CHAPTER 12

Overcoming Data Quirks to Design Trading Strategies

Stock price data are widely available, so you can expect that many trading strategies have already been uncovered by your competitors. Furthermore, the data you will use in performing event studies and backtesting strategies cannot be used without special care. This chapter will outline some of the main pitfalls in price data, and coping strategies to overcome those pitfalls. Our list is hardly exhaustive, but it will provide a sense of the nuances you must anticipate.

Actual Versus Adjusted Stock Price Data

Naive use of stock price data as reported by financial services—termed here *actual price data*—can be misleading. The reason is that stock prices often change for reasons that have nothing to do with market supply and demand. To properly account for the value returned by holding a stock, we must consider *splits* and *dividends issued*.

In back testing, these factors can be accounted for in a manual sort of way, provided one has a database that includes that information. However, most researchers fold the effects of these two factors into a special version of historical price called *adjusted price*. This streamlines backtesting by removing the complexity of manipulating that data.

Stock Splits

Company boards of directors occasionally believe that their stock's price has made it too expensive for some groups of investors, especially because most trades are for round lots of at least 100 shares. Options contracts are also issued to cover 100 shares at a time. A stock trading at $100 per

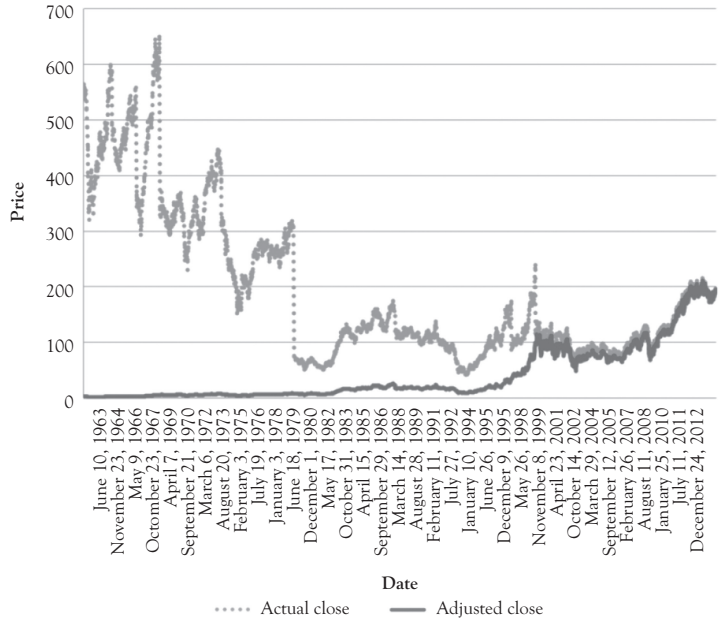

Figure 12.1 IBM's actual and split-adjusted stock price, 1963 to 2012

share would require $10,000 to open the smallest economic position—too large an amount for some retail investors. To keep the stock attractive to such purchasers, the board may elect to split the stock, dividing each old share into a larger number of new shares. The value of the company does not change, while each share is worth proportionately less (but each holder or an old share now owns *N* new shares [Figure 12.1]).

A 2 for 1 split, for instance, converts each old share into two new shares, so an owner of 100 shares before the split will own 200 shares after the split. But because the company's value hasn't changed, the split should not, in itself, change share values, beyond the arithmetic consequence of a split. A medium pizza is still a medium pizza, whether it is sliced into 4, 6, or 8 pieces.

Splits are infrequent, but they can still happen multiple times over a long time period. IBM, for example, has split several times in the 50 years between 1962 and 2012 (see Figure 12.1). Observe the steep drop in actual price in 1979 from about $300 per share to about $75 per share. This price change reflects a 4 for 1 split. Note that even though the price

changed substantially, if you held a portion of stock before the split, the total value of that stock, and your portion of ownership of the company did not change.

Historical prices are adjusted so you can more accurately and easily evaluate how portfolio values would have grown over time.

IBM's actual 1962 share price of over $500 after adjustment corresponds to about $2.50 per share whose fractional ownership corresponds with the same fraction a share of stock confers today—an over 200-fold difference. Note that in 2012, IBM was trading for about $200, so it had appreciated about 80-fold (from $2.50 to $200) over 50 years—a 9 percent annual return.

Stock price data services adjust for splits. If on January 1, 2012, a stock split 2 for 1, then all prices prior to that date will be halved to account for the effect of the split. Where actual prices would show a 50 percent reduction in the stock price in a single day, the adjusted price would reflect the real long-term trend, independent of a split. When you conduct event studies or backtest trading strategies, it is crucial that you use prices that have been adjusted for splits.

Some boards avoid splits, probably in the belief that a high stock price will tend to attract long-term investors and discourage casual traders. As of this writing, Apple (AAPL), for example, sells for about $500 per share, making each 100-share trade worth about $50,000. Berkshire Hathaway (BRKA), Buffett's firm, has never split despite 20 percent+ annual price appreciation for nearly 50 years, and is currently priced well above $100,000 per share. However, in the mid-2000s, Berkshire created a second share class (Class "B," BRKB) priced at 1/1500th of the "A" shares.

