FAMINE EARLY WARNING SYSTEMS NETWORK TECHNOLOGY SUPPORT CONTRACT (FEWS NET TSC)

FUTURE DIRECTIONS IN PRICE DATA MAPPING FOR FOOD SECURITY ANALYSIS

October 4, 2012

This publication was produced for review by the U.S. Agency for International Development. It was prepared by Kimetrica LLC under contract number AID-OAA-C-11-00171

FAMINE EARLY WARNING SYSTEMS NETWORK TECHNOLOGY SUPPORT CONTRACT (FEWS NET TSC)

FUTURE DIRECTIONS IN PRICE DATA MAPPING FOR FOOD SECURITY ANALYSIS

CONTENTS

ABSTRACT

This paper explains the basic strategies necessary to ensure that consumer price data can be used for famine early warning and food security analysis. Currently, price data is not used in a systematic fashion for determining early warning phase classifications or triggering emergency needs assessments. This is a fault of the data. Price datasets are both too patchy in terms of commodity/time/space coverage and too large (too many observations, rarely synthesized in a way that is useful for decision-making). However, methods exist to create usable price datasets. This paper describes methods that can be used to create consistent datasets, synthesize the data, and present clear policy metrics.

1. INTRODUCTION

This paper/presentation sets out a road map for generating meaningful "heat maps" of market prices that is similar to the WRSI and NDVI products. We propose to build tools to analysing price data which will include on and offline mapping capabilities. Advances in GIS make it possible to generate continuous (raster) map surfaces from point data.

Market food prices are a key determinant of food insecurity, but price data are not systematically considered in analyses. Stronger analysis would benefit policy-making both in terms of identification and quantification of the scale of the food security problem (i.e. spotting those areas where needs assessments need to be focused) and identifying the best responses.

A key reason for inadequate price data analysis is patchy data. It is proposed to develop tools that are applicable in data scarce FEWSNET countries, using a lowest common denominator i.e., data that are available in all FEWSNET countries. This will ensure that the methods and reporting are comparable across countries.

The paper/presentation proposes a pathway to produce meaningful map products for price analysis. It covers the methods for handling patchy data (interpolation), data summary methods, and analysis and drawing policy implications.

2. METHODS FOR PRICE DATA INTERPOLATION

2.1 WHAT IS INTERPOLATION?

Interpolation describes a toolkit of methods for "filling in" data where raw data do not exist. There are numerous interpolation techniques and models. Many have been developed for geo-spatial analysis and have not been applied to price analysis.

Interpolation can be spatial (filling geographic gaps) or temporal (filling gaps in a time series) or some combination of the two. Price data interpolation is generally temporal. In this paper we will look at spatial methods too.

2.2 WHY IS TEMPORAL INTERPOLATION NEEDED?

Interpolation is necessary in all FEWSNET countries because of missing and incomplete data. Comparison of prices over time and space require consistently defined data. There are several reasons

why, in the conventional "market information system" model – i.e. agent model, consistent time series are lacking.

- Problem of low transactional volumes (too few valid observations when small markets are quiet),
- Low compliance with data collection protocols by agricultural extensive officers, as a result of perverse or inadequate incentives
- Inconsistent investment in market information
- Data loss through poor coding (e.g. market/commodity definitions), identification, data management

SMS based price reporting will make things worse! The problems are likely to be compounded by the introduction of transactional-based information systems which will be based on data from SMS, online trades, or local exchanges, rather than agents collecting data from specific market points. The primary problem for the spatial analysis of this data is that the points of transaction keep changing.

Without interpolation, price data can't be compared with other indicators. Price data are point data i.e. defined at a market point or point of transaction. How can this be compared with raster or vector data that are commonly used for food security analysis? The differences in data make it difficult to use price data explicitly in composite indices or to inform consistent spatial or temporal phase classifications (such as the IPC or early warning stages). It would generally be wrong to assume that just because there is a market with a livelihood zone (LZ), this reflects market conditions within that particular LZ. Of course reporting markets are typically located in urban and peri-urban areas.

Interpolation allows visualization of price data in map form. It is hard to get a picture of the spatial distribution of market prices from a few scattered market points. Interpolation will allow the possibility of building maps that show market "hotspots" where prices are anomalous for some reason.

2.3 TEMPORAL INTERPOLATION METHODS

- Temporal interpolation for a given point: filling in the gaps (trend, seasonality)
  - Using autoregressive / lag models
  - General consumer price deflation, application of more specific CPIs (e.g. rural/urban, higher frequency)
  - Explicit estimation of seasonal effects using NDVI (i.e. association of intra annual variability of de-trended data with explicit determinant of seasonality)
  - Using NDVI based measures to predict price variability (Brown, Pinzon and Price, 2008)

2.4 TESTING INTERPOLATION METHODS:

- Regression modeling

3. INDEXING

3.1 INTRODUCTION

- **For price data to be useful for policy-making, it now needs to be summarized.** Single commodity data is hard to map and interpret.
- **The key method for summarizing price data is in indices.** The index aggregates price data from multiple commodities.
- **How the indexing is done depends on what story you want to tell**. For FEWSNET, the main concern is how food prices will affect consumers, so the relevant index takes consumer prices (retail) of food and summarizes them.
- **Need to interpolate first and then calculate the indices.** It is hard to build indices with missing data for specific commodities. For most price data sets, if you rely on raw price data you cannot build an index because there are too few months for which data are available on all commodities.
- **Work on mapping price indices rather than on mapping specific commodity prices.** However, in some countries, one commodity has such a major significance for consumption (e.g. maize in Kenya, Tanzania and southern Africa), that it will be useful to generate maps for that single commodity.

