Policy Research Working Paper 10768

A Data-Driven Approach for Early Detection of Food Insecurity in Yemen's Humanitarian Crisis

Steve Penson
Mathijs Lomme
Zacharey Carmichael
Alemu Manni
Sudeep Shrestha
Bo Pieter Johannes Andrée

WORLD BANK GROUP
Poverty and Equity Global Practice
May 2024

Policy Research Working Paper 10768

Abstract

The Republic of Yemen is enduring the world's most severe protracted humanitarian crisis, compounded by conflict, economic collapse, and natural disasters. Current food insecurity assessments rely on expert evaluation of evidence with limited temporal frequency and foresight. This paper introduces a data-driven methodology for the early detection and diagnosis of food security emergencies. The approach optimizes for simplicity and transparency, and pairs quantitative indicators with data-driven optimal thresholds to generate early warnings of impending food security emergencies. Historical validation demonstrates that warnings can be reliably issued before sharp deterioration in food security occurs, using only a few critical indicators that capture inflation, conflict, and agricultural productivity shocks. These indicators signal deterioration most accurately at five months of lead time. The paper concludes that simple data-driven approaches show a strong capability to generate reliable food security warnings in Yemen, highlighting their potential to complement existing assessments and enhance lead time for effective intervention.

This paper is a product of the Poverty and Equity Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at spenson@worldbank.org and bandree@worldbank.org.

A Data-Driven Approach for Early Detection of Food Insecurity in Yemen's Humanitarian Crisis

Steve Penson,[a*] Mathijs Lomme,[a,b] Zacharey Carmichael,[a] Alemu Manni,[c] Sudeep Shrestha,[b] Bo Pieter Johannes Andrée,[a*1]

Keywords: Agriculture and Food Security, Crisis, Early Warning Systems, Food Price Analysis, Vulnerability, Economic Monitoring

JEL: C01, C14, C25, C53, O10

1 a World Bank, b ACAPS, c FAO, *email: spenson@worldbank.org and bandree@worldbank.org. Funding by the World Bank's Food Systems 2030 (FS2030) Multi-Donor Trust Fund program (grants TF0C0728 and TF0C0828) is gratefully acknowledged. This paper has been prepared as background to the Joint Monitoring Report (JMR), a multi-partner monitoring initiative in Yemen. We would like to thank the peer reviewers on the Quality Enhancement Review for the Yemen Data Driven Identification of Food Security Crises through the JMR, Alan Fuchs, Alexandra Christina Horst, and José Lopez; and the peer reviewers on the Decision Review meeting on Expansion of Real Time Food and Energy Price Monitor, Sergiy Zorya, Nick Haan, Kamau Wanjohi, whose comments helped improve this paper. The authors would like to thank members of JMR Core Development Team including teams at ACAPS, UNICEF, FAO, WFP, WHO and the World Bank. Thanks also to those who contributed to the contextual analysis and methodology development and review including Francesca Marini, Artavazd Hakobyan, Faiza Hesham Hael Ahmed, Nic Parham, Maliha Hussein, Oleg Bilukha, Elijah Odundo, Ismail Kassim, Peter Hailey, Dan Maxwell, Rebecca Semmes, Riham Abuismail, Fawad Raza, Gaurav Singhal, Hussein Gadain, Andres Chamorro, Alia Jane Aghajanian, Olaf De Groot, Emily Henderson, Patrick Vercammen, Felix Leger and Seb Fouquet. Finally, we would like to thank participants to the Joint Monitoring Report workshops including the Yemen Social Fund for Development (SFD) as well as humanitarian, development, and donor partners who have supported the research throughout. This paper reflects the views of the authors and does not reflect the official views of the World Bank, its Executive Directors, or the countries they represent.

1 Introduction

Despite considerable humanitarian assistance, the food crisis in the Republic of Yemen remains one of the world's most dire humanitarian catastrophes in the world. Yemen currently is home to the 5th largest population in the world experiencing crisis levels of acute food insecurity (FSIN and Global Network Against Food Crises, 2023). The latest country-wide multi-partner Integrated food security Phase Classification (IPC) assessment of end of 2022 (IPC, 2022) estimated that 17 million people were in a food crisis, or worse situation, meaning that the population was unable to meet minimum dietary needs without resorting to irreversible coping strategies.

Food security can be assessed across four dimensions requiring that food is available, that individuals can access this food, that food supply and access are stable, and that food provides adequate nutrition (Food and Agriculture Organization (FAO), 2008). Acute food insecurity arises when these dimensions are severely impaired. The food security challenges faced by Yemen are significant, and the number of interrelated factors that can worsen acute food insecurity are many. Humanitarian disasters stem from intricate connections among conflict, poverty, extreme weather, climate, and food price shocks (Misselhorn, 2005; Headey, 2011; Singh, 2012; D'Souza and Jolliffe, 2013), exacerbated by enduring structural factors (Maxwell and Fitzpatrick, 2012), and ultimately lead to high levels of acute malnutrition and mortality in vulnerable populations.

The prevalence of acute severe malnutrition in Yemen has impacted the population, leading to increased vulnerability to health issues and diseases, such as cholera, stunting, wasting, and a variety of both physical and mental health consequences. Prior to the 2015 escalation of conflict, Yemen already had one of the world's highest malnutrition levels. The situation has been aggravated by escalating conflict and economic decline, and recently the overwhelming impact of the COVID-19 pandemic and the war in Ukraine. During this time, many aid projects, including emergency food assistance, WASH services, and malnutrition treatment programs that are highly dependent on continued funding from donor partners (UNICEF, 2020), have been disrupted periodically by funding shortfalls.

Malnutrition has particularly severe impacts on children, leading to long-term declines in cognitive development and potentially enduring health issues. The assessment of Black et al. (2013) on maternal and child undernutrition in low-income and middle-income countries revealed that nearly half of child deaths worldwide were linked to undernutrition. In 2021-2022, Gatti, et al. (2023) estimated that severe food price shocks in the MENA region resulted in hundreds of thousands of children facing long-term consequences, including stunted growth. Additionally, beyond immediate health consequences and loss of life, severe food crises inflict lasting harm on the children of affected families, resulting in intergenerational adverse health and educational outcomes (Galler & Barrett, 2001; Veenendaal, et al., 2013; Galler & Rabinowitz, 2014; Asfaw, 2016).

Recognizing these costs, the international community has responded to Yemen's food crisis with enormous humanitarian aid. According to World Bank data on aid and official development assistance, Yemen received 8 billion dollars in official development assistance alone in 2018. At the $1.90 dollar poverty line, this was sufficient to pay for almost 90% of the annual expenses of the population FAO estimates to have been malnourished that year. In a comprehensive review of 2020 aid programs, Ghorpade and Ammar (2021) estimated that the combined reach of humanitarian and development programs was enough to cover, and in fact exceed, the entire Yemeni population. Alongside aid, there has recently been a growing emphasis on prevention and targeted intervention, as it is often more cost-effective and sustainable to prevent humanitarian catastrophes rather than solely responding to them (Meerkatt, Kolo, & Renson, 2015; Mechler, 2016).

Moving from reactive to pro-active aid requires investment in close monitoring to enable early detection and rapid response when new food security risks emerge (Maxwell & Hailey, 2020). Different early warning and food security information systems already exist to support and inform humanitarian and

development programming, including FEWS NET and the Integrated Food Security Phase Classification (IPC). To date, IPC analyses have provided the primary and common means for tracking food insecurity risks. These major analyses, however, require significant resources and time to conduct and are typically updated on an annual or at best on a semi-annual basis. Large-scale household surveys are infrequent due to access constraints, security issues, and lack of funding. For instance, the World Bank has not carried out a Living Standards Measurement Study and poverty assessment in Yemen since the war broke out in 2015.

While comprehensive analyses are vital for informing programming, targeted humanitarian interventions require more frequent monitoring to mitigate potentially fast-moving developments. The need to enhance the current food security monitoring processes in Yemen is well documented and highlighted for instance by the IPC Famine Review (Maxwell, et al., 2022). To contribute to an improved capacity to predict when, where and how food insecurity escalates, this paper explores data-driven approaches for the early detection and diagnosis of food insecurity emergencies in Yemen. The approach optimizes for simplicity and transparency, and pairs quantitative indicators with data-driven optimal thresholds to generate early warnings of impending food insecurity emergencies.

