Estimating Food Price Inflation from Partial Surveys

Second technical workshop on nowcasting in international organizations, May 26, 2022

Bo Pieter Johannes Andrée, bandree@worldbank.org
DEC Analytics & Tools Unit

UKaid from the Department for International Development
UNHCR The UN Refugee Agency
THE WORLD BANK IBRD • IDA | WORLD BANK GROUP

Building the Evidence on Forced Displacement

Overview

I Background and Motivation

II Modeling approach

Missing data problem

Multivariate Imputation

III Results

Afghanistan

Papers associated with this work

- Andree, Bo Pieter Johannes. 2021. ***Estimating Food Price Inflation from Partial Surveys***. Policy Research Working Paper; No. 9886. World Bank, Washington, DC. http://hdl.handle.net/10986/36778
- Andree, Bo Pieter Johannes; Chamorro, Andres; Kraay, Aart; Spencer, Phoebe; Wang, Dieter. 2020. ***Predicting Food Crises.*** Policy Research Working Paper; No. 9412. World Bank, Washington, DC. http://hdl.handle.net/10986/34510
- Wang, Dieter; Andree, Bo Pieter Johannes; Chamorro, Andres Fernando; Girouard Spencer, Phoebe. 2020. ***Stochastic Modeling of Food Insecurity***. Policy Research Working Paper; No. 9413. World Bank, Washington, DC. http://hdl.handle.net/10986/34511

Explore the food price inflation estimates on the *World Bank Microdata Library*:

- Monthly food price inflation estimates by country, 25 countries

  https://doi.org/10.48529/2ZH0-JF55

Larger Objectives:

I Advance toward real-time data

II Track spatially explicit price dynamics

III Look beyond areas where formal measurement of prices takes place

CPI Data challenges:

- ⊖ Often lags, particularly during crises when prices deteriorate rapidly
- ⊖ Is not easily unbundled into prices of distinct goods (e.g. domestic wheat prices)
- ⊖ Highly urbanized, covers markets that do not serve the majority of poor

Desired capabilities:

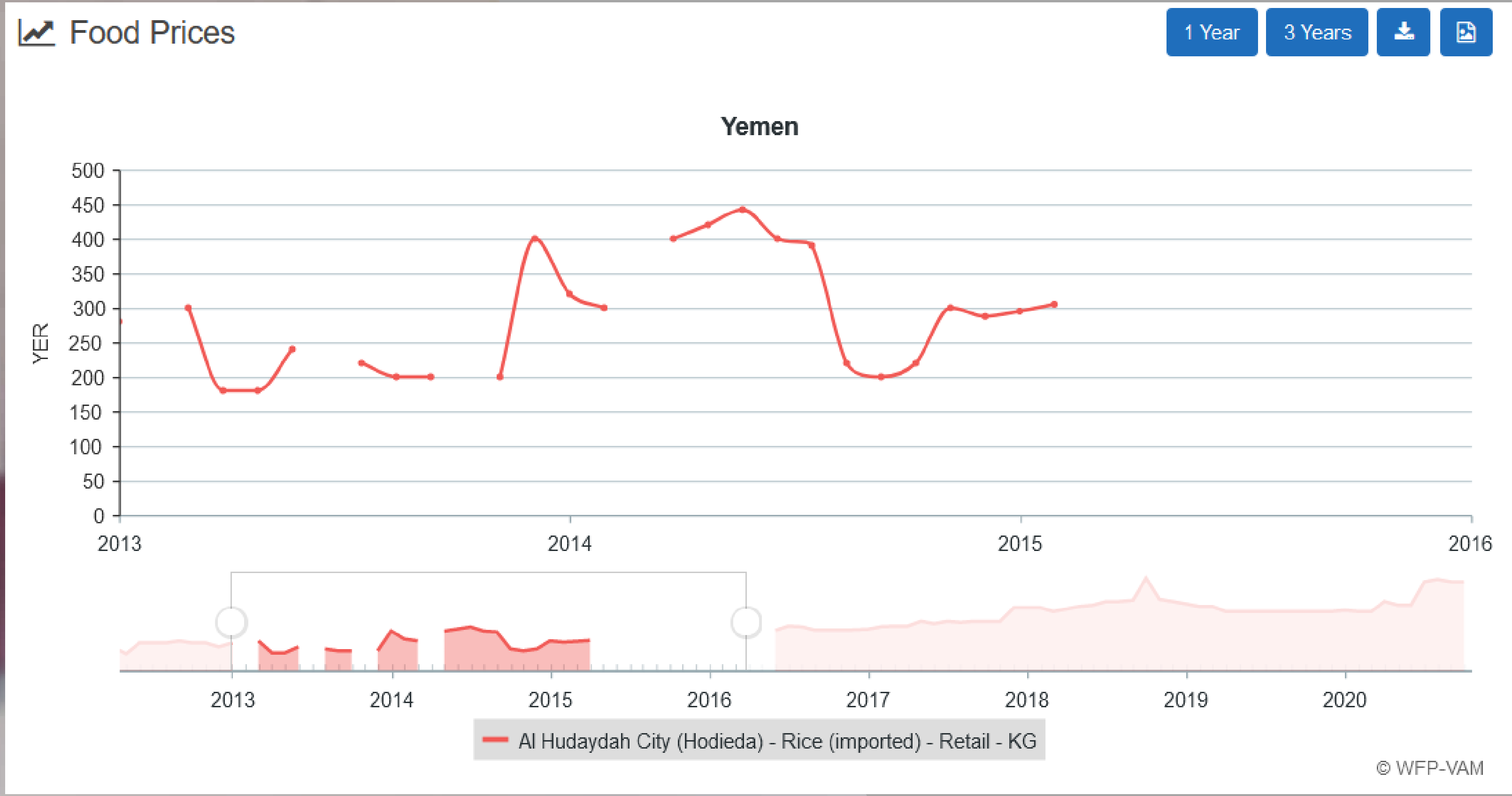
- + Gain real-time insight beyond what is formally measured
- + Deployable during crises, when deliberate sampling processes are prohibited
- + Commodity-specific

