Rigged Milk Prices

Introduction

Dataset and Variables

The data used in this project is the gotten from the bids recieved from diaries competing for the milk contracts between 1983 and 1991 of fluid milk products for the Northern Kentucky school districts. The dataset consists of the following characteristics(variables): Year, Market, Winner, District, WWBID, LFWBID, LFCBID. Variable Year represents the Year in which milk contact awarded. Variable Market represents the Northern Kentucky Market (Tri-COUNTY or SURROUND) Variable Winner represents the name of winning dairy. Variable WWBID represents the Winning bid price of whole white milk (dollars per half-pint). Variable LFWBID represents the Winning bid price of low-fat white milk (dollars per half-pint). Variable LFCBID represents the Winning bid price of low-fat chocolate milk (dollars per half-pint) School district number.

Types of Variables

The variable Year is a Ordinal Categorical variable. The variable Market is a nominal categorical variable. The variable Winner is also a nominal categrical variable. The variable District is also a nominal categrical variable. The variable WWBID, LFWBID, LFCBID is a continous numerical variable.

Descriptive Statistics

Year : This variable contains 9 categories, each year has a similar frequency between 39 and 51, except 1991 which is 18.

Market : This variable contains only 2 categories(Tri-county and Surround),Surround is a lot larger than tri-county, approximately 71% comapred to 29% respectively.

Winner : This variable contains 20 categories, Trauth and Meyer have a signifigantly larger frequency than the other Winners, approximately 27.55% and 22.45% resepctively. Borden, Falv-o-Rich, Cloverlead and Southern Belle are also have larger frequencies than the rest, 10.71%, 9.18%, 7.91% and 6.63% respectively. The other 14 are all under 3.1% relative frequency.

District : The number of distinct districts over the years in our data is 51, which increased from 1983 till 1986 where it became stable, it then began to drop from 1989 to 1990 and dropped drastically from 1990-1991. We suspect it was because of the drop in dairy price from 1990 to 1991.

WWBID, LFCBID, LFWBID : These variables are signficantly skewed to the right. There doesnt appear to be any outliers. The data being skewed right suggests that majority of the prices were above the competitive average price for low-fat chocolate milk. The sizes of these vaiables are over 40, by the central limit theorem all these variables have an approximate normal distribution.

Analysis and Main Findings

Objective

In several school districts in northern Kentucky (called the “tricounty” market), two suppliers-Meyer Dairy and Trauth Dairy - were accused of price-fixing - That is, conspiring to allocate the districts so that the winning bidder was predetermined and the price per pint was set above the competitive price. Our objective is to determine if the accusation is true or false.

Research Questions and Hypothesis

Research Questions

1. Did Meyer and Trauth conspire to rig their bids in the tricounty market? Economic theory states that, if so, the mean winning price in the rigged tricounty market will be higher than the mean winning price in the competitive surrounding market. Is there support for the claim that the dairies in the tricounty market participated in collusive practices?

2. Market allocation is a common form of collusive behaviour in bid-rigging conspiracies. Under collusion, the same dairy usually controls the same school districts year after year. The incumbency rate for a market is defined as the proportion of school districts that are won by the vendor that won the previous year. Past experience with milk bids in a competitive environment reveals that a typical incumbency rate is 0.7. That is, 70% of the school districts are expected to purchase their milk from the dairy that won the previous year. Incumbency rates of 0.9 or higher are strong indicators of collusive bidding. Over the years, when bid collusion was alleged to have occurred in northern Kentucky, there were 51 potential vendor transitions (i.e. changes in milk supplier from one year to the next year in a district.) in the tricounty market, and 134 potential vendor transitions in the surrounding market. By comparing the two incumbency rates of the tricounty and surrounding milk markets, can you get the results to further support for the bid collusion theory?

Hypothesis

The dairy companies Trauth and Meyer colluded in the tri-county market between the years 1984 and 1988

1. Null Hypothesis: the mean for the WWBID, LFCBID, and LFWBID in the Surround Market is greater than or equal to the yearly mean for the WWBID, LFCBID, and LFWBID in the Tri-County Market between the years 1984 and 1988.
   Alternative Hypothesis: the yearly mean for the WWBID, LFCBID, and LFWBID in the Surround Market is less than the yearly mean for the WWBID, LFCBID, and LFWBID in the Tri-County Market between the years 1984 and 1988.

2. Null Hypothesis: the incumbency rates in the Tri-county market between the years 1984 and 1988 is lesser than and equal to 70%.
   Alternative Hypothesis: the incumbency rates between the years 1984 and 1988 in the Tri-county market is greater than 70%.

Statistical Inference

Statistical Inference Methods

Assumptions

- The sample population has a distribution that is approximately normal
- Sample is random from a large population

Question 1.
See Part A in Appendix
x_1 : Surround price yearly average
x_2 : Tri-County price yearly average
$x_d = x_1 - x_2$
$\overline{x_d}$ = -0.01885896

s_d = 0.006932496
n_d = 5
Target population parameter : μ_d
Sample size : 5, Use Paired t-distribution
Statistical Inference : Test of Hypothesis
t statistic = $\frac{\overline{x_d}}{s_d(\frac{1}{\sqrt{n_d}})}$
= -6.0829

See Part B in Appendix
WWBID
x_1 : Surround WWBID price yearly average
x_2 : Tri-County WWBID price yearly average
$x_d = x_1 - x_2$
$\overline{x_d}$ = -0.017515
s_d = 0.006669253
n_d = 5
Target population parameter : μ_d
Sample size : 5, Use Paired t-distribution
Statistical Inference : Test of Hypothesis
t statistic = $\frac{\overline{x_d}}{s_d(\frac{1}{\sqrt{n_d}})}$
= -5.8724

