i times the standard deviation of j. For example, the covariance of the 20-year swap rate with the 30-year swap rate is .99 × 4.20 × 4.15, or 17.26. The variance-covariance matrix for the example of this appendix is

$$\mathbf{V} = \begin{pmatrix} 18.06 & 16.96 & 15.87 \\ 16.96 & 17.64 & 17.26 \\ 15.87 & 17.26 & 17.22 \end{pmatrix} \quad \textbf{(10.39)}$$

One use of a variance-covariance matrix is to write succinctly the variance of a particular portfolio of the relevant

TABLE 10-7 Correlation Matrix for Swap Rate Example

Term	10-Year	20-Year	30-Year
10-Year	1.00	0.95	0.90
20-Year	0.95	1.00	0.99
30-Year	0.90	0.99	1.00

securities. Consider a portfolio with a total *DV*01 of .50 in the 10-year swap, −1.0 in the 20-year swap, and .60 in the 30-year swap. Without matrix notation, then, the dollar variance of the portfolio, denoted by σ^2 would be given by

$$\begin{aligned}\sigma^2 &= .5^2 4.25^2 + (-1)^2\, 4.20^2 + .6^2 4.15^2\\ &\quad + 2 \times .5 \times (-1) \times .95 \times 4.25 \times 4.20\\ &\quad + 2 \times .5 \times .6 \times .90 \times 4.25 \times 4.15\\ &\quad + 2 \times (-1) \times .6 \times .99 \times 4.20 \times 4.15\\ &= .464^2 \end{aligned} \tag{10.40}$$

With matrix notation, letting the transpose of the vector w be w′ = (.5, −1, .6), the dollar variance of the portfolio is given more compactly by

$$\mathbf{w'Vw} = (.5 \quad -1 \quad .6)\begin{pmatrix} 18.06 & 16.96 & 15.87 \\ 16.96 & 17.64 & 17.26 \\ 15.87 & 17.26 & 17.22 \end{pmatrix}\begin{pmatrix} .5 \\ -1 \\ .6 \end{pmatrix} \tag{10.41}$$

Finally, note that the sum of the variances of the rates is $4.25^2 + 4.20^2 + 4.15^2 = 52.925$, or, for a measure of total volatility, take the square root of that sum to get 7.27 basis points.

Returning now to principal components, the idea is to create three factors that capture the same information as the variance-covariance matrix. The procedure is as follows. Denote the first principal component by the vector $\mathbf{a} = (a_1, a_2, a_3)'$. Then find the elements of this vector by maximizing $\mathbf{a'Va}$ such that $\mathbf{a'a} = \mathbf{1}$. As mentioned in the PCA overview, this maximization ensures that, among the three PCs to be found, the first PC explains the largest fraction of the variance. The constraint, $\mathbf{a'a} = \mathbf{1}$, along with a similar constraint placed on the other PCs, will ensure that the total variance of the PCs equals the total variance of the underlying data. Performing this maximization, which can be done with the solver in Excel, $\mathbf{a} = (.5758, .5866, .5696)$. Note that the variance of this first component is $\mathbf{a'Va} = 51.041$ which is 51.041/52.925 or 96.44% of the total variance of the rates.

The second principal component, denoted by the vector $\mathbf{b} = (b_1, b_2, b_3)$ is found by maximizing $\mathbf{b'Vb}$ such that $\mathbf{b'b} = \mathbf{1}$ and $\mathbf{b'a} = \mathbf{0}$. The maximization and the first constraint are analogous to those for finding the first principal component. The second constraint requires that the PC **b** is uncorrelated with the first PC, **a**. Solving, gives $\mathbf{b} = (-.7815, .1902, .5941)$. Note that $\mathbf{b'Vb} = 1.867$ which explains 1.867/52.925 or 3.53% of the total variance of the rates.

Finally, the third PC, denoted by $\mathbf{c} = (c_1, c_2, c_3)$ is found by solving the three equations, $\mathbf{c'c} = \mathbf{1}$; $\mathbf{c'a} = \mathbf{0}$; and $\mathbf{c'b} = \mathbf{0}$. The solution is $\mathbf{c} = (.2402, -.7872, .5680)$.

As will be clear in a moment, it turns out to be more intuitive to work with a different scaling of the PCs, namely, by multiplying each by its volatility. In the example, this means multiplying the first PC by $\sqrt{51.041}$ or 7.14; the second PC by $\sqrt{1.867}$ or 1.37; and the third by $\sqrt{.017}$ or .13. This gives the PCs, to be denoted $\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\mathbf{c}}$, as recorded in Table 10-8.

TABLE 10-8 Transformed PCs for the Swap Rate Example

Term	1st PC	2nd PC	3rd PC
10-Year	4.114	−1.068	.032
20-Year	4.191	.260	−.103
30-Year	4.069	.812	.075

Under this scaling the PCs have a very intuitive interpretation: a one standard-deviation increase of the first PC or factor is a 4.114 basis-point increase in the 10-year rate, a 4.191 basis-point increase in the 20-year rate, and a 4.069 basis-point increase in the 30-year rate. Similarly, a one standard-deviation increase of the second PC is a 1.068 basis-point drop in the 10-year rate, a .260 basis-point increase in the 20-year rate, and a .812 basis-point increase in the 30-year rate. Finally, a one standard-deviation increase of the third PC constitutes changes of .032, −.103, and .075 basis points in each of the rates, respectively.

To appreciate the scaling of the PCs in Table 10-8, note the following implications:

- By construction, the PCs are uncorrelated. Hence, the volatility of the 10-year rate can be recovered from Table 10-8 as

$$\sqrt{4.114^2 + (-1.068)^2 + .032^2} = 4.25 \tag{10.42}$$

And the volatilities of the 20- and 30-year rates can be recovered equivalently.

- The variance of each PC is the sum of squares of its elements, or, its volatility is the square root of that sum of squares. For the three PCs,

$$\sqrt{4.114^2 + 4.191^2 + 4.069^2} = 7.14 \tag{10.43}$$

$$\sqrt{(-1.068)^2 + .260^2 + .812^2} = 1.37 \tag{10.44}$$

$$\sqrt{.032^2 + (-.103)^2 + .075^2} = .13 \tag{10.45}$$

- The square root of the sum of the variances of the PCs is the square root of the sum of the variances of the rates, which quantity was given above as 7.27 basis points:

$$\sqrt{7.14^2 + 1.37^2 + .13^2} = \sqrt{52.925} = 7.27 \tag{10.46}$$

- The volatility of any portfolio can be found by computing its volatility with respect to each of the PCs and then taking the square root of the sum of the resulting variances. Returning to the portfolio with *DV*01 weights of $\mathbf{w}' = (.5, -1, .6)$, its volatility with respect to each of the PCs can be computed as in Equations (10.47) through (10.49). Then, adding the sum of these squares and taking the square root, gives a portfolio volatility of .464, as computed earlier from the variances and covariances.

$$\sqrt{(\mathbf{w}'\tilde{\mathbf{a}})^2} = \sqrt{(.5 \times 4.114 - 1 \times 4.191 + .6 \times 4.069)^2} = .3074 \tag{10.47}$$

$$\sqrt{(\mathbf{w}'\tilde{\mathbf{b}})^2} = \sqrt{(.5 \times (-1.068 - 1 \times .260 + .6 \times .812)^2} = .3068 \tag{10.48}$$

$$\sqrt{(\mathbf{w}'\tilde{\mathbf{c}})^2} = \sqrt{(.5 \times .032 - 1 \times (-.103) + .6 \times .075)^2} = .1640 \tag{10.49}$$

In summary, the PCs in Table 10-8 contain the same information as the variances and covariances, but have the interpretation of one standard-deviation changes in the level, slope, and short rate factors. Of course, the power of the methodology is evident not in a simple example like this, but when, as in the text, changes in 30 rates can be adequately expressed with changes in three factors.

The Science of Term Structure Models

11

Learning Objectives

After completing this reading you should be able to:

- Calculate the expected discounted value of a zero-coupon security using a binomial tree.
- Construct and apply an arbitrage argument to price a call option on a zero-coupon security using replicating portfolios.
- Define risk-neutral pricing and apply it to option pricing.
- Distinguish between true and risk-neutral probabilities, and apply this difference to interest rate drift.
- Explain how the principles of arbitrage pricing of derivatives on fixed income securities can be extended over multiple periods.
- Define option-adjusted spread (OAS) and apply it to security pricing.
- Describe the rationale behind the use of recombining trees in option pricing.
- Calculate the value of a constant maturity Treasury swap, given an interest rate tree and the risk-neutral probabilities.
- Evaluate the advantages and disadvantages of reducing the size of the time steps on the pricing of derivatives on fixed income securities.
- Evaluate the appropriateness of the Black-Scholes-Merton model when valuing derivatives on fixed income securities.
- Describe the impact of embedded options on the value of fixed income securities.

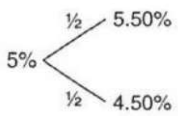
Excerpt is Chapter 7 of Fixed Income Securities, *Third Edition, by Bruce Tuckman.*

This chapter uses a very simple setting to show how to price interest rate contingent claims relative to a set of underlying securities by arbitrage arguments. Unlike the arbitrage pricing of securities with fixed cash flows, the techniques of this chapter require strong assumptions about how interest rates evolve in the future. This chapter also introduces *option-adjusted spread* (OAS) as the most popular measure of deviations of market prices from those predicted by models.

RATE AND PRICE TREES

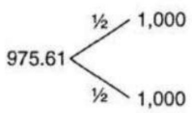
Assume that the six-month and one-year spot rates are 5% and 5.15% respectively. Taking these market rates as given is equivalent to taking the prices of a six-month bond and a one-year bond as given. Securities with assumed prices are called underlying securities to distinguish them from the contingent claims priced by arbitrage arguments.

Next, assume that six months from now the six-month rate will be either 4.50% or 5.50% with equal probability. This very strong assumption is depicted by means of a *binomial tree*, where "binomial" means that only two future values are possible:



Note that the columns in the tree represent dates. The six-month rate is 5% today, which will be called date 0. On the next date six months from now, which will be called date 1, there are two possible outcomes or *states of the world.* The 5.50% state will be called the *up-state* while the 4.50% state will be called the *down-state.*

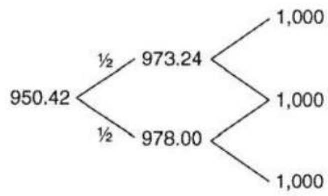
Given the current term structure of spot rates (i.e., the current six-month and one-year rates), trees for the prices of six-month and one-year zero-coupon bonds may be computed. The price tree for $1,000 face value of the six-month zero is



since $\$1{,}000/(1 + .05/2) = \$975.61$. (For easy readability, currency symbols are not included in price trees).

Note that in a tree for the value of a particular security, the maturity of the security falls with the date. On date 0 of the preceding tree the security is a six-month zero, while on date 1 the security is a maturing zero.

The price tree for $1,000 face value of a one-year zero is the following:



The three date 2 prices of $1,000 are, of course, the maturity values of the one-year zero. The two date 1 prices come from discounting this certain $1,000 at the then-prevailing six-month rate. Hence, the date 1 up-state price is $\$1{,}000/(1 + .055/2)$ or $973.2360, and the date 1 down-state price is $\$1{,}000/(1 + .045/2)$ or $977.9951. Finally, the date 0 price is computed using the given date 0 one-year rate of 5.15%: $\$1{,}000/(1 + .0515/2)^2$ or 950.423.

The probabilities of moving up or down the tree may be used to compute the average or expected values. As of date 0, the expected value of the one-year zero's price on date 1 is

$$\frac{1}{2}\$973.24 + \frac{1}{2}\$978.00 = \$975.62 \tag{11.1}$$

Discounting this expected value to date 0 at the date 0, six-month rate gives an *expected discounted value*[1] of

$$\frac{\frac{1}{2}\$973.24 + \frac{1}{2}\$978.00}{(1+\frac{.05}{2})} = \$951.82 \tag{11.2}$$

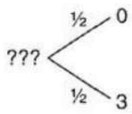
Note that the one-year zero's expected discounted value of $951.82 does not equal its given market price of $950.42. These two numbers need not be equal because investors do not price securities by expected discounted value. Over the next six months the one-year zero is a risky security, worth $973.24 half of the time and $978 the other half of the time for an average or expected value of $975.62. If investors do not like this price uncertainty, they would prefer a security worth $975.62 on date 1 with certainty. More specifically, a security worth $975.62 with

[1] Over one period, discounting the expected value and taking the expectation of discounted values are the same. But, over many periods the two are different and, with the approach taken by the short rate models, taking the expectation of discounted values is correct—hence the choice of the term "expected discounted value."

certainty after six months would sell for $\$975.62/(1 + \frac{.05}{2})$ or \$951.82 as of date 0. By contrast, investors penalize the risky one-year zero-coupon bond with an average price of \$975.62 after six months by pricing it at \$950.42. The next chapter elaborates further on investor *risk aversion* and how large an impact it might be expected to have on bond prices.

ARBITRAGE PRICING OF DERIVATIVES

The text now turns to the pricing of a derivative security. What is the price of a call option, maturing in six months, to purchase \$1,000 face value of a then six-month zero at \$975? Begin with the price tree for this call option:



If on date 1 the six-month rate is 5.50% and a six-month zero sells for \$973.23, the right to buy that zero at \$975 is worthless. On the other hand, if the six-month rate turns out to be 4.50% and the price of a six-month zero is \$978, then the right to buy the zero at \$975 is worth \$978 – \$975 or \$3. This description of the option's terminal payoffs emphasizes the derivative nature of the option: its value depends on the value of an underlying security.

A security is priced by arbitrage by finding and pricing its replicating portfolio. When, as in that context, cash flows do not depend on the levels of rates, the construction of the replicating portfolio is relatively simple. The derivative context is more difficult because cash flows do depend on the levels of rates, and the replicating portfolio must replicate the derivative security for any possible interest rate scenario.

To price the option by arbitrage, construct a portfolio on date 0 of underlying securities, namely six-month and one-year zero-coupon bonds, that will be worth \$0 in the up-state on date 1 and \$3 in the down-state. To solve this problem, let $F^{.5}$ and F^{1} be the face values of six-month and one-year zeros in the replicating portfolio, respectively. Then, these values must satisfy the following two equations:

$$F^{.5} + .97324F^{1} = \$0 \qquad \textbf{(11.3)}$$

$$F^{.5} + .97800F^{1} = \$3 \qquad \textbf{(11.4)}$$

Equation (11.3) may be interpreted as follows. In the up-state, the value of the replicating portfolio's now maturing six-month zero is its face value. The value of the once one-year zeros, now six-month zeros, is .97324 per dollar face value. Hence, the left-hand side of Equation (11.3) denotes the value of the replicating portfolio in the up-state. This value must equal \$0, the value of the option in the up-state. Similarly, Equation (11.4) requires that the value of the replicating portfolio in the down-state equal the value of the option in the down-state.

Solving Equations (11.3) and (11.4), $F^{.5} = -\$613.3866$ and $F^{1} = \$630.2521$. In words, on date 0 the option can be replicated by buying about \$630.25 face value of one-year zeros and simultaneously shorting about \$613.39 face amount of six-month zeros. Since this is the case, the law of one price requires that the price of the option equal the price of the replicating portfolio. But this portfolio's price is known and is equal to

$$.97561F^{.5} + .95042F^{1} = -.97561 \times \$613.3866 + .95042 \times \$630.2521 = \$.58 \qquad \textbf{(11.5)}$$

Therefore, the price of the option must be \$.58.

Recall that pricing based on the law of one price is enforced by arbitrage. If the price of the option were less than \$.58, arbitrageurs could buy the option, short the replicating portfolio, keep the difference, and have no future liabilities. Similarly, if the price of the option were greater than \$.58, arbitrageurs could short the option, buy the replicating portfolio, keep the difference, and, once again, have no future liabilities. Thus, ruling out profits from riskless arbitrage implies an option price of \$.58.

It is important to emphasize that the option cannot be priced by expected discounted value. Under that method, the option price would appear to be

$$\frac{.5 \times \$0 + .5 \times \$3}{1 + \frac{.05}{2}} = \$1.46 \qquad \textbf{(11.6)}$$

The true option price is less than this value because investors dislike the risk of the call option and, as a result, will not pay as much as its expected discounted value. Put another way, the risk penalty implicit in the call option price is inherited from the risk penalty of the one-year zero, that is, from the property that the price of the one-year zero is less than its expected discounted value. Once again, the magnitude of this effect is discussed in the next chapter.

This section illustrates arbitrage pricing with a call option, but it should be clear that arbitrage can be used to price any security with cash flows that depend on the six-month rate. Consider, for example, a security that, in six months, requires a payment of \$200 in the up-state but generates a payment of \$1,000 in the down-state. Proceeding as in the option example, find the portfolio of six-month and one-year zeros that replicates these two terminal payoffs, price this replicating portfolio as of date 0, and conclude that the price of the hypothetical security equals the price of the replicating portfolio.

A remarkable feature of arbitrage pricing is that the probabilities of up and down moves never enter into the calculation of the arbitrage price. See Equations (11.3) to (11.5). The explanation for this somewhat surprising observation follows from the principles of arbitrage. Arbitrage pricing requires that the value of the replicating portfolio matches the value of the option in both the up and the down-states. Therefore, the composition of the replicating portfolio is the same whether the probability of the up-state is 20%, 50%, or 80%. But if the composition of the portfolio does not depend directly on the probabilities, and if the prices of the securities in the portfolio are given, then the price of the replicating portfolio and hence the price of the option cannot depend directly on the probabilities either.

Despite the fact that the option price does not depend directly on the probabilities, these probabilities must have some impact on the option price. After all, as it becomes more and more likely that rates will rise to 5.50% and that bond prices will be low, the value of options to purchase bonds must fall. The resolution of this apparent paradox is that the option price depends indirectly on the probabilities through the price of the one-year zero. Were the probability of an up move to increase suddenly, the current value of a one-year zero would decline. And since the replicating portfolio is long one-year zeros, the value of the option would decline as well. In summary, a derivative like an option depends on the probabilities only through current bond prices. Given bond prices, however, probabilities are not needed to derive arbitrage-free prices.

RISK-NEUTRAL PRICING

Risk-neutral pricing is a technique that modifies an assumed interest rate process, like the one assumed at the start of this chapter, so that any contingent claim can be priced without having to construct and price its replicating portfolio. Since the original interest rate process has to be modified only once, and since this modification requires no more effort than pricing a single contingent claim by arbitrage, risk-neutral pricing is an extremely efficient way to price many contingent claims under the same assumed rate process.

In the example of this chapter, the price of a one-year zero does not equal its expected discounted value. The price of the one-year zero is \$950.42, computed from the given one-year spot rate of 5.15%. At the same time, the expected discounted value of the one-year zero is \$951.82, as derived in Equation (11.2) and reproduced here:

$$\frac{\frac{1}{2}\$973.24 + \frac{1}{2}\$978.00}{\left(1+\frac{.05}{2}\right)} = \$951.82 \qquad \textbf{(11.7)}$$

The probabilities of ½ for the up and down-states are the assumed true or real-world probabilities. But there are other probabilities, called *risk-neutral* probabilities, that do cause the expected discounted value to equal the market price. To find these probabilities, let the risk-neutral probabilities in the up and down-states be p and $(1-p)$, respectively. Then, solve the following equation:

$$\frac{\$973.24p + \$978.00(1-p)}{1+\frac{.05}{2}} = \$950.42 \qquad \textbf{(11.8)}$$

The solution is $p = .8024$. In words, under the risk-neutral probabilities of .8024 and .1976 the expected discounted value equals the market price.

In later chapters the difference between true and risk-neutral probabilities is described in terms of the *drift* in interest rates. Under the true probabilities there is a 50% chance that the six-month rate rises from 5% to 5.50% and a 50% chance that it falls from 5% to 4.50%. Hence the expected change in the six-month rate, or the drift of the six-month rate, is zero. Under the risk-neutral probabilities there is an 80.24% chance of a 50-basis point increase in the six-month rate and a 19.76% chance of a 50-basis point decline for an expected change of 30.24 basis points. Hence the drift of the six-month rate under these probabilities is 30.24 basis points.

As pointed out in the previous section, the expected discounted value of the option payoff is \$1.46, while the arbitrage price is \$.58. But what if expected discounted value is computed using the risk-neutral probabilities? The resulting option value would be:

$$\frac{.8024 \times \$0 + .1976 \times \$3}{1+\frac{.05}{2}} = \$.58 \qquad (11.9)$$

The fact that the arbitrage price of the option equals its expected discounted value under the risk-neutral probabilities is not a coincidence. In general, to value contingent claims by risk-neutral pricing, proceed as follows. First, find the risk-neutral probabilities that equate the price of the underlying securities with their expected discounted values. (In the simple example of this chapter the only risky, underlying security is the one-year zero.) Second, price the contingent claim by expected discounted value under these risk-neutral probabilities. The remainder of this section will describe intuitively why risk-neutral pricing works. Since the argument is a bit complex, it is broken up into four steps.

Step 1: Given trees for the underlying securities, the price of a security that is priced by arbitrage does not depend on investors' risk preferences. This assertion can be supported as follows.

A security is priced by arbitrage if one can construct a portfolio that replicates its cash flows. Under the assumed process for interest rates in this chapter, for example, the sample bond option is priced by arbitrage. By contrast, it is unlikely that a specific common stock can be priced by arbitrage because no portfolio of underlying securities can mimic the idiosyncratic fluctuations in a single common stock's market value.

If a security is priced by arbitrage and everyone agrees on the price evolution of the underlying securities, then everyone will agree on the replicating portfolio. In the option example, both an extremely risk-averse, retired investor and a professional gambler would agree that a portfolio of \$630.25 face of one-year zeros and −\$613.39 face of six-month zeros replicates the option. And since they agree on the composition of the replicating portfolio and on the prices of the underlying securities, they must also agree on the price of the derivative.

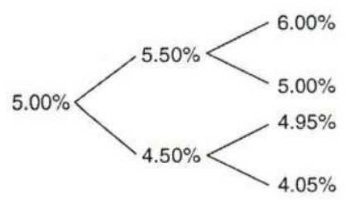
Step 2: Imagine an economy identical to the true economy with respect to current bond prices and the possible value of the six-month rate over time but different in that the investors in the imaginary economy are risk neutral. Unlike investors in the true economy, investors in the imaginary economy do not penalize securities for risk and, therefore, price securities by expected discounted value. It follows that, under the probabilities in the imaginary economy, the expected discounted value of the one-year zero equals its market price. But these probabilities satisfy Equation (11.8), namely the risk-neutral probabilities of .8024 and .1976.

Step 3: The price of the option in the imaginary economy, like any other security in that economy, is computed by expected discounted value. Since the probability of the up-state in that economy is .8024, the price of the option in that economy is given by Equation (11.9) and is, therefore, \$.58.

