Module 2D: Futures Prices, Spreads and Basis

# Learning Outcomes

After completing this module students should be able to:

* Describe at a general level the equations which comprise the eight-quarter pricing model.
* Describe the forecasting procedure which is incorporated into the model and used to generate futures pricing uncertainty.
* Explain how the forward curve of futures prices reflects the saw-toothed pricing pattern (i.e., pricing seasonality).
* Calculate price spreads and use regression to identify causal relationships between production and long term demand forecasts and the spreads in futures contract prices.
* Calculate the basis and explain the existence of non-anticipated variations in the basis over time.

# Goal of this Module

This module is the first in a series which investigates commodity futures and spot markets with particular reference to the U.S. corn market. In Module 2C an eight-quarter (two year) pricing model was constructed and calibrated to the U.S. corn market. There it was shown that a key feature of modeling commodity prices is convenience yield because the inclusion of this feature is required to generate realistic simulated price paths. In this module non-anticipated changes in USDA corn production forecasts and long term corn demand forecasts are added to the eight-quarter pricing model in order to generate realistic levels of randomness in the pricing outcomes. Most of this module is devoted to a detailed analysis of 5000 sets of eight-quarter random pricing outcomes.

An important feature of real world commodity markets is price risk. Indeed, price risk is largely what motivates hedgers to hedge and institutional investors who seek portfolio diversification to hold commodity futures. The analysis of risk is beyond the scope of this course and so we will restrict our attention to expected return. We examine expected return from the perspective of both a hedger and an institutional investor. However, a lot needs to be covered before we get to the point of explicitly analyzing expected returns.

# Background

It is important to have a good understanding of real-world commodity prices before attempting to model pricing outcomes. Commodity futures have two distinct dynamic components. The “snap shot” component consists of the prices of the set of contracts with various future expiry dates (e.g., December 2021, March 2022) at a particular point in time (e.g., October 31, 2021). The “time path” component consists of the time path of a particular futures contract (e.g., December 2021) over time. The first shows how the commodity’s price is expected to evolve over time and the second shows how the commodity’s price actually stochastically evolves over time. A graph of the snapshot of prices of all corn contracts at a particular point in time is commonly referred to as the commodity’s forward curve.

Recall from Module 2A that an important property of commodity prices is that the spot and futures price at the same location (e.g., Chicago) must be equal when the futures contract expires. This means that we can include the spot price as the first price in the forward curve. The basis is defined as the spot price minus the price of one of the futures contracts. The spread is defined as the price of a futures contract with a more distant expiry date minus the price of a futures contract with a less distant expiry date. The fact that basis and spreads are defined in opposite ways can be confusing. More confusing is that sites such as BarChart define spread as the less distant contract price minus the more distant contract price since this ensures a consistency with the definition of basis.

Much of this module will focus on spreads and the basis. In the next module the role of basis is expanded by redefining basis as the spot price in location X minus the futures price. In this more general case, the basis will contain both the forward curve component, which is the focus of this module, and a spatial LOP component, which was the focus of previous modules. A good place to examine how spreads evolve over time is [Barchart](https://www.barchart.com/futures/quotes/ZC*0/futures-spreads). Their interactive dashboard allows you to choose from a large number of contract expiry months and plot the spread data for up to one year. A similar tool does not exist for basis because there are hundreds of basis data being reported each day because each unique location will have a unique basis.

Spreads tell us a lot about how traders are perceiving the availability of stocks at different time periods. We will mainly focus on the spread for futures contracts which are in different crop years. We will learn that arbitrage prevents the spread from exceeding the carrying cost between the two contracting periods. However, arbitrage cannot prevent the spread from falling below the carrying cost and possibility taking on a negative value. A positive spread involving contracts which span crop years is indicative of sizeable stocks being carried over from one crop year to the next. On the other hand, a negative spread is indicative of a current shortage relative to the supplies which will be available in the future.

If you trade futures you can reduce your risk by trading the spread. To be successful you will need to understand which factors determine changes in the spread over time. For example, if forecasted corn yields are lower in a USDA crop production report, should you expect the cross-year spread to increase or to decrease? What about if China unexpectedly announces that it will purchase a larger volume of corn than what the market was anticipating? In the November 2, 2021 slide deck for this course a detailed analysis of the December 2021 to May 2022 price spread reveals much about the surge in demand for commodities as the global economy began to rebound after the worst of the COVID-19 crisis was over.

As will be shown in a later module, understanding the determinants of the basis is crucial for effective hedging. An important goal of hedging is to reduce risk but managing expected profit when constructing the hedge is also very important. The profit earned by a short hedge can be measured as the change in the basis between when the hedge was initiated and when it was lifted. The opposite is true for a long hedge. The basis changes systematically over time, depending on the particular saw-toothed pattern within the forward curve. The basis also changes randomly over time as the price of futures contracts unexpectedly increase or decrease relative the spot price in response to the release of new information to the market.

The November 2, 2021 lecture slides contain interesting and relevant background material on pricing relationships in the U.S. corn market. You are encouraged to review those slides before reading the rest of the material in this module.

# Theoretical Considerations

Two important theoretical restrictions on commodity prices concern the maximum spread between prices of future contracts with different expiry months, and the connection between futures prices and expected spot prices. Each is discussed in turn.

## Arbitrage and Price Spreads

A key theoretical result is that the price spread between a pair of contracts should not exceed the cost of carrying the commodity between the expiry date of the nearby contract and expiry date of the more distant contract. For example, suppose on November 1, 2021 a trader observes that December 2021 corn futures are trading at $5.20/bu and May 2022 corn futures are trading at $5.32/bu. Further suppose the cost of carrying corn between December 2021 and May 2022 is $0.10/bu. This situation can be arbitraged because on November 1 a trader could take a long position in the December market and a short positon in the May market. When December arrives the trader would accept delivery of the corn and make payment of $5.20/bu. The corn would be put into storage and when May 2022 arrives the corn will be delivered against the outstanding short contract and payment of $5.32/bu will be provided to the trader. The trader earns $0.12/bu on the pair of futures transactions (ignoring commission fees, interest on margin money, etc), which is $0.02/bu above the storage cost. If many traders do the same transaction the December 2021 futures price will be bid up and the March 2022 futures price will be bid down until profits on the spread trade disappear.

There is no corresponding arbitrage which ensures that the spread does not fall below the carrying cost, possibly to a negative value. For example, in the example provided above suppose the spread was $0.08/bu instead of $0.12/bu. Arbitrage would involve taking a short December 2021 position and a long May 2022 position. The trader would like to accept delivery in May 2022 and then return to December 2021 to make delivery against the short position. The negative of the storage costs would earned as revenue. Of course this going back in time transaction is not possible. For this reason there is no lower limit on the price spread between a pair of futures contracts.

## The Futures Price - Expected Spot Price Assumption

The eight quarter pricing model which was introduced in Module 2C had no uncertainty, and so the prices which were generated showed precisely how the spot prices will change over time. In this module uncertainty is introduced and so now the prices which we solve for beyond the current quarter must be interpreted as expected spot prices. The goals is to develop a model of futures prices and so we need a way to connect future prices to expected spot prices. We will assume the simplest case which is commonly referred to as the “expectations hypothesis”.

**Expectations Hypothesis: The price of a futures contract which expires in quarter is equal to the spot price which traders, on average, believe will exist in quarter .**

The expectations hypothesis is appropriate for risk neutral traders. If we ignore the small amount of capital that a trader has invested in her margin account, then it follows that the trader has zero opportunity cost of holding a futures contract. With risk neutrality, competition will drive the return from holding the futures contract down to zero. If a trader expects to earn zero from holding a futures contract then it must be the case that the futures price is expected to remain constant over time. If we combine this result with the arbitrage result from Module 1, which is that the spot price and the futures price must converge when the futures contract expires, we end up with the expectations hypothesis.

