# Palm Oil Case Study

## Learning Outcomes

* Describe key features of the global market for palm oil
* Use the UN ComTrade database to identify important trading patterns in palm oil
* Use R to retrieve and analyze select palm oil trade data from ComTrade
* Estimate palm oil prices and transportation costs for select trading partners
* Determine if prices and transportation cost conform to the spatial law-of-one price

## Exercise 1

Read <https://greenpalm.org/about-palm-oil/what-is-palm-oil/what-is-palm-oil-used-for>

Answer the following TRUE/FALSE questions.

1. Australia has an ideal climate for growing palm trees.
2. Palm oil has two harvesting periods, each lasting about six weeks.
3. About 50% of all supermarket products contain palm oil.
4. Palm oil is the base for dozens of different food ingredients such as stearic acid.
5. Indonesia and Malaysia together account for about one third of global palm oil production.
6. Expanding palm tree plantations is causing large losses in global biodiversity.
7. 90 percent of the world’s palm oil produers belong to the Roundtable on Sustainable Palm Oil

# Focus on Rotterdam

Rotterdam is a major port city in the Netherlands. Prior to 2004 Rotterdam was the world’s busiest port. The Wikipedia entry for Rotterdam indicates that the petrochemical industry is especially important in ROtterdam. In 2018 almost 30,000 vessels arrived at Rotterdam.

We see from [Index Mundi](https://www.indexmundi.com/agriculture/?commodity=palm-oil&graph=imports) that the EU-27 is the world’s third largest importer of palm oil behind India and China. We are interested to know what fraction of these European imports arrive at the Port of Rotterdam.

We can estimate receipts of palm oil at Rotterdam by using the UN ComTrade database to identify imports of palm oil by the Netherlands. Lets consider imports from the world as a whole and from the top palm oil exporters, which include Indonesia and Malaysia.

# Types of Palm Oil

We will mostly focus on crude palm oil (CPO). Palm olein is a liquid refined form of palm oil used in cooking and baking. Stearin is a hard refined form of palm oil for food applications. To use the ComTrade database we need to know the harmonized system (HS) commodity code. The four-digit HS code for palm oil is 1511: “Palm oil and its fractions, whether or not refined, but not chemically modified.”. We wil use a pair of six digit codes:

* 151110: “Crude palm oil”
* 151190: “Palm oil or fractions simply refined”

# UN ComTrade Database

Read about the UN ComTrade database [here](https://unstats.un.org/unsd/tradekb/knowledgebase/50075/what-is-un-comtrade):

Answer the following TRUE/FaLSE questions.

1. The data within the UN ComTrade database is collected via UN surveys.
2. The UN ComTrade database contains over 40 billion records.

To use the database you need to know the country and commodity codes. The country code is in column A of a [UN Country Code Excel File](https://unstats.un.org/unsd/tradekb/Knowledgebase/50377/Comtrade-Country-Code-and-Name). The commodity code can be looked up [here](https://comtrade.un.org/db/mr/rfCommoditiesList.aspx)

# Manual Query of UN ComTrade

Take a guess: which country exports more crude palm oil (CPO) to Canada: Malaysia or the USA? To determine if you answered correctly access the [UN ComTrade](https://comtrade.un.org/data/) database and manually generate data for 2019 and 2020. Include Canadian imports from the world as a whole to serve as a benchmark.

# Using R to Query UN ComTrade

The UN has written a get.ComTrade function in R to enable users to extract data using R. The queries are limited in size unless you have premium access, You can view the function in the section titled “A user-defined function to extract data from the UN Comtrade API” by clicking [here](https://comtrade.un.org/data/Doc/api/ex/r). This is a complex function and you are not responsible for knowing how it works.

This function has been saved to “ComTrade\_Function.R”. Ensure that this file resides in your working directory so that it is available for use.

