

Een analyse van correlaties tussen de productprijzen in ontwikkelingslanden

Luc Weytingh, Pim Meerdink, Jeroen Jagt, Andreas Berentzen

1. Inleiding

In de westerse wereld wordt rond de 10% van het huishoudensinkomen uitgegeven aan voedingsmiddelen. Dit staat in contrast met ontwikkelingslanden waar al snel meer dan de helft van het inkomen uit wordt gegeven aan eten.¹ Fluctuaties in de prijs van voedingswaren kunnen hierdoor grote gevolgen hebben in ontwikkelingslanden. Fluctuerende food prices hebben effecten op de lange en korte termijn. Korte termijn gevolgen zijn hongersnoden onder de armsten, lange termijn effecten zijn ernstiger voor de economie. Als kosten van geïmporteerde voedingsmiddelen stijgen, leidt dit tot een daling in handel en activiteit binnen de economie. Dit kan weer leiden tot werkloosheid en daling in welvaart.² Doordat voedselprijs fluctuaties een groot effect hebben op voornamelijk de armste wereldburgers is het van groot belang dat we grip kunnen krijgen op deze ontwikkelingen. Met behulp van de *'prevalence of undernourishment database'* van de Food and Agriculture Organization (FAO) proberen we beeld te krijgen van het verband tussen prijsfluctuaties en ondervoeding.

In dit onderzoek analyseren we de productprijzen van de afgelopen jaren aan de hand van de dataset 'global food prices database'³ van de World Food Programme (WFP) en beantwoorden daarmee de volgende onderzoeksvragen:

1. a. Zijn er productprijzen die een positieve/negatieve relatie tonen en bestaat deze correlatie door de jaren heen of alleen in een bepaalde periode?
b. Is het mogelijk om de ingrediënten van een product te detecteren in de prijzen van een ander product?
2. Laten landen in dezelfde regio's, vergelijkbare prijsverschillen zien? Is hier een verklaring voor te vinden?
3. Welke correlaties kunnen worden gevonden tussen de food price data en ondervoeding?

June 28, 2018

De hypothese voor de eerste deelvraag is dat er significante positieve en negatieve relaties bestaat tussen producten in de dataset. Wat betekent dat in de data tussen meerdere producten vergelijkbare of tegenovergestelde prijsfluctuaties te zien zullen zijn. We verwachten ingrediënten van producten te kunnen detecteren in hun prijzen. Er zijn ingrediënten van voedingsmiddelen die essentieel zijn voor de productie van dat voedingsmiddel (bijvoorbeeld graan en brood) zodat de prijs van het ingrediënt een sterk effect zal hebben op de prijs van het product. We verwachten ook duidelijke negatieve correlaties te zien tussen bijvoorbeeld producten die in tegenovergestelde seizoenen op de markt komen.

De hypothese voor de tweede deelvraag is dat landen in dezelfde regio's vergelijkbare prijsverschillen laten zien. Landen in dezelfde regio zijn vaak afhankelijk van gedeelde handelsroutes en importprijzen, hierdoor zullen in deze landen vergelijkbare fluctuaties in prijs te zien zijn. Bovendien hebben landen in dezelfde regio grofweg dezelfde ligging waardoor zij vaak beïnvloedt worden door dezelfde non-economische factoren (bijvoorbeeld vluchtelingenstromen of natuurgeweld). Dit zal dus inhouden dat voor een product er vergelijkbare prijsfluctuaties zullen zijn in landen die zich in dezelfde regio bevinden.

De hypothese voor de derde deelvraag is dat er niet een duidelijke correlatie gevonden kan worden tussen de *food prices* data en de ondervoeding data. Ondervoeding en product prijzen zijn twee ingewikkelde concepten die fundamenteel verwant zijn. Toch zijn er vermoedelijk te veel externe factoren zijn om duidelijke correlaties te kunnen detecteren.

2. Methode

De gebruikte *food prices* dataset bestond uit $\pm 783,000$ rijen. Hierbij representeert elke rij de prijs van een product per maand. Origineel bestond het bestand uit achttien kolommen. De kolomnamen zijn aangepast om de semantische waarde van de kolom naam af te stemmen op de kolom inhoud. Per rij werd voor een product het land, de markt, de valuta, de maand, het jaar en de eenheid gegeven.

2.1. Pre-Processing

Veel datasets zijn incompleet, onoverzichtelijk en chaotisch. Voordat er nuttige analyses gedaan kunnen worden met de dataset moet een dataset eerst 'gecleaned' worden. Dit houdt in het aanpassen of verwijderen van incomplete data en het aanpassen van de data zodat er efficiënter mee kan worden gewerkt.

Naast het *food prices* bestand zijn er nog twee externe datasets gebruikt. Als eerste het *US dollar exchange rates* bestand van de Bank for International Settlements.⁴ Dit is een comma-separated values (csv) bestand dat per maand, per jaar en per kwartaal de wisselkoers van 119 economieën beschrijft. De wisselkoersen worden allemaal beschreven ten opzichte van de USD (United States Dollar) en dit kon gebruikt worden om onze *food prices* te normaliseren zodat er producten uit verschillende landen vergeleken konden worden. Is niet door ons *gecleaned*, de dataset was dermate compleet dat er direct mee gewerkt kon worden.

Als tweede externe dataset is er de *Prevalence of Undernourishment Database* van het World Bank gebruikt.⁵ Dit is een csv bestand waar gegevens over de *prevalence of undernourishment* (het voorkomen van ondervoeding) van 150 verschillende landen over de afgelopen 25 jaar in staan. De *prevalence of undernourishment* beschrijft hoeveel procent van de bevolking van een bepaald land ondervoed is. Deze dataset is niet door ons *gecleaned*. Er was wel een probleem met de overeenkomst tussen de namen van landen. Congo stond bijvoorbeeld in ons *undernourishment* bestand als simpelweg 'Congo', maar in ons *food prices* bestand stond data over hetzelfde land onder 'Democratic Republic of the Congo'. Dit is opgelost door een dictionary toe te voegen die de namen met elkaar verbindt. Hiernaast stonden sommige van de landen uit *food prices* niet in het *undernourishment* bestand. Er stonden wel data over regio's zoals 'Sub-Saharan Africa', dus bij ontbrekende landen konden we de data van hun regio gebruiken.

De *food prices* dataset is complex en onoverzichtelijk. Er is weinig consistentie wat betreft eenheden en specificiteit wat betreft het product tussen landen. Vermoedelijk is dit gekomen doordat de data handmatig door verschillende personen is ingevoerd. Hierdoor zijn de volgende stappen ondernomen om de data te *cleanen*.

2.1.1. Verwijderen

Aangezien we beschikking hadden over heel veel data is er de keuze gemaakt om sommige data niet te gebruiken. Ten eerste zijn alle rijen over Somalië zijn verwijderd. Deze gebruikten een valuta waarvan de wisselkoers niet in ons *exchange rate* bestand was opgenomen en er was geen betrouwbare bron te vinden waar deze wisselkoers wel beschreven werd. Ten tweede was er veel data wat in onbruikbare of onduidelijke eenheden was beschreven, bijvoorbeeld de 'cuartilla'. Een eenheid van Spaanse afkomst wat volgens verschillende bronnen verschillende hoeveelheden beschreef. Er waren in de dataset meerdere van deze ambigue en onduidelijke eenheden zoals: 'Packet', 'Package', 'Course', 'MT' en 'Sack'. De data met deze eenheden is daarom verwijderd. Ook waren er in sommige landen data over producten te vinden die nergens anders beschreven werd. Over Turkije, bijvoorbeeld, was data

opgenomen over de kost van elektriciteit in kWh. Nergens anders in de dataset was er informatie over elektriciteitsprijzen te vinden, wegens overbodigheid hebben we dit soort data ook verwijderd. Ten slotte zaten er in de *food prices* dataset wisselkoersen van lokale valuta (LCU) Dit is ook wegens overbodigheid de dataset verwijderd omdat we een externe dataset hadden met wisselkoersen.

2.1.2. Normaliseren

Aan onze *food prices* dataset is een kolom toegevoegd met alle prijzen omgezet in USD met behulp van het *exchange* bestand. Dit is zodat er één centrale valuta was waar mee kon gerekend worden. Er zijn veel eenheden genormaliseerd naar één centrale, dit is handig om prijzen tussen landen en markten te vergelijken in verschillende hoeveelheden. Soms kwam informatie over de prijs van rijst bijvoorbeeld per 5 KG en soms per 2 KG. Er is voor gekozen om deze allemaal om te zetten naar 1 KG voor overzichtelijkheid en gemak in het plotten en vergelijken van de data. Op deze manier zijn alle massa's naar kilogram, alle volumes naar liters en alle hoeveelheden naar 'unit' omgezet.

Er kwamen vaak ook varianten van één product voor, bijvoorbeeld 'Rice (milled 80-20)' en Rice (denikassia, imported). Deze zijn allemaal omgezet naar simpelweg 'Rice'. Op deze manier zijn van 331 producten, 123 producten gemaakt. Zonder deze aanpassingen zou veel data niet vergeleken worden, gezien het dan als een ander product wordt gezien. Door deze normalisatie uit te voeren gaat er informatie over verschillende soorten van een product verloren, maar zonder deze aanpassing is de data te divers om mee te kunnen werken.

