

Voedselprijzen en vluchtelingen;
Een analyse van Afrika.

Chris Al Gerges, Diederik Salimans, Dion Custers, Geert Rien Bakker

28 Juni 2018

1 Introductie

Het World Food Programme, ookwel bekend als het WFP, is het onderdeel van de Verenigde Naties die zich met name bezig houdt met het verstrekken van voedsel aan ontwikkelingslanden en gebieden met veel vluchtelingen. Hiernaast houden ze maandelijks de Global Food Prices Database bij. Deze dataset bevat voedselprijzen van 76 verschillende landen die voedselhulp ontvangen van het World Food Programme en loopt bij sommige landen terug tot 1992. Er bevindt zich voornamelijk veel data over de landen in het continent Afrika. De prijzen lijken te fluctueren en makkelijk beïnvloedbaar te zijn. Voor het project Data Analysis and Visualization is er gekozen om de voedselprijzen in Afrika tussen 2001 en 2016 uit de Global Food Prices Database te analyseren en te visualiseren.

Het doel van dit verslag is het in kaart brengen van de effecten van vluchtelingenstromen op de voedselprijzen in kwetsbare markten en zou wellicht kunnen bijdragen aan het verbeteren van voedselhulp in deze kwetsbare gebieden.

Ten eerste, gaat er gekeken worden of er duidelijke negatieve of positieve correlaties zijn tussen de prijzen van verschillende producten binnen landen. De verwachting is dat er bij vergelijkbare producten, denk bijvoorbeeld aan een set van alle graanproducten of een set van alle soorten rijst, een correlatie te vinden zal zijn. Ook is de verwachting dat de prijzen van ingrediënten van een product correleren met de prijs van het product zelf.

Ten tweede, gaat er gekeken worden of er correlatie is in de verschillen in prijs in specifieke regio's. De verwachting is dat de correlatie tussen de prijzen van producten en de prijzen van diezelfde producten in de buurlanden en lokale regio's hoger is dan de correlatie tussen diezelfde producten in landen die verder weg van elkaar liggen. Mochten er duidelijke outliers gevonden worden, dan wordt getracht verklaring hiervoor te vinden.

Ten derde, gaat er gekeken worden of de prijzen gecorreleerd kunnen worden aan de dataset van de Verenigde Naties over vluchtelingenstromen. In deze dataset wordt per jaar aangegeven hoeveel vluchtelingen er een land in en uit zijn gegaan en waar deze vluchtelingen vandaan zijn gekomen. Net zoals de Global Food Prices Database bevat ook deze dataset vooral veel data over landen in Afrika. De verwachting is wanneer er veel vluchtelingen een land binnenstromen, dat gekenmerkt zal zijn door een stijging in de prijzen van de producten in dat land. Daarnaast is de verwachting ook dat het omgekeerde te zien is, namelijk, als er veel vluchtelingen een land uit gaan, dat de voedselprijzen van dat land zullen dalen.

2 Methode

De drie datasets die gebruikt zijn voor dit onderzoek, zijn de "United Nations Refugee Dataset" (UNRD) (data.un.org, 2018), de "Currency Exchange Dataset" (CED) (bis.org, 2018) en de World Food Programme Food Prices dataset (WFPFP) (data.humdata.org, 2018). Deze drie datasets zijn door ons aangepast en opgeschoond om de data te kunnen gebruiken.

Om de datasets te bewerken is gebruik gemaakt van Python 3 code en LibreOffice. Om de data te visualiseren is gebruik gemaakt van Bokeh en Plotly packages, in combinatie met Python 3 code. De volgende veranderingen zijn aan alle datasets aangebracht.

Allereerst is er gekozen om alle data voor voor 2001 en na 2016 weg te halen. Dit is gedaan omdat er voor veel landen weinig data was van voor 2001 en het bij nadere inspectie duidelijk werd dat de data van de UNRD minder accuraat was naarmate je verder het verleden in ging. Er was initieel ook voor het 2001 afkappunt gekozen omdat we wilden focussen op het midden-oosten, waar in begin 2001 de Afghaanse oorlog begon.