In the early literature on commodity futures Kenyes (1939) suggested that risk averse short hedgers (e.g., farmers) were willing to pay the risk premium which was demanded by long speculators. The risk premium biased the futures price below the expected future spot price, and the amount of the downward bias was the expected loss for the hedger and the expected gain for the speculator. This theory, which was popular for many years, was known as Keynsian normal backwardation. In modern portfolio theory, the amount of risk premium which is required by an investor who holds a continually-rolled long position (i.e., an index fund) depends on the amount of risk which cannot be diversified away. Some even suggest that investors are willing to pay a risk diversification premium to hold commodities in their portfolio. In that case the futures price will be above the expected spot price. There is a large empirical literature which attempts to estimate the size of the risk premium. The results are mixed, and at best the estimated risk premium is small. Consequently, for this analysis it is reasonable to proceed with the expectations hypothesis assumption.

# Non-stochastic Prices

The eight quarter simulation model, which was introduced in Module 2C, was built in Excel, and it has not yet been transferred to R. This means that simulated prices from Excel will be imported into R for a detailed analysis. Before beginning this analysis it is useful to a briefly review the eight quarter simulation model.

## Model Equations and Solution

The model consists of eight quarters with three equations per quarter. The first equation is the intertemporal LOP equation which requires prices to change over time according to the marginal carrying cost. The marginal carrying cost, which is equal to the marginal cost of storage net of the marginal convenience yield, adjusts according the level of outstanding aggregate stocks which are being carried through time. In this model the market never stocks out and so it is sufficient to use only the first part of the LOP. The relevant equation can be written as:

The second equation is the harvest and stock de-accumulation equation which ensures that for each time period the ending stock balance, , is equal to the beginning stock balance, , plus harvest if in Q1 or Q5, minus consumption, .

By assumption, the market begins in Q1 with incoming stocks, , plus Q1 harvest, . As well, the market must carry stocks out of quarter 8 equal to . Incoming stocks, , and outgoing stocks, , are “pipeline” stocks. These pipeline stocks do not play an important role in our analysis. They were added to for the purpose of model calibration, and throughout the full analysis .

The third equation is inverse demand, which links consumption and the price of the commodity.

It is important to keep in mind that stocks, consumption and price are fully endogenous (i.e., there is no one way causality within this set of variables). The model has four exogenous variables, two of which were discussed above. The remaining two exogenous variables are the size of the harvest in Q5, , and the level of stocks carried out of Q8 in excess of the pipeline stocks, . Both of these variables play a particularly important role in the analysis because they will be converted to random variables subject to forecasting. As will be shown, it is the on-going revisions to the forecasts for these two variables which generate pricing uncertainty within the model. For the base case, , = 14.377, and . All quantity variables are measured in billions of bushels and price is measured in dollars per bushel.

The system of 24 endogenous variables within 24 equations can be solved to generate 8 quarterly equilibrium prices, 8 quarterly equilibrium consumption levels and 8 quarterly equilibrium storage levels. If the values of any of the four exogenous variables are changed the model must be resolved to generate a new set of equilibrium prices, consumption and storage. In the simulations which are used to generate the flow of uncertain prices, the model is resolved and new equilibrium prices are generated at the beginning of each quarter because this is when the forecasts of the and variables are revised.

## Setting up the What If Analysis

In module 2C the non-stochastic version of the eight-quarter model was used to conduct a small number of what if analysis. Some additional what if analysis with the non-stochastic model is conducted in this section to shed light on how a key cross-year price spread is impacted by changes in the Q8 stock carry out as measured by , and changes in Q5 production as measured by . The outcomes of these two what if outcomes provide key intuition for the results below.

Five sets of eight quarterly prices were generated in the Excel model and saved to a csv file titled *price\_scenarios\_nonstochastic.csv*. The five scenarios are as follows:

* Base
* High Q8 stocks demand ()
* Low Q8 stocks demand ()
* Low Q5 production ()
* High Q5 production ()

After importing these prices into R the goal is calculate the Q4 to Q5 carry over, as measured by the variable. This carry over volume is important because it is responsible for the transmission of price shocks across the two crop years. We can calculate the Q4 to Q5 carry over volume by subtracting the sum of consumption for Q1 through Q4 from initial stocks, which are assumed to equal . Quarterly consumption can be calculated using the imported prices and the inverted demand equation: . The rather long code chunk below accomplishes these tasks.

pacman::p\_load(here, dplyr, ggplot2, gridExtra)  
  
# load the non-stochastic prices for the five scenarios  
P\_ns <- read.csv(here("Data", "price\_scenarios\_nonstochastic.csv"), header=TRUE, sep=",", stringsAsFactors = FALSE)  
  
# define demand curve parameters  
a <- 16.21  
b <- 3.5  
stock\_begin <- 14.38 + 2.015  
  
# create consumption values for the five price scenarios  
X\_ns <- P\_ns %>% mutate(  
 X\_base = a/b - 1/b\*P\_base,  
 X\_D\_high = a/b - 1/b\*P\_D\_high,  
 X\_D\_low = a/b - 1/b\*P\_D\_low,  
 X\_H5\_low = a/b - 1/b\*P\_H5\_low,  
 X\_H5\_high = a/b - 1/b\*P\_H5\_high) %>% select(X\_base,X\_D\_high,X\_D\_low,X\_H5\_low,X\_H5\_high)  
  
# calculate Q4 ending stocks for five price scenarios and create a row vector  
S4\_base <- stock\_begin-X\_ns$X\_base[1]-X\_ns$X\_base[2]-X\_ns$X\_base[3]-X\_ns$X\_base[4]  
S4\_D\_high <- stock\_begin-X\_ns$X\_D\_high[1]-X\_ns$X\_D\_high[2]-X\_ns$X\_D\_high[3]-X\_ns$X\_D\_high[4]  
S4\_D\_low <- stock\_begin-X\_ns$X\_D\_low[1]-X\_ns$X\_D\_low[2]-X\_ns$X\_D\_low[3]-X\_ns$X\_D\_low[4]  
S4\_H5\_low <- stock\_begin-X\_ns$X\_H5\_low[1]-X\_ns$X\_H5\_low[2]-X\_ns$X\_H5\_low[3]-X\_ns$X\_H5\_low[4]  
S4\_H5\_high <- stock\_begin-X\_ns$X\_H5\_high[1]-X\_ns$X\_H5\_high[2]-X\_ns$X\_H5\_high[3]-X\_ns$X\_H5\_high[4]  
   
s4 <- cbind(S4\_base,S4\_D\_high, S4\_D\_low, S4\_H5\_low, S4\_H5\_high)

## What If Results and Analysis

We can now display the imported prices and calculated stocks which carry over from Q4 to Q5.

print(P\_ns, digits = 4)

## Quarter P\_base P\_D\_high P\_D\_low P\_H5\_low P\_H5\_high  
## 1 1 3.488 4.289 2.687 4.33 2.645  
## 2 2 3.651 4.458 2.843 4.50 2.801  
## 3 3 3.706 4.527 2.884 4.57 2.841  
## 4 4 3.654 4.496 2.811 4.54 2.767  
## 5 5 3.494 4.365 2.623 4.41 2.578  
## 6 6 3.657 4.563 2.751 4.55 2.764  
## 7 7 3.712 4.661 2.762 4.59 2.834  
## 8 8 3.660 4.661 2.659 4.53 2.789

print(s4, digits = 4)

## S4\_base S4\_D\_high S4\_D\_low S4\_H5\_low S4\_H5\_high  
## [1,] 2.011 2.947 1.076 2.995 1.028

First consider the base case set of prices and Q4 carry over, which are in the first column. Notice that the prices in Q5 through Q8 are largely a repeat of the prices in Q1 through Q4. As well, the Q4 carry over amount, 2.011, is approximately equal to the pipeline stocks which are carried into Q1 and out of Q8. This is to be expected because the two years have been constructed to be essentially a repeat of each other. We will use the price in Q3 as representative of the first year, and the price in Q7 as representative of the second year. In the base case, the Q3 and Q7 prices are both approximately equal to 3.71.