2.1.3. Toevoegen

Om de onderzoeksvragen te beantwoorden was het nodig om vergelijkingen te maken tussen de datasets. Hiervoor was het het makkelijkst om kolommen toe te voegen aan onze *food prices* dataset waar de benodigde data in werd opgenomen. Naast de kolom met de prijs van alle producten omgezet tot USD zijn ook twee kolommen toegevoegd met informatie over de regio waar een land zich in bevindt. De eerste toegevoegde kolom, genaamd 'region_name', geeft de naam van de regio waar het land in ligt (bijvoorbeeld 'Sub-Sahara-Afrika'). De tweede is een cijfer corresponderend aan deze regio. Dit is makkelijker en efficiënter computationeel gezien dan werken met de strings in 'region_name'. Ook is er een kolom toegevoegd met de data uit de *prevalence of undernourishment* dataset. Deze geeft per rij de ondervoeding-index aan.

2.1.4. Packages

Voor het verwerken van de data zijn meerdere *packages* gebruikt. Ten eerste is er gebruik gemaakt van Pandas, een module binnen python gemaakt voor het werken met databestanden (csv). Daarnaast is er gebruik gemaakt van numpy voor het berekenen van statistieken over onze data en is er gebruik gemaakt van bokeh voor het plotten van de data. Ten slotte is er door middel van Chartkick een geografische weergave gemaakt van de data.

2.2. Data Analysis

2.2.1. Pearson correlatie-coëfficiënt (r)

De Pearson correlatie-coëfficiënt is een maat voor de lineaire correlatie tussen twee variabelen. Dit is een getal tussen -1 en 1, waarbij $r < 0$ een negatief verband aanduidt, en $r > 0$ een positief verband. Bij 0 is er geen verband aanwezig. Hoe sterker het verband, groter de absolute waarde van r (naar 1). Er is in dit onderzoek een implementatie gebruikt waarbij r wordt uitgerekend voor het verband tussen twee lijsten van dezelfde lengte. In dit onderzoek is voornamelijk de correlatie tussen prijzen van producten of prijsgemiddelden van producten op meerdere markten berekend, met uitzondering van vraag 3, waar de correlaties tussen prijzen van producten en het ondervoede deel van de bevolking binnen een land berekend worden.

2.2.2. Vraag 1

a) *Zijn er food prices die een positieve/negatieve relatie tonen en bestaat deze correlatie door de jaren heen of alleen in een bepaalde periode*

Voor het beantwoorden van deze vraag zijn er verscheidene stappen genomen: Per markt is iedere combinatie van producten genomen. Vervolgens is de set van datums waarvoor beide producten datapunten hebben berekend, en is tussen de bijbehorende prijzen van ieder product de correlatie-coëfficiënt r berekend. Zie ook onderstaande pseudo-code:

```

markets = all_markets_in(food_prices_data)
for market in markets:
    item_combinations = get_item_combinations(market)
    for item1, item2 in item_combinations:
        dates = get_shared_dates(item1, item2)
        item1_prices = item_prices_at_dates(item1, dates)
        item2_prices = item_prices_at_dates(item2, dates)
        r = pearson_correlation(item1_prices, item2_prices)
        # r is the correlation between item1 and item2 on market

```

Figure 1: *Verzamelen van product informatie van twee markten en berekenen correlatie coëfficiënt hiertussen*

Deze r is, samen met de markt en de twee producten, opgeslagen in een JSON dictionary. Om de grootste en positieve en negatieve correlaties te vinden, is deze dictionary gesorteerd op de waarde van de correlatie, en zijn de eerste x en de laatste x waarden van deze dictionary genomen. Om de producten te vinden die wereldwijd het meeste correleren, is per productpaar de gemiddelde correlatie en som van alle correlaties van dit productpaar in de dictionary berekend. Om de landen te vinden met de meest correlerende producten, is per land (gevonden door middel van de markt) de gemiddelde correlatie van alle productparen in de dictionary berekend.

b) *Is het mogelijk om de ingrediënten van een product te vinden in een ander product?* Van twee producten is het makkelijk af te leiden of één van beiden een ingrediënt is van het andere product, zoals bij milk en sour cream. Ingrediënten zullen positief met elkaar correleren; immers, als de prijs van het ingrediënt stijgt, zal de prijs van het eindproduct ook stijgen. De prijs van het eindproduct zal altijd groter zijn dan de prijs van het ingrediënt. Bij de grootst correlerende producten, en de grootste positieve correlaties, is gekeken naar mogelijke ingrediënt-eindproduct-paren.

2.2.3. Vraag 2

Laten landen in dezelfde regio's, vergelijkbare prijsverschillen zien? Is hier een verklaring voor te vinden?

Voor het beantwoorden van deze vraag zijn er verscheidene stappen genomen: In eerste instantie is er per product gezocht naar correlaties tussen landen. Hiervoor moet de data uit de *food prices* dataset in een bruikbaar formaat gezet worden. Aangezien er door de normalisatie van de data een aantal producten zijn in een land

die dubbele datapunten hebben moeten, om het gemiddelde van het land te berekenen, alle voorkomens van een product op een bepaalde datum in een land gesommeerd worden en gedeeld worden door het aantal voorkomens om zo het gemiddelde te berekenen van een product in een land (zie onderstaande pseudo-code).

```
data_set_of_item = all data about given item in given country
total_price_per_date = {}
result_dict = {}
for row in data_set_of_item:
    if the row contains the desired item:
        add item price (in USD) and corresponding date to
        date_dictionary

for date in total_price_per_date:
    add average of prices to result_dict
```

Figure 2: het berekenen van de gemiddelde prijs per product in een land

Vervolgens wordt de data opgeslagen in het volgende formaat:

```
All_data = [[avg_price_per_date, product_name, region_name],
             [avg_price_per_date2, product_name2, region_name],
             [avg_price_per_date3, product_name, region_name2] ...]
```

Figure 3: het formaat waarin de data wordt opgeslagen

Deze data wordt vervolgens gebruikt om te berekenen wat de correlaties zijn tussen de landen in dezelfde regio's. Alle landen zijn verdeeld tussen zes regio's: Zuid-Azië, Het Midden Oosten & Noord-Afrika, Europa Centraal-Azië, Sub-Sahara Afrika, Latijns-Amerika & het Caribisch gebied en Oost-Azië & de Stille Oceaan. Voor het berekenen van de correlaties is de pearson correlatiecoëfficiënt gebruikt. Er wordt door alle mogelijke combinaties van de landen gegaan, waarbij er in eerste instantie wordt bekeken wat de gemeenschappelijke data is, waarmee de gemeenschappelijke values gevonden kunnen worden. Vervolgens wordt de correlatiecoëfficiënt berekend en wordt die toegevoegd aan de lijst van correlaties (samen met de naam, regio en de twee landen tussen welke de correlatie was).

```

result = []
for country1 in all_data:
    for country2 in all_data:
        corresponding_values = corresponding_data between
        country1 and country2
        correlationcoefficient = correlationcoefficient between
        the corresponding values
        append the correlationcoefficient to the result list

```

Figure 4: de correlatiecoëfficiënt wordt berekend tussen twee landen

Ten slotte wordt er bepaald hoeveel van de correlaties onder of boven de 0,4 zijn. Deze grens van 0,4 is gekozen op basis van waarnemingen en meermalig testen, waaruit bleek dat zodra de correlatiecoëfficiënt onder de 0,4 komt, er geen (duidelijk zichtbare) correlatie meer is. Deze keuze heeft verder geen invloed op de resultaten, aangezien dezelfde waarde wordt gekozen bij de wereldwijde landenvergelijking (zonder regio's). Als er meer correlaties met een correlatiecoëfficiënt boven de 0,4 dan onder de 0,4 dan geeft dat aan dat prijs koersen in dezelfde regio's voor meer dan de helft vergelijkbaar zijn.

Dit proces wordt op dezelfde wijze herhaald, maar deze keer worden aan alle landen geen regio's toegekend, wat tot gevolg heeft dat alle productprij-correlaties, ook tussen landen die zich niet in dezelfde regio bevinden, worden berekend.

Door per regio de som van alle mogelijke correlaties te sommeren en delen door de hoeveelheid is er een staafgrafiek gemaakt met de gemiddelde correlatie per regio.

2.2.4. Vraag 3

'Welke correlaties kunnen worden gevonden tussen food prices en ondervoeding?'