Om significantie te waarborgen is er geopteerd om alleen gegevens te gebruiken waar minstens 30% van de data aanwezig is. Er is gekozen voor een minimum van 30% van aanwezige data omdat hierdoor er een goede balans ontstaat tussen de hoeveelheid van werkbare data en de zuiverheid van deze data. Dit betekent dat in de tijdreeks van 2001 tot 2016 er alleen data is gebruikt waar meer dan 4 jaar aan data voor aanwezig was ($p = 31.25\%$). Verder, zijn alleen de jaren waarin minstens 4 maanden aan data aanwezig was ($p = 33.33\%$) gebruikt om om te zetten in jaarlijkse data.

Daarnaast is er gekozen om alleen data te gebruiken van de jaren die in alledrie de datasets bruikbaar waren. Zo is Somalië eruit gehaald, omdat er geen accurate data over de valuta aanwezig was in de CED.

2.1 WFPFP dataset

Om de data per dataset bruikbaar te maken, zijn er ook individuele aanpassingen gemaakt. Allereerst is er gekozen om de WFPFP niet regionaal maar nationaal weer te geven. Hiervoor is gekozen omdat andere datasets ook op een nationaal niveau zijn weergegeven. De WFPFP dataset is van regionaal naar nationaal omgezet door de maandelijksse prijzen van de product van de regio's op te tellen en te delen door het aantal regio's van het betreffende land. Zo is er tot een nationale prijs per maand gekomen.

Vervolgens is er gekozen het gemiddelde te nemen van de maandelijksse data en om te zetten in jaarlijkse data. Daarbij is een kolom toegevoegd die de procentuele verandering ten opzichte van het voorgaande jaar laat zien. Zo kan er eenvoudig gekeken worden naar relatieve verhogingen of verlagingen in prijzen.

Daarbij, zijn de prijzen in de WFPFP dataset omgezet naar US Dollars doormiddel van de gemiddelde jaarlijkse wisselkoers van alle lokale valuta om te zetten naar de gemiddelde jaarlijkse waarde van de US Dollar. Deze data is gehaald uit de CED. Dit is gedaan om ervoor te zorgen dat de voedselprijzen gelinkt zijn aan een stabiele valuta en om inflatie te relateren aan één specifieke valuta. Hierdoor kunnen uitspraken gedaan worden over de waarden van de producten, ondanks dat de prijzen in verschillende valuta zijn uitgedrukt in de originele dataset. Bovendien is de dollar de meest geaccepteerde en gebruikte valuta ter wereld en geeft een duidelijker referentiepunt voor een prijs voor mensen die niet bekend zijn met de oorspronkelijke valuta.

2.2 United Nations Refugee Dataset

De UNRD is vergelijkbare bewerkingen ondergaan. Zo is er gekeken wat de jaarlijkse totale instroom en uitstroom is van refugees van alle landen, zodat er makkelijk gekeken kon worden waar veel activiteit was. Ook hier is er een kolom toegevoegd die de procentuele verandering ten opzichte van het voorgaande jaar laat zien.

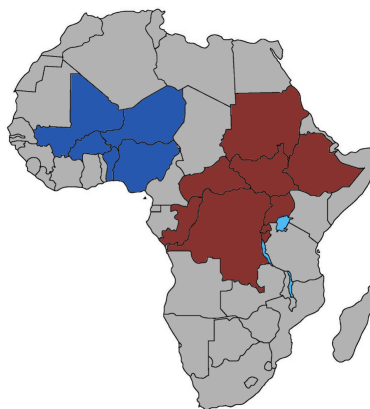
Tijdens de Exploratory Data Analysis is er voor gekozen om te focussen op het continent Afrika, omdat daar de meeste data gecentreerd was en de data het meest bruikbaar was. De periode 2001-2016 als afkappunt is aangehouden, wegens de andere eerder gegeven redenen.

Elk land vormt samen met zijn aangrenzende buurlanden de regio van dat land. Met uitzondering van Djibouti daar is naast Ethiopië ook Yemen aan de regio van toegevoegd, omdat deze landen enkel gescheiden worden door een zeestraat van 27 kilometer breed. Dit was voor ons een dusdanig kleine afstand dat er besloten is Yemen bij de regio van Djibouti te betrekken. Yemen wordt echter niet in de rest van Data meegenomen of gevisualiseerd omdat het niet in Afrika ligt maar in het Midden Oosten. Van de andere buurlanden van Djibouti (Eritrea en Somalië) was er onvoldoende data beschikbaar.

