Sample Chapter 6

Value-at-Risk

Theory and Practice

Glyn A. Holton

This is Chapter 6 of Glyn Holton's book **Value-at-risk: Theory and Practice**. For more information, visit www.value-at-risk.net.

Chapter 6

Market Data

6.1. MOTIVATION

When we design a VaR measure, one of the first steps is to choose a key vector ${}^1\boldsymbol{R}$. We need this before we can design a mapping procedure that will construct portfolio mappings ${}^1P = \theta({}^1\boldsymbol{R})$. We also need it before we can design an inference procedure that will characterize the conditional distribution of ${}^1\boldsymbol{R}$.

There are various issues to consider in selecting what financial variables to represent with key factors ${}^1R_i$. One of these is the availability of historical market data. An inference procedure requires historical data related to all key factors. If there is no historical data relating to a particular financial variable, it makes little sense to model that variable as a key factor.

In this chapter, we discuss types of historical market data that may be used by VaR measures. We describe how data is collected over time, how it is filtered and cleaned of errors, and how it is converted into forms usable by an inference procedure.

6.2. FORMS OF DATA

Data can represent market prices, interest rates, spreads, implied volatilities, etc. Any of these may be directly quoted in the markets, or they may be inferred from other quantities that are directly quoted. All are, in some sense, prices. Price data can vary with respect to type, method of collection, and source.

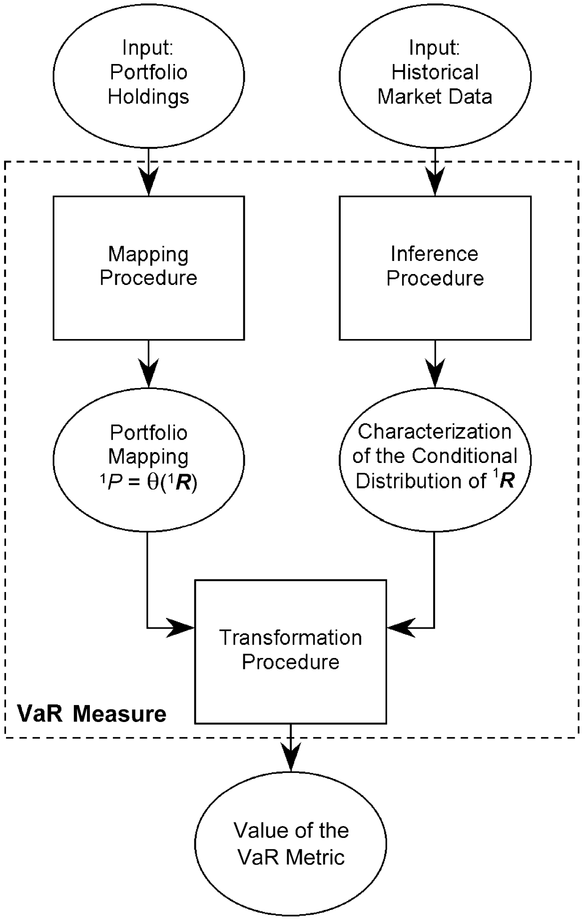

Exhibit 6.1 A reproduction of Exhibit 1.11, which is a general schematic for VaR measures. The availability of historical market data influences the selection of financial variables to be modeled with key vector $^1\boldsymbol{R}$. This, in turn, influences our design of both a mapping procedure and an inference procedure.

Types of Prices

There are essentially four types of prices:

- **Transaction prices** are prices at which actual transactions took place.
- **Firm prices** are bid or offer prices quoted by market participants, who are then obligated to transact at those prices if accepted by a counterparty. Firm bid and offer prices may be averaged to obtain firm mid-market prices.

- **Indicative prices** are bid, offer, or mid-market prices quoted by market participants—usually brokers or market makers—for informational purposes only. The quoting party has no obligation to transact.
- **Settlement prices** are prices specified by exchanges for purposes such as calculating daily margin requirements. Each exchange has its own rules for calculating settlement prices based upon transaction, firm, or indicative prices at the close of trading each day.

Collecting Data

Time series of historical prices are constructed by recording prices at a regular interval—each day, each week, each month, etc. For settlement prices, collecting data is as easy as recording the daily settlement prices released by the appropriate exchange.

Transaction prices are more problematic. For actively traded instruments, there can be thousands of transactions spaced at irregular intervals throughout each day. Other assets trade less frequently and may fail to trade at all on some days. Four techniques for collecting daily transaction data are:

- **opening prices**, which comprise the first transaction price following the market open each day;
- **closing prices**, which comprise the last transaction price prior to the market close each day;
- **high prices**, which comprise the highest transaction price each day; or
- **low prices**, which comprise the lowest transaction price each day.

For VaR applications, closing prices are most commonly used. Irrespective of how they are collected, transaction prices tend to be **nonsynchronous**. This means that prices for different assets will reflect transactions occurring at different times. On a given day, the last trade in one stock might occur at 2:20 PM with the last trade in another occurring at 4:15 PM.

If trading in an asset is light, several days or weeks may transpire between transactions. If there are no transactions in a given day, a missing data indicator may be entered in the time series in lieu of a value. Alternatively, some rule may be applied for filling in missing values. On days when an asset doesn't trade, its closing price might be set equal to the previous day's closing price.

With exchanges implementing evening trading sessions and electronic trading sessions, it is important to clarify which trading sessions opening, closing, high, and low prices are based upon. For markets that trade around the clock, such as foreign exchange, the notions of market open and market close are not meaningful. In such cases, two times can be selected to represent "market open" and "market close." In certain OTC markets, high and low prices may be exaggerated by outliers—transactions at significantly off-market prices. Some rule may be employed to discard outliers when calculating high and low prices.