Need the following country codes:

* Netherlands 528 (reporter)
* Indonesia 360 (partner)
* Malaysia 458 (partner)
* World 0 (partner)

In addition, to

* 151110: “Crude palm oil”
* 151190: “Palm oil or fractions simply refined”

we will work with

* 3826: “Biodiesel”

# Parameters for get.ComTrade function

The get.ComTrade function takes a large number of parameters. If you do not supply a particular parameter value then R will use the default value. The parameters we will use are as follows

* r ==> country code for reporting country (i.e., r=“528” for Netherlands)
* p ==> country codes for partner countries (e.g.,p=“360” for Indonesia )
* ps ==> year (e.g., ps=“2020” for the year 2020)
* rg ==> type (e.g., rg=1 for imports)
* cc ==> commodity (e.g., cc=“151110” for crude palm oil)
* fmt ==> data save format (e.g., fmt=“csv” for comma separated values)

In our case the function will look as follows:

get.Comtrade(r=“528”, p=“0,170,320,360, 458”, ps=“2020”, rg=1, cc=“151110”, fmt=“csv”)

# Running the code

After we retrieve the data we want it the form of a standard R data frame. The following code gives the desired results:

source("ComTrade\_Function.R")  
library(rjson)

## Warning: package 'rjson' was built under R version 4.0.3

A <- get.Comtrade(r="528", p="0, 170, 320, 360, 458", ps="2020",   
 rg=1, cc="151110")  
dfA <- as.data.frame(A$data)

If you print this data frame to the screen you will see it is long and rather messy. Let’s eliminate the columns which are not of interest, and give shorter and more descriptive titles to the remaining columns and report trade in tonnes rather than kilograms. Finally, we will sort the data from lowest to highest imports.

Imp.A <- dfA$rtTitle  
Exp.A <- dfA$ptTitle  
Comm.A <- dfA$cmdCode  
Year.A <- dfA$yr  
Type.A <- dfA$rgDesc  
Unit.A <- "tonnes"  
Weight.A <- as.numeric(dfA$TradeQuantity)/1000  
  
dfAsum <- as.data.frame(cbind(Imp.A, Exp.A, Comm.A, Year.A, Type.A, Unit.A, Weight.A))  
dfAsum$Weight.A <- as.numeric(dfAsum$Weight.A)  
   
A\_final <-dfAsum[order(dfAsum$Weight.A),]  
A\_final

## Imp.A Exp.A Comm.A Year.A Type.A Unit.A Weight.A  
## 3 Netherlands Guatemala 151110 2020 Import tonnes 126951.0  
## 2 Netherlands Colombia 151110 2020 Import tonnes 210570.2  
## 4 Netherlands Indonesia 151110 2020 Import tonnes 360133.7  
## 5 Netherlands Malaysia 151110 2020 Import tonnes 673268.0  
## 1 Netherlands World 151110 2020 Import tonnes 2180656.6

# Import Shares

Let’s create a new column in the dt\_final dataframe which shows the import shares of Indonesia and Malaysia. Specifically, we will create a new variable call *world*, which is total world imports at the bottom right of dt\_final.To create the share column we will divide each value in the Weight column bu the ^world\* variable. The code is as follows:

world <- A\_final[5,c("Weight.A")]  
A\_final$share <- A\_final$Weight.A/world  
A\_final

## Imp.A Exp.A Comm.A Year.A Type.A Unit.A Weight.A share  
## 3 Netherlands Guatemala 151110 2020 Import tonnes 126951.0 0.05821687  
## 2 Netherlands Colombia 151110 2020 Import tonnes 210570.2 0.09656275  
## 4 Netherlands Indonesia 151110 2020 Import tonnes 360133.7 0.16514921  
## 5 Netherlands Malaysia 151110 2020 Import tonnes 673268.0 0.30874554  
## 1 Netherlands World 151110 2020 Import tonnes 2180656.6 1.00000000

The revised dt\_final table shows that Malaysia supplies almost one third of Netherland’s imports, and Malaysia and Indonesia combined supply almost one half of Netherland imports. We do not have information about the market structure and in particular if these large market shares are likely to result in non-competitive pricing.

# Netherland Processing

The Netherlands is a small country and certainly does not need large volumes of palm oil for its own domestic consumption. Rather, palm oil shipped from Malaysia and Indonesia to European countries such as Germany and Switzerland will by imported through the Netherlands. This means we should expect relatively large volums of crude palm oil exports from the Netherlands. Germany is a large country and so we might expect particularly large volumes of crude palm oil from the Netherlands to Germany.

