Forecasting Crop Production: A Seasonal Regression Model

Decomposition of MAPE and SMAPE

Jean Baptiste HABYARIMANA
Agriculture Statistician
Ministry of Agriculture and Animal Resources
Kigali – Rwanda

The essentialities of crop production forecasting are huge. Crop production forecasts are mainly used to monitor food security and compile Gross Domestic Product. Due to the usefulness of crop production forecasts, this paper uses a Seasonal Regression Model to forecast crop production. Given the fact that one agricultural season differs from another in terms of number of the days, weather conditions (the amount of rains received), and crop growing conditions as a whole, this paper models agriculture production as function of agriculture season performance. This paper provides cereal crop specification end-of-season production forecasts. Production forecasts in this paper rely on crop production recorded in 32 last agriculture seasons and season performances. In sample production forecasts are evaluated using $R^2$, MAPE, and SMAPE. An out sample forecasts for six agricultural seasons ahead were computed for rice time series that satisfy the condition of MAPE and SMAPE less that 10% and which IMAE and ISMAE curves coincide at all fitted points.

Introduction

The majority of population in developing countries depends on agriculture production for household food consumption. The most predominant agricultural system practiced in these countries is subsistence agriculture. Commercial agriculture is less developed and insignificant to increase households’ purchasing power as source of income and contribute to households’ food consumption. To feed its population developing countries import food commodities to complement the gap that exists between domestic food demand and domestic food production. Domestic demand is function of the population size and food supply. Food supply is function of domestic food production, import and export of food commodities, and changes in domestic food stock. However, crop production forecasts can be a good tool to monitor food security situation in a country. They may be used to forecast food availability, understand food price variations, plan for import and/or export of food commodities, and monitor domestic food stock in relation with population size. They may also be useful when planning for the development of food processing industries in the country.

Crop production can also be used as raw materials in food processing industries. From this extend, the value added resulting from each step of processing can be used both for government revenue and computing Gross Domestic Product. As government revenue, expansion of food processing industries can increase the amount of taxes collected on value added VAT. The value added from crop production can be used when

The author wishes to acknowledge technical assistance generously given by Mr. Enock Chinganda Senior Expert in Statistics.
compiling Gross Domestic Product by production. For this, crop production forecasts allowing plan for human food consumption and food processing industries development can contribute to economic growth and measurement of economic growth. Seen that the uses of crop production forecast are huge, this paper is limited to the crop production forecasting.

**Literature**

Agricultural statisticians and economists have used a wide array of models to forecast crop production (Anneke and Todd, 2011). In the models based on crop yield, two terms are distinguished: Area Sown that refers to the area on which sowing or planting has been carried, for the crop under consideration, on the soil prepared for that purpose and Area Harvested that excludes the area from which, although sown or planted, there was no harvest due to damage, failure, etc (FAO, 2010). However, the earliest models to forecast crop production have been relied on the combination of crop yields estimates generated through crop cutting methods and harvested area estimates generated through list and/or area frame sampling (Spinks)\(^1\). With the evolution of agricultural statistics methods and techniques to crop production forecasting, (USDA, 1999) shows that crop production forecasts have two components: acres to be harvested and expected yield per acre and states that crop production forecasts are based on the current crop growing conditions and projected conditions. The theory of crop cuts to estimates crop yields and then crop production was endorsed by FAO in 1950s (FAO, 1982; Verma et al., 1988; Spencer, 1989; Murphy et al., 1991; Rozelle, 1991; Fielding and Riley, 1997; Norman et al., 1995; Anneke and Todd, 2011). As Crop cuts measure the biological yields, they were criticized of not taking into account postharvest losses and thus not reflect the economic yield (Poate, 1988; Anneke and Todd, 2011). The most common alternative method for crop cuts is the use of farmer estimates, which give a measure of economic yield to forecast crop production (David, 1978; Singh, 2003; Anneke and Todd, 2011). Farmer estimate method has been criticized of miss-recording and records keeping of data on production by farmers (Anneke and Todd, 2011). For years, it was assumed that farmer estimates were too subjective and unreliable to obtain reliable data on crop yield (Verma et al. 1988), whereas crop cuts were assumed to be unbiased (Murphy et al. 1991). With the improvement in data collection, computers and statistical programs and positive mathematical programming, crop forecasting models based research have expanded. To undertake crop production forecasting, (SHARIFI, 2000) combined crop area estimates generated through area frame sampling and processing of satellite data, and yields derived through the application of crop growth simulation model and statistical techniques.