Previous data-driven approaches to forecast impending emergencies have been pioneered for instance by Mellor (1986), who emphasized economic vulnerabilities, crop failures, and price signals as key indicators of famine. This provides a modeling template that remains in place today. Further insights into price signals and economic deterioration specifically were provided by Seaman and Holt (1980), Cutler (1984) and (Khan, 1994) in the context of the Ethiopian famine of 1972-1974 and the 1984-1985 famine in Niger, and by Andree (2022) who forecasts severe food insecurity in 191 countries based on macro-economic data. More recently, machine learning and time series methods have been employed for prediction at a granular level, as demonstrated by Andrée et al., who predicted local future food crisis using data on food prices, agricultural productivity shocks, and conflict. Using the same data, Wang, et al. model transitions across lower and higher food insecurity phases. Related approaches have since been developed to provide machine-learning driven high-frequency monitoring of food security (Martini et al., 2022).

The proposed methodology in this paper builds upon the existing food security modeling literature, with a specific focus on parsimony, simplicity, and transparency. The proposed approach is a response to recent calls emphasizing the importance of simplicity and transparency in food security modeling (e.g., Baylis et al., 2021; Zhou et al., 2022; McBride et al., 2022). The approach is also informed by the indicator-driven alert system developed by Somalia's Food Security and Nutrition Analysis Unit (FSNAU), which provides automated monitoring capabilities. The proposed methodology extends this system by optimizing alert thresholds at different levels of tolerance for false alerts, benchmarking different indicators, and selecting optimal approaches based on historical validation. This results in a lightweight but effective food security monitoring system that can cater to different targeting strategies simultaneously and supplement existing food insecurity assessments.

Historical validation of the warnings demonstrates that food insecurity emergencies can be reliably detected before they occur, using only a few critical indicators to capture inflation, conflict, and agricultural productivity shocks. The paper concludes that the simple data-driven approaches show a strong capability to detect impending food security emergencies, highlighting their potential to complement existing assessments and enhance lead time for effective, timely, and proactive response.

The paper is structured as follows. Section 2 introduces the data, section 3 develops a framework for the validation and calibration of indicators and thresholds, paying particular focus to balancing false positives and false negatives. Section 4 presents key results, and section 5 concludes. Additional results are found in the supplementary appendices.

2 Data

2.1 Target variable: Emergency outbreaks

This paper aims to predict transitions into critical states of food insecurity with sufficient lead time for action, solely using readily observable indicators. The strategy is to pair the indicators with optimized thresholds to issue reliable warnings before major escalations in food insecurity occur. This work particularly draws on the World Bank's research on Predicting Food Crises (Andrée et al., 2020).

Official IPC data is only available from 2018 to the present. Since the objective is to calibrate indicators to a historical time series of food insecurity situations, historical IPC-compatible data was gathered from FEWS NET covering periodic assessments conducted in 333 districts in Yemen from October 2014 to July 2023. The data quantifies food insecurity using the IPC-compatible analytical framework categorizing the severity of food insecurity and recommending risk mitigation policies (Hillbruner and Moloney (2012) provide a review of the process). The IPC scale distinguishes five phases of food insecurity: (1) minimal/none, (2) stressed, (3) crisis, (4) emergency, and (5) famine/catastrophe. When food insecurity reaches crisis levels, the IPC scale advises a significant policy shift. Specifically, for conditions of stress (2) and below, the focus is on risk management, while at crisis level (IPC Phase 3) and above, it shifts to urgent action to mitigate outcomes (IPC, 2021).

FEWS NET IPC data are reported at a sub-national livelihood zone level. To obtain a consistent time series, the data were mapped to a comparable district level using a spatial overlay and population density was calculated using Meta 2018 high resolution population density maps (Meta, 2024). Table 1 shows summary statistics.

Table 1: Summary statistics of FEWS NET IPC classifications from 2014 to 2023.

Frequency of FEWS NET IPC observations: 8,646. Number of districts: 333.

FEWS NET IPC phase adjusted for aid	Frequency (n=8,646)
1	0
2	678
3	4213
4	3751
5	4
FEWS NET IPC phase transition adjusted for aid	**Frequency (n=8,464)**
1 to 2	0
2 to 3	194
3 to 4	444
4 to 5	4
% of time each district spent in IPC Phase 3+	**Frequency (n=333)**
0% - 20%	9
20% - 40%	0
40% - 60%	0
60% - 80%	30
80% - 100%	294

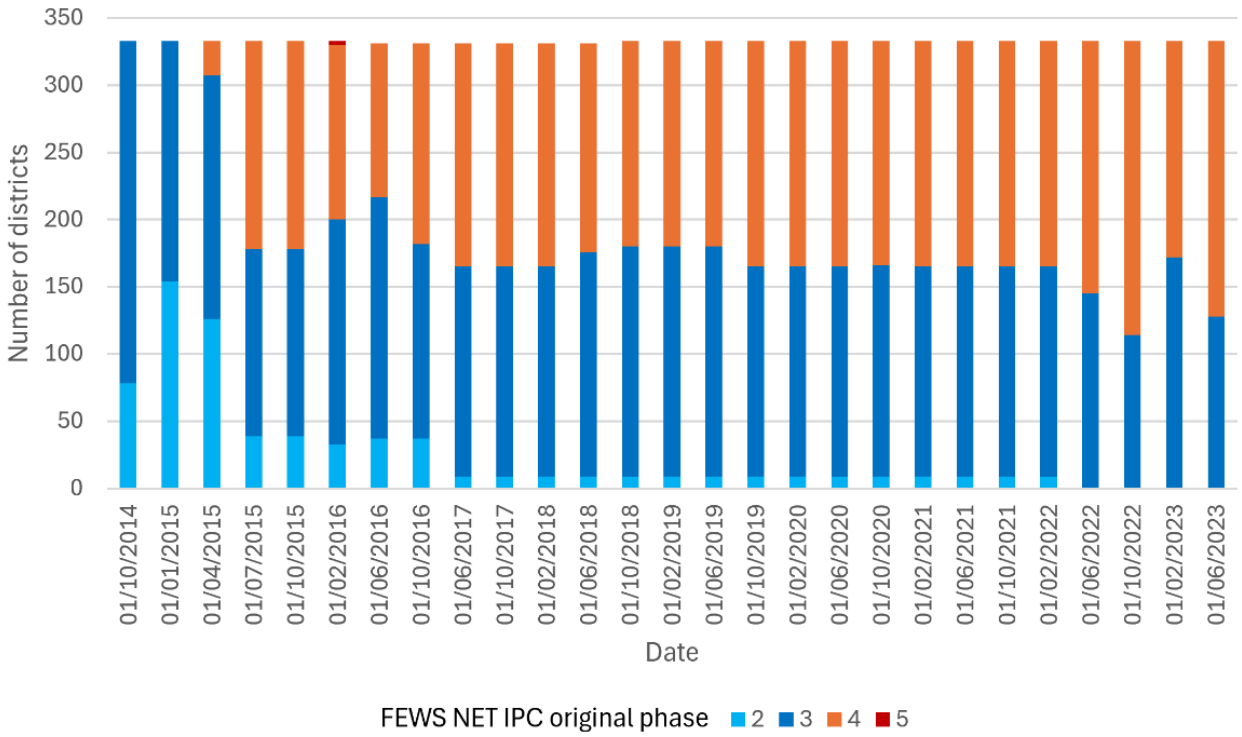
The summary statistics reveal several important insights:

- 92% of IPC observations were in IPC Phase 3+.
- 7% of IPC observations could be classified as IPC escalations when compared to the previous period.
- 88% of districts spent 80%-100% of the time in IPC Phase 3+.

Given most observations fall into IPC Phase 3 or above, the focus of the application is on preventing the transition from IPC Phase 3 to 4. The goal is to issue warnings of impending IPC Phase 4, bolstering prevention efforts. Figure 1 shows the FEWS NET IPC phase distribution between 2014 and 2023.

Figure 1: FEWS NET IPC Phase distribution.

The phase data is netted from humanitarian impacts, the values indicate the five phases of food insecurity: (1) minimal/none, (2) stressed, (3) crisis, (4) emergency, and (5) famine/catastrophe.



2.2 Food security risk indicators

To identify key indicators of worsening food security, datasets with comprehensive spatial and temporal coverage were reviewed. A thorough review of 26 available datasets was carried out to select food security risk indicators. These indicators were reviewed through several factors:

- **Data quality:** The data quality of the indicator.
- **Relation to food security/nutrition:** Whether the indicator is related to food security or nutrition.
- **Risk/outcome indicator analysis:** Whether the indicator is a risk indicator or an outcome indicator.
- **Suitability:** Whether the indicator is suitable for food security risk alert modeling based on expert consultation.

The goal is to leverage this data to spot signs of impending emergencies through historical comparison, issuing alerts at two specific thresholds. These thresholds are adjusted to balance the trade-off between false positives and false negatives effectively. Table 2 highlights the food security risk indicators used by the analysis. These indicators align with food and nutrition security dimensions used by IPC: food access, food availability and food stability (IPC, 2021). Direct measures of nutrition and food utilization were not available.

