Begin by load the packages

pacman::p\_load(here, dplyr, ggplot2)

# 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 8 quarter (two year) pricing model was constructed and calibrated to the U.S. corn market. A key feature of the model is convenience yield. Indeed, a combination of storage costs and convenience yield results in realistic simulated price paths. In this module random sequences of spot and futures prices for corn are generated and analyzed in order to shed light on key features of spot and futures prices.

An important feature of commodity markets is the relative riskiness of spot and futures prices. The analysis of risk is beyond he scope of this module 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 point of explicitly analyzing expected returns.

# Background

It is important to have a good understanding of real-world commodity prices before attempting to model them. Commodity futures have two important dynamic components. The “snap shot” component consists of the set of prices of the contracts with 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. A graph of the set of prices of all corn contracts at particular point in time is commonly referred to as a forward curve. It turns out that the slope of the forward curve is an important determinant of expected profits for both hedgers and institutional investors.

This background section is not yet complete. For now you need to refer to the material in the class slides.

# 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 how the spot prices changed 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”.

**Expecations 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 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 that an investor who holds a continually rolling over 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.

# Eight Quarter Model

The eight quarter simulation model was build 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 brief review of 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 the beginning stock balance plus harvest minus consumption is equal to the stocks which are carried forward through time.

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: initial stocks inclusive of Q1 harvest, , the size of the harvest in Q5, , the level of pipeline stocks which must be carried out of Q8, , and , which is carry out stocks in excess of (positive or negative) due to temporary surges of slumps in longer term demand for the commodity. For the base case, , = 14.377, and . All quantity variables are measured in billions of bushels and price is measure 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. This framework lends itself to “what if” analysis. More importantly, in the analysis below two of the exogenous variables will be treated as random variables, and each new quarterly outcome of these two variables will result in a new set of equilibrium prices, consumption and storage. This is the way through which random prices are generated for detailed analsis.

## Forecasting

Our goal is use the 24 equation model to generate the types of random prices that we observe in the real world. We do this by assuming that is random but its Q5 value can be forecasted at the beginning of Q1 through Q5 (the Q5 forecast is identical to the Q5 actual outcome). Similarly, we assume that is random but its Q8 value can be forecasted at the beginning of Q1 through Q8. We make the strong assumption that each quarterly forecast of and is perfectly accurate. This means that we can replace the actual values of and in the model with their respective forecasts and then solve the model for equilibrium prices, consumption and storage levels. With the arrival of new forecasts at the beginning of the next quarter, the model is resolved with the revised forecasts replacing the actual values of and to generate a new set of prices, consumption and storage levels. The random arrival of revised forecasts for and results in a sequence of random prices over time.

Let and denote the forecasts for and at the beginning of quarter . For we have and for we have . An important assumption is that forecasts for and follow a random walk. This means that a trader’s best estimate of future forecasts is the current forecast. For example, in Q1 when the trader observes and she assumes that all future forecasts of and will take on these same values. In Q2 when the forecasts are revised to and traders will now assume that all all future forecasts of and will take on these same revised values.

The random walk assumption is both realistic and important because it means that with the arrival of a new forecast at the beginning of a new quarter, the set of equilibrium prices which comes from the solution to the model can be interpreted as the current spot price and the set of expected future spot prices. If we combine this outcome with our expectations hypothesis it follows that with a given forecast for and at the beginning of a particular quarter, the values of the equilibrium prices beyond the current quarter can be interpreted as futures prices.

## Simulation

The random forecasts and model solution 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 a random set of forecasts for and , one spot price and seven futures prices for Q1, one spot price and six futures prices for Q2, etc. Vercammen (2021) shows that the summary statistics of the 5000 simulated prices are reasonably reflective 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 nine .csv files (one file for each of the eight quarters and a ninth file which summarizes the spot prices). These files can now be pulled into this current session for 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 500, this function returns the set of futures and spot 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 chunck to follow loads *price-reformat.R* and then imports the forecast data.

source(here("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

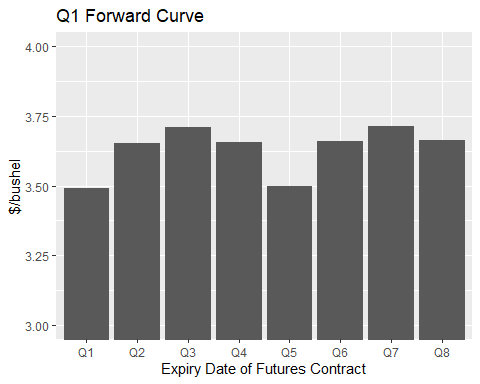
It is important to read this pricing data correctly. Each row corresponds to a time period and a given set of forecast information. The diagonal elements are the spot prices and the elements to the right of the diagonal are expected values of the future spot prices. For example, in the fourth row which 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 Q5 is 3.730, 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. 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.