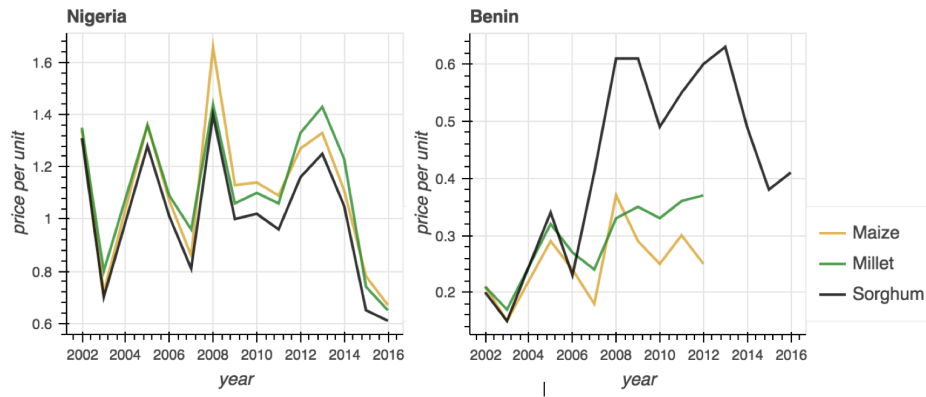
Verder is er gekozen om twee extra regio's te definiëren. Zo is er gekozen voor de regio bestaande uit Nigeria, Niger, Mali, Benin en Burkina Faso (blauw) en voor de regio bestaande uit de Democratisch Republiek van de Congo, Soedan, Zuid Soedan, Ethiopië, Oeganda, Rwanda, Burundi en Centraal Afrikaanse Republiek (rood), beide te zien in [figuur 1].

Er is gekozen voor deze twee regio's omdat tijdens de exploratory data analysis van de WFPFP er een sterke correlatie leek te zijn tussen de de voedselprijzen van drie specifieke producten, namelijk "Maize" (maïs), "Sorghum" (sorghum gierst) en "Millet" (pluimgierst), in de blauw aangegeven regio. Om een duidelijk beeld te krijgen van de stabiliteit van de prijzen in deze landen is er besloten als aparte regio te beschouwen en uitgebreid te analyseren. Ook is er voor gekozen om te focussen op deze drie graanproducten, aangezien deze vergelijkbaar zijn in hoe ze geteeld en gebruikt worden. De tweede regio is gedefinieerd naar aanleiding van de exploratory data analysis van de UNRD. Het viel op dat de in- en uitstroom van vluchtelingen in deze regio erg groot was, vandaar dat er gekozen is om deze regio nader te onderzoeken.

Tijdens het onderzoek is er geen gebruik gemaakt van lineaire regressie, clustering of t-SNE. Hiervoor is gekozen omdat deze methodes geen nieuwe inzichten zouden geven of omdat de data niet voldoende geschikt was. Voor lin-