Table 2: Food security indicators

For each indicator, the dimension corresponds to those recognized by the IPC framework. The method and window have been selected based on historical calibration. Alert and alarm indicate thresholds at which warnings are issued, with alarms indicating critical risks compared to alerts.

Indicator	Description	Dimension	Method	Window	Alert	Alarm
Food prices	Average of top 5 performing food items (YER)	Access	Percentage change from exponential moving average (EMA)	4 months	7.0%-14.2% increase	>14.2% increase
Fuel prices	Average of petrol and diesel price (YER)	Access	Percentage change from moving average (MA)	4 months	15.8%-35.1% increase	>35.1% increase
Exchange rate	YER to USD exchange rate	Access	Relative Strength Index (RSI)	8 months	67.1-75.2	>75.2
Drought	Standardized Precipitation Index (SPI)	Availability	SPI	1 month	-0.88 to-0.12	<-0.88
Conflict	District and neighboring districts averaged conflict fatalities	Stability	RSI of the EMA	14-month RSI over 12 month EMA	64.5-90.2	>90.2
Displacement	Summed displacements (from and to)	Stability	RSI of the EMA	14-month RSI over 6 month EMA	55.1-68.7	>68.7

Below, each data source is discussed and detail the different indicators that were constructed. Details on the formulas used to implement the different indicator methods are available in Annex I.

2.2.1 Food prices

In Yemen, the affordability of key food commodities is a critical indicator of food security levels. Rising food prices, currency devaluation, disruptions in public salary payments, and diminished job opportunities have significantly decreased purchasing power, leaving more people unable to meet their basic needs (ACAPS, Mercy Corps, 2020) (ACAPS, 2023). Several food price indicators were explored to capture these drivers including individual food item analysis, custom food item baskets and the humanitarian Standard Minimum Expenditure Basket (SMEB) (Cash and Markets Working Group Yemen, 2022). Data is used from the World Bank's RTFP data derived from WFP surveys, capturing monthly food prices at a district level (Andrée, 2021, Andrée, 2023a, Andrée, 2023b, Adewopo et al., 2024). This methodology integrates actual data and machine learning estimates to monitor continuous food prices across, filling in gaps where direct market data collection was not possible due to access issues in real time.

Three methods were tested to analyze food prices: the percentage deviation from the Moving Average (MA), the percentage deviation from the Exponential Moving Average (EMA), both across time-windows ranging from 1 to 12 months, and the Relative Strength Index (RSI), across time-windows ranging from 6 to 14 months. Both YER and USD food prices were modeled but separating these as two indicators worked better. In addition to the five food categories in the SMEB, prices for individual Real Time Food Prices (RTFP)[2] food items were analyzed. Results indicated a strong predictive power of certain food categories, specifically imported foods such as beans, millet, sorghum, sugar, and wheat flour. A 5-item basket consisting of these food items reached almost 10% lower loss than the SMEB basket. This resulted

[2] RTFP: https://microdata.worldbank.org/index.php/catalog/study/WLD_2021_RTFP_v02_M

in our selection of the average price of these five food items as our target indicator. The optimal method and time-window found was the percentage deviation from a four-month EMA. For detailed statistical outcomes, refer to Annex II.

2.2.2 Fuel prices

Fuel affordability remains a critical issue affecting food security in Yemen, with rising fuel prices increasing food distribution costs and, consequently, food prices (ACAPS, 2023). Yemen's dependence on fuel imports makes it susceptible to fluctuations in international oil prices, impacting internal fuel prices and driving up overland transportation costs and food prices. Additionally, higher fuel prices escalate the cost of living, pushing more people below the basic needs affordability threshold.

Modeling employed the World Bank's RTEP[3] data (Andrée, 2021, Andrée, 2023a, Andrée, 2023c) derived from WFP surveys, capturing monthly food prices at the district level. For fuel price analysis, the same methods were used to developed indicators from the food price data. Diesel and petrol showed high predictive accuracy individually compared to gas prices as well as all other indicators. To capture robust signals from the data, a fuel basket was specified consisting of an equal part of petrol and diesel. The percentage deviation from the 4-month EMA emerged as the most effective method. For detailed statistical outcomes, refer to Annex II.

2.2.3 Exchange rate

Exchange rate volatility significantly influences food price changes, and particularly in areas controlled by the Internationally Recognized Government (IRG) of Yemen. In January 2020, the Central Bank of Yemen (CBY) in Sana'a prohibited the use of new Yemeni rial (YER) banknotes issued by the IRG-controlled CBY in Aden, leading to a dual currency system. The printing of new YER banknotes by CBY Aden to finance the IRG's budget deficit has depreciated the YER against the USD, increasing the cost of goods and services for households. This depreciation affects purchasing power and, consequently, food security, underscoring the importance of continuous monitoring (ACAPS, 2023). The new notes are not accepted in areas controlled by the Ansar Allah (AA), resulting in a dual currency system with a pronounced inflation differential.

Two data sources for the exchange rate were considered: a Telegram source recording rates several times a week in both Aden and Sana'a, and the World Bank's RTFX data (Andrée, 2021, Andrée, 2023a, Andrée, 2023d) derived from WFP surveys, which provides monthly averaged rates per governorate. The World Bank's Real Time Exchange Rates (RTFX)[4] data, offered more detail, better data coverage, and better modeling results and was selected as the exchange rate indicator for the analysis. The same indicators were constructed as for the fuel and food prices. Analysis over 6 to 14-month time windows revealed an 8-month RSI as most effective in signaling escalation risks. For comprehensive statistical details, see Annex II.

2.2.4 Drought

Drought significantly impacts food security in Yemen, exacerbating water scarcity, reducing crop yields and livestock productivity, increasing reliance on expensive food imports, and intensifying the humanitarian crisis. It can also cause displacement and resource conflicts, further destabilizing the region. Despite imported food accounting for 83% of Yemenis' daily caloric intake (ACAPS, 2023), making drought a relatively less impactful food security risk compared to food prices or exchange rates, drought remains critical due to the role of agricultural in over half of the households' employment and the significance of cash crops like Qat on income and expenditure, affecting household purchasing power (UNDP, 2022).

[3] RTEP: https://microdata.worldbank.org/index.php/catalog/study/WLD_2023_RTEP_v01_M
[4] RTFX: https://microdata.worldbank.org/index.php/catalog/study/WLD_2023_RTFX_v01_M

For drought analysis, rainfall data (Funk et al., 2015), Normalized Difference Vegetation Index (NDVI) (USGS, 2023), and Standardized Precipitation Index (SPI) (Guttman, 2007) were evaluated. Various methods, including Z-score (mean and median), moving averages of the anomaly rates, and the SPI, were tested. Crop calendar information for Yemen was used (FAO-FSIS, Government of Yemen, 2018) to determine the critical months for rainfall in crop growth, and indicators were tested that used the non-crop months to fix the signals at 0 in the data. Z-scores considering a crop calendar resulted generally in improved results. However, the best results came from the year-round data SPI method, which outperformed crop-growing only months. The predefined general thresholds recommend for the SPI to indicate severe droughts are a value of -1.3 and -1.6 to indicate extreme droughts. The optimized thresholds given by our model are higher than the standard SPI thresholds. For detailed statistical outcomes, refer to Annex II.

2.2.5 Conflict

Conflict affects general security and the movement of people, crucial for agricultural and fisheries production and market access. Globally in 2017, 60% of undernourished people and 79% stunted children lived in conflict-affected areas (FAO, IFAD, UNICEF, WHO, and WFP, 2017). Studies indicate that conflict's impact on food security and nutrition worsens with prolonged conflict and weak institutional response (Holleman, Jackson, Sanchez, & Vos, 2017), as seen in Yemen.

ACLED data, tracking monthly conflict incidents and fatalities at the district level, was utilized. Conflict intensity was gauged by fatalities and incidents per district. Incidents include battles, explosions/remote violence, or violence against civilians. The analysis considered conflict at both the district and surrounding district levels to capture exposure to nearby conflict, defined as occurring within either one or two levels of proximity.

Methods tested for both direct and neighboring district conflict data included percentage deviations from MA and EMA, and the RSI of these EMAs. For neighboring conflict, the averages over the neighboring districts were constructed first, before calculating the moving averages over time. The interpretation of this is that the RSI measures whether the EMA (compounding trend) of exposure to local/regional conflict is surging or easing. Of the indicators tested, neighboring conflict within one level proximity outperformed both local and two-level proximity indicators. The best performing was neighboring conflict fatalities using RSI over a 14-month period based on the EMA of the last 12 months. This indicator picks up on escalations, and peaks in reaction to the compounding impacts of sustained violence. Detailed statistical results are available in Annex II.