Firm or indicative prices are easier to collect synchronously than transaction prices. In inactive markets, indicative quotes may be solicited from brokers or dealers at a fixed time each day. In more active markets, firm or indicative quotes are distributed electronically throughout each trading day. Problems may arise during periods of active trading when dealers may be too busy to update indicative quotes. Lyons (1995) observes that, during periods of volatility in the foreign exchange market, indicative bid and offer prices on Reuters FXFX screens may fail to bracket transaction prices.

A shortcoming of indicative quotes is questionable quality. If a relationship exists with the quoting party, that party will want to provide quality quotes. However, market makers have no financial stake in indicative quotes, so they are unlikely to prepare them as carefully as they would firm quotes. A less common problem is indicative quotes being intentionally made off-market as a means of influencing markets.

One technique for addressing such problems is to obtain indicative quotes from multiple parties. The high and low quotes may be discarded and the remainder averaged. The British Bankers Association (BBA) uses such a formula in preparing its daily Libor indicative quotes.

The particular types of data we choose to use in a VaR analysis will depend upon the foregoing issues as well as availability. Transaction prices are generally available for exchange-traded instruments, but are more difficult to obtain in OTC markets. Settlement prices are only available for exchange-traded instruments such as futures and options. In some OTC markets, firm quotes may be available. In others, it may only be possible to obtain indicative quotes. Indeed, in many markets, data of any sort may be difficult to obtain.

Data Sources

There are many sources for data. The most common include:

- exchanges,
- broker or dealer quotes,
- data vendors,
- real-time data feeds, and
- trade tickets.

All exchanges record detailed information on transactions. This is used for various purposes, including:

- audit,
- dispute resolution,
- distribution of real-time price information via data feeds, and
- determination of settlement prices.

Not all information is available publicly, but most exchanges distribute price histories indicating, among other things, high, low, closing, and settlement prices for each trading day.

Indicative prices may be obtained directly from brokers or dealers. It is best to standardize this process so the operational definition of the time series remains stable over time. Quotes may be obtained from one or several parties. Be candid with quoting parties. Let them know that the quotes are for risk management purposes only, and seek their commitment to provide the quotes on an ongoing basis. Many brokers or dealers are happy to provide this service for institutions with whom they have a profitable business relationship.

Data vendors collect data from multiple sources, preprocess it, and distribute it as time series. Costs vary, but are generally modest compared to the cost of constructing and maintaining one's own time series.

It is important to understand what sources data vendors use for data as well as what preprocessing they perform. If possible, it is best to purchase data prepared specifically for risk analyses. Some data vendors distribute covariance matrices for key factors instead of time series. If possible, avoid using these. As we shall discuss in Chapter 7, different analyses may yield very different covariance matrices from a given time series. The choice of technique is a significant design decision for a VaR model that should not be left to a data vendor.

Most trading organizations purchase real-time data feeds for their traders. Values can be captured from these periodically—perhaps at an appointed time each day—and used to construct time series. Data collected in this manner may require extensive cleaning. Because the data is delivered in real time, quality controls are minimal. Even in liquid markets, data feeds may lag the markets. Also, most vendors place restrictions upon use of their real-time data, and these may prohibit the accumulation of historical databases.

If an institution is active in a particular market, it may capture transaction prices directly from its own trade tickets, or subsequently from its trade accounting system. This should be considered only as a last resort if alternative data sources are unavailable. It is feasible only for market participants who trade in volume—primarily market makers. Cleaning the data is expensive and labor intensive. There is the risk of losing the data source if, for some reason, the institution is forced to stop trading in a particular market for a period.

6.3. NONSYNCHRONOUS DATA

A time series $\{{}^{-\alpha}\boldsymbol{q}, {}^{-\alpha+1}\boldsymbol{q}, \ldots, {}^{0}\boldsymbol{q}\}$ is said to be **synchronous** if components ${}^{t}q_i$ for each term ${}^{t}\boldsymbol{q}$ are realized simultaneously. Otherwise, it is said to be **nonsynchronous**. Financial time series are typically nonsynchronous for one of two reasons:

- **Trading effects** relate to instruments that trade infrequently or fail to trade for a period of time. It also encompasses certain effects such as brokers failing to provide timely indicative quotes during a period of heavy trading volumes.
- **Timing effects** relate to instruments trading in different time zones or according to different schedules. Some exchanges—such as the Singapore Exchange—have different closing times for different instruments trading in the same time zone. This causes settlement prices to be nonsynchronous.

We have already mentioned trading effects. They primarily affect opening or closing transaction price data. As we indicated above, they may also affect firm or indicative quotes. High and low transaction prices are not expected to be synchronous since different assets will attain their high and low prices at different times each day. If instruments do not trade for a day or more, the resulting nonsynchronicity can be a serious issue. Trading effects generally do not affect settlement prices. Each exchange has its own methodology for determining settlement prices. Most are based upon some average of transaction prices occurring immediately prior to the market close. If trading is not active at the close, a common practice is to poll traders and base the settlement price on some average of the indicative quotes.

As an illustration of a timing effect, the Tokyo, London, and New York stock exchanges stop trading each day at 3:00 PM, 5:00 PM, and 4:00 PM, respectively, in their local times. Because of time zone differences, Tokyo actually closes 11 hours before London, and London closes 4 hours before New York.[1] Closing prices collected from the respective exchanges are nonsynchronous.

Impact of Nonsynchronous Data

Nonsynchronous data complicates the task of assigning a current value ${}^0p$ to a portfolio if the current value ${}^0\boldsymbol{r}$ of the key factors is collected nonsynchronously. Cross-hedged positions may appear to not be hedged. Arbitrage conditions that should hold may appear to not hold. The resulting value for ${}^0p$ may be misleading or even nonsensical.

If time series models are not implemented specifically to address nonsynchronous data, inferred correlations will be lower, in absolute value, than they would be if the data were collected synchronously. Spurious autocorrelations between different risk factors will also arise.