The second column shows the pricing and Q4 carry over impacts which result from a large surge in demand for Q8 carry out stocks (increasing from 0 to 2). Of course we expect prices to go up given this demand surge but it is worth asking what exactly causes prices in the first year to increase given that the demand surge is happening at the end of the second year. The answer lies with the Q4 stock carry over variable. Notice that this variable has increased from 2.011 in the base case to 2.947 with the demand surge. The higher volume of stocks carried out of year 1 result in lower consumption and thus higher prices in year 1.

The additional stocks carried out of Q4 as compared to the base case is equal to . This is slightly less than half of the added amount which is carried out of Q8 as compared to the base case. The outcome that new stocks arriving in Q5 are less than new stocks leaving Q8 is expected given that storage is costly. The differential in stock movement means that there will be less overall consumption in year 2 as compared to year 1. With lower consumption in year 2, the prices must be higher in year 2 than in year 1. This can be readily confirmed by observing the prices in the second column. The Q3 to Q7 price spread has increased from approximately 0 in the base case to . The price spreads are positive for the other quarters as well. In general, the demand curve has caused the pricing schedule to slope up. This is an important result that is the key to the analysis of institutional investment in commodity futures, which is the topic of Module 2F.

The third column of the above pricing results is the mirror image of the second column. In this case the Q8 carry out demand is very weak because the value of has been reduced from 0 to -2. The weak Q8 carry out demand causes prices in year 2 to fall. The lower year 2 prices induces traders in year 1 to carry less stock from Q4 to Q5, and this results in lower prices in year 1. The second entry in the *s4* table shows that Q4 carry over decreases from 2.011 in the base case to 1.076 for the case with weak Q8 carry out demand. Similar to the previous case of a demand surge, the reduction in the Q4 stock carry out is a little less than half of the reduced demand in year 2. This imbalance causes the prices in year 2 to fall further than the prices in year 1. For example, the Q3 to Q7 price spread has decreased from approximately zero in the base case to . In general, the weak demand for Q8 carry out stocks has created a negative Q3 to Q7 price spread, and has caused the entire pricing schedule to slope down.

It is tempting to conclude that a 2 unit reduction in Q5 production will have the same pricing impact as a 2 unit increase in Q8 carry out stocks because both reduce available supply in year 2 by the same amount. The same holds for a 2 unit increase in Q5 production and a 2 unit decrease in Q8 carry out stocks. The last two columns of the above set of prices correspond to a 2 unit decrease and increase in Q5 production, respectively. The prices respond in a similar way to the demand shocks but there is an important difference. Notice that with a 2 unit change in Q5 production, the change in the Q3 to Q7 price spread is much smaller than the analogous impact in the case of the demand shock. In other words, the slope of the pricing schedule changes very little in response to the sizable positive and negative shocks to Q5 production.

To understand this result we need to discuss the role of carrying costs as a determinant of prices and price spreads. In the case of the surge in demand for Q8 carry out stocks the higher set of prices which result in both years reflect not only the higher demand but also the cost of carrying stocks through time to meet the higher Q8 demand. As a rough approximation you can think of the added storage costs being equal in both years. In year 1 stocks are carried forward and combined with the Q5 stockpile. In year 2 stocks are carried forward and added to the Q8 stockpile. This means that carrying costs result in somewhat higher prices when there is a demand surge but they do not change the tilt of the pricing schedule.

When there is a Q5 production shortfall of the same size as the Q8 demand surge, prices will rise in both years but will rise more in the second year, similar to that of the Q8 demand surge. However, the lower production implies less overall stocks moving from quarter to quarter and thus lower overall carrying charges. The reduced carrying charges partially reverses the price increase rather than strengthening it as in the case of the Q8 demand surge. The critical piece to the puzzle is that carrying charges are reduced more in the second year than in the first year because it is the second year where the shortfall in production takes place. The direct effect of the shortfall in production is to increase the Q3 to Q7 price spread for the same reasons why a surge in demand for Q8 carry out stocks increased the Q3 to Q7 price spread. The indirect effect of the shortfall in production is to decrease the Q3 to Q7 price spread because carrying costs are reduced more in the second year than in the first year. The net effect, which is illustrated in the fourth column above, is that the impact on the Q3 to Q7 price spread is much smaller with a Q5 production decrease as compared to a Q8 carry out stock increase. The same logic holds and the results are the mirrow image of the current results for the case where Q5 is 2 units above its base case value (see the last column in the pricing data above).

# Stochastic Prices

For the remainder of this module we will work with the results from a stochastic version of the eight-quarter pricing model. The model has been calibrated to generate prices with realistic levels of pricing uncertainty. Working with this type of pricing brings us much closer to real world commodity prices. In real world markets, it is the sequence of new information which is continually arriving in the market which creates variability over time in the futures prices. The variability affects both the level of prices and the pattern of prices such as the price spreads between different contract months. Some of this new information is in the form of USDA forecasts about the size of the up-and-coming harvest, and the longer term demand for stocks (e.g., export demand).

For our analysis we assume that the outcome of the Q5 production variable, , is unknown in Q1 through Q4, and the outcome of the Q8 stock carry out demand variable, , is unknown in Q1 through Q7. The USDA provides a forecast of these variables at the beginning of each quarter. We will not model different beliefs by different speculators, which is a key condition for active trading of long and short positions. Rather, we will keep things simple by assuming that all traders believe the USDA forecast is 100 percent accurate. The spot price and the set of futures prices which we assume the competitive traders (rather trivially) negotiate will equal the set of prices which would be obtained if the forecasts are plugged into the non-stochastic version of the pricing model and then the model is solved. This equivalency is another application of the invisible hand since competition ensures a LOP outcome.

At the beginning of the next quarter when the USDA releases a revised forecast, the model is solved again with the revised forecasts in order to generate a revised set of prices. As before, these revised prices are the same as what we would expect to see emerge with competitive traders negotiating the spot price and prices in the future. By repeating this process for each of the eight quarters we end up with eight price series, one for each of the eight futures contracts. The prices are not reported for expired contracts and so the series of futures prices are truncated. Specifically, there will be a one-price series for a futures contract which expires in Q1, a two-price series for a futures contract which expires in Q2, etc. The full sequence for all eight futures contracts (including the Q1 contract which is expiring when the simulation begins) has 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 = 36 prices.

In the above discussion we assumed that the prices which emerge from the solution to the pricing model can be viewed as the prices of futures contracts which trade in a competitive market. It is worth digressing a bit to explain why this assumption is acceptable. Remember that futures prices are the expected spot price in the period that the futures contract expires. This means that the revised prices which emerge from the pricing model at the beginning of each quarter are expected spot prices. This only makes sense if the forecasts do not systematically trend up or down over time. That is, the forecasts must follow a random walk.

To model the random walk for Q5 production, let where is a random error term (note that the “hat” on the variable indicates that it is a forecast). Using time series methods, the variance of can be estimated. Now values for can be simulated, and this will generate a sequence of Q5 production forecast revisions which are required for the pricing model. A similar procedure is used for the Q8 carry out demand variable, .

## Simulation

The simulated forecasts and sets of equilibrium prices were previously generated in an Excel macro-enabled workbook. The simulated forecasts come from an econometric analysis of the USDA crop production and WASDE monthly forecasts (see Vercammen 2021 for details). One simulation consists of randomly generated forecasts for and , and a set of equilibrium prices for each of the eight quarters. Vercammen (2021) shows that the summary statistics of the 5000 simulated prices are reasonably similar to the summary statistics of real-world corn prices (e.g., same average price and a similar level of volatility).