2.2.6 Displacement

Displacement amplifies demand for food and services in host areas and often signifies severe livelihood disruptions. Research indicates significant impacts of forced migration on wages, household income, consumption, wellbeing measures, and employment in both origin and destination communities (Calderón-Mejía & Ibáñez, 2016; Foged & Peri, 2016; Kreibaum, 2016; Maystadt & Duranton, 2019; George & Adelaja, 2022; Esen & Binatli, 2017; Ruiz & Silva, 2015). The UN International Organization for Migration's Displacement Tracking Matrix (IOM-DTM) provides monthly data on displacements to and from districts since 2014. In modeling displacement, total displacements to, from, and the aggregate of both were examined, using the same analytical methods as for conflict modeling.

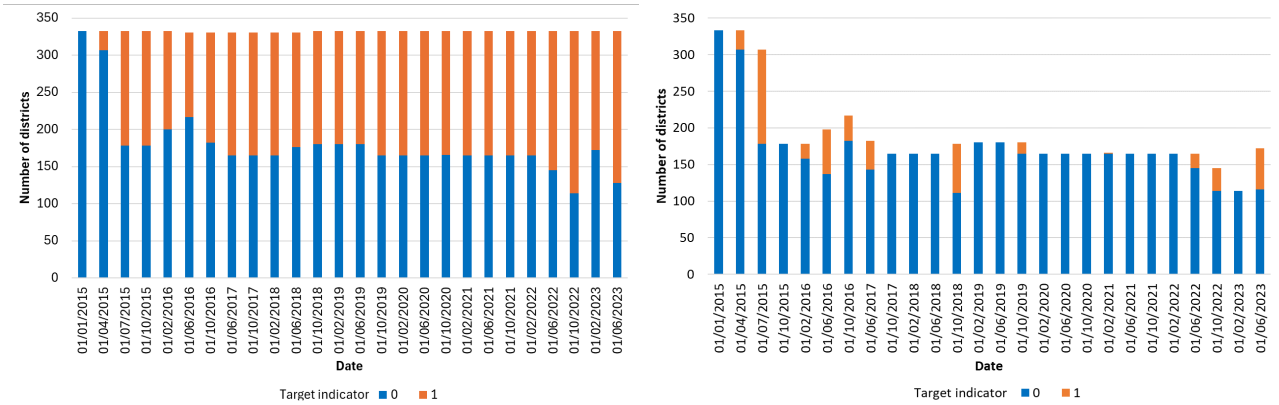
The combined total of displacements to and from a district was chosen as the indicator, with the RSI over an EMA identified as the optimal method. The analysis employed a 6-month time window for the EMA calculation and a 14-month time window for the RSI calculation. For detailed statistical outcomes, refer to Annex II.

3 Methods

To set and test thresholds for food security indicators, a binary target variable was created from the IPC phases, adjusting for humanitarian impacts. Areas in crisis (IPC Phase 3) that would reach emergency status (IPC Phase 4) without aid were marked as emergencies. This approach aims to predict intervention needs rather than outcomes. Thus, IPC Phases ≤3 were coded as 0 (non-emergency) and ≥4 as 1 (emergency). A focused sample was then taken that highlights district escalations, targeting instances where the prior period's target was 0, ensuring only genuine food security escalations are tracked. Figure 2 displays the data. Thresholds for each indicator were then fine-tuned to best front run these new food insecurity emergencies.

Figure 2: Distribution the target variable.

Distribution of binarized indicator used to calibrate alerts, left). Distribution of escalation events, used to validate the early warning capability, right).



Notation-wise, this can be detailed as follows. Specifically, the target variable takes the value 1 if a food security emergency is observed, defined as IPC categories 4, or 5, and 0 otherwise. For convenience, 0 and 1 can be referred to generically as "class labels," with observations corresponding to food security emergencies being the "positive class." Y was defined as a column vector containing the class labels, encompassing all districts and months where an IPC rating was observed. X is the food security indicator defined as a corresponding vector with indicator values used to generate warnings. $\hat{Y}$ was defined as the corresponding column vector of binary predictions generated according to the following threshold rule:

$$\hat{Y} = \begin{cases} 1 & if\ X > c \\ 0 & else \end{cases} \quad (1)$$

c is defined as a numeric threshold. Warnings for time $t+1$ are always generated with information available at time t. In other words, c represents the threshold at which an observed value for any given indicator is high enough to predict that the IPC rating in the next time step will be at least IPC Phase 4. The threshold is optimized (see Table 2 for values) to minimize a loss function that evaluates the ability to front run transitions into emergencies by validating against a specific subset of data: district/month observations with an IPC phase rating of 4 and above, provided the prior rating was below 4. More specifically, let Y^* and $\hat{Y}^*$ be defined as the sub-vectors of Y and $\hat{Y}$ that included the entries preceded by non-emergencies. A conformable vector was used of ones denoted as I and a scalar weight denoted as w to evaluate predictive performance using this prediction loss function:

$$L = w\frac{Y'(I-\hat{Y})}{Y'Y} + (1-w)\frac{(I-Y)'\hat{Y}}{(I-Y)'(I-Y)} = w * FNR + (1-w) * FPR \quad (2)$$

This loss function is a weighted average of the False Positive Rate (FPR) and False Negative Rate (FNR) measured against emergency escalations. L is oriented so that lower values correspond to better predictions. The special cases where w is equal to the rate of occurrence of the positive class or equal to 0.5, the value for $1 - L$ equals the standard Accuracy and common Balanced Accuracy rates. The weight factor determines if the FNR or the FPR is more penalised. Values of w that place increasing weights on false negatives are used to determine thresholds to generate warnings for different risk levels. The settings correspond to:

- w = ⅓: failing to recognize an emergency (false negative) is half as costly as raising a false warning
- w = ½: failing to recognize an emergency (false negative) is just as costly as raising a false warning
- w = ⅔: failing to recognize an emergency (false negative) is twice as costly as raising a false warning

For each risk indicator, the weight factor w = ½ was used to find an optimal method and time-window. The solution spaces per indicator were trimmed down following several filtering steps. First, we filter out all solutions where the Loss (w = ½) >= 0.5 and where the FPR >= 0.5. This excludes calibration results that are uninformative, or attain low average loss by raising warnings most of the time potentially leading fatigue. Solutions where FNR > 3*FPR, and solutions where FPR > 2*FNR, were removed as well, to avoid extremely unbalanced error profiles that may be undesirable for similar reasons. Finally, two method-specific filtering steps were followed to remove implausible solutions. First, for RSI indicators, solutions involving threshold values below 50, which indicate decreases instead of increases, were excluded. Second, for Z-score methods, solutions involving threshold values above 0.5 were excluded. The Z-scores are used for the drought indicators, where positive values indicate an increase relative to the benchmark values, which is the opposite of what these indicators are intended to capture. After these filtering steps, the optimal performing method/time-window combination is selected, with which we determine the two thresholds.

Having found an optimal method and time-window, weight factors of 1/3 and 2/3 respectively were used to determine threshold values that minimize loss in two scenarios. For ***alerts*** *indicating heightened risks*, emphasis was on reducing false negatives (w = ⅔:), leading to more frequent warnings to catch potential emergencies early. For ***alarms*** *indicating critical risks*, the focus was on lowering false positives (w = ⅓), resulting in fewer but more certain warnings. The two-step approach ensures that alarms are generated from the same distribution as alerts, only at higher risk levels. This means that heightened alerts capture the majority of potential emergencies, while the subset of more conservative critical alerts capture those that are most likely to escalate.

4 Results

Two sets of results were produced from the analysis. Initially, the indicators were examined individually to evaluate their predictive capabilities and to establish thresholds for generating alerts (signaling elevated risks) and alarms (indicating critical risks). Following the identification of optimized thresholds, a multivariate analysis was conducted. This analysis aimed to determine the most effective way to combine these optimized indicators, assess the incremental accuracy of each indicator, and ascertain the ideal number of indicators for monitoring. This comprehensive approach ensures a nuanced understanding of each indicator's contribution to an overall risk assessment and facilitates the development of a robust model that can accurately predict areas at risk of worsening food insecurity.

4.1 Individual indicators

Table 3 presents the univariate validation results of the six main indicators from Table 2. The statistical validation underscores varied performance and differential sensitivity and specificity across indicators.

Recall that alarms are issued based on a more conservative calibration of thresholds compared to alerts. This trades false and true positives, aiming to minimize the occurrence of false alarms while maintaining the capacity to detect true emergencies.