Figuur 1: De besproken regio in Afrika



Figuur 2

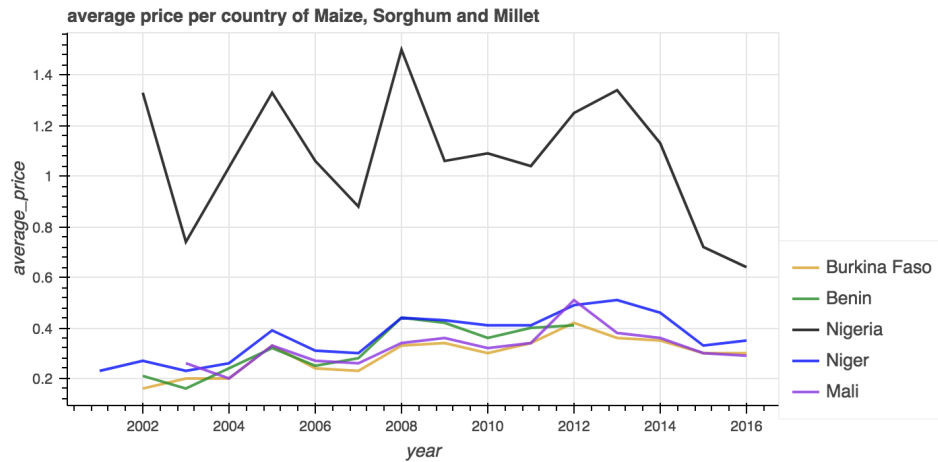
aire regressie waren er te weinig individuele datapoints per grafiek, aangezien we met jaarlijkse data werken. Zo konden er in optimale omstandigheden maximaal 16 datapoints worden gebruikt per grafiek. Als de data dan ook nog gesplitst zou moeten worden tussen trainingsdata en testdata, zouden de resultaten zeer onbetrouwbaar zijn wegens de kleine dataset. Clustering is niet gebruikt aangezien er maar één specifieke grafiek was waar het mogelijk was om deze techniek te gebruiken en bij nadere inspectie werd het duidelijk dat er hieruit maar één significante cluster gevormd zou worden. Aangezien deze cluster zich rond de oorsprong van de grafiek vormt (zoals te zien is in [figuur 5]), waar niets significant van geconcludeerd kan worden, leek het overbodig om dit nader te onderzoeken. T-SNE is niet gebruikt omdat er geen gebruik is gemaakt van hoger-dimensionale data.

3 Resultaten

Om te kijken of er duidelijke negatieve of positieve correlaties zijn tussen de prijzen van verschillende producten, is er voor gekozen om te focussen op de regio van Nigeria, Niger, Mali, Benin en Burkina Faso. Om een duidelijk beeld te krijgen van de stabiliteit van de prijzen in deze regio is er gekeken naar drie specifieke producten, namelijk "Maize", "Sorghum" en "Millet".

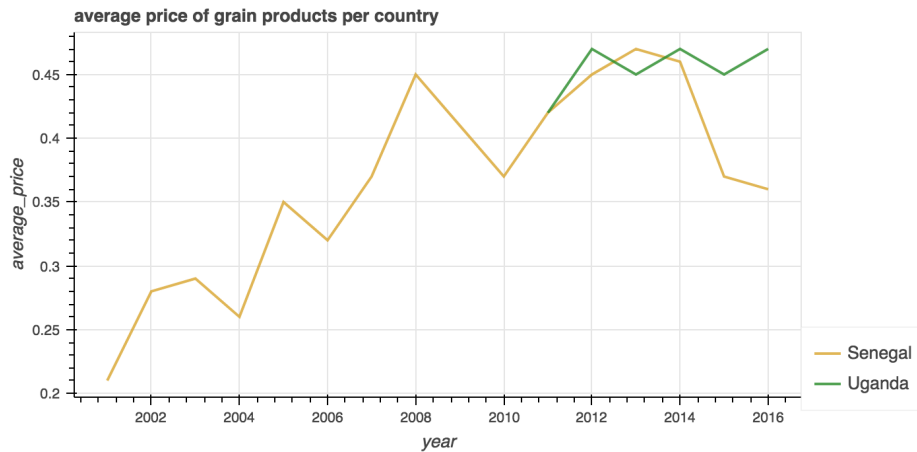
In [figuur 2] is te zien dat de prijzen van deze producten erg sterk gecorreleerd zijn binnen Nigeria, met een gemiddelde correlatie van 0,97. Het land waar deze producten het minst sterk gecorreleerd waren is Benin, waar een gemiddelde correlatie van 0,79 aanwezig is. Als de gemiddelden genomen worden van deze drie producten, dan zien we in [figuur 3] een lagere, maar nog steeds significante correlatie binnen de gehele regio, namelijk een gemiddelde correlatie van 0,48, met als laagste correlatie tussen de voedselprijzen van Nigeria en de voedselprijzen van Mali, waartussen een correlatie van -0,17 ligt.