The data from 5000 independent simulations were saved into eight .csv files (one file for each of the eight quarters). These files can now be pulled into this current session for formal analysis. Rather than importing the .csv files in the usual way, a separate R program called *price-reformat.R* imports the data and then pulls the data into a function called *get\_simulated()*. When this function is called in the current session by supplying a row number ranging from 1 to 5000, this function returns the set of futures prices in a pre-formatted matrix.

In additon to loading *price-reformat.R* it is necessary to import two .csv files which contain the quarterly forecasts for the and variables. The code chunk to follow loads the required packages and imports the forecast data.

source(here("Code","price-reformat.R"))  
harvest <- read.csv(here("Data", "demand\_forecast.csv"), header=TRUE, sep=",", stringsAsFactors = FALSE)  
demand <- read.csv(here("Data", "harvest\_forecast.csv"), header=TRUE, sep=",", stringsAsFactors = FALSE)

As noted, the simulation data has 5000 rows where each row corresponds to one complete set of simulated forecasts and prices. The value we assign to the *row\_select* variable identifies which of the 5000 simulated outcomes is selected for analysis. We begin by using the *get\_simulated* function to retrieve the set of futures prices.

row\_select <-3  
prices <- get\_simulated(row\_select)  
print(prices, digits = 4)

## pQ1 pQ2 pQ3 pQ4 pQ5 pQ6 pQ7 pQ8  
## [1,] 3.49 3.653 3.708 3.656 3.497 3.659 3.714 3.662  
## [2,] NA 3.548 3.602 3.548 3.386 3.549 3.605 3.552  
## [3,] NA NA 3.640 3.587 3.425 3.590 3.647 3.596  
## [4,] NA NA NA 3.889 3.730 3.901 3.967 3.927  
## [5,] NA NA NA NA 3.764 3.932 3.995 3.953  
## [6,] NA NA NA NA NA 3.629 3.689 3.642  
## [7,] NA NA NA NA NA NA 3.497 3.448  
## [8,] NA NA NA NA NA NA NA 3.397

It is important to read this pricing data correctly. Each row corresponds to a time period and a given set of forecast information. The diagonal elements are the spot prices and the elements to the right of the diagonal are expected values of the future spot prices. For example, in the fourth row which corresponds to the fourth quarter, the Q4 spot price is 3.889, the spot price which traders are expecting for Q5 is 3.730, the spot price which traders are expecting for Q6 is 3.901, etc. We can see by much the actual spot price is different from what traders were expecting the previous period by moving down one row. For example, in Q4 traders were expecting the spot price in Q5 to equal 3.730. However, the diagonal element in the fifth row shows that the actual Q5 spot price was 3.764. This means the information which arrived in Q5 was slightly bullish because the Q5 price turned out to be slightly higher than previously expected.

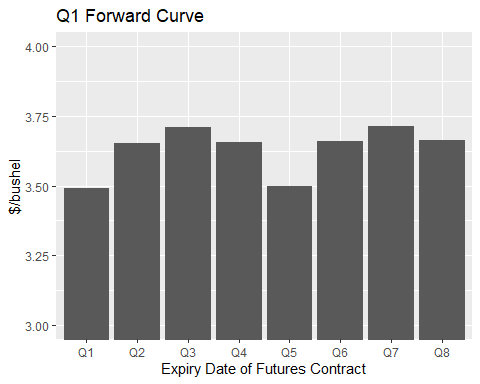
If we invoke the expectations hypothesis we can interpret the expected future spot prices as futures prices. This means that each column of the above pricing matrix can be interpreted as a distinct futures contract. Reading a column from top to bottom provides the price history of that contract. As before, the difference between a pair of cells in a column represents the extent that actual futures price turned out to be different than what traders were expecting the previous quarter. We do not expect the intertemporal LOP to hold for the prices within a column because the forecast information is different for each row.

Reading a row from left to right is the forward curve for a particular point in time. The forward curve provides a snap shot of the set of futures prices at a particular point in time. The prices which make up the forward curve must satisfy the intertemporal LOP because all of these prices are based on the same pair of forecasts (i.e., there is no pricing uncertainty within each row).

## Saw-Toothed Pricing Pattern

In Q1 the forecast is and , which are the base case values. We know two things: (1) the futures prices in the Q1 forward curve are the expected values of the spot prices when the respective contracts expire; and (2) the expected spot prices are the same as the simulated spot prices in a model with no uncertainty where and . These two outcomes together imply that the set of prices which make up the Q1 forward curve are necessarily identical to the base case prices which we closely examined in Module 2C. We can plot the Q1 forward curve to verify that this is the case using the following code:

forward <- as.data.frame(prices[1,])  
colnames(forward)<- "ForwardQ1"  
labels <- c("Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8")  
forward <- cbind(labels,forward)  
  
forward\_plot <-ggplot(data=forward, aes(x=labels, y=ForwardQ1)) +  
 geom\_bar(stat="identity") + ggtitle("Q1 Forward Curve") +  
 labs(x= "Expiry Date of Futures Contract") + labs(y= "$/bushel") +  
 coord\_cartesian(ylim=c(3, 4))  
forward\_plot



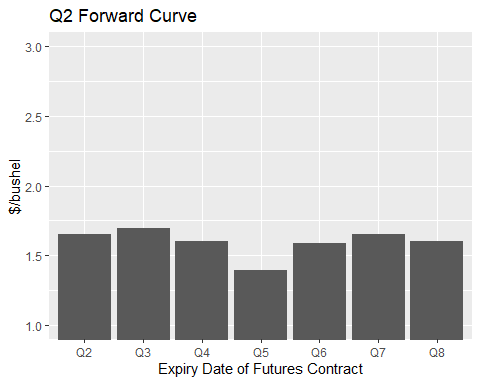
The above column chart shows that prices in the eight quarter have the saw-toothed pricing pattern with the convenience yield modification. Recall from Module 2C that convenience yield causes the prices to drop more gradually versus an abrupt drop as the first crop year finishes. Moreover, convenience yield allows the prices to drop even though there is positive carry over from the old crop year to the new crop year. In the above chart, prices rise when moving from Q1 through Q3, and when moving from Q5 through Q7 because of the cost of storage is higher than the convenience yield. Prices drop when moving from Q3 to Q5 and when moving from Q7 to Q8 because convenience yield is higher than the cost of storage.

We should not expect this particular pattern to remain intact when we move to Q2 and beyond. This is because the forecast in Q2 will typically be different than the forecast in Q1. A different forecast will typically involve a different volume of corn carried across crop years, and this will affect the set of expected future spot prices and thus the pricing pattern in the Q2 forward curve.

To illustrate how the pattern of the forward curve can change with the arrival of new information, let’s consider a more extreme forecasting outcome. In particular, we will retrieve the prices from the eighth row of the matrix of 5000 simulated forecasts and prices. Row 8 contains a highly unusual pair of forecasts for Q2. In particular, harvest is forecasted to take on a record high value of 17.74 billion bushels and the demand for Q8 carry out stocks in excess of the 2.015 billion bushels of pipeline inventory is forecasted to equal -0.736 billion bushels. This is a perfect storm for low prices because of the level of production in Q5 is forecasted to be exceptionally high and the level of Q8 stock demand is forecasted to be exceptionally low.

