## Modelling the relationship between maize yields and precipitation and temperature

**Time period:** 1999- 2014

* At the moment, the aggregated climate/weather data which I was using are for the period 1999-2018. The yields data are from 1970-2014

**Weather data:**

1. Precipitation:
   * Z-score at the location where precipitation corresponds to the 10th percentile in each county (I opted for z-scores rather than the raw data because the maximum likelihood algorithm often fails to converge using the raw data.)
   * Monthly frequency
2. Temperature:
   * Z-score at the location where temperature corresponds to the 90th percentile in each county
   * Monthly frequency

**Maize yield data:**

* MT (= metric tons) per hectare
* Yearly frequency
* Obtained from Gideon Galu from the FEWS NET. The original source of the most of the data is the Ministry of Agriculture.

**Precipitation and temperature data aggregation:**

* The frequency of the yields data is yearly while the weather data are monthly. We need the frequency of both datasets to be equal for estimating the panel models.

-> temperature and precipitation need to be aggregated to obtain yearly values

* For each county and year, we aggregate the weather data over the months of planting and growing based no the following seasonal calendar: <http://fews.net/east-africa/kenya/seasonal-calendar/december-2013>
* According to the seasonal calendar, the counties can be divided into two groups: Eastern and Western.
  + **Eastern counties:**
    - Provinces: Eastern, North Eastern, Coast
    - Two planting and harvesting seasons
    - The yearly climate measures obtained as averages over **November and December of the previous year and January, February, March, April, May, June, July, August and September of the current year** (the months of planting and growing seasons).
  + **Western counties:**
    - Provinces: Rift Valley, Western, Nyanza, Central
    - One harvesting and planting season
    - Yearly climate measures obtained as averages over **May, June, July, August and September** (the months of planting and growing seasons.)
* Besides the average climate measures above, the **coefficients of variation** for both climate and temperature were included in the first models.
  + The coefficients of variation were calculated over the same months as described above for the means (based on the seasonal calendar: <http://fews.net/east-africa/kenya/seasonal-calendar/december-2013>)
  + The coefficients of variation turn out to be insignificant

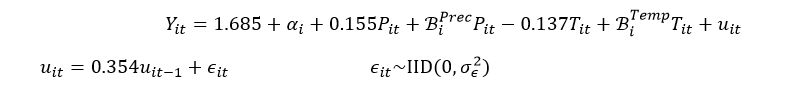
## First results: Selected specifications of the mixed-effects models

## 1. No weights

The best specifications of the error structure based on LR tests of serial correlation, and LR tests of random effects:

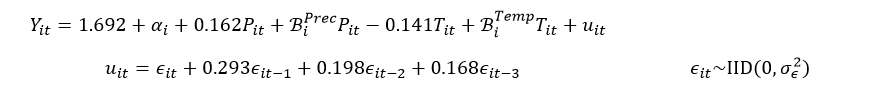
**a)** AR(1) errors

lme(Yield~1+PrecZscore +TempZscore , random= ~ PrecZscore + TempZscore|ID, correlation = corAR1(0, form= ~ Year|ID))



**b)** MA(3) errors

lme(Yield~1+PrecZscore +TempZscore , random= ~ PrecZscore + TempZscore|ID, correlation = corARMA(0, form = ~ as.numeric(Year)|ID, p=0,q=3))



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No weights | **Error structure** | | | R function *pvcm* from the *plm* package (no AR or MA) | |
| **Fixed effects** | **No AR** | **AR(1)** | **MA(3)** | **Swamy (1970)** | **Counties separately (mean)** |
| Intercept | 1.647\*\*\* | 1.685\*\*\* | 1.692\*\*\* | 1.610\*\*\* | 1.636 |
| Precipitation  (z – score) | 0.171\*\*\* | 0.155\*\*\* | 0.162\*\*\* | 0.196\*\*\* | 0.201 |
| Temperature  (z – score) | -0.109\* | -0.137\*\* | -0.141\*\* | -0.089  (p-val= 0.141) | -0.106 |
| AIC | 1180.105 | 1130.211 | 1122.896 |  |  |
| BIC | 1225.900 | 1180.522 | 1182.354 |  |  |

ANOVA: Precipitation explains much bigger part of the variation than temperature (the values of the F-statistics: 44.4 and 7.8).

