this point. The idea is to get you familiar with handling financial data and seeing some of the ideas we have discussed.

To start, you will need to download and install R, a free software implementation of the S statistical analysis language. Information on how to do this is in Appendix A.

R is widely used in academia and industry, so learning to use it is beneficial.[18] That's not to say that Python is not useful; however, the learning curve for R is more gentle than that for Python. Furthermore, R is the language in which academic statisticians largely develop new methodology.

Instructions

Use the code below to answer the questions. We will use the `xts` package for handling the time series since it aligns dates and makes a lot of drudgery easy. We will also use the `quantmod` package and `Quandl` package to download data.

> ⚠ **Beware!** *At the time of this writing, Google Finance data (a common source for stock and mutual fund prices) has been turned off and Yahoo Finance data has sometimes been limited. You can still use* `quantmod` *to access data from Yahoo (for now) as well as other sources. The* `Quandl` *package can access some stock prices updated by a community of users as well as inexpensive end-of-day closing prices. Other companies offer similar services, such as Alpha Vantage, YCharts, and* `eodhistoricaldata.com`.

Look at the data for the following instruments:

- Fixed Income: We want to look at some US Treasuries over time. Unfortunately bonds mature; and, there may not be a bond maturing in exactly 2 years on a given day. So Fed economists kindly create pseudo-bonds to replicate the behavior of bonds for certain key maturities. Yields for these constant-maturity treasuries are often quoted as measures of the bond market. Also grab the Moody's seasoned Baa corporate bond yield index.
- Real Estate: Download near futures for the Dow Jones US Real Estate index and 3 REITs: Annally (a US mortgage REIT), Equity Residential (a US equity REIT), and Canadian Apartment Properties (a Canadian equity REIT).
- Commodities: Download settlement prices of near futures for the GSCI index, corn, soybeans, WTI crude oil, and US natgas (Chicago Mercantile Exchange contracts); and, copper, and steel rebar (Shanghai Futures Exchange contracts).
- Equity Indices: Fetch the adjusted close price for the S&P 500, Russell 2000, TSX Composite, IPC, FTSE 100, Euro Stoxx 50, SMI, Hang Seng, Straits Times, and Nikkei 225. Why not grab the TSX 60, CSI 300 or KOSPI? Permissioning on downloads made them difficult to get.

[18] I might be biased, having used R for over 15 years and having been a co-organizer of the R/Finance conference (`http://www.rinfinance.com`) since 2009.

- FX Rates: Download settlement prices of near futures for USD/JPY, USD/EUR, USD/CAD, and AUD/NZD.

1. Using the R code, explore the above instruments. Start by creating a summary of prices (or yields). Which instruments seem to be fairly stable? Which are not so stable — and tend to drift widely? Which tend to stay in a small range but have rare outliers?

2. The R code also creates plots. Which instruments move similarly? Do they sometimes move around much more than at other times? Do they sometimes wander far away from their typical level? You may even see a couple of data errors, for example a series with an inexplicable spike in it.

3. I previously mentioned problems with geometric indices. We can see this with some simple simulation. Simulate five series r_{it} of simulated returns: random variates for $i \in \{1,\ldots,5\}$ and $t \in \{1,\ldots,10000\}$. The variates should be iid normal $r_{it} \overset{iid}{\sim} N(0, 0.04^2)$. You can generate a sequence of returns with R's `rnorm` function: `r1 <- rnorm(10000, 0, 0.04)`

The geometric index return is $r_{gi,t} = \sqrt[5]{\prod_{i=1}^{5}(1+r_{it})}$ while an equally-weighted portfolio would have return of $r_{ew,t} = \frac{1}{5}\sum_{i=1}^{5}(1+r_{it})$. Run the simulations and compute summary statistics (using `summary`). Would you rather hold the index or the equally-weighted portfolio?

4. Can you prove the behavior you saw in the previous question is always true? Often true? Hint: An inequality might help.

R Code

```
library(xts)
library(Quandl)
library(quantmod)

# Get constant-maturity (US) Treasuries
ust.tickers <- c("FRED/DGS3MO", "FRED/DGS2", "FRED/DGS10", "FRED/DGS30")
ust.raw <- Quandl(ust.tickers, type="xts")
ust.colnames <- c("UST3M", "UST2Y", "UST10Y", "UST30Y")
colnames(ust.raw) <- ust.colnames

corpbond.raw <- Quandl("FED/RIMLPBAAR_N_B", type="xts")
colnames(corpbond.raw) <- c("USBaaCorp")

re.tickers <- c("CHRIS/CME_JR1", "NLY", "EQR", "CAR-UN.TO")
re.raw <- Quandl(re.tickers[1], type="xts")["Settle"]
# Sadly, there is a data error: a few days reported 10x price
re.raw[re.raw>1000] <- re.raw[re.raw>2000]/10
for (j in 2:length(re.tickers)) {
    tmp <- getSymbols(re.tickers[j], src="yahoo", env=NULL)
    adj.col <- last(colnames(tmp))
    re.raw <- cbind(re.raw, tmp[,adj.col])
}
colnames(re.raw) <- c("DJUSRE", "NLY", "EqRes", "CdnApt")
```

