Module 1D: Pricing Integration and Fertilizer Case Study

## Learning Outcomes

* Describe the concept of pricing integration for spatially-separated markets, as it relates to commodity pricing in Sub-Saharan Africa.
* Construct a full information econometric model of pricing integration with the goal of quantifying pricing integration.
* Show that a standard cross-sectional model of pricing integration suffers from omitted variable bias.
* Describe why the fixed effects panel model cannot be used to eliminate the omitted variable bias in the pricing integration model.
* Review the estimation results of commodity pricing in central and eastern Africa: Brenton et al. (2014).
* Use R to create a bubble map of northwest Africa and of southeast Africa, both of which show regional fertilizer prices as different sized bubbles. .

## Rationale for this Module

In Modules 1A through 1C we focused on the spatial LOP in the context of international trade. In this module we will instead focus on the spatial LOP within a country (or between countries within a land mass) where roads are the primary form of commodity transportation.

In a developed country such as Canada where most roads are high quality, road length and road quality are not particularly important when trying to understanding regional price differences. For example, in the grain producing region of Western Canada, grain is transported by semi-truck for hundreds of kilometers to competing inland grain terminals which are owned by the “ABCD” grain multinationals. The additional cost for transporting the grain an additional 100 km is small in comparison to the value of the product. Consequently, the price differences between pairs of delivery points is generally quite small.

The situation is very different in developing countries such as those in Africa. Cities and villages are connected by roads, some of which are poor in quality. As well, agricultural commodities are often transported in small vehicles rather then cost efficient semi-trucks. This means that the cost of moving the commodity from a surplus region to a deficit region can be sizeable relative to the value of the commodity. In other words, transportation costs are often a dominant feature in the intraregional spatial pricing of commodities.

This module is divided into two parts. In Part A we will review a 2014 [World Bank](https://openknowledge.worldbank.org/handle/10986/19340) paper written by Brenton et al., which reports the econometric results from the analysis of spatially-separated commodity prices (maize, rice and sorghum) in central and eastern Africa. The authors propose that it is the combination of distance-dependent transportation costs and fixed administrative transaction costs which explain the intraregional pricing differences. A road quality variable is included to account for transportation cost heterogeneity. When the trade route involves crossing a national border (e.g., Malawi and Mozambique) the authors allow for an additional fixed administrative cost.

In Part B of this module we focus on the spatial distribution of fertilizer prices in Sub-Saharan Africa. A 2020 study by [Cedrez et al](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0227764) made available their full data set, which includes fertilizer prices at the town level, geocoded by longitude and latitude. The focus of Part B is using R’s mapping functions to create a visualization of the spatial distribution of fertilizer prices.

Of particular interest in Part B is the extent that fertilizer prices are significantly higher for inland locations versus more coastal locations. This is very relevant for fertilizer, which is mainly imported rather than being produced domestically, and which is costly to transport to inland locations. According to the [Break Through](https://thebreakthrough.org/issues/food/expand-fertilizer-manufacturing-across-africa) blogsite, in 2016 about 80 percent of nitrogen fertilizer was imported by African countries despite sizable deposits of that natural gas which is required to produce nitrogen fertilizer.

# Part A: Commodity Market Integration in Central and Eastern Africa

## The Concept of Market Integration

The 2014 World Bank paper by Brenton et al. that we will review is titled “Food Prices, Road Infrastructure, and Market Integration in Central and Eastern Africa”, This title contains the phrase “market integration”. What exactly does that mean? According to Wikipedia:

“Market integration occurs when prices among different locations or related goods follow similar patterns over a long period of time. Groups of goods often move proportionally to each other and when this relation is very clear among different markets it is said that the markets are integrated. Thus, market integration is an indicator that explains how much different markets are related to each other.”

This definition of market integration focuses on the time series properties of prices. However it follows that if markets are integrated then if one were to take a snapshot of the prices at two different points in time the absolute prices across the pair of snapshots may be quite different but the relative prices would be quite similar. The Brenton et al. (2014) study uses a snapshot approach rather than a time series approach to examine pricing integration.