Step 4: Step 1 implies that given the prices of the six-month and one-year zeros, as well as possible values of the six-month rate, the price of an option does not depend on investor risk preferences. It follows that since the real and imaginary economies have the same bond prices and the same possible values for the six-month rate, the option price must be the same in both economies. In particular, the option price in the real economy must equal \$.58, the option price in the imaginary economy. More generally, the price of a derivative in the real economy may be computed by expected discounted value under the risk-neutral probabilities.

ARBITRAGE PRICING IN A MULTI-PERIOD SETTING

Maintaining the binomial assumption, the tree of the previous section might be extended for another six months as follows:

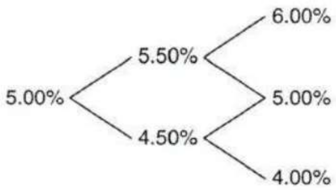


When, as in this tree, an up move followed by a down move does not give the same rate as a down move followed by an up move, the tree is said to be *nonrecombining*. From an economic perspective, there is nothing wrong with this kind of tree. To justify this particular tree, for example, one might argue that when short rates are 5% or higher they tend to change in increments of 50 basis points. But when rates fall below 5%, the size of the change starts to decrease. In particular, at a rate of 4.50%, the short rate may change by only 45 basis points.

A volatility process that depends on the level of rates exhibits *state-dependent* volatility.

Despite the economic reasonableness of nonrecombining trees, practitioners tend to avoid them because such trees are difficult or even impossible to implement. After six months there are two possible states, after one year there are four, and after N semiannual periods there are 2^N possibilities. So, for example, a tree with semiannual steps large enough to price 10-year securities will, in its rightmost column alone, have over 500,000 nodes, while a tree used to price 20-year securities will in its rightmost column have over 500 billion nodes. Furthermore, as discussed later in the chapter, it is often desirable to reduce substantially the time interval between dates. In short, even with modern computers, trees that grow this quickly are computationally unwieldy. This doesn't mean, by the way, that the effects that give rise to nonrecombining trees, like state-dependent volatility, have to be abandoned. It simply means that these effects must be implemented in a more efficient way.

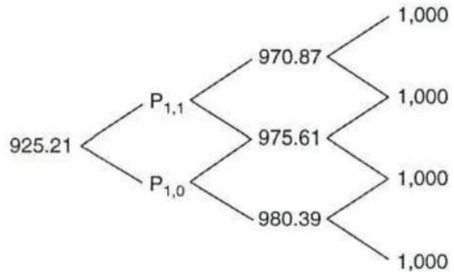
Trees in which the up-down and down-up-states have the same value are called *recombining trees.* An example of this type of tree that builds on the two-date tree of the previous sections is



Note that there are two nodes after six months, three after one year, and so on. A tree with weekly rather than semiannual steps capable of pricing a 30-year security would have only 52 × 30 + 1 or 1,561 nodes in its rightmost column. Evidently, recombining trees are much more manageable than nonrecombining trees from a computational viewpoint.

As trees grow it becomes convenient to develop a notation with which to refer to particular nodes. One convention is as follows. The dates, represented by columns of the tree, are numbered from left to right starting with 0. The states, represented by rows of the tree, are numbered from bottom to top, also starting from 0. For example, in the preceding tree the six-month rate on date 2, state 0 is 4%. The six-month rate on state 1 of date 1 is 5.50%.

Continuing where the option example left off, having derived the risk-neutral tree for the pricing of a one-year zero, the goal is to extend the tree for the pricing of a 1.5-year zero assuming that the 1.5-year spot rate is 5.25%. Ignoring the probabilities for a moment, several nodes of the 1.5-year zero price tree can be written down immediately:



On date 3, the zero with an original term of 1.5 years matures and is worth its face value of $1,000. On date 2, the value of the then six-month zero equals its face value discounted for six months at the then-prevailing spot rates of 6%, 5%, and 4% in states 2, 1, and 0, respectively:

$$\frac{\$1{,}000}{1+\frac{.06}{2}} = \$970.87 \qquad \textbf{(11.10)}$$

$$\frac{\$1{,}000}{1+\frac{.05}{2}} = \$975.61 \qquad \textbf{(11.11)}$$

$$\frac{\$1{,}000}{1+\frac{.04}{2}} = \$980.39 \qquad \textbf{(11.12)}$$

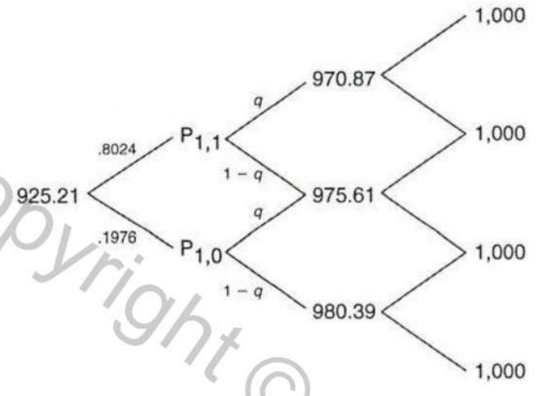
Finally, on date 0, the 1.5-year zero equals its face value discounted at the given 1.5-year spot rate:

$$\frac{\$1{,}000}{\left(1+\frac{.0525}{2}\right)^3} = \$925.21 \qquad \textbf{(11.13)}$$

The prices of the zero on date 1 in states 1 and 0 are denoted $P_{1,1}$ and $P_{1,0}$ respectively. The then one-year zero prices are not known because, at this point in the development, possible values of the one-year rate in six months are not available.

The previous section showed that the risk-neutral probability of an up move on date 0 is .8024. Letting q be the risk-neutral probability of an up move on date 1,[2] the tree becomes

[2] For simplicity alone, this example assumes that the probability of moving up from state 0 equals the probability of moving up from state 1.



By definition, expected discounted value under risk-neutral probabilities must produce market prices. With respect to the 1.5-year zero price on date 0, this requires that

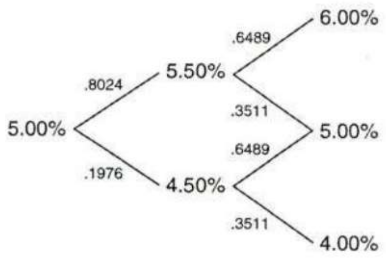
$$\frac{.8024P_{1,1} + .1976P_{1,0}}{1+\frac{.05}{2}} = \$925.21 \tag{11.14}$$

With respect to the prices of a then one-year zero on date 1,

$$P_{1,1} = \frac{\$970.87q + \$975.61(1-q)}{1+\frac{.055}{2}} \tag{11.15}$$

$$P_{1,0} = \frac{\$975.61q + \$980.39(1-q)}{1+\frac{.045}{2}} \tag{11.16}$$

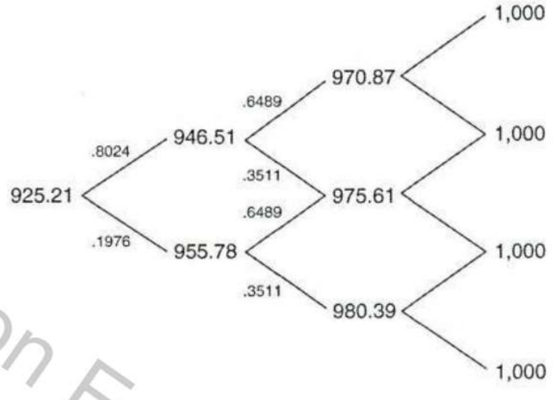
While Equations (11.14) through (11.16) may appear complicated, substituting (11.15) and (11.16) into (11.14) results in a linear equation in the one unknown, q. Solving this resulting equation reveals that $q = .6489$. Therefore, the risk-neutral interest rate process may be summarized by the following tree:



Furthermore, any derivative security that depends on the six-month rate in six months and in one year may be priced by computing its discounted expected value along this tree. An example appears in the next section.

The difference between the true and risk-neutral probabilities may once again be described in terms of drift. From dates 1 to 2, the drift under the true probabilities is zero. Under the risk-neutral probabilities the drift is computed from a 64.89% chance of a 50-basis point increase in the six-month rate and a 35.11% chance of a 50-basis point decline in the rate. These numbers give a drift or expected change of 14.89 basis points.

Substituting $q = .6489$ back into Equations (11.15) and (11.16) completes the tree for the price of the 1.5-year zero:



It follows immediately from this tree that the one-year spot rate six months from now may be either 5.5736% or 4.5743% since

$$\$946.51 = \frac{\$1{,}000}{(1+\frac{5.5736\%}{2})^2} \tag{11.17}$$

$$\$955.78 = \frac{\$1{,}000}{(1+\frac{4.5743\%}{2})^2} \tag{11.18}$$

The fact that the possible values of the one-year spot rate can be extracted from the tree is at first surprising. The starting point of the example is the date 0 values of the .5-, 1-, and 1.5-year spot rates as well as an assumption about the evolution of the six-month rate over the next year. But since this information, in combination with arbitrage or risk-neutral arguments, is sufficient to determine the price tree of the 1.5-year zero, it is sufficient to determine the possible values of the one-year spot rate in six months. Considering this fact from another point of view, having specified initial spot rates and the evolution of the six-month rate, a modeler may not make any further assumptions about the behavior of the one-year rate.

The six-month rate process completely determines the one-year rate process because the model presented here has only one factor. Writing down a tree for the evolution of the six-month rate alone implicitly assumes that prices of all fixed income securities can be determined by the evolution of that rate.

Just as some replicating portfolio can reproduce the cash flows of a security from date 0 to date 1, some other replicating portfolios can reproduce the cash flows of a security from date 1 to date 2. The composition of these replicating portfolios depends on the date and state. More specifically, the replicating portfolios held on date 0, on state 0 of date 1, and on state 1 of date 1 are usually different. From the trading perspective, the replicating portfolio must be adjusted as time passes and as interest rates change. This process is known as *dynamic replication*, in contrast to the *static replication strategies*. As an example of static replication, the portfolio of zero-coupon bonds that replicates a coupon bond does not change over time nor with the level of rates.

Having built a tree out to date 2 it should be clear how to extend the tree to any number of dates. Assumptions about the future possible values of the short-term rate have to be extrapolated further into the future and risk-neutral probabilities have to be calculated to recover a given set of bond prices.

EXAMPLE: PRICING A CONSTANT-MATURITY TREASURY SWAP

Equipped with the last tree of interest rates in the previous section, this section prices a particular derivative security, namely $1,000,000 face value of a stylized *constant-maturity Treasury (CMT) swap* struck at 5%. This swap pays

$$\$1{,}000{,}000\frac{y_{CMT}-5\%}{2} \quad \textbf{(11.19)}$$

every six months until it matures, where y_{CMT} is a semiannually compounded yield, of a predetermined maturity, on the payment date. The text prices a one-year CMT swap on the six-month yield. In practice, CMT swaps trade most commonly on the yields of the most liquid maturities, i.e., on 2-, 5- and 10-year yields.

Since six-month semiannually compounded yields equal six-month spot rates, rates from the tree of the previous section can be substituted into (11.19) to calculate the payoffs of the CMT swap. On date 1, the state 1 and state 0 payoffs are, respectively,

$$\$1{,}000{,}000\frac{5.50\%-5\%}{2}=\$2{,}500 \quad \textbf{(11.20)}$$

$$\$1{,}000{,}000\frac{4.50\%-5\%}{2}=-\$2{,}500 \quad \textbf{(11.21)}$$

Similarly on date 2, the state 2, 1, and 0 payoffs are, respectively,

$$\$1{,}000{,}000\frac{6\%-5\%}{2}=\$5{,}000 \quad \textbf{(11.22)}$$

$$\$1{,}000{,}000\frac{5\%-5\%}{2}=\$0 \quad \textbf{(11.23)}$$

$$\$1{,}000{,}000\frac{4\%-5\%}{2}=-\$5{,}000 \quad \textbf{(11.24)}$$

The possible values of the CMT swap at maturity, on date 2, are given by Equations (11.22) through (11.24). The possible values on date 1 are given by the expected discounted value of the date 2 payoffs under the risk-neutral probabilities plus the date 1 payoffs given by (11.20) and (11.21). The resulting date 1 values in states 1 and 0, respectively, are

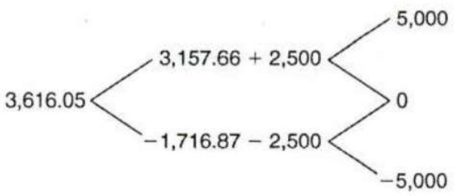
$$\frac{.6489\times\$5{,}000+.3511\times\$0}{1+\frac{.055}{2}}+\$2{,}500=\$5{,}657.66 \quad \textbf{(11.25)}$$

$$\frac{.6489\times 0+.3511\times(-\$5{,}000)}{1+\frac{.045}{2}}-\$2{,}500=-\$4{,}216.87 \quad \textbf{(11.26)}$$

Finally, the value of the swap on date 0 is the expected discounted value of the date 1 payoffs, given by (11.25) and (11.26), under the risk-neutral probabilities:

$$\frac{.8024\times\$5{,}657.66+.1976\times(-\$4{,}216.87)}{1+\frac{.05}{2}}=\$3{,}616.05 \quad \textbf{(11.27)}$$

The following tree summarizes the value of the stylized CMT swap over dates and states:



A value of $3,616.05 for the CMT swap might seem surprising at first. After all, the cash flows of the CMT swap are zero at a rate of 5%, and 5% is, under the real probabilities, the average rate on each date. The explanation, of course, is that the risk-neutral probabilities, not the real probabilities, determine the arbitrage price of the swap. The expected discounted value of the swap under the real probabilities can be computed by following the steps leading to (11.25) through (11.27) but using .5 for all up and down moves. The result of these calculations does give a value close to zero, namely −\$5.80.

The expected cash flow of the CMT swap on both dates 1 and 2, under the real probabilities, is zero. It follows immediately that the discounted value of these expected cash flows is zero. At the same time, the expected discounted value of the CMT swap is −\$5.80.

OPTION-ADJUSTED SPREAD

Option-adjusted spread (OAS) is a widely-used measure of the relative value of a security, that is, of its market price relative to its model value. OAS is defined as the spread such that the market price of a security equals its model price when discounted values are computed at risk-neutral rates plus that spread. To illustrate, say that the market price of the CMT swap in the previous section is \$3,613.25, \$2.80 less than the model price. In that case, the OAS of the CMT swap turns out to be 10 basis points. To see this, add 10 basis points to the discounting rates of 5.5% and 4.5% in Equations (11.25) and (11.26), respectively, to get new swap values of

$$\frac{.6489 \times \$5{,}000 + .3511 \times \$0}{1+\frac{.056}{2}} + \$2{,}500 = \$5{,}656.13 \quad \textbf{(11.28)}$$

$$\frac{.6489 \times 0 + .3511 \times (-\$5{,}000)}{1+\frac{.046}{2}} - \$2{,}500 = -\$4{,}216.03 \quad \textbf{(11.29)}$$

Note that, when calculating value with an OAS spread, rates are only shifted for the purpose of discounting. Rates are not shifted for the purposes of computing cash flows. In the CMT swap example, cash flows are still computed using Equations (11.20) through (11.24).

Completing the valuation with an OAS of 10 basis points, use the results of (11.28) and (11.29) and a discount rate of 5% plus the OAS spread of 10 basis points, or 5.10%, to obtain an initial CMT swap value of

$$\frac{.8024 \times \$5{,}656.13 + .1976 \times (-\$4{,}216.03)}{1+\frac{.051}{2}} = \$3{,}613.25 \quad \textbf{(11.30)}$$

Hence, as claimed, discounting at the risk-neutral rates plus an OAS of 10 basis points produces a model price equal to the given market price of \$3,613.25.

If a security's OAS is positive, its market price is less than its model price, so the security trades cheap. If the OAS is negative, the security trades rich.

Another perspective on the relative value implications of an OAS spread is the fact that the expected return of a security with an OAS, under the risk-neutral process, is the short-term rate plus the OAS per period. Very simply, discounting a security's expected value by a particular rate per period is equivalent to that security's earning that rate per period. In the example of the CMT swap, the expected return of the fairly-priced swap under the risk-neutral process over the six months from date 0 to date 1 is

$$\frac{.8024 \times \$5{,}657.66 - .1976 \times \$4{,}216.87 - \$3{,}616.05}{\$3{,}616.05} = 2.5\% \quad \textbf{(11.31)}$$

which is six month's worth of the initial rate of 5%. On the other hand, the expected return of the cheap swap, with an OAS of 10 basis points, is

$$\frac{.8024 \times \$5{,}656.13 - .1976 \times \$4{,}216.03 - \$3{,}613.25}{\$3{,}613.25} = 2.55\% \quad \textbf{(11.32)}$$

which is six month's worth of the initial rate of 5% plus the OAS of 10 basis points, or half of 5.10%.

PROFIT AND LOSS ATTRIBUTION WITH AN OAS

We introduced profit and loss (P&L) attribution. This section gives a mathematical description of attribution in the context of term structure models and of securities that trade with an OAS.

By the definition of a one-factor model, and by the definition of OAS, the market price of a security at time *t* and a factor value of *x* can be written as $P_t(x, OAS)$. Using a first-order Taylor approximation, the change in the price of the security is

$$dP = \frac{\partial P}{\partial x}dx + \frac{\partial P}{\partial t}dt + \frac{\partial P}{\partial OAS}dOAS \quad \textbf{(11.33)}$$

Dividing by the price and taking expectations,

$$E\left[\frac{dP}{P}\right] = \frac{1}{P}\frac{\partial P}{\partial x}E[dx] + \frac{1}{P}\frac{\partial P}{\partial t}dt \qquad \textbf{(11.34)}$$

Since the OAS calculation assumes that OAS is constant over the life of the security, moving from (11.33) to (11.34) assumes that the expected change in the OAS is zero.

As mentioned in the previous section, if expectations are taken with respect to the risk-neutral process,[3] then, for any security priced according to the model,

$$E\left[\frac{dP}{P}\right] = rdt \qquad \textbf{(11.35)}$$

But Equation (11.35) does not apply to securities that are not priced according to the model, that is, to securities with an OAS not equal to zero. For these securities, by definition, the cash flows are discounted not at the short-term rate but at the short-term rate plus the OAS. Equivalently, as argued in the previous section, the expected return under the risk-neutral probabilities is not the short-term rate but the short-term rate plus the OAS. Hence, the more general form of (11.35), is

$$E\left[\frac{dp}{P}\right] = (r + OAS)dt \qquad \textbf{(11.36)}$$

Combining these pieces, substitute (11.34) and (11.36) into (11.33) and rearrange terms to break down the return of a security into its component parts:

$$\frac{dP}{P} = (r + OAS)dt + \frac{1}{P}\frac{\partial P}{\partial x}(dx - E[dx]) + \frac{1}{P}\frac{\partial P}{\partial OAS}dOAS \qquad \textbf{(11.37)}$$

Finally, multiplying through by P,

$$dP = (r + OAS)Pdt + \frac{\partial P}{\partial x}(dx - E[dx]) + \frac{\partial P}{\partial OAS}dOAS \qquad \textbf{(11.38)}$$

In words, the return of a security or its P&L may be divided into a component due to the passage of time, a component due to changes in the factor, and a component due to the change in the OAS. The terms on the right-hand side of (11.38) represent, in order, carry-roll-down,[4] gains or losses from rate changes, and gains or losses from spread change. For models with predictive power, the OAS converges or tends to zero, or, equivalently, the security price converges or tends toward its fair value according to the model.

The decompositions (11.37) and (11.38) highlight the usefulness of OAS as a measure of the value of a security with respect to a particular model. According to the model, a long position in a cheap security earns superior returns in two ways. First, it earns the OAS over time intervals in which the security does not converge to its fair value. Second, it earns its sensitivity to OAS times the extent of any convergence.

The decomposition equations also provide a framework for thinking about relative value trading. When a cheap or rich security is identified, a relative value trader buys or sells the security and hedges out all interest rate or factor risk. In terms of the decompositions, $\partial P/\partial x = 0$. In that case, the expected return or P&L depends only on the short-term rate, the OAS, and any convergence. Furthermore, if the trader finances the trade at the short-term rate, i.e., borrows P at a rate r to purchase the security, the expected return is simply equal to the OAS plus any convergence return.

REDUCING THE TIME STEP

To this point this chapter has assumed that the time elapsed between dates of the tree is six months. The methodology outlined previously, however, can be easily adapted to any time step of Δt years. For monthly time steps, for example, $\Delta t = 1/12$ or .0833, and one-month rather than six-month interest rates appear on the tree. Furthermore, discounting must be done over the appropriate time interval. If the rate of term Δt is r, then discounting means dividing by $1 + r\,\Delta t$. In the case of monthly time steps, discounting with a one-month rate of 5% means dividing by $1 + .05/12$

In practice, there are two reasons to choose time steps smaller than six months. First, a security or portfolio of securities rarely makes all of its payments in even six-month intervals from the starting date. Reducing the time step to a month, a week, or even a day can ensure that all cash flows are sufficiently close in time to some date in the tree. Second, assuming that the six-month rate can take on only two values in six months, three values

[3] Taking expected values with respect to the true probabilities would add a risk premium term to the right-hand side of this equation. See Chapter 12.

[4] For expositional simplicity, no explicit coupon or other direct cash flows have been included in this discussion.

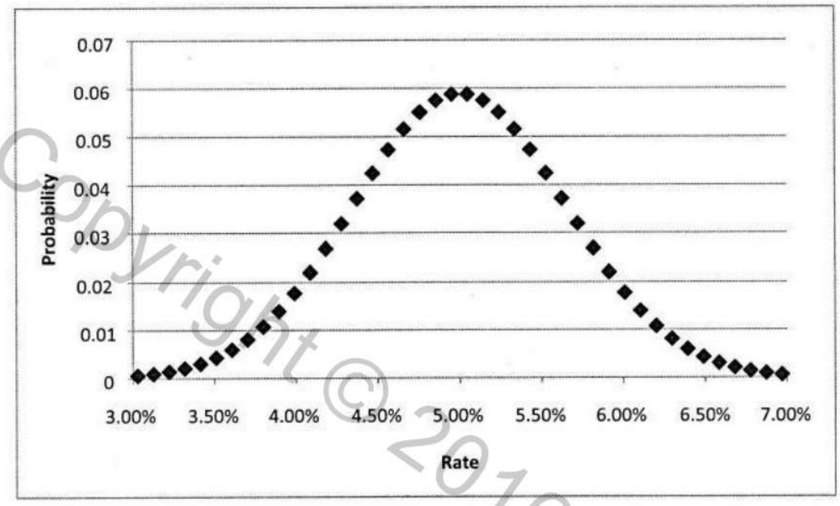


FIGURE 11-1 Sample probability distribution of the six-month rate in six months with daily time steps.

option: that tree would imply only two possible rates on the option expiration date.