The Netherlands is also a large-scale processor of petrochemicals and so it is reasonable to assume that the Netherlands is also a large-scale processor of crude palm oil into refined palm olein (liquid) and palm stearin (solid). The UN ComTrade database combines both of these into one HS category called “Palm oil or fractions simply refined” (151190). We also expect significant refining of crude palm oil into biodiesel, which as HS code 3826.

# Netherland Trade by Commodity Type

two separate tables. The first is 2020 Netherland imports of crude palm oil (151110), refined palm oil (151190) and biodiesel (3826) for the world as a whole. The second is the same except it is for Netherland exports. Later we will use these values to isolate Nethland’s domestic consumption of palm oil.

# Code for Imports and Exports

B <- get.Comtrade(r="528", p="0", ps="2020",   
 rg="1,2", cc="151110,151190,3826")  
dfB <- as.data.frame(B$data)  
  
Imp.B <- dfB$rtTitle  
Exp.B <- dfB$ptTitle  
Comm.B <- dfB$cmdCode  
Year.B <- dfB$yr  
Type.B <- dfB$rgDesc  
Unit.B <- "tonnes"  
Weight.B <- as.numeric(dfB$TradeQuantity)/1000  
  
dfBsum <- as.data.frame(cbind(Imp.B, Exp.B, Comm.B, Year.B, Type.B, Unit.B, Weight.B))  
dfBsum$Weight.B <- as.numeric(dfBsum$Weight.B)  
dfBsum

## Imp.B Exp.B Comm.B Year.B Type.B Unit.B Weight.B  
## 1 Netherlands World 151110 2020 Import tonnes 2180656.6  
## 2 Netherlands World 151110 2020 Export tonnes 224945.9  
## 3 Netherlands World 151190 2020 Import tonnes 338144.8  
## 4 Netherlands World 151190 2020 Export tonnes 956350.8  
## 5 Netherlands World 3826 2020 Import tonnes 3980055.6  
## 6 Netherlands World 3826 2020 Export tonnes 4733541.2

# Exercise 2

In R you can select the ith row and jth column of a dataframe X using X[i,c(j)] or X[i,c(“col\_name”)]. This means that if I wanted to know what was in the second row and third column of the dfBsum dataframe (i.e., the commodity code) I would use one of the following:

dfBsum[2,c(3)]

## [1] "151110"

dfBsum[2,c("Comm.B")]

## [1] "151110"

It is approximately true that the production of refined palm oil and biodiesel from crude palm oil preserves the original weight. It is also true that the Netherlands does not produce any crude palm oil This means that the difference between total imports and total exports of the three products is approximately equal to the weight of the three commodities which were consumed in the Netherlands. Your job is to calculate the weight of this domestic consumption.

# Exercise 2 Solution

import\_tot <- dfBsum[1,c("Weight.B")] + dfBsum[3,c("Weight.B")] + dfBsum[5,c("Weight.B")]  
export\_tot <- dfBsum[2,c("Weight.B")] + dfBsum[4,c("Weight.B")] + dfBsum[6,c("Weight.B")]  
import\_tot

## [1] 6498857

export\_tot

## [1] 5914838

domestic <- import\_tot - export\_tot  
domestic

## [1] 584019.2

# The Law of One Price

For the remainder of this module we will focus on palm oil prices. The spatial law-of-one-price (LOP) tells us that for a homogenous good such as crude palm oil, in the long run:

1. If two regions are trading with each other then the price difference must equal the transportation cost.
2. If two regions are not trading then the price difference must be less than the transportation cost.

If the price difference is greater than the transportation cost than arbitrage will take place and prices will be driven together until the LOP holds. Specifically, the excess demand in the low price region will drive up the price and the excess supply in the high price region will drive down the price.

# Local Versus U.S. Dollar Prices

Like many commodities, crude palm oil (CPO) trades in U.S. dollars in international markets. If you are a trader who deals only in U.S. dollars, then the price of CPO in the local currency (e.g. Malaysian Ringgit (RM) for the case of Malaysia) and the corresponding exchange rate is not important. If instead you operated in the local market, or in simultaneously in the local and international market, then you must consider the effect of the exchange rate on the local price.