LFWBID
x_1 : Surround LFWBID price yearly average
x_2 : Tri-County LFWBID price yearly average
$x_d = x_1 - x_2$
$\overline{x_d}$ = -0.01997653
s_d = 0.006669772
n_d = 5
Target population parameter : μ_d
Sample size : 5, Use Paired t-distribution
Statistical Inference : Test of Hypothesis
t statistic = $\frac{\overline{x_d}}{s_d(\frac{1}{\sqrt{n_d}})}$
= -6.6972

LFCBID x_1 : Surround LFCBID price yearly average
x_2 : Tri-County LFCBID price yearly average
$x_d = x_1 - x_2$
$\overline{x_d}$ = -0.02102761
s_d = 0.007325544
n_d = 5
Target population parameter : μ_d
Sample size : 5, Use Paired t-distribution
Statistical Inference : Test of Hypothesis
t statistic = $\frac{\overline{x_d}}{s_d(\frac{1}{\sqrt{n_d}})}$ = -6.4185

Question 2.
See Part C in Appendix
\overline{x = 0.7653061
$\mu_0 = 0.7$
$n = 4$
$s = 0.03534798$
Target population parameter : μ
Sample size : 4, Use Paired t-distribution
Statistical Inference : Test of Hypothesis
t statistic = \frac{\overline{x} - μ_{0}}{s(\frac{1}{√n)} = 3.695

Results: Analysis and Interpretation

Analysis

Question 1.
See Part A in Appendix
$\alpha = .05$
p-value = 0.001846 We cannot reject the null hypothesis since p-value < α

See Part B in Appendix
WWBID p-value = 0.0021
LFWBID p-value = 0.001293
LFCBID p-value = 0.001514
We reject the null hypothesis since WWBID p-value < α, LFWBID p-value < α, and LFCBID p-value < α

Question 2.
See Part C in Appendix
$\alpha = .05$
p-value = 0.0172
We reject the null hypothesis since p-value < α

Interpretation

Question 1 Based on the yearly average prices, there is evidence at the 5% level that the yearly mean for the WWBID, LFCBID, and LFWBID in the Surround Market is less than the yearly mean for the WWBID, LFCBID, and LFWBID in the Tri-County Market between the years 1984 and 1988.

Question2 Based on the incumbency rate of the tri-county market between the years 1984 and 1988, there is evidence at the 5% level the incumbency rates in the Tri-county market exceeds 70% which is a sign of non-collusive behaviour..

Conclusion

There is a collusive environment in the Tri-County market and the market shares, bid dispersions, Comparison of average winning bid prices and incumbency rates all suggest any collusive bidding patterns

Appendix1

#Appendix2 **Tables Graphs**

##Question 1 Did Meyer and Trauth conspire to rig their bids in the tricounty market? Economic theory states that, if so, the mean winning price in the rigged tricounty market will be higher than the mean winning price in the competitive surrounding market. Is there support for the claim that the dairies in the tricounty market participated in collusive practices?

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──
```

```
## ✓ ggplot2 3.3.2     ✓ purrr   0.3.4
## ✓ tibble  3.0.3     ✓ dplyr   1.0.0
## ✓ tidyr   1.1.0     ✓ stringr 1.4.0
## ✓ readr   1.3.1     ✓ forcats 0.5.0
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
MILK <- read_csv("MILK.CSV" )
```

```
## Parsed with column specification:
## cols(
##   Year = col_double(),
##   Market = col_character(),
##   Winner = col_character(),
##   WWBID = col_double(),
##   LFWBID = col_double(),
##   LFCBID = col_double(),
##   District = col_double()
## )
```

```
MILK$District <- as.character(MILK$District)
MILK$Year <- as.character(MILK$Year)
summary(MILK)
```