Typical survey data is incomplete

There are existing systems by WFP, FAO, FEWS NET, IFPRI, and more

Prices are often not reliably available across all goods, markets, and time

Example WFP data on rice prices in Yemen



Combining survey and imputation

Multi-step modeling process to get to real-time data

I Filter the data for outliers

Relies on non-parametric methods that search for irregularities in distributions. Outliers will be converted to missing data points

II Multivariate-imputation

A regression-chain approach that differs from existing approaches such as MICE in that each step involves a fully optimized and cross-validated machine learning model.

III Intra-month price spreads are estimated

Involves a Fractionally Integrated GARCH model (Generalized Auto-Regressive Conditional Heteroskedasticity) using a Generalized Error Distribution. Then integrating under the distribution around each price point to obtain Conditional Expectations for intra-month High and Low points.

The missing data problem

Three hypothetical vectors of price time series, with elements a_t, b_t and c_t being individual price quotes indexed by time t.

Blank entries represent missing observations.

The challenge is to estimate change rates ΔP of the basket price vector $P = A + B + C$, for all $t = 1, \ldots, 6$.

A (Apples)	*B* (Bananas)	*C* (Coconuts)	*P* (Basket price)
a_1	b_1		
	b_2	c_2	
a_3		c_3	
a_4	b_4^*		
	b_6	c_6	

Element b_4^* is an example outlier price which needs to be removed and replaced with an estimate.

Problems with traditional methods cont.

- A regression imputation is impossible:
  - 55% of entries have an element
  - 0% of rows are complete

- Ideally, we still want to exploit the information held in price ratios:

Pairs (a_1, b_1), (b_2, c_2), (a_3, c_3)
and so forth.

A (Apples)	*B* (Bananas)	*C* (Coconuts)	*P* (Basket price)
a_1	b_1		
	b_2	c_2	
a_3		c_3	
a_4	b_4^*		
	b_6	c_6	

Algorithm sketch

- Set all missing entries at some sensible value.
- Iterate the following:
  - Replace previous imputations in Apples using predictions based on Bananas and Coconuts
  - Replace previous imputations in Bananas using predictions based on Apples and Coconuts
  - Replace previous imputations in Coconuts using predictions based on Apples and Bananas
- Stop when the predictions don’t get any better.

Some general observations:

- ✓ Typically, more price items results in better accuracy. Prediction quality also improves when inflation (deflation) is strong compared to volatility.
- ✓ More markets, increases heterogeneity so demands more price items and observations to maintain performance.
- ✓ When the majority of data is missing, the quality of imputations can still be comparable to the accuracy of direct measurement.