Om te onderzoeken of we correlaties kunnen vinden tussen productprijzen en ondervoeding moeten we de correlaties tussen producten in een land en de ondervoeding in dat land kunnen berekenen. Hiervoor is de functie `coeff_country_undernourish` geschreven. De pseudo code van deze functie staat in figuur 5.


```

def coeff_country_undernourish(country, limit)
    coeffs = {}
    products = all_products_in(country)
    for product in products:
        data = get_all_data_of(product)
        undernourishdata = get_undernourishment_data_of(data)
        itemdata = get_item_data_of(data)
        if all elements of undernourishdata is the same:
            coeffs[product] = 0
        elif itemdata.length > limit:
            corr = pearson_correlation(itemdata, undernourishdata)
            coeffs[product] = corr
    return coeffs

```

Figure 5: Deze functie berekent voor alle producten in een gegeven land de correlatie coëfficiënt

Deze functie geeft een dictionary met de structuur {product1 : corrcoeff, product2: corrcoeff} Een belangrijk onderdeel van de functie is dat deze checkt of er genoeg itemdata is. Deze check voert hij uit door te kijken of itemdata langer is dan limit. Limit is een hoeveelheid maanden, een integer dus. Elk element in itemdata is de prijs van een gegeven product in één maand. Dit limiet wordt ingesteld om te garanderen dat de lijnen lang genoeg zijn om iets aan de correlatie te hebben. Er staat ook een check of alle elementen van de undernourishment index niet hetzelfde is. Het is belangrijk om te checken dat niet alle elementen hetzelfde zijn omdat het wiskundig onmogelijk is om correlatie tussen een lijn en een horizontale lijn te berekenen, dit geeft NaN ('Not a Number').

Om de data te kunnen analyseren en er een grip op te krijgen moest er gekeken kunnen worden naar de resultaten op een systematische manier, dit gebeurde met de functie biggest_correlator. Deze functie keek over de hele wereld naar de sterkste correlaties tussen producten. Als argument nam de functie alleen een limiet, hetzelfde limiet als in coeff_country_undernourish. De psuedo code van biggest_correlator staat in figuur 6

```

def biggest_correlator(limit)
    countries = list of all countries
    res = []
    for country in countries:
        countrydict = coeff_country_undernourish(country, limit)
        for product in countrydict:
            coefficient = coefficient of product and country
            res.append ((product, country),coefficient) # tuple
    sortedres = sort_by_coefficient(res)
    return sortedres

```

Figure 6: Deze functie geeft een gesorteerde lijst op correlatie coëfficiënt tussen een bepaald product en de ondervoeding in een bepaald land

Deze functie returned een gesorteerde lijst van alle producten in alle landen en hun correlatie coëfficiënt Deze lijst heeft de volgende structuur: [((Country , Product), coëfficiënt)),.....]. Elk element van de lijst bevat dus het land, het product en de coëfficiënt. Het resultaat is slechts afhankelijk van het limiet dat wordt ingesteld, als er een hoger limiet wordt ingesteld worden er minder producten meegerekend omdat het restrictiever is met betrekking tot de hoeveelheid benodigde data. Om op grotere schaal te kunnen kijken naar correlatie tussen een product en de ondervoeding van het land waar het product in voorkomt is er een functie geschreven die per product van alle correlaties met de ondervoeding in hun respectievelijke landen het gemiddelde neemt. Deze heet `avg_undernourishment..` De ingevoerde corrlist heeft dezelfde structuur als de sortedres die `biggest_correlator` returned: [((Country , Product), coëfficiënt)),.....]. De pseudo code van deze functie staat in figuur 7.

```

def avgundernourishment(corrlist)
    cumdict = {}
    for correlation in corrlist:
        # correlation is a tuple containing product, country and coeff
        product = correlation[0][1]
        coeff = correlation[1]
        if product in cumdict.keys():
            append coeff to cumdict[product] # append to list
        else:
            cumdict[product] = [coeff] # list with coeff as value
    reslist = []
    for product, coefflist in cumdict: # iterate over the cum dictionary
        numberscount = amount of numbers in coefflist
        sum = sum of all numbers in coefflist
        average = sum/numberscount
        reslist.append((product,average)) # append tuple
    res = sort_by_average(reslist)
    return res

```

Figure 7: Deze functie geeft een gesorteerde lijst van gemiddelde correlatie coëfficiënten van producten met de ondervoeding

Deze functie returned een gesorteerde lijst met tuples, deze tuples bevatte de naam van het product en zijn gemiddelde coëfficiënt. De functie stelt eerst een dictionary samen met als key het product en als value een lijst van al zijn correlatie coëfficiënten. Vervolgens neemt hij het gemiddelde van de lijsten met coëfficiënten en koppelt deze aan de producten. Dit zet hij in een lijst, deze lijst wordt op de gemiddelden gesorteerd. Het is belangrijk om de NaN's niet mee te nemen in de berekening, anders komt er vervolgens weer NaN uit.

3. Resultaten

3.1. Vraag 1

Van ieder paar producten die minstens één markt delen, is de gemiddelde (r_{avg}) en gesommeerde (r_{sum}) Pearson correlatie-coëfficiënt r berekend. Er zijn een aantal

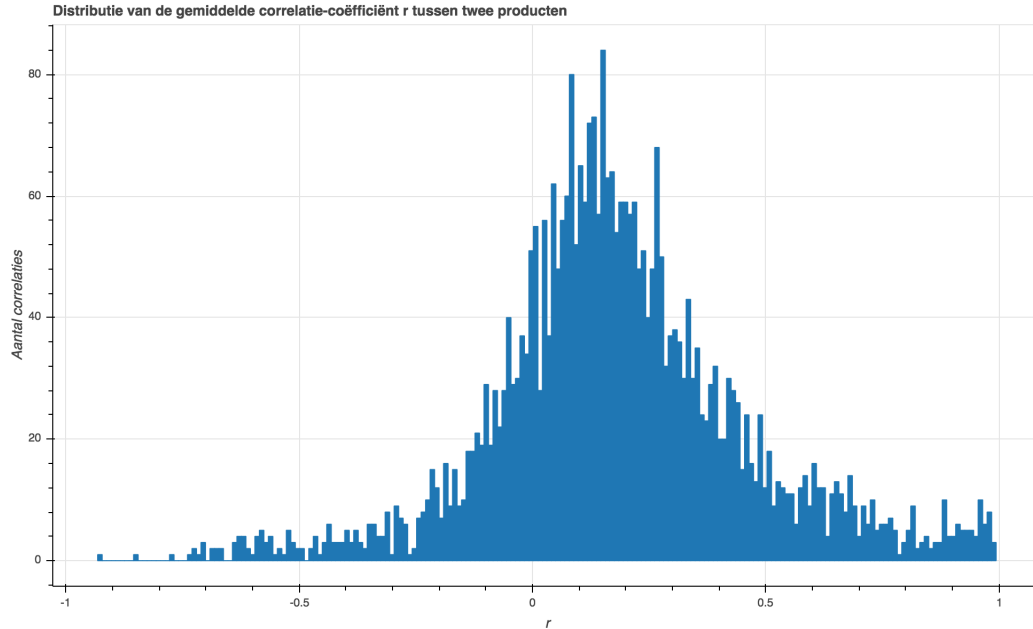


Figure 8: *De distributie van de gemiddelde r tussen de prijzen in USD tussen twee producten*

'perfecte' correlaties ($r = 1.0$ of $r = -1.0$). De prijzen van de producten met een dergelijke coëfficiënt zijn allemaal simpele, bijna helemaal rechte lijnen, wat komt door een gebrek aan data. Het zou kunnen dat deze producten correleren (en vaak zijn ze duidelijk gerelateerd; bijvoorbeeld twee soorten vis, of olie), maar de data is zo invariant dat deze correlaties niet significant zijn. Deze correlaties zijn eruit gefilterd.

Er zijn in totaal 3559 productparen waar een correlatie tussen berekend kon worden. Van de gemiddelde correlaties hiervan zijn 2791 positief, en 768 negatief. Er zijn 897 correlaties die tussen -0.1 en 0.1 liggen, wat een zeer zwak verband aanduidt.

Table 1: top 20 hoogste positieve gemiddelde correlaties wereldwijd

Product 1	Product 2	Som	Gemiddelde
Milk	Sour cream	25.64	0.99
Sour cream	Butter	25.55	0.98
Milk	Butter	25.49	0.98
Curd	Sour cream	25.47	0.98
Curd	Milk	25.46	0.98
Ghee	Oil (cooking)	4.87	0.97
Curd	Butter	25.24	0.97
Beans	Catarino	14.50	0.97
Oil (cotton)	Oil (vegetable)	4.83	0.97
Chicken	Sausage	24.69	0.95
Buffalo	Beef	10.43	0.95
Pasta	Tea	4.73	0.95
Wheat	Oil (vegetable)	3.78	0.95
Carrots	Beetroots	24.36	0.94
Sesame	Oil (vegetable)	4.68	0.94
Bread	Sausage	24.33	0.94
Beef	Sausage	24.18	0.93
Milk	Sausage	23.89	0.92
Cowpeas (white)	Millet	13.75	0.92
Curd	Sausage	23.79	0.91
Fonio	Cashew	1.82	0.91

Table 2: top 20 hoogste negatieve gemiddelde correlaties wereldwijd

Product 1	Product 2	Som	Gemiddelde
Oil (cooking)	Beans	-3.80	-0.76
Oil (cooking)	Sugar	-3.65	-0.73
Cow	Lamb	-1.33	-0.67
Lentils	Sorghum	-1.33	-0.66
Goat	Milk	-1.22	-0.61
Lentils	Milk (powder)	-16.62	-0.59
Pasta	Goat	-1.17	-0.58
Potatoes	Oil (olive)	-1.12	-0.56
Petrol	Sugar (brown)	-2.18	-0.55
Sweet potatoes	Cashew	-1.60	-0.53
Ghee	Beans	-3.18	-0.53
Milk	Cocoa	-1.05	-0.52
Sorghum flour	Cucumbers	-1.05	-0.52
Soybeans	Sorghum	-6.66	-0.51
Fonio	Diesel	-3.58	-0.51
Oil (vegetable)	Potatoes	-11.20	-0.45
Catfish	Snake head	-2.23	-0.45
Plantains	Cucumbers	-1.33	-0.44
Eggs	Poultry	-2.20	-0.44
Milk (powder)	Lentils	-3.01	-0.43
Cassava flour	Cassava leaves	-18.90	-0.43

Milk, Sour Cream, Curd en Butter zijn producten die alleen in Oekraïne zijn gemeten, en hebben allemaal een hoge correlatie met elkaar, op verschillende markten. Figuur 9 laat de prijs over tijd van deze vier producten zien.