```
##      Year              Market             Winner              WWBID       
##  Length:392         Length:392         Length:392         Min.   :0.1041  
##  Class :character   Class :character   Class :character   1st Qu.:0.1231  
##  Mode  :character   Mode  :character   Mode  :character   Median :0.1313  
##                                                           Mean   :0.1359  
##                                                           3rd Qu.:0.1475  
##                                                           Max.   :0.1800  
##                                                           NA's   :38      
##      LFWBID           LFCBID         District        
##  Min.   :0.0987   Min.   :0.1010   Length:392        
##  1st Qu.:0.1139   1st Qu.:0.1132   Class :character  
##  Median :0.1242   Median :0.1234   Mode  :character  
##  Mean   :0.1280   Mean   :0.1283                     
##  3rd Qu.:0.1394   3rd Qu.:0.1410                     
##  Max.   :0.1690   Max.   :0.1690                     
##  NA's   :20       NA's   :63
```

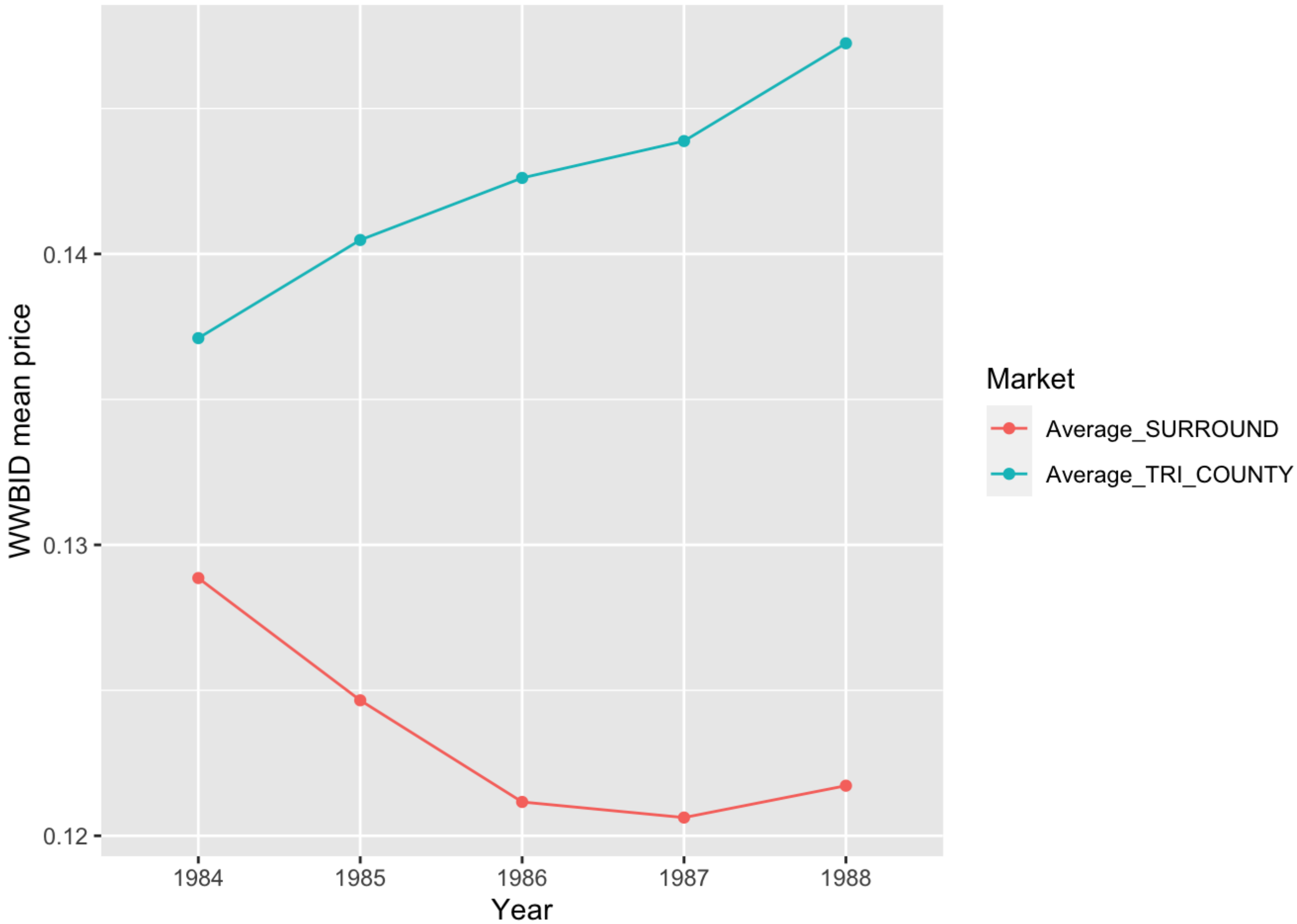
A

```
MILK4 <- filter(MILK, !is.na(MILK$WWBID) )
MILK4 <- filter(MILK4, !is.na(MILK4$LFCBID) )
MILK4 <- filter(MILK4,  !is.na(MILK4$LFWBID))
ave <- data.frame(matrix(ncol = 3, nrow = 0))
colnames(ave) <- c("Year", "Average")
for (i in c(1984:1988)){
  ave <- rbind(ave, c(i, mean(c(mean(filter(MILK4, Year == i & Market == "SURROUND")$WWBID),mean(filter(MILK4, Year == i & Market == "SURROUND")$LFWBID),mean(filter(MILK4, Year == i & Market == "SURROUND")$LFCBID)))
                      ,mean(c( mean(filter(MILK4, Year == i & Market == "TRI-COUNTY")$WWBID),mean(filter(MILK4, Year == i & Market == "TRI-COUNTY")$LFCBID),mean(filter(MILK4, Year == i & Market == "TRI-COUNTY")$LFWBID)))))
}
colnames(ave) <- c("Year", "Average_SURROUND", "Average_TRI_COUNTY")
ave
```

```
##   Year Average_SURROUND Average_TRI_COUNTY
## 1 1984        0.1288596          0.1371067
## 2 1985        0.1246571          0.1404800
## 3 1986        0.1211630          0.1426133
## 4 1987        0.1206265          0.1438810
## 5 1988        0.1217216          0.1472417
```

```
ave1 <- gather(ave, key = "Market", value = "Mean_price", "Average_SURROUND", "Average_TRI_COUNTY")
ggplot(data = ave1, aes(x = Year, y = Mean_price, color = Market)) +
  geom_line()+
  geom_point()+
  labs(y = "WWBID mean price") +
  scale_x_discrete(limits = ave$Year)
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
#95% t.test Paired Data TOH
t.test(ave$Average_SURROUND-ave$Average_TRI_COUNTY,alternative="less",mu=0,conf.level=1-
.05)
```

```
##
##  One Sample t-test
##
## data:  ave$Average_SURROUND - ave$Average_TRI_COUNTY
## t = -6.0829, df = 4, p-value = 0.001846
## alternative hypothesis: true mean is less than 0
## 95 percent confidence interval:
##         -Inf -0.01224959
## sample estimates:
##   mean of x
## -0.01885896
```