Consider the stocks of two companies that trade on the same exchange and tend to move in tandem. The first trades actively; the second does not. News arriving late in the day will not be reflected in the inactive stock's closing price if that stock fails to trade subsequent to the arrival of the news. This tendency

[1]Time differences reflect periods when daylight savings time is nowhere in effect.

will suppress the inferred return correlation between the two stocks and induce a positive autocorrelation between the stocks' returns lagged by one day. If the inactive stock sometimes fails to trade for a day or more, such autocorrelations may persist for several days.

Eurodollar futures are traded on both the CME in Chicago and the Singapore Exchange. The two contracts are essentially identical. For a given expiration date, we may treat closing prices in Chicago and Singapore as being for the same contract, but collected at different times.[2] By comparing these closing prices to closing prices of some other contract, we may assess the impact of nonsynchronous data.

We collect daily settlement price data from May 3, 1999, to February 1, 2000, for the following contracts:

- CBOT June 2000 Treasury bond future,
- CME June 2000 Eurodollar future, and
- Singapore Exchange June 2000 Eurodollar future.

The CBOT and CME are both in Chicago, so settlement prices for the CBOT Treasury bond contract and CME Eurodollar contract are synchronous.[3] The settlement price for the Singapore Eurodollar contract is set 8 hours earlier.[4] Also, 16 hours after the Chicago settlement prices are set, the subsequent day's Singapore Eurodollar settlement price is set (assuming this is also a trading day.) This is illustrated in Exhibit 6.2:

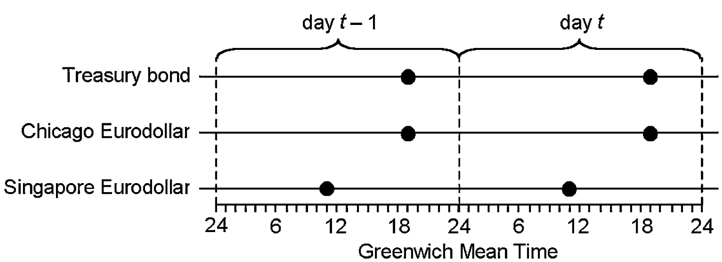

Exhibit 6.2 Time lines illustrate when prices for the three contracts are collected. The Chicago settlement prices are set 8 hours after the Singapore settlement prices. They are set 16 hours before the Singapore settlement prices for the subsequent day, assuming it is also a trading day.

We expect the Treasury bond contract to be positively correlated with both Eurodollar contracts, but because the Singapore Eurodollar contract is nonsynchronous, we expect its correlation with the Treasury bond contract to be lower.

[2]The two exchanges have an offset arrangement that allows an open contract on one exchange to be closed on the other, so the two exchanges' contracts are truly fungible.

[3]Open outcry trading for both contracts ends at 2:00 PM each day.

[4]Trading closes in Singapore at 7:00 PM local time.

The Treasury bond contract should also exhibit a modest autocorrelation with the Singapore Eurodollar contract. To test these hypotheses, we calculate return correlations and return autocorrelations as indicated in Exhibit 6.3:[5]

Daily Return Correlations

	Chicago Eurodollar settlement price for day *t*	**Singapore Eurodollar** settlement price for day *t*
Chicago Treasury bond settlement price for day *t*	.808	.142

Daily Return Autocorrelations

	Chicago Eurodollar settlement price for day *t*	**Singapore Eurodollar** settlement price for day *t*
Chicago Treasury bond settlement price for day $t - 1$	-.036	.634
Chicago Treasury bond settlement price for day $t + 1$	-.034	.043

Exhibit 6.3 The top table shows daily return correlations between the CBOT Treasury bond contract and the two Eurodollar contracts. The bottom table shows corresponding return autocorrelations with the CBOT Treasury bond contract either leading or lagging by a day.

Prices for Treasury bond and Eurodollar futures move in tandem, as indicated by the .808 correlation for the two contracts traded in Chicago. However, nonsynchronicity causes the Chicago Treasury bond future and Singapore Eurodollar future to have a correlation of just .142. If the Treasury bond future is lagged a day, the autocorrelation is .634.

The effect of nonsynchronicity on correlations tends to diminish if data is collected at longer intervals. If prices are collected monthly, the effect of the nonsynchronicity will be modest, as long as the prices are not too mean-reverting.

The most effective way to address nonsynchronicity is to avoid it in the first place. If transaction prices are highly nonsynchronous because of nontrading, consider using firm, indicative, or settlement prices instead. To the extent possible, try to collect prices in a single time zone. If this is impossible, consider collecting most prices in a particular time zone, and obtain prices from other time zones as broker quotes, which offer some flexibility in the timing.

---

[5]Log returns are used. Data from days when any of the exchanges were closed is omitted from the calculation. Results are inferred using the method of uniformly weighted moving averages (UWMA) discussed in Chapter 7.

6.4. DATA ERRORS

Data errors are data values that are, in some sense, erroneous. **Filtering** is any procedure for identifying data values as erroneous. **Cleaning** is any procedure that corrects erroneous data values or replaces them in the data series with some missing data indicator.

Errors

Data errors frequently arise when data is first placed in electronic form, either through manual keying or through a process of scanning and optical character recognition. With manual keying, forward prices may be entered for the wrong maturity. Bid and offer prices may be transposed. Exchange rates may be inverted or entered for the wrong currency pair. Decimal places can be omitted or shifted; digits can be transposed; incorrect digits may be entered; digits may be dropped or extra digits entered.

Optical character recognition software introduces different types of errors. Characters may be mistaken for other characters, with particular mistakes depending upon the font in which scanned data is printed as well as the quality of the printing. The letter O may be read as the number 0. In certain fonts, the numerals 6, 8, and 9 may be mistaken for one another, as may be 5 and 6. Stray marks may be interpreted as text or decimal points. Legitimate decimal points may be overlooked.