## Importance of Market Integration

Market integration is crucial for food security in developing nations such as those in Sub-Saharan Africa. Nations in this region often face food shortages due to drought and other climate-related crop failures. If markets are well integrated, product will quickly move from commodity surplus regions to commodity deficit regions, and in doing so the impact of the supply shock will be lessened. More generally, price shocks will be dampened in well integrated markets and this dampening can significantly improve food security outcomes. According to Brenton et al. (2014) well integrated markets will also improve the speed of agricultural transformation and economic growth. Similarly, social safety net programs are more effective and efficient when markets are well integrated.

According to Brenton et al. (2014):

“In well-integrated and well-functioning markets, price differences should be arbitraged away through intra-regional trade so that price differences between towns should be bounded by transportation costs between the markets. Yet, policy regulations (e.g., export bans), border barriers, and other impediments to trade may prevent arbitrage and lead to large sustained price differences between locations.”

The analysis of Brenton et al. (2014) identify impediments to trade which includes road quality and “thick” borders. In the international trade literature, a “thick” border refers to a situation where trading within a region with no administrative borders (e.g., British Columbia and Ontario) is less costly than distance-adjusted trade within a region that does involve an administrative border (e.g., British Columbia and California).

## Theoretical Framework

The theoretical framework of Brenton et al. (2014) is consistent with the spatial law-of-one-price framework that we are utilizing in FRE 501. Following Baulch (1997), the authors indicate that:

1. “When transfer costs equal the inter-market price differential and there are no impediments to trade between markets, trade will cause prices in the two markets to move on a one-for-one basis”

and

1. “When transfer costs exceed the … difference of prices, trade will not occur… In that case, the price difference between two locations is lower than the transaction cost”.

Brenton et al. (2014) note that these two conditions are consistent with “competitive spatial price equilibrium and market integration.”

Brenton et al. (2014) describe lack of market integration in a way that we have not made so precise in the previous FRE 501 case studies. In particular, they indicate that

1. “When price differences exceed transfer costs, the spatial arbitrage conditions are violated whether or not trade occurs.”

Brenton et al. (2014) suggest a number of reasons why there may be a violation of the spatial arbitrage conditions. For example, inter-market price differences may reflect constraints other than pure transfer costs such as government controls on commodity flows, transportation bottlenecks and non-competitive pricing. The time cost associated with shipment delays are not discussed by Brenton et al. (2014) but these implicit costs are also important determinants of price differentials.

In our study of palm oil we saw considerable variations in the Malaysia - Rotterdam price gap over time. Unfortunately, it was not possible to identify when a particularly large price gap belonged to category (1) and its large size was due to an unusually high transportation cost, and when it belonged to category (3) and its large size was due to a violation of the spatial arbitrage condition.

## Quantifying Market Integration

It should be clear from the above discussion that if spatially-separated markets for a particular commodity are fully integrated then the spatial law-of-one-price (LOP) holds exactly. More generally, market integration can be viewed as a measure of the extent that the spatial LOP holds, or, alternatively, as a measure of the extent that spatial arbitrage conditions fail to hold.

It is worthwhile asking if market integration can be quantified. Brenton et al. indicate that volume of trade is often used as a measure of market integration, with higher trade volumes indicating a higher level of market integration. This may be true but as we know from our study of palm oil and tomatoes, the price correlation across regions may not be particularly strong despite high annual trade volumes.

A better approach to quantifying pricing integration is to estimate a model where pricing differentials across pairs of markets is the dependent variable and the full cost of transporting the commodity between markets is the explanatory variable. The goodness of fit of the regression and the level of significance of the explanatory variables will serve as indicators of the strength of pricing integration.

Let denote the absolute price difference between a pair of trading markets, and let denote the full cost of moving a unit of the product between the two markets. In this specification, the cost of moving the product from market to market is assume to equal the cost of moving the product from market to market . This allows us to use the absolute price difference as the measure of the price gap rather than keeping track of which market is selling and which market is buying.