The Q1 forward curve for this new set of prices is the same as the previous set of prices because by assumption the Q1 forward curve contains the base case set of prices. Of interest is the Q2 forward curve and how its shape differs from the shape of the Q1 forward curve. The following code chunk generates a column chart of the Q2 forward curve in this unusual case.

row\_select2 <-8  
prices2 <- get\_simulated(row\_select2)  
  
forward2 <- as.data.frame(prices2[2,-1])  
  
colnames(forward2)<- "ForwardQ2"  
labels2 <- c("Q2","Q3","Q4","Q5","Q6","Q7","Q8")  
forward2 <- cbind(labels2,forward2)  
  
forward\_plot2 <-ggplot(data=forward2, aes(x=labels2, y=ForwardQ2)) +  
 geom\_bar(stat="identity") + ggtitle("Q2 Forward Curve") +  
 labs(x= "Expiry Date of Futures Contract") + labs(y= "$/bushel") +  
 coord\_cartesian(ylim=c(1, 3))  
forward\_plot2



The pricing level and saw-toothed pricing pattern in the above chart has changed significantly from the base case. As expected, prices are much lower due to the higher level of forecasted production and the weaker level of long term demand. Carry over from Q4 to Q5 is 0.278 billion tonnes, which is much lower than the 2.015 billion tonnes in the base case. This outcome is expected because traders have very low convenience yield and thus a very weak incentive to store when facing the prospects of much lower prices in the second year. The large drop in storage implies that consumption in Q2, Q3 and Q4 is much higher than in the base case, and this higher consumption is associated with substantially lower prices in these three quarters.

The reduced carry over from Q4 to Q5 also significantly alters the saw-toothed pattern. First, the price increase from Q2 to Q3 is much smaller than the base case because less is being carried over from Q4 to Q5 (the reduced storage reduces the marginal storage cost). Second, the price drop from the peak price in Q3 to the Q5 post-harvest price is much larger than in the base case. Recall that a similar result was observed in the estimation of the dummy variable model for hay prices. Specifically, in those years when hay stocks were low the price drop when transitioning from the old crop year to the new crop year was much larger. Finally, the price increase between Q5 and Q7 is larger as compared to the base case because of the higher volume of corn which is in storage during this period of time.

# Empirical Examination of Futures Price Spreads

Earlier in this module we used the non-stochastic pricing model to examined how changes in and affected price levels, price spreads and the overall tilt of the pricing schedule. In this revised model we are interpreting the pricing schedule as the forward curve populated with futures prices. We are interested in the same question. Is there a causal link between higher (lower) forecasted demand for Q8 carry out stocks and/or lower (higher) forecasted Q5 production and the level of futures prices and price spreads.

Recall from the previous analysis that we observed that a revision to the Q5 production forecast had a much smaller impact on the Q3 to Q7 price spread (and also the tilt of the pricing schedule) as compared to the impact of an equal-sized revision to the Q8 stock demand forecast. The reason for this asymmetric response was due to the fact that higher Q8 carry out demand raises the carrying charges by an approximate equal amount in both years. In contrast, lower Q5 production results a larger decrease in the year 2 carrying charge versus the year 1 carrying charge. Of interest is whether similar results can be observed with an econometric analysis of the 5000 simulated price series.

To begin, let’s calculate the mean and standard deviation of the two forecast variables, the Q3 and Q7 futures prices as well as the Q3 to Q7 price spread. To accomplish this task we need to merge the second columns of the following data frames (the second column corresponds to Q2): *price3*, *price7*, *harvest* and *demand*. After the merge the two forecast variables should be renamed.

futures\_2\_3 <- price3 %>% select(P\_2\_3)  
futures\_2\_7 <- price7 %>% select(P\_2\_7)  
futures\_2\_3\_7 <- cbind(futures\_2\_3,futures\_2\_7,harvest[,2],demand[,2])  
colnames(futures\_2\_3\_7)[3:4] <- c("D\_frcst","H5\_frcst")

We can now create the Q3 to Q7 spread variable for the Q2 forward curve.

futures\_2\_3\_7 <- futures\_2\_3\_7 %>%   
 mutate(Sprd\_3\_7 = P\_2\_7 - P\_2\_3)  
  
head(round(futures\_2\_3\_7,4))

## P\_2\_3 P\_2\_7 D\_frcst H5\_frcst Sprd\_3\_7  
## 1 3.5890 3.5869 -0.1437 14.4810 -0.0021  
## 2 3.8987 3.9470 0.7378 14.6945 0.0484  
## 3 3.6021 3.6046 -0.0636 14.5307 0.0024  
## 4 4.7930 4.8446 0.7943 12.9442 0.0516  
## 5 3.7688 3.7370 -0.6622 13.6253 -0.0319  
## 6 4.5087 4.5286 0.2411 12.9918 0.0199

With this data in place we can calculate the mean and standard deviation of the price spread across the simulated values as well as the maximum and minimum observed spreads in the sample.

mean\_2\_3\_7 <- futures\_2\_3\_7 %>% summarise\_if(is.numeric, mean)  
sd\_2\_3\_7 <- futures\_2\_3\_7 %>% summarise\_if(is.numeric, sd)  
print(mean\_2\_3\_7, digits = 4)

## P\_2\_3 P\_2\_7 D\_frcst H5\_frcst Sprd\_3\_7  
## 1 3.717 3.724 0.007922 14.37 0.006537

print(sd\_2\_3\_7, digits = 4)

## P\_2\_3 P\_2\_7 D\_frcst H5\_frcst Sprd\_3\_7  
## 1 0.6469 0.6632 0.6156 1.171 0.03526

max(futures\_2\_3\_7$Sprd\_3\_7)

## [1] 0.1286598

min(futures\_2\_3\_7$Sprd\_3\_7)

## [1] -0.126366

These summary statistics confirm that the mean values of the Q3 and Q7 variables in the sequence of 5000 simulated futures prices are approximately equal to the values of these two prices in the base case, as reflected by the Q1 forward curve. As well, the mean values of the two forecast variables are approximately equal to the base case values of these two variables. More important is the outcome that the mean value of the Q3 to Q7 price spread is approximately equal to zero. This outcome is expected given the discussion above and the spread equation ,

The standard deviation, maximum and minimum values of the 5000 simulated price spreads is 0.0353, 0.128 and -0.126, respectively. The variation in the spread values are small relative to what is observed in the real world. The main reason for this outcome is that prices in this model are stationary in the long run whereas this is not the case with real world commodity prices.

# Regression Results

We can test whether the Q3 to Q7 price spread responds different to Q8 carry out demand forecasts versus Q5 production forecasts using regression analysis. Let’s begin by regressing the Q3 to Q7 price spread on the outcomes of the two forecast variables (keep in mind that all variables are measured in Q2).

reg\_sprd <- lm(Sprd\_3\_7 ~ H5\_frcst + D\_frcst, data = futures\_2\_3\_7)  
summary(reg\_sprd)

##   
## Call:  
## lm(formula = Sprd\_3\_7 ~ H5\_frcst + D\_frcst, data = futures\_2\_3\_7)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.016e-09 -2.939e-10 -2.690e-12 2.968e-10 1.013e-09   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.083e-03 7.126e-11 8.537e+07 <2e-16 \*\*\*  
## H5\_frcst -3.019e-12 4.944e-12 -6.110e-01 0.541   
## D\_frcst 5.729e-02 9.406e-12 6.090e+09 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.094e-10 on 4997 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.855e+19 on 2 and 4997 DF, p-value: < 2.2e-16

The regression results reveal a highly statistically significant and positive relationship between the size of the Q8 carry out stocks forecast, *D\_frcst*, and the Q3 to Q7 price spread. Despite being highly statistically significant, the size of *D\_frcst*, 0.0573, is rather small. Indeed, a one standard deviation change in *D\_frcst*, which we know from above is equal to 0.616, results in a change in the Q3 to Q7 price spread, on average, equal to .

The regression also reveals that the estimated coefficient for the production forecast is very small and statistically insignificant. We were expecting a smaller value for this coefficient given our theoretical predictions, but we were not expecting that the two effects which determine the Q3 to Q7 price spread are almost fully offsetting. While the intuition for this outcome is not obvious it is clear that the carrying charge is at the heart of the explanation.