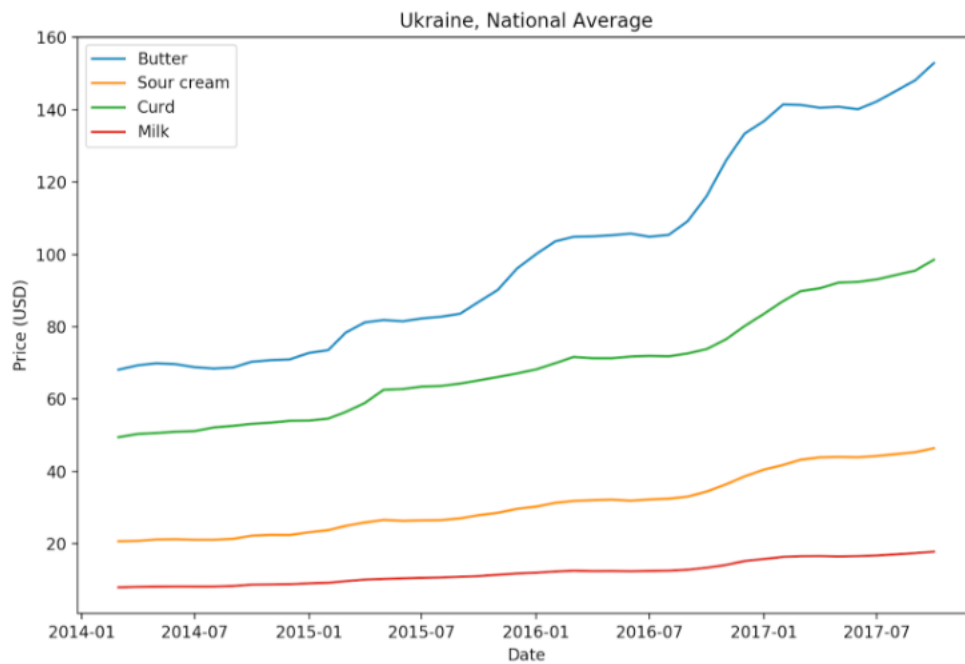


Figure 9: de prijzen van vier correlerende producten Milk, Sour cream, Curd en Butter in het National Average van Oekraïne in USD

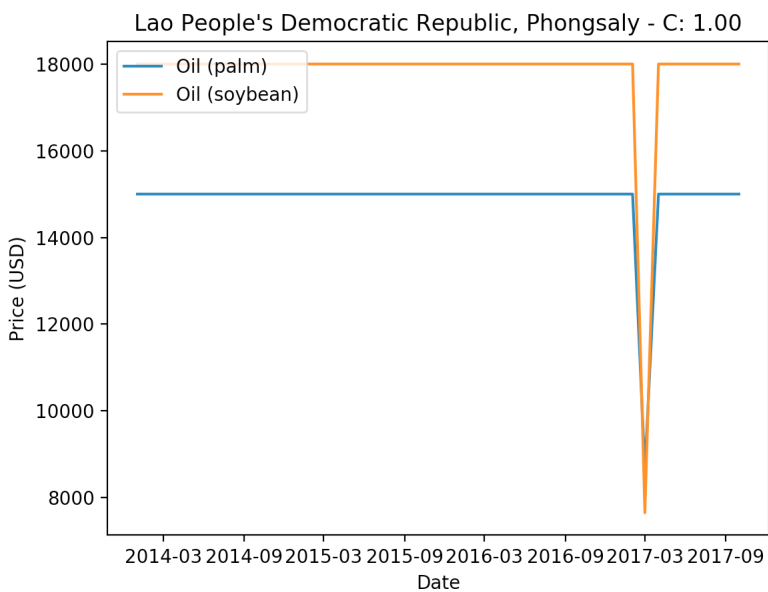


Figure 10

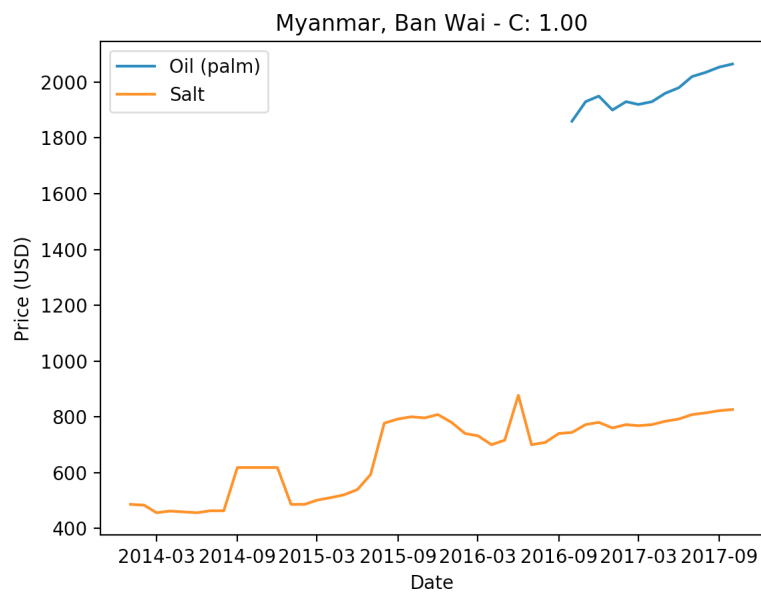


Figure 11

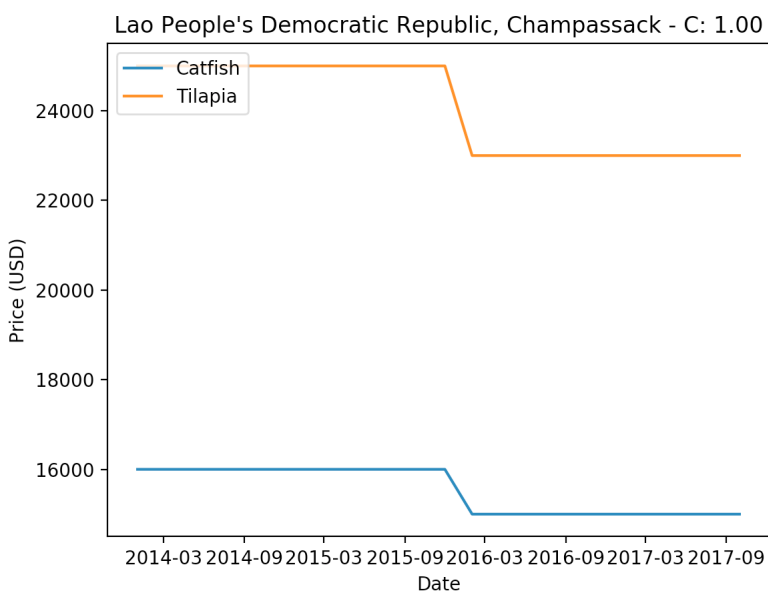


Figure 12

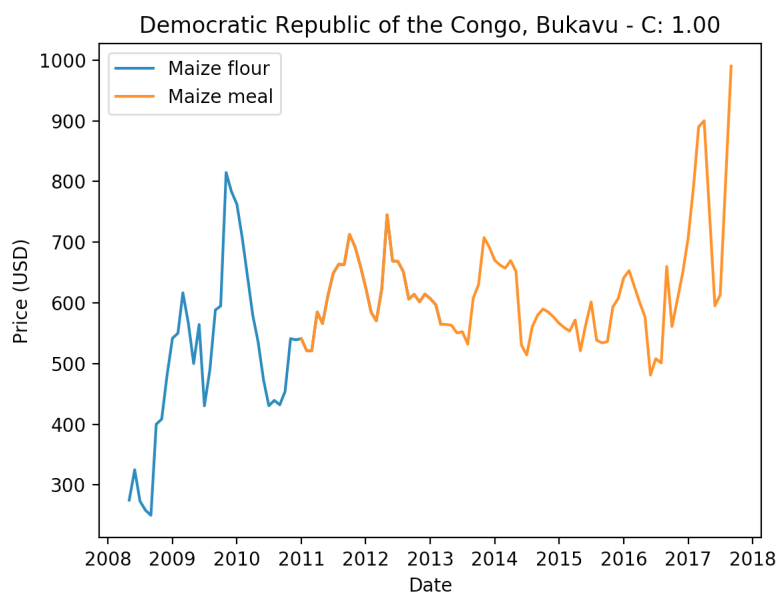


Figure 13

Figuur 10-13: De vijf productparen met de grootste positieve correlatie (C). NB: C is in iedere figuur lager dan 1.00, dit is een afronding.