To see how this works, monthly average CPO prices were collected FOB Malaysia (expressed in RM) and FOB Rotterdam (expressed in US$). The monthly prices were aggregated from daily prices, found [here](https://bepi.mpob.gov.my/admin2/price_local_daily_view_cpo_msia.php?more=Y&jenis=1Y&tahun=2019) for the Malaysian prices and [here](https://www.investing.com/commodities/crude-palm-oil-cif-rotterdam-futures-historical-data) for the Rotterdam prices. The average Ringgit-US dollar exchange rate was obtained from the [U.S. Federal Reserve](https://fred.stlouisfed.org/series/DEXMAUS).

The code below reads in the data and converts the data column into R format. The first six months of the sample, which runs from January of 2012 to June of 2016, is displayed. The columnn titled “balticDry” will be explained below.

graphics.off() # close all plots  
data <- read.csv(file="./Data/palm\_data.csv", header=TRUE, sep=",", stringsAsFactors = FALSE)  
data$date <- seq(as.Date("2012/1/1"), by = "month", length.out = 114)  
head(data)

## date rotterdam malaysia exchange balticDry cornShip  
## 1 2012-01-01 1062.5 3182.5 3.1092 680 51.75  
## 2 2012-02-01 1135.0 3108.5 3.0220 750 49.00  
## 3 2012-03-01 1175.0 3278.0 3.0444 934 49.80  
## 4 2012-04-01 1180.0 3480.5 3.0586 1155 52.75  
## 5 2012-05-01 1030.0 3188.5 3.0978 923 50.75  
## 6 2012-06-01 1015.0 2955.5 3.1783 1004 52.32

The following code generates side-by-side graphs of the two price series. To run this code you need to have the ggplot2 and gridExtra packages installed.

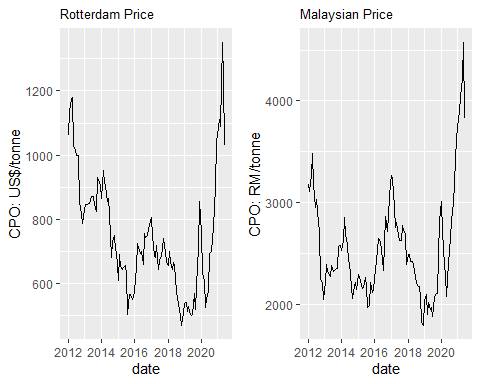
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.5

library(gridExtra)

## Warning: package 'gridExtra' was built under R version 4.0.5

plot\_rotterdam <- ggplot(data, aes(x = date, y = rotterdam)) +   
 geom\_line() +ggtitle("Rotterdam Price")+labs(y= "CPO: US$/tonne")+ theme(plot.title = element\_text(size=10))  
plot\_malaysia <- ggplot(data, aes(x = date, y = malaysia)) +   
 geom\_line() +ggtitle("Malaysian Price")+labs(y= "CPO: RM/tonne")+ theme(plot.title = element\_text(size=10))  
grid.arrange(plot\_rotterdam, plot\_malaysia, ncol = 2)

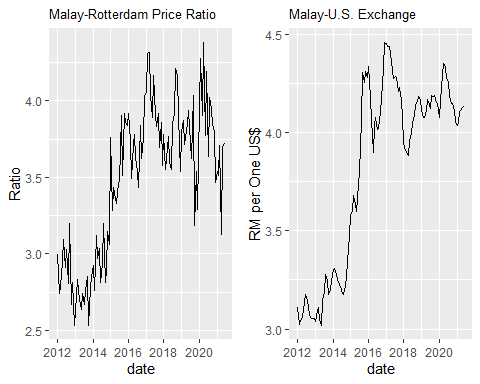


# Isolating the Exchange Rate Impacts

The pair of graphs reveal that the Rotterdam price of CPO in US dollars and the Malaysian price of CPU in Ringgits move closely together but they are far from being perfectly positively correlated. Suppose the cost of transporting CPO from Malaysia to Rotterdam is zero, and the cost of instantly arbitraging the two markets is also zero. In that case we should expect $P\_R^{US} = P\_M^{RM}/E $ where is the Rotterdam price expressed in U.S. dollars, is the Malaysian price expressed in Malaysian Ringgits and is the exchange rate, expressed as the number of Ringgits required to purchase one U.S. dollar.