Among the indicators, price data emerge as the most important group of indicators to signal deteriorating food security conditions. Food prices emerge as the most predictive indicator of lower risk levels (alerts) and demonstrate a relatively balanced performance with a moderate FPR of 0.27 and FNR of 0.33 for alerts, and a significantly reduced FPR of 0.05 for alarms, highlighting a successful calibration towards conservatism for critical risks. This is mirrored by the loss values, which shift from 0.31 for alerts to a more favorable 0.28 for alarms, indicating an effective rebalancing between sensitivity and specificity at critical risk levels. Fuel prices produce similar alerts but emerge as the more predictive indicator for critical risks (alarms). Fuel price alarms result in a lower FNR (0.45 as opposed to 0.73 for food prices), with slightly higher FPR (0.16, compared to 0.06 for food price alarms). Together, this provides an improved calibration for critical risks (loss value of 0.26).

Table 3: Food security risk indicator validation

Summary of statistical validation for food security indicators, delineating calibrated thresholds for 'Alerts' and 'Alarms'. 'Alerts' are optimized to lower false negatives, enhancing the frequency of warnings for early emergency detection. 'Alarms' focus on reducing false positives to ensure the reliability of critical risk warnings, facilitating targeted and timely responses.

Risk indicator	**Food prices**		**Fuel prices**		**Exchange Rates**	
Method	Percentage of EMA		Percentage of MA		RSI	
Risk level	**Alert**	**Alarm**	**Alert**	**Alarm**	**Alert**	**Alarm**
True positives	298	121	286	246	321	248
False positives	1,144	234	1,196	687	1,923	1,464
True negatives	3,131	4,041	3,079	3,588	2,352	2,811
False negatives	146	323	158	198	123	196
Kappa	0.20	0.24	0.18	0.26	0.10	0.09
Accuracy	0.73	0.88	0.71	0.81	0.57	0.65
Balanced Accuracy	0.70	0.61	0.68	0.70	0.64	0.61
F1 Score	0.32	0.30	0.30	0.36	0.24	0.23
Precision	0.21	0.34	0.19	0.26	0.14	0.14
FPR	0.27	0.05	0.28	0.16	0.45	0.34
FNR	0.33	0.73	0.36	0.45	0.28	0.44
Loss (Alert, w=2/3; Alarm, w=1/3)	0.31	0.28	0.33	0.26	0.33	0.38
Risk indicator	**Drought**		**Conflict**		**Displacement**	
Method	SPI		RSI of EMA		RSI of EMA	
Risk level	**Alert**	**Alarm**	**Alert**	**Alarm**	**Alert**	**Alarm**
True positives	258	129	228	105	212	129
False positives	1,492	647	1,217	557	952	548
True negatives	2,783	3,628	3,058	3,718	3,323	3,727
False negatives	186	315	216	339	232	315
Kappa	0.10	0.10	0.11	0.09	0.15	0.13
Accuracy	0.64	0.80	0.70	0.81	0.75	0.82
Balanced Accuracy	0.62	0.57	0.61	0.55	0.63	0.58
F1 Score	0.24	0.21	0.24	0.19	0.26	0.23
Precision	0.15	0.17	0.16	0.16	0.18	0.19
FPR	0.35	0.15	0.28	0.13	0.22	0.13
FNR	0.42	0.71	0.49	0.76	0.52	0.71
Loss (Alert, w=2/3; Alarm, w=1/3)	0.40	0.34	0.42	0.34	0.42	0.32

Exchange rates, analyzed through RSI, exhibit the highest FPR among all indicators reaching 0.45 for alerts, and decreasing to 0.34 for alarms. Despite this reduction, the high initial FPR points to a significant over-triggering tendency, potentially leading to alarm fatigue. Nevertheless, its loss values, from 0.33 for alerts to 0.38 for alarms, indicate moderate predictive value, with room for improving specificity. Our further analysis shows that the indicator performs much better when taking a North/South divide into consideration.

Drought indicators, utilizing the SPI at default thresholds showed a stark contrast in FPR between alert (0.05) and alarm (0.03) levels, suggesting a very conservative approach that unfortunately results in a high false negative rate (FNR) of 0.94 for alerts and 0.96 for alarms. This extreme conservatism, reflected in its loss values (0.64 for alerts and 0.34 for alarms), suggests standard threshold values may be under-predicting true risks, limiting its utility in timely emergency responses. Optimizing the thresholds had a strong impact on improving the results lowering loss (0.40 for alerts and 0.34 for alarms albeit with a more balanced error mixture), reaching similar loss values as the conflict and displacement indicators.

Conflict and displacement indicators, employing RSI of the EMA of local and neighboring conflict fatalities, highlight a different challenge. With notably high FNRs at the alert level (0.49 for conflict and 0.52 for displacements), which increase to 0.76 and 0.71 for alarms respectively, these indicators show a propensity to under-predict emergencies. Their loss values reflect this, with a significant reduction in predictive error from alerts to alarms, for conflict (from 0.42 to 0.34) and displacements (from 0.42 to 0.32), driven by lower FPRs and a better performance at the critical risk level.

To examine the impact of slight adjustments in thresholds, kernel density plots for the six indicators were generated (Figure 3), based on continuous values with orange lines for alert triggers and red for critical alarms. The plots reveal that alerts trigger more frequently, and allow some deterioration in indicator values to occur before escalating to alarms.

Figure 3: Kernel density plots for the six indicators.

Orange line = alert. Red line = alarm. Kernel densities are estimated and indicative of smooth approximated distributions and may indicate some mass at impossible values, such as RSI values outside of the 0 -100 range.

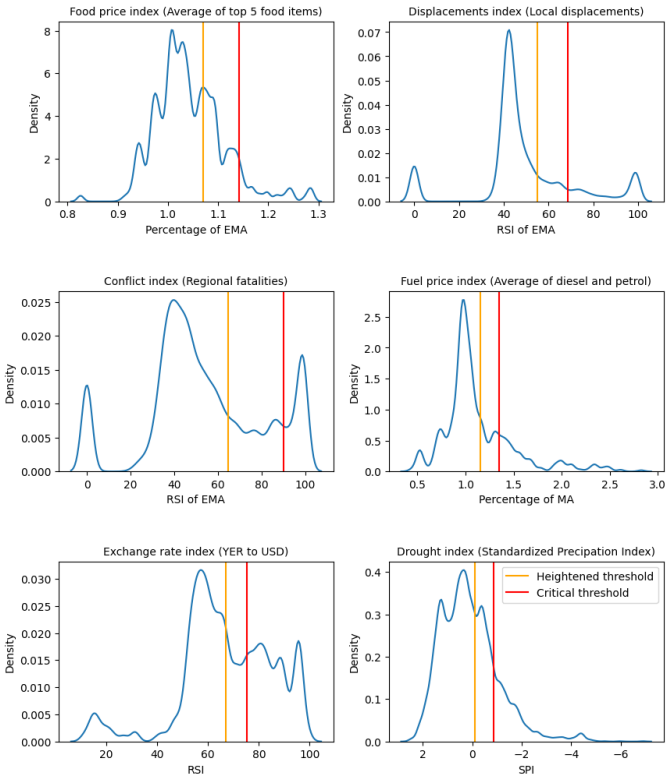


Figure 4: Historic food security warnings 2009 to 2024

The graph displays the percentage of districts for which alerts and alarms are issued. Alerts are plotted in a cumulative sense in that they remain in place when the risks transition into alarms. Annex III contains a detailed timeline outlining critical food security events from 2014 to 2023, as compiled by food security experts. Accompanying this timeline are enlarged versions of the figure below, focusing on specific periods to provide clearer insights.