3.2. Vraag 2

In tabel 3 is een deel van de resultaten van de correlaties tussen landen in dezelfde regio's te zien. In totaal zijn van de 1524 verschillende correlaties, 904 met een correlatiecoëfficiënt onder de 0,4 en 620 met een correlatiecoëfficiënt boven de 0,4.

Table 3: een deel van de correlaties tussen de prijskoersen van producten in verschillende landen, die zich in dezelfde regio bevinden

Product	Correlatie tussen	Regio	r
Wheat	India and Nepal	South Asia	0.96
Oil (sunflower)	Lesotho and Swaziland	Sub-Saharan Africa	0.96
Beef	Kyrgyzstan and Tajikistan	Europe & Central Asia	0.96
Oil (palm)	Bangladesh and India	South Asia	0.95
.[...]	[...]	[...]	[...]
Rice	Djibouti and Yemen	Middle East & North Africa	-0.86
Oil (vegetable)	Guatemala and Haiti	Latin America & Caribbean	-0.87
Wheat flour	Iraq and Egypt	Middle East & North Africa	-0.89
Milk	Tajikistan and Turkey	Europe & Central Asia	-0.89

In tabel 4 is een deel van de correlaties te zien tussen landen over de hele wereld. Hier zijn, van de in totaal 4665 correlaties, 2911 correlaties met een correlatiecoëfficiënt onder de 0,4 en 1754 correlaties met een correlatiecoëfficiënt boven de 0,4.

Table 4: een deel van de correlaties tussen de prijskoersen van producten in verschillende landen, die zich niet in dezelfde regio bevinden

Product	Correlatie tussen	Regio	r
Diesel	Lao People's Democratic Republic and Ukraine	East Asia & Pacific, Europe & Central Asia	0.97
Petrol	Colombia and Lao People's Democratic Republic	Latin America & Caribbean, East Asia & Pacific	0.97
Wheat	flour Cape Verde and Sri Lanka	Sub-Saharan Africa, South Asia	0.97
Petrol	Colombia and Ukraine	Latin America & Caribbean, Europe & Central Asia	0.97
.[...]	[...]	[...]	[...]
Oil (vegetable)	Guatemala and Madagascar	Latin America & Caribbean, Sub-Saharan Africa	-0.91
Rice	Nigeria and Timor-Leste	Sub-Saharan Africa, East Asia & Pacific	-0.93
Salt	Tajikistan and Yemen	Europe & Central Asia, Middle East & North Africa	-0.93
Rice	Tajikistan and Yemen	Europe & Central Asia, Middle East & North Africa	-0.94

Om een beter beeld te krijgen van welke regio's een hoge correlatie hebben tussen dezelfde producten is er een staafgrafiek gemaakt van de gemiddelde correlatiecoëfficiënten per regio (zie figuur 14).

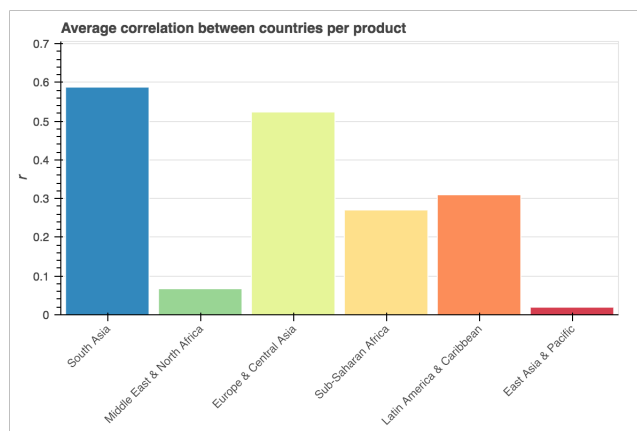


Figure 14: de gemiddelde correlatie tussen de prijs koersen van een product in verschillende landen die zich in dezelfde regio bevinden

3.3. Vraag 3

In tabel 5 staan de wereldwijde top tien hoogste positieve correlaties, dit zijn correlaties tussen het aangegeven product en de ondervoeding in aangegeven het land. In tabel 6 staan de top tien negatieve correlaties wereldwijd tussen ondervoeding en product.

Table 5: Top tien grootste negatieve correlaties tussen product prijs en ondervoeding in het aangegeven land

Product	Land	r
Diesel	Kenya	0.98
Petrol	Kenya	0.97
Kerosene	Kenya	0.96
Fish	Congo	0.95
Petrol	Timor-Leste	0.94
Sugar	Guinea	0.93
Sugar	Armenia	0.93
Groundnuts	Congo	0.92
Petrol	Laos	0.92
Petrol	Ethiopia	0.91

Table 6: Top tien grootste positieve correlaties tussen product prijs en ondervoeding in het aangegeven land

Product	Land	r
Cashew	Gambia	-0.99
Petrol	Madagascar	-0.97
Diesel	Madagascar	-0.96
Cowpeas	Nigeria	-0.94
Beef	Tajikistan	-0.92
Cassava	Congo	-0.92
Fish	Congo	-0.92
Rice	Iraq	-0.92
Chicken	Peru	-0.91
Diesel	Armenia	-0.91

In tabel 7 staan deze statistieken maar dan met een limiet ingesteld van 60. Dat wil zeggen dat er minstens 60 datapunten over de product prijs moet zijn, anders wordt het product niet meegerekend.

Table 7: Top tien grootste positieve correlaties tussen product prijs en ondervoeding in het aangegeven land met een 'limit' van 60

Product	Land	r
Rice	Myanmar	0.85
Potatoes	Guatemala	0.84
Maize	Swaziland	0.82
Oil (palm)	Myanmar	0.82
Beef	Guatemala	0.82
Rice	Djibouti	0.82
Plantains	Guatemala	0.79
Bananas	Guatemala	0.79
Salt	Myanmar	0.79
Eggs	Guatemala	0.77

In tabel 8 staan de top tien hoogste gemiddelde negatieve correlaties per product, en in figuur 9 de top tien hoogste gemiddelde positieve.

Table 8: Top tien producten met de grootste negatieve gemiddelde correlatie met ondervoeding
 Table 9: Top tien producten met de grootste positieve gemiddelde correlatie met ondervoeding

Product	Average r	Product	Average r
Pork	0.26	Noodles	-0.85
Bananas	0.26	Passion fruit	-0.84
Milk (powder)	0.25	Kerosen	-0.84
Plantains	0.23	Oil (cotton)	-0.71
Coffee	0.23	Oil (mustard)	-0.60
Peas	0.18	Tortilla	-0.55
Tomatoes	0.11	Chili (green)	-0.47
Chickpeas	0.10	Lamb	-0.44
Fonio	0.10	Millet	-0.42
Cheese	0.10	Wheat	-0.41

In figuur 15 staat de wisselkoers van de Myanmar Kyat uitgezet tegen de tijd. In figuur 16 staan de prijzen van alle producten uit Guatemala uitgezet tegen de tijd en figuur 17 laat per regio de gemiddelde undernourishment index uitgezet tegen de tijd.