If we had high quality data on and then for those markets which are experiencing a positive level of trade we could estimate the following equation: . Perfect market integration and the spatial LOP requires and . A positive value for and a statistically significant value of less than one would indicate less than perfect market integration. Finally, a positive value for and a value for which is not statistically significant from zero would indicate no market integration.

To implement the previous approach it is necessary to have data which indicates if the trade flow between a particular pair of markets is positive or zero. Is there a way to estimate the reqression equation in the previous paragraph without having this type of indicator data? In a 2001 paper by Goodwin and Piggott titled “Spatial Market Integration in the Presence of Threshold Effects” the authors use threshold regression to allow for a “neutral” pricing band. Within the pricing band a pair of prices from different markets can fluctuate without attracting arbitrage because the price difference is less than the cost of arbitraging the price gap. If the pair of prices move out of this pricing band then the regression equation recognizes that arbitrage will occur, in which case an equation similar to will be estimated.

An important shortcoming of the Brenton et al. (2014) analysis is that they do not use threshold regression to account for the fact that trade between many markets is zero since the price difference is much less than the transportation cost. The data set involves 150 towns, which means there are over 10,000 unique town pairings. It must be the case that a very large majority of these 10,000 town pairings involve zero trade. Threshold regression or an equivalent mechanism to distinguish between trading and non trading towns is needed to accurately test for the level of market integration.

Can we say anything about the bias in the model’s estimated parameters which is likely to arise because of the threshold problem described above? By including data that involves relatively small price differences and relatively large transportation costs it necessarily follows that the relationship between transportation costs and price differences will be biased downward.

To better understand this bias the first 10 hypothetical data points in the table below represent situations where there is positive trade (the price gap is equal to the transportation cost plus a random error) and the last 10 data points represent situations where there is zero trade (the price gap is consistently less than the transportation cost). If we regress the price difference on the transportation cost with just the first 10 data points the estimated intercept is 0.738 and the estimated slope (statistically significant) is 0.789. In contrast, if we include all 20 data points then the estimated intercept is 1.483 and the estimated slope (statistically significant) is 0.565. This confirms that mixing price and transportation cost pairs that do not involve active trade with price and transportation cost pairs that do involve active trade will necessarily create a downward bias in the estimate of the strength of the relationship between distance and the pricing gap.

|  |  |  |
| --- | --- | --- |
| Cost | Price Gap | Trade? |
| 3 | 3.5 | Yes |
| 2.75 | 3.1 | Yes |
| 4.2 | 4.3 | Yes |
| 4.05 | 3.75 | Yes |
| 3.4 | 2.95 | Yes |
| 4.4 | 4.1 | Yes |
| 3.05 | 2.6 | Yes |
| 3.6 | 3.65 | Yes |
| 3.35 | 3.55 | Yes |
| 3.9 | 4.05 | Yes |
| 5.5 | 4.4 | No |
| 5.05 | 3.95 | No |
| 4.8 | 4.6 | No |
| 5.15 | 3.75 | No |
| 4.75 | 3.7 | No |
| 5.4 | 5 | No |
| 6.05 | 4.9 | No |
| 5.15 | 4.85 | No |
| 4.85 | 3.6 | No |
| 4.95 | 4.75 | No |

## Alternative Estimation Approach Due to Lack of Data

Let’s follow the author’s lead by ignoring the threshold problem described above. Later when we interpret the authors’ results will revisit the issue of bias.

It is generally not possible to obtain accurate data for the full cost of transportation, , because of non-observable fixed administrative costs which are important determinants of the size of the price gap. Keep in mind that the fixed administrative cost includes explicit and implicit time costs and possibly the cost of bribes and other implicit transaction costs.

In light of this data deficiency, an alternative model is where is the distance between the two trading markets measured in kilometers. If all trading pairs in the data set have the same fixed administrative fee, then the estimate of the coefficient in the above equation would be our estimate of this common fixed administrative fee. In this case, assuming perfect market integration (i.e., exact spatial LOP) and the same distance-related cost for all trading markets, then our estimate of the coefficient would be the additional cost of transporting the commodity one additional kilometer. If we knew that the actual cost of transporting the commodity one additional kilometer was equal to then we could test the strength of market integration by determining the extent that our estimate of lies below .