If we rearrange this expression as we see that in this hypothetical world, the price ratio will be equal to the exchange rate. The following code generates graphs of the ratio of the monthly average prices and the monthly average exchange rate.

data$priceRatio <- data$malaysia/data$rotterdam  
plot\_PriceRatio <- ggplot(data, aes(x = date, y = priceRatio)) +   
 geom\_line() +ggtitle("Malay-Rotterdam Price Ratio")+labs(y= "Ratio")+ theme(plot.title = element\_text(size=10))  
plot\_Exchange <- ggplot(data, aes(x = date, y = exchange)) +   
 geom\_line() +ggtitle("Malay-U.S. Exchange")+labs(y= "RM per One US$")+ theme(plot.title = element\_text(size=10))  
grid.arrange(plot\_PriceRatio, plot\_Exchange, ncol = 2)



The left graph shows that the RM price of CPO in Malaysia strengthened relative to the U.S. dollar price of CPO in Rotterdam over the years 2013 to 2017. This strengthening occurred because over this time period the Ringgit weakened against the U.S. dollar. In other words, Malaysian exporters are being paid U.S. dollars for their CPO shipments and when these U.S. dollars are converted back to Ringgits the amount received is higher.

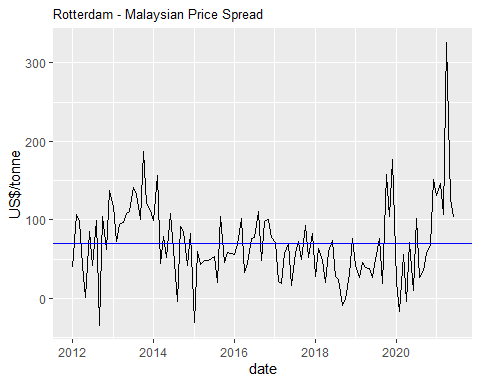
The two plots are not identical, which means that fails to hold exactly. This failure occurs for three reasons: (1) transportation costs are positive; (2) transportation costs are changing over time; and (3) arbitrage is relatively slow and so what we are observing is the short run and not the long run.

# Analysis of Price Spread

We know from the first part of this module that Malaysia is constantly shipping CPO to Rotterdam. The spatial LOP tells us that in the long run and when all prices are expressed in U.S. dollars, then the Rotterdam price should equal the Malaysian price plus the cost of transporting the CPO between these two regions. In other words, .

The following code calculates the price spread by dividing the Malaysian price by the exchange rate and then subtracting the resulting value from the Rotterdam price. The calculated spread is then plotted.

data$spread <- data$rotterdam -data$malaysia/data$exchange  
plot\_Spread <- ggplot(data, aes(x = date, y = spread)) +   
 geom\_line() +ggtitle("Rotterdam - Malaysian Price Spread")+labs(y= "US$/tonne")+ theme(plot.title = element\_text(size=10))+geom\_hline(yintercept = mean(data$spread), color="blue")  
plot\_Spread



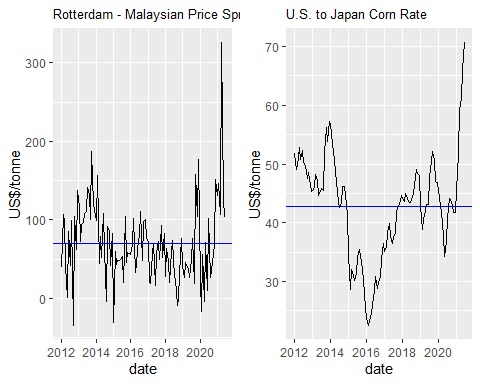
This graph makes it clear that the price spread (after for the exchange rate) is quite variable. One possibility is that the transportation cost is relatively stable, and it is lack of arbitrage in the short run which is causing the variable price spread. A second possibility is that arbitrage is effective but it is highly variable transportation costs which are responsible for the variable price spreads.