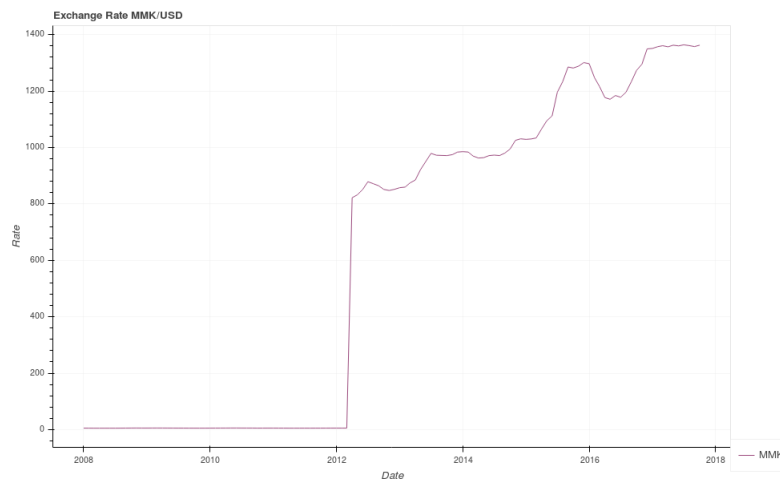


Figure 15: De wisselkoers van de MMK ten opzichte van de US dollar

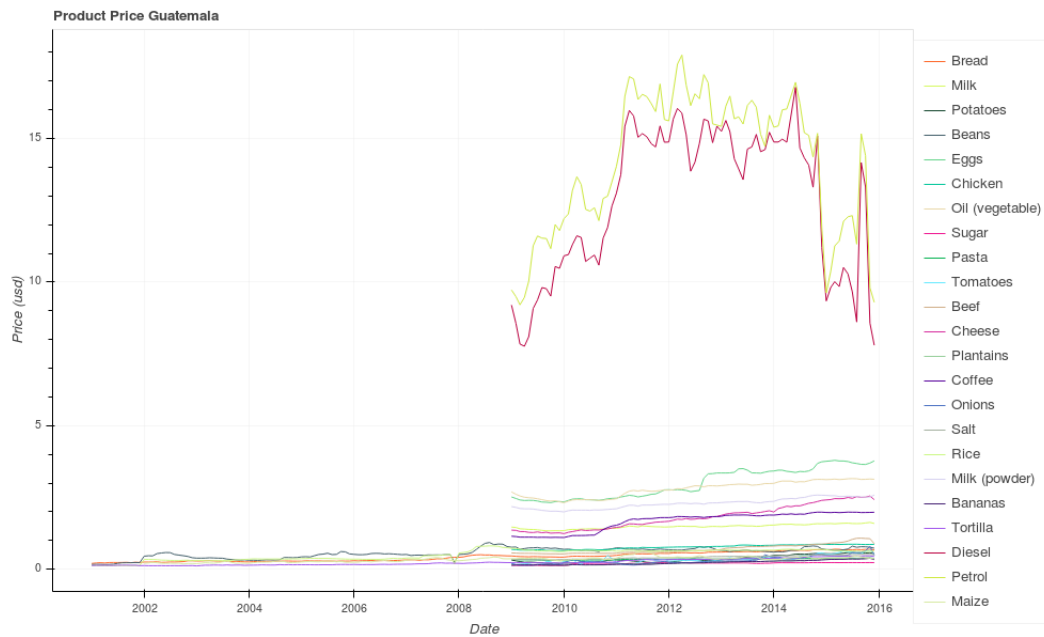


Figure 16: De gemiddelde prijzen van alle producten in Guatemala

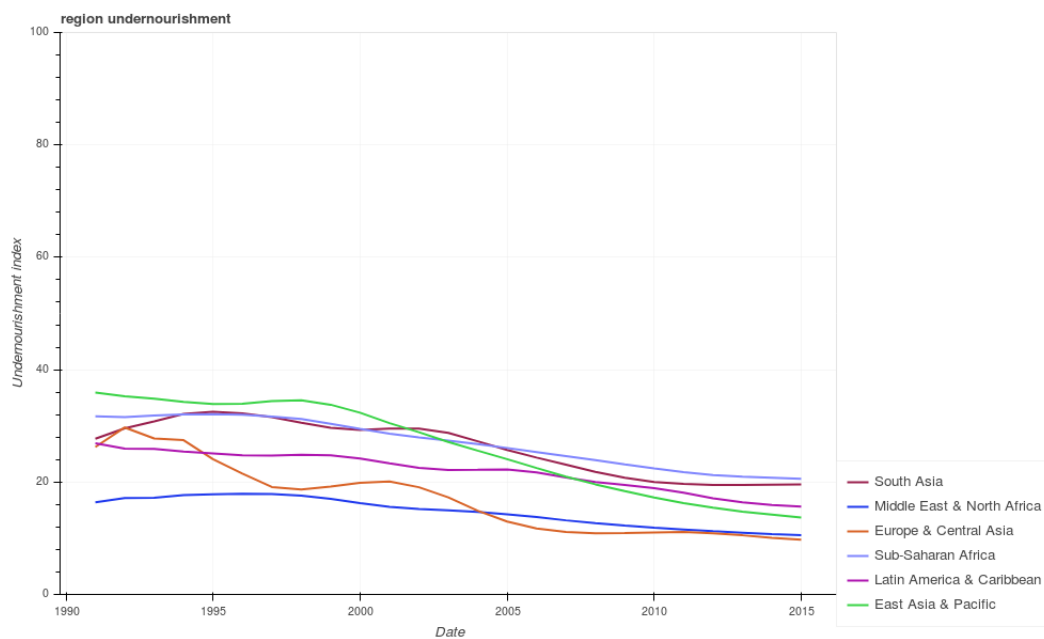


Figure 17: De gemiddelde undernourishment index per regio

3.4. Losstaande interessante resultaten

Bij het analyseren van de food prices data kwamen er een aantal interessante resultaten aan het licht die geen toegevoegde waarde hebben voor het beantwoorden van de onderzoeksvraag maar wel de moeite waren om te onderzoeken.

3.4.1. De negatieve correlatie van tomaten

Bij het berekenen van de top correlaties van food prices tussen regio's, kwam er een interessante negatieve correlatie aan het licht: de negatieve correlatie van de prijs van tomaten tussen Zuid-Azië en Europa & Centraal-Azië (zie figuur 18).

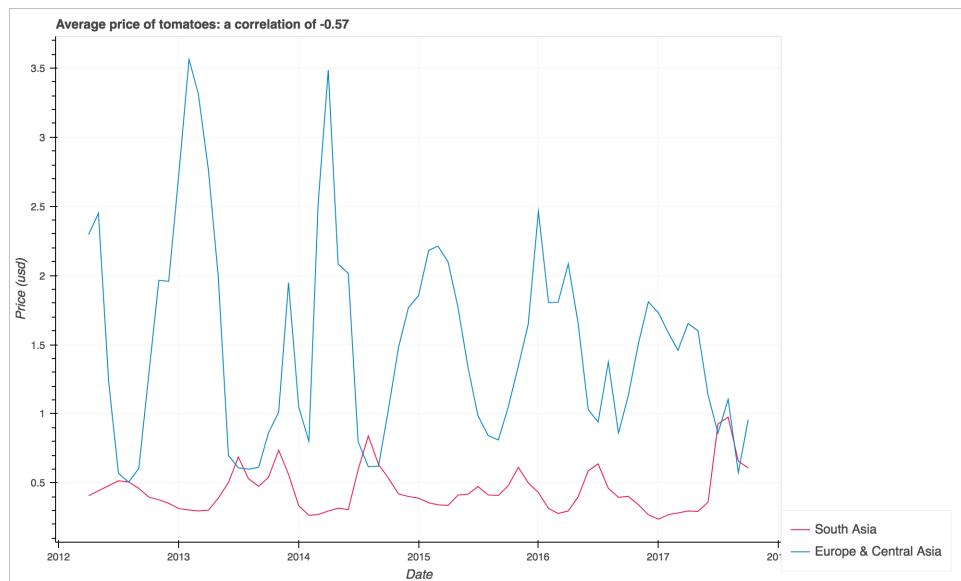


Figure 18: De opvallend negatieve correlatie tussen de prijs koersen van tomaten in Zuid Azië en Europa & Centraal Azië

Deze negatieve correlatie blijkt na verder onderzoek te zijn tussen twee specifieke landen: Armenië en India (zie figuur 19).

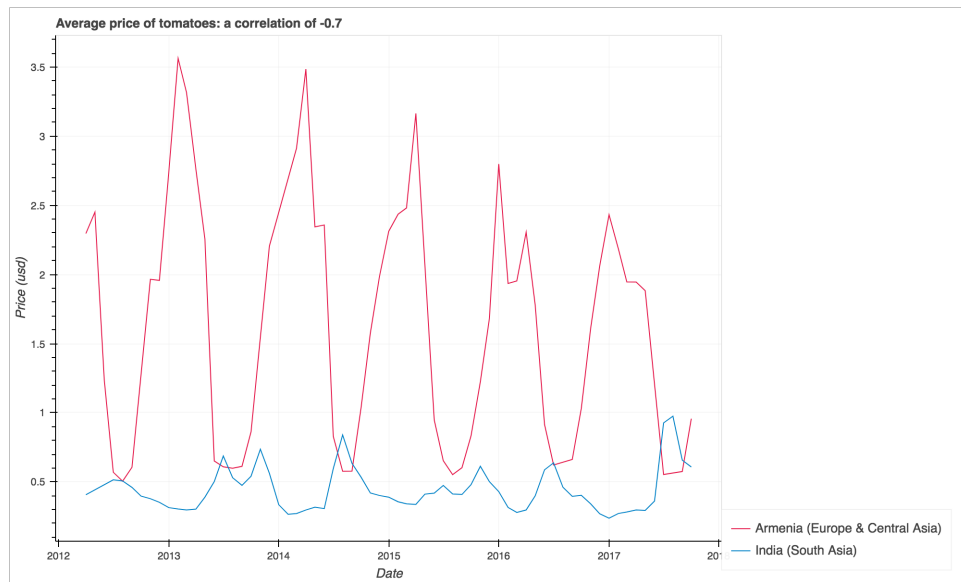


Figure 19: De opvallend negatieve correlatie tussen de prijs koersen van tomaten in India en Armenië

3.4.2. Zuid Soedan productprijzen

In figuur 20 staan de product prijzen van alle producten in de stad Rumbek in Zuid Soedan.

