A Tour of Financial Modeling

Dale W.R. Rosenthal

1 June 2018
R/Finance 2018
Introduction

Who am I?

- Head partner, Q$_{36}$: quant finance firm.
- Was: finance professor at UIC; now teach MSF class at Notre Dame.
- Presented @ ECB, FDIC, CFTC, BdF, BoFinland, NB Slovakia, RBNZ.
- 3 years: Proprietary algo trader, Equity Trading Lab, Morgan Stanley
- 5 years: Equity derivatives strategist, Long-Term Capital Management
- Interned at Goldman Sachs
- Co-organizer of R/Finance conference since start in 2009.
Roadmap

- We often talk about fundamentals and modeling.
- Must unite these to make business/trading decisions.
- Over the next hour we are going to explore a few markets.
- The goal: see how to build a model and infer business knowledge.
- A beautiful aspect of financial theory is its wide reach.
- We are going to take an idea and see how far we can reach.
- We will try modeling different aspects of a spread:
  - A “brew” spread. Thoroughly.
- Will try to tease out spread, how it has changed.
- Maybe see implications for trading, IO, valuation?
A brew spread (or “corn crush”) is ethanol less inputs.

Typically, we look at the two biggest inputs:

- Corn and natgas (to distill corn mash into ethanol).

Typical inputs: 1 bushel of corn and 72.8k BTUs of natgas

Typical outputs: 2.8 gal ethanol, 17 lbs of DDGs, 0.7 lbs corn oil.

Researchers at UIUC (Farm Doc!) suggests 42 cents/gal of costs

Ethanol, corn, natgas: all traded futures; DDGs no longer.

Does ratio change over time or with corn price levels?

\[\text{farmdocdaily.illinois.edu/2017/02/}\]
\[\text{the-profitability-of-ethanol-production-in-2016.html} \]
Data

- To answer these questions, we will look at US daily data.
- Specifically: Futures end-of-day settlement prices.
- Obviously, daily data ignores intraday spread dynamics.
- Data limitations and contract changes:
  - Natgas e-mini contract introduced in 2014; still low-volume
## Contract Conversions?

- Must know units for our prices vs. units for contracts.

<table>
<thead>
<tr>
<th>Contract</th>
<th>Sym</th>
<th>Quoted</th>
<th>Contract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn</td>
<td>C</td>
<td>$/bu</td>
<td>5000 bu</td>
</tr>
<tr>
<td>Natgas</td>
<td>NG</td>
<td>$/mmBTU</td>
<td>10000 mmBTU</td>
</tr>
<tr>
<td>Mini Natgas</td>
<td>QG</td>
<td>$/mmBTU</td>
<td>2500 mmBTU</td>
</tr>
<tr>
<td>Railcar Ethanol</td>
<td>EH</td>
<td>$/gal</td>
<td>29000 gal</td>
</tr>
<tr>
<td>Platt’s Ethanol</td>
<td>CU</td>
<td>$/gal</td>
<td>42000 gal</td>
</tr>
</tbody>
</table>

- Why 42000 gallons? Equal to 1000 barrels (bbl).
- So maybe we should trade 15–5–2 corn–ethanol–mini natgas?
We load the data using R and Quandl. Blend ethanol prices:

```r
library(xts)
library(Quandl)
futures.tickers <- paste("CHRIS/CME_", c("NG1", "C1", "CU1", "EH1"), sep="")
settle.column <- "Settle"
futs.raw <- Quandl(futures.tickers[1], type="xts")[,settle.column]
for (i in 2:length(futures.tickers)) {
  tmp <- Quandl(futures.tickers[i], type="xts")[,settle.column]
  futs.raw <- cbind(futs.raw, tmp)
}
colnames(futs.raw) <- c("natgas", "corn", "platts.ethanol", "railcar.ethanol")

## Make smooth transition from railcar ethanol to Platt’s
transition.length <- length(index(futs.raw["20170601/20170831"]))
transition.wts <- (1:transition.length)/transition.length
futs.raw$ethanol <-
  rbind(futs.raw$railcar.ethanol["/20170531"],
         futs.raw$railcar.ethanol["20170601/20170831"]*(1-transition.wts)+
         futs.raw$platts.ethanol["20170601/20170831"]*transition.wts,
         futs.raw$platts.ethanol["20170901/"])
futs <- futs.raw["20111001/"]
```
The first step is always to look at the data.
Are there spikes? data errors? comovement?
For spreads, it helps to plot them on similar scales:

```r
plot(futs$corn, col="black")
plot(futs$natgas, col="blue")
plot(futs$railcar.ethanol, col="darkgreen")
plot(futs$platts.ethanol, col="lightgreen")
plot(futs$ethanol, col="green")
```
Observation: Plot
We can notice a few things from these plots.