The important recent contribution to the analysis on crop production forecasting concerns the impact of climate change on agriculture and livestock activities and is related to the specification of seasonal components with multiple effects on crop production (Richard, 1998; Parry et al., 2004; Claudia et al., 2010; Oyiga, 2011). Recent evidences show that climate change will worsen the living conditions of farmers-dependent people. In addition to this, climate change and food security analysts have recently developed models to show how more frequent and more intense extreme weather events will have adverse impacts on crop production (Claudia et al., 2010). In all these, agricultural seasons have been modeled as a repeating pattern often found in food production, distribution, demand, and price data (Abera et al., 2010; Pedram et al., 2011; Alessandro, 2012).

---

\(^1\) [http://ageconsearch.umn.edu/bitstream/8933/1/24010018.pdf](http://ageconsearch.umn.edu/bitstream/8933/1/24010018.pdf)
Most applications in the literature assume that regions where agriculture is rainfall fed and that rainfall received was enough for crop water requirement and growing conditions, harvesting corresponds in most of the case to better crop production (Martin et al.\textsuperscript{2}, Masvaya et al.\textsuperscript{3}).

In a rainfall fed agriculture system, discussed methods used to forecast crop production are limited to the estimation of crop production at the end of the agriculture seasonal in the progress and they are weak to provide crop production forecasts beyond agricultural season under observation. To generate out-sample forecasts, this paper first evaluate in sample forecasts. Taking into account climate change and long run seasonal changes, this paper also assumes that seasonal performance for the 6 posterior seasons would change as they had changed in the 32 prior to the out sample forecasting.

**Empirical Modeling**

For the purpose of forecasting crop production, this paper uses a dummy variable regression model. As Rwanda has two main agricultural seasons, season A and Season B, the dummy variable regression model is stated as follow:

\[
y = \beta_0 + \beta_1 t + \beta_2 S_A + \beta_3 S_B + \epsilon,
\]

Where \(y\) is forecasted value, \(\beta_i, i=1,\ldots,3\) are regression coefficients, \(t\) is period (2002A period 1 up to 2017B period 32), \(S_A\) is Season A and \(S_B\) is season B, and \(\epsilon\) is regression errors. Since in this model there is 2-period seasonal variation, 2 dummy variables are assigned as follow:

- For Season A data: \(S_A = 1\) \(S_B = 0\)
- For Season B data: \(S_A = 0\) \(S_B = 1\)

As the main purpose of this paper is to produce crop production forecasts for 6 period ahead, this paper assumes that seasonal performance for the 6 periods ahead would change in the same conditions as they have changed during the observed periods.

To evaluate goodness of fit of used model and determine data suited to furcating procedures, this paper uses Mean Absolute Percentage Error (MAPE) and Symmetric Mean Absolute Percentage Error (SMAPE). To detect errors for each recorded actual value that can cause superior errors in the forecasted values; we decomposed MAPE and SMAPE into Individual Mean Absolute Error (IMAE) and Individual Symmetric Mean Absolute Error (ISMAE).

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|A_i - F_i|}{A_i} * 100;
\]

\[
\text{SMAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|A_i - F_i|}{(A_i + F_i)/2} * 100
\]

\[
\text{IMAE} = \frac{|A_i - F_i|}{A_i} * 100;
\]

\[
\text{ISMAE} = \frac{|A_i - F_i|}{(A_i + F_i)/2} * 100
\]

Where: \(A_i = \text{Actual Value}, F_i = \text{Forecast Value}, n = \text{Number of fitted points}\)

When IMAE and ISMAE are plotted together and that the two curves coincide it shows how the forecast