# 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 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



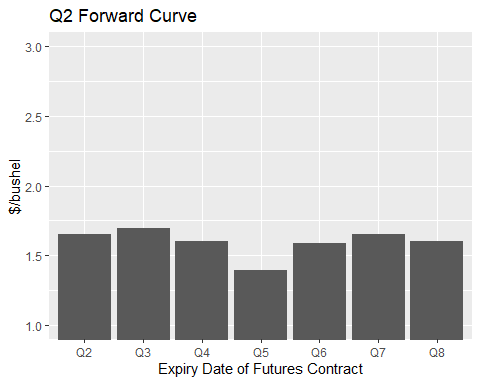
The above column chart shows that prices in the eight quarter have the saw-toothed pricing pattern with the convenience yield modification. Recall from Module 2C that convenience yield causes the prices to drop more gradually versus an abrupt drop as the first crop year finishes. Moreover, convenience 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 likely be different than the forecast in Q1. A different forecast will likely 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.

Let’s choose another row in the matrix of 5000 simulated forecasts and prices. Row 8 is 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.

Let’s view the 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 storage costs are lower. 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.

## Futures Price Spreads - Theory

A price spread is defined as the futures price of a contract with a more distant expiry date minus the futures price of a contract with a less distance expiry date (e.g., the Q3 futures price minus the Q2 futures price). It is important to keep this in mind because a Q2 to Q3 price spread means the Q3 price minus the Q2 price and not vice versa. The size and the sign of the price spreads reflect the slope of the forward curve. For example, large positive values for the price spread imply a steep upward sloping forward curve, which is typically associates with large stocks and thus a high cost of storage.

Price spreads across marketing years are of particular importance because they reflect the relative abundance and scarcity of the commodity across neighboring crop years. In this eight quarter model it is useful to examine the spread in the prices of the Q3 and Q7 contracts when in Q2. Before examining the simulation results it is useful to consider the predictions based on theory alone.

Let’s begin with the case case set of Q1 forward prices, which are displayed in the first column chart above. Here it can be seen that the Q3 and Q7 prices have almost identical values. You might be thinking that the Q3 to Q7 price spread should equal the cost of carrying stocks from Q3 to Q7 because stocks are being carried over from Q4 to Q5. But it doesn’t work this way. We know and where and is the carrying cost from Q1 to Q3 and from Q5 to Q7, respectively. If the subtract the first equation from the second we get . In the base case stocks are equally available in both cases and so we should expect and . Thus, in the base case we should expect the Q3 to Q7 price spread to approximately equal to zero. The prices from the base case as shown in teh first column chart above confirm that this is indeed the case.

Let’s now consider forecasts which are different than the base case. We will keep things simple by assuming four forecast combinations.

* Well above average forecast for and high positive value for .
* Well above average forecast for and high negative value for .
* Well below average forecast for and high positive value for .
* Well below average forecast for and high negative value for .

In the first case where Q5 forecasted production and Q8 carry out demand are both high the net abundance of corn stocks in Q7 relative to Q3 is ambiguous. In this case the Q3 to Q7 price spread may be positive or negative, depending on whether the larger than expected Q5 harvest outstrips the larger than expected Q8 carry out demand or vice versa.

In the second case where Q5 forecasted production is high and Q8 carry out demand is weak the Q7 price would be much lower than the Q3 price if carry over stocks from Q4 to Q5 are not adjusted. Carry over stocks from Q4 to Q5 are reduced in response to this negative price spread and this reduction will make the spread less negative. However, stock carry over cannot become negative and so it may be the case that in the final equilibrium the Q7 futures will trade below the Q3 futures, and the equilibrium spread is negative.

In the third case where Q5 forecasted production is low and Q8 carry out demand is strong traders have a strong incentive to carry over additional stocks from Q4 to Q5. Using the price spread equation that we derived above we have . In this case and is likely to have a negative value but will be small in absolute value. Consequently, the Q3 to Q7 price spread will take on a positive value, which is necessarily less than the cost of storage between Q3 and Q7.

In the fourth and final case where Q5 forecasted production is low and Q8 carry out demand is weak the situation is once again ambiguous. If the reduced harvest outstrips the weak demand then stocks in the second year will be scarce relative to the first year. In this case additional stocks will be carried from Q4 to Q5 and the Q3 to Q7 price spread will be similar to the previous case (i.e., positive and approximately equal to the Q1 and Q5 price difference). If instead the weak demand outstrips the reduced harvest then second year stocks will be relatively abundant. In this case the Q3 price may rise above the Q7 price although not by as much as would be expected in the second case above.