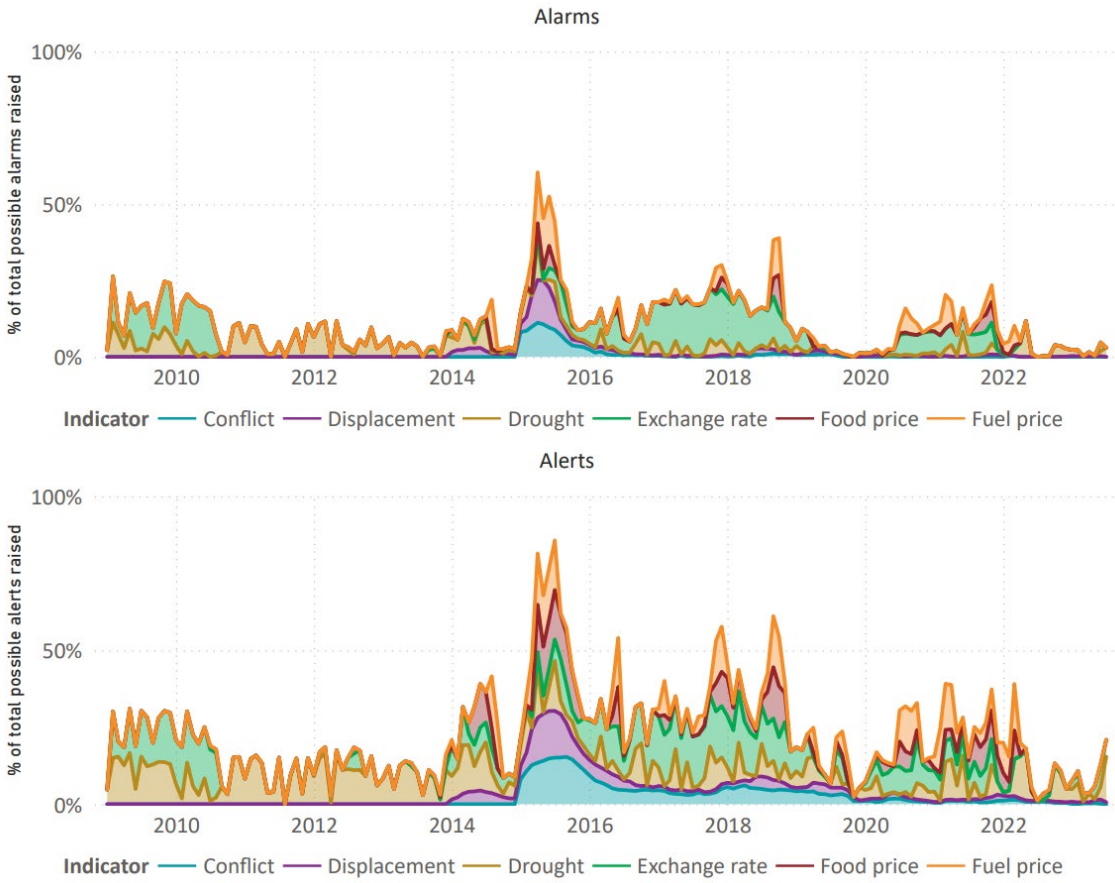


Figure 4 illustrates the nationwide historical distribution of districts receiving alerts and alarms over time, indicating that alarms that signal higher risk levels are generally issued more conservatively and follow after alerts have already indicated preceding risk levels. Several periods stand out. Notably, the period of 2010-2011, marked by a sharp currency devaluation reflective of Yemen's severe internal conflict and the onset of the Arab Spring protests. The subsequent period from 2011 to 2014 appears relatively stable with a sporadic issuance of drought alerts. It is crucial to acknowledge that displacement data only begins in 2014, and conflict data in 2015, implying that the full spectrum of potential alerts and alarms prior to these years is not captured. From 2014 onward, the displacement and conflict data contribute to a stark rise in the issuance of alerts and alarms, with the food and fuel price data highlighting additional risks during the Saudi-led port blockade in 2015. A significant increase in currency devaluation and inflation from 2017, coupled with escalating conflict, culminated in a peak in alarm issuance in 2018. Lastly, a widespread inflationary surge across all price indicators triggered another spike in alarms around 2022. Finally, while the major food insecurity periods are marked by escalations in economic and conflict indicators, the drought alarms do not center on any specific year and the highest count of alarms occurs when economic and conflict risks coincide with periods of drought.

4.2 Multivariate indicator analysis

The univariate validation exercise confirmed the relevance of each indicator while also underscoring the limited predictive capability of relying solely on one alone. In practice, monitoring multiple indicators

raises the question of how best to integrate these measures and evaluate the relative value of additional indicators. Annex IV includes correlation matrices for the various indicators, revealing some correlations between different warnings. Notably, food and fuel prices show a stronger correlation with each other, whereas displacement and conflict warnings are more closely related. This indicates that combining either a conflict or displacement indicator with food prices may yield a more comprehensive overview than pairing food and fuel price indicators.

To systematically evaluate this, the six indicators were analyzed using a Logit model against the target indicator. The analysis focused on how the integration of individual indicators enhanced the model's performance in predicting the target indicator. To explore the incremental accuracy of each indicator, we conducted (inverse) Recursive Feature Elimination (RFE). Leveraging the outcomes from the Generalized Linear Models (GLM), we ranked the indicators by their importance, with the most influential indicator positioned first. We then forecasted the outcome, determining the optimal threshold for dichotomizing the outcome into binary results and computing the Balanced Accuracy. Subsequently, the second-ranked indicator was incorporated, and the procedure was iterated, recalculating the Balanced Accuracy with each additional indicator until the metric was derived for the full set of indicators. The findings are depicted in Figure 5. The analysis indicates that sequentially adding indicators in order of importance improves predictive performance, with the most substantial improvements observed upon adding the initial indicators. However, performance saturates or may even drop beyond a certain point, indicating that combining multiple indicators may lead to over-fitting.

Table 4 presents the regression outcomes for the GLM employing RFE. The initial three models are based on: (1) normalized continuous indicator data, (2) binary alert data, and (3) binary alarm data. A Brier score nearing zero and a pseudo R-squared value, computed as 1− log loss/ uninformative log loss, approximating 0.6 indicate that these simple models possess a commendable predictive capability regarding actual escalations in food security.

Table 4: Foundational GLM results

Regression results for three models: 1) normalized continuous indicator data, 2) binary alert data, and 3) binary alarm data. To assess model performance, the Brier score, pseudo R-squared and a weighted average of error types, optimizing the probability cut-off used of classification, were calculated.

	(1) GLM Continuous	(2) GLM Alert	(3) GLM Alarm
Food price	3.805***	1.061***	0.812***
	(0.298)	(0.138)	(0.149)
Fuel price		0.737***	1.453***
		(0.137)	(0.139)
Exchange rate		1.213***	0.868***
		(0.119)	(0.113)
Displacement			0.261*
			(0.153)
Conflict			-0.048
			(0.160)
Drought		0.888***	0.343**
		(0.111)	(0.140)
Pseudo R2	0.574	0.621	0.615
Brier	0.082	0.074	0.075
Loss w=1/2	0.298	0.252	0.275
Loss w=1/3	0.289	0.275	0.256
Loss w=2/3	0.308	0.230	0.295
Note:			*p<0.1; **p<0.05; ***p<0.01

The results in Table 4 help understand the interplay of different indicators when used jointly to track overall food insecurity risks. Strikingly, when using continuous indicator values and employing RFE, the best forecasting power was achieved when only using the food price indicator. In contrast, models (2) and (3) that respectively use binary indicators with optimized thresholds as inputs, reach considerably

lower loss and utilize more indicators. Combining alerts yields the lowest loss values (for instance 0.23 for w=2/3, which is considerably below the univariate scores in Table 3). In conclusion, the dichotomization into alerts and alarms does not only provide more interpretable warnings as compared to the level readings of the continuous indicators, it also helps improve forecasting in a simple linear modeling framework. Furthermore, the correlation matrices in Annex IV reveal that the dichotomization reduces correlations between the data, which ensures that fewer warnings raise simultaneously.

Figure 5: Marginal accuracy of basic indicators

The vertical axis plots cross-validated balanced accuracy using a Generalized Linear Model (GLM), the horizontal axis shows how prediction performance evolves as indicators are combined. From right to left, indicators are dropped in order of significance, the first indicator thus being the most dominant. Results are for Model (2) (left), Model (3) (right).

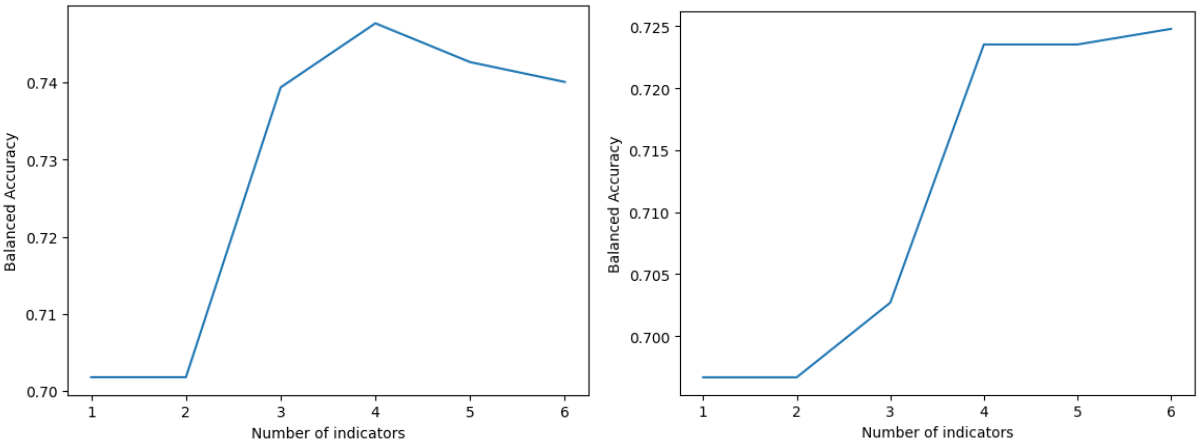
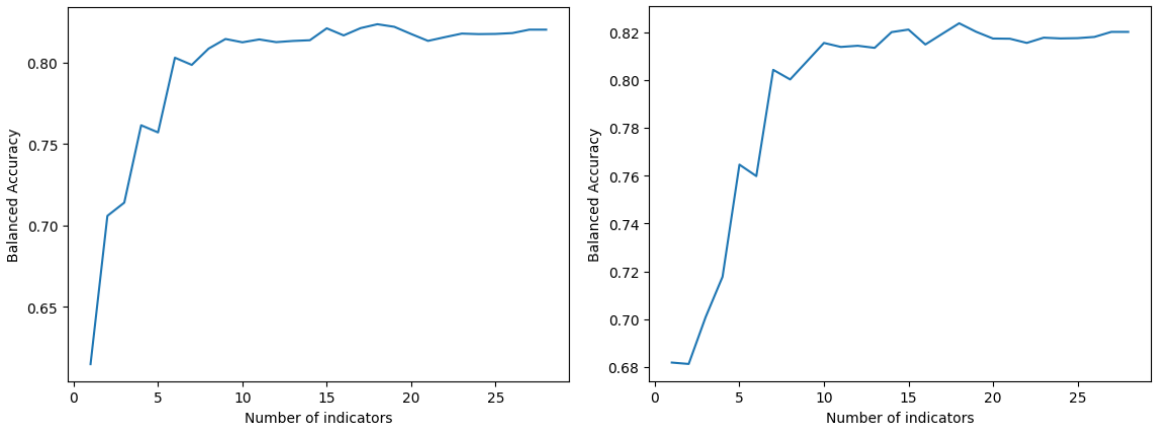