```
sd(ave$Average_SURROUND-ave$Average_TRI_COUNTY)
```

```
## [1] 0.006932496
```

B

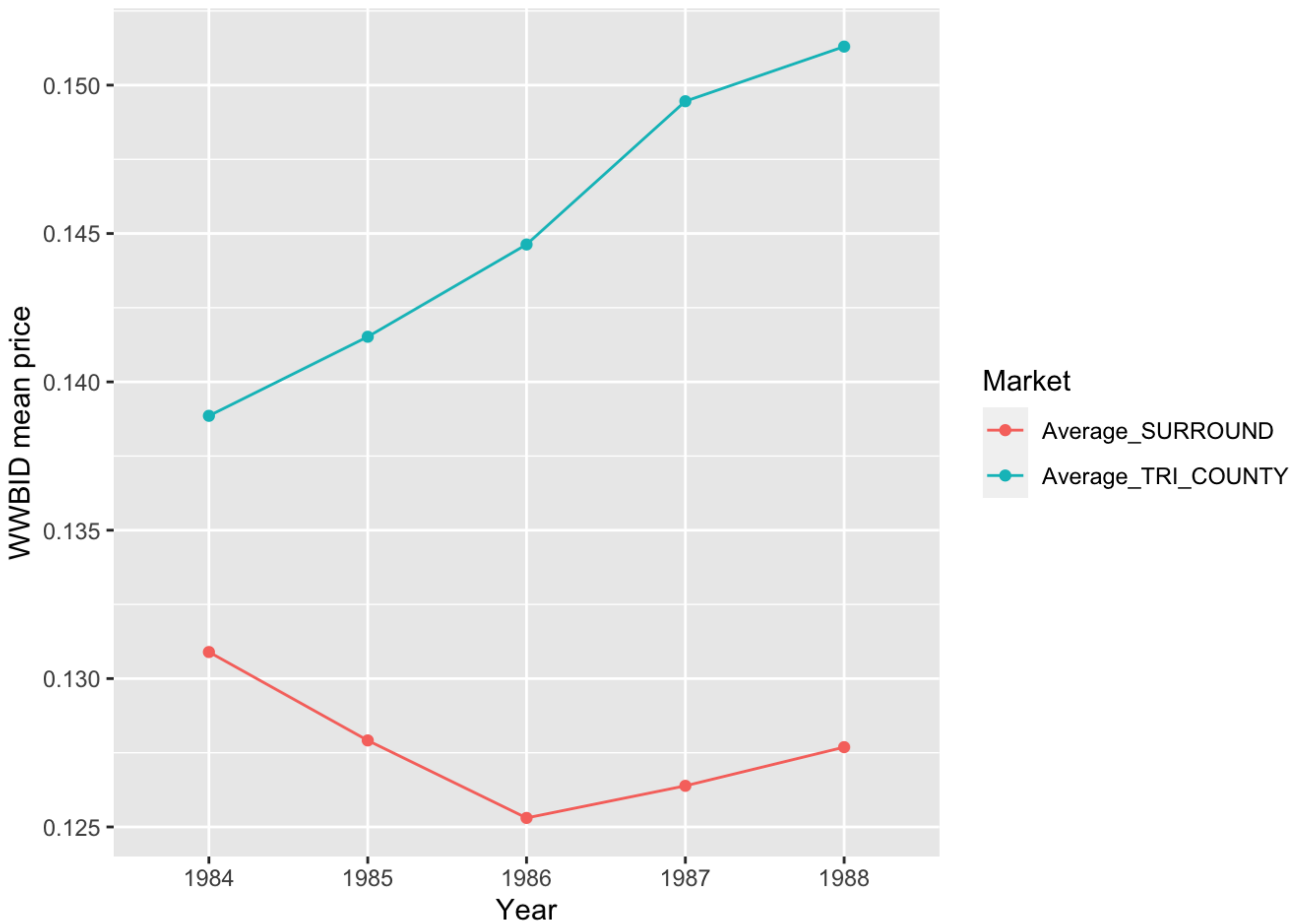
Comparing the mean

```
MILK1 <- filter(MILK, !is.na(MILK$WWBID))
ave <- data.frame(matrix(ncol = 3, nrow = 0))
colnames(ave) <- c("Year", "Average")
for (i in c(1984:1988)){
  ave <- rbind(ave, c(i, mean(filter(MILK1, Year == i & Market == "SURROUND")$WWBID), me
an(filter(MILK1, Year == i & Market == "TRI-COUNTY")$WWBID)))
}
colnames(ave) <- c("Year", "Average_SURROUND", "Average_TRI_COUNTY")
ave
```

```
##   Year Average_SURROUND Average_TRI_COUNTY
## 1 1984        0.1308955          0.1388556
## 2 1985        0.1279154          0.1415200
## 3 1986        0.1253030          0.1446300
## 4 1987        0.1263861          0.1494583
## 5 1988        0.1276889          0.1513000
```

```
ave1 <- gather(ave, key = "Market", value = "Mean_price", "Average_SURROUND", "Average_T
RI_COUNTY")
ggplot(data = ave1, aes(x = Year, y = Mean_price, color = Market)) +
  geom_line()+
  geom_point()+
  labs(y = "WWBID mean price") +
  scale_x_discrete(limits = ave$Year)
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
#95% t.test Paired Data TOH
t.test(ave$Average_SURROUND-ave$Average_TRI_COUNTY,alternative="less",mu=0,conf.level=1-.05)
```

```
##
##  One Sample t-test
##
## data:  ave$Average_SURROUND - ave$Average_TRI_COUNTY
## t = -5.8724, df = 4, p-value = 0.0021
## alternative hypothesis: true mean is less than 0
## 95 percent confidence interval:
##        -Inf -0.0111566
## sample estimates:
## mean of x
## -0.017515
```