It would be good to econometrically distinguish between the two offsetting effects which are responsible for the lack of a statistically significant relationship between the Q5 production forecast and the Q3 to Q7 price spread. One possibility is to include the Q4 to Q5 carry over variable (i.e., the “S4” variable) as a right side explanatory variable. The inclusion of this variable will control for the direct effect of changes in the Q5 production forecast, which means that the estimated coefficient on the forecast variable can be interpreted as a measure of the indirect impact (i.e., the change attributable to the carrying charge).

The main problem with this proposal is that it is not appropriate to use an endogenous variable as an explanatory variable. An alternative is to replace the variable in the regression with an instrument. The theory of instrumental variables tells us that an instrument must satisfy the relevance and exclusion restrictions. A relevant instrument is one which has a causal impact on the endogenous variable which is being replaced. The exclusion restriction ensures that the instrument affects the dependent variable only through changes in the endogenous variable which is being replaced. That is, there must not be a direct causal link between the instrument and the dependent variable.

Let’s check to see if the Q8 carry out stock demand variable, *D\_frst*, can be used as an instrument. First, we know that *D\_frst* has a strong causal impact on the the variable, and so we can be assured that the relevance criteria is satisfied. Second, we know that the exclusion restriction is satisfied because the impact on the Q3 to Q7 price spread from a change in *D\_frst* can be fully controlled for (i.e., made to vanish) through an appropriately-sized adjustment in the Q4 carry over variable, .

Given that *D\_frst* is an appropriate instrument, we can implement the two stage least squares procedure. In the first stage we regress the endogenous variable of interest, , on the instrument, *D\_frst*, and on the other exogenous variable, *H5\_frcst*. In the second stage we use the in-sample forecasts of the endogenous variable (i.e., the fitted values) as an explanatory variable in the main regression, which has the Q3 to Q7 price spread as the dependent variable.

To implement this procedure we first need to create the stock carry over column of data.

stocks\_4 <- stock\_begin - (4\*a/b - (1/b)\*(priceSpot$spot1 + priceSpot$spot2 + priceSpot$spot3 + priceSpot$spot4))

Now the first stage regression equation can be estimated as follows.

reg\_sprd2 <- lm(stocks\_4 ~ D\_frcst + H5\_frcst, data = futures\_2\_3\_7)  
summary(reg\_sprd2)

##   
## Call:  
## lm(formula = stocks\_4 ~ D\_frcst + H5\_frcst, data = futures\_2\_3\_7)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.97735 -0.15503 0.00565 0.16698 0.95448   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.104123 0.042452 190.90 <2e-16 \*\*\*  
## D\_frcst 0.255499 0.005604 45.59 <2e-16 \*\*\*  
## H5\_frcst -0.423146 0.002945 -143.67 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2439 on 4997 degrees of freedom  
## Multiple R-squared: 0.819, Adjusted R-squared: 0.8189   
## F-statistic: 1.131e+04 on 2 and 4997 DF, p-value: < 2.2e-16

stocks\_fit2 <- reg\_sprd2$fitted.values

And now the second stage.

reg\_sprd3 <- lm(Sprd\_3\_7 ~ H5\_frcst + stocks\_fit2 , data = futures\_2\_3\_7)  
summary(reg\_sprd3)

##   
## Call:  
## lm(formula = Sprd\_3\_7 ~ H5\_frcst + stocks\_fit2, data = futures\_2\_3\_7)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.016e-09 -2.939e-10 -2.690e-12 2.968e-10 1.013e-09   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.811e+00 3.062e-10 -5.914e+09 <2e-16 \*\*\*  
## H5\_frcst 9.488e-02 1.630e-11 5.819e+09 <2e-16 \*\*\*  
## stocks\_fit2 2.242e-01 3.682e-11 6.090e+09 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.094e-10 on 4997 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.855e+19 on 2 and 4997 DF, p-value: < 2.2e-16

The estimated coefficients for the first and second stage regression both have the correct sign, as suggested by theory. In the first stage, higher demand for Q8 carry out stocks increases the amount carried over from Q4 to Q5. In contrast, higher Q5 production decreases the Q4 carry over.

The signs of the second stage coefficients are more challenging to interpret. The estimated coefficient for the fitted Q4 stocks variable has a positive sign because a smaller harvest results in more Q4 stocks and a higher price spread. The estimated coefficient for the Q5 production forecast has a positive sign because a smaller year 2 harvest forecast reduces the year 2 carrying costs, which in turn reduces the price spread.

# Basis

In this last section we begin to integrate the spot and futures price of a commodity. We will see that regional spot prices are constructed by adding a basis to the Chicago futures price. The basis reflects the transportation costs, carrying costs (i.e., storage and convenience yield) and local temporary supply and demand imbalances. In this module we will focus only the carrying cost component of the basis.

## Basis Overview

In commodity markets the basis is defined as the spot price minus the price of a designated futures contract. With this definition, the basis in Chicago is often negative because there is no adjustment for location and futures contracts will trade above the Chicago spot price when stocks are plentiful. The fact that the basis is generally negative means that it is very important to be clear when talking about a basis becoming “bigger” or “smaller”. To avoid confusion it is a good idea to refer to the “absolute basis” when referring to changes in the size of the basis.

One can view basis as the calculated difference between an observed futures price and an observed spot price. However, basis can also be viewed as a price which is set by a grain merchant on a daily basis. The spot price (also called the cash bid price) is then calculated by adding the basis to the observed futures price. In either case the basis must be associated with an underlying futures contract.

The basis may be reported with the underlying futures contract made explicit or it may be reported with respect to a nearby futures on a rolling basis. Grain merchants generally use the first method (more details below) and the USDA, which reports the daily basis for informational purposes, generally use the second method.

Consider the USDA’s information reporting task. They begin with the cash bids which are reported by hundreds of regional country elevators, in-land terminal elevators, in-land processing elevators and export terminal elevators. The USDA first categorizes the cash bids by type of elevator and local region. It then converts all explicit reference to an underlying commodity futures to a nearby commodity futures. This allows the USDA to report the basis and the associated cash bid on a nearby- contract rolling basis, and thus without the need to explicitly identify the underlying futures contract. A good example of USDA reporting is the daily cash bids for the state of [Kansas](https://www.ams.usda.gov/mnreports/ams_2886.pdf).

In general, commodity basis data is very important and widely used in the grain trade because this data reflects the link between the futures market and local cash markets. However, the study of basis is complicated for the following three reasons:

* The basis reflects the location of the grain merchant because futures contracts are priced with a Chicago delivery point and the cash bid/spot price assumes local delivery of the commodity.
* The basis reflects underlying carrying costs because futures contracts assume delivery when the contract expires and the cash bid/spot price assumes either immediate delivery or delivery in a time slot which precedes the expiry of the futures contract.
* The basis reflects local demand and supply imbalances and grain transportation constraints (this is examined in greater detail in Module 2E).

The focus of Module 2E is the first of these three bullets. In this course little will be said about the third bullet because a theoretical framework is not in place to formally examine temporary movements away from the spatial and intertemporal LOP. Rather, the temporary deviations in the basis due to local demand and supply shocks is what comprises the error term when examining the basis in an econometric framework. For students who take FRE 517 in term 2, real-world basis is central to the study of commodity markets. The second bullet point, which emphasizes the intertemporal component of the basis, is the focus of this current module.