It is unlikely that fixed administrative costs are the same for all trading pairs. Moreover, it is likely that trading pairs which have a greater distance between them have a higher fixed administrative fee. This is because language and cultural differnces tend to increase with distance, as do the number of adminstrative borders and thus the number of administrative fees (and possibly bribes) which must be paid. The fact that fixed administrative fees may be positively corrlated with distance is a big problem because it means that the coefficient will be biased upward. This outcome is classic omitted variable bias which emerges with unobserved heterogeneity. The upward bias means that the estimate of may be pushed closer to the true unit cost , which means we may incorrectly conclude that the market has strong integration when in fact it is weak or non-existant.

## Panel Data to the Rescue (or not!)

As you may already know, and certainly will know if you take FRE 529 next semester, estimating a panel model with fixed effects can eliminate the omitted variable bias. For the current situation, a fixed effects model is essentially equivalent to adding a dummy variable for every pair of trading markets. This will allow markets which are separated by a longer distance to have a higher fixed administrative fee. Doing this will ensure that the estimate is not biased upward due to the positive association between distance between markets and the fixed administrative cost of shipping.

The World Bank data set is in panel format because it consists of monthly observations of regional price differences. Letting denote the fixed effect, the panel data model can be expressed as

Unfortunately, the fixed effect model as specified above does not allow the slope coefficient to be estimated. This is because the fixed effect and the distance variable are indistinguishable. That is, will be treated as a single variable. The general rule with fixed effect estimation with panel data is that only variables which vary over time can be included in addition to the fixed effect variable. For example, in a model of international trade where tariffs have been gradually declining over time it would be possible to use a fixed effects model to estimate how the price gap depends on the tariff level. Distance between markets does not change over time and so it is not possible to examine the impact of distance on the price gap when using the fixed effects panel model.

It would not be very interesting or useful to estimate the model without having an estimate of . Thus, the fixed effect panel model cannot be used in the current study. The next best option is to estimate the model as a pooled cross section - time series model without fixed effects but to include control variables to reduce the bias. A good control will reduce the positive association between the distance variable and the fixed administrative fee variable. For example, a dummy variable which indicates if the trade involves crossing a national border is expected to reduce this positive association. As well, a common-language dummy variable is also expected to reduce the positive association between the size of the fixed administrative fee and the distance between the two trading partners.

## Data

The prices analyzed in the 2014 Brenton et al. (World Bank) study come from 150 towns in the following 13 countries:

* Burundi
* Djibouti
* DRC
* Ethiopia
* Kenyna
* Mosambique
* Malawi
* Rwanda
* Sudan
* Somalia
* Tanzania
* Uganda
* Zambia

The commodities included in the analysis include maize, rice and sorghum. The monthly data was collected over the period May 2008 to September 2009,

For one commodity, if all 150 towns have a recorded price then we can determine the number of price pairs which exist in the data. We need to use the formula for a “combination” (i.e., 150 choose 2):

Substituting in and gives:

The actual number of observations are 10,232 for maize, 10,832 for rice and 9,1118 for sorghum (see Table 2). In comparison to the 11,175 maximum number of pairs it can be seen that most of the towns are included in the sample for each of the three commodities.

## Log of Price Ratio

In their econometric analysis, Brenton et al. (2014) use the following expression for their dependent variable:

In this expression, is the price of commodity at location in month . It should be clear that the authors are using the absolute value of the log of the price ratio as the dependent variable. The absolute value is used because they don’t want to keep track of which is the low-priced selling market, and which is the high-priced buying market.

Some of you will recognize that the log of the price ratio is equal to the percent difference in the two prices assuming that the difference is not excessively large. To understand this, recall from finance that with continuous compounding the per-period rate of interest, , goes to zero and the number of compounding periods goes to infinity. In this extreme case . Take the natural log of both sides to get $r = ln(1+r) $. Since is the percent return over one compounding period we can write this expression as

In this expression is the price at the beginning of the compounding period and is the price at the end of the compounding period. Thus, the log of the price ratio is approximately equal to the percent difference in the pair of prices assuming that the difference is relatively small.