non-bank activities (like insurance or investment banking) at any of their subsidiary firms. Sometimes, there were turf wars: The SEC and CFTC could not agree on which would regulate futures on single-stocks or narrow indices. In 1982, they encoded this dispute into law by banning these futures with the **Shad-Johnson Accord**, named for the heads of the SEC and CFTC.

3.9 Quiz

Try answering the following questions in five minutes.

1. What is the difference between the primary and secondary market?

2. The NYSE and Nasdaq stock markets use continuous trading, and not auctions, to set prices. | T F

3. Circle one for each choice:
A (limit held market stop loss) order trades immediately.
A (limit held market stop loss) order sacrifices immediacy to specify a worst-possible execution price.

4. What are the names for the prices at which an intermediary (*e.g.* a specialist or market maker) will buy and sell?

Intermediary buys at the _____.

Intermediary sells at the _____.

5. What are the names for the prices at which a retail investor sending a market order will buy and sell?

Retail investor's market order buys at the _____.

Retail investor's market order sells at the _____.

6. Some people think all this high-speed trading is crazy. They propose that we should just have one auction each day for trading in each stock. If we were to make this change and move to one auction/day:

a) What effects would we likely see?

b) Why would we see those effects?

7. A market maker collecting limit buy and sell orders is "bookbuilding." | T F

8. Which act, over 70 years old, split commercial and investment banking? (You need not give the year for the act.)

9. Theory and data show that self regulation tends to yield greater economic growth than regulation by the government — even if both have the same goal. Why might this be?

10. Suppose I want to short shares of Motorola. It is big and in the S&P 500, so liquidity is not too much of a concern. Nonetheless: What do I need to do before I can sell those shares short?

3.10 Exercises

To flesh out a couple of the ideas we discussed, we will look at data from a couple of different types of markets and asset classes as well as some data for short selling.

Instructions

Modify the *R* code from Chapter 2 to answer the following questions.

Finding information about bid-ask spreads is a bit difficult in daily data; however, there is some data available. We will again use the `xts` and `Quandl` packages. Look at the data for the following instruments:

- Fixed Income: Download data from Quandl for Canadian government bond futures on the Montréal Exchange. Grab the bid and ask prices for the five- and ten-year bond futures: `CHRIS/MX_CGF1` and `CHRIS/MX_CGB1`. Unfortunately, these data can be noisy so you will need to eliminate any days where the bid or ask is 0.
- Commodities: Download data from Quandl for metals on the London Metal Exchange. Grab the buyer and seller prices for "cash" buyer and seller (bid and ask) for copper, aluminum, and tin: `LME/PR_CU`, `LME/PR_AL`, and `LME/PR_TN`.
- Equity Indices: Download data from Quandl for equity index futures on the Hong Kong and Montréal Exchanges. Grab the bid and ask for the Hang Seng and TSX 60: `CHRIS/HKEX_HSI1` and `CHRIS/MX_SXM1`
- FX Rates: Everyone is always interested in crypto-currencies. So download data from Quandl for the BitFinEx. Grab the bid and ask exchange rate between Bitcoin and USD: `BITFINEX/BTCUSD`.
- Equities: Finally, look at short volume versus stock price for three names: AMD, RGR, and TSLA. Download short volume from Quandl (`FINRA/FNSQ_AMD`, `FINRA/FNSQ_RGR`, and `FINRA/FNSQ_TSLA`) and adjusted close prices from Yahoo.

need not give the year for the act.)

9. Theory and data show that self regulation tends to yield greater economic growth than regulation by the government — even if both have the same goal. Why might this be?

10. Suppose I want to short shares of Motorola. It is big and in the S&P 500, so liquidity is not too much of a concern. Nonetheless: What do I need to do before I can sell those shares short?

3.10 Exercises

To flesh out a couple of the ideas we discussed, we will look at data from a couple of different types of markets and asset classes as well as some data for short selling.

Instructions

Modify the *R* code from Chapter 2 to answer the following questions.