While the trees in this chapter assume that the step size is the same throughout the tree, this need not be the case. Sophisticated implementations of trees allow step size to vary across dates in order to achieve a balance between realism and computational concerns.

in one year, and so on, produces a tree that is too coarse for many practical pricing problems. Reducing the step size can fill the tree with enough rates to price contingent claims with sufficient accuracy. Figure 11-1 illustrates this point by showing a relatively realistic-looking probability distribution of the six-month rate in six months from a tree with daily time steps, a drift of zero, and a horizon standard deviation of 65 basis points.

While smaller time steps generate more realistic interest rate distributions, it is not the case that smaller time steps are always desirable. First, the greater the number of computations in pricing a security, the more attention must be paid to numerical issues like round-off error. Second, since decreasing the time step increases computation time, practitioners requiring quick results cannot make the time step too small. Customers calling market makers in options on swaps, or *swaptions*, for example, expect price quotations within minutes if not sooner. Hence, the time step in a model used to price swaptions must be consistent with the market maker's required response time.

The best choice of step size ultimately depends on the problem at hand. When pricing a 30-year callable bond, for example, a model with monthly time steps may provide a realistic enough interest rate distribution to generate reliable prices. The same monthly steps, however, will certainly be inadequate to price a one-month bond

FIXED INCOME VERSUS EQUITY DERIVATIVES

While the ideas behind pricing fixed income and equity derivatives are similar in many ways, there are important differences as well. In particular, it is worth describing why models created for the stock market cannot be adopted without modification for use in fixed income markets.

The famous Black-Scholes-Merton pricing analysis of stock options can be summarized as follows. Under the assumption that the stock price evolves according to a particular random process and that the short-term interest rate is constant, it is possible to form a portfolio of stocks and short-term bonds that replicates the payoffs of an option. Therefore, by arbitrage arguments, the price of the option must equal the known price of the replicating portfolio.

Say that an investor wants to price an option on a five-year bond by a direct application of this logic. The investor would have to begin by making an assumption about how the price of the five-year bond evolves over time. But this is considerably more complicated than making assumptions about how the price of a stock evolves over time. First, the price of a bond must converge to its face value at maturity while the random process describing the stock price need not be constrained in any similar way. Second, because of the maturity constraint, the volatility of a bond's price must eventually get smaller as the bond approaches maturity. The simpler assumption that the volatility of a stock is constant is not so appropriate for bonds. Third, since stock volatility is very large relative to short-term rate volatility, it may be relatively harmless to assume that the short-term rate is constant. By contrast, it can be difficult to defend the assumption that a bond price follows some random process while the short-term interest rate is constant.[5]

These objections led researchers to make assumptions about the random evolution of the interest rate rather than of the bond price. In that way bond prices would naturally approach par, price volatilities would naturally approach zero, and the interest rate would not be assumed to be constant. But this approach raises another set of questions. Which interest rate is assumed to evolve in a particular way? Making assumptions about the 5-year rate over time is not particularly helpful for two reasons. First, 5-year coupon bond prices depend on shorter-term rates as well. Second, pricing an option on a 5-year bond requires assumptions about the bond's future possible prices. But knowing the 5-year rate over time is insufficient because, in a very short time, the option's underlying security will no longer be a 5-year bond. Therefore, one must often make assumptions about the evolution of the entire term structure of interest rates to price bond options and other derivatives. In the one-factor case described in this chapter it has been shown that modeling the evolution of the short-term rate is sufficient, combined with arbitrage arguments, to build a model of the entire term structure. In short, despite the enormous importance of the Black-Scholes-Merton analysis, the fixed income context does demand special attention.

Having reached the conclusion at the end of the previous paragraph, there are some contexts in which practitioners invoke assumptions so that the Black-Scholes-Merton models can be applied in place of more difficult-to-implement term structure models.

[5] Because these three objections are less important in the case of short-term options on long-term bonds, practitioners do use stock-like models in this fixed income context. Also, it is often sufficient to assume, somewhat more satisfactorily, that the relevant discount factor is uncorrelated with the price of the underlying bond.

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

The Evolution of Short Rates and the Shape of the Term Structure

12

Learning Objectives

After completing this reading you should be able to:

- Explain the role of interest rate expectations in determining the shape of the term structure.
- Apply a risk-neutral interest rate tree to assess the effect of volatility on the shape of the term structure.
- Estimate the convexity effect using Jensen's inequality.
- Evaluate the impact of changes in maturity, yield, and volatility on the convexity of a security.
- Calculate the price and return of a zero coupon bond incorporating a risk premium.

Excerpt is Chapter 8 of Fixed Income Securities, *Third Edition, by Bruce Tuckman.*

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

This chapter presents a framework for understanding the shape of the term structure. In particular, it is shown how spot or forward rates are determined by expectations of future short-term rates, the volatility of short-term rates, and an interest rate risk premium. To conclude the chapter, this framework is applied to swap curves in the United States and Japan.

INTRODUCTION

From assumptions about the interest rate process for the short-term rate and from an initial term structure implied by market prices, Chapter 11 showed how to derive a risk-neutral process that can be used to price all fixed income securities by arbitrage. Models that follow this approach, i.e., models that take the initial term structure as given, are called *arbitrage-free* models. A different approach, however, is to start with assumptions about the interest rate process and about the risk premium demanded by the market for bearing interest rate risk and then derive the risk-neutral process. Models of this sort do not necessarily match the initial term structure and are called *equilibrium* models.[1]

This chapter describes how assumptions about the interest rate process and about the risk premium determine the level and shape of the term structure. For equilibrium models, an understanding of the relationships between the model assumptions and the shape of the term structure is important in order to make reasonable assumptions in the first place. For arbitrage-free models, an understanding of these relationships reveals the assumptions implied by the market through the observed term structure.

Many economists might find this chapter remarkably narrow. An economist asked about the shape of the term structure would undoubtedly make reference to such macroeconomic factors as the marginal productivity of capital, the propensity to save, and expected inflation. The more modest goal of this chapter is to connect the dynamics of the short-term rate of interest and the risk premium with the shape of the term structure. While this goal does fall short of answers that an economist might provide, it is more ambitious than the derivation of arbitrage restrictions on bond and derivative prices given underlying bond prices.

EXPECTATIONS

The word *expectations* implies uncertainty. Investors might expect the one-year rate to be 10%, but know there is a good chance it will turn out to be 8% or 12%. For the purposes of this section alone, the text assumes away uncertainty so that the statement that investors expect or forecast a rate of 10% means that investors assume that the rate will be 10%. The sections to follow reintroduce uncertainty.

To highlight the role of interest rate forecasts in determining the shape of the term structure, consider the following simple example. The one-year interest rate is currently 10% and all investors forecast that the one-year interest rate next year and the year after will also be 10%. In that case, investors will discount cash flows using forward rates of 10%. In particular, the price of one-, two- and three-year zero-coupon bonds per dollar face value (using annual compounding) will be

$$P^1 = \frac{1}{1.10} \tag{12.1}$$

$$P^2 = \frac{1}{(1.10)(1.10)} = \frac{1}{1.10^2} \tag{12.2}$$

$$P^3 = \frac{1}{(1.10)(1.10)(1.10)} = \frac{1}{1.10^3} \tag{12.3}$$

From inspection of Equations (12.1) through (12.3), the term structure of spot rates in this example is flat at 10%. Very simply, investors are willing to lock in 10% for two or three years because they assume that the one-year rate will always be 10%.

Now assume that the one-year rate is still 10%, but that all investors forecast the one-year rate next year to be 12% and the one-year rate in two years to be 14%. In that case, the one-year spot rate is still 10%. The two-year spot rate, $\hat{r}(2)$, is such that

$$P^2 = \frac{1}{(1.10)(1.12)} = \frac{1}{(1+\hat{r}(2))^2} \tag{12.4}$$

Solving, $\hat{r}(2) = 10.995\%$. Similarly, the three-year spot rate, $\hat{r}(3)$, is such that

[1] This nomenclature is somewhat misleading. Equilibrium models, in the context of their assumptions, which do not include market prices for the initial term structure, are also arbitrage-free.

$$P^3 = \frac{1}{(1.10)(1.12)(1.14)} = \frac{1}{(1+\hat{r}(3))^3} \quad \textbf{(12.5)}$$

Solving, $\hat{r}(3) = 11.998\%$. Hence, the evolution of the one-year rate from 10% to 12% to 14% generates an upward-sloping term structure of spot rates: 10%, 10.995%, and 11.988%. In this case, investors require rates above 10% when locking up their money for two or three years because they assume one-year rates will be higher than 10%. No investor, for example, would buy a two-year zero at a yield of 10% when it is possible to buy a one-year zero at 10% and, when it matures, buy another one-year zero at 12%.

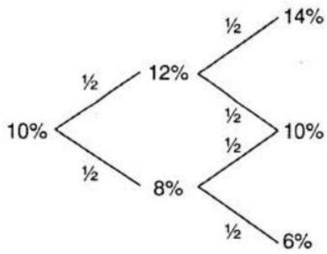
Finally, assume that the one-year rate is 10%, but that investors forecast that it will fall to 8% in one year and to 6% in two years. In that case, it is easy to show that the term structure of spot rates will be downward-sloping. In particular, $\hat{r}(1) = 10\%$, $\hat{r}(2) = 8.995\%$, and $\hat{r}(3) = 7.988\%$.

These simple examples reveal that expectations can cause the term structure to take on any of a myriad of shapes. Over short horizons, the financial community can have very specific views about future short-term rates. Over longer horizons, however, expectations cannot be so granular. It would be difficult, for example, to defend the position that the expectation for the one-year rate 29 years from now is substantially different from the expectation of the one-year rate 30 years from now. On the other hand, an argument can be made that the long-run expectation of the short-term rate is 5%: 3% due to the long-run real rate of interest and 2% due to long-run inflation. Hence, forecasts can be very useful in describing the shape and level of the term structure over short-term horizons and the level of rates at very long horizons. This conclusion has important implications for extracting expectations from observed interest rates (see the application at the end of this chapter) and for choosing among term structure models.

VOLATILITY AND CONVEXITY

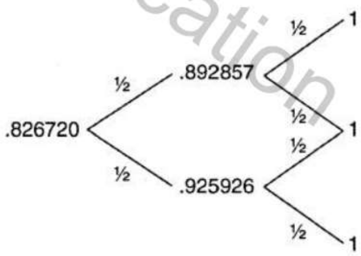
This section drops the assumption that investors believe that their forecasts will be realized and assumes instead that investors understand the volatility around their expectations. To isolate the implications of volatility on the shape of the term structure, this section assumes that investors are risk-neutral so that they price securities by expected discounted value. The next section drops this assumption.

Assume that the following tree gives the true process for the one-year rate:



Note that the expected interest rate on date 1 is $.5 \times 8\% + .5 \times 12\%$ or 10% and that the expected rate on date 2 is $.25 \times 14\% + .5 \times 10\% + .25 \times 6\%$ or 10%. In the previous section, with no volatility around expectations, flat expectations of 10% imply a flat term structure of spot rates. That is not the case in the presence of volatility.

The price of a one-year zero is, by definition, $\frac{1}{1.10}$ or .909091, implying a one-year spot rate of 10%. Under the assumption of risk-neutrality, the price of a two-year zero may be calculated by discounting the terminal cash flow using the preceding interest rate tree:



Hence, the two-year spot rate is such that $.82672 = (1 + \hat{r}(2))^{-2}$, implying that $\hat{r}(2) = 9.982\%$.

Even though the one-year rate is 10% and the expected one-year rate in one year is 10%, the two-year spot rate is 9.982%. The 1.8-basis point difference between the spot rate that would obtain in the absence of uncertainty, 10%, and the spot rate in the presence of volatility, 9.982%, is the effect of convexity on that spot rate. This convexity effect arises from the mathematical fact, a special case of *Jensen's Inequality*, that

$$E\left[\frac{1}{1+r}\right] > \frac{1}{E[1+r]} = \frac{1}{1+E[r]} \quad \textbf{(12.6)}$$

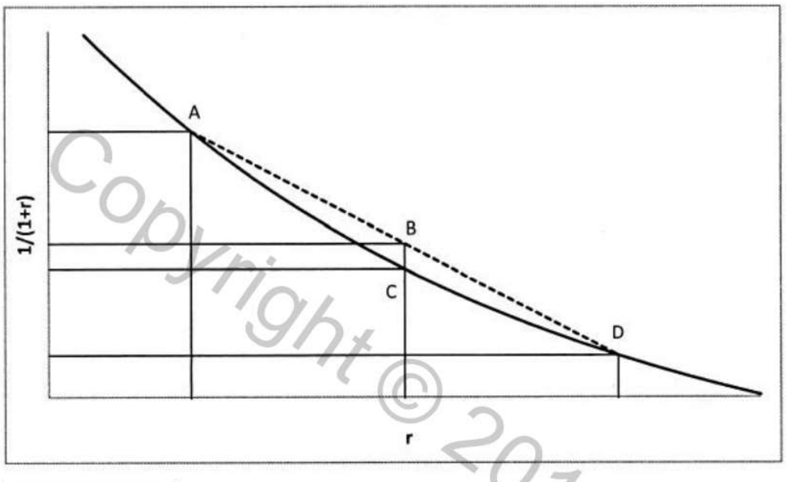


FIGURE 12-1 An illustration of convexity.

Figure 12-1 graphically illustrates this equation. There are two possible values of r and, consequently, of the function $\frac{1}{1+r}$ in the figure,[2] shown as points A and D. The height or vertical-axis coordinate of point B is the average of these two function values. Under the assumption that the two possible values of r occur with equal probability, this average can be thought of as $E[\frac{1}{1+r}]$ in (12.6). And under the same assumption, the horizontal-axis coordinates of the points B and C can be thought of as $E[r]$ so that the height of point C can be thought of as $\frac{1}{1+E[r]}$. Clearly, the height of B is greater than that of C, or $E[\frac{1}{1+r}] > \frac{1}{1+E[r]}$. To summarize, Equation (12.6) is true because the pricing function of a zero-coupon bond, $\frac{1}{1+r}$, is convex rather than concave.

Returning to the example of this section, Equation (12.6) may be used to show why the one-year spot rate is less than 10%. The spot rate one year from now may be 12% or 8%. According to (12.6),

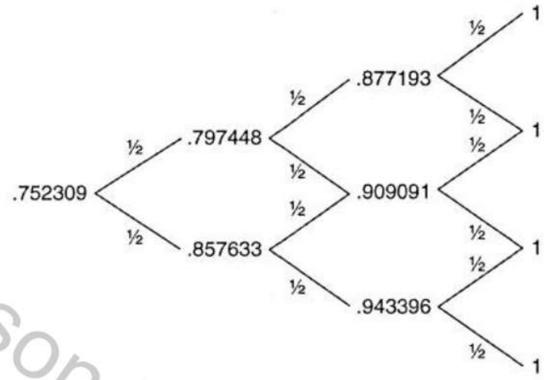
$$.5 \times \frac{1}{1.12} + .5 \times \frac{1}{1.08} > \frac{1}{.5 \times 1.12 + .5 \times 1.08} = \frac{1}{1.10} \quad \textbf{(12.7)}$$

Dividing both sides by 1.10,

$$\frac{1}{1.10}\left[.5 \times \frac{1}{1.12} + .5 \times \frac{1}{1.08}\right] > \frac{1}{1.10^2} \quad \textbf{(12.8)}$$

The left-hand side of (12.8) is the price of the two-year zero-coupon bond today. In words, then, Equation (12.8) says that the price of the two-year zero is greater than the result of discounting the terminal cash flow by 10% over the first period and by the expected rate of 10% over the second period. It follows immediately that the yield of the two-year zero, or the two-year spot rate, is less than 10%.

The tree presented at the start of this section may also be used to price a three-year zero. The resulting price tree is



The three-year spot rate, such that $.752309 = (1 + \hat{r}(3))^{-3}$, is 9.952%. Therefore, the value of convexity in this spot rate is 10% − 9.952% or 4.8 basis points, whereas the value of convexity in the two-year spot rate was only 1.8 basis points.

It is generally true that, all else equal, the value of convexity increases with maturity. This will become evident shortly. For now, suffice it to say that the convexity of the pricing function of a zero maturing in N years, $(1 + r)^{-N}$, increases with N. In terms of Figure 12-1, the longer the maturity of the illustrated pricing function, the more convex the curve.

Securities with greater convexity perform better when yields change a lot and perform worse when yields do not change by much. The discussion in this section shows that convexity does, in fact, lower bond yields. The mathematical development in a later section ties these observations together by showing exactly how the advantages of convexity are offset by lower yields.

The previous section assumes no interest rate volatility and, consequently, yields are completely determined by forecasts. In this section, with the introduction of volatility, yield is reduced by the value of convexity. So it may be said that the value of convexity arises from volatility.

[2] The curve shown is actually a power of $\frac{1}{1+r}$; i.e., the price of a longer-term zero-coupon bond, so that the curvature is more visible.

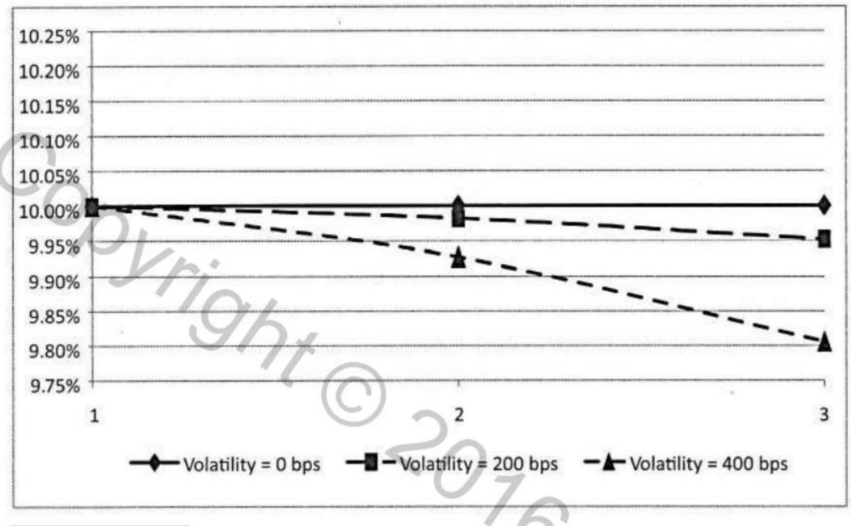
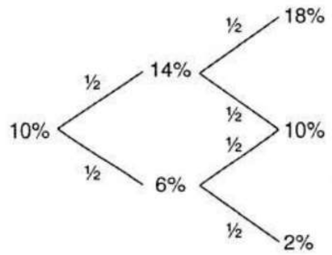


FIGURE 12-2 Volatility and the shape of the term structure in three-date binomial models.

Furthermore, the value of convexity increases with volatility. In the tree introduced at the start of the section, the standard deviation of rates is 200 basis points a year.[3] Now consider a tree with a standard deviation of 400 basis points:



The expected one-year rate in one year and in two years is still 10%. Spot rates and convexity values for this case may be derived along the same lines as before. Figure 12-2 graphs three term structures of spot rates: one with no volatility around the expectation of 10%; one with a volatility of 200 basis points a year (the tree of the first example); and one with a volatility of 400 basis points per year (the tree preceding this paragraph). Note that the value of convexity, measured by the distance between the rates assuming no volatility and the rates assuming volatility, increases with volatility. Figure 12-2 also illustrates that the value of convexity increases with maturity.

For very short terms and realistic levels of volatility, the value of convexity is quite small. But since simple examples must rely on short terms, convexity effects would hardly be discernible without raising volatility to unrealistic levels. Therefore, this section had to make use of unrealistically high volatility. The application at the end of this chapter uses realistic volatility levels to present typical convexity values.

RISK PREMIUM

To illustrate the effect of risk premium on the term structure, consider again the second interest rate tree presented in the preceding section, with a volatility of 400 basis points per year. Risk-neutral investors would price a two-year zero by the following calculation:

$$.827541 = \frac{.5\left[\frac{1}{1.14} + \frac{1}{1.06}\right]}{1.10} = \frac{.5[.877193 + .943396]}{1.10} \quad \textbf{(12.9)}$$

By discounting the expected future price by 10%, Equation (12.9) implies that the expected return from owning the two-year zero over the next year is 10%. To verify this statement, calculate this expected return directly:

$$.5\frac{.877193 - .827541}{.827541} + .5\frac{.943396 - .827541}{.827541} = .5 \times 6\% + .5 \times 14\% = 10\% \quad \textbf{(12.10)}$$

Would investors really invest in this two-year zero offering an expected return of 10% over the next year? The return will, in fact, be either 6% or 14%. While these two returns do average to 10%, an investor could, instead, buy a one-year zero with a certain return of 10%. Presented with this choice, any risk-averse investor would prefer an investment with a certain return of 10% to an investment with a risky return that averages 10%. In other words, investors require compensation for bearing interest rate risk.[4]

Risk-averse investors demand a return higher than 10% for the two-year zero over the next year. This return can be

[3] Chapter 13 describes the computation of the standard deviation of rates implied by an interest rate tree.

[4] This is an oversimplification. See the discussion at the end of the section.

effected by pricing the zero-coupon bond one year from now at less than the prices of $\frac{1}{1.14}$ and $\frac{1}{1.06}$. Equivalently, future cash flows could be discounted at rates higher than the possible rates of 14% and 6%. The next section shows that adding, for example, 20 basis points to each of these rates is equivalent to assuming that investors demand an extra 20 basis points for each year of duration risk. Assuming this is indeed the fair market *risk premium*, the price of the two-year zero would be computed as follows:

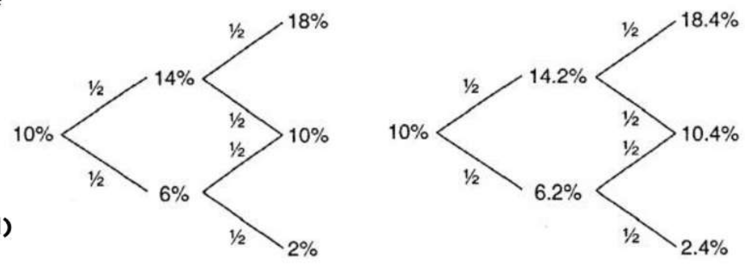
$$.826035 = \frac{.5\left[\frac{1}{1.142} + \frac{1}{1.062}\right]}{1.10} \qquad \textbf{(12.11)}$$

The price in (12.11) is below the value obtained in (12.9) which assumes that investors are risk-neutral. Put another way, the increase in the discounting rates has increased the expected return of the two-year zero. In one year, if the interest rate is 14%, then the price of a one-year zero will be $\frac{1}{1.14}$ or .877193. If the rate is 6%, then the price will be $\frac{1}{1.06}$ or .943396. Therefore, the expected return of the two-year zero priced at .826035 is

$$\frac{.5[.877193 + .943396] - .826035}{.826035} = 10.20\% \qquad \textbf{(12.12)}$$

Hence, recalling that the one-year zero has a certain return of 10%, the risk-averse investors in this example demand 20 basis points in expected return to compensate them for the one year of duration risk inherent in the two-year zero.[5]

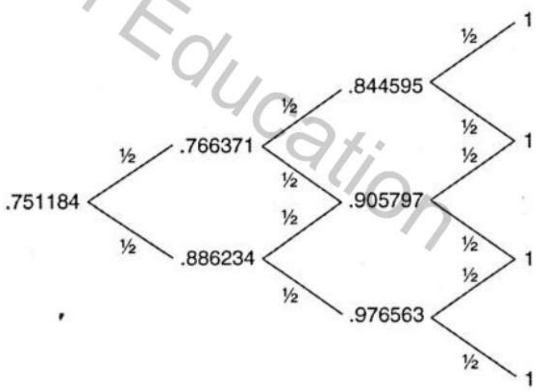
Continuing with the assumption that investors require 20 basis points for each year of duration risk, the three-year zero, with its approximately two years of duration risk,[6] needs to offer an expected return of 40 basis points. The next section shows that this return can be effected by pricing the three-year zero as if rates next year are 20 basis points above their true values and as if rates the year after next are 40 basis points above their true values. To summarize, consider trees (a) and (b) below. If tree (a) depicts the actual or true interest rate process, then pricing with tree (b) provides investors with a risk premium of 20 basis points for each year of duration risk. If this risk premium is, in fact, embedded in market prices, then, by definition, tree (b) is the risk-neutral interest rate process.