## Econometric Equation

The econometric equation of Brenton et al. (2014) (with some variable names abbreviated) can be expressed as

In the right side of this expression, and are town-product dummy interaction terms, and is a product-month interaction term. The first pair of terms is intended to capture town specific reasons why the price of a particularly commodity might be particularly high or low. For example, independent of which other towns it trades with, a particular town might have an above average price for a particular commodity due to above average local product quality or a connection to an export market that other towns do not have. The product-month dummy interaction is intended to control for seasonality in prices and price trends over time.

Regarding the other variables on the right side of the above equation is the distance between towns and , and is the quality of the roads which connect towns and . Finally, is a dummy variable which indicates if the trade route between the two town crosses a national border.

## Descriptive Statistics

There is considerable cross-country variation in the mean travel distance within each country. For example, the mean distance is 154.5 km in Rwanda, and 604.2 km in Sudan. The mean within-country distance across all 13 countries is 206.2 km. The mean between-country distance is 370.2 km. Regarding road quality, the mean value is 81.5 percent within the 13 countries and 78.3 percent between countries. The variation in road quality is much less than the variation in the mean distance. The average percent price difference is 20.0 percent within the 13 countries, and 33.0 percent between countries.These rather large values for the average price gap is our first indication that pricing integration may be problematic.

## Intra-Country Results

The table below shows the econometric results for model (2) in Table 3 of Brenton et al. (2014).

|  |  |  |
| --- | --- | --- |
|  | Abs % Price Diff |  |
| ln(D\_ij) | 0.0618\*\*\* | (0.0061) |
| Q\_ij | -0.0407\*\* | (0.0140) |
| Constant | 0.109 | (0.0898) |

Notice that the distance and road quality variables are both highly statistically significant, and both have the anticipated signs.

Consider first the impact of distance. It is useful to hold constant and allow to increase with distance between the two markets. Similar results would be obtained in the more general case where both and adjusted with a longer distance. If we totally differentiate the previous econometric equation with respect to and we obtain:

Rearranging this expression gives us the elasticity of price with respect to distance:

This elasticity is quite small. The mean within-country distance between markets is approximately 200 km. Thus, an additional 50 km of distance represents 25 percent relative to the average. Using the above formula, we can infer that moving from 200 km to 250 km results in an increase in the price gap equal to 25\* 0.0618 = 1.54 percent on average. Even a doubling of the distance from 200 km to 400 km would result in a price differential increase of 100\* 0.0618 = 6.18 percent. This suggests that at most distance contributes about on third of the 20 percent price gap. The remaining two thirds can be attributed to variations in road quality and the administrative fixed fee.

We can also calculate an elasticity for road quality. In this case the elasticity is the percent change in given a one percentage point increase in road quality:

Here we see that if road quality improves from the mean value of 81.5 percent to 90 percent, the price gap will decrease on average by -0.0407\*8.5 = 0.346 percent. Similar to distance, the variations in road quality do not appear to explain much of the variation in price differences across markets.

## Between Country Results

The between country results, which are given by Model 3 in Brenton et al. (2014), are shown in the table below. Similar to the previous intra-country results, the signs of the distance and road quality variables are statistically significant and have the anticipated signs. The coefficient on the Border dummy variable is positive and statistically significant. The 0.0742 implies that all else equal, the price difference between two markets is about 7.5 percentage points higher if the trade crosses a national border versus does not cross a national border.

|  |  |  |
| --- | --- | --- |
|  | Abs % Price Diff |  |
| ln(D\_ij) | 0.0352\*\*\* | (0.00516) |
| Q\_ij | -0.0260\*\* | (0.0131) |
| Border | 0.0742\*\*\* | (0.00576) |
| Constant | 0.247\*\* | (0.102) |

The average price difference with between country trade is 33.0 percent. This suggests that about one quarter of this price difference can be explained by border-specific fixed administrative fees. As previously noted, this fixed administrative fee can consist of financial fees for cross-border shipments as well as implicit time costs.