\textsuperscript{2} http://www.fao.org/docrep/016/i2800e/i2800e02.pdf
\textsuperscript{3} http://www.waternetonline.ihe.nl/challengeprogram/ir34%20masvaya%20rainfall%20strategy.pdf
values are closer to actual values. As we have the same numerator for both IMAE and ISMAE, IMAE and ISMAE values will be closer when IMAE’ denominator and ISMAE denominator are closer. For a given recorded actual value where the corresponding IMAE and ISMAE curves do not coincide, this explains that there is a strong deviation between actual and forecast value at that point. When deviation between IMAE and ISMAE curves is observed, the detection in actual value shows that there is some strong differences in actual values around that record where the deviation was observed which can allow strong errors in forecasted values and a model with this characteristic cannot generate good out-sample forecasts. This paper consider that data suited to in-sample and out-sample forecasting are those which MAPE and SMAPE are less than 10% and which IMAE curve coincides with ISMAE curve at all fitted points when there are plotted together.

Case Study

This paper provides forecasts on four important cereals grown in Rwanda that include Sorghum, Maize, Rice, and Wheat. Rwanda has three agricultural seasons in a year that commonly known as Season A, Season B, and Season C. Season A that starts in September of one year to January the following year and Season B that starts in February to July on the same year are the main agriculture seasons available in Rwanda. All selected cereals are grown in the both main agricultural seasons A and B. The main purpose of selecting cereals in this study is that cereal production can be used for human consumption without transformation and can be used as raw materials for food processing industries. In addition to this, cereals are the main food crops that can be transformed into different multiple products of food for human consumption and for other uses like animal feed. Processing cereals may be extended to the production of flour, bread, biscuits, beer, etc … and other different animal foods. At each stage of transformation, there is an added value that contributes to Gross Domestic Product. Both cereals consumed without any transformation and those consumed after transformation depend on domestic production and imports.

Population Size determines a country’s food need. The population of Rwanda increased by 22.6% between 2002 and 2012 (NISR, 2012). This growth translates into a constant shift factor for food demand. Increase in population size increases domestic food demand and may affect the quantity available to industrial transformation. In a country with subsistence agricultural system, increase in food demand may lead to decrease in domestic food stock, food price variations and may also be a barrier to the sustainable development of domestic agricultural industries. The increase of food demand for human consumption limits the quantity of crop production available to agricultural industries as raw materials. This situation leads to food processing industries development ambiguity where the supply of raw materials for transformation is less than the production capacity of those industries and remains to be the main challenge to food processing industries productivity and profitability.

The development of food industry capable to produce food in different forms depends on available quantity of food crops production after human and animal consumption. To satisfy domestic food demand, there is a need of increasing cereal production in the country and the development of food industry can focus on cereals transformation that offers a wide food transformation than other food crops. To respond to both cases, crop production as food for human consumption and crop production as source of raw materials for industries the main problem remain to be the development of reliable crop production forecasting method.

This paper provides a statistical forecasting method that can be used to forecast food production. The used
method can also be used when planning for human food consumption and developing strategies for the development of food processing industries. When percentages for human and animal consumption need are available, the model used here can offer an easier way to plan for crop production available to food processing industries. This paper also contributes to the existing literature in the field of crop production forecasting.

**Empirical Strategy and Results Discussion**

We first determine whether a stable long-run relationship exists among the seasonal time series. The analysis of Adjusted R Squares of 0.93, 0.92, 0.89, and 0.93 for Sorghum, Maize, Wheat, and Rice respectively show that actual values are suited to forecasting techniques. To detect the stability in long run, we analyzed MAPE and SMAPE. MAPE and SMAPE values for sorghum, maize, and wheat show that these time series are not suited to forecasting techniques because they are too large and greater than 10%. While MAPE and SMAPE for rice shows that this time series is suited to forecasting techniques because are small and less than 10%. Results for sorghum, maize, and wheat show that the difference of MAPE (=25%) and SMAPE (=39%) of 14%, MAPE (=29%) and SMAPE (=43%) of 14%; MAPE (=29.5) and SMAPE (=36.1) of 7% are very big. These differences show that there are some actual values that disturb the used forecasting technique which may result in poor out-sample forecasts. But the results for rice show that the difference between MAPE (=7.5%) and SMAPE (=8.5%) of 1% is very small. This shows that errors in the used technique are very small to influence forecasted values.