Toward real-time local food prices:

Example results Afghanistan

Food Price Estimates Afghanistan I

Example simple 3-component index:

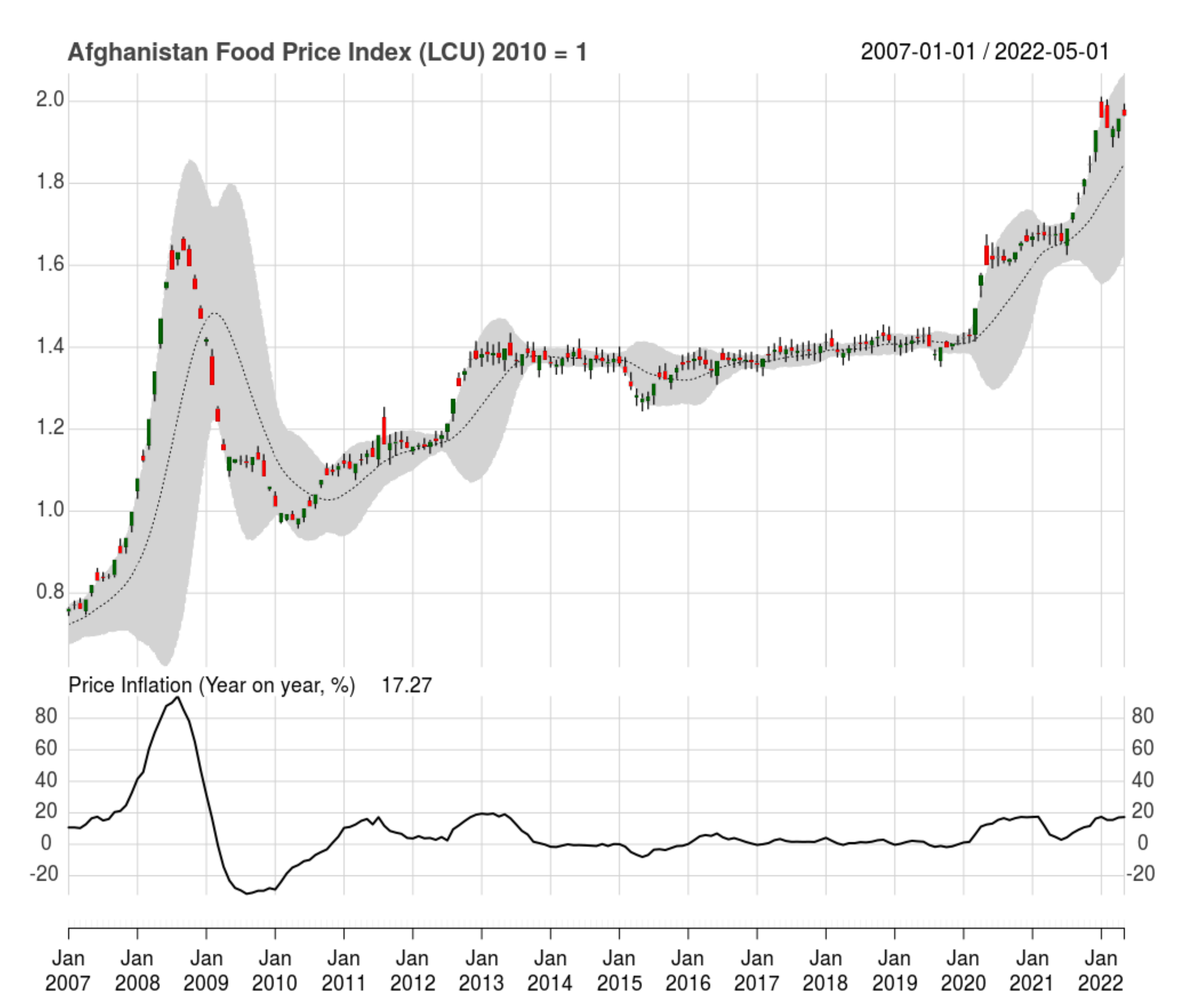
Bread (1 KG), Rice (Low Quality) (1 KG), Wheat (1 KG)

Averaged over 40 markets.

Estimates are updated automatically in the MicroData Library

https://doi.org/10.48529/2ZH0-JF55

Each candle consists of an Open, High, Low and Close price, with candle size reflecting intra-month price spreads. The black dotted line is a 12-month moving average, the gray shaded are envelopes plotted at a 12-month standard deviation level above and below the 12-month moving average and can help determine whether prices are high or low on a relative basis.