## Reminder about Bias

Recall that there are two likely sources of bias concerning the size of the relationship between distance and the size of the price gap. First, by not accounting for the threshold effects (i.e., mixing positive and zero trade market outcomes) the estimate of the relationship is likely biased down. Second, by not controlling for the positive correlation between distance and unobserved fixed administrative fees the relationship is likely biased up. This means that the bias could go in either direction. Given the very small distance elasticity which was estimated by Brenton et al. (2014) is it is quite likely that the net effect is a downward bias.

# Part B: Fertilizer Prices in Sub-Saharan Africa

The 2020 paper by Cedrez et al. is titled “Spatial variation in fertilizer prices in Sub-Saharan Africa”. The authors recognize that low yields are often the result of low fertilizer use. Fertilizer use depends on fertilizer prices and so the goal of the analysis is to investigate the extent that fertilizer prices vary spatially. The authors have 1789 geo-coded price observations in two separate regions: northwest Africa and southeast Africa. The data was collected for select years between 2010 and 2018. A key result is that fertilizer prices are higher in the land-locked countries of northwest and southeast Africa versus the coastal countries.

There are many interesting results in this paper, most of which have some connetion to the spatial LOP and pricing integration. Unfortunately time does not permit a detailed review of the paper. Rather, after a brief review of the Introduction of Cedrez et al, (2020) we will focus on using R to map the geo-coded price data which was published along with the paper.

## Review of Introduction of Cedrez et al. (2020)

The authors indicate that the average crop yield in Sub-Saharan Africa is 1266 kg per hectare while the global average is 3745 kg hectare. An important reason for low crop yields is the low use of fertilizers such as urea nitrogen. The average use of fertilizer in Sub-Saharan Africa is 14 kg per hectare, which is much lower than the 141 kg per hectare used in South Asia and the 154 kg per hectare used in the European Union.

There are many reasons for the low use of fertilizer in Sub-Saharan Africa. Reasons such as inadequate farm-level cash reserves and low return from investing in fertilizer are connected to the price of fertilizer. If there is wide variation in the price of fertilizer in different regions then it should be possible to test the extent that high fertilizer prices are responsible for low fertlizer use.

The authors obtained their data from the publication African Fertilizer, which reports the prices of 25 kg and 50 kg bags of Urea nitrogen by town. The authors geo-coded the latitude and longitude of each town. The final data set includes 7823 observations for 878 locations in seventeen countries:

* Benin
* Burundi
* Burkina Fas
* Coˆte d’Ivoire
* Ghana
* Kenya
* Mali
* Malawi
* Mozambique
* Niger
* Nigeria
* Rwanda
* Senegal
* Togo
* Tanzania
* Uganda
* Zambia

## Preliminary Coding

The analysis begins with the usual clearing of the objects from R’s memory and clearing all plots.

rm(list = ls())  
graphics.off()

The next step is to retrieve the various packages which will be used for the mapping exercise:

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

## Creating the Map Data Frames

To achieve a desired scale of the pricing maps it is useful to create separate maps for the southern countries and the northern countries from the previous list. The two lists are created as follows:

south <- c(  
 "Tanzania", "Burundi",   
 "Kenya", "Malawi", "Mozambique",  
 "Rwanda",   
 "Uganda"  
)  
  
north <- c(  
 "Burkina Faso", "Ivory Coast", "Ghana",  
 "Mali",  
 "Nigeria", "Senegal"  
)

The next step is to use R’s map\_data() function to create the mapping data frames which can be plotted with R’s ggplot() function. The arguments of the map\_data() function include: (1) the map which is to be displayed (e.g., North America, France, World); and (2) the specific regions to be shown in the displayed map. In this case the World map works well, and the regions to be displayed are contained in the two lists (south and north) as shown above. The required code is

southMap <- map\_data("world", region = south)  
northMap <- map\_data("world", region = north)

The next step is a little complicated. We want to take the names of the countries which are in *southMap* and *northMap* and use them to label each country in our pair of maps. We need to tell R where to put the labels. Let’s place them at the center of each country.

southLabel <- southMap %>%  
 group\_by(region) %>%  
 summarise(long = mean(long), lat = mean(lat))  
  
northLabel <- northMap %>%  
 group\_by(region) %>%  
 summarise(long = mean(long), lat = mean(lat))