```
sd(ave$Average_SURROUND-ave$Average_TRI_COUNTY)
```

```
## [1] 0.006669253
```

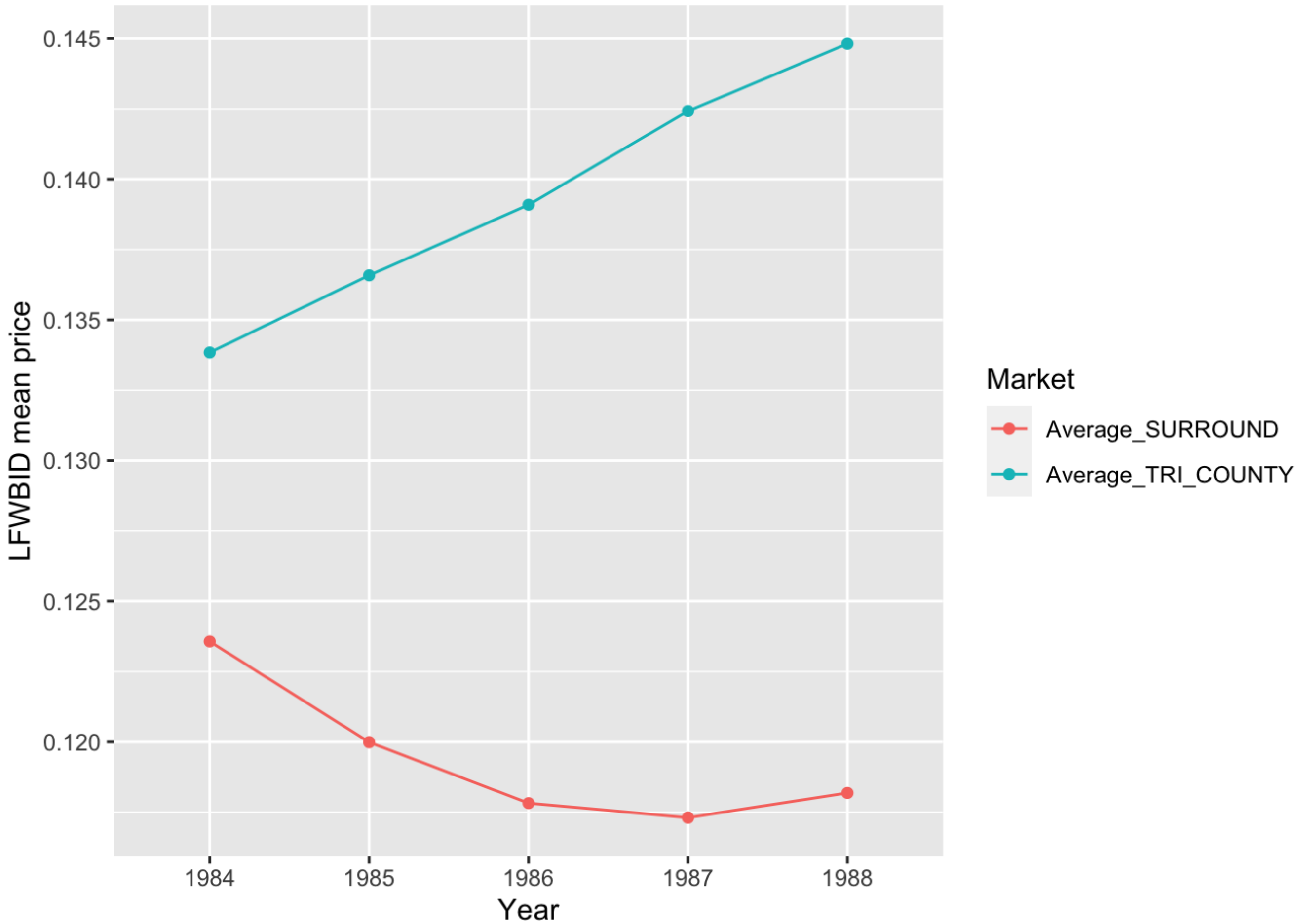
As can be seen in the graph, the mean TRI-COUNTY white milk bidding price is always higher than that of the SURROUND. Based on the mean of both TRI_COUNTY and SURROUND white milk bidding price, there is evidence at the 5% level that the mean winning price in the rigged tricounty market will be higher than the mean winning price in the competitive surrounding market for the white milk bidding price.

```
MILK2 <- filter(MILK, !is.na(MILK$LFWBID))
ave <- data.frame(matrix(ncol = 3, nrow = 0))
colnames(ave) <- c("Year", "Average")
for (i in c(1984:1988)){
  ave <- rbind(ave, c(i, mean(filter(MILK2, Year == i & Market == "SURROUND")$LFWBID), m
ean(filter(MILK2, Year == i & Market == "TRI-COUNTY")$LFWBID)))
}
colnames(ave) <- c("Year", "Average_SURROUND", "Average_TRI_COUNTY")
ave
```

```
##   Year Average_SURROUND Average_TRI_COUNTY
## 1 1984        0.1235692          0.1338417
## 2 1985        0.1199897          0.1365846
## 3 1986        0.1178212          0.1390923
## 4 1987        0.1173086          0.1424231
## 5 1988        0.1181857          0.1448154
```

```
ave1 <- gather(ave, key = "Market", value = "Mean_price", "Average_SURROUND", "Average_T
RI_COUNTY")
ggplot(data = ave1, aes(x = Year, y = Mean_price, color = Market)) +
  geom_line()+
  geom_point()+
  labs(y = "LFWBID mean price")  +
  scale_x_discrete(limits = ave$Year)
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
#95% t.test Paired Data TOH
t.test(ave$Average_SURROUND-ave$Average_TRI_COUNTY,alternative="less",mu=0,conf.level=1-
.05)
```

```
##
##  One Sample t-test
##
## data:  ave$Average_SURROUND - ave$Average_TRI_COUNTY
## t = -6.6972, df = 4, p-value = 0.001293
## alternative hypothesis: true mean is less than 0
## 95 percent confidence interval:
##        -Inf -0.01361763
## sample estimates:
##   mean of x
## -0.01997653
```