# Analysis of Transportation Cost

Begin by reading (or at least skimming) the [report](https://www.agflow.com/commodity-trading-101/a-guide-to-chemical-tanker-types-for-palm-and-edible-oils/) on ocean transport of crude palm oil. Use [Sea-distances.org](https://sea-distances.org/) to determine that the shortest route from Malaysia to Rotterdam is approximately 8000 nautical miles through the Suez Canal. This compares to approximately 9500 nautical miles from the U.S. Gulf Coast to Japan through the Panama Canal. There is no data on monthly shipping costs for CPO from Malaysia to Rotterdam. As an alternative monthly shipping costs for corn from the U.S. Gulf Coast to Japan will be used as a proxy. The data comes from Figure 17 in a [USDA](https://www.ams.usda.gov/services/transportation-analysis/gtr-datasets) report.

The corn freight rates are in the column titled “cornShip” of R’s “data” dataframe. The following code generates a pair of the graphs: the previous price spread graph on the left and the corn freight rate graph on the right.

plot\_Corn <- ggplot(data, aes(x = date, y = cornShip)) +   
 geom\_line() +ggtitle("U.S. to Japan Corn Rate")+labs(y= "US$/tonne")+ theme(plot.title = element\_text(size=10))+geom\_hline(yintercept = mean(data$cornShip), color="blue")  
   
grid.arrange(plot\_Spread, plot\_Corn, ncol = 2)



The pair of graphs show the average cost of shipping corn from the U.S. to Japan (horizontal blue line in right graph) is significantly lower than the average Rotterdam - Malaysia price spread (horizontal blue line in left graph). In the spatial LOP tells us that the long run average cost of shipping the CPO is approximately equal to the long run price spread. This makes it clear that despite a similar voyage distances, the cost of shipping corn is significantly less than the cost of shipping crude palm oil. This is not surprising given the specialized (heated) tanks which are required for the shipment of crude palm oil.

# Regression of Spread on Corn Transport Cost

The previous pair of graphs reveal a weak correlation between the monthly cost of shipping corn from the U.S. to Japan and the Malaysia - Rotterdam price spread. We can be more specific by regressing the CPO price spread on the corn shipping costs. The following code generates the regression results.

lmCorn <- lm(data$spread~data$cornShip, data = data)  
summary(lmCorn)

##   
## Call:  
## lm(formula = data$spread ~ data$cornShip, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -112.696 -32.251 0.094 22.406 222.002   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -8.7997 21.7096 -0.405 0.686004   
## data$cornShip 1.8387 0.4968 3.701 0.000335 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 46.91 on 112 degrees of freedom  
## Multiple R-squared: 0.109, Adjusted R-squared: 0.101   
## F-statistic: 13.7 on 1 and 112 DF, p-value: 0.0003345

The regression results show that the corn transportation cost is positive and highly statistically significant. This outcome supports the spatial LOP hypothesis that a higher transportation cost must lead to a higher price spread in the long run. However, the of 0.109 is quite low. Indeed, variation in the cost of shipping corn only explains about 10 percent of the variation in the price spread. We have controlled for exchange rate movements and a sensible measure of the cost of shipping but yet about 90 percent of the variation in the price spread remains unexplained.

This is a good lesson because it should remind us that concepts which are applicable in the long run may not be particularly helpful in the short run. If you were a trader whose job was to purchase raw palm oil in Malaysia, transport it to Rotterdam and then resell it in the European market, you will clearly want to hedge the price risk and not rely on the long-run spatial LOP concept!

# Optional R Exercise

Calculate the mean value of data$spread and the mean value of data$cornShip. Subtract the latter from the former and call the resulting value “adjust”. Create a new column in the “data” dataframe by adding “adjust” to each value in the data$cornShip column. Now plot on the SAME graph this new value along with data$spread. Doing this will allow you to have a much better visualization of the extent that the price spread and the cost of shipping corn move together over time. See Krisha if you need some help!