Figure 6: Marginal accuracy of all indicators

The vertical axis plots cross-validated balanced accuracy using a GLM, the horizontal axis shows how prediction performance evolves as indicators are combined. From right to left, indicators are dropped in order of significance, the first indicator thus being the most dominant. The predictors include all alerts and alarms with interaction terms for IRG/AA divide. Optimal number of indicators is identified when graph reached maximum balanced accuracy. Results are for Model (7) (left) and Model (8, final) (right).



Given the current geopolitical and economic context in Yemen, characterized by a division between areas controlled by IRG and AA, the study explored region-specific dynamics. Models (4) – (6) in Table 5 build upon the frameworks of Models (1) – (3), integrating region-specific interaction effects. Following this integration, RFE was employed to eliminate non-informative predictors and optimize predictions for out-of-sample data. The findings highlight a significant interaction effect across the IRG/AA divide. For the models utilizing alerts and alarms, these interactions center on the inflation and drought indicators. Critically, the extremely parsimonious alerts Model (5) suggests that food security in the IRG area is driven by inflation, while food security in the AA areas is driven by droughts. This outcome aligns the

existence of a dual currency system in the country when currency devaluation is rampant in IRG, and the AA areas are associated with increased agriculture and susceptibility to droughts. It is also worth noting that the alarms Model (6) tracks multiple dimensions of food security, and now outperforms the alerts model, as shown by the better loss values, suggesting that the use of alerts alone in Model (5) leads to an oversimplified representation of true food security drivers.

Table 5: GLM results with an IRG/AA interaction

Regression results for three models 4-6, extended using a regional interaction effect across the IRG/AA divide, and optimized using recursive feature elimination. To assess model performance, the Brier score and pseudo R-squared tests were calculated and a weighted average of error types was calculated, optimizing the probability cut-off used of classification.

	(4) GLM Continuous IRG/AA	(5) GLM Alerts IRG/AA	(6) GLM Alarms IRG/AA
Food price			-3.706***
			(0.781)
Fuel price	-6.659***		
	(1.266)		
Exchange rate	6.829***		
	(1.101)		
Displacement	7.608***		
	(1.097)		
Conflict	4.619***		1.195***
	(0.829)		(0.459)
Drought	-4.507***	3.782***	3.498***
	(1.163)	(0.271)	(0.404)
Food price with IRG/AA interaction	6.186***	1.971***	5.228***
	(0.968)	(0.133)	(0.836)
Fuel price with IRG/AA interaction	7.430***		1.751***
	(1.526)		(0.155)
Exchange rate with IRG/AA interaction	-6.434***	1.643***	1.012***
	(1.231	(0,145)	(0.135)
Displacement with IRG/AA interaction	-7.702***		
	(1.254)		
Conflict with IRG/AA interaction	-4.070***		-1.502**
	(0.953)		(0.587)
Drought with IRG/AA interaction	4.287***	-3.460***	-4.038***
	(1.385)	(0.317)	(0.546)
Pseudo R2	0.597	0.603	0.640
Brier	0.078	0.077	0.068
Loss w=1/2	0.271	0.277	0.253
Loss w=1/3	0.242	0.25	0.219
Loss w=2/3	0.299	0.304	0.286
Note			*p<0.1; **p<0.05; ***p<0.01

In our final comprehensive analysis, we combined both alerts and alarms and applied RFE across all predictors and interaction effects. Models (1) – (6) revealed that, while each indicator has a foundation in theory and is predictive on a univariate basis, optimal forecasting specifications typically result from discarding certain variables altogether. We, therefore, created a simple meta indicator which sums over the different alerts and alarms, to signal the number of warnings of any type. We applied a 4-month MA, to capture possible delays in impacts. Using this additional indicator, we apply RFE across all possible interactions. The results are in Table 6 and the RFE results for these models are in Figure 6. As depicted in Figure 6, the model's prediction performance improves significantly with the inclusion of the initial set of predictors and then flattens. Model (7) identifies the optimal combination of predictors, while Model 8 is designed to select the best predictors within the constraint that the meta-indicators for alerts and are maintained. This approach ensures that all dimensions are kept in the model and is justified by the theoretical relevance of each indicator to food security and their demonstrated univariate

significance. Retaining a relationship with all indicators in the model, despite causing a potential marginal decrease in historical prediction performance, guarantees the inclusion of essential food security aspects in future analyses, thereby enhancing the model's resilience to evolving food security trends. The pseudo R-squared (0.698), Brier score (0.057) and Balanced Accuracy (0.824) values of Model (8) finally remained identical at the triple digit level when compared to the purely statistically optimized Model (7), yielding a significant improvement over models (1) - (6). This observation underscores that incorporating additional dimensions of food security can be done without incurring a measurable performance trade-off, effectively capturing broader food security considerations without impact on overall model effectiveness.

Table 6: Comprehensive assessment and final GLM specification

Regression results for two models that nest models 5-6. Model 7 employs recursive feature elimination across all predictors, and model 8 only across interaction effects. To assess model performance, the Brier score and pseudo R-squared tests were calculated and a weighted average of error types was calculated, optimizing the probability cut-off used of classification.

	(7) GLM combined IRG/AA		(8) GLM combined IRG/AA keeping base indicators	
Predictors:	**Alerts**	**Alarms**	**Alerts**	**Alarms**
Meta indicator		2.369***	-0.634*	3.042***
		(0.207)	(0.342)	(0.441)
Food price		-4.538***		-4.401***
		(0.861)		(0.877)
Fuel price		0.785***		0.811***
		(0.213)		(0.214)
Exchange Rate	-1.331***		-0.971**	
	(0.378)		(0.434)	
Displacement				
Conflict	-2.713***	1.332**	-1.980***	
	(0.629)	(0.667)	(0.479)	
Drought	2.196***	2.006***	2.537***	1.921***
	(0.494)	(0.526)	(0.509)	(0.534)
Meta indicator with IRG/AA interaction	0.601***	-2.898***	1.299***	-3.648***
	(0.131)	(0.310)	(0.398)	(0.533)
Food price with IRG/AA interaction		6.786***		6.643***
		(0.913)		(0.928)
Fuel price with IRG/AA interaction	1.023***		0.972***	
	(0.226)		(0.226)	
Exchange rate with IRG/AA interaction	3.649***	-0.510**	3.219***	-0.542**
	(0.490)	(0.223)	(0.551)	(0.222)
Conflict rate with IRG/AA interaction	3.264***	-1.980**	2.456***	-0.478**
	(0.707)	(0.793)	(0.558)	(0.877)
Drought with IRG/AA interaction	-1.521***	-2.502***	-1.911***	-2.458***
	(0.582)	(0.685)	(0.596)	(0.690)
Pseudo R2	0.698		0.698	
Brier	0.057		0.057	
Loss w=1/2	0.176		0.176	
Loss w=1/3	0.167		0.168	
Loss w=2/3	0.186		0.184	
Note			*p<0.1; **p<0.05; ***p<0.01	

4.3 Estimates of populations in emergency

Using the GLM, it is possible to express the combined risk assessment in terms of the expected number of people in areas at risk of experiencing an IPC Phase 4+. For each district and time period, the GLM gives a probability between 0 and 1. Specifically, the total population per district can be multiplied by

the GLM-modeled probability to calculate the population-weighted average IPC Phase 4+ risk. This in turn can be scaled to match historical population counts in IPC Phase 4+ areas. To estimate the expected number of people in areas at risk of experiencing a deterioration into IPC Phase 4, model 8 was chosen.