```
sd(ave$Average_SURROUND-ave$Average_TRI_COUNTY)
```

```
## [1] 0.006669772
```

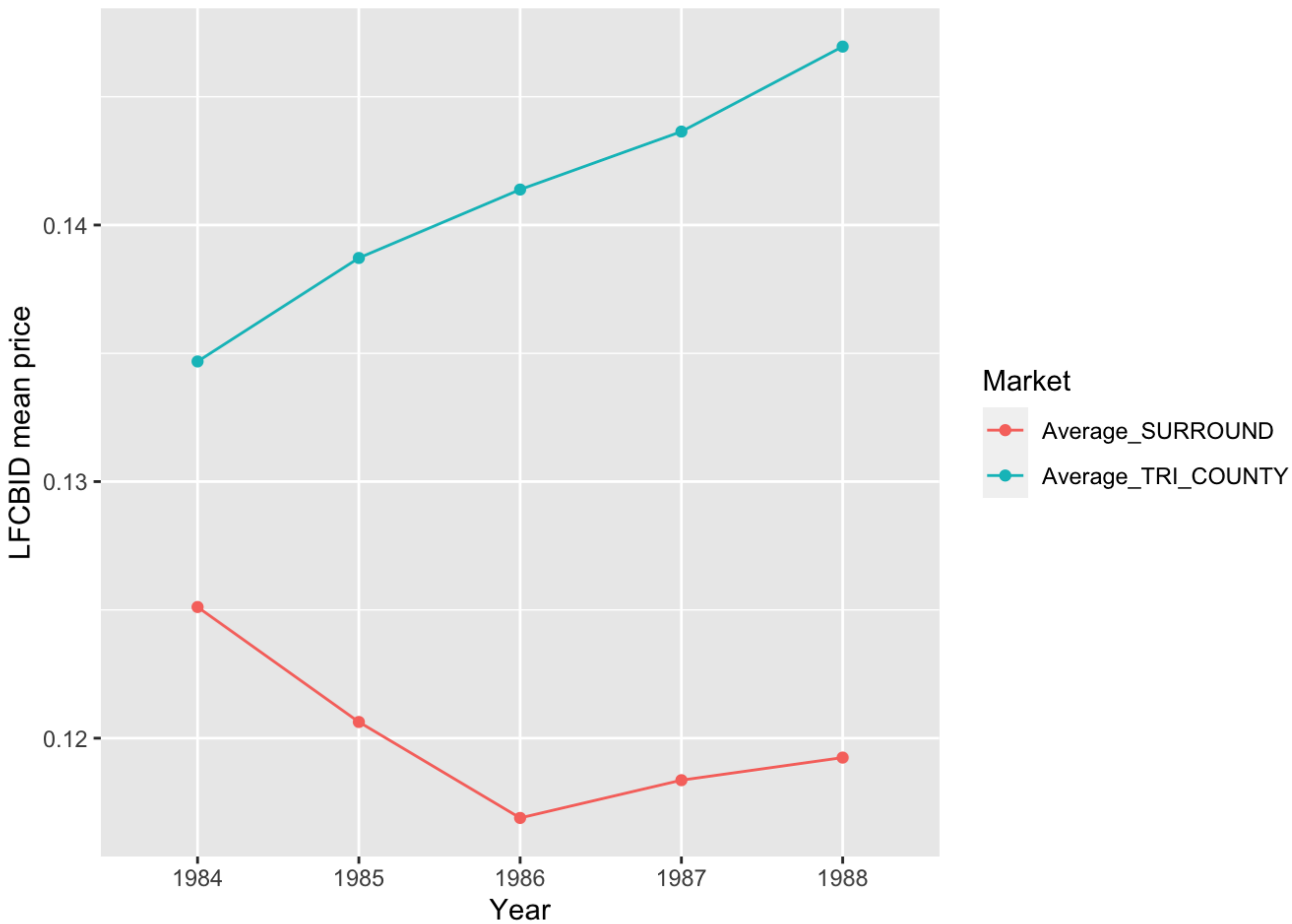
As can be seen in the graph, the mean TRI-COUNTY winning bid price of low-fat white milk average is always higher than that of the SURROUND. Based on the mean of both TRI_COUNTY and SURROUND winning bid price of low-fat white milk , there is evidence at the 5% level that the mean winning bid price of low-fat white milk in the rigged tricounty market will be higher than the mean winning price in the competitive surrounding market for the winning bid price of low-fat white milk.

```
MILK3 <- filter(MILK, !is.na(MILK$LFCBID))
ave <- data.frame(matrix(ncol = 3, nrow = 0))
colnames(ave) <- c("Year", "Average")
for (i in c(1984:1988)){
  ave <- rbind(ave, c(i, mean(filter(MILK3, Year == i & Market == "SURROUND")$LFCBID), m
ean(filter(MILK3, Year == i & Market == "TRI-COUNTY")$LFCBID)))
}
colnames(ave) <- c("Year", "Average_SURROUND", "Average_TRI_COUNTY")
ave
```

```
##   Year Average_SURROUND Average_TRI_COUNTY
## 1 1984        0.1251120          0.1346833
## 2 1985        0.1206321          0.1387167
## 3 1986        0.1168912          0.1413833
## 4 1987        0.1183639          0.1436429
## 5 1988        0.1192444          0.1469556
```

```
ave1 <- gather(ave, key = "Market", value = "Mean_price", "Average_SURROUND", "Average_T
RI_COUNTY")
ggplot(data = ave1, aes(x = Year, y = Mean_price, color = Market)) +
  geom_line()+
  geom_point()+
  labs(y = "LFCBID mean price") +
  scale_x_discrete(limits = ave$Year)
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
#95 % t.test Paired Data TOH
t.test(ave$Average_SURROUND-ave$Average_TRI_COUNTY,alternative="less",mu=0,conf.level=1-
.05)
```