Food Price Estimates Afghanistan II

Example of individual price item

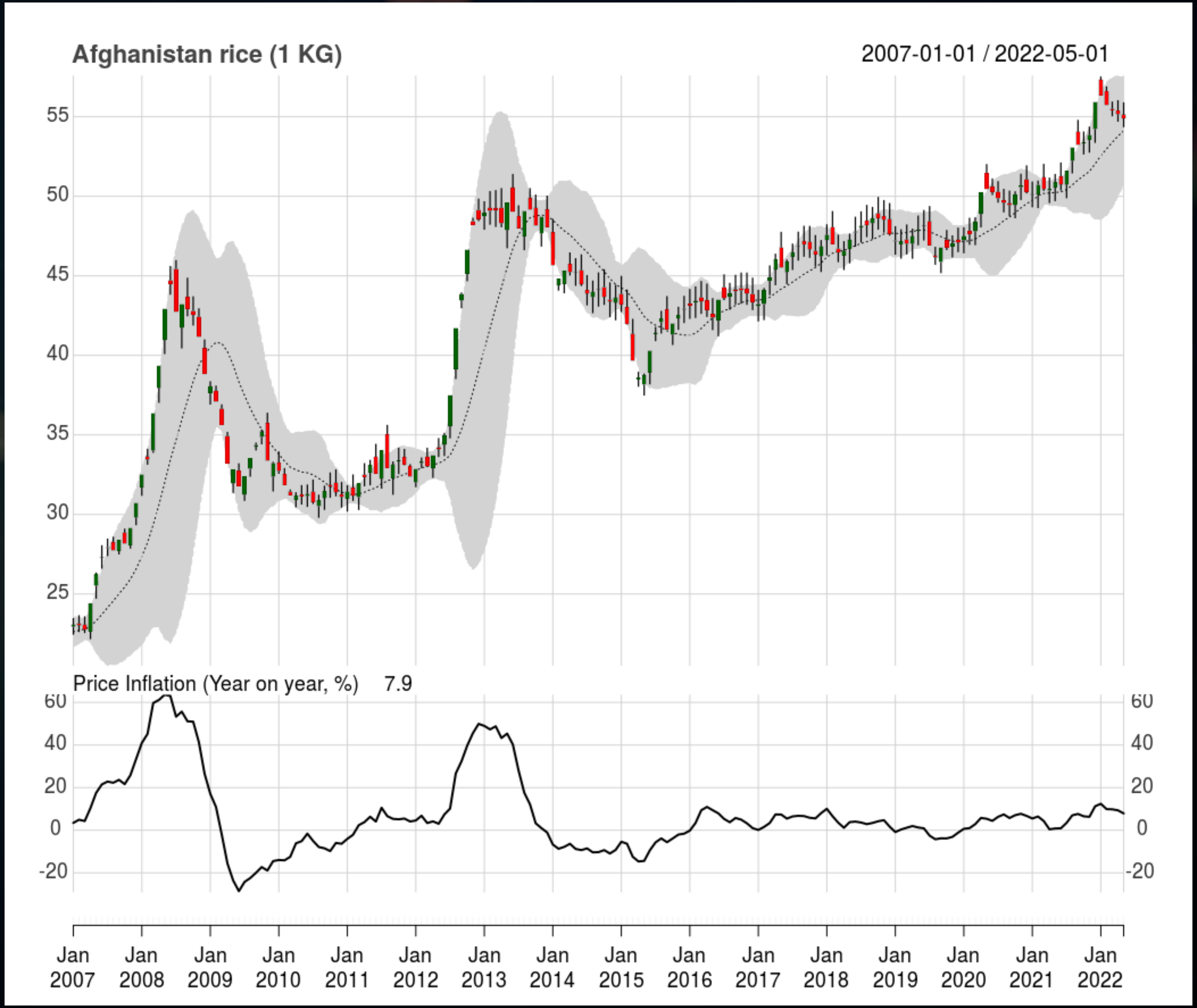
Rice (Low Quality) (1 KG) in LCU.

Averaged over 40 markets.

Estimates are updated automatically in the Microdata Library

https://doi.org/10.48529/2ZH0-JF55

Each candle consists of an Open, High, Low and Close price, with candle size reflecting intra-month price spreads. The black dotted line is a 12-month moving average, the gray shaded are envelopes plotted at a 12-month standard deviation level above and below the 12-month moving average and can help determine whether prices are high or low on a relative basis.