Opvallend is, dat ondanks de significante correlatie er een groot verschil is tussen de absolute waarden van de prijs in Nigeria relatief aan de andere landen. Als Nigeria uit de regio wordt gehaald, en de correlatie tussen de voedselprijzen van de landen wordt herberekend, dan word er een correlatie van 0,69 gevonden.



Figuur 3

Om te kijken of er significante correlaties te vinden zijn in de verschillen in prijs in specifieke regio's, is het van belang te kijken of er een significant verschil is tussen de correlatie van de productprijzen van een regio en van de productprijzen van landen die niet bij elkaar in de buurt liggen. Hierboven is al beschreven hoe de voedselprijzen binnen een regio met elkaar correleren. Om twee landen die ver weg van elkaar liggen te vergelijken, is er gekeken naar de correlatie tussen de gemiddelde prijs van "Maize", "Sorghum" en "Millet" in Senegal en in Oeganda. De correlatie van de gemiddelde prijzen tussen deze landen is 0,03, wat terug te zien is in [figuur 4] hieronder.

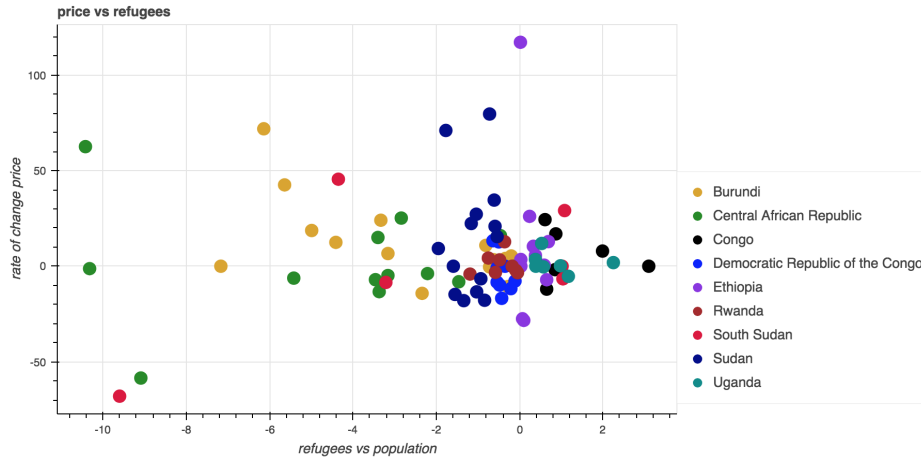


Figuur 4

Deze grafiek laat de data zien van Senegal van 2001 tot 2016, maar van Oeganda van 2011 tot 2016. Om de correlatie te berekenen, zijn alleen de jaren waar beide landen data voor hebben meegerekend. Deze correlatie komt uit op 0,03.

Om te kijken of er correlaties te vinden zijn tussen de voedselprijzen en vluchtelingenstromen, is ervoor gekozen om te focussen op een andere regio in Afrika, namelijk de regio van de Congo, de Democratisch Republiek van de Congo, Soedan, Zuid Soedan, Ethiopië, Oeganda, Rwanda, Burundi en Centraal Afrikaanse Republiek. Deze regio is gekozen omdat deze landen een grote instroom of uitstroom van vluchtelingen had.

In [figuur 5] wordt de procentuele verandering per jaar van de gemiddelde prijs van alle producten in een land tegenover het percentage van de vluchtelingenstroom relatief aan de populatie van het desbetreffende land weergegeven. Er is geen duidelijke correlatie te vinden, aangezien er “outliers” aan alle kanten van de grafiek bevinden. Wat wel opvallend is, is dat het voornamelijk rond het nulpunt van beide assen geclusterd lijkt te zijn.



Figuur 5

Daarnaast, is er gekeken naar de correlatie tussen de procentuele verandering in de gemiddelde voedselprijzen van alle producten per land en de procentuele verandering in de in- of uitstroom van vluchtelingen. De hoogste correlatie was te vinden in Benin, waar een correlatie van 0,69 was, met als laagste in Liberia, waar een correlatie van -0,96 was. Opvallend is, dat het gemiddelde van de correlaties van alle landen -0,04 is.