Even after data is in electronic form, software or hardware may introduce errors. If a real-time data feed fails, a user may misinterpret bid and offer prices at the time of the failure as being effective throughout the period when the system is down. A system might transmit only the third and fourth decimal places of a price except when the second decimal place changes. If the second decimal place changes during a period when the system is down, a user may be unaware of the change after the system is restored. Some systems may transmit dummy data when they are turned on or being tested. In some markets, automated quotation systems are designed to indicate "reasonable" prices as if they were actual prices during periods of trading inactivity.

Off-market prices are actual prices that are not, in our subjective opinion, reflective of markets at the time they are quoted. A broker may carelessly quote an indicative price. An unsophisticated market participant may accept an unreasonable price. A trader may mistakenly trade at an off-market price. Occasionally, a trader may transact at off-market prices to manipulate markets or distort his portfolio's mark-to-market value. **Wash trades** are identical offsetting trades between two counterparties, which can be performed at off-market prices. **Ramping** is the performance of very small transactions at off-market prices.

The pricing of one transaction may influence the pricing of other transactions with the same counterparty. If a trade accounting system cannot handle swaps, a trader may transact a swap as a strip of individual forwards, each at the same price. Recorded in a time series, the identical prices will appear to indicate a flat forward curve.

Data Filtering

Data filtering may entail one or more of:

- computer algorithms,
- human review, or
- data comparisons.

Many data errors—especially keying errors—are blatant. If the decimal point is accidentally dropped from 1.72, it becomes 172.00. If we mistakenly transpose the leading digits in 19.02, it becomes 91.02. Such blatant errors are easily identified in the context of a time series. Computer algorithms can sift through large volumes of data to locate such outliers. For more subtle errors, algorithms may employ statistical inference, pattern-recognition techniques, or arbitrage relationships—such as put-call parity or interest-rate parity—to identify suspect data values. Spreads or other price relationships can be checked to see if they conform to historical patterns.

Simple filtering algorithms are easy to construct, but more sophisticated algorithms require careful design as well as some fine-tuning over time. Seasonality and heteroskedasticity complicate designs. Tests relating to specific price relationships or patterns must be customized.

Subtle data errors can be identified manually by traders or other professionals who follow market developments. Such reviews should be performed soon after data is recorded, while a reviewer's memory is fresh. Also, if a computer algorithm identifies data values as suspect, humans may perform a final review to determine which of these are actual data errors.

Finally, if data for certain risk factors is available from several independent sources, the data from these sources can be compared for consistency.

Data Cleaning

Once data errors have been identified, they must be cleaned—either set equal to values we believe to be correct or deleted. Corrected values should be obtained from the same source from which the data originated. For keying errors, refer to the documents from which the data was keyed. If erroneous data is obtained from an exchange, the exchange should be able to correct it. If transaction prices are erroneous, contact the counterparties to the trade. If data is obtained from a data vendor, you won't have access to original sources, but you can request that the vendor obtain corrected values.

6.5. DATA BIASES

As market theoreticians, we like to think of historical prices as forming an unbroken series of numbers, each representing the price at which an efficient market cleared at a particular point in time. Data may contain errors, but we filter and clean it to eliminate these.

As practitioners, we accept a different view. While certain data values are clearly erroneous, others are not so easily categorized. From an operational standpoint, data is erroneous if our filtering procedure identifies it as such. Otherwise it is correct. Subjectively, things are not so stark. We perceive a continuum of gradations between erroneous and correct data. A filtering procedure discards certain values as erroneous, but the remainder may reflect a variety of modest biases or distortions. Isolated distortions, so long as they are minor, have little effect on VaR measurements. More problematic are modest biases of a systematic nature.

In any market, there are bid-ask spreads. These introduce a bias into transaction prices—biasing them upward/downward when a market maker is selling/buying. Such biases are inconsequential in liquid markets if daily price fluctuations dwarf bid-ask spreads. They are more problematic in illiquid markets, where bid-ask spreads can be large.

Transaction costs reduce arbitrage opportunities, which increases the potential for modest biases or discrepancies to persist in markets.

If an asset is credit sensitive, different transaction prices may reflect differing counterparty credit qualities or differing collateralization arrangements. Even if credit quality is not an issue, liquidity effects can introduce biases. Two bonds may be issued by similar credits—or both by the same credit—but one trades at a premium to the other because it has a more active secondary market. This is evident in the US Treasury bond market, where the most recently issued "on the run" bonds trade at a premium.

In most commodity markets, there is some choice as to the method of settlement, and prices vary depending upon which method is selected. Some standard methods are:

- free on board (FOB)—the commodity is cleared through customs and delivered on board the recipient's ship at the port of departure.
- free alongside (FAS)—FOB delivery, except the commodity is delivered alongside the recipient's ship.
- cost, insurance, freight (CIF)—FOB delivery as well as insurance and shipping. Essentially, the commodity is delivered to the destination port, but actual settlement is through delivery of the ship's bill of lading.
- ex-dock—the commodity is cleared through customs and delivered on the dock of the destination port.
- in-warehouse—the commodity is cleared through customs and delivered in a warehouse at the destination port.

- ex-warehouse—the delivering party provides in-warehouse delivery and pays the cost of moving the commodity to the warehouse exit.

In addition, transaction prices may reflect additional services such as storage, inventory management, or balancing fees. Payment terms may also influence prices.

The biases described above are all associated with transaction prices. As an alternative, firm or indicative prices may be available reflecting some standard settlement and payment terms, and mid-market prices can be calculated from these. Another solution is to apply standard adjustments to transaction prices to make them comparable. Prices reflecting different settlement methods might be adjusted to make them all consistent with, say, ex-dock settlement. If a market is active, it may be reasonable to use some average of prices, say recording each day's closing price as the average price of all transactions completed during the last minute of trading. Individual prices may reflect specific biases, but these will cancel somewhat in the averaged price.

Substituting the price of one asset for that of a closely related asset can introduce biases. Forward and future prices for the same underlier often move in tandem, especially if they have identical or similar settlement terms. However, they can diverge. One reason is market segmentation. If large transactions that would move the futures market are always transacted on the forward market, arbitraging between the two markets may be infeasible. Another reason is a phenomenon that has come to be known as **convexity bias**.