Food Price Estimates Afghanistan III

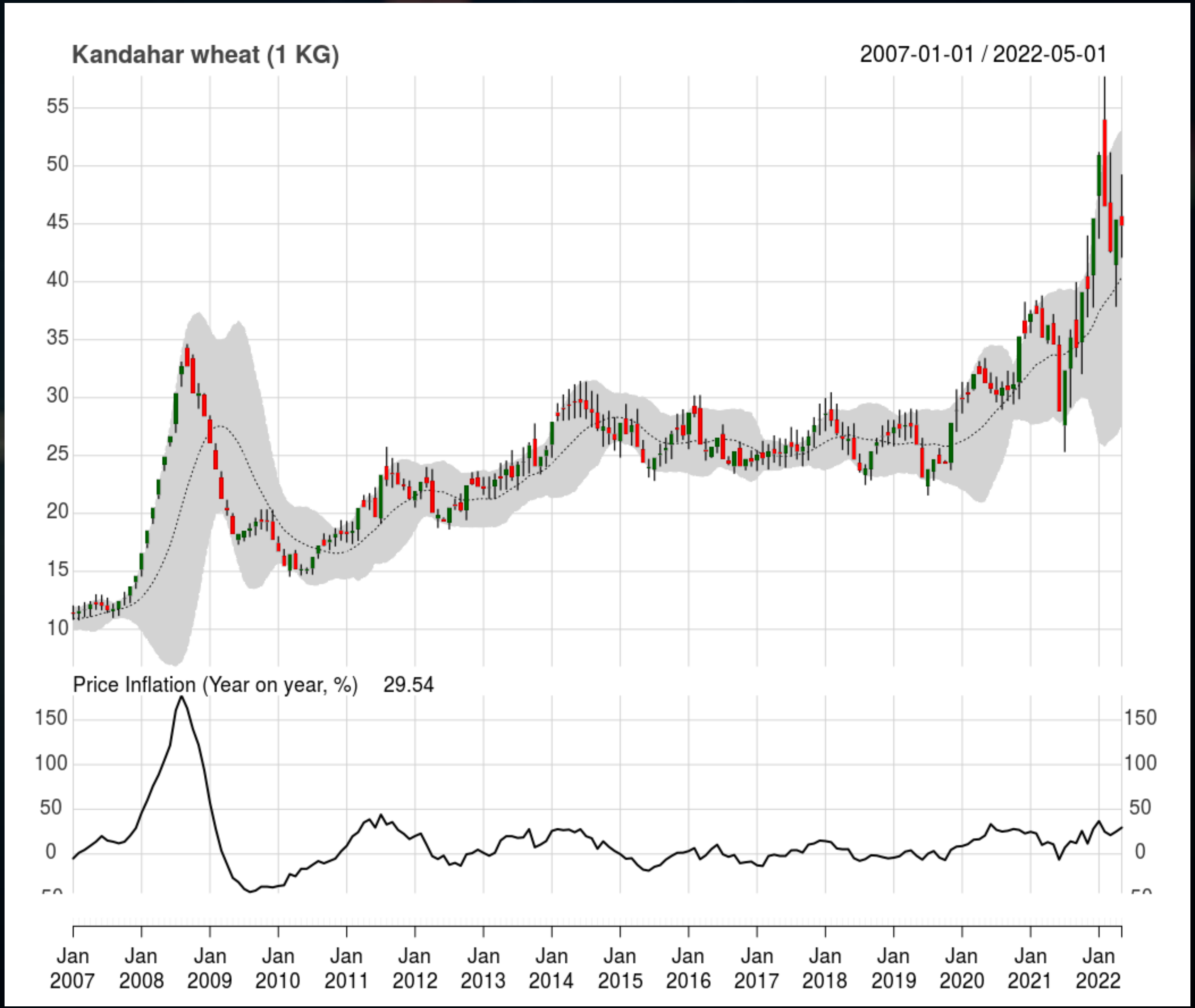
Example of individual price item at a local market

Wheat (1 KG) in LCU.

Estimates are updated automatically in the Microdata Library

https://doi.org/10.48529/2ZH0-JF55

Each candle consists of an Open, High, Low and Close price, with candle size reflecting intra-month price spreads. The black dotted line is a 12-month moving average, the gray shaded are envelopes plotted at a 12-month standard deviation level above and below the 12-month moving average and can help determine whether prices are high or low on a relative basis.