(a)　　(b)

The text now verifies that pricing the three-year zero with the risk-neutral process does offer an expected return of 10.4%, assuming that rates actually move according to the true process.

The price of the three-year zero can be computed by discounting using the risk-neutral tree:



To find the expected return of the three-year zero over the next year, proceed as follows. Two years from now the three-year zero will be a one-year zero with no interest rate risk.[7] Therefore, its price will be determined by discounting at the actual interest rate at that time: $\frac{1}{1.18}$ or .847458, $\frac{1}{1.10}$ or .909091, and $\frac{1}{1.02}$ or .980392. One year from now, however, the three-year zero will be a two-year zero with one year of duration risk. Therefore, its price at that time will be determined by using the risk-neutral rates of

[5] The reader should keep in mind that a two-year zero has one year of interest rate risk only in this stylized example: it has been assumed that rates can move only once a year. In reality, rates can move at any time, so a two-year zero has two full years of interest rate risk.

[6] A three-year zero has two years of interest rate risk only in this stylized example. See the previous footnote.

[7] Once again, this is an artifact of this example in which rates change only once a year.

14.20% and 6.20%. In particular, the two possible prices of the three-year zero in one year are

$$.769067 = \frac{.5(.847458 + .909091)}{1.142} \quad \textbf{(12.13)}$$

and

$$.889587 = \frac{.5(.909091 + .980392)}{1.062} \quad \textbf{(12.14)}$$

Finally, then, the expected return of the three-year zero over the next year is

$$\frac{.5(.769067 + .889587) - .751184}{.751184} = 10.40\% \quad \textbf{(12.15)}$$

To summarize, in order to compensate investors for two years of duration risk, the return on the three-year zero is 40 basis points above a one-year zero's certain return of 10%.

Continuing with the assumption of 400-basis-point volatility, Figure 12-3 graphs the term structure of spot rates for three cases: no risk premium, a risk premium of 20 basis points per year of duration risk, and a risk premium of 40 basis points. In the case of no risk premium, the term structure of spot rates is downward-sloping due to convexity. A risk premium of 20 basis points pushes up spot rates while convexity pulls them down. In the short end, the risk premium effect dominates and the term structure is mildly upward-sloping. In the long end, the convexity effect dominates and the term structure is mildly downward-sloping. The next section clarifies why risk premium tends to dominate in the short end while convexity tends to dominate in the long end. Finally, a risk premium as large as 40 basis points dominates the convexity effect and the term structure of spot rates is upward-sloping. The convexity effect is still evident, however, from the fact that the curve increases more rapidly from one to two years than from two to three years.

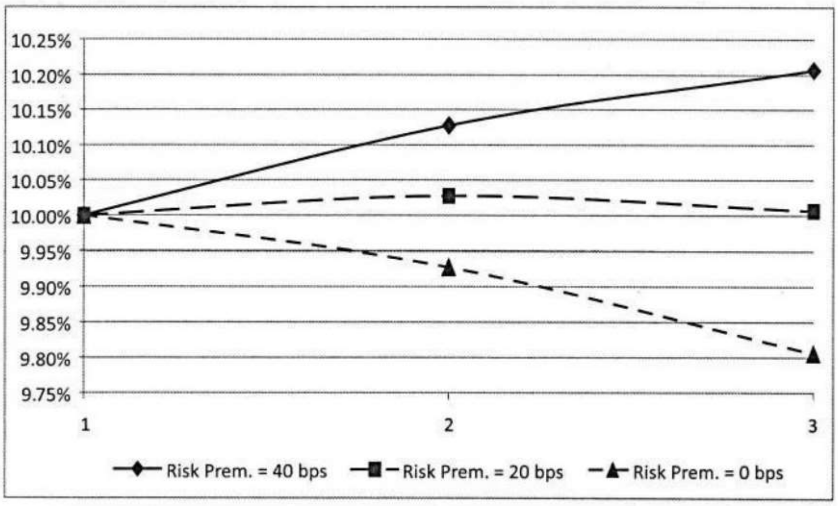


FIGURE 12-3 Volatility, risk premium, and the shape of the term structure in three-date binomial models.

Just as the section on volatility uses unrealistically high levels of volatility to illustrate its effects, this section uses unrealistically high levels of the risk premium to illustrate its effects. The application at the end of this chapter focuses on reasonable magnitudes for the various effects in the context of the USD and JPY swap markets.

Before closing this section, a few remarks on the sources of an interest rate risk premium are in order. Asset pricing theory (e.g., the Capital Asset Pricing Model, or CAPM) teaches that assets whose returns are positively correlated with aggregate wealth or consumption will earn a risk premium. Consider, for example, a traded stock index. That asset will almost certainly do well if the economy is doing well and poorly if the economy is doing poorly. But investors, as a group, already have a lot of exposure to the economy. To entice them to hold a little more of the economy in the form of a traded stock index requires the payment of a risk premium; i.e., the index must offer an expected return greater than the risk-free rate of return. On the other hand, say that there exists an asset that is negatively correlated with the economy. Holdings in that asset allow investors to reduce their exposure to the economy. As a result, investors would accept an expected return on that asset below the risk-free rate of return. That asset, in other words, would have a negative risk premium.

This section assumes that bonds with interest rate risk earn a risk premium. In terms of asset pricing theory, this is equivalent to assuming that bond returns are positively correlated with the economy or, equivalently, that falling interest rates are associated with good times. One argument supporting this assumption is that interest rates fall when inflation and expected inflation fall and that low inflation is correlated with good times.

The concept of a risk premium in fixed income markets has probably gained favor more for its empirical usefulness than for its theoretical solidity. On average, over the past 75 years, the term structure of interest rates has sloped upward.[8] While the market may from time to time expect that interest rates will rise, it is hard to believe that the market expects interest rates to rise on average. Therefore, expectations cannot explain a term structure of interest rates that, on average, slopes upward. Convexity, of course, leads to a downward-sloping term structure. Hence, of the three effects described in this chapter, only a positive risk premium can explain a term structure that, on average, slopes upward.

An uncomfortable fact, however, is that over earlier time periods the term structure has, on average, been flat.[9] Whether this means that an interest rate risk premium is a relatively recent phenomenon that is here to stay or that the experience of persistently upward-sloping curves is only partially due to a risk premium is a question beyond the scope of this book. In short, the theoretical and empirical questions with respect to the existence of an interest rate risk premium have not been settled.

[8] See, for example, Homer, S., and Richard Sylla, *A History of Interest Rates*, 3rd Edition, Revised, Rutgers University Press, 1996, pp. 394–409.

[9] Ibid.

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

The Art of Term Structure Models: Drift

13

Learning Objectives

After completing this reading you should be able to:

- Construct and describe the effectiveness of a short term interest rate tree assuming normally distributed rates, both with and without drift.
- Calculate the short-term rate change and standard deviation of the rate change using a model with normally distributed rates and no drift.
- Describe methods for addressing the possibility of negative short-term rates in term structure models.
- Construct a short-term rate tree under the Ho-Lee Model with time-dependent drift.
- Describe uses and benefits of the arbitrage-free models and assess the issue of fitting models to market prices.
- Describe the process of constructing a simple and recombining tree for a short-term rate under the Vasicek Model with mean reversion.
- Calculate the Vasicek Model rate change, standard deviation of the rate change, expected rate in *T* years, and half-life.
- Describe the effectiveness of the Vasicek Model.

Excerpt is Chapter 9 of Fixed Income Securities, *Third Edition, by Bruce Tuckman.*

Chapters 11 and 12 show that assumptions about the true and risk-neutral short-term rate processes determine the term structure of interest rates and the prices of fixed income derivatives. The goal of this chapter is to describe the most common building blocks of short-term rate models. Selecting and rearranging these building blocks to create suitable models for the purpose at hand is the art of term structure modeling.

This chapter begins with an extremely simple model with no drift and normally distributed rates. The next sections add and discuss the implications of alternate specifications of the drift: a constant drift, a time-deterministic shift, and a mean-reverting drift.

MODEL 1: NORMALLY DISTRIBUTED RATES AND NO DRIFT

The particularly simple model of this section will be called Model 1. The continuously compounded, instantaneous rate r_t is assumed to evolve according to the following equation:

$$dr = \sigma dw \tag{13.1}$$

The quantity dr denotes the change in the rate over a small time interval, dt, measured in years; σ denotes the annual *basis-point volatility* of rate changes; and dw denotes a normally distributed random variable with a mean of zero and a standard deviation of $\sqrt{dt}$.[1]

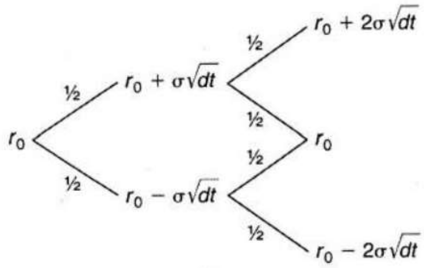
Say, for example, that the current value of the short-term rate is 6.18%, that volatility equals 113 basis points per year, and that the time interval under consideration is one month or $\frac{1}{12}$ years. Mathematically, $r_0 = 6.18\%$; $\sigma = 1.13\%$; and $dt = \frac{1}{12}$. A month passes and the random variable dw, with its zero mean and its standard deviation of $\sqrt{\frac{1}{12}}$ or .2887, happens to take on a value of .15. With these values, the change in the short-term rate given by (13.1) is

$$dr = 1.13\% \times .15 = .17\% \tag{13.2}$$

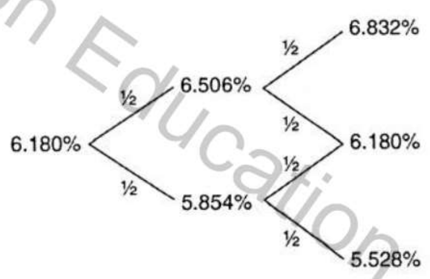
or 17 basis points. Since the short-term rate started at 6.18%, the short-term rate after a month is 6.35%.

Since the expected value of dw is zero, (13.1) says that the expected change in the rate, or the drift, is zero. Since the standard deviation of dw is $\sqrt{dt}$, the standard deviation of the change in the rate is $\sigma\sqrt{dt}$. For the sake of brevity, the standard deviation of the change in the rate will be referred to as simply the standard deviation of the rate. Continuing with the numerical example, the process (13.1) says that the drift is zero and that the standard deviation of the rate is $\sigma\sqrt{dt}$, which is $1.13\% \times \sqrt{\frac{1}{12}} = .326\%$ or 32.6 basis points per month.

A rate tree may be used to approximate the process (13.1). A tree over dates 0 to 2 takes the following form:



In the case of the numerical example, substituting the sample values into the tree gives the following:



To understand why these trees are representations of the process (13.1), consider the transition from date 0 to date 1. The change in the interest rate in the up-state is $\sigma\sqrt{dt}$ and the change in the down-state is $-\sigma\sqrt{dt}$. Therefore, with the probabilities given in the tree, the expected change in the rate, often denoted $E[dr]$, is

$$E[dr] = .5 \times \sigma\sqrt{dt} + .5 \times -\sigma\sqrt{dt} = 0 \tag{13.3}$$

The variance of the rate, often denoted $V[dr]$, from date 0 to date 1 is computed as follows:

$$\begin{aligned} V[dr] &= E[dr^2] - \{E[dr]\}^2 \\ &= .5 \times (\sigma\sqrt{dt})^2 + .5 \times (-\sigma\sqrt{dt})^2 - 0 \\ &= \sigma^2 dt \end{aligned} \tag{13.4}$$

Note that the first line of (13.4) follows from the definition of variance. Since the variance is $\sigma^2 dt$, the standard deviation, which is the square root of the variance, is $\sigma\sqrt{dt}$.

[1] It is beyond the mathematical scope of the text to explain why the random variable dw is denoted as a change. But the text uses this notation since it is the convention of the field.

Equations (13.3) and (13.4) show that the drift and volatility implied by the tree match the drift and volatility of the interest rate process (13.1). The process and the tree are not identical because the random variable in the process, having a normal distribution, can take on any value while a single step in the tree leads to only two possible values. In the example, when *dw* takes on a value of .15, the short rate changes from 6.18% to 6.35%. In the tree, however, the only two possible rates are 6.506% and 5.854%. Nevertheless, as shown in Chapter 10, after a sufficient number of time steps the branches of the tree used to approximate the process (13.1) will be numerous enough to approximate a normal distribution. Figure 13-1 shows the distribution of short rates after one year, or the *terminal distribution* after one year, in Model 1 with $r_0 = 6.18\%$ and $\sigma = 1.13\%$. The tick marks on the horizontal axis are one standard deviation apart from one another.

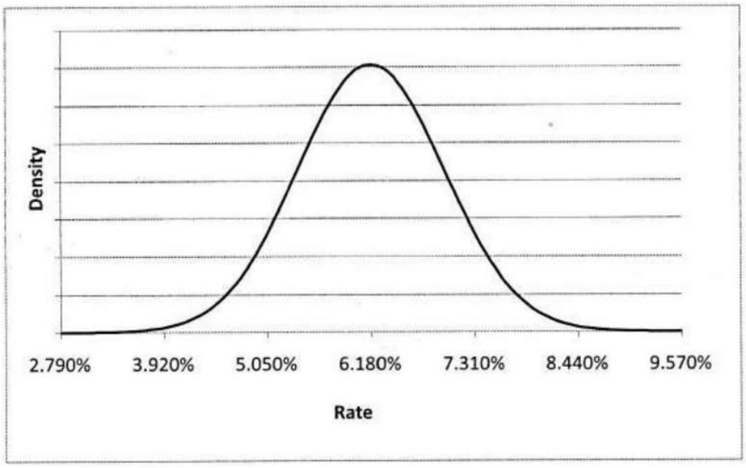


FIGURE 13-1 Distribution of short rates after one year, Model 1.

Models in which the terminal distribution of interest rates has a normal distribution, like Model 1, are called *normal* or *Gaussian* models. One problem with these models is that the short-term rate can become negative. A negative short-term rate does not make much economic sense because people would never lend money at a negative rate when they can hold cash and earn a zero rate instead.[2] The distribution in Figure 13-1, drawn to encompass three standard deviations above and below the mean, shows that over a horizon of one year the interest rate process will almost certainly not exhibit negative interest rates. The probability that the short-term rate in the process (13.1) becomes negative, however, increases with the horizon. Over 10 years, for example, the standard deviation of the terminal distribution in the numerical example is $1.13\% \times \sqrt{10}$ or 3.573%. Starting with a short-term rate of 6.18%, a random negative shock of only $^{6.18\%}/_{3.573\%}$ or 1.73 standard deviations would push rates below zero.

The extent to which the possibility of negative rates makes a model unusable depends on the application. For securities whose value depends mostly on the average path of the interest rate, like coupon bonds, the possibility of negative rates typically does not rule out an otherwise desirable model. For securities that are asymmetrically sensitive to the probability of low interest rates, however, using a normal model could be dangerous. Consider the extreme example of a 10-year call option to buy a long-term coupon bond at a yield of 0%. The model of this section would assign that option much too high a value because the model assigns too much probability to negative rates.

The challenge of negative rates for term structure models is much more acute, of course, when the current level of rates is low, as it is at the time of this writing. Changing the distribution of interest rates is one solution. To take but one of many examples, lognormally distributed rates, as will be seen in Chapter 14, cannot become negative. As will become clear later in that chapter, however, building a model around a probability distribution that rules out negative rates or makes them less likely may result in volatilities that are unacceptable for the purpose at hand.

Another popular method of ruling out negative rates is to construct rate trees with whatever distribution is desired, as done in this section, and then simply set all negative rates to zero.[3] In this methodology, rates in the original tree are called the shadow rates of interest while the rates in the adjusted tree could be called the observed rates of interest. When the observed rate hits zero, it can remain

[2] Actually, the interest rate can be slightly negative if a security or bank account were safer or more convenient than holding cash.

[3] Fischer Black, "Interest Rates as Options," *Journal of Finance*, Vol. 50, 1995, pp. 1371–1376.

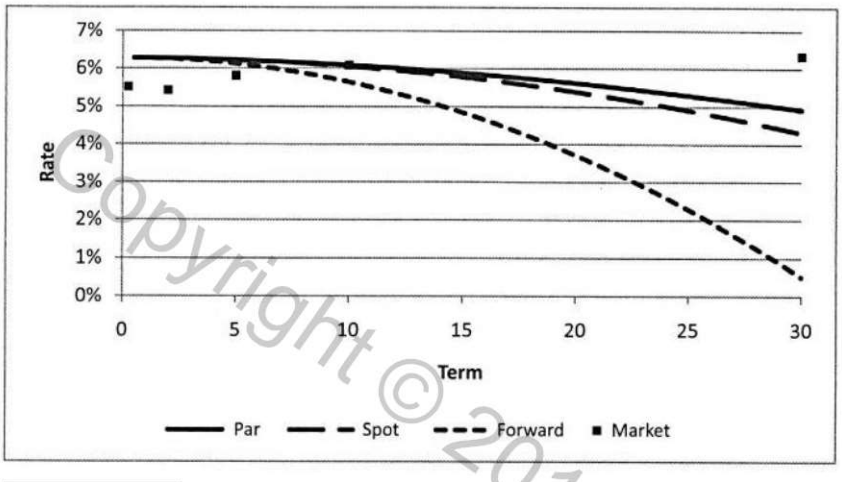


FIGURE 13-2 Rate curves from Model 1 and selected market swap rates, February 16, 2001.

TABLE 13-1 Convexity Effects on Par Rates in a Parameterization of Model 1

Term (years)	Convexity (bps)
2	−0.8
5	−5.1
10	−18.8
30	−135.3

there for a while until the shadow rate crosses back to a positive rate. The economic justification for this framework is that the observed interest rate should be constrained to be positive only because investors have the alternative of investing in cash. But the shadow rate, the result of aggregate savings, investment, and consumption decisions, may very well be negative. Of course, the probability distribution of the observed interest rate is not the same as that of the originally postulated shadow rate. The change, however, is localized around zero and negative rates. By contrast, changing the form of the probability distribution changes dynamics across the entire range of rates.

Returning now to Model 1, the techniques of Chapter 11 may be used to price fixed coupon bonds. Figure 13-2 graphs the semiannually compounded par, spot, and forward rate curves for the numerical example along with data from U.S. dollar swap par rates. The initial value of the short-term rate in the example, 6.18%, is set so that the model and market 10-year, semiannually compounded par rates are equal at 6.086%. All of the other data points shown are quite different from their model values. The desirability of fitting market data exactly is discussed in its own section, but Figure 13-2 clearly demonstrates that the simple model of this section does not have enough flexibility to capture the simplest of term structure shapes.

The model term structure is downward-sloping. As the model has no drift, rates decline with term solely because of convexity. Table 13-1 shows the magnitudes of the convexity effect on par rates of selected terms.[4] The numbers are realistic in the sense that a volatility of 113 basis points a year is reasonable. In fact, the volatility of the 10-year swap rate on the data date, as implied by options markets, was 113 basis points. The convexity numbers are not necessarily realistic, however, because, as this chapter will demonstrate, the magnitude of the convexity effect depends on the model and Model 1 is almost certainly not the best model of interest rate behavior.

The term structure of volatility in Model 1 is constant at 113 basis points per year. In other words, the standard deviation of changes in the par rate of any maturity is 113 basis points per year. As shown in Figure 13-3, this implication fails to capture the implied volatility structure in the market. The volatility data in Figure 13-3 show that the term structure of volatility is humped, i.e., that volatility initially rises with term but eventually declines. As this shape is a feature of fixed income markets, it will be revisited again in this chapter and in Chapter 14.

The last aspect of this model to be analyzed is its factor structure. The model's only factor is the short-term rate. If this rate increases by 10 semiannually compounded basis points, how would the term structure change? In this simple model, the answer is that all rates would increase by 10 basis points. (See the closed-form solution for spot rates in Model 1 in the Appendix in Chapter 14). Therefore, Model 1 is a model of parallel shifts.

[4] The convexity effect is the difference between the par yield in the model with its assumed volatility and the par yield in the same structural model but with a volatility of zero.

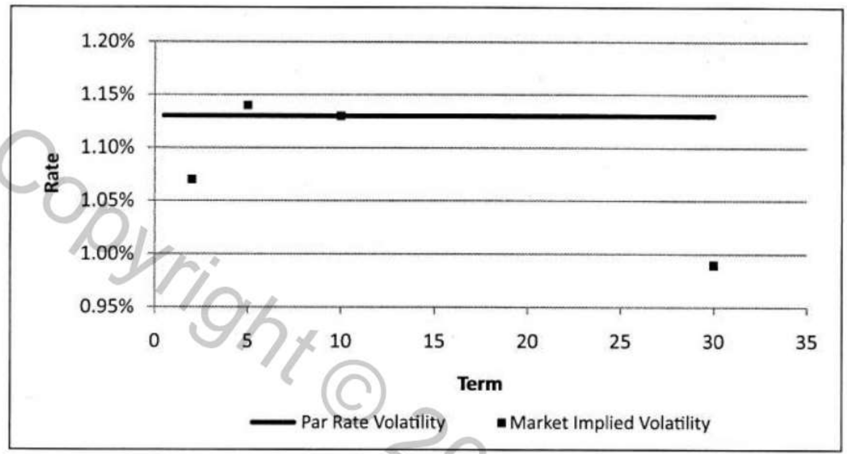


FIGURE 13-3 Par rate volatility from Model 1 and selected implied volatilities, February 16, 2001.