```
##
##  One Sample t-test
##
## data:  ave$Average_SURROUND - ave$Average_TRI_COUNTY
## t = -6.4185, df = 4, p-value = 0.001514
## alternative hypothesis: true mean is less than 0
## 95 percent confidence interval:
##         -Inf -0.01404351
## sample estimates:
##   mean of x
## -0.02102762
```

```
sd(ave$Average_SURROUND-ave$Average_TRI_COUNTY)
```

```
## [1] 0.007325544
```

As can be seen in the graph, the mean TRI-COUNTY winning bid price of low-fat chocolate milk is always higher than that of the SURROUND. Based on the mean of both TRI_COUNTY and SURROUND winning bid price of low-fat chocolate milk, there is evidence at the 5% level that the mean winning bid price of low-fat chocolate milk in the rigged tricounty market will be higher than the mean winning price in the competitive surrounding market for winning bid price of low-fat chocolate milk.

C

```
SURROUND <- filter(MILK, Market == "SURROUND")
TRI_COUNTY <- filter(MILK, Market == "TRI-COUNTY")
TRI<- TRI_COUNTY %>% gather(key = "Milk", value = "Price", WWBID, LFWBID, LFCBID)  %>% s
elect(Year, District, Winner, Milk) %>% group_by(Year, District) %>% mutate(grouped_id =
row_number()) %>% spread(Year, Winner) %>%
  select(-grouped_id)
SUR <- SURROUND %>% gather(key = "Milk", value = "Price", WWBID, LFWBID, LFCBID) %>% sel
ect(Year, District, Winner, Milk) %>% group_by(Year, District) %>% mutate(grouped_id = r
ow_number())%>% spread(Year, Winner) %>%
  select(-grouped_id)
TRI
```

```
## # A tibble: 49 x 11
## # Groups:   District [13]
##    District Milk  `1983` `1984` `1985` `1986` `1987` `1988` `1989` `1990` `1991`
##    <chr>    <chr> <chr>  <chr>  <chr>  <chr>  <chr>  <chr>  <chr>  <chr>  <chr> 
##  1 134      LFCB… MEYER  MEYER  MEYER  MEYER  MEYER  MEYER  MEYER  MEYER  MEYER 
##  2 134      LFWB… MEYER  MEYER  MEYER  MEYER  MEYER  MEYER  MEYER  MEYER  MEYER 
##  3 134      WWBID MEYER  MEYER  MEYER  MEYER  MEYER  MEYER  MEYER  MEYER  MEYER 
##  4 147      LFCB… <NA>   <NA>   MEYER  MEYER  TRAUTH TRAUTH TRAUTH <NA>   TRAUTH
##  5 147      LFCB… <NA>   <NA>   <NA>   <NA>   <NA>   <NA>   <NA>   MEYER  <NA>  
##  6 147      LFCB… <NA>   <NA>   <NA>   <NA>   <NA>   <NA>   <NA>   TRAUTH <NA>  
##  7 147      LFWB… <NA>   <NA>   MEYER  MEYER  TRAUTH TRAUTH TRAUTH <NA>   TRAUTH
##  8 147      LFWB… <NA>   <NA>   <NA>   <NA>   <NA>   <NA>   <NA>   MEYER  <NA>  
##  9 147      LFWB… <NA>   <NA>   <NA>   <NA>   <NA>   <NA>   <NA>   TRAUTH <NA>  
## 10 147      WWBID <NA>   <NA>   MEYER  MEYER  TRAUTH TRAUTH TRAUTH MEYER  TRAUTH
## # … with 39 more rows
```

```
SUR
```

```
## # A tibble: 124 x 11
## # Groups:   District [38]
##    District Milk  `1983` `1984` `1985` `1986` `1987` `1988` `1989` `1990` `1991`
##    <chr>    <chr> <chr>  <chr>  <chr>  <chr>  <chr>  <chr>  <chr>  <chr>  <chr> 
##  1 101      LFCB… HILL'… HILL'… FRENC… HILL'… THOMP… TRAUTH THOMP… THOMP… <NA>  
##  2 101      LFWB… HILL'… HILL'… FRENC… HILL'… THOMP… TRAUTH THOMP… THOMP… <NA>  
##  3 101      WWBID HILL'… HILL'… FRENC… HILL'… THOMP… TRAUTH THOMP… THOMP… <NA>  
##  4 11       LFCB… <NA>   <NA>   SOUTH… BORDEN BORDEN BORDEN HOLLA… SOUTH… <NA>  
##  5 11       LFWB… <NA>   <NA>   SOUTH… BORDEN BORDEN BORDEN HOLLA… SOUTH… <NA>  
##  6 11       WWBID <NA>   <NA>   SOUTH… BORDEN BORDEN BORDEN HOLLA… SOUTH… <NA>  
##  7 121      LFCB… SOUTH… WESTO… WESTO… TRAUTH TRAUTH TRAUTH TRAUTH TRAUTH <NA>  
##  8 121      LFWB… SOUTH… WESTO… WESTO… TRAUTH TRAUTH TRAUTH TRAUTH TRAUTH <NA>  
##  9 121      WWBID SOUTH… WESTO… WESTO… TRAUTH TRAUTH TRAUTH TRAUTH TRAUTH <NA>  
## 10 13       LFCB… CLOVE… CLOVE… CLOVE… CLOVE… CLOVE… CLOVE… MEYER  MEYER  <NA>  
## # … with 114 more rows
```