Food Price Estimates Afghanistan IV

Example of price controls breaking

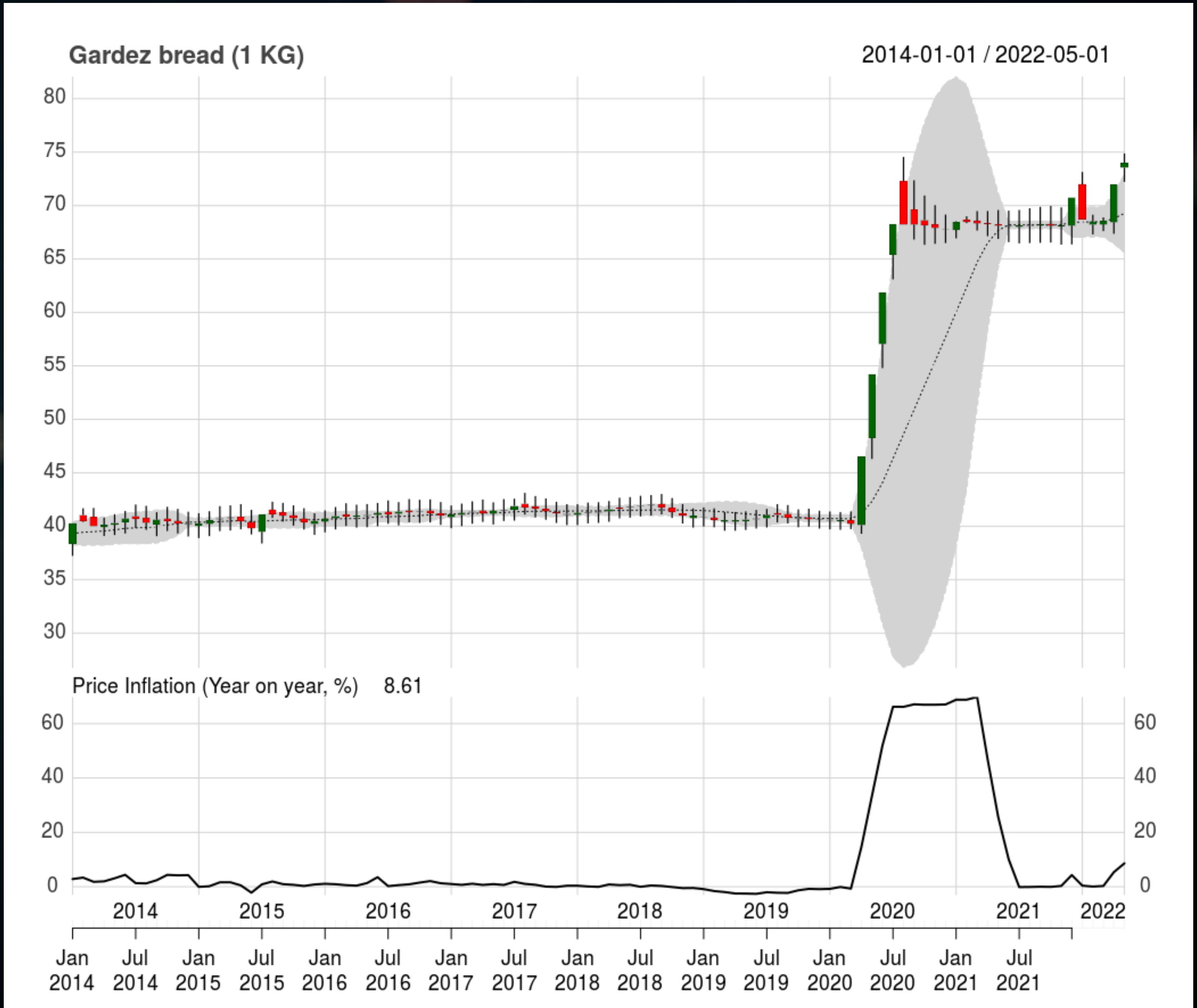
Bread (1 KG) in LCU.

Price at a single market.

Estimates are updated automatically in the Microdata Library

https://doi.org/10.48529/2ZH0-JF55

Each candle consists of an Open, High, Low and Close price, with candle size reflecting intra-month price spreads. The black dotted line is a 12-month moving average, the gray shaded are envelopes plotted at a 12-month standard deviation level above and below the 12-month moving average and can help determine whether prices are high or low on a relative basis.



Papers associated with this work

- Andree, Bo Pieter Johannes. 2021. ***Estimating Food Price Inflation from Partial Surveys*.** Policy Research Working Paper; No. 9886. World Bank, Washington, DC. http://hdl.handle.net/10986/36778

- Andree, Bo Pieter Johannes; Chamorro, Andres; Kraay, Aart; Spencer, Phoebe; Wang, Dieter. 2020. ***Predicting Food Crises.*** Policy Research Working Paper; No. 9412. World Bank, Washington, DC. http://hdl.handle.net/10986/34510

- Wang, Dieter; Andree, Bo Pieter Johannes; Chamorro, Andres Fernando; Girouard Spencer, Phoebe. 2020. ***Stochastic Modeling of Food Insecurity***. Policy Research Working Paper; No. 9413. World Bank, Washington, DC. http://hdl.handle.net/10986/34511