- Corn prices were higher up to mid-2013, now lower. (Tech shock?)
- Natgas has weird spikes in Jan+Feb (Polar Vortex) and late-2014.
- Spikes in ethanol about 1.5 months later; inventory implications.
- Spike transmission also suggests many ethanol producers unhedged.
- Natgas spike in late 2017 does not affect ethanol prices. Hedged?
- Blending of ethanol prices does not seem to have caused any issues.
We noted earlier that we could trade a brew spread: 15-5-2.

Theoretical spread (IA State) would be 15-5-2.18 (almost identical).

However, prices have different ratios. Plot brew spreads:

```r
brew.spread <- 2.8*futs$ethanol-futs$corn/100-72.8/1000*futs$natgas/15
plot(brew.spread, col="brown", main="Brew Spread (Theory)"
```

Oscillated mostly b/w -$1/bu and $2/bu except in 2014.

More recently has largely stayed b/w 0 and $1/bu.;

---

3 In brown, since maybe it is a nice brown ale.
Estimating an O-U Model

- Recall the Ornstein-Uhlenbeck SDE:
  \[ dS_t = \gamma(\mu - S_t)dt + \sigma dW_t. \] (1)

- Solving this and discretizing time, we get:
  \[ S_{i+1} = \mu(1 - e^{-\gamma\Delta t}) + S_i e^{-\gamma\Delta t} + \sigma \sqrt{1 - e^{-2\gamma\Delta t}} \frac{1}{2\gamma} \epsilon_i \] (2)
  where \( \epsilon_i \sim N(0, 1) \).

- So we can fit a linear model: \( S_{i+1} = a + bS_i + \eta_i \).

- Then we convert to O-U parameters:
  \[ \gamma = \frac{-\log(b)}{\Delta t}, \quad \mu = \frac{a}{1 - b}, \quad \sigma = \hat{\sigma}_\eta \sqrt{-\frac{2\log(b)}{\Delta t(1 - b^2)}} \] (3)

- Implied half-life of convergence: \( \log(2)/\gamma \).
Estimating an O-U Model: Code

- Fitting this linear model is easy with *R* and *xts*:
  ```r
  ou.brew <- lm(brew.spread~lag(brew.spread))
  summary(ou.brew)
  > ... 
  Estimate Std. Error t value Pr(>|t|)
  (Intercept) 0.008847 0.004190 2.111 0.0349 *
  lag(brew.spread) 0.982637 0.004600 213.636 <2e-16 ***
  ---
  Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
  Residual standard error: 0.1379 on 1650 degrees of freedom
  (24 observations deleted due to missingness)
  Multiple R-squared: 0.9651, Adjusted R-squared: 0.9651
  F-statistic: 4.564e+04 on 1 and 1650 DF, p-value: < 2.2e-16
  ```

- Good to see both intercept and slope are significant.
- \( R^2 = \text{terrible} \) metric, but good we explain 96.5% of variation.
Estimating an O-U Model: Parameters

- We then use these to solve for brew spread O-U parameters: \(^4\)

