

**RE: Data sources**

Annemie Maertens

**Sent:** Monday, March 19, 2018 5:45 PM**To:** Monika Novackova; Pedram Rowhani**Attachments:** SoilGuidebook\_full\_web\_final.pdf (2 MB) ; Geovariables.Description.pdf (110 KB)

Dear Monika,

Thank you for these updates. A bit of feedback below:

- On the soil data: integrating into datasets has been done by the LSMS team in the WB. Now, Kenya does not have an LSMS, but their technique, which uses the same raw data you identified can be used. I attach the datafile for Malawi and a working paper on the topic from the bank. In general, you need to distinguish between generic soil classification (based on colonial maps) and recent results of topsoil test (much more rare, but major progress has been made in the last 5-10 years).
- On the Data that Pedram can obtain. I have already discussed your points (2) and (3) with him - so we are all good on those.
- On the other variables you have identified as missing; I added a source next to each one of them (my best guess):
  - locations of historical railways (or distance from railways) - yes Alex
  - agricultural subsidies - this is national level - see documents from ministry of ag.
  - food aid - from household surveys
  - export and import - not sure - national level I presume?
  - population density - leading econ indicators
  - natality - leading econ indicators
  - mortality (also children under 5, infant) - leading econ indicators
  - life expectancy - unsure
  - education attainment - leading econ indicators
  - access to and use of ICT facilities - household surveys
  - access to social amenities - not sure what is meant
  - costs of production - e.g. seeds, labour, chemicals/fertilizers, maintenance of farm buildings, motor vehicles and farm equipment used for farm business - some in leading econ indicators - others in household surveys

In addition to the DHS (note all rounds are freely downloadable via the DHS website), there is the KIBHS, which seems to be very close to the LSMS. These data are also available.

Annemie

Annemie Maertens

Senior Lecturer, Department of Economics

Jubilee Building 260, University of Sussex

Falmer, Brighton

BN1 9SL, UK

Phone: 44 (0)1273 678840

E-mail: A.Maertens@sussex.ac.uk

Homepage: <https://sites.google.com/site/maertensannemie/>

---

From: Monika Novackova

Sent: 18 March 2018 20:56

To: Pedram Rowhani; Annemie Maertens

Subject: FW: Data sources

Dear Pedram and Annemie,

I have done some additional data research to update what data we may need, see if I would be able to find some and mainly if there is something that Pedram could potentially try to get in Kenya (see bottom of this email). I was hoping to do this in couple of hours, but unfortunately it took me a bit longer (just going through all the publications of KNBS takes a long time), so I am sorry for sending this email later than I expected. But I have some interesting findings.

A good news is that it seems that the soil data should be available - see this link:

<http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/>

However, for me to be able to actually use them, it would probably require substantial data work first. I would need to read the manual carefully (see attached), and then I could probably manipulate the data using GIS and I may also need to use MS Access (the data is in mdb format). However, I think that this would be worth doing, perhaps after I finish the models of nutrition outcomes and basic first models of production. Most of the production studies I have read include soil data - they are relatively important controls...

While looking for the additional data at the webpage of KNBS, I found potentially interesting county level data from surveys or censuses. I realised that Annemie mentioned them in the email that she sent from Kenya, but I didn't have time to go through them properly as I was busy with the NDMA reports, market data, coding, DRiSL meeting and other... The drawback of most of the KNBS data is, that they are only available for one or two years (some of them for five or six years usually between 2010 and 2016), but perhaps they could still be included in some models as fixed effect or we could at least use them for comparison with the data from the NDMA reports. The county level data I found are:

Economic Survey: - these data are mostly for years between 2012-2016

- student enrolment by sex and university (this would need to be matched to counties as there is not a University in each county, at least not with data in this report)
- full immunisation coverage
- registered deaths
- fertility rate, number of births
- number of health institutions
- annual county government revenue
- contraception methods
- household possession of insecticide treat nets
- number of registered voters

KDHS - Kenya Demographic Health Survey: - these data are just for year 2014, maybe also for 2003

- population: rural, urban, total
- height for age, weight for age, weight for height, BMI,
  - presence of iodised salt in household
  - child vaccination
  - fertility, maternity rate
  - children under 5, who are underweight
  - percentage of people who slept under mosquito net last night (including percentage of children under 5)
  - percentage of children with fever and their treatment
  - knowledge of AIDS and application of HIV prevention methods, testing status, risking sexual behaviour,..(men and women separately)

Housing Survey Report 2012-2013:

- Number of Water service providers in a county providing informal settlements with water
- Total cubic meters of water dispensed to residential areas from water companies, 2010-2011
- Household headship by sex and county ,also by age, economic activity
- Household members by economic activity
- Households by rent paid per month
- Roof material, floor material, wall material
- Reported Commercial banks deposits, loans and housing loans

Statistical Abstract (this seems to be something else than the County Statistical Abstracts)

2016

- Land area, water area, population by sex, population density - , all at county levels but just for 2009. these are actually from the housing census 2009
- Highest level of education from Housing census 2009
- reported births and deaths for years 2011-2015
- mortality indicators for 2009
- Outpatient Morbidity for Persons of Age 5 Years and Above by County, 2015 by reason of death (may need some data manipulation-data formatting)

- Primary and secondary School enrolment by county for the years 2010-2016
- Annual County Governments Revenue, FY 2014/15 - 2015/16
- Consolidated County Governments Budget Allocation, FY 2015/16
- Economic classification of county government expenditure, at the end of 2014/15
- Categories of Agricultural land, 2013 (High potential, Medium potential, low potential). This seemed to be very promising as proxy for the soil characteristics, but then I realised that it is probably only based on annual cumulative rainfall
- Registered voters by county and sex 2015

2007 (I look into some random years)

- municipal counties percentage expenditure on main services 2004/05 and 2005/06
- wage employment 2001 - 2006 but it would require data manipulation as it is just copied from a data book

Population and density per counties in the Population and Housing Census, but these just one in 10 years

Data, which would be useful and could potentially be obtained in Kenya:

1. At the moment, we only have MUAC, CSI, and the Early warning phase which I digitalized from the NDMA reports for the 23 ASAL counties, starting at year 2014. It would be good to get them for the other counties (the Phase classification may not be available for the counties which are not ASAL, but maybe we could at least get MUAC and CSI for the remaining counties?) and for the previous years.

2. Leading economic indicators from KNBS these are the price data I have downloaded and digitalized from year 2006 onwards. They are not online for previous years, but should be available in the KNBS library. Citing from the document that Annemie sent us 23.January:

'Leading economic indicators: This is the coolest thing ever. This is a monthly publication by the Central Bureau of Statistics. I am unsure of the start date - this was an entire bookcase and they were not in order - but a random selection revealed that this seems to go back to the 90s at least. While they focus on macro-economic indicators, they have the market data you are looking for: Monthly market data for the various agricultural outputs (I spotted maize, beans for 43 market towns). This is available online from 2007 onwards. Copying them for earlier years would likely take about half a day if we were to do this ourselves'

3. County Statistical Abstracts (at the webpage of KNBS (<https://www.knbs.or.ke/publications/>)) include data on livestock population and livestock production. But the data are only for 2013 and 2014 and also there is a pdf file for each county so it would need to be digitalised. But based on the email from Annemie (sent 23.Jan 2018) these data could be available for previous years in County Development Profiles (precursors of the County Statistical Abstract) which should be in the KNBS library. This would, however require manual data entry from the library of KNBS.

With regard to the list of data which could possibly used as controls below: I don't think that we need to have all or most of it. But if it is possible to get them as a county level panels, it would be useful. They could help us to get more precise estimates (smaller standard errors) and reduce the risk of confounding (ascribing effect of something else to an incorrect factor). However, if we don't get them it should not be a big problem unless it is strongly correlated with both dependent variable and some of our main explanatory variable of interest.

- household income/ GDP or even better per capita GDP (I have made an official request via the website of KNBS and also the World Bank and the Central Bank of Kenya, but I am not sure if I will get a useful response)
- locations of historical railways (or distance from railways) ?? - this is my random idea - in some studies they use it as an instrument for GDP. I could also ask Alex Moradi if he could point us where to find them, or even share them
- agricultural subsidies (for example subsidies for fertilisers. But I am not sure if this is relevant as I don't know what is the situation with agricultural subsidies in Kenya)

- food aid
- export and import
- population density
- natality
- mortality (also children under 5, infant)
- life expectancy
- education attainment
- access to and use of ICT facilities
- access to social amenities
- costs of production - e.g. seeds, labour, chemicals/fertilizers, maintenance of farm buildings, motor vehicles and farm equipment used for farm business. These variables would allow for the replication of Deschenes and Greene (2007), which I find interesting. I imagine that obtaining these data at county level over reasonable time period and with reasonable frequency would be difficult. So having these would be ideal situation but we can estimate other models without these

See you soon. Pedram, have a good stay in Kenya.

Monika