[4]

4 Discussie

Wegens de grootte van de dataset en van het onderzoeksterrein zijn er natuurlijk veel correlaties te vinden waar niks over te concluderen valt. Dat betekent dat de uiteindelijke relevantie van onze conclusies zeer beperkt zal zijn, met betrekking tot de specificiteit van de geanalyseerde data. Zo kan de vraag "zijn er duidelijke negatieve of positieve correlaties tussen de prijzen van verschillende producten" op duizenden verschillende manieren worden beantwoord. Dat gezegd hebbende, zijn er wel interessante dingen naar voren gekomen bij de analyse van deze datasets.

Zo zijn er wel degelijk landen waar een erg sterke correlatie te vinden is tussen de prijzen van producten. In alle individuele landen, met uitzondering op Benin, in de regio van Nigeria, Niger, Mali, Benin en Burkina Faso, zijn er significante correlaties naar voren gekomen over de prijzen van "Maize", "Sorghum" en "Millet". Dit komt overeen met onze verwachting, aangezien deze drie producten vergelijkbaar zijn in hoe ze geteeld worden en hoe ze gebruikt worden.

Opvallend is, dat ondanks de significante correlatie er een groot verschil is tussen de absolute waarden van de prijs in Nigeria. Als bij nader onderzoek Nigeria uit de regio wordt gehaald, en de correlatie tussen de voedselprijzen van

de landen wordt herberekend, dan word er een correlatie van 0,69 gevonden, in tegenstelling tot een correlatie van 0,48, wat een significant verschil is. Een mogelijke verklaring hiervoor kan zijn dat Nigeria een vele malen hogere GDP per capita heeft dan de andere landen uit de regio, wat een effect kan hebben op de regionale handelsmarkten.

Daarnaast is er in de algehele regio ook een sterke correlatie tussen de gemiddelden van deze prijzen te zien. Daarentegen is te zien dat in verre landen, namelijk Senegal en Oeganda, er een significant lagere correlatie is tussen vergelijkbare producten. Een voor de hand liggende verklaring hiervoor is dat landen binnen een regio een meer uitgebreid handelsnetwerk hebben met elkaar dan met verre landen dat met elkaar hebben. Dit ondersteunt de verwachting dat de landen binnen lokale regio's een sterke correlatie hebben in voedselprijzen.

Een valide kritiekpunt op het hierboven genoemde argument is dat er maar twee verschillende landen zijn vergeleken, namelijk Senegal en Oeganda. Er is hiervoor gekozen omdat er binnen de data geen andere landen waren die ver weg van elkaar lagen en die data had van dezelfde of vergelijkbare producten van dezelfde jaren.

Daarnaast moet er ook rekening gehouden worden met het feit dat de WFP dataset alleen data bevat van landen waar voedselhulp wordt aangeboden. Wat voor effect voedselhulp heeft op voedselprijzen van lokale markten is onduidelijk.

Verder, is er vrijwel niets significant te concluderen over de relatie tussen vluchtelingen en voedselprijzen. Zo is er geen duidelijke correlatie te zien in de procentuele verandering per jaar in prijs, gerelateerd aan procentuele verandering aan vluchtelingen of aan vluchtelingen als functie van de huidige lokale populatie. Dit gaat tegen onze verwachtingen in, aangezien er geen bewijs gevonden kan worden voor enige vorm van significante correlatie.

Het is onduidelijk waarom er geen correlatie te vinden is, maar een mogelijke verklaring is dat er alleen gekeken is naar de stroom van vluchtelingen vanaf 2001 tot 2016. Dat betekent dat er geen data is meegenomen over de vluchtelingen die zich al bevonden in andere landen van voor 2001. Ook is er geen data gebruikt over de reguliere migratie en is er geen rekening gehouden met de reden waarom mensen vluchten.

5 Bronnen

- Currency Exchange Dataset. (2018). Verkregen van www.bis.org/statistics/xrusd
- World Food Programme Global Food Prices Dataset. (2018). Verkregen van <https://data.humdata.org/dataset/wfp-food-prices>
- United Nations Refugee Dataset. (2018). Verkregen van <http://data.un.org/Data.aspx?d=UNHCRf=indID%3AType-Ref>