<table>
<thead>
<tr>
<th>Spread</th>
<th>(\hat{a})</th>
<th>(\hat{b})</th>
<th>(\hat{\sigma}_\eta)</th>
<th>(\hat{\mu})</th>
<th>(\hat{\gamma})</th>
<th>(\hat{\sigma})</th>
<th>HL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory</td>
<td>0.0088</td>
<td>0.9826</td>
<td>0.138</td>
<td>0.510</td>
<td>0.0175</td>
<td>0.139</td>
<td>39.6</td>
</tr>
</tbody>
</table>

- **Half-life**: mean time to revert halfway to \(\mu\).
- Half-life of about 40 *business days* = 2 calendar months.
- If spread not converging by half-life: structural change?
- Is the theoretical spread best? Tough to say. Let’s look deeper.

\(^4\)For \(\Delta t = 1\) — which means our units will be in days.
Finding Data Implied Spreads: Cointegration

- We start by asking if the prices are cointegrated.
- Use the `urca` (Unit-Root Cointegration Analysis) package.
- Must tell the procedures which model form we care about.
- Usually: want “transitory” form, “constant” spread.

```r
library(urca)
vars2study <- c("ethanol", "corn", "natgas")
ci.model <- ca.jo(futs[,vars2study], type="trace",
                 ecdet="const", spec="transitory")
summary(ci.model)
```
Finding Data Implied Spreads: Cointegration?

Test type: trace statistic, without linear trend and constant in cointegration

Eigenvalues (lambda):
[1] 2.870931e-02 4.026871e-03 1.398593e-03 1.718671e-18

Values of test statistic and critical values of test:
<table>
<thead>
<tr>
<th>test</th>
<th>10pct</th>
<th>5pct</th>
<th>1pct</th>
</tr>
</thead>
<tbody>
<tr>
<td>r &lt;= 2</td>
<td>2.32</td>
<td>7.52</td>
<td>9.24</td>
</tr>
<tr>
<td>r &lt;= 1</td>
<td>9.03</td>
<td>17.85</td>
<td>19.96</td>
</tr>
<tr>
<td>r = 0</td>
<td>57.41</td>
<td>32.00</td>
<td>34.91</td>
</tr>
</tbody>
</table>

- Is there a cointegrating relationship? Yes, just 1.
- Reject test for 0 relationships: 57.41 > 34.91, 41.07.
- Can’t reject ≤ 1 relationships. (9.03 < 19.96)
- So there does seem to be a cointegrating relationship.
Finding Data Implied Spreads: Cointegration...

```r
cajorls(ci.model)
> $beta
>    ect1  ethanol.l1  corn.l1  natgas.l1  constant
>        1.000000000 -0.002377774 -0.236428486 0.002311395
```

- What does the data suggest? Strangeness.
- Long 1 gal ethanol, short 0.24 bu corn, short 236kBTU natgas.
- So brewers used 2/3 of typical corn and 3.25 as much heat?
- Should we trade this? Probably not. You have a brain; use it.
Finding Data Implied Spreads: Cointegration Co-problems

- Cointegration doesn’t know that corn + natgas $\Rightarrow$ ethanol.
- When a model fails, your responsibility is to use your head.
- Economics drives the industry. Maybe the data is wrong.
- Data are settlement prices; set at end-of-day.
- People might try to manipulate settle prices.
- Is this likely? If settle prices are off, how to arb them?
- More likely: maybe the industry has changed over time.
Perhaps the crack spread is varying with time.

Break things up by year to investigate.

```r
years <- paste(2011:2018)
for (yr in years) {
  tmp.brew <- lm(brew.spread[yr] ~ lag(brew.spread[yr]))
  gamma.brew <- -log(coef(tmp.brew)[2])  # -log of the slope
  mu.brew <- coef(tmp.brew)[1]/(1-coef(tmp.brew)[2])
  print(sprintf("%s Brew: sd=%0.3f mu=%0.3f gamma=%0.3f hl=%0.2f",
                yr, sd(brew.spread[yr], na.rm=TRUE),
                mu.brew, gamma.brew, log(2)/gamma.brew))
}
```
Time Varying Spreads? Results