The decomposition of MAPE and SMAPE into IMAE and ISMAE shows that when IMAE and ISAME for sorghum, maize, and wheat are plotted together their related curves do not coincide at all points and they show deviations between IMAE and ISMAE. The detection on problem in actual recorded values shows that records that are allowing forecasting problems in our model for sorghum are those recorded in 2005 Season A and 2006 Season A. For maize time series are those recorded in 2004 Season B, 2005 Season B, 2006 Season B, 2007 Season B, and 2008 Season B. While for wheat are those recorded in 2006 Season A, 2006 Season B, and 2007 Season A (see annex)
Figure 1. Curves of IMAE and ISMAE plotted together for each time series.
I this paper we consider that a time series suited to forecasting techniques with best forecast values is the one that can provides MAPE and SMAPE less than 10% and which IMAE and ISMAE curves coincide at least at all points of observation. In our study, only rice time series at least respect these conditions (See the above Graphs).

**Out sample forecasting**

The model evaluation and in sample forecasting show that rice seasonal time series is suited to forecasting techniques. The figure below plots actual values (in blue line) together with forecasted values (in red line). This paper forecasts 6 periods ahead the actual values. Forecasted values show that at the first and second period, the rice production would decrease when compared to their previous values. While from the third up to the sixth period the forecasted values will increase when compared to the actual values.

Assuming that the percentages of the quantity of rice required both for human consumption and food industries transformation remain constant from the first up to the sixth period of forecasting, the decrease of rice production in the first and second period (forecasted value) would explain that quantity of rice available to both human consumption and food processing industries will decrease while it will increase from the third up to the sixth period when compared to the previous recorded values.

![Figure 2. Actual (blue) and forecasted (red) values.](image)

**Conclusion and policy implication**

This paper has forecasted crop production using seasonal regression model. the major findings that emerge from this study are that Adjusted R Square is not the best measure of goodness of fit of forecasting models; Mean Absolute Percentage Error and Symmetric Mean Absolute Percentage Error can provide the best measures of goodness of fit of model; Individual Mean Absolute Error and Individual Symmetric Mean Absolute Error curves coincide at all points when a time series’ forecasted values and actual values are very closer. This paper empirically demonstrates that rice time series was suited to forecasting technique and that sorghum, maize, and wheat time series showed forecasting problems rising from errors in recorded actual values. Forecasts for rice production showed than during the two first periods, the production will decrease to increase in the third to the sixth period. This could be the best raison to plan for import during the two first periods to satisfy human consumption and food processing industries demands.
References

United States Department of Agriculture (USDA). (1999). Understanding USDA Crop Forecasting. http://www.nass.usda.gov/Education_and_Outreach/Understanding_Statistics/pub1554.pdf

SHARIFI M.A. (2000). Crop Inventory and Production Forecasting Using Remote Sensing and Agrometeorological Models: The Case of Major Agricultural Commodities in Hamadan Province, Iran. http://www.isprs.org/proceedings/xxxiii/congress/part7/1364_XXXIII-part7.pdf

FAO. (2010). Data Structure, Concepts and Definitions common to FAOSTAT and CountrySTAT framework. http://www.countrystat.org/resources/metadata_en.pdf

Spinks G.R. (.). Uses and Methods of Crop Forecasting. http://ageconsearch.umn.edu/bitstream/8933/1/24010018.pdf

Anneke Fermont and Todd Benson. (2011). Estimating Yield of Food Crops Grown by Smallholder Farmers. http://www.ifpri.org/sites/default/files/publications/ifpridp01097.pdf

FAO. 1982. The Estimation of Crop Areas and Yields in Agricultural Statistics. Economic and Social Development Paper No. 22. Rome, Italy.

Murphy, J., D. J. Casley, and J. J. Curry. 1991. Farmers’ Estimations as a Source of Production Data. World Bank. Technical Paper 132. Washington, DC: World Bank.

David, I. P. 1978. “Non-sampling Errors in Agricultural Surveys. Review, Current Findings, and Suggestions for Future Research.” Paper presented at the Philippine Statistical Association Annual Conference, June 19, Manila, Philippines. www.adb.org/Documents/EDRC/Occasional_Papers/non_samplingerrors.pdf

Spencer, D. S. C. 1989. Micro-level Farm Management and Production Economics Research among Traditional African Farmers: Lessons from Sierra Leone. African Rural Employment Paper No. 3. East Lansing, MI, Michigan State University.