MODEL 2: DRIFT AND RISK PREMIUM

The term structures implied by Model 1 always look like Figure 13-2: relatively flat for early terms and then downward sloping. Chapter 12 pointed out that the term structure tends to slope upward and that this behavior might be explained by the existence of a risk premium. The model of this section, to be called Model 2, adds a drift to Model 1, interpreted as a risk premium, in order to obtain a richer model in an economically coherent way.

The dynamics of the risk-neutral process in Model 2 are written as

$$dr = \lambda dt + \sigma dw \tag{13.5}$$

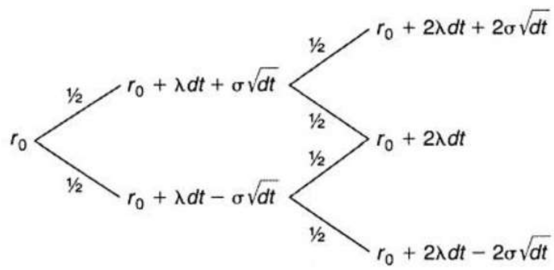
The process (13.5) differs from that of Model 1 by adding a drift to the short-term rate equal to λdt. For this section, consider the values $r_0 = 5.138\%$, $\lambda = .229\%$, and $\sigma = 1.10\%$. If the realization of the random variable dw is again .15 over a month, then the change in rate is

$$dr = .229\% \times \frac{1}{12} + 1.10\% \times .15 = .1841\% \tag{13.6}$$

Starting from 5.138%, the new rate is 5.322%.

The drift of the rate is $.229\% \times \frac{1}{12}$ or 1.9 basis points per month, and the standard deviation is $1.10\% \times \sqrt{\frac{1}{12}}$ or 31.75 basis points per month. As discussed in Chapter 12, the drift in the risk-neutral process is a combination of the true expected change in the interest rate and of a risk premium. A drift of 1.9 basis points per month may arise because the market expects the short-term rate to increase by 1.9 basis points a month, because the short-term rate is expected to increase by one basis point with a risk premium of .9 basis points, or because the short-term rate is expected to fall by .1 basis points with a risk premium of two basis points.

The tree approximating this model is



It is easy to verify that the drift and standard deviation of the tree match those of the process (13.5).

The terminal distribution of the numerical example of this process after one year is normal with a mean of $5.138\% + 1 \times .229\%$ or 5.367% and a standard deviation of 110 basis points. After 10 years, the terminal distribution is normal with a mean of $5.138\% + 10 \times .229\%$ or 7.428% and a standard deviation of $1.10\% \times \sqrt{10}$ or 347.9 basis points. Note that the constant drift, by raising the mean of the terminal distribution, makes it less likely that the risk-neutral process will exhibit negative rates.

Figure 13-4 shows the rate curves in this example along with par swap rate data. The values of r_0 and λ are calibrated to match the 2- and 10-year par swap rates, while the value of σ is chosen to be the average implied volatility of the 2- and 10-year par rates. The results are satisfying in that the resulting curve can match the data much more closely than did the curve of Model 1 shown in Figure 13-2. Slightly unsatisfying is the relatively high value of λ required. Interpreted as a risk premium alone, a value of .229% with a volatility of 110 basis points implies a relatively high Sharpe ratio of about .21. On the other hand, interpreting λ as a combination of true drift and risk premium is difficult in the long end where, as argued in Chapter 12, it is difficult to make a case for rising expected rates. These interpretive difficulties arise because Model 2 is still not flexible enough to explain the shape of the term structure in an economically meaningful way. In fact, the use of r_0 and λ to match the 2- and 10-year rates in this relatively inflexible model may explain why the model curve overshoots the 30-year par rate by about 25 basis points.

Moving from Model 1 with zero drift to Model 2 with a constant drift does not qualitatively change the term structure of volatility, the magnitude of convexity effects, or the parallel-shift nature of the model.

Models 1 and 2 would be called equilibrium models because no effort has been made to match the initial term structure closely. The next section presents a generalization of Model 2 that is in the class of arbitrage-free models.

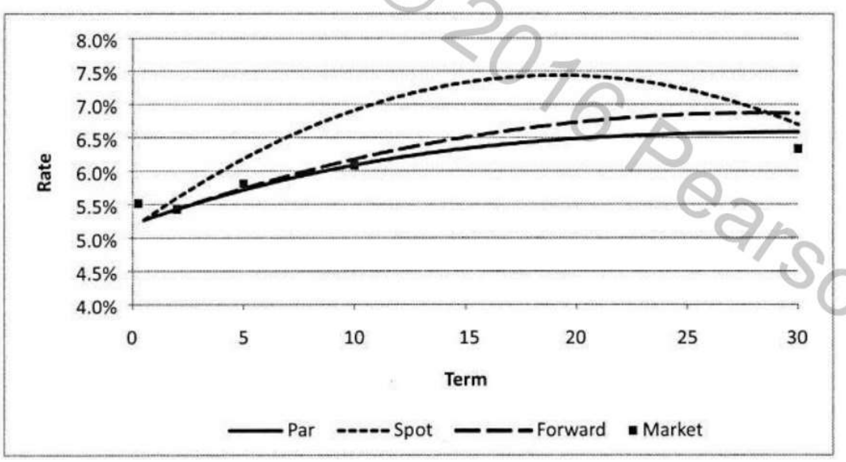


FIGURE 13-4 Rate curves from Model 2 and selected market swap rates, February 16, 2001.

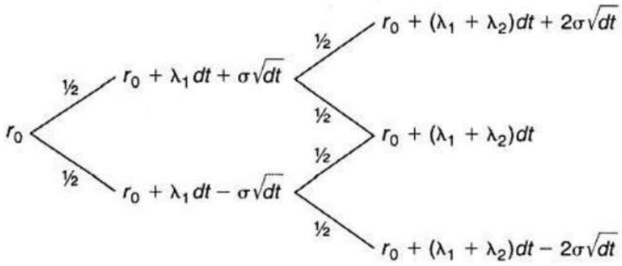
THE HO-LEE MODEL: TIME-DEPENDENT DRIFT

The dynamics of the risk-neutral process in the Ho-Lee model are written as

$$dr = \lambda_t dt + \sigma dw \tag{13.7}$$

In contrast to Model 2, the drift here depends on time. In other words, the drift of the process may change from date to date. It might be an annualized drift of −20 basis points over the first month, of 20 basis points over the second month, and so on. A drift that varies with time is called a *time-dependent* drift. Just as with a constant drift, the time-dependent drift over each time period represents some combination of the risk premium and of expected changes in the short-term rate.

The flexibility of the Ho-Lee model is easily seen from its corresponding tree:



The free parameters λ_1 and λ_2 may be used to match the prices of securities with fixed cash flows. The procedure may be described as follows. With $dt = 1/12$, set r_0 equal to the one-month rate. Then find λ_1 such that the model produces a two-month spot rate equal to that in the market. Then find λ_2 such that the model produces a three-month spot rate equal to that in the market. Continue in this fashion until the tree ends. The procedure is very much like that used to construct the trees in Chapter 11. The only difference is that Chapter 11 adjusts the probabilities to match the spot rate curve while this section adjusts the rates. As it turns out, the two procedures are equivalent so long as the step size is small enough.

The rate curves resulting from this model match all the rates that are input into the model. Just as adding a constant drift to Model 1 to obtain Model 2 does not affect the shape of the term structure of volatility nor the parallel-shift characteristic of the model, adding a time-dependent drift does not change these features either.

DESIRABILITY OF FITTING TO THE TERM STRUCTURE

The desirability of matching market prices is the central issue in deciding between arbitrage-free and equilibrium models. Not surprisingly, the choice depends on the purpose of building the model in the first place.

One important use of arbitrage-free models is for quoting the prices of securities that are not actively traded based on the prices of more liquid securities. A customer might ask a swap desk to quote a rate on a swap to a particular date, say three years and four months away, while liquid market prices might be observed only for three- and four-year swaps, or sometimes only for two- and five-year swaps. In this situation, the swap desk may price the odd-maturity swap using an arbitrage-free model essentially as a means of interpolating between observed market prices.

Interpolating by means of arbitrage-free models may very well be superior to other curve-fitting methods, from linear interpolation to more sophisticated approaches. The potential superiority of arbitrage-free models arises from their being based on economic and financial reasoning. In an arbitrage-free model, the expectations and risk premium built into neighboring swap rates and the convexity implied by the model's volatility assumptions are used to compute, for example, the three-year and four-month swap rate. In a purely mathematical curve fitting technique, by contrast, the chosen functional form heavily determines the intermediate swap rate. Selecting linear or quadratic interpolation, for example, results in intermediate swap rates with no obvious economic or financial justification. This potential superiority of arbitrage-free models depends crucially on the validity of the assumptions built into the models. A poor volatility assumption, for example, resulting in a poor estimate of the effect of convexity, might make an arbitrage-free model perform worse than a less financially sophisticated technique.

Another important use of arbitrage-free models is to value and hedge derivative securities for the purpose of making markets or for proprietary trading. For these purposes, many practitioners wish to assume that some set of underlying securities is priced fairly. For example, when trading an option on a 10-year bond, many practitioners assume that the 10-year bond is itself priced fairly. (An analysis of the fairness of the bond can always be done separately.) Since arbitrage-free models match the prices of many traded securities by construction, these models are ideal for the purpose of pricing derivatives given the prices of underlying securities.

That a model matches market prices does not necessarily imply that it provides fair values and accurate hedges for derivative securities. The argument for fitting models to market prices is that a good deal of information about the future behavior of interest rates is incorporated into market prices, and, therefore, a model fitted to those prices captures that interest rate behavior. While this is a perfectly reasonable argument, two warnings are appropriate. First, a mediocre or bad model cannot be rescued by calibrating it to match market prices. If, for example, the parallel shift assumption is not a good enough description of reality for the application at hand, adding a time-dependent drift to a parallel shift model so as to match a set of market prices will not make the model any more suitable for that application. Second, the argument for fitting to market prices assumes that those market prices are fair in the context of the model. In many situations, however, particular securities, particular classes of securities, or particular maturity ranges of securities have been distorted due to supply and demand imbalances, taxes, liquidity differences, and other factors unrelated to interest rate models. In these cases, fitting to market prices will make a model worse by attributing these outside factors to the interest rate process. If, for example, a large bank liquidates its portfolio of bonds or swaps with approximately seven years to maturity and, in the process, depresses prices and raises rates around that maturity, it would be incorrect to assume that expectations of rates seven years in the future have risen. Being careful with the word *fair*, the seven-year securities in this example are fair in the sense that liquidity considerations at a particular time require their prices to be relatively low. The seven-year securities are not fair, however, with respect to the expected evolution of interest rates and the market risk premium. For this reason, in fact, investors and traders might buy these relatively cheap bonds or swaps and hold them past the liquidity event in the hope of selling at a profit.

Another way to express the problem of fitting the drift to the term structure is to recognize that the drift of a risk-neutral process arises only from expectations and risk premium. A model that assumes one drift from years 15 to 16 and another drift from years 16 to 17 implicitly assumes one of two things. First, the expectation today of the one-year rate in 15 years differs from the expectation today of the one-year rate in 16 years. Second, the risk premium in 15 years differs in a particular way from the risk premium in 16 years. Since neither of these assumptions is particularly plausible, a fitted drift that changes dramatically from one year to the next is likely to be erroneously attributing non-interest rate effects to the interest rate process.

If the purpose of a model is to value bonds or swaps relative to one another, then taking a large number of bond or swap prices as given is clearly inappropriate: arbitrage-free models, by construction, conclude that all of these bond or swap prices are fair relative to one another. Investors wanting to choose among securities, market makers looking to pick up value by strategically selecting hedging securities, or traders looking to profit from temporary mispricings must, therefore, rely on equilibrium models.

Having starkly contrasted arbitrage-free and equilibrium models, it should be noted that, in practice, there need not be a clear line between the two approaches. A model might posit a deterministic drift for a few years to reflect relatively short-term interest rate forecasts and posit a constant drift from then on. Another model might take the prices of 2-, 5-, 10- and 30-year bond or swap rates as given, thus assuming that the most liquid securities are fair while allowing the model to value other securities. The proper blending of the arbitrage-free and equilibrium approaches is an important part of the art of term structure modeling.

THE VASICEK MODEL: MEAN REVERSION

Assuming that the economy tends toward some equilibrium based on such fundamental factors as the productivity of capital, long-term monetary policy, and so on, short-term rates will be characterized by *mean reversion.* When the short-term rate is above its long-run equilibrium value, the drift is negative, driving the rate down toward this long-run value. When the rate is below its equilibrium value, the drift is positive, driving the rate up toward this value. In addition to being a reasonable assumption about short rates,[5] mean reversion enables a model to capture several features of term structure behavior in an economically intuitive way.

The risk-neutral dynamics of the Vasicek model[6] are written as

$$dr = k(\theta - r)dt + \sigma dw \qquad \textbf{(13.8)}$$

The constant θ denotes the long-run value or central tendency of the short-term rate in the risk-neutral process and the positive constant k denotes the speed of mean reversion. Note that in this specification, the greater the difference between r and θ, the greater the expected change in the short-term rate toward θ.

Because the process (13.8) is the risk-neutral process, the drift combines both interest rate expectations and risk premium. Furthermore, market prices do not depend on how the risk-neutral drift is divided between the two. Nevertheless, in order to understand whether or not the parameters of a model make sense, it is useful to make assumptions sufficient to separate the drift and the risk premium. Assuming, for example, that the true interest rate process exhibits mean reversion to a long-term value r_∞ and, as assumed previously, that the risk premium enters into the risk-neutral process as a constant drift, the Vasicek model takes the following form:

$$\begin{aligned} dr &= k(r_\infty - r)dt + \lambda dt + \sigma dw \\ &= k\left(\left[r_\infty + \frac{\lambda}{k}\right] - r\right)dt + \sigma dw \end{aligned} \qquad \textbf{(13.9)}$$

The process in (13.8) is identical to that in (13.9) so long as

$$\theta \equiv r_\infty + \frac{\lambda}{k} \qquad \textbf{(13.10)}$$

Note that very many combinations of r_∞ and λ give the same θ and, through the risk-neutral process (13.8), the same market prices.

For the purposes of this section, let $k = .025$, $\sigma = 126$ basis points per year, $r_\infty = 6.179\%$, and $\lambda = .229\%$. According to (13.10), then, $\theta = 15.339\%$. With these parameters, the process (13.8) says that over the next month the expected change in the short rate is

$$.025 \times (15.339\% - 5.121\%)\frac{1}{12} = .0213\% \qquad \textbf{(13.11)}$$

or 2.13 basis points. The volatility over the next month is $126 \times \sqrt{\frac{1}{12}}$ or 36.4 basis points.

Representing this process with a tree is not quite so straightforward as the simpler processes described

[5] While reasonable, mean reversion is a strong assumption. Long time series of interest rates from relatively stable markets might display mean reversion because there happened to be no catastrophe over the time period, that is, precisely because a long time series exists. Hyperinflation, for example, is not consistent with mean reversion and results in the destruction of a currency and its associated interest rates. When mean reversion ends, the time series ends. In short, the most severe critics of mean reversion would say that interest rates mean revert until they don't.

[6] O. Vasicek, "An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5, 1977, pp. 177–188. It is appropriate to add that this paper started the literature on short-term rate models. The particular dynamics of the model described in this section, which is commonly known as the Vasicek model, is a very small part of the contribution of that paper.

previously because the most obvious representation leads to a nonrecombining tree. Over the first time step,

$$5.121\% \xrightarrow{1/2} 5.121\% + \frac{.025(15.339\% - 5.121\%)}{12} + \frac{.0126}{\sqrt{12}} = 5.5060\%$$

$$5.121\% \xrightarrow{1/2} 5.121\% + \frac{.025(15.339\% - 5.121\%)}{12} - \frac{.0126}{\sqrt{12}} = 4.7786\%$$

To extend the tree from date 1 to date 2, start from the up state of 5.5060%. The tree branching from there is

$$5.5060\% \xrightarrow{1/2} 5.506\% + \frac{.025(15.339\% - 5.5060\%)}{12} + \frac{.0126}{\sqrt{12}} = 5.8902\%$$

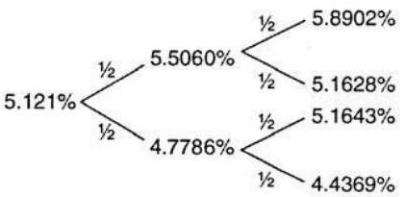
$$5.5060\% \xrightarrow{1/2} 5.506\% + \frac{.025(15.339\% - 5.5060\%)}{12} - \frac{.0126}{\sqrt{12}} = 5.1628\%$$

while the tree branching from the date 1 down-state of 4.7786% is

$$4.7786\% \xrightarrow{1/2} 4.7786\% + \frac{.025(15.339\% - 4.7786\%)}{12} + \frac{.0126}{\sqrt{12}} = 5.1643\%$$

$$4.7786\% \xrightarrow{1/2} 4.7786\% + \frac{.025(15.339\% - 4.7786\%)}{12} - \frac{.0126}{\sqrt{12}} = 4.4369\%$$

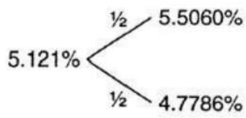
To summarize, the most straightforward tree representation of (13.8) takes the following form:



This tree does not recombine since the drift increases with the difference between the short rate and θ. Since 4.7786% is further from θ than 5.5060%, the drift from 4.7786% is greater than the drift from 5.5060%. In this model, the volatility component of an up move followed by a down move does perfectly cancel the volatility component of a down move followed by an up move. But since the drift from 4.7786% is greater, the move up from 4.7786% produces a larger short-term rate than a move down from 5.5060%.

There are many ways to represent the Vasicek model with a recombining tree. One method is presented here, but it is beyond the scope of this book to discuss the numerical efficiency of the various possibilities.

The first time step of the tree may be taken as shown previously:



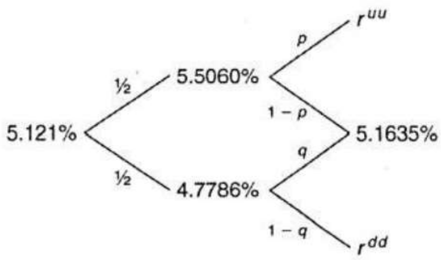
Next, fix the center node of the tree on date 2. Since the expected perturbation due to volatility over each time step is zero, the drift alone determines the expected value of the process after each time step. After the first time step, the expected value is

$$5.121\% + .025\,(15.339\% - 5.121\%)\,\frac{1}{12} = 5.1423\% \tag{13.12}$$

After the second time step, the expected value is

$$5.1423\% + .025\,(15.339\% - 5.1423\%)\,\frac{1}{12} = 5.1635\% \tag{13.13}$$

Take this value as the center node on date 2 of the recombining tree:



The parts of the tree to be solved for, namely, the missing probabilities and interest rate values, are given variable names.

According to the process (13.8) and the parameter values set in this section, the expected rate and standard deviation of the rate from 5.5060% are, respectively,

$$5.5060\% + .025\,(15.339\% - 5.5060\%)\,\frac{1}{12} = 5.5265\% \quad \textbf{(13.14)}$$

and

$$1.26\%\sqrt{\frac{1}{12}} = .3637\% \quad \textbf{(13.15)}$$

For the recombining tree to match this expectation and standard deviation, it must be the case that

$$p \times r^{uu} + (1 - p) \times 5.1635\% = 5.5265\% \quad \textbf{(13.16)}$$

and, by the definition of standard deviation,

$$\sqrt{p(r^{uu} - 5.5265\%)^2 + (1-p)(5.1635\% - 5.5265\%)^2} = .3637\% \quad \textbf{(13.17)}$$

Solving Equations (13.16) and (13.17), $r^{uu} = 5.8909\%$ and $p = .4990$.

The same procedure may be followed to compute r^{dd} and q. The expected rate from 4.7786% is

$$4.7786\% + .025\,(15.339\% - 4.7786\%)\,\frac{1}{12} = 4.8006\% \quad \textbf{(13.18)}$$

and the standard deviation is again 36.37 basis points. Starting from 4.7786%, then, it must be the case that

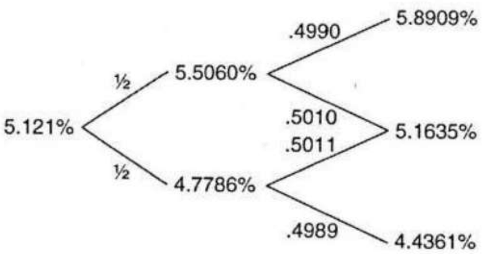
$$q \times 5.1635\% + (1 - q) \times r^{dd} = 4.8006\% \quad \textbf{(13.19)}$$

and

$$\sqrt{q(5.1635\% - 4.8006\%)^2 + (1-q)(r^{dd} - 4.8006\%)^2} = .3637\% \quad \textbf{(13.20)}$$

Solving Equations (13.19) and (13.20), $r^{dd} = 4.4361\%$ and $q = .5011$.

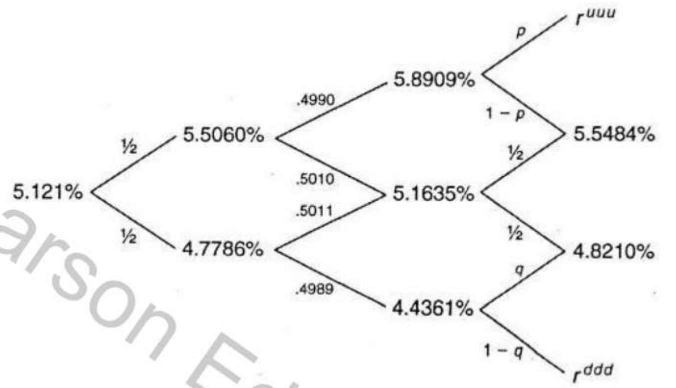
Putting the results from the up- and down-states together, a recombining tree approximating the process (13.8) with the parameters of this section is



To extend the tree to the next date, begin again at the center. From the center node of date 2, the expected rate of the process is

$$5.1635\% + .025 \times (15.339\% - 5.1635\%)\,\frac{1}{12} = 5.1847\% \quad \textbf{(13.21)}$$

As in constructing the tree for date 1, adding and subtracting the standard deviation of .3637% to the average value 5.1847% (obtaining 5.5484% and 4.8210%) and using probabilities of 50% for up and down movements satisfy the requirements of the process at the center of the tree:



The unknown parameters can be solved for in the same manner as described in building the tree on date 2.