<table>
<thead>
<tr>
<th>Year</th>
<th>sd</th>
<th>$\hat{a}$</th>
<th>$\hat{b}$</th>
<th>$\hat{\sigma}_\eta$</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{h}\ell$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>0.519</td>
<td>-0.015</td>
<td>0.955</td>
<td>0.149</td>
<td>-2.753</td>
<td>0.005</td>
<td>128.89</td>
</tr>
<tr>
<td>2012</td>
<td>0.332</td>
<td>-0.012</td>
<td>0.980</td>
<td>0.070</td>
<td>-0.614</td>
<td>0.020</td>
<td>34.69</td>
</tr>
<tr>
<td>2013</td>
<td>0.726</td>
<td>0.037</td>
<td>0.943</td>
<td>0.230</td>
<td>0.646</td>
<td>0.059</td>
<td>11.81</td>
</tr>
<tr>
<td>2014</td>
<td>0.688</td>
<td>0.080</td>
<td>0.949</td>
<td>0.226</td>
<td>1.556</td>
<td>0.053</td>
<td>13.18</td>
</tr>
<tr>
<td>2015</td>
<td>0.242</td>
<td>0.013</td>
<td>0.969</td>
<td>0.056</td>
<td>0.414</td>
<td>0.032</td>
<td>21.96</td>
</tr>
<tr>
<td>2016</td>
<td>0.297</td>
<td>0.017</td>
<td>0.978</td>
<td>0.062</td>
<td>0.763</td>
<td>0.023</td>
<td>30.47</td>
</tr>
<tr>
<td>2017</td>
<td>0.208</td>
<td>0.014</td>
<td>0.972</td>
<td>0.056</td>
<td>0.502</td>
<td>0.029</td>
<td>24.19</td>
</tr>
<tr>
<td>2018</td>
<td>0.118</td>
<td>0.019</td>
<td>0.903</td>
<td>0.052</td>
<td>0.194</td>
<td>0.102</td>
<td>6.78</td>
</tr>
</tbody>
</table>

- Seems brew spread has decreased in variance, mean; faster reverting.
- Structural break after 2012? Check the news!
- Sure enough: Heavy losses; consolidation and plant closures.
- Reversion half-lives of 7–34 biz days < 39 biz days.
- Why larger full-sample half-life estimate? Structural breaks!
Estimating Industry Profitability: Data

- Now let’s switch to looking at industry profitability.
- Tough to get industry-wide refining profits; we will proxy.
- Use profits (calendar year, in millions) of ethanol refining firms:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>3997</td>
<td>3623</td>
<td>3889</td>
<td>4768</td>
<td>3966</td>
<td>3619</td>
<td>3506</td>
</tr>
<tr>
<td>ANDE</td>
<td>353</td>
<td>358</td>
<td>365</td>
<td>397</td>
<td>376</td>
<td>346</td>
<td>319</td>
</tr>
<tr>
<td>BG</td>
<td>2627</td>
<td>2573</td>
<td>2760</td>
<td>2621</td>
<td>2693</td>
<td>2410</td>
<td>1764</td>
</tr>
<tr>
<td>CHSCP</td>
<td>1734</td>
<td>1917</td>
<td>1603</td>
<td>1842</td>
<td>1311</td>
<td>900</td>
<td>910</td>
</tr>
<tr>
<td>GPRE</td>
<td>172</td>
<td>97</td>
<td>173</td>
<td>426</td>
<td>207</td>
<td>281</td>
<td>261</td>
</tr>
<tr>
<td>INGR</td>
<td>1126</td>
<td>1238</td>
<td>1131</td>
<td>1115</td>
<td>1242</td>
<td>1402</td>
<td>1473</td>
</tr>
<tr>
<td>MGPI</td>
<td>7</td>
<td>25</td>
<td>21</td>
<td>28</td>
<td>59</td>
<td>65</td>
<td>76</td>
</tr>
<tr>
<td>PEIX</td>
<td>19</td>
<td>-20</td>
<td>36</td>
<td>112</td>
<td>10</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>REGI</td>
<td>127</td>
<td>58</td>
<td>239</td>
<td>161</td>
<td>111</td>
<td>172</td>
<td>84</td>
</tr>
<tr>
<td>REX</td>
<td>35</td>
<td>14</td>
<td>64</td>
<td>142</td>
<td>51</td>
<td>71</td>
<td>44</td>
</tr>
<tr>
<td>Total</td>
<td>10197</td>
<td>9883</td>
<td>10309</td>
<td>11612</td>
<td>10024</td>
<td>9319</td>
<td>8442</td>
</tr>
</tbody>
</table>