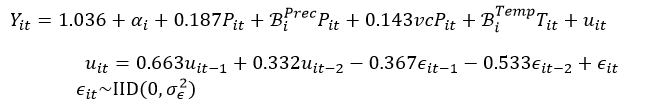
## 

## 2. Weights: Area of cropland

lme(Yield~1+PrecZscore +cv\_Prec, random= ~PrecZscore+TempZscore|ID , weights=~Area, correlation= corARMA(form = ~ Year|ID, p=2,q=2))

The best specification of the error structure based on LR tests of serial correlation, and LR tests of random effects:

ARMA(2,2) errors:



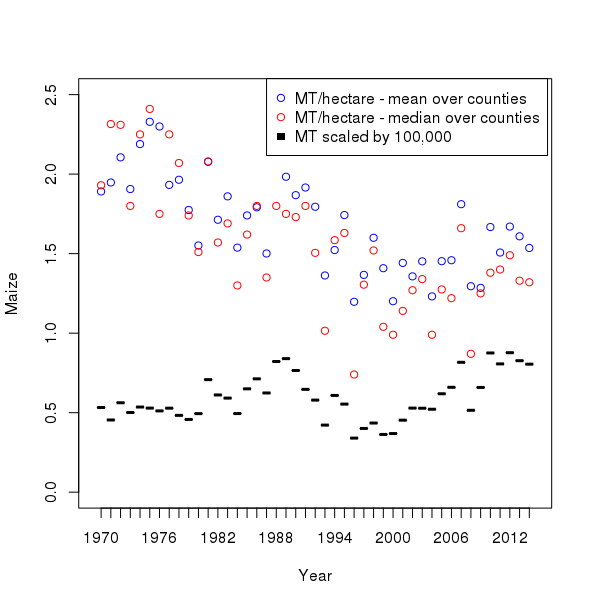
**AIC:** 1817.509 **BIC:** 1881.509

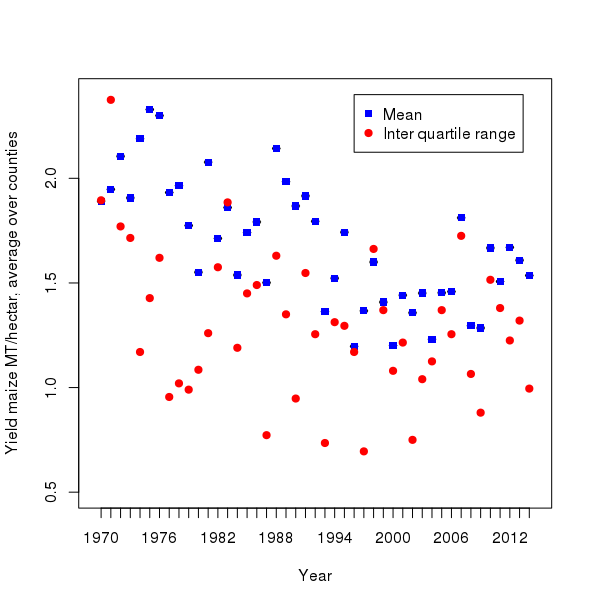
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Weights = Area* | **Error structure** | | R function *pvcm* from the *plm* package (no AR or MA) | |
| **Fixed effects** | **No AR** | **ARMA(2,2)** | **Swamy (1970)** | **Counties separately (mean)** |
| Intercept | 1.237\*\*\* | 1.036\*\*\* | 1.515\*\*\* | 1.519 |
| Precipitation  (Z – score) | 0.214\*\*\* | 0.187\*\*\* | 0.218\*\*\* | 0.224 |
| Precipitation –  coef. of variation | 0.207\*\* | 0.143\* | 0.107 | 0.198 |
| AIC | 1868.164 | 1817.509 |  |  |
| BIC | 1913.900 | 1881.540 |  |  |

## 

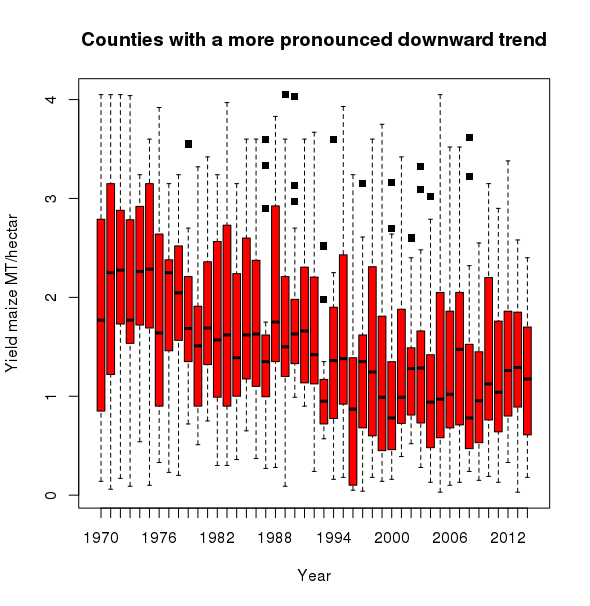
## 

## Descriptive analysis of maize yields data from 1970

* Interesting point: **a strong downward trend in yields**, at least until 1990. Although the production (MT) exhibit an upward trend.