## A Note about Longitude and Latitude

Let’s display the contents of the *northLabel* data frame to better understand how the labels are placed on the map:

northLabel

## # A tibble: 6 x 3  
## region long lat  
## <chr> <dbl> <dbl>  
## 1 Burkina Faso -1.59 12.0   
## 2 Ghana -1.06 8.32  
## 3 Ivory Coast -5.77 7.97  
## 4 Mali -5.38 14.8   
## 5 Nigeria 7.75 8.55  
## 6 Senegal -14.4 14.0

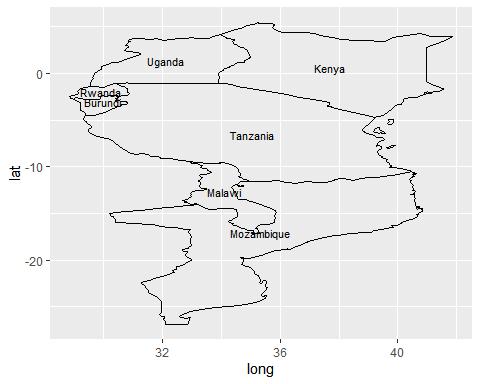
The first column is the country labels for the northern group of African countries. The second column is the longitude and the third column is the latitude of the center of the country in question. For example, the longitude of the center of Burkina Faso is -1.59 and the latitude is 12.0. Lines of longitude run north – south, beginning with prime meridian which runs through Greenwich, England. The positive longitude values are degrees west of the prime meridian and these run from 0 to 180. The negative longitude values are degrees east of the prime meridian and these run from 0 to -180. If you look at a map of the world you will notice that the center of the Burkina Fasa is slightly to the east of a vertical line which passes through Greenwich, England. This is why the longitude of the center of Burkina Faso is -1.59.

Lines of latitude are those which run parallel to the equator. Positive values indicate degrees north of the equator, running from 0 to 90. Negative values indicate degrees south of the equation, running from 0 to -90. Notice that the center of Burkina Faso is 12.0, which implies that it is 12 degrees north of the equator. As a point of comparison, Nicaragua in Central America has approximately the same latitude as the center of Burkina Faso.

## Plotting a Map without the Price Data

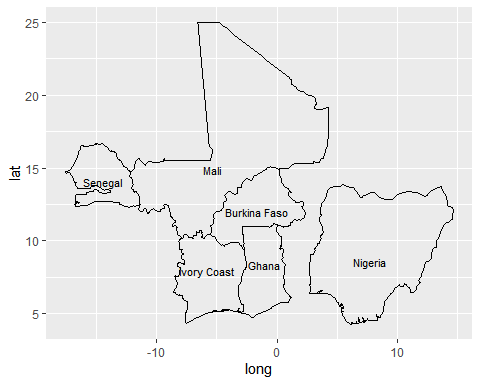
We will use R’s ggplot function to generate the maps for the southern and northern group of countries. The code for generating the map for the southern group of countries is as follows:

south\_plot <- ggplot() +  
 geom\_path(data = southMap, aes(x = long, y = lat, group = group)) +  
 geom\_text(aes(x = long, y = lat, label = region), data = southLabel, size = 3, hjust = 0.5)  
  
south\_plot



The code for the northern group of countries is similar

north\_plot <- ggplot() +  
 geom\_path(data = northMap, aes(x = long, y = lat, group = group)) +  
 geom\_text(aes(x = long, y = lat, label = region), data = northLabel, size = 3, hjust = 0.5)  
  
north\_plot



## Importing the Price Data

To ensure that the map does not become over crowded with pricing information, ony a subset of the fertilizer pricing data from 2018 is plotted. The data is in the csv file titled *fertilizer\_price.csv*. The file is opened in the usual way:

myData <- read.csv(here("Data", "fertilizer\_price.csv"))  
head(myData)