- Sum these by year, see what predictive power spread has.
- Also need to compute average spread.
Now calculate yearly mean and sd of brew spreads.

```r
## clean up period.apply tagging w/EOY date
brew.mean <- period.apply(brew.spread, endpoints(brew.spread,"years",1),
                          mean, na.rm=TRUE)["/2017"]
brew.sd <- period.apply(brew.spread, endpoints(brew.spread,"years",1),
                        sd, na.rm=TRUE)["/2017"]
pl$mean.spread <- xts(coredata(brew.mean), order.by=pl.years)
pl$sd.spread <- xts(coredata(brew.sd), order.by=pl.years)
```

This yields the following brew spread stats:

<table>
<thead>
<tr>
<th>Year</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.72</td>
<td>-0.49</td>
<td>0.51</td>
<td>1.60</td>
<td>0.45</td>
<td>0.65</td>
<td>0.60</td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.52</td>
<td>0.33</td>
<td>0.73</td>
<td>0.69</td>
<td>0.24</td>
<td>0.30</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Add post-2014 dummy; spread dynamics seem to change then.
Profits often quadratic in mispricing; size ↑ as profits ↑.

We can fit a linear model with post-2014 dummy, mean-squared:

```r
pl$post2014 <- index(pl) >= as.POSIXct("20150101", format="%Y%m%d")
lm(total ~ brew.mean + brew.sd, data=pl)
lm(total ~ brew.mean + brew.sd + post2014, data=pl)
lm(total ~ brew.mean + I(brew.mean^2) + brew.sd, data=pl)
```

Unfortunately, results are not good:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>t-stat</td>
<td>Est.</td>
</tr>
<tr>
<td>Intercept</td>
<td>8.45 bn</td>
<td>12.8</td>
<td>9.76 bn</td>
</tr>
<tr>
<td>Spread Mean</td>
<td>0.22 bn</td>
<td>0.4</td>
<td>0.63 bn</td>
</tr>
<tr>
<td>Spread Std Dev</td>
<td>3.24 bn</td>
<td>2.0</td>
<td>0.65 bn</td>
</tr>
<tr>
<td>Post-2014</td>
<td>3.24 bn</td>
<td>2.0</td>
<td>0.65 bn</td>
</tr>
<tr>
<td>Spread Mean$^2$</td>
<td>-1.02 bn</td>
<td>-0.8</td>
<td></td>
</tr>
</tbody>
</table>

What went wrong?? Look at the companies we chose.

ADM and Bunge are large multi-commodity traders.

Ingredion, however, was once known as Corn Products.
So create a total without ADM and BG, refit.

\[ \text{pl}\$\text{revtotal } <- \text{pl}\$\text{total} - \text{pl}\$\text{ADM} - \text{pl}\$\text{BG} \]