Cox, Ingersoll, and Ross (1981) and Jarrow and Oldfield (1981) suggest that daily margin payments on futures may cause forward and futures prices to diverge. If there is a correlation between daily futures prices and interest rates, one party to a futures contract will tend to receive margin payments on days when interest rates rise and make margin payments on days when interest rates decline. On average, she can expect to invest the margin payments she receives at interest rates that are higher than those at which she finances the margin payments she makes. The other party can expect the opposite experience. This should cause a divergence in forward and futures prices, with the effect depending upon the maturity of contracts, the magnitude of correlations, and the volatility of the future's prices.

Empirical studies by Cornell and Reinganum (1981), French (1983), and Park and Chen (1985) confirm a modest convexity bias in gold, silver, silver coins, platinum, copper, and plywood prices, but fail to find one for various currencies.

As we might expect, the effect is most pronounced when a future's underlier is an interest rate or fixed-income instrument that exhibits a high correlation with applicable interest rates. In this context, Burghardt and Hoskins (1995) coined the name "convexity bias." They and Gupta and Subrahmanyam (2000) discuss the convexity bias in the pricing of interest-rate swaps. The swap market traditionally priced swaps directly off Eurocurrency futures without recognizing any convexity bias, but this started to change in the early 1990s. Today, swap prices reflect a significant convexity bias relative to Eurocurrency futures.

Taxes may also introduce biases. Discrepancies can exist across jurisdictions or within a jurisdiction.

Between 1953 and 1963, the US Treasury issued a number of Treasury bonds with a special feature. If tendered as payment of federal estate taxes, the bonds would be valued at par, irrespective of their current market value. Because of their association with estate taxes and funerals, the bonds came to be called "flower bonds." When interest rates rose during the 1970s and 1980s, flower bonds traded at a premium. The last flower bond had a coupon of 3.5% and matured in 1998.

Many tax-related biases are unintentional. If they offer opportunities for tax arbitrage, they are likely to soon be legislated out of existence. Cornell (1981) and Viswanath (1989) describe one such effect that existed until 1981 in the US Treasury bill futures market.

6.6. FUTURES

Futures prices are generally collected as settlement prices. Because the contracts mature, and because multiple contracts usually trade simultaneously for a given underlier, futures prices do not form continual time series. Instead, they form multiple time series corresponding to different contracts. Each one begins when the corresponding contract starts to trade and ends when that contract expires. This is illustrated in Exhibit 6.4.

Time (month)	Contract (indicated by month of expiration)				
	Jan-00	Mar-00	May-00	Jul-00	Sep-00
Apr-99	241.2				
May-99	244.1				
Jun-99	243.6	248.8			
Jul-99	239.3	244.0	246.0		
Aug-99	245.7	248.5	251.5		
Sep-99	237.3	242.3	246.5	252.0	
Oct-99	230.0	235.0	239.0	243.0	
Nov-99	223.5	225.2	228.9	232.4	235.5
Dec-99	216.7	218.5	221.6	224.7	231.5
Jan-00		221.7	225.0	229.0	233.0
Feb-00		221.8	222.1	225.0	230.0
Mar-00			231.0	232.5	240.5
Apr-00			238.2	241.5	244.5
May-00				224.0	227.0
Jun-00				218.0	218.0
Jul-00					214.8
Aug-00					219.0

Exhibit 6.4 Monthly prices for flaxseed futures traded on the Winnipeg Commodities Exchange (WCE). Prices are quoted in CAD/ton and represent the last settlement price for the respective month. Source: WCE.

Techniques of time series analysis require data to be in continual time series. There are two ways to convert discontinual futures price data into continual time series. We can construct nearbys or constant-maturity price series.

Nearbys

The standard means of obtaining continual time series from futures prices is to use **nearby series** or simply **nearbys**. Consider futures trading on a particular exchange for a given underlier. At any point in time, there will be contracts trading for several expirations. A **first nearby** is a time series comprising the price, at each point in time, of the nearest-to-expiration contract. The **second nearby** comprises the price, at each point in time, of the second nearest-to-expiration contract, etc.

Exhibit 6.5 indicates monthly settlement price data for municipal bond index futures traded on the CBOT. Prices are in USD and reflect the last settlement price for each month. Data is provided for six contracts maturing between September 1998 and December 1999. Exhibit 6.6 presents a first and second nearby constructed from the data. To clarify how the nearbys are constructed, we have used shading in Exhibits 6.5 and 6.6 to indicate the prices used to form the second nearby.

Time (month)	Contract (indicated by month of expiration)					
	Sep-98	Dec-98	Mar-99	Jun-99	Sep-99	Dec-99
Jun-98	124.4688	124.0625				
Jul-98	124.0000	123.5625				
Aug-98	126.6250	126.4063				
Sep-98		128.3750	128.2188			
Oct-98		125.4688	125.0938			
Nov-98		126.4688	126.0938			
Dec-98			125.0000	124.5313		
Jan-99			126.7500	126.2813	125.8125	125.3438
Feb-99			124.6250	123.6875	123.2188	122.7500
Mar-99				123.5938	122.7188	122.2500
Apr-99				123.1875	122.3125	121.8438
May-99				121.9688	120.4688	119.5313
Jun-99					118.7813	117.7500
Jul-99					118.0625	116.4375
Aug-99					114.9688	114.0000

Exhibit 6.5 Monthly prices for municipal bond index futures traded on the CBOT. Prices are quoted in USD and represent the last settlement price for the respective month. The purpose of the shaded numbers is explained in the text. Source: CBOT.

