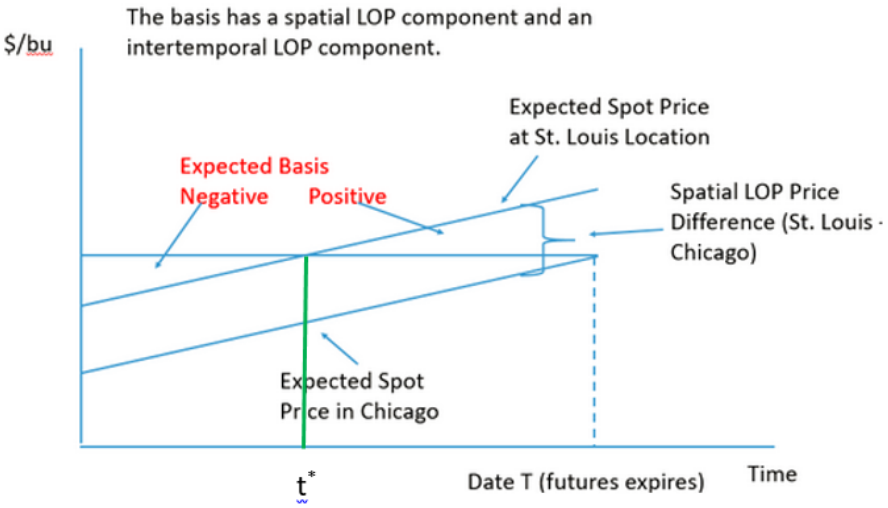
FRE 501
Module 2 Assignment, Q4
Due: 11:59 pm November 28, 2021

Introduction

This question will allow you to estimate the basis diagram found in your November 9th lecture slides, with Central Texas replacing St. Louis. The diagram (see below) has been modified by including t^*, which is the number of months it takes for the basis to equal zero (see the vertical green line).



The left side of the graph is assumed to equal January 1, and the right side (date T) is assumed to equal June 30th. The basis is calculated using the July corn futures. You should assume that the July futures expires on June 30th instead of the middle of July, which is a reasonable approximation. Similar to the diagram shown above, the slope of the spot price schedule is the same in Chicago as it is in Texas.

Part A of the analysis requires econometric estimation of the slope of the pair of expected spot pricing schedules in the above diagram. You will also use coefficient estimates of various dummy variables which make up your econometric model to estimate the initial Central Texas basis. To obtain an accurate measure of the slope you will pool together the basis data from three regions in Texas over a seven-year period. The basis in the graph is calculated with the July futures but you will use basis data for the July, September and December futures for the estimation procedure. The pooling of basis data over three regions, seven years and three futures contracts will allow you to obtain a highly accurate estimate of β_P, which is the slope of the expected spot price schedule. More details about the estimation procedure is provided below.

A secondary task in Part A is to construct a pivot table for the purpose of estimating the average futures price for each of the six months between January and June. Some of these results will be used in Part B.

In Part B of the analysis you will use some of the results from Part A to construct the graph which is shown at the top of this question sheet, and which is customized for Texas rather than St. Louis. Some hints for constructing this graph are provided below.

Part A: Econometric Procedure

The data comes from http://ageco-basis.tamu.edu/results. The three included regions: (1) Central Texas South (shortened to "Central"); (2) North of the Canadian (shortened to "North"); and (3) Eastern Panhandle (shortened to "East"). The data is weekly, running from the first week in January to the last week in June, from January of 2014 to the end of June, 2020. The data for 2021 is available but it is omitted because it should be viewed as a COVID outlier. The data consists of the regional spot prices and the futures prices for the July, September and December futures contracts (same calendar year as the spot price).

If you are interesting in knowing how the multiple .csv files were extracted from the above website, combined and then formatted into long data format, watch the Zoom recording created by Krisha. She shows how this task can be accomplished using Excel's power query function or R's looping routine.

There are generally 26 weeks for each of the seven years in the three regions. The central region is missing data for one week in 2018 and the eastern region is missing data for seven weeks in 2014. Thus, the total number of observations is 3*26*7 - 8 = 538 for each of the three futures contracts. With three futures contracts the total number of observations in the data set is 538*3 = 1614.

The data is in the file named "Q4_data.csv". The data cleaning, estimation and pivot table steps can easily be done in R, and you are encouraged to follow Krisha's template for carrying out these steps (you can find links to these templates in the Canvas Q4 submission page). Krisha shows how the graph at the top of this question sheet can be constructed in R. However, many of you will want to do the econometric estimation in R and the graphing part of this question (i.e., Part B) in Excel because constructing the above figure in R is somewhat complex.

For Excel Users

Your first step should be to import "Q4_data.csv" into an Excel workbook. Copy the data to a new sheet titled "Regression". Add a new column named "Year", which is the year of the observation. Add a second column named "Month", which is the month of the observation. Add a third column named "Trend" which takes on a value of 0 for January, a value of 1 for February, etc. until a value of 5 for June.