These results are better:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>t-stat</td>
<td>Est.</td>
<td>t-stat</td>
<td>Est.</td>
<td>t-stat</td>
</tr>
<tr>
<td>Intercept</td>
<td>3.02 bn</td>
<td>12.6</td>
<td>3.79 bn</td>
<td>7.0</td>
<td>3.13 bn</td>
<td>42.0</td>
</tr>
<tr>
<td>Spread Mean</td>
<td>0.04 bn</td>
<td>0.2</td>
<td>0.28 bn</td>
<td>1.2</td>
<td>-0.36 bn</td>
<td>-4.1</td>
</tr>
<tr>
<td>Spread Std Dev</td>
<td>1.22 bn</td>
<td>2.1</td>
<td>-0.30 bn</td>
<td>-0.3</td>
<td>0.87 bn</td>
<td>4.7</td>
</tr>
<tr>
<td>Post-2014</td>
<td>-0.60 bn</td>
<td>-1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spread Mean(^2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.42 bn</td>
<td>6.4</td>
</tr>
</tbody>
</table>

So maybe we have found an OK model: Model 3.

Post-2014 dummy does not seem to be helpful.

If we had profits/gallon, results would be much stronger.

2018? Brew spread so far: mean=0.60, sd=0.21

\[ \implies \$3.2 \text{ bn profit for firms}\{\text{ADM,BG}\} \]
Valuing a Simple Refinery

- Consider building a simple ethanol refinery (100 mn gal).
- How would we estimate the value of a refinery?
- If we had a model of profits/gal, we could use that.
- We sort of do: FarmDoc (UIUC) suggests costs around $0.42/gal.

\[
\text{profit/gal} = (\bar{S}_{yr,brew} - 0.42)^+.
\]  

But this is inherently an Asian option-like payoff.

We can simulate the mean-reverting spread as:

\[
S_T = \hat{\mu} + e^{-\gamma T}(S_0 - \hat{\mu}) + \sigma \sqrt{\frac{1 - e^{-2\gamma T}}{2\gamma}} Z_T
\]  

where \( Z_T \sim iid N(0,1) \)
Valuing a Simple Refinery: Code

- Use overall O-U numbers (since post-2014 dummy not significant).
- So: $\hat{\mu} = 0.51$, $\hat{\gamma} = 0.0175$, $\hat{\sigma} = 0.139$.
- Price as of 1 Jan 2017; use $S_0 = 0.822$. (Or 2018: $S_0 = 0.07$...)
- But this is an Asian option; can approximate:
  - $\bar{S}_{0 \to T} \approx S_T - \frac{\sigma^2}{2}; \sigma \bar{S} \approx \sigma \sqrt{\frac{1}{3}}$.

```r
T <- 1; r <- 0.02  # 1 year option, 2% int. rates
s0 <- 0.822-0.139^2/2; mu <- 0.51; gamma <- 0.0175
sigma.brew <- 0.139/sqrt(3); K <- 0.42
num.sims <- 100000  # 100,000 simulations
z <- rnorm(num.sims)  # simulated Z_T values
s.sim <- mu + exp(-gamma*T)*(s0-mu) +
  sigma.brew*sqrt((1-exp(-2*gamma*T))/(2*gamma))*z
profit.sim <- pmax(s.sim - K, 0)
opt.value <- mean(profit.sim)*exp(-r*T)
opt.sd <- sd(profit.sim*exp(-r*T))
```
Valuing a Simple Refinery: Real Option Results

- For Ornstein-Uhlenbeck with Asian option approximation:
  - Refinery value: $0.379/gal; sd: $0.078/gal = $37.89 mn/year.
- For 100 mn gal/year refinery: profit = $37.9 mn/year; sd = $7.8 mn.
- $37.9 mn perpetuity value: \( \frac{37.9\text{mn}}{r_f=0.02} = 1.9 \text{ bn} \), sd $391 mn
- Compare to DCF: \( \frac{0.822-0.42}{WACC=0.08} \times 100 \text{ mn} = 503 \text{ mn} \)
- If we price as of 1 Jan 2018: \( S_0 = 0.07\ldots \) value is 0.
- Why? Hard to model politicians breaking long-established policy.

No time to look at CAPM, factor models; Jensen’s Inequality correction!
Want More?

- Buy the book! Amazon, B&N, etc.
- Come to the book launch! Email firm (info@q36llc.com)
- Check out the slides: www.q36llc.com/think/investments