## Deferred Delivery Contracts

Before starting the formal analysis of the intertemporal component of the basis it is useful to examine the basis associated with deferred delivery contracts. We will use the data provided by [Mid Iowa Cooperative](https://www.midiowacoop.com/grains/cash-bids/) to illustrate the concept. The current-day prices can be downloaded in csv format. The chunk of code to follow pulls in this data so we can take a closer look. The top six rows correspond to cash bids for corn by Mid Iowa Cooperative in the town of Beaman.

cashbids <- read.csv(here("Data", "cashbids\_iowa.csv"), header=TRUE, sep=",", stringsAsFactors = FALSE)   
cashbids <- subset(cashbids, select = -c(Location, Notes))  
head(cashbids, 6)

## Commodity Delivery.Start Delivery.End Symbol Futures Basis Cash.Price  
## 1 Corn 11/1/2021 11/30/2021 ZCZ21 553-0 -19 $5.34   
## 2 Corn 1/1/2022 1/31/2022 ZCH22 562-2 -25 $5.37   
## 3 Corn 3/1/2022 3/31/2022 ZCH22 562-2 -30 $5.32   
## 4 Corn 5/1/2022 5/31/2022 ZCK22 567-2 -27 $5.40   
## 5 Corn 7/1/2022 7/31/2022 ZCN22 568-6 -30 $5.39   
## 6 Corn 10/1/2022 10/31/2022 ZCZ22 540-4 -40 $5.00

The displayed data shows that as of Friday, November 5th a farmer has a number of deferred delivery options for corn sales, each one with a unique basis, underlying futures price and cash bid price. The top row shows that if the farmer wants to deliver in the month of November (i.e., immediate), she would sign a contract which guaranteed payment upon delivery equal to $5.34/bu. This cash bid price is derived by adding a -$0.19/bu basis to the November 5th price of the December 2021 futures contract (symbol ZCZ21). The -$\0.19 basis partially reflects the location of the town of Beaman (i.e., transportation costs to move the corn to market) and partially reflects local supply and demand conditions.

If instead the farmer wanted to delivery in March of 2022, she would sign a contract with a deferred delivery price equal to $5.32/bu. This cash price is derived by adding the -$0.30/bu basis to the November 5th price of the March 2022 futures contract (symbol ZCH22) equal to $5.62/bu. Notice that as compared to immediate delivery the cash price to be paid to the farmer with March delivery is slightly lower. We see that the March 2022 futures price is trading about $0.09/bu higher than the December 2021 futures price. This means that in the larger market the spot price is expected to rise by $0.09/bu over the next three months, presumably to cover the carrying cost over this period. However, largely offsetting the higher futures is a larger absolute basis ($0.30/bu versus $0.19/bu). Both delivery windows are just before the expiry of the underlying futures contract and so it must be that the higher absolute basis is reflecting a local supply and demand imbalance. In particular, Mid Iowa Cooperative is paying a higher price for immediate delivery rather than delivery next March.

The last row of the displayed data shows that if the farmer wishes to deliver her corn in October of 2022 (presumably from the new crop) then she will be paid $5.00/bu, which is $0.34/bu lower than the price she would receive with immediate delivery. This $0.34/bu reduction can be attributed to both a difference in the pair futures contracts with underlie the two delivery periods (December 2021 for immediate delivery versus December 2022 for October of 2022 delivery) and a larger absolute basis for next October delivery, which is assigned by the Mid Iowa Cooperative.

In the analysis below we will focus on the cash bid/spot price with the transportation component eliminated by assuming Chicago is the spot market location, and the temporary supply and demand disequilibrium eliminated by assuming the basis satisfies the intertemporal LOP restriction.

## Chicago Basis with Non Stochastic Prices

We will once again leave the real world and examine prices which come from the eight-quarter model. Let’s start with the case where prices are non-stochastic. To keep things simple, assume that all basis calculations are made with respect to the Q7 futures contract. Keeping in mind that the spot price is the price of an expiring futures contract, we can define basis in quarter as the quarter price of the quarter futures contract which is just expiring minus the quarter price of the Q7 contract.

To better visualize the basis data it is a good idea to re-display the pricing matrix that we introduced earlier (i.e., the one which comes from the third row of the 5000 sets of simulated prices which we previously imported).

print(round(prices, 4))

## pQ1 pQ2 pQ3 pQ4 pQ5 pQ6 pQ7 pQ8  
## [1,] 3.4905 3.6532 3.7083 3.6563 3.4966 3.6593 3.7144 3.6624  
## [2,] NA 3.5479 3.6021 3.5483 3.3859 3.5495 3.6046 3.5516  
## [3,] NA NA 3.6403 3.5868 3.4250 3.5899 3.6466 3.5956  
## [4,] NA NA NA 3.8891 3.7299 3.9010 3.9666 3.9273  
## [5,] NA NA NA NA 3.7641 3.9321 3.9950 3.9531  
## [6,] NA NA NA NA NA 3.6288 3.6890 3.6419  
## [7,] NA NA NA NA NA NA 3.4966 3.4478  
## [8,] NA NA NA NA NA NA NA 3.3975

The top row of prices in this matrix are the base case (non-stochastic) prices. Assume these prices repeat each quarter until the Q7 futures contract expires in Q7. This means that in the top row, the price in the pQ1 column minus the price in the pQ7 column is the basis for a Q7 contract when in Q1. As well, the price in the pQ2 column minus the price in the pQ7 column is the basis for a Q7 contract when in Q2, and so forth. If we put these calculated basis values in a vector we will have the time path of the basis for a Q7 contract, ranging from Q1 through to Q7.

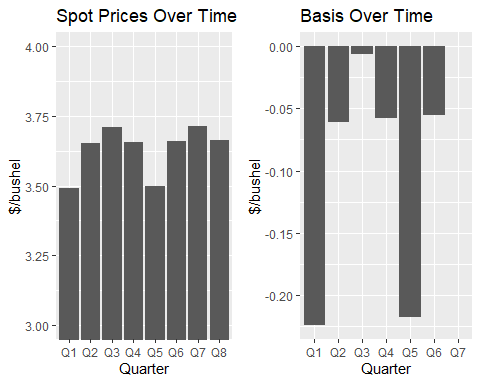
The code required to generate this vector of basis results is as follows:

basis7\_ns <- c()   
basis7\_ns[1] <- prices[1,c("pQ1")]-prices[1,c("pQ7")]  
basis7\_ns[2] <- prices[1,c("pQ2")]-prices[1,c("pQ7")]  
basis7\_ns[3] <- prices[1,c("pQ3")]-prices[1,c("pQ7")]  
basis7\_ns[4] <- prices[1,c("pQ4")]-prices[1,c("pQ7")]  
basis7\_ns[5] <- prices[1,c("pQ5")]-prices[1,c("pQ7")]  
basis7\_ns[6] <- prices[1,c("pQ6")]-prices[1,c("pQ7")]  
basis7\_ns[7] <- prices[1,c("pQ7")]-prices[1,c("pQ7")]  
print(round(basis7\_ns, 5))

## [1] -0.22392 -0.06119 -0.00608 -0.05813 -0.21779 -0.05511 0.00000

To better visualize the basis results let’s plot them beside a plot of the non-stochastic price series.

forward <- as.data.frame(prices[1,])  
colnames(forward)<- "ForwardQ1"  
labels <- c("Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8")  
forward <- cbind(labels,forward)  
  
basis7\_ns <- as.data.frame(basis7\_ns)  
colnames(basis7\_ns)<- "BasisQ1"  
labels7 <- c("Q1","Q2","Q3","Q4","Q5","Q6","Q7")  
basis <- cbind(labels7,basis7\_ns)  
  
forward\_plot <-ggplot(data=forward, aes(x=labels, y=ForwardQ1)) +  
 geom\_bar(stat="identity") + ggtitle("Spot Prices Over Time") +  
 labs(x= "Quarter") + labs(y= "$/bushel") +  
 coord\_cartesian(ylim=c(3, 4))  
  
basis\_plot <-ggplot(data=basis, aes(x=labels7, y=BasisQ1)) +  
 geom\_bar(stat="identity") + ggtitle("Basis Over Time") +  
 labs(x= "Quarter") + labs(y= "$/bushel")  
  
grid.arrange(forward\_plot, basis\_plot, ncol = 2)



It should be clear from these two plots that the basis has the negative of the mirror image of the price series. This means that the saw-toothed pricing pattern is reflected by the basis as well as the prices themselves. During the Q1 and Q5 harvest period, the basis takes on large negative values because the spot price is well below the price of a Q7 futures contact. A basis with a relatively wide negative value is said to be a weak basis. As the spot price increases when moving from Q1 to Q3, and from Q5 to Q7, the basis becomes less negative. The basis comes close to vanishing in Q3 and of course does vanish in Q7. A basis which takes on a small negative value is said to be a strong basis, and a basis which is becoming less negative over time is said to be strengthening basis. It should obvious that the basis is weakening when the spot price is trending lower while transitioning from Q4 to Q5.