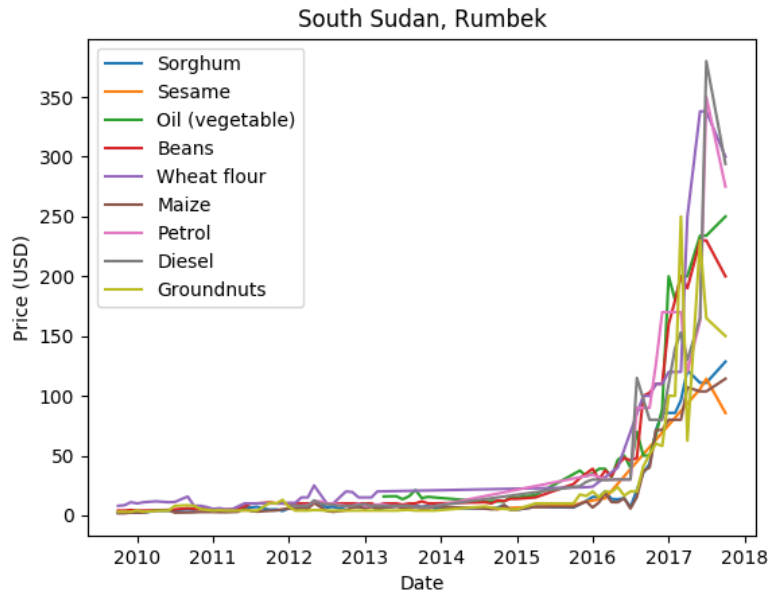


Figure 20: Productprijzen Rumbek, Zuid-Sudan

4. Discussie

4.1. Vraag 1

a. *Zijn er food prices die een positieve/negatieve relatie tonen en bestaat deze correlatie door de jaren heen of alleen in een bepaalde periode?*

Zoals in figuur 8 te zien is, zijn er beduidend meer positieve dan negatieve correlaties. Dit is niet onverwacht, aangezien veel producten inherent positief correleren: zoals vergelijkbare producten, of producten die met andere producten gemaakt worden, terwijl dit niet het geval is met negatief correlerende producten.

Van de 3559 correlerende productparen, zijn er slechts 897 zeer zwak (tussen -0.1 en 0.1). De piek van de distributie (figuur 8) is 0.15, en de distributie gelijkend aan een normaalverdeling. De meeste productparen hebben dus een lichte positieve correlatie; tussen de 0 en 0.3.

Een mogelijke verklaring dat deze piek rond 0.15 ligt in plaats van 0, is dat twee producten van één markt altijd licht zullen correleren omdat ze beiden de schommeling van de waarde van het geld volgen, en er iets kan gebeuren wat de hele markt beïnvloedt (zoals in Zuid-Sudan).

Tabel 2 laat de gemiddelde correlatie van producten per land zien. De twee landen met de hoogste correlatie, Bhutan en Iran, hebben beide maar één markt met enkele producten. Er is verder geen verband gevonden tussen de gemiddelde correlatie van een land en andere variabelen.

Veel van de producten die een hoge positieve correlatie hebben (figuur 2), zijn gerelateerd, zoals Beef en Lamb, Beans en Catarino (Catarino is een type boon), verschillende soorten olie (cotton, vegetable, etc.), Carrots en Beetroot. Dit is een verwacht resultaat, aangezien gerelateerde producten op eenzelfde manier worden geproduceerd/verbouwd, waardoor factoren op beide prijzen hetzelfde resultaat hebben, zoals een slechte oogst.

De producten die een grote positieve correlatie hadden, bleven met elkaar correleren gedurende de hele periode waar data beschikbaar was. Soms was er een onderbreking van de correlatie, maar dit kwam exclusief door incomplete data, d.w.z. gedurende periodes waar steeds dezelfde prijs was ingevoerd. Voor enkele producten was de correlatie sterker in een bepaalde periode (voornamelijk in Zuid-Sudan: zie het betreffende deel van de discussie).

Zoals in figuur 10-13 te zien is, zijn de hoogste positieve correlaties veelal dankzij zeer invariante prijzen. Wij wijten dit aan een suboptimale dataset, aangezien het niet aannemelijk is dat de prijzen van een product gedurende een periode exact gelijk blijven, zeker niet op lokale markten in ontwikkelingslanden. Echter denken wij niet dat dit onze resultaten al te veel beïnvloedt, omdat dit uitzonderingen op de norm zijn, en bijna niet voorkomen in de gemiddelde hoge correlaties, zoals te zien is in figuur 9, waar de prijzen wel variant zijn maar waar alsnog een hoge correlatie is.

Er is geen verband gevonden tussen de producten die voorkomen in de hoogste negatieve gemiddelde correlaties (tabel 2). De meeste producten hebben intuïtief geen verband. Een mogelijke verklaring is dat de producten in tegenovergestelde seizoenen verbouwd/geoogst worden.

b. Is het mogelijk om de ingrediënten van een product te vinden in een ander product?

De producten die wereldwijd de hoogste gemiddelde correlatie hadden (tabel 1), zijn er een aantal ingredient-eindproduct-paren:

1. Milk is een ingredient van Sour cream, Curd en Butter (in Oekraïne).
2. Beef en/of Chicken zijn waarschijnlijk ingredienten van Sausage.
3. Maize is een ingredient van Maize meal (meel van mais).

Milk, Sour cream, Curd en Butter zijn producten die exclusief in Oekraïne zijn gemeten. In figuur 9 is de hoge correlatie te zien, alsmede de prijsverschillen die

ieder product heeft. Aangezien Milk de laagste prijs heeft, is dit het basisingredient voor de overige producten.

Een opmerkelijk negatief verband is de onderste rij van tabel 2: Cassava flour en Cassava leaves hebben een gemiddeld negatief verband van -0.43. Deze twee producten zijn geen ingredienten van elkaar, maar hebben een gemeenschappelijke plant: de cassave (Cassava flour wordt gemaakt van de wortel van cassave). Een negatief verband is onverwacht, aangezien er een positief verband is tussen Cassava flour en cassave, en Cassava leaves en cassave. Hier is geen verdere verklaring voor gevonden.

4.2. Vraag 2

Laten landen in dezelfde regio's, vergelijkbare prijsverschillen zien? Is hier een verklaring voor te vinden?

Van de 1524 verschillende correlaties tussen landen in dezelfde regio zijn er 620 met een correlatiecoëfficiënt boven de 0,4. Van de 4665 correlaties tussen landen over de hele wereld zijn er 1754 correlaties met een correlatiecoëfficiënt boven de 0,4. Dat betekent dat zo'n 41% van de prijs koersen tussen landen in dezelfde regio vergelijkbaar zijn en zo'n 38% van de prijs koersen tussen landen die niet in dezelfde regio zijn vergelijkbaar zijn. Hieruit valt af te leiden dat er tussen landen in dezelfde regio's, meer vergelijkbare prijsverschillen zijn dan tussen landen die zich niet in dezelfde regio bevinden. Dit resultaat is in lijn met de hypothese.

Een mogelijke verklaring hiervoor zijn de weersomstandigheden; landen in dezelfde regio zijn afhankelijk van dezelfde weersomstandigheden. Aangezien het overgrote deel van de producten in onze database eten is en de landen in onze database grotendeels ontwikkelingslanden zijn zullen veel producten afhankelijk zijn van de oogst (op het land). Weersomstandigheden zullen dan wel hetzelfde effect hebben op landen die in de buurt van elkaar liggen, maar niet op landen ver van elkaar.

Dit wordt verder ondersteund door het feit dat van de top tien food price correlaties tussen alle landen van de wereld, zes correlaties diesel of petrol betreft. Producten waar over de hele wereld vraag naar is en die niet afhankelijk zijn van weersomstandigheden.

In figuur 14 zijn de gemiddelde correlatiecoëfficiënten tussen landen met hetzelfde product te zien. Opvallend is dat Zuid Azië en Centraal Azië gemiddeld een hoge correlatiecoëfficiënt hebben, dit is mogelijk te verklaren met de grote hoeveelheid aan lokale markten en winkels. Het Midden Oosten & Noord Afrika hebben in contrast een erg lage gemiddelde correlatiecoëfficiënt. Deze landen zijn grote spelers in de olie-industrie, een markt die meer op globaal, dan regionaal niveau speelt. Een mogelijke verklaring voor de lage gemiddelde correlatiecoëfficiënt van Oost Azië en

de Stille Oceaan is dat veel van de landen in die regio (gedeeltelijk) eilanden zijn, die mogelijk slechter bereikbaar zijn waardoor er minder handel is tussen eilanden. Hierdoor kan er eerder schaarste ontstaan als de oogst in een van de landen slecht is.