The text now turns to the effects of mean reversion on the term structure. Figure 13-5 illustrates the impact of mean reversion on the terminal, risk-neutral distributions of the short rate at different horizons. The expectation or mean

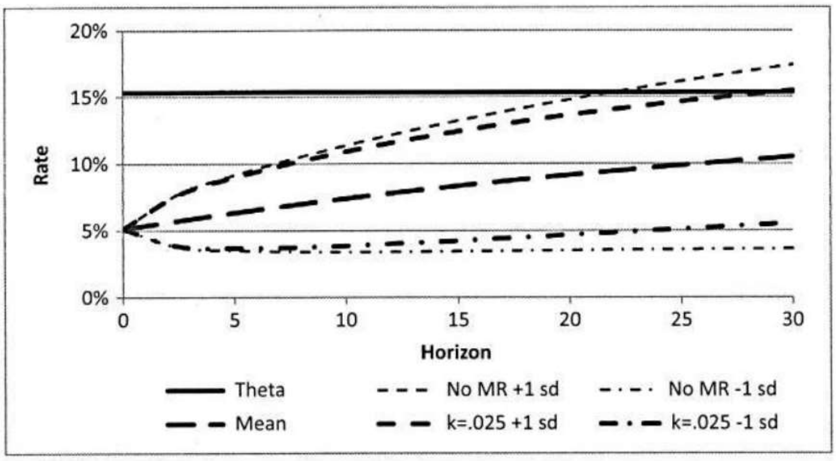


FIGURE 13-5 Mean reversion and the terminal distribution of short rates.

of the short-term rate as a function of horizon gradually rises from its current value of 5.121% toward its limiting value of θ = 15.339%. Because the mean-reverting parameter k = .025 is relatively small, the horizon expectation rises very slowly toward 15.339%. While mathematically beyond the scope of this book, it can be shown that the distance between the current value of a factor and its goal decays exponentially at the mean-reverting rate. Since the interest rate is currently 15.339% − 5.121% or 10.218% away from its goal, the distance between the expected rate at a 10-year horizon and the goal is

$$10.2180\% \times e^{-.025\times 10} = 7.9578\% \tag{13.22}$$

Therefore, the expectation of the rate in 10 years is 15.3390% − 7.9578% or 7.3812%.

For completeness, the expectation of the rate in the Vasicek model after T years is

$$r_0 e^{-kT} + \theta(1 - e^{-kT}) \tag{13.23}$$

In words, the expectation is a weighted average of the current short rate and its long-run value, where the weight on the current short rate decays exponentially at a speed determined by the mean-reverting parameter.

The mean-reverting parameter is not a particularly intuitive way of describing how long it takes a factor to revert to its long-term goal. A more intuitive quantity is the factor's *half-life*, defined as the time it takes the factor to progress half the distance toward its goal. In the example of this section, the half-life of the interest rate, τ, is given by the following equation:

$$(15.339\% - 5.121\%)\, e^{-.025\tau} = \frac{1}{2}(15.339\% - 5.121\%) \tag{13.24}$$

Solving,

$$e^{-.025\tau} = \frac{1}{2}$$

$$\tau = \frac{\ln(2)}{.025}$$

$$\tau = 27.73 \tag{13.25}$$

where ln (·) is the natural logarithm function. In words, the interest rate factor takes 27.73 years to cover half the distance between its starting value and its goal. This can be seen visually in Figure 13-5 where the expected rate 30 years from now is about halfway between its current value and θ. Larger mean-reverting parameters produce shorter half lives.

Figure 13-5 also shows one-standard deviation intervals around expectations both for the mean-reverting process of this section and for a process with the same expectation and the same σ but without mean reversion ("No MR"). The standard deviation of the terminal distribution of the short rate after T years in the Vasicek model is

$$\sqrt{\frac{\sigma^2}{2k}(1 - e^{-2kT})} \tag{13.26}$$

In the numerical example, with a mean-reverting parameter of .025 and a volatility of 126 basis points, the short rate in 10 years is normally distributed with an expected value of 7.3812%, derived earlier, and a standard deviation of

$$\sqrt{\frac{.0126^2}{2\times .025}(1 - e^{-2\times .025\times 10})} \tag{13.27}$$

or 353 basis points. Using the same expected value and σ but no mean reversion the standard deviation is $\sigma\sqrt{T} = 1.26\%\sqrt{10}$ or 398 basis points. Pulling the interest rate toward a long-term goal dampens volatility relative to processes without mean reversion, particularly at long horizons.

To avoid confusion in terminology, note that the mean-reverting model in this section sets volatility equal to 125 basis points "per year." Because of mean reversion, however, this does not mean that the standard deviation of the terminal distribution after T years increases with the square root of time. Without mean reversion, this is the case, as mentioned in the previous paragraph. With mean reversion, the standard deviation increases with horizon more slowly than that, producing a standard deviation of only 353 basis points after 10 years.

Figure 13-6 graphs the rate curves in this parameterization of the Vasicek model. The values of r_0 and θ were calibrated to match the 2- and 10-year par rates in the market. As a result, Figure 13-6 qualitatively resembles Figure 13-4. The mean reversion parameter might have been used to make the model fit the observed term structure more closely, but, as discussed in the next paragraph, this parameter was used to produce a particular term structure of volatility. In conclusion, Figure 13-6 shows that the model as calibrated in this section is probably not flexible enough to produce the range of term structures observed in practice.

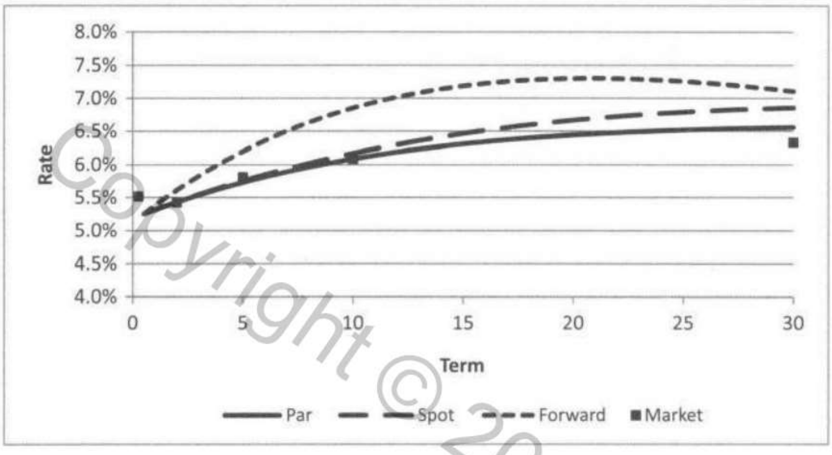


FIGURE 13-6 Rate curves from the Vasicek model and selected market swap rates, February 16, 2001.

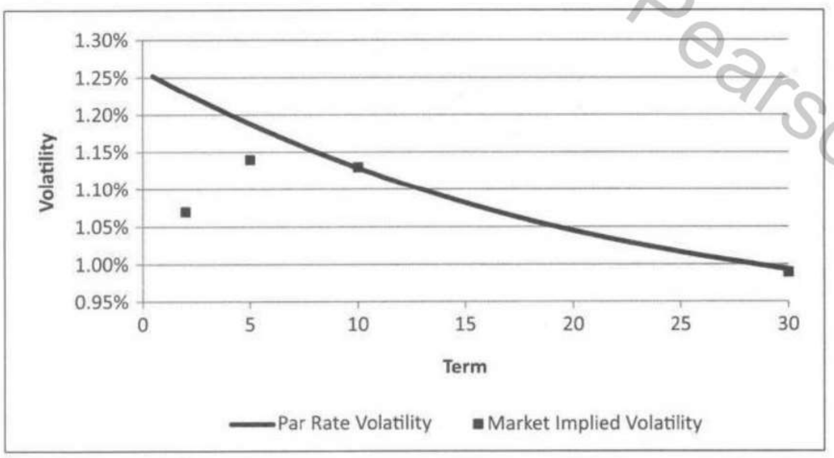


FIGURE 13-7 Par rate volatility from the Vasicek model and selected implied volatilities, February 16, 2001.

A model with mean reversion and a model without mean reversion result in dramatically different term structures of volatility. Figure 13-7 shows that the volatilities of par rates decline with term in the Vasicek model. In this example the mean reversion and volatility parameters are chosen to fit the implied 10- and 30-year volatilities. As a result, the model matches the market at those two terms but overstates the volatility for shorter terms. While Figure 13-7 certainly shows an improvement relative to the flat term structure of volatility shown in Figure 13-3, mean reversion in this model generates a term structure of volatility that slopes downward everywhere.

Since mean reversion lowers the volatility of longer-term par rates, it must also lower the impact of convexity on these rates. Table 13-2 reports the convexity effect at several terms. Recall that the convexity effects listed in Table 13-1 are generated from a model with no mean reversion and a volatility of 113 basis points per year. Since this section sets volatility equal to 126 basis points per year and since mean reversion is relatively slow, the convexity effects for terms up to 10 years are slightly larger in Table 13-2 than in Table 13-1. But by a term of 30 years the dampening effect of mean reversion on volatility manifests itself, and the convexity effect in the Vasicek model of about 75 basis points is substantially below the 135 basis point in the model without mean reversion.

Figure 13-8 shows the shape of the interest rate factor in a mean-reverting model, that is, how the spot rate curve is affected by a 10-basis point increase in the short-term rate. By definition, short-term rates rise by about 10 basis points but longer term rates are impacted less. The 30-year spot rate, for example, rises by only 7 basis points. Hence a model with mean reversion is not a parallel shift model.

The implications of mean reversion for the term structure of volatility and factor shape may be better understood by reinterpreting the assumption that short rates tend toward a long-term goal. Assuming that short rates move as a result of some news or shock to the economic system, mean reversion implies that the effect of this

TABLE 13-2 Convexity Effects on Par Rates in a Parameterization of the Vasicek Model

Term (years)	Convexity (bps)
2	−1.0
5	−5.8
10	−19.1
30	−74.7

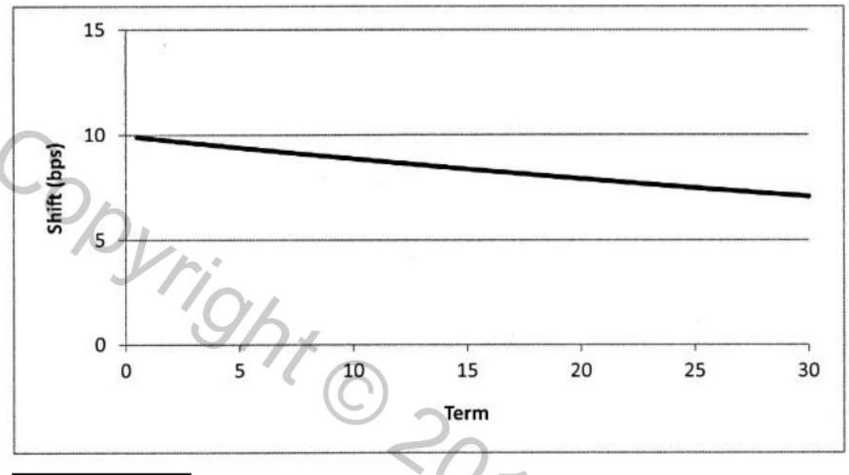


FIGURE 13-8 Sensitivity of spot rates in the Vasicek model to a 10-basis-point change in the factor.

shock eventually dissipates. After all, regardless of the shock, the short rate is assumed to arrive ultimately at the same long-term goal.

Economic news is said to be *long-lived* if it changes the market's view of the economy many years in the future. For example, news of a technological innovation that raises productivity would be a relatively long-lived shock to the system. Economic news is said to be *short-lived* if it changes the market's view of the economy in the near but not far future. An example of this kind of shock might be news that retail sales were lower than expected due to excessively cold weather over the holiday season. In this interpretation, mean reversion measures the length of economic news in a term structure model. A very low mean reversion parameter, i.e., a very long half-life, implies that news is long-lived and that it will affect the short rate for many years to come. On the other hand, a very high mean reversion parameter, i.e., a very short half-life, implies that news is short-lived and that it affects the short rate for a relatively short period of time. In reality, of course, some news is short-lived while other news is long-lived, a feature captured by the multi-factor Gauss+ model.

Interpreting mean reversion as the length of economic news explains the factor structure and the downward-sloping term structure of volatility in the Vasicek model. Rates of every term are combinations of current economic conditions, as measured by the short-term rate, and of long-term economic conditions, as measured by the long-term value of the short rate (i.e., θ). In a model with no mean reversion, rates are determined exclusively by current economic conditions. Shocks to the short-term rate affect all rates equally, giving rise to parallel shifts and a flat term structure of volatility. In a model with mean reversion, short-term rates are determined mostly by current economic conditions while longer-term rates are determined mostly by long-term economic conditions. As a result, shocks to the short rate affect short-term rates more than longer-term rates and give rise to a downward-sloping term structure of volatility and a downward-sloping factor structure.

The Art of Term Structure Models: Volatility and Distribution

14

Learning Objectives

After completing this reading you should be able to:

- Describe the short-term rate process under a model with time-dependent volatility.
- Calculate the short-term rate change and determine the behavior of the standard deviation of the rate change using a model with time-dependent volatility.
- Assess the efficacy of time-dependent volatility models.
- Describe the short-term rate process under the Cox-Ingersoll-Ross (CIR) and lognormal models.
- Calculate the short-term rate change and describe the basis point volatility using the CIR and lognormal models.
- Describe lognormal models with deterministic drift and mean reversion.

Excerpt is Chapter 10 of Fixed Income Securities, *Third Edition, by Bruce Tuckman.*

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

This chapter continues the presentation of the elements of term structure modeling, focusing on the volatility of interest rates and on models in which rates are not normally distributed.

TIME-DEPENDENT VOLATILITY: MODEL 3

Just as a time-dependent drift may be used to fit many bond or swap rates, a time-dependent volatility function may be used to fit many option prices. A particularly simple model with a time-dependent volatility function might be written as follows:

$$dr = \lambda(t)dt + \sigma(t)dw \quad \textbf{(14.1)}$$

Unlike the models presented in Chapter 13, the volatility of the short rate in Equation (14.1) depends on time. If, for example, the function $\sigma(t)$ were such that $\sigma(1) = 1.26\%$ and $\sigma(2) = 1.20\%$, then the volatility of the short rate in one year is 126 basis points per year while the volatility of the short rate in two years is 120 basis points per year.

To illustrate the features of time-dependent volatility, consider the following special case of (14.1) that will be called Model 3:

$$dr = \lambda(t)dt + \sigma e^{-\alpha t}dw \quad \textbf{(14.2)}$$

In (14.2), the volatility of the short rate starts at the constant σ and then exponentially declines to zero. Volatility could have easily been designed to decline to another constant instead of zero, but Model 3 serves its pedagogical purpose well enough.

Setting σ = 126 basis points and α = .025, Figure 14-1 graphs the standard deviation of the terminal distribution of the short rate at various horizons.[1] Note that the standard deviation rises rapidly with horizon at first but then rises more slowly. The particular shape of the curve depends, of course, on the volatility function chosen for (14.2), but very many shapes are possible with the more general volatility specification in (14.1).

Deterministic volatility functions are popular, particularly among market makers in interest rate options. Consider

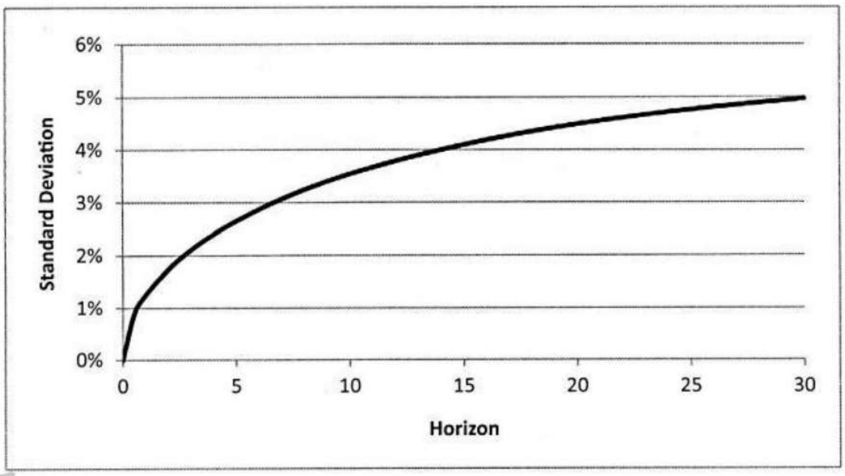


FIGURE 14-1 Standard deviation of terminal distributions of short rates in Model 3.

the example of *caplets*. At expiration, a caplet pays the difference between the short rate and a strike, if positive, on some notional amount. Furthermore, the value of a caplet depends on the distribution of the short rate at the caplet's expiration. Therefore, the flexibility of the deterministic functions $\lambda(t)$ and $\sigma(t)$ may be used to match the market prices of caplets expiring on many different dates.

The behavior of standard deviation as a function of horizon in Figure 14-1 resembles the impact of mean reversion on horizon standard deviation in Figure 13-5. In fact, setting the initial volatility and decay rate in Model 3 equal to the volatility and mean reversion rate of the numerical example of the Vasicek model, the standard deviations of the terminal distributions from the two models turn out to be identical. Furthermore, if the time-dependent drift in Model 3 matches the average path of rates in the numerical example of the Vasicek model, then the two models produce exactly the same terminal distributions.

While these parameterizations of the two models give equivalent terminal distributions, the models remain very different in other ways. As is the case for any model without mean reversion, Model 3 is a parallel shift model. Also, the term structure of volatility in Model 3 is flat. Since the volatility in Model 3 changes over time, the term structure of volatility is flat at levels that change over time, but it is still always flat.

The arguments for and against using time-dependent volatility resemble those for and against using a time-dependent drift. If the purpose of the model is to quote

[1] This result is presented without derivation.

fixed income options prices that are not easily observable, then a model with time-dependent volatility provides a means of interpolating from known to unknown option prices. If, however, the purpose of the model is to value and hedge fixed income securities, including options, then a model with mean reversion might be preferred for two reasons.

First, while mean reversion is based on the economic intuitions outlined earlier, time-dependent volatility relies on the difficult argument that the market has a forecast of short-term volatility in the distant future. A modification of the model that addresses this objection, by the way, is to assume that volatility depends on time in the near future and then settles at a constant.

Second, the downward-sloping factor structure and term structure of volatility in mean-reverting models capture the behavior of interest rate movements better than parallel shifts and a flat term structure of volatility. It may very well be that the Vasicek model does not capture the behavior of interest rates sufficiently well to be used for a particular valuation or hedging purpose. But in that case it is unlikely that a parallel shift model calibrated to match caplet prices will be better suited for that purpose.

THE COX-INGERSOLL-ROSS AND LOGNORMAL MODELS: VOLATILITY AS A FUNCTION OF THE SHORT RATE

The models presented so far assume that the basis-point volatility of the short rate is independent of the level of the short rate. This is almost certainly not true at extreme levels of the short rate. Periods of high inflation and high short-term interest rates are inherently unstable and, as a result, the basis-point volatility of the short rate tends to be high. Also, when the short-term rate is very low, its basis-point volatility is limited by the fact that interest rates cannot decline much below zero.

Economic arguments of this sort have led to specifying the basis-point volatility of the short rate as an increasing function of the short rate. The risk-neutral dynamics of the Cox-Ingersoll-Ross (CIR) model are

$$dr = k(\theta - r)dt + \sigma\sqrt{r}\,dw \tag{14.3}$$

Since the first term on the right-hand side of (14.3) is not a random variable and since the standard deviation of dw equals $\sqrt{dt}$ by definition, the annualized standard deviation of dr (i.e., the basis-point volatility) is proportional to the square root of the rate. Put another way, in the CIR model the parameter σ is constant, but basis-point volatility is not: annualized basis-point volatility equals $\sigma\sqrt{r}$ and increases with the level of the short rate.

Another popular specification is that the basis-point volatility is proportional to rate. In this case the parameter σ is often called *yield volatility*. Two examples of this volatility specification are the Courtadon model,

$$dr = k(\theta - r)dt + \sigma r dw \tag{14.4}$$

and the simplest *lognormal model*, to be called Model 4, a variation of which will be discussed in the next section:

$$dr = ardt + \sigma rdw \tag{14.5}$$

In these two specifications, yield volatility is constant but basis-point volatility equals σr and increases with the level of the rate.

Figure 14-2 graphs the basis-point volatility as a function of rate for the cases of the constant, square root, and proportional specifications. For comparison purposes, σ is set in all three cases such that basis-point volatility equals 100 at a short rate of 8%. Mathematically,

$$\sigma^{bp} = .01 \tag{14.6}$$

$$\sigma^{CIR} \times \sqrt{8\%} = 1\% \Rightarrow \sigma^{CIR} = .0354 \tag{14.7}$$

$$\sigma^{y} \times 8\% = 1\% \Rightarrow \sigma^{y} = 12.5\% \tag{14.8}$$

Note that the units of these volatility measures are somewhat different. Basis-point volatility is in the units of an interest rate (e.g., 100 basis points), while yield volatility is expressed as a percentage of the short rate (e.g., 12.5%).

As shown in Figure 14-2, the CIR and proportional volatility specifications have basis-point volatility increasing with rate but at different speeds. Both models have the basis-point volatility equal to zero at a rate of zero.

The property that basis-point volatility equals zero when the short rate is zero, combined with the condition that the drift is positive when the rate is zero, guarantees that the short rate cannot become negative. In some respects this is an improvement over models with constant basis-point volatility that allow interest rates to become negative. It should be noted again, however, that choosing a model depends on the purpose at hand. Consider a trader who believes the following. One, the assumption of constant volatility is best in the current economic environment. Two, the possibility of negative rates has a small impact on the

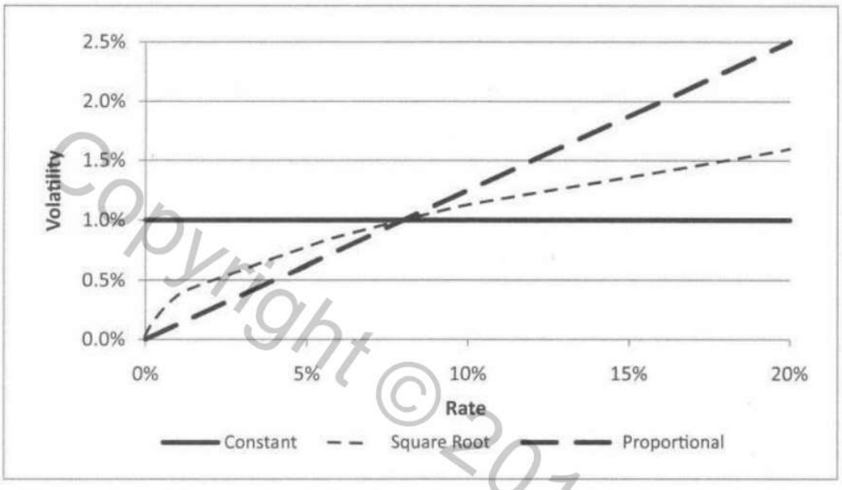


FIGURE 14-2 Three volatility specifications.