Create six dummies, one for each of 2015 through to 2020 (we are leaving "2014" out and treating it as the base year). Create two region dummies, one for North and one for East (we are leaving "Central" out and treating it as the base region). Create two contract dummies, one for September and one for December (we are leaving "July" out and treating it as the base contract). Finally, add a column named "Basis", which is the cash price minus the futures price. Your data should look as follows:

Date	Month	Region	Cash	Contract	Futures	Year	D-2015	D-2016	D-2017	D-2018	D-2019	D-2020	D-North	D-East	D-Sep	D-Dec	Trend	Basis
1/9/2014	1	Central	4.34	Dec	4.41	2014	0	0	0	0	0	0	0	0	0	1	0	-0.07
1/16/2014	1	Central	4.45	Dec	4.52	2014	0	0	0	0	0	0	0	0	0	1	0	-0.07
1/23/2014	1	Central	4.46	Dec	4.49	2014	0	0	0	0	0	0	0	0	0	1	0	-0.03
1/30/2014	1	Central	4.47	Dec	4.5	2014	0	0	0	0	0	0	0	0	0	1	0	-0.03

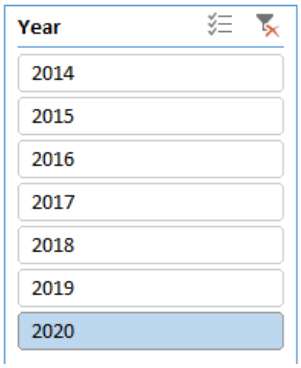
Estimate the model with Basis as the dependent variable. The explanatory variables are Trend and the ten dummies.

When estimating this model keep in mind that the inclusion of the various dummies allows for unique intercepts when regressing the basis on the trend variable (i.e., the intercept varies across the three regions, the three contracts and the seven years). This formulation of the econometric model minimizes

the potential for omitted variable bias. We could have estimated the model using a weekly time trend instead of a monthly time trend but doing that would complicate the graphical part of the analysis. The estimated coefficient on the Trend variable in your regression, which was estimated with the 1614 pooled observations, should be interpreted as the average monthly change in the basis from January to June, averaged over six years, three regions and three futures contracts.

Construct a pivot table with months along the rows, the three different contracts along the columns and the average weekly futures price across the seven years in the body of the table. Add a slicer and choose 2020 since this will be the reference year for this exercise. Your completed table should look like:

Average of Futures	Column Labels			
Row Labels	Jul	Sep	Dec	Grand Total
1	3.965	3.9585	3.9845	3.969333333
2	3.836875	3.824375	3.869375	3.843541667
3	3.644375	3.655625	3.715	3.671666667
4	3.295	3.342	3.435	3.357333333
5	3.201875	3.25	3.3475	3.266458333
6	3.2675	3.310625	3.393125	3.32375
Grand Total	**3.542403846**	**3.564038462**	**3.630673077**	**3.579038462**



Part B: Constructing the Graph

You will construct the graph with a perspective of a trader on January 1 of 2020 who is forming expectations about how the spot and futures price of corn will evolve from January 1 to when the July contract expires at the end of June of 2020 (the July contract actually expires in the middle of July but assuming an expiry at the end of June is a reasonable approximation). Assume that on January 1 of 2020 July corn futures are trading at $\bar{F}_{\text{July}}$ and the basis in Central Texas (calculated with the July futures) is equal to $\bar{B}_{\text{Jan}}$. The value you should use for $\bar{F}_{\text{July}}$ is the value in the top left corner of the pivot table since this corresponds to the average weekly value of the July contract in January of 2020. You will estimate $\bar{B}_{\text{Jan}}$ as the sum of the estimated intercept and the estimated coefficient on the 2020 dummy variable in your econometric model (details below).

The estimated slope of the expected spot price schedule, β_P, together with your estimates of $\bar{F}_{\text{July}}$ and $\bar{B}_{\text{Jan}}$ is sufficient to reconstruct the above diagram with the X axis ranging from January 1 to July 1. The trick is to recognize that the Central Texas spot price schedule intersects with vertical axis at point $\bar{P}_{Jan}^{Texas}$, and this value can be calculated using of $\bar{F}_{\text{July}}$ and $\bar{B}_{\text{Jan}}$. You can now use , β_P and the six month remaining life of the July contract to solve for $\bar{B}_{\text{June}}$, which is the expected basis for Central Texas at the end of June (i.e., at the point where the December futures contract expires).

The Chicago spot price schedule intersects with the vertical axis at point $\bar{P}_{Jan}^{Chicago}$, and this value can be calculated using $\bar{B}_{\text{June}}$ and $\bar{P}_{Jan}^{Texas}$ (use the diagram to make this connection).

To graph the data create a column called Month and add seven X axis labels “Jan 1”, “Feb 1”, …, “Jul 1”. Now create three columns named “Futures”, “Texas Spot” and “Chicago Spot”. It should be obvious that the value for $\bar{F}_{\text{July}}$ from the pivot table repeats throughout all seven entries of the “Futures” column. In the first row corresponding to Month = “Jan 1”, add the values for $\bar{P}_{Jan}^{Texas}$ and $\bar{P}_{Jan}^{Chicago}$ in the Texas

Spot and Chicago Spot columns. Add the remaining values in these two columns using your estimate of the slope of the pricing schedules as measured by β_P. You can now create a line graph of the Futures, Texas Spot and Chicago spot columns with Month as the horizontal axis labels. Be sure to calculate t^* and include this value on the graph (you can add t^* and the vertical green line using the line tool in Excel).

Label your graphs in a way which is similar to how the graph at the top of this question sheet is labeled.

Hints. To complete the steps described above you will require:.

1. A value for β_P, which is the slope of the expected spot price schedule. This value is the estimated coefficient on the Trend variable from Part A.
2. A value for $\bar{P}_{Jan}^{Texas}$, which is the sum of the estimated intercept and the estimated coefficient on the 2020 dummy variable. With this specification $\bar{P}_{Jan}^{Texas}$ can be interpreted as the expected corn basis in Central Texas in January of 2020.

Optional (highly recommended) Question (place answer in a textbox in Excel worksheet)

Identify the estimated slope of the December futures dummy and divide that value by 6. Compare your answer to the slope estimate, for β_P. Explain the strong connection between the two estimated outcomes.