Finding information about bid-ask spreads is a bit difficult in daily data; however, there is some data available. We will again use the `xts` and `Quandl` packages. Look at the data for the following instruments:

- Fixed Income: Download data from Quandl for Canadian government bond futures on the Montréal Exchange. Grab the bid and ask prices for the five- and ten-year bond futures: `CHRIS/MX_CGF1` and `CHRIS/MX_CGB1`. Unfortunately, these data can be noisy so you will need to eliminate any days where the bid or ask is 0.
- Commodities: Download data from Quandl for metals on the London Metal Exchange. Grab the buyer and seller prices for "cash" buyer and seller (bid and ask) for copper, aluminum, and tin: `LME/PR_CU`, `LME/PR_AL`, and `LME/PR_TN`.
- Equity Indices: Download data from Quandl for equity index futures on the Hong Kong and Montréal Exchanges. Grab the bid and ask for the Hang Seng and TSX 60: `CHRIS/HKEX_HSI1` and `CHRIS/MX_SXM1`
- FX Rates: Everyone is always interested in crypto-currencies. So download data from Quandl for the BitFinEx. Grab the bid and ask exchange rate between Bitcoin and USD: `BITFINEX/BTCUSD`.
- Equities: Finally, look at short volume versus stock price for three names: AMD, RGR, and TSLA. Download short volume from Quandl (`FINRA/FNSQ_AMD`, `FINRA/FNSQ_RGR`, and `FINRA/FNSQ_TSLA`) and adjusted close prices from Yahoo.

1. Using your modified *R* code, create a column to hold the bid-ask spread. If you name the columns for the bid and the ask correctly, you can just do this:

```
alldata$CGB1.spread <- alldata$CGB1.ask - alldata$CGB1.bid
```

Create a spread column for each of the fixed income, commodity, equity index, and FX instruments. Plot the spreads, all over the same time period, and comment on trends, unusual events, and commonalities.

2. Create fractional spreads, like so:

```
alldata$CGB1.fracspread <- log(alldata$CGB1.ask) - log(alldata$CGB1.bid)
```

and plot these spreads as in the prior question. Comment on trends, unusual events, and commonalities.

3. Look at the summaries for the spreads and fractional spreads. Which seem to be better behaved: arithmetic or log-spreads? Why?

4. Finally, plot the short volume and adjusted close for the equities, making sure all plots are for the same time scale. Do you notice any relationships between the short volume and price plots?

2. Electronic markets have higher fixed/startup costs: | T | F

3. What type of market venue is known for the fastest executions, shows limit orders placed by others, and matches buyers and sellers when they agree on price?

4. Give two reasons why SEC Rules 11Ac1-4–6, decimalization, and (ultimately) Reg NMS were passed.

5. How much did spreads change (indicate up/down and size of change) in response to these rule changes?

6. What is meant by *program trading*?

7. I want exposure to the S&P 500 index. I can buy ETFs (traded at the NYSE or an ECN) or index futures (traded at the CME). I am indifferent between these choices, but I worry about grossly overpaying if I don't monitor both markets in deciding where to trade.

a) Should I be very concerned about this? (Or, put another way: am I likely to grossly overpay at one venue versus another?)

b) Why should I or why should I not be concerned?

8. High frequency traders use high-speed optimization and complex mathematics to make money on most trades: | T | F

9. The Flash Crash (2010) and the trading glitches at Knight Capital (2012) would seem to suggest that electronic mishaps have increased greatly since a decade ago. But Gao and Mizrach (2016) show this is not the case. Why do glitches *seem* to be more common?

10. (Bonus!) Which three acts have affected the split between commercial and investment banking? Hint: One act split them, another reunited them, and a third curtails their ability to take trading risks.

4.9 Exercises

Instructions

Download the 200204, 200604, 201004, 201404, and 201804 execution quality reports at `ftp://doe.chx.com`.

You will examine the execution quality of marketable orders using field definitions

at `http://www.chx.com/trading-information/execution-quality/`. First, however, we will look at a few applications of statistics to what we just covered.

1. If we look at Figure 3 on page 1823 of the Christie and Schultz paper, we see the histogram for 372,625 NASDAQ quotes. Assume the even eighths each have a frequency of 21% and the odd eighths each have a frequency of 4%. What is the probability of this divergence from a uniform distribution occuring randomly?

2. Assume the unconditional density for prices p has a finite expectation p_0 and a *cádlág* cdf F. A large order comes in to the order book at $p_0 + c$. Suppose this shifts some fraction γ of the probability mass between c and ∞ to a point mass at c. Show that this imparts a negative bias to the (conditional) price process.

3. Using the downloaded execution data and the R code below as a starting point, compute for each month and size group the overall average of the following metrics for marketable stock orders:

   a) average realized spread for covered order executions;

   b) average effective spread for covered order executions; and,

   c) share-weighted average time from order receipt to execution (at the quote).

Plot each of the three average metrics across time with each size group on the same plot (so that you have three plots).