```
rate1 <- c(
           sum(TRI$`1984` == TRI$`1985`, na.rm = T)/49,
           sum(TRI$`1985` == TRI$`1986`, na.rm = T)/49,
           sum(TRI$`1986` == TRI$`1987`, na.rm = T)/49,
           sum(TRI$`1987` == TRI$`1988`, na.rm = T)/49
)

rate2 <- c(
           sum(SUR$`1984` == SUR$`1985`, na.rm = T)/124,
           sum(SUR$`1985` == SUR$`1986`, na.rm = T)/124,
           sum(SUR$`1986` == SUR$`1987`, na.rm = T)/124,
           sum(SUR$`1987` == SUR$`1988`, na.rm = T)/124)
incumbency_rate_tri <- data.frame(c(1984:1991), rate1)
colnames(incumbency_rate_tri) <- c("Year", "Incumbency_rate")
incumbency_rate_sur <- data.frame(c(1984:1991), rate2)
colnames(incumbency_rate_sur) <- c("Year", "Incumbency_rate")
t.test(rate1,alternative = "greater", mu = .7, conf.level = 1-.05)
```

```
##
##  One Sample t-test
##
## data:  rate1
## t = 3.695, df = 3, p-value = 0.0172
## alternative hypothesis: true mean is greater than 0.7
## 95 percent confidence interval:
##  0.7237128       Inf
## sample estimates:
## mean of x
## 0.7653061
```

```
sd(rate1)
```

```
## [1] 0.03534798
```

Checking if there is a collusive environment

Collusive Market Environment: certain economic features of a market create an environment in which collusion may be found. These basic features include the following: – Few sellers and high concentration. Only a few diaries control all or nearly all of the milk business in the market.

```
TRI_COUNTY <- filter(MILK, Market == "TRI-COUNTY")
summarise(group_by(TRI_COUNTY, Winner), freq = n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 2
##   Winner  freq
##   <chr>  <int>
## 1 MEYER     52
## 2 TRAUTH    63
```

– Homogeneous products. The products sold are essentially the same from the standpoint of the buyer, i.e., the school district. – Inelastic demand. Demand is relatively insensitive to price. Note: the quantity of milk required by a school district is primarily determined by school enrollment, not price. – similar costs. The dairies bidding for the milk contracts face similar cost conditions. Note: approximately 60% of a dairy's production cost is raw milk, which is federally regulated. Meyer and Trauth are dairies of similar size, and both bought their raw milk from the same supplier.

Collusive Bidding Patterns: The analyses of patterns in sealed bids reveal much about the level of competition, or lack thereof, among the vendors serving the market. Con- sider the following bid analyses:

– Market shares. A market share for a dairy is the number of milk half-pint supplied by the diary over a given school year, divided by the total number of half-pints supplied to the entire market. One sign of potential collusive behavior is stable, nerly equal market shares over time for the dairies under investigation.

```
TRI_COUNTY <- filter(MILK, Market == "TRI-COUNTY")
market_share <- data.frame(matrix(ncol = 3, nrow = 0))
for (i in c(1983:1991)){
  market_share <- rbind(market_share, c(i, nrow(filter(TRI_COUNTY, Year == i & Winner ==
"MEYER")), nrow(filter(TRI_COUNTY, Year == i & Winner == "TRAUTH"))))
}
colnames(market_share) <- c("Year", "MEYER", "TRAUTH")
market_share
```

```
##   Year MEYER TRAUTH
## 1 1983     6      5
## 2 1984     5      7
## 3 1985     6      7
## 4 1986     6      7
## 5 1987     5      8
## 6 1988     5      8
## 7 1989     7      6
## 8 1990     6      8
## 9 1991     6      7
```

– Bid levels and dispersion. In competitive sealed bid markets, vendors do not share information about their bids. Consequently, more dispersion or variability among the bids is observed than in collusive markets., where vendors communicate about their bids and have a tendency to submit bids in close proximity to one another in an attempt to make the bidding appear competitive. Furthermore, in competitive markets the bid dispersion tends to be directly proportional to the level of the bid: when bids are submitted at relatively high levels, there is more variability among the bids than when they are submitted at or near marginal cost, which will be approximately the same among dairies in the same geographic market.

```
SURROUND <- filter(MILK, Market == "SURROUND")
SURROUND <- filter(SURROUND, Year > 1983 )
SURROUND <- filter(SURROUND,  Year< 1988)
TRI_COUNTY <- filter(MILK, Market == "TRI-COUNTY")
TRI_COUNTY <- filter(TRI_COUNTY, Year > 1983 )
TRI_COUNTY <- filter(TRI_COUNTY, Year< 1988 )
var(TRI_COUNTY$WWBID, na.rm = T)
```

```
## [1] 3.951745e-05
```

```
var(SURROUND$WWBID, na.rm = T)
```

```
## [1] 0.0001408355
```

```
var(TRI_COUNTY$LFWBID, na.rm = T)
```

```
## [1] 4.452827e-05
```

```
var(SURROUND$LFWBID, na.rm = T)
```

```
## [1] 0.0001774669
```

```
var(TRI_COUNTY$LFCBID, na.rm = T)
```

```
## [1] 4.812893e-05
```

```
var(SURROUND$LFCBID, na.rm = T)
```

```
## [1] 0.000201219
```

– Bid sequence. School milk bids are submitted over the spring and summer months, generally at the end of one school year and before the beginning of the next. When the bids are examined in sequence in competitive markets, the level of bidding is expected to fall as the bidding season progresses. (This phenomenon is attributable to the learning process that occurs during the season, with bids adjusted accordingly. Dairies may submit relatively high bids early in the season to “test the market”, confident that volume can be picked up later if the early high bids lose. But, dairies who do not win much business early in the season are likely to become more aggressive in their bidding as the season progresses, driving price levels down.) Constant or slightly increasing price patterns of sequential bids in a market where a single dairy wins year after year is considered another indication of collusive behaviour. True See A and B