## Futures Price Spreads - Simulations

Let’s begin the analysis of the simulated price spreads by re-examining the prices from row 8 of the 5000 pricing outcomes. To following code displays the set of eight forward curves for this particular case.

print(prices2, digits = 4)

## pQ1 pQ2 pQ3 pQ4 pQ5 pQ6 pQ7 pQ8  
## [1,] 3.49 3.653 3.708 3.656 3.497 3.659 3.7144 3.6624  
## [2,] NA 1.654 1.692 1.606 1.394 1.588 1.6560 1.5996  
## [3,] NA NA 1.904 1.820 1.612 1.812 1.8882 1.8420  
## [4,] NA NA NA 1.345 1.133 1.339 1.4171 1.3685  
## [5,] NA NA NA NA 1.022 1.218 1.2868 1.2273  
## [6,] NA NA NA NA NA 0.903 0.9687 0.9038  
## [7,] NA NA NA NA NA NA 1.0983 1.0345  
## [8,] NA NA NA NA NA NA NA 1.5031

The discussion in the previous paragraph implies that the top row of the pricing matrix can be interpreted as the Q1 forward curve. That is, the top row is the snap shot of the set of futures prices as of the beginning of Q1. Equivalently, the top row is the set of expected spot prices as the market evolves over time. The forward curve prices in the top row equal the base case set of prices which were examined in Module 2C because the Q1 forecasts for and take on base case values by assumption.

If we progressively work down the rows of the pricing matrix, we see how the forward curve evolves over time within one 8 quarter simulation. It is important to note the intertemporal LOP holds for each forward curve in a particular quarter. For example, in the fifth row the forward curve consists of the spot price for Q5 and the price of futures contracts which expire in Q6, Q7 and Q8. If you were to check you would see that , , etc. You could also verify that the stock deaccumulation equation holds in the fifth row of the pricing matrix. If you substitute the Q5 through Q8 prices from the fifth row into the inverse demand equation you will have recovered the equilibrium levels of consumption for Q5 through Q8. Aggregate consumption over this period must equal the difference between stocks at the beginning of Q5 and the end of Q8.

Let’s begin by setting *row\_select = 3* and thus extracting the third row from the *harvest*, *demand* and *priceSpot* data frames.

row\_select <-3  
H5 <- harvest[row\_select,]  
D <- demand[row\_select,]  
spot <- priceSpot[row\_select,]

Each row of this matrix corresponds to a new quarter moving through time. In contrast, the data in the *harvest*, *demand* and *priceSpot* data frames are such that each column corresponds to a new quarter moving through time. To merge the data frames we need to extract the third row from *demand* data frame, transpose it and then bind it to the front of the *prices* data frame. This process can then repeated for the *harvest* data frame. It is also be repeated for the *priceSpot* data frame but this data should be bound to the end rather than to the beginning of *prices*. After creating the enlarged data frame the column headers of the new column entries should be renamed. The following code chunk achieves the various tasks.

prices <- cbind(as.data.frame(t(H5)),as.data.frame(t(D)),prices, as.data.frame(t(spot)))  
colnames(prices)[1:2] <- c("H5For","DFor")  
colnames(prices)[ncol(prices)] <- "Spot"  
print(prices,digits = 4)

## H5For DFor pQ1 pQ2 pQ3 pQ4 pQ5 pQ6 pQ7 pQ8 Spot  
## Q1 0.00000 14.38 3.49 3.653 3.708 3.656 3.497 3.659 3.714 3.662 3.490  
## Q2 -0.06362 14.53 NA 3.548 3.602 3.548 3.386 3.549 3.605 3.552 3.548  
## Q3 0.01422 14.54 NA NA 3.640 3.587 3.425 3.590 3.647 3.596 3.640  
## Q4 0.49689 14.57 NA NA NA 3.889 3.730 3.901 3.967 3.927 3.889  
## Q5 0.42073 14.46 NA NA NA NA 3.764 3.932 3.995 3.953 3.764  
## Q6 0.15771 NA NA NA NA NA NA 3.629 3.689 3.642 3.629  
## Q7 0.04728 NA NA NA NA NA NA NA 3.497 3.448 3.497  
## Q8 0.03289 NA NA NA NA NA NA NA NA 3.397 3.397

The spot prices for each of the eight quarters are along the diagonal of the pricing matrix. These prices can be gathered into one vector and displayed as follows.

It is important to recognize that these spot prices are changing over time for two reasons. First, even if the forecasts remained perfectly constant over time, the spot prices would change due to seasonal pattern of pricing (i.e., rise from Q1 to Q2 due to the positive storage cost, and then fall for Q3 to Q4 due to a rising convenience yield). In fact, the top row of the pricing matrix shows how the spot price is expected to evolve over time. Second, the uncertainty in the and forecasts implies that the spot price will evolve over time according to the evolving impacts of the revised forecasts. For example, the top row and fourth column of the pricing matrix indicates that the Q4 spot price was expected to equal $3.656/bu but due to large demand forecast shock in Q4 the spot price was much higher at $3.889/bu.

Each column in the pricing matrix corresponds to a particular futures contract. For example, the fifth column with header “pQ5” is the set of prices o