Desondanks er een duidelijk verband tussen food prices en regio's is, zijn er nog steeds veel landen die in dezelfde regio liggen maar geen vergelijkbare prijs koers hebben. Een verklaring hiervoor is dat er een groot aantal onvoorspelbare factoren zijn die de voedselprijs beïnvloeden.

4.3. Vraag 3

Welke correlaties kunnen worden gevonden tussen de food price data en ondervoeding?

Zoals te zien is in figuur 5 zijn zes van de top tien hoogste correlaties tussen de producten in onze database en ondervoeding petroleumproducten. Deze hebben allemaal een r boven de 0.91. Dit resultaat lijkt te komen door twee factoren, de eerste heeft er mee te maken dat deze producten allemaal pas recenter worden ingevoerd in de database wat betekent dat er minder datapunten zijn om te vergelijken. Hierdoor ontstaat er sneller een hoge of lage correlatie, dit wordt bevestigd door de tien hoogste negatieve correlatie coëfficiënten. Deze bevat namelijk drie petroleum producten, de reden dat dit lager ligt dan de top tien heeft te maken met twee wereldwijde trends: het dalen van olieprijsen in de jaren 2013-2015⁶ en het dalen van ondervoeding wereldwijd wat duidelijk te zien is in figuur 17. De tweede reden dat er significant veel petroleum producten te vinden zijn in de top tien is de hoge volatiliteit van petroleumproducten. De prijzen van petroleum producten zijn afhankelijk van olie prijzen, en olie prijzen staan bekend als extreem volatiel.⁷ Dit duidt er dus op dat er met deze resultaten niet valide conclusies vallen te trekken wat betreft correlatie tussen *food prices* en ondervoeding.

Het is dus nodig om een limiet in te stellen voor hoeveel data er vereist is om een correlatie aan te kunnen wijzen. Dit leidt er wel toe dat we data uit sluiten, gezien lang niet alle producten jarenlang worden bijgehouden, wat wel een vereiste is. Met een limiet van 3 jaar (36 maanden dus) wordt al bijna de helft van de producten uitgesloten. Dit komt omdat er de vorige 3-5 jaar een stijging is geweest in de hoeveelheid producten die ingevoerd worden, wat resulteert in veel producten met relatief weinig data. We hebben er voor gekozen om een limiet in te stellen van 5 jaar oftewel 60 maanden, dit betekent dat we per land gemiddeld ongeveer vier producten uitsluiten. Ook betekent dit dat we 18 landen compleet uitsluiten gezien er voor geen producten genoeg data bestaat.

Zoals te zien is in tabel 7 geeft het invoeren van een limiet compleet andere resultaten, geen van de originele top tien blijven over als we filteren op meer dan

60 maanden. Dit benadrukt het belang van het stellen van een ondergrens qua de hoeveelheid data. Het is opmerkelijk dat acht van de top tien correlaties uit Myanmar en Guatemala komen. Het vele voorkomen van Myanmar is te verklaren door de herinstelling van de valuta (de MMK) in 2012⁸. In dit land is de zogenaamde 'street rate' (de werkelijke werkbare waarden van de valuta) anders dan de werkelijke koers. Dit heeft ertoe geleid dat de prijzen in de markten rond deze tijd niet hebben veranderd, maar de wisselkoers waar we mee rekenen wel. We hebben geen gangbare verklaring kunnen vinden voor het vele voorkomen van Guatemala in de top tien, er zijn 21 (niet petroleum) producten in Guatemala. In totaal zijn er gemiddeld per land gemiddeld 4.07 meegenomen, en 70 landen. Dit betekent dat dat 21/290 van de correlaties in Guatemala zijn. We verwachten dus dat 1/10 van de top tien uit Guatemala komt. In werkelijkheid is dit 1 / 2. Er is duidelijk in Guatemala een onverwacht sterke correlatie tussen productprijzen en ondervoeding.

Het is ook mogelijk om door per product de gemiddelde correlatie coëfficiënt te berekenen te bekijken of er een product is wat wereldwijd significant correleert met ondervoeding. In tabel 8 staan de resultaten hiervan. Opmerkelijk ligt de hoogste correlatie relatief tot tabel 7 heel laag. Dit duidt erop dat er geen producten te vinden zijn die in een groot aantal landen een hoge correlatie hebben met de ondervoeding.

De hoogste negatieve gemiddelde correlaties uit tabel 9 zijn wel significant. Een correlatiecoëfficiënt van 0.84 zoals die van noodles duidt op een sterke correlatie. Deze top tien gemiddelden zijn allemaal berekend aan de hand van één correlatiecoëfficiënt, wat conclusies eruit trekken meteen buiten sluit. Er zijn aan de hand van onze analyse weinig sterke correlaties gevonden tussen productprijzen en ondervoeding. Veel data moest buitengesloten worden omdat er niet overlap was met betrekking tot de datums van de data in de twee datasets. Ook waren de datasets niet goed op elkaar afgestemd. Een voorbeeld van slechte afstemming was het feit dat onze productprijzen maandelijks waren ingevoerd, en de undernourishment indexen alleen jaarlijks waren berekend. Dit leidde tot hoge grenzen (60 rijen) en steeds minder goed vergelijkbare data. Dit laat zien dat ondanks het feit dat we een overvloed aan data leken te hebben het belangrijk is dat je data hebt dat afgestemd is op elkaar.

4.4. Interessante losstaande resultaten

Bij het onderzoeken van de negatieve correlatie tussen tomaten in India en Armenië (zie figuur 19) is er één mogelijke verklaring aan het licht gekomen. Een oorzaak voor het verband zou het verschil in oogsting periodes kunnen zijn. Door het verschil in oogsting periodes zouden er op tegenovergestelde momenten van het jaar meer of minder tomaten op de markt kunnen komen waardoor de prijs tegen-

overgesteld schommelt. Maar aangezien India het hele jaar warm is (tomaten zouden in principe het hele jaar door kunnen groeien) en in Armenië een landklimaat is waar juist alleen na de zomer geoogst kan worden (ook te zien in de figuur 18 en 19), na van het jaar is er telkens een daling van de prijs), wordt deze verklaring grotendeels tegengesproken. Mogelijk is het simpelweg toeval.

Bij onderzoek naar correlaties tussen voedselprijzen in een bepaald land werd figuur 20 gegenereerd. We zien in deze grafiek de prijs van verschillende producten in de hoofdstad van Zuid Sudan. Rond 2016 beginnen voedselprijzen na een lange tijd constant te hebben gebleven beginnen te stijgen. Deze stijging loopt gelijk met de 2016 'Juba clashes' (<https://www.theguardian.com/world/2016/jul/10/south-sudan-capital-juba-violence-salva-kiir>). Een serie conflicten in Juba in de burgeroorlog in Zuid Sudan, er zijn hierdoor rond 300 mensen omgekomen. Dit leidde tot verhoogde spanningen in de regio, en dus ook tot verhoogde voedselprijzen.

4.5. Gebrekkige data

Er kwam bij de food prices dataset veel cleaning aan te pas. Toch bleek zelfs na het cleanen vaak nog dat er data inconsistent of onlogisch was. Bekijk als voorbeeld figuur 10, hierin staat een enorme daling en stijging van de prijs rond maart 2017 tegen een verder geheel onveranderde prijs koers (ook onrealistisch). Dit soort onregelmatigheden in de data kwamen regelmatig voor, dit duidt op een systematisch probleem bij het verzamelen en/of invoeren van de data in de food prices dataset. De inconsistentie in data maakte het juist visualiseren en analyseren van de data een stuk gecompliceerder.

4.6. Toekomstig onderzoek

In dit onderzoek is de correlatie tussen de ondervoeding en de voedselprijzen berekend met behulp van de 'global food prices database' en de 'prevalence of undernourishment' database. Er konden uit dit onderzoek door verschillende redenen niet concrete conclusies getrokken worden. Om deze reden zou een interessant vervolgonderzoek verder in kunnen gaan op het verband tussen ondervoeding en de food prices in de wereld.

References

- ¹ A. Gray, Which countries spend the most on food? this map will show you.
URL <https://www.weforum.org/agenda/2016/12/this-map-shows-how...../>

² R. Holmes, N. Jones, Understanding the impact of food prices on children, Children in Focus.

³ Global food prices database (wfp) (Dec 2015).

URL <https://data.humdata.org/dataset/wfp-food-prices>

⁴ Us dollar exchange rates bis (Jun 2018).

URL <https://www.bis.org/statistics/xrusd.htm>

⁵ Prevalence of undernourishment (% of population).

URL <https://data.worldbank.org/indicator/SN.ITK.DEFC.ZS>

⁶ Wti crude oil prices - 10 year daily chart.

URL <http://www.macrotrends.net/2516/wti-crude-oil-prices-10-year-daily-chart>

⁷ G. Bornstein, P. Krusell, S. Rebelo, What makes oil prices so volatile?

URL <https://insight.kellogg.northwestern.edu/article/what-makes-oil-prices-so-volatile>

⁸ BBC, Burma sets currency exchange rate as it floats the kyat (Apr 2012).

URL <https://www.bbc.com/news/business-17581115>