Time months	Nearby First	Nearby Second
Jun-98	124.4688	124.0625
Jul-98	124.0000	123.5625
Aug-98	126.6250	126.4063
Sep-98	128.3750	128.2188
Oct-98	125.4688	125.0938
Nov-98	126.4688	126.0938
Dec-98	125.0000	124.5313
Jan-99	126.7500	126.2813
Feb-99	124.6250	123.6875
Mar-99	123.5938	122.7188
Apr-99	123.1875	122.3125
May-99	121.9688	120.4688
Jun-99	118.7813	117.7500
Jul-99	118.0625	116.4375
Aug-99	114.9688	114.0000

Exhibit 6.6 The first two nearbys constructed from the data of Exhibit 6.5. The first nearby represents, for each time t, the price of the future that was closest to expiration at that time. The second nearby represents, for each time t, the price of the future that was second closest to expiration at that time. To clarify the construction of nearbys, the futures prices used to construct the second nearby have been shaded in both Exhibits 6.5 and 6.6.

Nearbys and Distortions

Nearbys exhibit certain distortions related to contract expirations. Depending upon the underlying commodity, price standard deviations may rise or fall as the last trade date approaches. Correlations with other prices may also be affected. If the same commodity is traded on two geographically separate exchanges, the correlation between futures prices on the two exchanges will tend to decline as the contracts approach expiration. Localized supply and demand imbalances will become evident and it will become increasingly infeasible to transport the commodity from one location to the other in time for delivery. During the final days prior to a contract's last trade date, distortions can be pronounced as positions are closed out, open interest declines, and liquidity migrates to the subsequent contract.

If futures prices are formed into nearbys, such expiration-related behavior manifests itself as a cyclical pattern in each nearby. This may be especially pronounced in the first nearby. One way to mitigate this effect in the first nearby is to construct nearbys by rolling over contracts prior to each last-trade date.

The **rollover date** for a set of nearbys is the date when each nearby is switched to reflect prices of the subsequent futures contract. The last possible rollover date is the last trade date of the expiring contract, but earlier dates, such as the first day of the expiring contract's delivery month, are possible.

The earlier we make a rollover date, the more we can mitigate expiration-related distortions in the first nearby. If we set our rollover date 2 weeks prior to the last-trade date, prices during each contract's last 2 weeks of trading will be discarded. They will appear in none of the nearbys.

Early rollover of nearbys raises a problem if a portfolio holds futures contracts to expiration. We don't want to stop recording data for a soon-to-expire contract while it is still held by the portfolio! A simple solution is to track data right up until contracts expire. Only after they expire do we go back and modify nearbys to reflect the earlier rollover date.

The judicious selection of a rollover date is just one technique that addresses a specific form of distortion in the first nearby. Other distortions may remain across all nearbys. Exhibit 6.7 is a graph of the daily second nearby for IPE natural gas futures. The first trading day of each delivery month is used as a rollover date. The graph covers the second half of 1998.

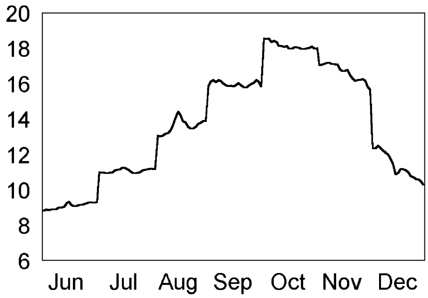

Exhibit 6.7 Second nearby for International Petroleum Exchange (IPE) natural gas futures for June 1998 through December 1998. Prices are settlement prices in USD. The first trading day of each delivery month is used as a rollover date.

The nearby prices exhibit a pattern. This is more pronounced in Exhibit 6.8, which shows daily returns for the same nearby. The six largest (positive or negative) returns in Exhibit 6.8 are evenly spaced in time. Indeed, each occurs on the first day of a month. These returns do not reflect market events. They are artifacts of how the nearby was constructed. IPE natural gas futures expire monthly. Every month, each nearby rolls over to the next contract. Because of seasonality effects, there is usually a price jump between consecutive contracts. These are incorporated into each nearby. The monthly price jumps are evident in Exhibit 6.7. They cause the large evenly spaced returns in Exhibit 6.8.

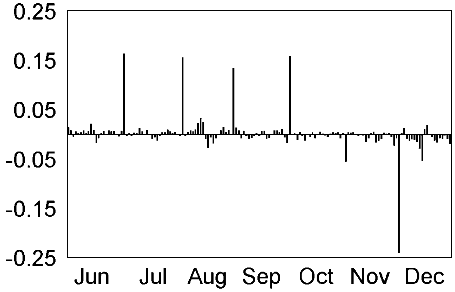

Exhibit 6.8 Daily log returns for the nearby price data of Exhibit 6.7.

Such artifacts can be addressed in various ways. One approach is to remove artifact price jumps by directly adjusting nearby prices upward or downward as needed prior to each rollover date. Prices prior to the most recent rollover date are adjusted by the price difference between the new and old contracts. Prices prior to the previous rollover date are adjusted by that amount plus the price difference between the new and old contracts for that rollover date, etc. Nearbys constructed in this manner are called **price-adjusted nearbys**. This solution eliminates artifact price jumps. However, the accumulation of price adjustments may cause older nearby prices to stray significantly from their true levels. Indeed, if we go back far enough, nearby prices may even become negative.

Alternatively, we may remove artifact returns by directly adjusting those returns that straddle a rollover date. The artifact returns arise because, when a return straddles a rollover date, it is calculated using a price from the old contract and another price from the new contract. The solution is to calculate such returns exclusively from the new contract's prices. Based upon these returns, and returns that can be calculated from the rest of the nearby's prices, we obtain a series of returns for the nearby. We then convert these into what is called a **return-adjusted nearby** by starting with the most recent futures price and applying the most recent return to obtain the previous price. Working backwards in this manner, we obtain an entire series of prices. Prior to the most recent rollover date, these may not reflect actual price levels or actual price changes. However, prices will not become negative, and historical returns will be correctly represented.

Through the judicious selection of a rollover date and the use of price-adjusted or return-adjusted nearbys, we can mitigate the most obvious distortions associated with nearbys. More subtle distortions can remain. Primarily, these will be cyclical patterns in standard deviations or correlations. If contracts expire frequently, say every month, these will be modest. With bimonthly or quarterly expirations, they may be more severe. Such distortions are not easily mitigated.