## Basis with Stochastic Prices

When the basis is calculated with stochastic prices the time path of the basis will have a random component. The basis is equivalent to a special case of a price spread and so we should expect fluctuations in the basis similar to the fluctuation in the Q3 to Q7 spread that we examined above. The procedure for calculating the basis with stochastic prices is the same as with non-stochastic prices except now all rows of the previous pricing matrix should be used rather than just the top row. The code which generates the random time path for the seven basis values is as follows.

basis7 <- c()   
basis7[1] <- prices[1,c("pQ1")]-prices[1,c("pQ7")]  
basis7[2] <- prices[2,c("pQ2")]-prices[2,c("pQ7")]  
basis7[3] <- prices[3,c("pQ3")]-prices[3,c("pQ7")]  
basis7[4] <- prices[4,c("pQ4")]-prices[4,c("pQ7")]  
basis7[5] <- prices[5,c("pQ5")]-prices[5,c("pQ7")]  
basis7[6] <- prices[6,c("pQ6")]-prices[6,c("pQ7")]  
basis7[7] <- prices[7,c("pQ7")]-prices[7,c("pQ7")]  
print(round(basis7, 4))

## [1] -0.2239 -0.0566 -0.0063 -0.0775 -0.2309 -0.0602 0.0000

It is useful to display both sets of prices so that we can visually identify similarities and differences with the two approaches.

print(t(basis7\_ns), digits = 4)

## [,1] [,2] [,3] [,4] [,5] [,6] [,7]  
## BasisQ1 -0.2239 -0.06119 -0.006083 -0.05813 -0.2178 -0.05511 0

print(round(basis7, 4))

## [1] -0.2239 -0.0566 -0.0063 -0.0775 -0.2309 -0.0602 0.0000

A comparison of the stochastic and non-stochastic basis values reveals relatively small differences. This suggests the quarterly spot prices and the price of a Q7 futures contract are strongly positively correlated. To get a more accurate estimate of basis uncertainty we should use the full 5000 sets of simulated prices.

In Module 1F on hedging we will create a mini case study where a merchant places a short hedge with a Q7 contract at the beginning of Q4 and offsets that hedge at the beginning of Q6. It will be shown that the profits for the short hedger are equal to the change in the basis between when the hedge was place (i.e., Q4) and when it was lifted (i.e., Q6). For this reason it is of interest to use the 5000 sets of prices to calculate the mean and standard deviation of the Q7 basis in Q4 and then again in Q6, which is when the hedge is lifted. With this data in hand it is straight forward to calculate summary statistics for the change in the basis over the Q4 to Q6 time period.

## Q7 basis in Q4 and Q6

The following code creates the two basis data series and the difference between the two series.

b7Q4 <- price4$P\_4\_4 - price7$P\_4\_7  
b7Q6 <- price6$P\_6\_6 - price7$P\_6\_7  
b7\_ch <- b7Q6 - b7Q4

The data series can be bound together and displayed using the following code.

b7 <- cbind(b7Q4,b7Q6,b7\_ch)  
b7 <- as.data.frame(b7)  
head(round(b7,4))

## b7Q4 b7Q6 b7\_ch  
## 1 -0.0642 -0.0724 -0.0083  
## 2 -0.0898 -0.0570 0.0328  
## 3 -0.0775 -0.0602 0.0173  
## 4 -0.0764 -0.0633 0.0131  
## 5 -0.0193 -0.0359 -0.0166  
## 6 -0.0769 -0.0622 0.0147

Now we can calculate and examine the summary statistics.

mean\_b7 <- b7 %>% summarise\_if(is.numeric, mean)  
sd\_b7 <- b7 %>% summarise\_if(is.numeric, sd)  
mean\_b7

## b7Q4 b7Q6 b7\_ch  
## 1 -0.05740165 -0.0549387 0.002462954

sd\_b7

## b7Q4 b7Q6 b7\_ch  
## 1 0.03572191 0.02052691 0.02014423

The first two entries in the top set of summary statistics shows the mean level of the basis for the Q7 contract calculated in Q4 and Q6. Both values are negative because on average both the Q4 and Q7 futures contracts upon expiry trade below the Q7 futures contracts in any given time period. The slightly more negative value for the Q4 basis indicates that the Q4 futures trades slightly below the Q6 futures, on average.

The third entry in the top set of summary statistics is the average difference in the basis for a Q7 contract when in Q6 versus Q4. When we model hedging outcomes in Module 2F we will assume that the hedge is placed in Q4 and lifted in Q6. The expected profit from the short hedge is equal to the average change in the basis when moving from Q4 to Q6. The data here reveal that the expected profits for the short hedge is equal to $0.0025/bu. This very small value suggests that the hedge will mostly achieve reduced risk rather than generating higher profits on average.

The second of the two sets of outcomes above show the standard deviation of the basis for Q7 contract in Q4 and Q6. The third entry shows the standard deviation of the change in the basis for the Q7 contract between Q4 and Q6. This last value is important because it is a measure of the variation in profits for the short hedger. In this particular case the variation in profit is not very large since the standard deviation of the change in the basis is equal to 0.020.

# Conclusions

In this module we covered a lot of ground. Most importantly we used the eight-quarter pricing model to simulate commodity futures prices with realistic properties. When analyzing the simulated prices we saw that for each time period, the forward curve provides a snapshot view of the futures prices, which are the spot prices we expect to observe as we move through time. The saw-toothed pattern which is inherent in every forward curve reveals the seasonal pattern of spot prices over the eight-quarter time period.

We chose to use the Q3 to Q7 spread in futures prices as a measure of the overall tilt of the forward curve. A surge in demand for stocks leaving Q8 causes the price spread to increase, and thus the forward curve to tilt upward. Weak demand for stocks leaving Q8 would have the opposite effect. In the summer of 2021 the forward curve for corn tilted sharply. We believe this happened because of strong summer of 2021 demand for corn relative to future time periods. The eight quarter model is not designed to model a surge in current demand. Nevertheless, by interpreting a positive surge in current demand as having the same outcome as a negative surge in long term demand we obtain the intuition as to why the forward curve for corn tilted sharply down in the summer of 2021.

We also learned that a shock to forecasted production which leaves more or less stock available for consumption in year 2 has a very different impact on the Q3 to Q7 price spread as compared to an equal-sized change in demand for stocks carried out of Q8. Indeed, regression with the simulated data showed that the Q3 to Q7 price spread changes very little, even with comparatively large changes in forecasted harvest levels. The reason is that changes in harvest have a direct effect on the year 2 carrying cost (and an indirect effect on the price spread), and this effect is not present in the case of a demand surge.

This module concluded with an analysis of the basis. The basis is similar to a spread in the futures price when the spot market is located in Chicago, However, when analyzing the basis we are generally interested in how it changes over time. This is different than our interest in the price spread, which is mainly associated with the one-time tilt in the forward curve. We saw that variation in the futures prices over time translate into uncertainty in how the basis will change over time. As will be shown in Module 2F, this uncertainty translates into pricing risk for short and long hedgers.