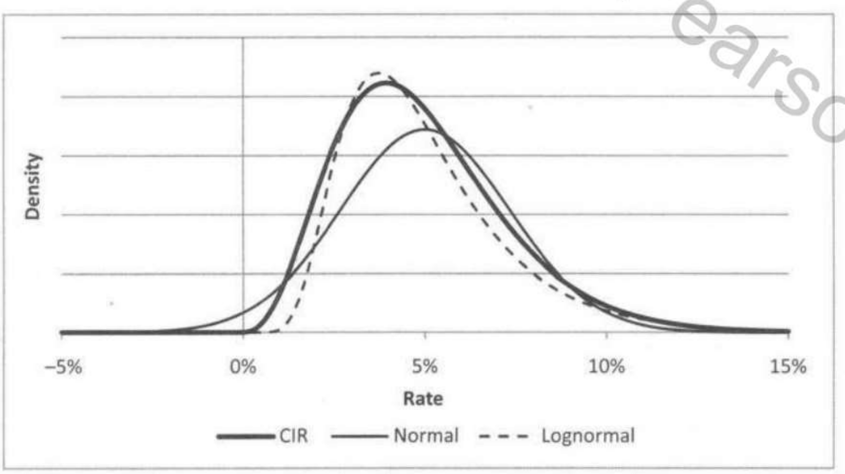


FIGURE 14-3 Terminal distributions of the short rate after ten years in CIR, normal, and lognormal models.

pricing of the securities under consideration. And three, the computational simplicity of constant volatility models has great value. This trader might very well opt for a model that allows some probability of negative rates.

Figure 14-3 graphs terminal distributions of the short rate after 10 years under the CIR, normal, and lognormal volatility specifications. In order to emphasize the difference in the shape of the three distributions, the parameters have been chosen so that all of the distributions have an expected value of 5% and a standard deviation of 2.32%. The figure illustrates the advantage of the CIR and lognormal models with respect to not allowing negative rates.

The figure also indicates that out-of-the-money option prices could differ significantly under the three models. Even if, as in this case, the mean and volatility of the three distributions are the same, the probability of outcomes away from the means are different enough to generate significantly different options prices. More generally, the shape of the distribution used in an interest rate model is an important determinant of that model's performance.

TREE FOR THE ORIGINAL SALOMON BROTHERS MODEL

This section shows how to construct a binomial tree to approximate the dynamics for a lognormal model with a deterministic drift, a model attributed here to researchers at Salomon Brothers in the '80s. The dynamics of the model are as follows:

$$dr = \tilde{a}(t)rdt + \sigma rdw \quad \textbf{(14.9)}$$

By Ito's Lemma, which is beyond the mathematical scope of this book,

$$d[\ln(r)] = \frac{dr}{r} - \frac{1}{2}\sigma^2 dt \quad \textbf{(14.10)}$$

Substituting (14.9) into (14.10),

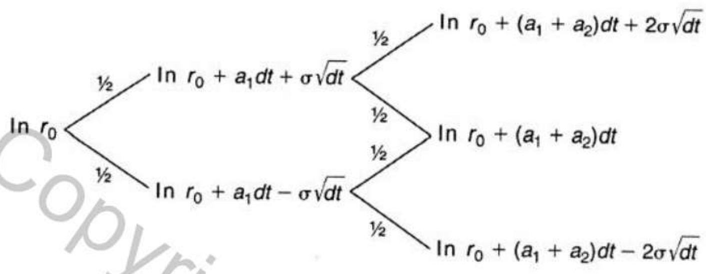
$$d[\ln(r)] = \left[\tilde{a}(t) - \frac{1}{2}\sigma^2\right]dt + \sigma dw \quad \textbf{(14.11)}$$

Redefining the notation of the time-dependent drift so that $a(t) = \tilde{a}(t) - \frac{1}{2}\sigma^2$, Equation (14.11) becomes

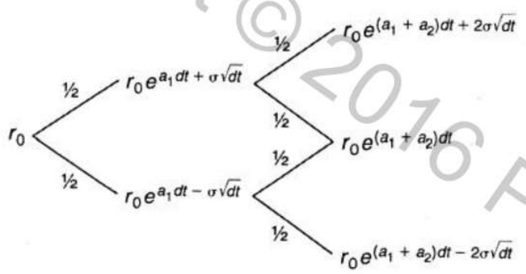
$$d[\ln(r)] = a(t)dt + \sigma dw \quad \textbf{(14.12)}$$

Equation (14.12) says that the natural logarithm of the short rate is normally distributed. Furthermore, by definition, a random variable has a lognormal distribution if its natural logarithm has a normal distribution. Therefore, (14.12) implies that the short rate has a lognormal distribution.

Equation (14.12) may be described as the Ho-Lee model based on the natural logarithm of the short rate instead of on the short rate itself. Adapting the tree for the Ho-Lee model accordingly, the tree for the first three dates is



To express this tree in rate, as opposed to the natural logarithm of the rate, exponentiate each node:



This tree shows that the perturbations to the short rate in a lognormal model are multiplicative as opposed to the additive perturbations in normal models. This observation, in turn, reveals why the short rate in this model cannot become negative. Since e^x is positive for any value of x, so long as r_0 is positive every node of the lognormal tree results in a positive rate.

The tree also reveals why volatility in a lognormal model is expressed as a percentage of the rate. Recall the mathematical fact that, for small values of x, $e^x \approx 1 + x$. Setting $a_1 = 0$ and $dt = 1$, for example, the top node of date 1 may be approximated as

$$r_0e^{\sigma} \approx r_0(1 + \sigma) \qquad \textbf{(14.13)}$$

Volatility is clearly a percentage of the rate in equation (14.13). If, for example, $\sigma = 12.5\%$, then the short rate in the up-state is 12.5% above the initial short rate.

As in the Ho-Lee model, the constants that determine the drift (i.e., a_1 and a_2) may be used to match market bond prices.

THE BLACK-KARASINSKI MODEL: A LOGNORMAL MODEL WITH MEAN REVERSION

The final model to be presented in this chapter is a lognormal model with mean reversion called the Black-Karasinski model. The model allows volatility, mean reversion, and the central tendency of the short rate to depend on time, firmly placing the model in the arbitrage-free class. A user may, of course, use or remove as much time dependence as desired.

The dynamics of the model are written as

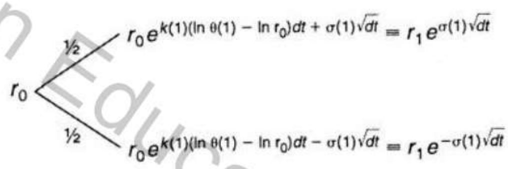
$$dr = k(t)(ln\tilde{\theta}(t) - \ln r)dt + \sigma(t)rdw \qquad \textbf{(14.14)}$$

or, equivalently,[2] as

$$d[\ln r] = k(t)(\ln \theta(t) - \ln r)dt + \sigma(t)dw \qquad \textbf{(14.15)}$$

In words, Equation (14.15) says that the natural logarithm of the short rate is normally distributed. It reverts to $\ln \theta(t)$ at a speed of $k(t)$ with a volatility of $\sigma(t)$. Viewed another way, the natural logarithm of the short rate follows a time-dependent version of the Vasicek model.

As in the previous section, the corresponding tree may be written in terms of the rate or the natural logarithm of the rate. Choosing the former, the process over the first date is



The variable r_1 is introduced for readability. The natural logarithms of the rates in the up and down-states are

$$\ln r_1 + \sigma(1)\sqrt{dt} \qquad \textbf{(14.16)}$$

and

$$\ln r_1 - \sigma(1)\sqrt{dt} \qquad \textbf{(14.17)}$$

respectively. It follows that the step down from the up-state requires a rate of

$$r_1e^{\sigma(1)\sqrt{dt}}e^{k(2)[\ln\theta(2) - \{\ln r_1 + \sigma(1)\sqrt{dt}\}]dt - \sigma(2)\sqrt{dt}} \qquad \textbf{(14.18)}$$

while the step up from the down-state requires a rate of

$$r_1e^{-\sigma(1)\sqrt{dt}}e^{k(2)[\ln\theta(2) - \{\ln r_1 - \sigma(1)\sqrt{dt}\}]dt + \sigma(2)\sqrt{dt}} \qquad \textbf{(14.19)}$$

A little algebra shows that the tree recombines only if

$$k(2) = \frac{\sigma(1) - \sigma(2)}{\sigma(1)dt} \qquad \textbf{(14.20)}$$

[2] This derivation is similar to that of moving from Equation (14.9) to Equation (14.12).

Imposing the restriction (14.20) would require that the mean reversion speed be completely determined by the time-dependent volatility function. But these elements of a term structure model serve two distinct purposes. As demonstrated in this chapter, mean reversion controls the term structure of volatility while time-dependent volatility controls the future volatility of the short-term rate (and the prices of options that expire at different times). To create a model flexible enough to control mean reversion and time-dependent volatility separately, the model has to construct a recombining tree without imposing (14.20). To do so it allows the length of the time step, dt, to change over time.

Rewriting Equations (14.18) and (14.19) with the time steps labeled dt_1 and dt_2 gives the following values for the up-down and down-up rates:

$$r_1 e^{\sigma(1)\sqrt{dt_1}} e^{k(2)[\ln\theta(2) - \{\ln r_1 + \sigma(1)\sqrt{dt_1}\}]dt_2 - \sigma(2)\sqrt{dt_2}} \quad \textbf{(14.21)}$$

$$r_1 e^{-\sigma(1)\sqrt{dt_1}} e^{k(2)[\ln\theta(2) - \{\ln r_1 - \sigma(1)\sqrt{dt_1}\}]dt_2 + \sigma(2)\sqrt{dt_2}} \quad \textbf{(14.22)}$$

A little algebra now shows that the tree recombines if

$$k(2) = \frac{1}{dt_2}\left[1 - \frac{\sigma(2)\sqrt{dt_2}}{\sigma(1)\sqrt{dt_1}}\right] \quad \textbf{(14.23)}$$

The length of the first time step can be set arbitrarily. The length of the second time step is set to satisfy (14.23), allowing the user freedom in choosing the mean reversion and volatility functions independently.

APPENDIX

Closed-Form Solutions for Spot Rates

This appendix lists formulas for spot rates, without derivation, in various models mentioned in the text. These can be useful for some applications and also to gain intuition about applying term structure models. The spot rates of term T, $\hat{r}(T)$, are continuously compounded rates.

Model 1

$$\hat{r}(T) = r_0 - \frac{\sigma^2 T^2}{6} \quad \textbf{(14.24)}$$

Model 2

$$\hat{r}(T) = r_0 + \frac{\lambda T}{2} - \frac{\sigma^2 T^2}{6} \quad \textbf{(14.25)}$$

Vasicek

$$\hat{r}(T) = \theta + \frac{1-e^{-kT}}{kT}(r_0 - \theta) - \frac{\sigma^2}{2k^2}\left(1 + \frac{1-e^{-2kT}}{2kT} - 2\frac{1-e^{-kT}}{kT}\right) \quad \textbf{(14.26)}$$

Model 3 with $\lambda(t) = \lambda$

$$\hat{r}(T) = r_0 + \frac{\lambda T}{2} - \sigma^2 \frac{2\alpha^2 T^2 - 2\alpha T + 1 - e^{-2\alpha T}}{8\alpha^3 T} \quad \textbf{(14.27)}$$

Cox-Ingersoll-Ross

Let $P(T)$ be the price of a zero-coupon bond maturing at time T (from which the spot rate can be easily calculated). Then,

$$P(T) = A(T)e^{-B(T)r_0} \quad \textbf{(14.28)}$$

where

$$A(T) = \left[\frac{2he^{(k+h)T/2}}{2h + (k+h)(e^{hT} - 1)}\right]^{2k\theta/\sigma^2} \quad \textbf{(14.29)}$$

$$B(T) = \frac{2(e^{hT} - 1)}{2h + (k+h)(e^{hT} - 1)} \quad \textbf{(14.30)}$$

$$h = \sqrt{k^2 + 2\sigma^2} \quad \textbf{(14.31)}$$

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

15 OIS Discounting

Learning Objectives

After completing this reading you should be able to:

- Explain the main considerations in choosing a risk-free rate for derivatives valuation.
- Describe the OIS rate and the LIBOR-OIS spread, and explain their uses.
- Evaluate the appropriateness of the OIS rate as a proxy for the risk-free rate.
- Describe how to use the OIS zero curve in determining forward LIBOR rates and valuing swaps.

Excerpt is from Chapter 9 of Options, Futures, and Other Derivatives, *Ninth Edition, by John C. Hull.*

This chapter discusses a number of issues that have become important in derivatives markets since the credit crisis of 2007. The first of these concerns the choice of a risk-free discount rate. This is important because the valuation of almost any derivative involves discounting expected cash flows at a risk-free rate. Prior to the credit crisis, market participants usually used LIBOR/swap rates as proxies for risk-free rates. They constructed a zero curve from LIBOR rates and LIBOR-for-fixed swap rates and used this to provide risk-free zero rates. Since the crisis, they have started to use other proxies in some circumstances.

THE RISK-FREE RATE

The standard procedure for valuing a derivative involves setting up a risk-free portfolio and arguing that in a no-arbitrage world it should earn the risk-free rate. Swaps are portfolios of FRAs or forward contracts and so their valuations rely on risk-free discounting. Indeed, as our understanding of derivatives develops, we will see that the valuation of almost any derivative requires risk-free discounting. This makes the choice of the risk-free rate important.

In the United States, the rates on Treasury bills, Treasury notes, and Treasury bonds might be thought to be natural candidates for risk-free rates. These instruments are issued by the US government and denominated in US dollars. Most analysts consider it extremely unlikely that the US government will ever default on the instruments as it always has the opportunity of increasing the money supply (which can be thought of as "printing more money") in order to repay lenders. Similar arguments can be made for instruments issued by other governments in their own currencies.[1]

In fact, derivatives market participants do not use treasury rates as risk-free rates. This is because treasury rates are generally considered to be artificially low. Some of the reasons for this are listed in Box 15-1. Pre-2008, market participants used LIBOR rates and LIBOR-for-fixed swap rates as risk-free rates. LIBOR is the short-term (1 year or less) rate of interest at which creditworthy banks (typically those rated AA or better) can borrow from other banks. Prior to the credit crisis that started in 2007, LIBOR was thought to be close to risk-free. The chance of a bank defaulting on a loan lasting 1 year or less, when the bank is rated AA at the time the loan is granted, was thought to be very small.

During the credit crisis, LIBOR rates soared because banks were reluctant to lend to each other. The TED spread, which is the excess of 3-month Eurodollar deposit rate (which like 3-month LIBOR is an interbank borrowing rate) over the 3-month US Treasury bill rate, is less than 50 basis points in normal market conditions. Between October 2007 and May 2009, it was rarely lower than 100 basis points and peaked at over 450 basis points in October 2008. Clearly banks did not regard loans to other banks as close to risk-free during this period!

It might be thought that their experience during the credit crisis would lead to practitioners looking for a better proxy for the risk-free rate when valuing derivatives. However, this is not exactly what has happened. Following the credit crisis, most banks have changed their risk-free

BOX 15-1 What Is the Risk-Free Rate?

Derivatives dealers argue that the interest rates implied by Treasury bills and Treasury bonds are artificially low because:

1. Treasury bills and Treasury bonds must be purchased by financial institutions to fulfill a variety of regulatory requirements. This increases demand for these Treasury instruments driving the price up and the yield down.
2. The amount of capital a bank is required to hold to support an investment in Treasury bills and bonds is substantially smaller than the capital required to support a similar investment in other instruments with very low risk.
3. In the United States, Treasury instruments are given a favorable tax treatment compared with most other fixed-income investments because they are not taxed at the state level.

Traditionally derivatives dealers have assumed that LIBOR rates are risk-free. But LIBOR rates are not totally risk-free. Following the credit crisis that started in 2007, many dealers switched to using overnight indexed swap (OIS) rates as risk-free rates, at least for collateralized transactions. These rates and the ways they are used are explained in this chapter.

[1] Note that the argument does not apply to eurozone countries, that is, countries which use the euro as their currency. This is because any one eurozone country, such as Italy or Spain, does not have control over the European Central Bank.

discount rates for collateralized transactions from LIBOR to what are known as overnight indexed swap (OIS) rates (see next section). But for non-collateralized transactions they continue to use LIBOR, or an even higher discount rate. This reflects a belief that the discount rate used by a bank for a derivative should represent its average funding costs, not a true risk-free rate. The average funding costs for a non-collateralized derivative is considered to be at least as high as LIBOR. Collateralized derivatives are funded by the collateral, and OIS rates, as we shall see, provide an estimate of the funding cost for these transactions.

THE OIS RATE

The fed funds rate is an overnight unsecured borrowing rate of interest between financial institutions in the US. A broker usually matches borrowers and lenders. The weighted average of the rates in brokered transactions (with weights proportional to transaction size) is termed the *effective federal funds rate*. Other countries have similar systems to the US. For example, in the UK the average of brokered overnight rates is termed the sterling overnight index average (SONIA) and, in the eurozone, it is termed the euro overnight index average (EONIA). The overnight rate in a country is monitored by the central bank, which may intervene with its own transactions in an attempt to raise or lower it.

An overnight indexed swap (OIS) is a swap where a fixed rate for a period (e.g., 1 month or 3 months) is exchanged for the geometric average of the overnight rates during the period. (The overnight rates are the average of the rates in brokered transactions as just described.) If, during a certain period, a bank borrows funds at the overnight rate (rolling the interest and principal forward each day), the interest rate it pays for the period is the geometric average of the overnight interest rates. Similarly, if it lends money at the overnight interest rate every day (rolling the interest and principal forward each day), the interest it earns for the period is also the geometric average of the overnight interest rates. An OIS therefore allows overnight borrowing or lending for a period to be swapped for borrowing or lending at a fixed rate for the period. The fixed rate in an OIS is referred to as the *OIS rate*. If the geometric average of daily rates for the period proves to be less than the fixed rate, there is a payment from the fixed-rate payer to the floating-rate payer at the end of the period; otherwise, there is a payment from the floating-rate payer to the fixed-rate payer at the end of the period.

Example 15.1

Suppose that in a US 3-month OIS the notional principal is $100 million and the fixed rate (i.e., the OIS rate) is 3% per annum. If the geometric average of overnight effective federal funds rates during the 3 months proves to be 2.8% per annum, the fixed-rate payer has to pay $0.25 \times (0.030 - 0.028) \times \$100{,}000{,}000$ or $50,000 to the floating-rate payer. (This calculation does not take account of the impact of day count conventions.)

Overnight indexed swaps tend to have relatively short lives (often 3 months or less). However, transactions that last as long as 5 to 10 years are becoming more common. An OIS lasting longer than 1 year is typically divided into 3-month subperiods. At the end of each subperiod the geometric average of the overnight rates during the subperiod is exchanged for the OIS rate. The swap rate in a plain vanilla LIBOR-for-fixed swap is a continually refreshed LIBOR rate (i.e., the rate that can be earned on a series of short-term loans to AA-rated financial institutions). Similarly, the OIS rate is a continually refreshed overnight rate (i.e., it is the rate that can be earned by a financial institution from a series of overnight loans to other financial institutions).

Suppose that Bank A engages in the following transactions:

1. Borrow $100 million in the overnight market for 3 months, rolling the interest and principal on the loan forward each night.
2. Lend the $100 million for 3 months at LIBOR to another bank, Bank B.
3. Use an OIS to exchange the overnight borrowings for borrowings at the 3-month OIS rate.

This will lead to Bank A receiving the 3-month LIBOR rate and (assuming its creditworthiness remains acceptable to the overnight market) paying the 3-month overnight indexed swap rate. We might therefore expect the 3-month overnight indexed swap rate to equal the 3-month LIBOR rate. However, it is generally lower. This is because Bank A requires some compensation for the risk it is taking that Bank B will default on the 3-month LIBOR loan. The overnight lenders to Bank A bear much less risk than Bank A does when it lends to Bank B for 3 months.

This is because they have the option of ceasing to lend to Bank A if Bank A's credit quality declines.

The excess of the 3-month LIBOR rate over the 3-month overnight indexed swap rate is known as the 3-month *LIBOR-OIS spread*. It is often used as a measure of stress in financial markets. Its values between 2002 and 2013 are shown in Figure 15-1. In normal market conditions, it is about 10 basis points. However, it rose sharply during the 2007–2009 credit crisis because banks became less willing to lend to each other for 3-month periods. In October 2008, the spread spiked to an all time high of 364 basis points. By a year later it had returned to more normal levels. But it has since increased in response to stresses and uncertainties in financial markets. For example, it rose to about 50 basis points at the end of December 2011 as a result of concerns about the economies of European countries such as Greece.

The OIS rate is a good proxy for the risk-free rate. The OIS rate is not totally risk-free, but it is very close to risk-free. Two sources of risk can be identified, both very small. The first is that there might be a default on an overnight loan between two financial institutions. The chance of this is very small because any hint of an imminent credit problem is likely to lead to a financial institution being excluded from the overnight market. The second is that there might be a default on the OIS swap itself. However, the adjustment to an OIS swap rate to reflect default possibilities is generally very small (particularly if the OIS is collateralized).

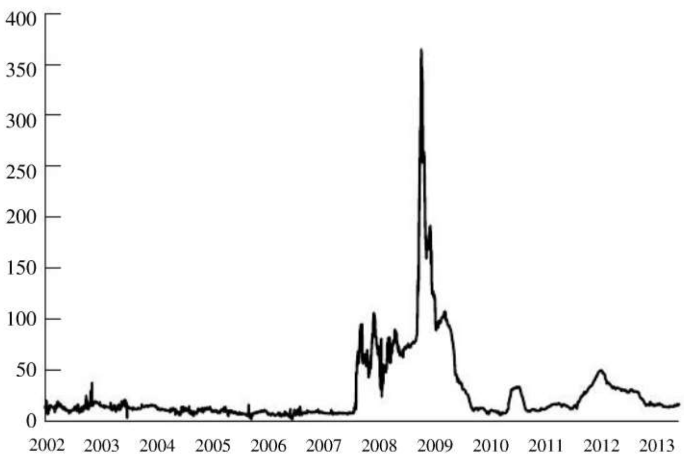


FIGURE 15-1 The LIBOR-OIS spread from January 2002 to May 2013.

Determining the OIS Zero Curve

The bootstrap method can be used to calculate the LIBOR/swap zero curve. LIBOR-for-fixed swap rates define a series of par yield bonds. A key point here is that, for the swap rates to define a series of par yield bonds, it is necessary for the rates being bootstrapped to be the same as the rates being used for discounting.