Constant-Maturity Futures Prices

As an alternative to nearbys, futures price data can be merged into continual time series as **constant-maturity** prices. A constant-maturity price series indicates, for each time t, an interpolated price reflecting a specific time-to-expiration that is constant over time.

To illustrate, let's construct monthly 60-day, 120-day, and 180-day constant-maturity prices from the CME random-length lumber futures prices of Exhibit 6.9. Data for contracts maturing between September 1998 and November 1999 are indicated. Prices are USD settlement prices for the last trading day of each month. The corresponding number of days to expiration is indicated for each contract at each time t.

Interpolation is performed with a cubic spline. Exhibit 6.10 illustrates the interpolation for the data of October 1998, which is shaded in Exhibit 6.9. Performing such an interpolation for each month yields the constant-maturity price series of Exhibit 6.11.

Constant-maturity prices are appealing because they don't exhibit some of the cyclical distortions of nearbys, but they have other limitations. One is the fact that forward contracts are typically quoted as spreads to specific "benchmark" futures contracts. This is done in oil, natural gas, coffee, cocoa, and other markets. For this reason, it is desirable to model the actual futures prices that are used as benchmarks. Converting data to a constant-maturity form is a step in the wrong direction.

Another problem is the availability of futures prices from which to interpolate constant-maturity prices. Active futures markets, such as the Eurodollar or Henry Hub natural gas markets, have numerous contracts trading simultaneously. Less active markets may offer just a handful of maturities—perhaps three, or sometimes only two. Futures curves may have irregular shapes. Supply and demand issues as well as seasonality effects cause them to rise and fall with irregular patterns. Trying to interpolate constant-maturity prices based upon two or three contracts is likely to be a unsettling experience!

The most useful application of constant-maturity futures prices is as a proxy for forward prices. As long as enough futures maturities are traded to facilitate reasonable interpolation, and there is no significant convexity bias, such use may be reasonable. For most purposes, nearbys are a better alternative.

Exercises

6.1 Use the data of Exhibit 6.9 to construct first and second nearbys for random-length lumber. Use last reported trade dates as rollover dates. Perform the calculations twice:
   a. Calculate price-adjusted nearbys.
   b. Calculate return-adjusted nearbys.

Time (month)	Contract (indicated by month of expiration)													
	Sep-98		Nov-98		Jan-99		Mar-99		May-99		Jul-99		Sep-99	
	days	price	days	price	days	price	days	price	days	price	days	price	days	price
Dec-97	258	317.8												
Jan-98	228	325.8	287	320.0										
Feb-98	200	317.4	259	312.0										
Mar-98	168	326.0	227	320.0	290	321.7								
Apr-98	138	313.0	197	312.5	260	323.0	319	333.9						
May-98	109	280.5	168	285.6	231	300.5	290	313.0						
Jun-98	77	292.9	136	294.9	199	302.2	258	307.5	318	305.1				
Jul-98	46	301.3	105	297.4	168	303.7	227	310.5	287	311.0				
Aug-98	15	287.9	74	275.9	137	284.5	196	291.5	256	300.0				
Sep-98			44	255.8	107	268.9	166	281.9	226	292.5	288	300.5		
Oct-98			14	260.7	77	270.0	136	282.0	196	286.2	258	292.7		
Nov-98					46	290.5	105	297.2	165	299.0	227	301.6		
Dec-98					15	305.6	74	309.2	134	307.0	196	307.1		
Jan-99							45	343.5	105	322.7	167	317.9	229	316.4
Feb-99							17	334.4	77	333.8	139	323.5	201	321.6
Mar-99									44	337.5	106	329.5	168	324.2
Apr-99									14	343.9	76	334.8	138	328.3
May-99											48	349.2	110	341.0
Jun-99											15	403.6	77	386.6
Jul-99													47	372.9

Exhibit 6.9 Monthly prices and days-to-expiration for random length lumber futures traded on the CME. Prices are quoted in USD and represent the last settlement price for the respective month. Source: CME.

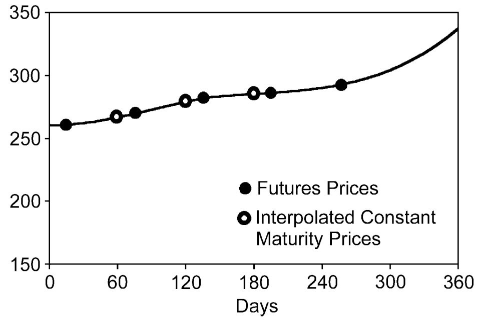

Exhibit 6.10 The interpolation of constant-maturity prices is illustrated. Random length lumber futures prices for October 1998 (shaded in Exhibit 6.9) are used to interpolate 60-day, 120-day, and 180-day constant-maturity prices. Interpolation is performed with a cubic spline.

Time	Constant-maturity Price		
(month)	60-day	120-day	180-day
Jul-98	299.05	298.28	305.25
Aug-98	276.21	281.33	289.57
Sep-98	258.85	271.82	284.62
Oct-98	266.62	279.06	285.31
Nov-98	292.83	297.90	299.33
Dec-98	309.16	307.62	306.35
Jan-99	336.11	320.49	317.77
Feb-99	335.65	326.62	320.16
Mar-99	334.99	328.19	323.20
Apr-99	336.72	330.19	323.05

Exhibit 6.11 Using the data of Exhibit 6.9, 60-day, 120-day, and 180-day constant-maturity random-length lumber prices are interpolated using cubic splines. Results are shown for July 1998 through April 1999.

6.7. IMPLIED VOLATILITIES

If a portfolio holds options, it is exposed to changes in implied volatilities. These can be modeled with key factors just like any other risk factors, but an inference procedure will require historical implied volatility data. Implied volatilities are rarely quoted directly in the market. An exception is the OTC currency options market, but in most markets, implied volatilities must be inferred from option prices.