Reverse splits are also possible. In a "reverse split"—say, 1 for 3—the stock price rises to reflect a share's increased ownership—in this example, a share controls three times as much as it did before the reverse split. A stock priced at $30 before a 1 for 3 reverse split would be priced at $90 afterward.

Dividends

Shareholders can earn income from stocks without selling them if the company board of directors declares a dividend. Dividends are a return to

the shareholders of cash that is a portion of the company's annual earnings. A stock's *payout ratio* indicates this proportion. Boards can declare regular dividends (usually quarterly), or special, one-time dividends.

This means that the long-term value of owning a stock is not represented simply by its actual price, but also by the value shed over time in the terms of these dividends. The income-generating potential for a stock is measured by its dividend yield: the amount of its annual dividend per share divided by the share price. The average dividend yield presently is about 2 percent. Some companies pay dividends as high as 10 percent, and others (like BRK, or AAPL until quite recently) pay no dividend at all.

When a company pays a dividend, it distributes cash from its holdings to shareholders. So the company's assets are reduced, which typically reduces the share price by an equivalent amount. Even though the share price was reduced, the holder of the stock still captures the corresponding value. IBM share prices, note how the actual and adjusted prices differed before February 6, 2014. The difference is due to a dividend of $0.95 paid on that date.

To account for the benefit of the dividend paid on that date, we correct the adjusted price before that time proportionally. As an example, in the case of IBM's dividend of about 0.55 percent we adjust all previous, historical actual prices downward by that same amount. This ends up having the effect, in backtesting, of reaping a 0.55 percent "reward" on the date of the dividend.

Breaks in Series and Missing Data

A given stock, or all stocks, may not show price data for intervals if its trading was suspended (e.g., by its exchange), or if all trading is suspended. All trading was halted on the NASDAQ for just over 3 hours on August 22, 2013, because of faulty communications between the exchange's computers and those of the NYSE. In such situations, there will be no price data for some periods of time. Figure 12.2 shows the break in trading in the NASDAQ on that day.

Solvers such as QSTK cannot handle missing data; they will return error messages. So it is necessary to fill the data with reasonable guesses

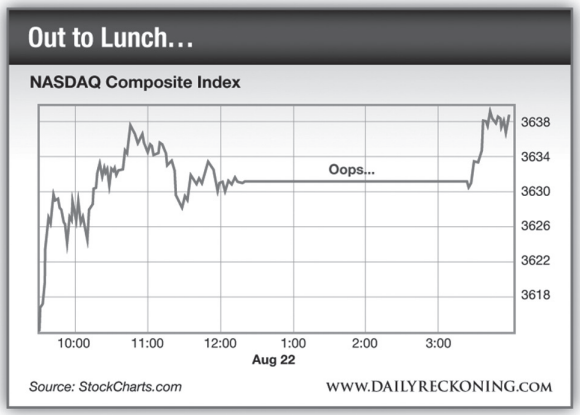

Figure 12.2 Three-hour break in NASDAQ stock price series due to unplanned trading suspension, August 22, 2013

as to what the data "would have been" if trading had occurred during the absent periods.

The common approach is to *fill forward*: to treat missing values as the same level as the last known value. So for minute-by-minute NASDAQ price data for August 22, 2013, the analyst would treat all values between 12:14 p.m. and 3:10 p.m. as the same value as at 12:13 p.m. (the last traded value). That's filling forward. as shown in Figure 12.2.

If values are missing at the beginning of the series, filling forward isn't an option. Then you would need to *fill backward*—use the first known value as also applying in the prior, missing periods.

The general rule for missing values is "fill forward first, and fill backward where you can't fill forward."

Missing Ticker Symbols

Stocks may be first listed in the middle of your time series, or delisted at some point, leading to missing data before or after the period where the stock existed. In the first instance, the company may have first gone public (begun trading its shares on an exchange). In the latter case, the company could have been acquired by another, gone private, or gone out of business entirely. Occasionally, companies change symbols; for example,

Sun Microsystems had at least three symbols in the roughly 25-year span between its listing in the 1980s and its acquisition by Oracle.

Analysts need to check to determine if sudden appearances and reappearances of symbols actually represent public company births and deaths, or merely labeling changes. For whole data sets, or for evaluations of managers, there can be a bias introduced if stocks or companies are omitted during a series. Commonly that company or fund has run into trouble, often fatal. (Hedge funds that shut down often have far underperformed their peers: according to research firm eVestment, only 30 percent of funds that existed 10 years ago are still in business today.) Analysts can scrub their sample to include only those companies that operated throughout the period, but this tends to omit more failures than successes, causing *survivor bias*.

Conclusion

These data problems are unavoidable. Capitalism is a dynamic process, which wreaks *creative destruction*—strangling poor performers, and (sometimes) buying up good performers. And market glitches happen—sometimes caused by hedge funds, as discussed in a later chapter. An analyst's goal should not be to choose a sample that avoids these data problems, because that would render his or her analysis virtually irrelevant. The challenge will be to refrain from accepting data uncritically, and also checking data for sanity.