Verma, V., T. Marchant, and C. Scott. 1988. Evaluation of Crop-Cut Methods and Farmer Reports for Estimating Crop Production: Results of a Methodological Study in Five African Countries. London: Longacre Agricultural Development Centre Ltd.

Singh, R. 2003. “Use of Satellite Data and Farmers Eye Estimate for Crop Yield Modeling.” New Delhi: Indian Agricultural Statistics Research Institute. www.gisdevelopment.net/application/agriculture/yield/mi04035p1.htm

Rozelle, S. 1991. Rural Household Data Collection in Developing Countries: Designing Instruments and Methods for Collecting Farm Production Data. Working Papers in Agricultural Economics 91–17. Ithaca, NY: Cornell University.

Poate, C.D. 1988. “A Review of Methods for Measuring Crop Production from Smallholder Producers. Experimental Agriculture 24: 1–14.

Norman, D. W., F. D. Worman, J. D. Siebert, and E. Modiaqgotla. 1995. The Farming Systems Approach to Development and Appropriate Technology Generation. FAO Farm System Management Series 10. Rome: FAO.

Murphy, J., D. J. Casley, and J. J. Curry. 1991. Farmers’ Estimations as a Source of Production Data. World Bank Technical Paper 132. Washington, DC: World Bank.

Fielding, W. J., and J. Riley. 1997. “How Big Should On-Farm Trials Be and How Many Plots Should Be Measured?” PLA Notes 29:19–22. International Institute of Economic Development, London.

Parry M.L., Rosenzweig C., Iglesias A., Livermore M., and Fischer G. (2004). Effects of Climate Change on Global Food Production under SRES emission and socio-economic scenarios. Global Environment Change. http://www.preventionweb.net/files/1090_foodproduction.pdf

Oyiga Benedict Chijioke, Mekbib Haile, and Christine Waschkeit (2011). Implication of Climate on Crop Yield and Food Accessibility in Sub-Sahara Africa. http://www.zef.de/fileadmin/downloads/forum/docprog/Termpapers/2011_1_Oyiga_Haile_Waschkeit.pdf

Richard M. Adams, Brian H. Hurd, Stephanie Lenhart, Neil Leary. (1998). Effects of global climate change on agriculture: and interpretative review. CLIMATE RESEARCH. http://www.int-res.com/articles/cr/11/c011p019

Claudia Ringler, Tingju Zhu, Ximing Cai, Jawoo Koo, Dingbao Wang. (2010). Climate Change Impacts on Food Security in Sub-Saharan Africa. Internation Food Policy Research Institute (IFPRI). http://www.ifpri.org/sites/default/files/publications/ifpridp01042.pdf

Alessandro de Pinto, Ulac Demirag, Akiko Haruna, Jawoo Koo, Marjan Asamoah. (2012). Climate Change, Agriculture, and Food Production in Ghana. http://www.ifpri.org/sites/default/files/publications/gssppn3.pdf
Pedram Rowhani, David B. Lobell, Marc Linderman, Navin Ramankutty. (2011). Climate Variability and Crop Production in Tanzania. Agriculture and Forest Meteorology. http://jatropha.pro/PDF%20bestanden/Climate%20variability%20and%20crop%20production%20in%20Tanzania.pdf

Abera Birhanu Demeke, Alwin Keil, Manfred Zeller. (2010). Using Panel Data to Estimate the Effect of Rainfall Shock on Smallholder Food Security and Vulnerability in Rural Ethiopia. https://www.uni-hohenheim.de/fileadmin/einrichtungen/eaae-fsc/Friday/Demeke_panel_Conference.pdf

Martin Smith, Pasquale Steduto. Yield response to water: the original FAO water production function. http://www.fao.org/docrep/016/i2800e/i2800e02.pdf