Explore the food price inflation estimates on the *World Bank Microdata Library*

- Monthly food price inflation estimates by country, 25 countries
  https://doi.org/10.48529/2ZH0-JF55

Building the Evidence on Forced Displacement

- https://www.worldbank.org/en/programs/building-the-evidence-on-forced-displacement

Annex

Problems with traditional methods

- Interpolation (say from c_3 to c_6) can only be done **after** c_6 is observed.
- a_4 can only be carried forward.
- Last-Observation-Carried-Forward, rapidly becomes unrealistic.
- Moving averages introduce biases

A (Apples)	*B* (Bananas)	*C* (Coconuts)	*P* (Basket price)
a_1	b_1		
	b_2	c_2	
a_3		c_3	
a_4	b_4^*		
	b_6	c_6	

State space methods are discussed in the paper, but a general specification that works fully automatic across 1000+ markets is too challenging.
Moreover, there is the issue of modeling an interpolation through b_4^*

Algorithm sketch cont.

- Suppose we start with an estimate for a_2 equal to a_1.
- In the first step we estimate a_2 will be replaced by a prediction based on b_2 and c_2.
- In the next step, imputing b_3 benefits from the “upgraded” vector A.
- Prediction performance can be monitored by cross-validating at each step to provide a stopping criterion.
- The process can be repeated several times so that by averaging, any randomness in the process cancels out.

A (Apples)	*B* (Bananas)	*C* (Coconuts)
a_1	b_1	
	b_2	c_2
a_3		c_3
a_4	b_4^*	
	b_6	c_6

Algorithm sketch

Simplified algorithm (see paper for more detailed exposition)

- Column-bind A, B, C into a matrix $\mathbf{P}$
- Iteratively scroll over the columns c of $\mathbf{P}$ and estimate $\mathbf{P}^{c}_{observed} = f^{c}(\mathbf{P}^{-c}_{observed})$ and generate $\widehat{\mathbf{P}}^{c}_{missing}$
- Use $\widehat{\mathbf{P}}^{c}_{missing}$ to replace the previous imputations for $\mathbf{P}^{c}_{missing}$
- Iterate, keep updating imputations.
- Repeat multiple times to generate $\mathbf{P}^{1}_{completed}, \ldots \mathbf{P}^{2}_{completed}, \mathbf{P}^{m}_{completed}$

- Create a final ensemble prediction $\mathbf{P}_{completed}$ by averaging $\mathbf{P}^{1}_{completed}, \ldots \mathbf{P}^{2}_{completed}, \mathbf{P}^{m}_{completed}$
- We can now extract $(A, B, C)_{completed}$ and calculate $P_{completed}$ by taking the basket sum

Validation results, 25 FCS countries

Food Price Data overview (August 2021)

Typical data completeness ranges from 30% to 70%.

Number of markets varies, it is usually more than in CPI.

The paper focuses only on food commodities, but the methods easily extend to non-foods.

Typically, more price items results in better accuracy.

More markets, increases heterogeneity so demands more price items and observations to maintain performance.

TABLE A1—SUMMARY OF RAW FOOD PRICE DATA.

Country	Currency	Markets	Items	Data Coverage	Start date
Afghanistan	AFN	9/40	4	69.27%	Jan-07
Burkina Faso	XOF	63/65	3	47.79%	Jan-07
Burundi	BIF	61/68	7	35.93%	Jan-07
Cameroon	XAF	12/51	11	20.82%	Jan-07
Central Afr. Rep.	XAF	18/40	3	36.89%	Jan-08
Chad	XAF	35/56	3	39.17%	Jan-07
Congo, Rep.	XAF	5/11	7	44.47%	May-10
Congo, Dem. Rep.	CDF	26/83	10	35.43%	Nov-07
Gambia, The	GMD	15/28	11	41.91%	Jan-07
Guinea-Bissau	XOF	3/45	7	56.63%	Feb-16
Haiti	HTG	9/9	6	68.83%	Jan-07
Iraq	IQD	18/18	13	48.02%	May-11
Lao PDR	LAK	17/17	5	46.04%	Feb-12
Lebanon	LBP	26/26	15	61.49%	Mar-12
Liberia	LRD	18/24	3	49.44%	Mar-07
Mali	XOF	77/126	6	58.37%	Jan-07
Mozambique	MZN	25/52	7	63.13%	Jan-07
Myanmar	MMK	36/165	3	43.00%	Apr-07
Niger	XOF	68/79	4	79.60%	Jan-07
Nigeria	NGN	33/35	16	27.59%	May-12
Somalia	SOS	18/28	4	55.49%	Jan-07
South Sudan	SSP	9/20	8	52.35%	Jan-07
Sudan	SDG	14/14	3	58.01%	Jan-07
Syrian Arab Rep.	SYP	36/91	15	56.99%	Aug-11
Yemen, Rep.	YER	24/24	12	45.48%	Nov-08
Average		27/49	7	51.47%	