The procedure for constructing the OIS zero curve when OIS rates are used for discounting is similar to that used to construct the LIBOR zero curve when LIBOR rates are used for discounting. The 1-month OIS rate defines the 1-month zero rate, the 3-month OIS rate defines the 3-month zero rate, and so on. When there are periodic settlements in the OIS contract, the OIS rate defines a par yield bond. Suppose, for example, that the 5-year OIS rate is 3.5% with quarterly settlements. (This means that at the end of each quarter $0.25 \times 3.5\% = 0.875\%$ is exchanged for the geometric average of the overnight rates during the quarter.) A 5-year bond paying a quarterly coupon at a rate of 3.5% per annum would be assumed to sell for par.

Although OIS swaps are becoming more liquid, they do not trade for maturities that are as long as the more common LIBOR-for-fixed interest rate swaps. If the OIS zero curve is required for long maturities, a natural approach is to assume that the spread between an OIS rate and the corresponding LIBOR/swap rates is the same at the long end as it is for the longest OIS maturity for which there is reliable data. Suppose, for example, that there are no reliable data on OIS swaps for maturities longer than 5 years. If the 5-year OIS rate is 4.7% and the 5-year LIBOR-for-fixed swap rate is 4.9%, OIS rates could be assumed to be 20 basis points less than the corresponding LIBOR/swap rates for all maturities beyond 5 years. An alternative approach for extending the OIS zero curve is to use basis swaps where 3-month LIBOR is exchanged for the average federal funds rate. These swaps have maturities as long as 30 years in the US.[2]

[2] If the swap rate for a 30-year LIBOR interest rate swap is 5% and LIBOR is swapped for the average federal funds rate plus 20 basis points, it might be assumed that the 30-year OIS rate is 4.8% (assuming appropriate adjustments have been made for day counts). Unfortunately, this would involve an approximation as a swap of fed funds for LIBOR involves the arithmetic average (not geometric average) of overnight rates during a period being swapped for the LIBOR rate applicable to the period. A "convexity adjustment" is in theory necessary. See, for example, K. Takada, "Valuation of Arithmetic Average of Fed Funds Rates and Construction of the US Dollar Swap Yield Curve," 2011, SSRN-id981668.

VALUING SWAPS AND FRAs WITH OIS DISCOUNTING

Once the OIS zero curve has been determined, many types of derivatives can be valued using OIS rates as the risk-free discount rates. For example, the value of a forward contract on an asset can be calculated with *r* equal to the OIS zero rate for a maturity of *T* years. To value swaps and FRAs we have a little more work to do. It is first necessary to calculate forward LIBOR in a way that is consistent with OIS discounting.

Determining Forward LIBOR Rates with OIS Discounting

LIBOR-for-fixed swaps can be valued assuming that forward LIBOR rates are realized. LIBOR-for-fixed swaps, if transacted at today's mid-market swap rates, are worth zero. This provides a way of determining LIBOR forward rates. The LIBOR forward rates given by OIS discounting are different from those given by LIBOR discounting. We will illustrate this for a simple situation. Example 15.2 calculates a forward LIBOR rate assuming that LIBOR rates are used for discounting. Example 15.3 calculates the same forward LIBOR rate assuming that OIS rates are used for discounting.

Example 15.2

Suppose that the 1-year LIBOR rate is 5% and the 2-year LIBOR-for-fixed swap rate with annual payments is 6%. Both rates are annually compounded. A bank uses LIBOR rates for discounting. Suppose that *R* is the 2-year LIBOR/swap zero rate. Because a bond providing a coupon of 6% is a par yield bond we must have

$$\frac{6}{1.05} + \frac{106}{(1+R)^2} = 100$$

Solving this gives $R = 6.030\%$. Suppose that *F* is the forward LIBOR rate for the 1-year period beginning in 1 year. We can calculate it from the zero rates:

$$F = \frac{1.06030^2}{1.05} - 1 = 7.0707\%$$

As a check of this result, we can calculate *F* so that it makes the value of the swap zero. The exchange in 1 year to the party receiving fixed is worth +1 per 100 of principal. (This is because the party receives 6 and pays 5.) Assuming forward rates are realized, the exchange in 2 years is $6 - 100F$ per 100 of principal. The value of the swap is

$$\frac{1}{1.05} + \frac{6-100F}{1.06030^2}$$

per 100 of principal. Setting this equal to zero and solving for *F*, we see that $F = 7.0707\%$ as before.

Example 15.3

As in Example 15.2, suppose that the 1-year LIBOR rate is 5% and the 2-year swap rate with annual payments is 6%. (Both rates are annually compounded.) A bank uses OIS rates for discounting. Assume that the OIS zero curve has been calculated as described in the previous section and the 1- and 2-year OIS zero rates are 4.5% and 5.5% with annual compounding. (In this situation, OIS zero rates are therefore about 50 basis points lower than LIBOR zero rates.) Suppose that *F* is the forward LIBOR rate for the 1-year period beginning in 1 year. Swaps can be valued assuming that forward LIBOR rates are realized. Because a swap where 6% is received and LIBOR is paid is worth zero, we must have

$$\frac{1}{1.045} + \frac{6-100F}{1.055^2} = 0$$

Solving this gives $F = 7.0651\%$.

In Examples 15.2 and 15.3, when we switch from LIBOR discounting to OIS discounting, the forward LIBOR changes from 7.0707% to 7.0651%. The change is a little more than half a basis point. It is small, but is something that traders would not want to ignore. In practice, the impact of the switch depends on the steepness of the zero curve and the maturity of the forward rate (see DerivaGem 3.00).

Calculations of the sort we have given in Example 15.3 enable a forward LIBOR curve to be constructed when OIS rates are used as risk-free discount rates. Using a series of swaps where exchanges are made every 3 months enables the 3-month forward rates as a function of maturity (i.e., as a function of the start of the 3-month period) to be constructed; using swaps where exchanges are made every 6 months enables the 6-month forward rate as a function of maturity to be constructed; and so on.[3] (Interpolation

[3] Basis swaps where, for example, 1-month LIBOR is exchanged for 6-month LIBOR, provide extra information to assist in the compilation of a complete set of LIBOR forward curves corresponding to different accrual periods.

between calculated forward rates is used to determine complete forward LIBOR curves.)

When a swap is valued using OIS discounting, the forward rates corresponding to the swap's cash flows are obtained from the appropriate forward LIBOR curves. Cash flows on the swap are then calculated assuming these forward rates will occur and the cash flows are discounted at the appropriate OIS zero rates.

OIS VS. LIBOR: WHICH IS CORRECT?

As already mentioned, most derivatives dealers now use discount rates based on OIS rates when valuing collateralized derivatives (i.e., derivatives where there is a collateral agreement) and discount rates based on LIBOR when valuing non-collateralized derivatives.[4] The reason most commonly given for this concerns funding costs. Collateralized derivatives are funded by the collateral and the federal funds rate (which, as we have explained, is linked to the OIS rate) is the overnight interest rate most commonly paid on collateral. In the case of non-collateralized transactions, it is argued that funding costs are higher and the discount rate should reflect this.

As explained later, arguments based on funding costs are questionable because it is a long-established principle in finance that the evaluation of an investment should not depend on the way it is funded. It is the risk of the investment and its expected cash flows that are important. Finance theory leads to the conclusion that we should always use the best proxy available for the risk-free rate when discounting in situations where riskless portfolios have been set up. Arguably the OIS zero curve is as close to risk-free as we can get. It should therefore be used for discounting regardless of whether the transaction is collateralized.[5]

[4] LCH.Clearnet is a large CCP that was clearing interest rate swap transactions with a total notional principal of over $350 trillion in 2013. Its transactions are collateralized with initial margin and variation margin. Following the practice of dealers, it now uses OIS discounting rather than LIBOR discounting.

[5] For further discussion of this, see J. Hull and A. White, "LIBOR vs. OIS: The Derivatives Discounting Dilemma," *Journal of Investment Management*, 11, 3 (2013), 14–27.

SUMMARY

We saw in earlier chapters that the credit crisis that started in 2007 has led to the over-the-counter derivatives markets being regulated much more heavily than before. In this chapter, we have seen that it has also caused derivatives market participants to carefully review their practices. Prior to the credit crisis, LIBOR was assumed to be a reasonable proxy for the risk-free rate. (This was convenient, as it made the valuation of an interest rate swap where LIBOR is exchanged for a fixed rate of interest relatively easy.) Since the credit crisis, practitioners have switched their risk-free proxy from the LIBOR rate to the OIS rate—at least for collateralized derivatives transactions.

The OIS rate is a rate swapped for the geometric average of the overnight federal funds rate. It is not perfectly risk-free because a default on an overnight loan or the swap is always possible. However, it is much closer to risk-free than LIBOR.

Using OIS rates rather than LIBOR rates for discounting changes estimates of forward LIBOR rates. When OIS discounting is used, forward LIBOR rates must be estimated so that all LIBOR-for-fixed swaps if entered into today at the mid-market swap rate have zero value.

Banks and other derivatives dealers have for many years been concerned about counterparty credit risk. Two adjustments are currently made for bilaterally cleared transactions. The credit value adjustment (CVA) is an adjustment for the possibility that the counterparty will default and reduces the value of a derivatives portfolio. The debit (or debt) value adjustment (DVA) is an adjustment for the possibility that the bank will default and increases the value of a derivatives portfolio. In addition, for collateralized portfolios, a further adjustment can be necessary if the interest paid on cash collateral is different from the risk-free rate.

Finance theory shows that the way a project is funded should not influence its valuation. In spite of this, some banks do make what is termed a funding value adjustment (FVA) so that a derivatives portfolio which requires (generates) funding is charged with (given credit for) an amount reflecting the bank's average funding cost. FVAs are controversial and have the potential to lead to disagreements between accountants, analysts, and traders.

Further Reading

Demiralp, S., B. Preslopsky, and W. Whitesell. "Overnight Interbank Loan Markets," Manuscript, Board of Governors of the Federal Reserve, 2004.

Filipovic, D., and A. Trolle. "The Term Structure of Interbank Risk," *Journal of Financial Economics*, 109, 3 (September 2013): 707–33.

Hull, J., and A. White. "The FVA Debate," *Risk*, 25th anniversary edition (July 2012): 83–85.

Hull, J., and A. White. "LIBOR vs. OIS: The Derivatives Discounting Dilemma," *Journal of Investment Management*, 11, 3 (2013): 14–27.

Hull, J., and A. White. "OIS Discounting and the Pricing of Interest Rate Derivatives," Working Paper, University of Toronto, 2013.

Smith, D. "Valuing Interest Rate Swaps Using OIS Discounting," *Journal of Derivatives*, 20, 4 (Summer 2013): 49–59.

Volatility Smiles

16

Learning Objectives

After completing this reading you should be able to:

- Define volatility smile and volatility skew.
- Explain the implications of put-call parity on the implied volatility of call and put options.
- Compare the shape of the volatility smile (or skew) to the shape of the implied distribution of the underlying asset price and to the pricing of options on the underlying asset.
- Describe characteristics of foreign exchange rate distributions and their implications on option prices and implied volatility.
- Describe the volatility smile for equity options and foreign currency options, and provide possible explanations for its shape.
- Describe alternative ways of characterizing the volatility smile.
- Describe volatility term structures and volatility surfaces and how they may be used to price options.
- Explain the impact of the volatility smile on the calculation of the "Greeks."
- Explain the impact of a single asset price jump on a volatility smile.

Excerpt is Chapter 20 of Options, Futures, and Other Derivatives, *Ninth Edition, by John C. Hull.*

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

How close are the market prices of options to those predicted by the Black-Scholes-Merton model? Do traders really use the Black-Scholes-Merton model when determining a price for an option? Are the probability distributions of asset prices really lognormal? This chapter answers these questions. It explains that traders do use the Black-Scholes-Merton model—but not in exactly the way that Black, Scholes, and Merton originally intended. This is because they allow the volatility used to price an option to depend on its strike price and time to maturity.

A plot of the implied volatility of an option with a certain life as a function of its strike price is known as a *volatility smile*. This chapter describes the volatility smiles that traders use in equity and foreign currency markets. It explains the relationship between a volatility smile and the risk-neutral probability distribution being assumed for the future asset price. It also discusses how option traders use volatility surfaces as pricing tools.

WHY THE VOLATILITY SMILE IS THE SAME FOR CALLS AND PUTS

This section shows that the implied volatility of a European call option is the same as that of a European put option when they have the same strike price and time to maturity. This means that the volatility smile for European calls with a certain maturity is the same as that for European puts with the same maturity. This is a particularly convenient result. It shows that when talking about a volatility smile we do not have to worry about whether the options are calls or puts.

Put-call parity provides a relationship between the prices of European call and put options when they have the same strike price and time to maturity. With a dividend yield on the underlying asset of q, the relationship is

$$p + S_0e^{-qT} = c + Ke^{-rT} \tag{16.1}$$

As usual, c and p are the European call and put price. They have the same strike price, K, and time to maturity, T. The variable S_0 is the price of the underlying asset today, and r is the risk-free interest rate for maturity T.

A key feature of the put-call parity relationship is that it is based on a relatively simple no-arbitrage argument. It does not require any assumption about the probability distribution of the asset price in the future. It is true both when the asset price distribution is lognormal and when it is not lognormal.

Suppose that, for a particular value of the volatility, p_{BS} and c_{BS} are the values of European put and call options calculated using the Black-Scholes-Merton model. Suppose further that p_{mkt} and c_{mkt} are the market values of these options. Because put-call parity holds for the Black-Scholes-Merton model, we must have

$$p_{BS} + S_0e^{-qT} = c_{BS} + Ke^{-rT}$$

In the absence of arbitrage opportunities, put-call parity also holds for the market prices, so that

$$p_{mkt} + S_0e^{-qT} = c_{mkt} + Ke^{-rT}$$

Subtracting these two equations, we get

$$p_{BS} - p_{mkt} = c_{BS} - c_{mkt} \tag{16.2}$$

This shows that the dollar pricing error when the Black-Scholes-Merton model is used to price a European put option should be exactly the same as the dollar pricing error when it is used to price a European call option with the same strike price and time to maturity.

Suppose that the implied volatility of the put option is 22%. This means that $p_{BS} = p_{mkt}$ when a volatility of 22% is used in the Black-Scholes-Merton model. From Equation (16.2), it follows that $c_{BS} = c_{mkt}$ when this volatility is used. The implied volatility of the call is, therefore, also 22%. This argument shows that the implied volatility of a European call option is always the same as the implied volatility of a European put option when the two have the same strike price and maturity date. To put this another way, for a given strike price and maturity, the correct volatility to use in conjunction with the Black-Scholes-Merton model to price a European call should always be the same as that used to price a European put. This means that the volatility smile (i.e., the relationship between implied volatility and strike price for a particular maturity) is the same for European calls and European puts. More generally, it means that the volatility surface (i.e., the implied volatility as a function of strike price and time to maturity) is the same for European calls and European puts. These results are also true to a good approximation for American options.

Example 16.1

The value of a foreign currency is $0.60. The risk-free interest rate is 5% per annum in the United States and 10% per annum in the foreign country. The market price of a European call option on the foreign currency with a maturity of 1 year and a strike price of \$0.59 is 0.0236. DerivaGem shows that the implied volatility of the call is 14.5%. For there to be no arbitrage, the put-call parity relationship in Equation (16.1) must apply with q equal to the foreign risk-free rate. The price p of a European put option with a strike price of \$0.59 and maturity of 1 year therefore satisfies

$$p + 0.60e^{-0.10\times 1} = 0.0236 + 0.59e^{-0.05\times 1}$$

so that $p = 0.0419$. DerivaGem shows that, when the put has this price, its implied volatility is also 14.5%. This is what we expect from the analysis just given.

FOREIGN CURRENCY OPTIONS

The volatility smile used by traders to price foreign currency options has the general form shown in Figure 16-1. The implied volatility is relatively low for at-the-money options. It becomes progressively higher as an option moves either in the money or out of the money.

In the appendix at the end of this chapter, we show how to determine the risk-neutral probability distribution for an asset price at a future time from the volatility smile given by options maturing at that time. We refer to this as the *implied distribution*. The volatility smile in Figure 16-1 corresponds to the implied distribution shown by the solid line in Figure 16-2. A lognormal distribution with the same mean and standard deviation as the implied distribution is shown by the dashed line in Figure 16-2. It can be seen that the implied distribution has heavier tails than the lognormal distribution.[1]

To see that Figures 16-1 and 16-2 are consistent with each other, consider first a deep-out-of-the-money call option with a high strike price of K_2. This option pays off only if the exchange rate proves to be above K_2. Figure 16-2 shows that the probability of this is higher for the implied probability distribution than for the lognormal distribution. We therefore expect the implied distribution to give a relatively high price for the option. A relatively high price leads to a relatively high implied volatility—and this

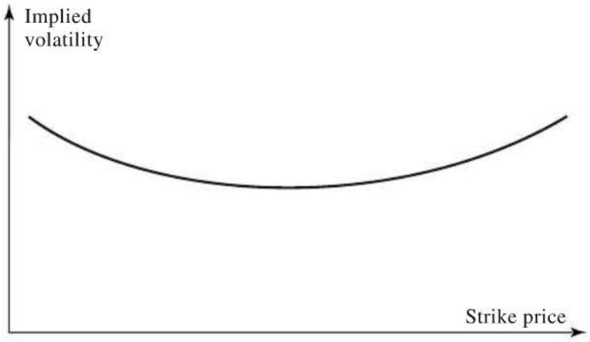


FIGURE 16-1 Volatility smile for foreign currency options.

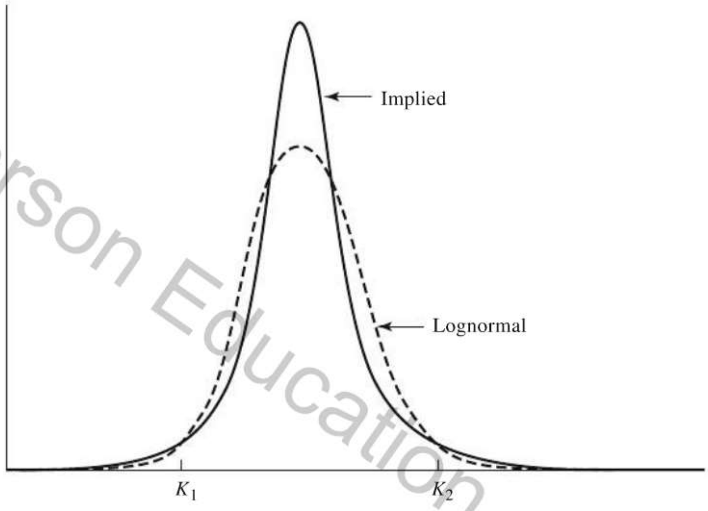


FIGURE 16-2 Implied and lognormal distribution for foreign currency options.

is exactly what we observe in Figure 16-1 for the option. The two figures are therefore consistent with each other for high strike prices. Consider next a deep-out-of-the-money put option with a low strike price of K_1. This option pays off only if the exchange rate proves to be below K_1. Figure 16-2 shows that the probability of this is also higher for the implied probability distribution than for the lognormal distribution. We therefore expect the implied distribution to give a relatively high price, and a relatively high implied volatility, for this option as well. Again, this is exactly what we observe in Figure 16-1.

Empirical Results

We have just shown that the volatility smile used by traders for foreign currency options implies that they consider

[1] This is known as *kurtosis*. Note that, in addition to having a heavier tail, the implied distribution is more "peaked." Both small and large movements in the exchange rate are more likely than with the lognormal distribution. Intermediate movements are less likely.

TABLE 16-1 Percentage of Days When Daily Exchange Rate Moves Are Greater than 1, 2, . . . , 6 Standard Deviations (SD = standard deviation of daily change)

	Real World	Lognormal Model
>1 SD	25.04	31.73
>2 SD	5.27	4.55
>3 SD	1.34	0.27
>4 SD	0.29	0.01
>5 SD	0.08	0.00
>6 SD	0.03	0.00

that the lognormal distribution understates the probability of extreme movements in exchange rates. To test whether they are right, Table 16-1 examines the daily movements in 12 different exchange rates over a 10-year period.[2] The first step in the production of the table is to calculate the standard deviation of daily percentage change in each exchange rate. The next stage is to note how often the actual percentage change exceeded 1 standard deviation, 2 standard deviations, and so on. The final stage is to calculate how often this would have happened if the percentage changes had been normally distributed. (The lognormal model implies that percentage changes are almost exactly normally distributed over a one-day time period.)

Daily changes exceed 3 standard deviations on 1.34% of days. The lognormal model predicts that this should happen on only 0.27% of days. Daily changes exceed 4, 5, and 6 standard deviations on 0.29%, 0.08%, and 0.03% of days, respectively. The lognormal model predicts that we should hardly ever observe this happening. The table therefore provides evidence to support the existence of heavy tails (Figure 16-2) and the volatility smile used by traders (Figure 16-1). Box 16-1 shows how you could have made money if you had done the analysis in Table 16-1 ahead of the rest of the market.

[2] The results in this table are taken from J. C. Hull and A. White, "Value at Risk When Daily Changes in Market Variables Are Not Normally Distributed." *Journal of Derivatives*, 5, No. 3 (Spring 1998): 9–19.

BOX 16-1 Making Money from Foreign Currency Options

Black, Scholes, and Merton in their option pricing model assume that the underlying's asset price has a lognormal distribution at future times. This is equivalent to the assumption that asset price changes over a short period of time, such as one day, are normally distributed. Suppose that most market participants are comfortable with the Black-Scholes-Merton assumptions for exchange rates. You have just done the analysis in Table 16-1 and know that the lognormal assumption is not a good one for exchange rates. What should you do?

The answer is that you should buy deep-out-of-the-money call and put options on a variety of different currencies and wait. These options will be relatively inexpensive and more of them will close in the money than the lognormal model predicts. The present value of your payoffs will on average be much greater than the cost of the options.

In the mid-1980s, a few traders knew about the heavy tails of foreign exchange probability distributions. Everyone else thought that the lognormal assumption of Black-Scholes-Merton was reasonable. The few traders who were well informed followed the strategy we have described—and made lots of money. By the late 1980s everyone realized that foreign currency options should be priced with a volatility smile and the trading opportunity disappeared.

Reasons for the Smile in Foreign Currency Options

Why are exchange rates not lognormally distributed? Two of the conditions for an asset price to have a lognormal distribution are:

1. The volatility of the asset is constant.
2. The price of the asset changes smoothly with no jumps.

In practice, neither of these conditions is satisfied for an exchange rate. The volatility of an exchange rate is far from constant, and exchange rates frequently exhibit jumps.[3] It turns out that the effect of both a nonconstant

[3] Sometimes the jumps are in response to the actions of central banks.