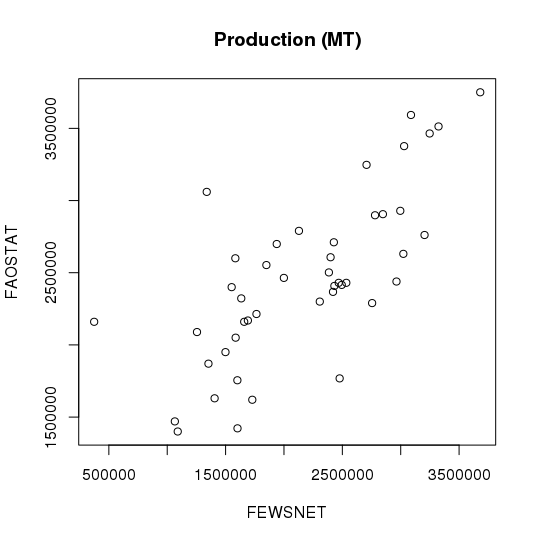
* It appears that the data are more reliable from year 1991 because of the following reasons:
  1. Relatively many n/a’s before 1991
  2. Strong downward trend in yields before 1991
  3. The interquartile range appears to be smaller and more stable after 1991
  4. T-tests show that there is a significant difference between the mean before 1991 and after 1991
  5. Before 1991, no separation between short and long seasons. Only one data entry for each county and year. Between 1991 and 2001 there are three data entries for most counties. These are long rains, short rains and together.
* To investigate, if there are some counties without the downward trend, I plotted each county separately:
  1. Counties **without** (or with much less pronounced) **downward trend**: Nyandarua, Kwale, Lamu, Tana River, Kitui, Mandera, Homa Bay, Migori, Elgeyo Marakwet, Narok, Samburu, Turkana, Uasin Gishu, West Poko, Bungoma, Kakamega , Vihiga , Meru, Tharaka Nithi, Nandi

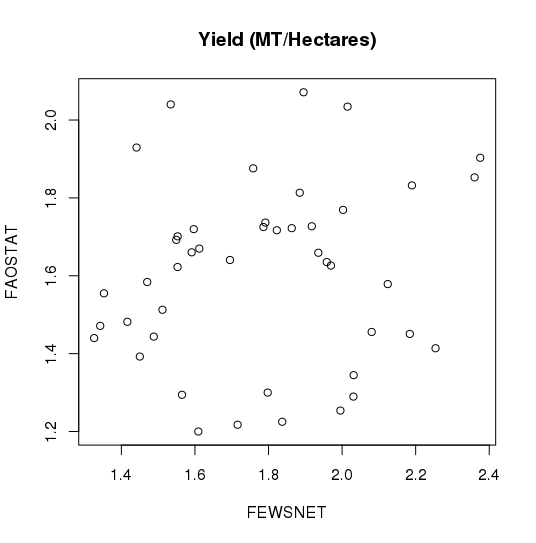
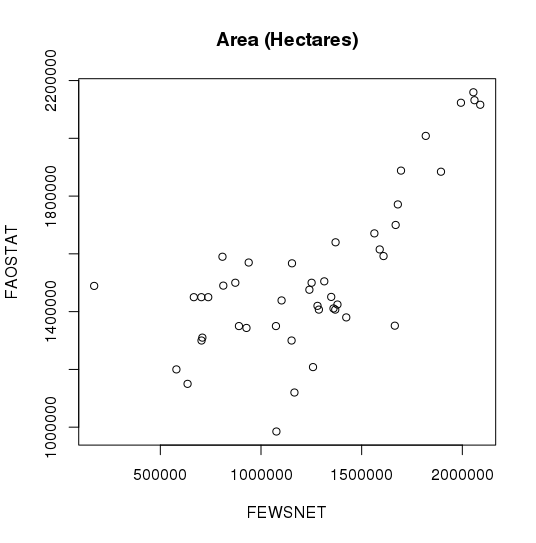