Masvaya E.N., Mupangwa W., Twomlow, Rainfall Variability Impacts on Farmers’ Crop Management Strategies. http://www.waternetonline.ihe.nl/challengeprogram/ir34%20masvaya%20rainfall%20strategy.pdf
| CROP | Sorghum IMAE | Sorghum ISMAE | Maize IMAE | Maize ISMAE | Wheat IMAE | Wheat ISMAE | Rice IMAE | Rice ISMAE |
|------|--------------|--------------|-----------|-----------|----------|----------|---------|---------|
| 2002A | 0.0352       | 0.0359       | 0.3751    | 0.3159    | 0.8163   | 1.3794   | 0.5296  | 0.7204  |
| 2002B | 0.0124       | 0.0123       | 0.2580    | 0.2285    | 0.4946   | 0.3966   | 0.0550  | 0.0566  |
| 2003A | 0.1457       | 0.1572       | 0.1553    | 0.1441    | 0.0062   | 0.0062   | 0.0715  | 0.0690  |
| 2003B | 0.0758       | 0.0731       | 0.4054    | 0.5084    | 0.1935   | 0.2142   | 0.3729  | 0.4584  |
| 2004A | 0.3012       | 0.3546       | 0.1341    | 0.1256    | 0.1110   | 0.1175   | 0.0196  | 0.0194  |
| 2004B | 0.1215       | 0.1145       | 1.1939    | 2.9622    | 0.2245   | 0.2018   | 0.0735  | 0.0709  |
| 2005A | 1.0438       | 2.1833       | 0.0044    | 0.0044    | 0.2198   | 0.2470   | 0.0210  | 0.0208  |
| 2005B | 0.0462       | 0.0473       | 0.7394    | 1.1732    | 0.3176   | 0.2740   | 0.0262  | 0.0259  |
| 2006A | 1.3376       | 4.0385       | 0.1580    | 0.1715    | 0.7115   | 1.1043   | 0.0180  | 0.0178  |
| 2006B | 0.0273       | 0.0277       | 0.7951    | 1.3199    | 0.5932   | 0.8434   | 0.0395  | 0.0388  |
| 2007A | 0.0182       | 0.0180       | 0.4622    | 0.3755    | 0.9313   | 1.7428   | 0.0129  | 0.0128  |
| 2007B | 0.0315       | 0.0310       | 0.8257    | 1.4062    | 0.4170   | 0.5269   | 0.0734  | 0.0708  |
| 2008A | 0.0042       | 0.0042       | 0.1245    | 0.1172    | 0.0277   | 0.0273   | 0.0636  | 0.0616  |
| 2008B | 0.0669       | 0.0648       | 0.6781    | 1.0259    | 0.8145   | 0.5788   | 0.1067  | 0.1013  |
| 2009A | 0.2122       | 0.1919       | 0.0419    | 0.0428    | 0.1190   | 0.1123   | 0.0125  | 0.0125  |
| 2009B | 0.0335       | 0.0340       | 0.0606    | 0.0588    | 0.0729   | 0.0704   | 0.0431  | 0.0422  |
| 2010A | 0.3976       | 0.3316       | 0.1153    | 0.1224    | 0.0618   | 0.0599   | 0.0465  | 0.0476  |
| 2010B | 0.0411       | 0.0419       | 0.2056    | 0.1864    | 0.0200   | 0.0198   | 0.0317  | 0.0312  |
| 2011A | 0.2691       | 0.2372       | 0.0096    | 0.0096    | 0.1109   | 0.1051   | 0.0481  | 0.0469  |
| 2011B | 0.1242       | 0.1324       | 0.1833    | 0.1679    | 0.0673   | 0.0696   | 0.0355  | 0.0362  |
| 2012A | 0.2610       | 0.2308       | 0.0298    | 0.0302    | 0.1472   | 0.1371   | 0.0641  | 0.0622  |
| 2012B | 0.2104       | 0.1904       | 0.1716    | 0.1580    | 0.1181   | 0.1256   | 0.0383  | 0.0390  |
| 2013A | 0.5612       | 0.4382       | 0.0027    | 0.0027    | 0.1173   | 0.1247   | 0.0134  | 0.0135  |
| 2013B | 0.0894       | 0.0935       | 0.1013    | 0.0964    | 0.0765   | 0.0796   | 0.0497  | 0.0510  |
| 2014A | 0.7830       | 0.5627       | 0.0604    | 0.0623    | 0.5795   | 0.4493   | 0.0183  | 0.0185  |