Food price inflation nowcasting and monitoring

Second technical workshop on nowcasting in international organizations (UNIDO/CCS-UN), May 25-26, 2022

Christian Mongeau
Food and Agriculture Organization of the United Nations (FAO)
Structure of the presentation

- Some context (FAO’s Data Lab and why we are in this business)
- Data
- Some technicalities
- Methodology
- Overview of the dashboard
- Brief live demo of the dashboard
- Limitations/improvements
FAO’s Data Lab

- Created at the end of 2019 with the aim to modernise the statistical process within FAO’s Statistics Division (ESS) by the use of new methodologies and data sources

- As Covid-19 appeared, timeliness of data as well as the capacity to quickly and automatically draw insights from data for policy making became an essential issue

- This has generated an increased need for timely, possibly real-time, information from non-conventional sources and its automated analysis

- Current methods adopted:
  - Scraping data from the web and literature (e.g., news articles from media outlets, data from NSOs, literature-mining for food loss and waste)
  - Text analytics (e.g., topic modelling, sentiment computation)
  - Statistical modelling (“standard” and new methodologies)

- The team is currently composed by seven people (statisticians, data scientists, IT), and it is growing
Why we are nowcasting food prices?

- Information on food-related indicators is a crucial element for FAO (Food and Agriculture Organization)
- The Organization has a “control room” where a series of country-specific indicators (food prices included) are displayed in specific dashboards with the aim of providing a general overview of the most recent situation in a country
- FAOSTAT’s prices domain is updated every three months and usually shows some time lag as data is standardised/validated
- National Statistical Offices provide information with at least one month lag (35 countries with info up to March, 94 up to April)
- Given that we want “current” information, we generate nowcasts to assess actual/potential trends
Data


- “Real-time” (or current) information we use:
  - the LCU/USD exchange rate
  - the average price of oil (Brent/WTI)
  - price of the 14 food items obtained from a crowdsourced database (more on this later)
  - sentiment index from Twitter (more on this later)

  Period starts on April 2020 (when we started collecting prices)
We use Numbeo’s Food Prices dataset: https://www.numbeo.com/food-prices/

It contains information on 14 commodities and are entered by users (“crowdsourced”) and manually collected data from authoritative sources (websites of supermarkets, taxi company websites, governmental institutions, newspaper articles, other surveys, etc.)

Information is available also by (main) city, but we use the national average

Numbeo performs automatic and semi-automatic filters to remove noise data
Twitter sentiment index

- We collect daily all tweets posted by 504 news outlets around the world: 189 countries, almost 3 accounts on average for each country
- The links posted on each tweet get saved and the article is extracted
- Sentiment is calculated on the article as $PW / (PW + NW) \times 100$ where $PW$ and $NW$ are all the positive and negative words found in the article, respectively (the index goes from 0 to 100)
Technicalities

- Daily cron jobs collect:
  - *Trading Economic* prices, which as of today make a PostgreSQL table of 68,595 data points for 171 countries (we take the latest data for each month)
  - Numbeo daily prices, which as of today are
  - Exchange rates, from Yahoo! Finance and PACIFIC Exchange Rate Service ([https://fx.sauder.ubc.ca/](https://fx.sauder.ubc.ca/))
  - Oil prices, from PACIFIC Exchange Rate Service
  - Tweets (and articles), which are currently nearly 16 million since Jan. 2020 of almost 20k per day; these are stored in a NoSQL database (solr), where more information is added (sentiment, topic, etc.)
Methodology (nowcasts)

- The model is built as a fixed effects dynamic panel model:

\[ \tilde{F}_{i,t} = \alpha + \beta_1 \tilde{F}_{i,t-1} + \beta_2 \tilde{P}_{i,t} + \beta_3 \tilde{E}_{i,t} + \beta_4 S_{i,t} + \delta_r \tilde{C}_t + \epsilon_{i,t} \]

where:

- a tilde over a variable (e.g., \( \tilde{X} \)) indicates the monthly variation of the variable
- \( \tilde{F}_{i,t} \) is the food consumer price index for country \( i \) at time \( t \)
- \( \tilde{P}_{i,t} \) is the average price of the 14 food items obtained from a crowdsourced database (more on this later) and aggregated by using the share of availability of the items (item \( i \) over sum of items)
- \( \tilde{E}_{i,t} \) is the LCU/USD exchange rate
- \( S_{i,t} \) is the Twitter sentiment index (more on this later)
- \( \tilde{C}_t \) is the average crude oil price (WTI and Brent; note: this enters as region \( r \) effect)
Checks on model

- We carried out standard tests to check whether the model behaves better than benchmark models (only autoregressive): (1) country-fixed effects; (2) region-fixed effects.

<table>
<thead>
<tr>
<th>model</th>
<th>MAE</th>
<th>rel_1</th>
<th>rel_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>1.86</td>
<td>0.0%</td>
<td>19.2%</td>
</tr>
<tr>
<td>[2]</td>
<td>1.56</td>
<td>-16.1%</td>
<td>0.0%</td>
</tr>
<tr>
<td>[3]</td>
<td>1.34</td>
<td>-28.0%</td>
<td>-14.1%</td>
</tr>
</tbody>
</table>

Results show that there is indeed some improvement (rel_1 and rel_2 are MAEs relative to (1) and (2) respectively; (3) is the final model).

- Comparing forecasts made by using a model for which the last 3 months were removed to actual values indicate that there is a relatively good match of forecasts.
Methodology (food prices acceleration; by Luís Silva e Silva)


- The idea is to compute a “normal” growth rate for prices and then check whether current growth deviates from it

- What “normal” inflation means is a 30-day compound growth:
  \[ CGR_t = \left( \frac{P_{tn}}{P_{t0}} \right)^{1/(t_n - t_0)} - 1 \]

- This is then standardised (see box) and deviations define whether there are abnormal variations:
  - Regular: less than 0.5 standard-deviation
  - Moderate: between 0.5 and 1.0 standard-deviation
  - Watch: between 1 and 2.5 standard-deviation
  - High: greater than 2.5 standard-deviation
Food price inflation nowcasting and monitoring

Online application (nowcasting)

Regional aggregation of food inflation

Interactive map with nowcasts

Table with details

(clicking on maps/table shows more info)
Online application (food prices acceleration)

Interactive map with food price acceleration

Top 20 countries by food prices acceleration

All data
The two methods are available in a public dashboard

- From FAO’s Data Lab website:
  
  https://www.fao.org/datalab/website/web/food-prices

- Or directly from the Shiny server:
  
  https://foodandagricultureorganization.shinyapps.io/dl_foodprices/

  (the former contains an iframe to the latter)

- They are updated daily by nightly cron jobs

- The output in the dashboard is in the process of being shared with the control room by means of API calls to the server and are integrated into “cards” in country-specific pages
Limitations (to see possible improvements)

- **Modeling:**
  - For nowcasts, we are currently using a relatively standard, or traditional, model, and thus not exploiting new methodological advances in the field (though, sometimes simple is good)

- **Numbeo:**
  - The methodology to “filter out noise data” is not available (though, we use either the monthly average or a smoothed version of the daily index)
  - Crowdsourced prices may reflect more urban (or, eventually, touristic) areas, thus may not be representative of more rural countries (where internet access is lower)
  - For a similar reason, prices for countries with conflicts or natural catastrophes may be biased

- **Twitter sentiment:**
  - A somehow basic approach to extract sentiment (we are reviewing it)
  - It uses a general-use dictionary; we may need to use a domain-specific dictionary (we are on it)
Thanks!

https://www.fao.org/datalab

Christian.Mongeau@fao.org