### 

### Comparison of the county level yield data (FEWSNET) with the country level FAOSTAT data

* I aggregated the FEWSNET data as follows:
  + For every year, I calculated the sum of production in MT over the counties
  + For every year, I calculated the sum of harvested area in hectares over the counties
  + I obtained the country level yields for every year by dividing the production (obtained as described above) by area (obtained as described above)
  + Then I compared the country level aggregates with the FAOSTAT data





* For each variable (production, area and yield) I calculated relative differences between the datasets in percentage as follows:

100\*(FAOSTAT-FEWSNET)/FEWSNET

* Histograms of the relative differences are below

# Climate measures in the literature:

* Erin Lentz et al. (2017): ***prices and consumer strategy index as dependent variables rather than yield***
  + Total rainfall during the last year’s rainy season (October to April for Malawi)
  + Timing of the beginning of the rains for the prior year
    - Number of days following the first October when the rains began
    - The beginning of days defined as when it rains for at least three of the past five days for a total accumulation of at least 10 millimeters
  + Dry spells: maximum number of days without rain during last year’s rainy season (Oct-April)
  + Maximum daily precipitation that month in regions that are susceptible to floods
* Abraha and Savage (2006)
  + Wet and dry day counts (probably per month)
    - Definition of `wet and dry days’ not clear from the paper
  + Monthly total rainfall and its variances
  + Daily and monthly mean and SD of wet day precipitation
  + Min and max air temperatures. Probably daily extremes, these then used for monthly mean and SD (not entirely clear from the paper)
* Adejuwon (2005)
  + Total rainfall for
    1. The first month of the period from sowing to harvesting (June)
    2. The first two months of the period from sowing to harvesting (June and July).
       - this turns out to be the most important in this case
    3. The three months of the period from sowing to harvesting (June, July and August)
  + Minimum and maximum temperature of each of the growing season months
* Ben Mohamed et al. (2002)
  + Sea surface temperature anomalies at various locations (see paper for the particular sites) - ***significant***
  + Number of rainy days (details not specified)-***significant***
  + Daily amount of rainfall (as such not usable in our case - not consistent with our yearly frequency)
  + Amount of rainfall in July, August and September - ***significant***
  + Maximum air temperature in the hottest month April (Niger)
  + Minimum air temperature in the coldest month January (Niger)
  + The length of rainy season:
    1. The difference between the dates of the beginning and the end of the season.
       - The beginning of the growing season is defined as being when the amount of rainfall in three consecutive days is at least 25mm and no dry spell of more than seven days duration occurs in the following 30 days.
       - The end of the rainy season is that rainy day after which rain recorded during 20 days is less than 5 mm.

* Blignaut et al. (2009)
  + Rainfall as the annual sum of the provincial monthly average
  + Temperature as the annual averages of the daily maximum temperatures in two or three towns or cities per province
    - The data received as monthly averages of daily maxima, these then used to compute annual averages

* Chipanshi et al. (2003)
  + Daily maximum and minimum temperatures
    - I think that these were used as inputs into a GCM (together with daily solar radiation calculated based on daily sunshine hours). Then climate change simulated->then hypothetical temperature as ***mean monthly values and rainfall as departures from normal*** obtained from GCM under climate change. These then used as intermediate inputs to estimate effects of climate change on yields
* Gbetibouo and Hassan (2005): ***Dependent variable = farm net revenue (Ricardian approach)***
  + *‘The normal climate variables based on 30 years average of temperatures and precipitation observed over 1970-2000’*
  + Rainfall pattern (mm)
    - Summer and winter separately
    - Summer and winter months not specified
    - Way of aggregation not specified in more detail, probably cumulative
  + Average temperature (°C)
    - Summer and winter separately
    - Summer and winter months not specified
    - Way of aggregation not specified in more detail
* Giannakopoulos et al. (2009)
  + General circulation model predicts climate change of C2. This used as an input for the CROPSYST model
  + CROPSYST model predicts effects of climate change on crop productivity
    - Measures used as inputs for the CROPSYST (not specified in more detail):
      * Tmax
      * Tmin
      * Total rainfall
  + Other interesting climate measures used to describe climate changes. But probably not used as inputs for the crop productivity model here.
    - Number of summer days defined as Tmax>25C
    - Number of hot days Tmax>30C
    - Number of heatwave days Tmax> 35C
    - Number of tropical nights Tmin>20C
    - Number of frost nights Tmin<0C
    - Number of dry days (dry day if daily precipitation amounts to less than 0.5mm = a typical threshold value) and spells
    - Annual maximum running total rainfall over 3 days (potential to cause local flooding)