A challenge in compiling historical data for implied volatilities is the sheer volume of data. For a single underlier, options may trade for multiple maturities and multiple strikes. Also, different implied volatilities may apply to different options structures. Put-call parity ensures close compatibility between put and call implied volatilities for a given underlier, but this is not true of other structures. Interest-rate caps and interest-rate swaptions ostensibly have the same underliers, but they are different instruments. Implied volatilities for one are remotely related to those of the other.

To provide a sense of the volume of raw data involved, consider put and call options traded on the Coffee, Sugar and Cocoa Exchange[6] (CSCE). On April 26, 2001, the exchange listed settlement prices for 536 distinct options on coffee futures, 247 options on sugar futures, and 328 options on cocoa futures. This is data generated in a single day for just three underliers! Faced with such an avalanche of data, we must find some means of organizing and synthesizing it into a form that is amenable to time-series analysis.

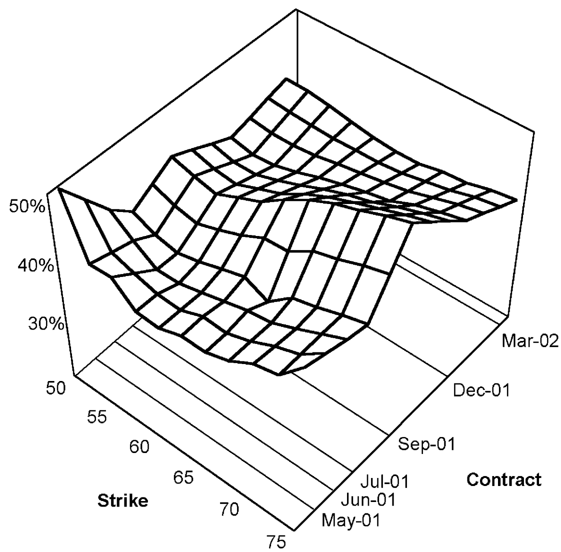

Exhibit 6.12 The implied volatility surface for CSCE coffee call options on March 12, 2001. Skew is apparent, as is a spike in volatilities for the September contract.

Implied volatilities vary by expiration. In most markets, they also vary by strike—a phenomenon that is referred to as **skew**. If data is to reflect both expiration and strike dependencies, it must comprise an entire **volatility surface** for each trading day. Such a surface is illustrated for options on coffee futures in Exhibit 6.12. A volatility surface is generally recorded as a set of implied volatilities for specific expiration-strike pairs. Below, we discuss how to choose those pairs for tracking historical implied volatility data.

[6]The CSCE is a subsidiary of the New York Board of Trade (NYBOT).

Expirations

If options are OTC, implied volatilities are generally tracked for constant-maturity expirations. For exchange-traded options, we may track implied volatilities either for constant-maturity expirations or as nearbys. Although there are exceptions, most exchange-traded options are either on equities or futures.

For exchange-traded equity options, as long as there are sufficient expirations to facilitate interpolation, it makes sense to track constant-maturity implied volatilities. This will avoid the nonstationarities and artifact returns associated with nearbys. Since equity markets exhibit no pronounced seasonalities, constant-maturity implied volatilities are generally well behaved.

For exchange-traded options on futures, it is usually best to treat options in the same manner as you do the underlying futures. If you track the futures as nearbys, track the implied volatilities as nearbys. Options on futures may expire as much as a month prior to the underlying future. Accordingly, rollover dates for implied volatility nearbys are likely to precede rollover dates for the futures. As long as a mapping procedure recognizes when an implied volatility from one nearby corresponds to the futures price of the *subsequent* nearby, this should present no problem. Another challenge is that there may be more options expirations than futures expirations. The CSCE has futures expiring on only certain months, but options on those futures expire monthly. Such issues can be addressed on a case-by-case basis through suitable construction of both nearbys and the mapping procedure.

Many of the distortions that arise with futures nearbys also arise with volatility nearbys. Analogous solutions generally apply.

Strikes

If options exhibit skew, we will want to track implied volatilities for multiple strikes. We might simply choose a few specific strikes—say GBP 40, GBP 45, and GBP 50—and track, for each expiration, implied volatilities at those strikes. Doing so raises two problems:

1. We need the strikes to bracket the current underlier price. Strikes that do so today may fail to do so in the future.
2. Because of skew, observed implied volatilities for a given strike will rise and fall as that strike moves in and out of the money.

A standard solution is to track implied volatilities for strikes corresponding to specific normalized[7] deltas. For call options, we might track implied volatilities

[7] A normalized delta is an option's delta divided by the option's notional amount. For vanilla options, normalized deltas are between –1 and 1.

for deltas of .25, .50, and .75. Over time, the corresponding strike prices will change, but the deltas will remain constant. We obtain time series for out-of-the-money, at-the-money, and in-the-money implied volatilities. Note that the .75 delta will correspond to a *lower* strike price than the .25 delta. Also, by put-call parity, call-option implied volatilities for deltas of .25, .50, and .75 should correspond to put implied volatilities for deltas of –.75, –.50, and –.25, respectively. In most markets, it is only necessary to track implied volatilities for only calls or only puts, as put-call parity will ensure close correspondence between the respective volatilities. Violations of put-call parity may occur for far in-the-money or far out-of-the money options. For options with significant time value, such violations may become apparent at call deltas of .25 and .75. A solution is to track implied volatilities for a more narrow range of deltas, say .35, .50, and .65. Corresponding put deltas would be –.65, –.50, and –.35, respectively.

6.8. FURTHER READING

See Davidson (1996) and Zangari (1996a) for issues related to sourcing of data. See Goldman Sachs and SBC Warburg Dillon Read (1998) for data issues primarily from a systems standpoint. Ma, Mercer, and Walker (1992) and Geiss (1995) discuss the construction of futures nearbys.