Note: The first column reports the local currency in which prices are measured, the second column reports the number of markets from which data is used as a fraction of all known market location for which predictions are made, the third columns reports the number of food items for which data is used to construct the food price index, the fourth columns reports the total number of price observations as a share of all *market* × *time* combinations where *markets* is the first number in the third column, the final column reports when the estimated price index starts.
Source: The statistics have been prepared by the author for this paper based on end-of-August (2021) food price data from World Food Program.

Validation Results, 25 FCS countries

The paper uses an outlier-robust out-of-sample R-squared calculated from prediction errors.

The typical performance is around 0.85.

With more prices (adding non-foods) this can be improved.

Prediction quality improves when inflation (deflation) is strong compared to volatility.

When the majority of data is missing, the quality of imputations can still be comparable to the accuracy of direct measurement.

TABLE A4—SUMMARY OF ESTIMATION RESULTS.

Country	Avg. inflation	Max. draw-down	Avg. volatility	CV-score
Afghanistan	5.76%	-44.02%	8.69%	0.87
Burkina Faso	4.08%	-37.74%	15.25%	0.77
Burundi	3.90%	-27.74%	13.26%	0.77
Cameroon	1.25%	-20.40%	7.01%	0.95
Central Afr. Rep.	1.07%	-29.92%	19.36%	0.75
Chad	4.17%	-45.92%	19.02%	0.69
Congo, Rep.	1.56%	-24.17%	11.76%	0.87
Congo, Dem. Rep.	7.63%	-16.67%	8.64%	0.87
Gambia, The	4.92%	-15.31%	10.52%	0.71
Guinea-Bissau	1.62%	-18.56%	14.51%	0.81
Haiti	7.60%	-34.82%	11.98%	0.82
Iraq	-1.07%	-25.58%	5.12%	0.90
Lao PDR	0.83%	-3.44%	1.56%	0.93
Lebanon	20.35%	-21.96%	14.07%	0.92
Liberia	11.67%	-17.95%	10.57%	0.93
Mali	2.46%	-23.97%	8.63%	0.84
Mozambique	9.10%	-29.33%	9.40%	0.84
Myanmar	1.82%	-38.76%	11.83%	0.74
Niger	4.03%	-23.71%	9.84%	0.87
Nigeria	7.82%	-20.89%	7.44%	0.92
Somalia	5.54%	-48.67%	12.95%	0.83
South Sudan	43.03%	-29.55%	25.11%	0.87
Sudan	47.16%	-30.24%	18.93%	0.87
Syrian Arab Rep.	35.25%	-23.49%	17.01%	0.88
Yemen, Rep.	9.55%	-23.84%	14.13%	0.78

Note: The first three columns respectively report average annualized inflation, maximum draw-down, and average annual realized volatility in percentages. The final column reports the cross-validated confidence score that ranges from 0 to 1 for the final food price index using the calculations from the paper. Additional cross-validation statistics can be found on the World Bank Data Catalog page where a live version of the data base is maintained.
Source: The statistics have been prepared by the author for this paper based on end-of-August (2021) food price data from World Food Program.

Toward real-time local food prices:

Example results Yemen

Food Price Estimates Yemen (Feb. 2022)

Equal weighted Food Price Index composed of Oil (Vegetable) (1 L), Rice (Imported) (1 KG), Salt (1 KG), Sugar (1 KG), Wheat (1 KG), Wheat Flour (1 KG), Beans (Kidney Red, 1 KG), Eggs (1 Unit), Lentils (1 KG), Millet (1 KG), Onions (1 KG), Peas (Yellow, Split, 1 KG), Potatoes (1 KG), Sorghum (1 KG), Tomatoes (1 KG). The index consists of average market prices (24 markets).

January and February 2022 prices are preliminary estimates based on time-series estimates.

Each candle consists of an Open, High, Low and Close price, with candle size reflecting intra-month price spreads. The black dotted line is a 12-month moving average, the gray shaded are envelopes plotted at a 12-month standard deviation level above and below the 12-month moving average and can help determine whether prices are high or low on a relative basis.