## year iso ppp\_price\_kg longitude latitude nat\_avg south  
## 1 2018 TZA 0.84 39.20833 -6.792354 0.88 1  
## 2 2018 TZA 0.82 37.66149 -6.816536 0.88 1  
## 3 2018 TZA 0.88 37.56535 -3.002242 0.88 1  
## 4 2018 TZA 0.88 36.69402 -3.372658 0.88 1  
## 5 2018 TZA 0.88 34.76669 -9.333318 0.88 1  
## 6 2018 TZA 0.88 35.64237 -10.646216 0.88 1

The last column of the imported data is an indicator variable for identifying whether the country in question belongs on the southern map or the norther map. This indicator can now be used to create a southern country price series and a norther country price series.

southData <- myData[(myData$south == 1), ]  
northData <- myData[(myData$south == 0), ]

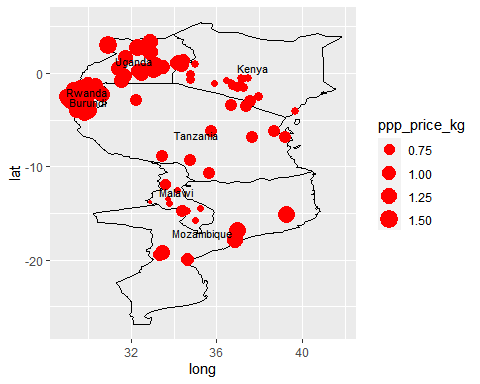
## Mapping the Price Data

Mapping the price data involves using ggplot to map the countries as shown above, and then adding the *geom\_point()* function to map the prices. The *geom\_point()* function identifies where the price marker should be located based on the longitude and latitude of the town. This function also uses the *size* parameter to represent prices as different sized bubbles.

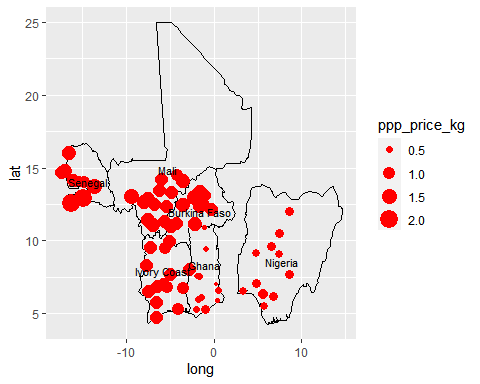
south\_plot2 <- ggplot() +  
 geom\_path(data = southMap, aes(x = long, y = lat, group = group)) +  
 geom\_point(data = southData, aes(x = longitude, y = latitude, size = ppp\_price\_kg), color = "red") +  
 geom\_text(aes(x = long, y = lat, label = region), data = southLabel, size = 3, hjust = 0.5)  
  
north\_plot2 <- ggplot() +  
 geom\_path(data = northMap, aes(x = long, y = lat, group = group)) +  
 geom\_point(data = northData, aes(x = longitude, y = latitude, size = ppp\_price\_kg), color = "red") +  
 geom\_text(aes(x = long, y = lat, label = region), data = northLabel, size = 3, hjust = 0.5)

We can now display the completed pair of maps.

south\_plot2



north\_plot2



## Analysis of Mapped Prices

The main purpose of this second half of Module 1D was to learn how to create spatial price maps. For this reason only brief comments will be provided about the pricing outcomes.

The previous map of the southern countries reveals that Kenya and Malawi have the lowest prices, and Uganda and Mozambique have the highest prices. Unfortunately, due to crowding of the markers it is not possible to determine if the prices in Rwanda and Burundi are above or below average. Uganda is a land-locked country and so higher prices are expected. Mozambique has a major port in the city of Maputo and so the fact that this country has relatively high prices is unexpected.

The previous map of the northern countries reveals a similar situation. In this case Nigeria and Ghana have relatively low prices, and Burkina Faso and Senegal have relatively high prices. Burkina Faso is land locked and so high prices are expected. In contrast, the relatively high prices in Senegal are not expected because the port of Dakar is the third largest in western Africa.

Overall the proposition that land locked countries in Sub-Saharan Africa have higher fertilizer prices has some empirical support in the pair of maps but the evidence is not very strong.