The following notation clarifies the calculation. Let $\hat{P}_{pop} = \sum_{i=1}^{n} \hat{P}_i * pop_i \,/\, \sum_{i=1}^{n} pop_i$ be the population-weighted IPC Phase 4+ probability. Using the binary FEWS NET IPC Phase 4+ data, the share of the population in IPC Phase 4+ areas can be calculated $Y_{pop} = \sum_{i=1}^{n} Y * pop_i \,/\, \sum_{i=1}^{n} pop_i$. A linear rescaling is then required between and $\hat{P}_{pop}$ and Y_{pop} to remove aby bias stemming from different scales that results from optimizing for balanced accuracy, rather than for minimizing RMSE against the population totals directly. To rescale between the FEWS NET data and the GLM, a single scaling parameter is calculated using Least Squares. To provide some robustness to outliers or spikes in the modeled data, the scaling parameter is calculated using a centered moving average of the modeled probabilities.

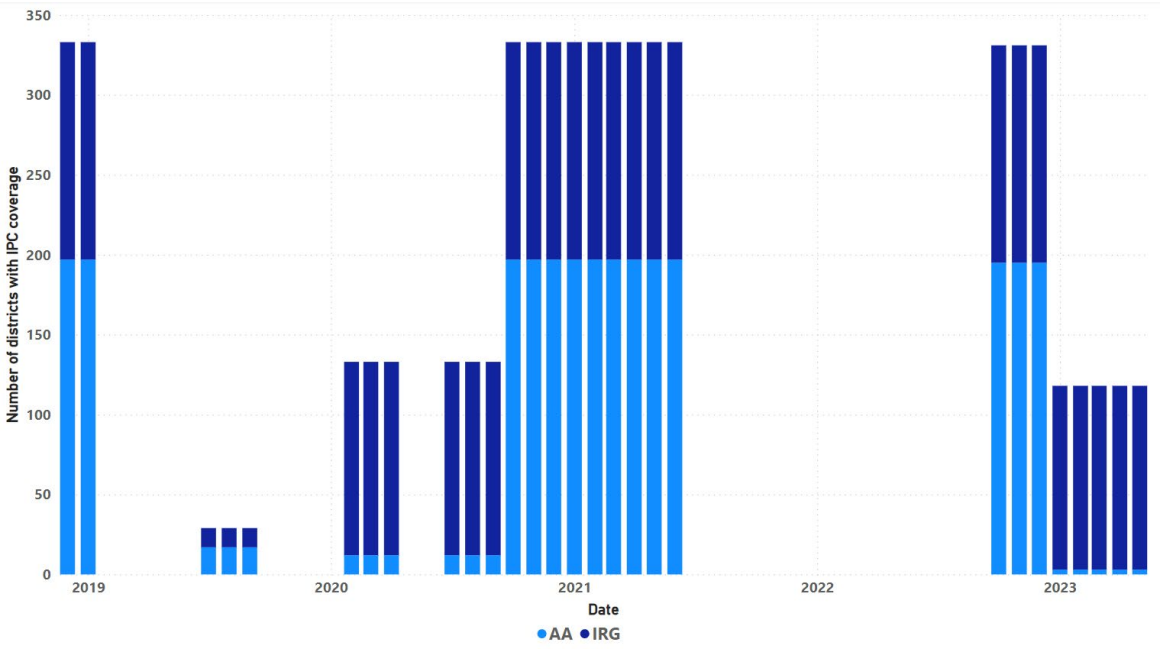
$$Y_{pop} = \beta \times L\big(CMA\big(\hat{P}_{pop}, 3\big), k\big) \tag{3}$$

In which $CMA(x, 3)$ denotes a 3-period centered moving average of x and $L(\cdot, k)$ is a k-period lag operator. The regression is fitted for $k = 1, \dots, 12$, and the optimal lag is selected by minimizing MAE. This $\hat{\beta}$ is then used to calculate the final population that are predicted to be in areas experiencing IPC Phase 4+.

$$pop|_{\widehat{IPC4_+}} = \hat{\beta}_{\arg\min k \in 1,\dots 12\ MAE\left(Y_{pop}, \hat{Y}_{pop}\right)} \times \hat{P}_{pop} \tag{4}$$

Figure 7: availability of IPC assessments

IPC assessments are contingent on sufficient evidence for high-confidence declarations, with no phases declared when information availability does not meet confidence criteria. The graph therefore highlights the likely access issues that may have also impacted FEWS NET assessments.



This final metric enables the monitoring of how alerts and alarms collectively correspond to the expected number of individuals in areas at risk of reaching IPC Phase 4+ conditions. To calculate this, we utilized Y_{pop}, incorporating data from both FEWS NET (starting from 2014) and IPC data (from 2019 onwards). The data integration was performed due to several challenges with the underlying assessment

data. Notably, in regions controlled by AA, we observed minimal variation in FEWS NET data across assessments post-COVID-19 pandemic. This is likely due to the assessment process, which involves projecting a most likely scenario and then during the next assessment cycle adjusting it based on new evidence. Reduced access and information scarcity have likely led to minimal adjustments, possibly causing the initial pre-pandemic assessment to carry over into later analyses. Concurrently, IPC data indicated considerably lower proportions of the population in IPC Phase 4+ areas since inception. It is important to note that IPC assessments are contingent on sufficient evidence for high-confidence declarations, with no phases declared when information availability does not meet confidence criteria. The sporadic availability of IPC data, particularly the absence of comprehensive coverage in AA areas highlighted in Figure 7, evidences the access issues that may have contributed to the persistence of initial IPC Phase 4 ratings in FEWS NET reports. Consequently, we noted that our modeled population estimates line up closely to FEWS NET figures in the pre-COVID-19 period, while lining up more closely to the IPC data in the period after.

Recognizing the potential biases and coverage issues that lead to discrepancies in both data sets, we applied Last Observation Carried Forward (LOCF) imputation to district-level IPC data. $Y_{pop}$was then calculated using FEWS NET data for the period before IPC data availability and a linear average of both datasets afterwards, as shown in the accompanying visuals. Y_{pop} thus represents an estimated percentage of populations in areas experiencing IPC Phase 4, giving equal weight to the last known IPC and last known FEWS NET assessments. We then determined the optimal scaling and lag parameters, finding the optimal lag value k to be 5. This suggests our final estimate can predict food security conditions most reliably five months in advance, indicating that alerts and alarms can preemptively signal deterioration into IPC Phase 4 conditions by at least five months.

Figure 8 presents the detailed comparison of IPC Phase 4+ figures from IPC and FEWS NET sources against the model's estimates of the population at risk of reaching IPC Phase 4+ conditions, based on data-driven alerts and alarms. Notably, the trends in the data-driven visuals precede the official declarations, aligning with the previously identified optimal lead-time of five months in IRG and four (but not significantly different from five) in AA. Specifically, in areas controlled by AA, the model's results closely align with FEWS NET data before 2020 and more closely follow IPC data thereafter although our indicators point to an overall lower severity than what is suggested by either FEWS Net or IPC data. It is important to note that for the entirety of 2022, there are no IPC assessments and a subsequent lower confidence in the FEWS NET phases. The modeled results in turn showed a sharp increase in estimates for the IRG areas in 2022 compared to 2020. This could similarly be attributed to overall uncertainties associated with the official IPC phase data not meeting confidence thresholds in both 2020 and 2022 as shown in Figure 7, making the relative increase captured by the model estimates less apparent in the official data. Finally, by 2023, the model estimates decreased, highlighting a period with notably fewer alerts and alarms, whereas the gap between FEWS NET and IPC data widened, particularly in IRG areas.

The difference between FEWS NET and IPC data highlights the importance of analyzing not only the ultimate population estimates from official or modeled sources but also monitoring the underlying alerts and alarms. This approach provides a thorough understanding of the factors driving food security. Overall, the findings indicate that combining indicators can reasonably express IPC phase exposure, although there are challenges related to the availability and potential biases in the underlying phase data. While the final population estimate is a valuable meta-indicator, the results emphasize the significance of evaluating the contributions of individual indicators to improve real-time food security monitoring and enable prompt interventions.

Figure 8: Estimated population in areas experiencing IPC Phase 4+

Comparison between official IPC Phase 4+ figures, FEWS NET and the calculated estimation of people at risk of experiencing a deterioration into IPC Phase 4+ in IRG and AA areas.