3.2 WHAT ARE PRICE INDICES?

- **Price indices summarise price data from multiple commodities into a single measure or index.** The basic idea is to make the information more amenable to analysis. It is one way of dealing with large data.
- **Consumer price indices summarise retail prices** for multiple consumer goods, based on consumption or expenditure shares: the more a commodity is consumed, the more influence the price has on the index.
- **Producer price indices summarise the producer output prices** which for FEWSNET purposes mean the farm-gate prices for food produce. Each output price is weighted by its share in production or farm revenue.
- **Food price indices are consumer price indices for a subset of food commodities.** These are probably the most relevant to FEWSNET. They should, in general, reflect the prices of the main staples.

3.3 WHY ARE INDICES IMPORTANT?

- **Synthesis of data:** It is hard to gain a clear picture from multiple price time series, especially when commodity prices are not all moving in the same direction. Since price data is collected over time and multiple commodities, it can be hard to gain an overall picture.
- **Local consumption habits vary**: One crop may be an important staple in one area but a secondary famine food in another (e.g. maize in parts of Ethiopia). It is possible to misinterpret the significance of a single commodity price change in a given context if prices are not indexed or if the indexing is inaccurate.

- **Significance of market purchases varies**: High prices may actually be beneficial for food security in areas that depend on food aid as it increases the resale value of the food aid. Likewise, In surplus and semi-subsistence areas, high prices are beneficial.

3.4 KEY INDEXING METHODS

- **Methods for consumer price indexing are based on expenditure shares**. They are used in all countries to estimate general price inflation. Methods vary in levels of sophistication depending on how much expenditure data is available.
- **Early warning requires disaggregated index methods.** National CPIs are usually based on aggregate national rural and urban expenditure shares. Where there is high variation in food consumption patterns between areas, national weights may be inaccurate.
- **It is possible to develop more refined and locally relevant indices.** National income and expenditure data can be processed (easily) to calculate local (generally admin level 3) expenditure shares
- **For EW purposes, disaggregation by wealth/poverty group may be desirable.** Since we are particularly interested in how price changes affect the most vulnerable, there may be an argument for calculating indices specifically for poorer groups.
- **It is possible to calculate indices for specific wealth groups**. In most FEWSNET countries, household income and expenditure data is sufficient to all the calculation of expenditure shares for specific wealth groups. Some livelihood zone or Household Economy datasets also include commodity breakdowns for specific wealth groups.

4. SPATIAL INTERPOLATION

4.1 INTRODUCTION

We can use a range of spatial interpolation techniques to create a raster surface.

4.2 INTERPOLATION METHODS

-Linear interpolation based on the linear (Euclidian) distance to the nearest markets

-Kriging

-Co-integration models that estimate the relationship between neighbouring markets over time.

As with the temporal interpolation it is also possible to use non-price data to increase accuracy. The basic idea is for spatial price differentials to be associated with known / measurable local and temporal characteristics for which data exist. These variables may be selected according to how they influence supply or demand in a given time and location.

The equation below allows for prices in a given location to be influenced by inertia (lagged effects), conditions in neighbouring markets (integration effects) and location supply and demand conditions

$P_{ijt} = f(p_{ijt-1}, p_{ijt}, q^{S}_{ijt}, q^{D}_{ijt}, e_{ijt})$

- Supply side determinations
- Cost surface modelling
- Demand

-Population density

-Broad market characterization

-Food aid as a share of consumption

-Poverty rates

4.3 TESTING METHODS

Methods exist for estimating the error of spatial interpolations. These include jack-knife techniques that compare predicted values and actual values for data points. However, in some countries there are too few data points to be able to use jack-knifing for accurate estimates.

Ground-truth exercises would be useful. Data can be collected on a once off basis from small markets (between main markets) and regression models run to test the predictive power of different estimation approaches.

The method can improve over time. The software can include intelligent algorithms that continuously test approaches and refine the estimates to minimise estimation error.

4.4. SOFTWARE ISSUES

There are considerable software challenges. A key problem will be the volume of data. Creating raster imagery on a monthly basis will require very large processing capacity. It might be possible to start at low resolutions and then build up processing capacity over time.

On-the-fly processing is unlikely. It will probably be necessary to pre-process and tile the monthly images and then deliver, rather than attempting to process images for the user's selected geographic extent and time period as processing times and data volumes will be high.

5. SPOTTING AND ANALYZING ANOMALIES AND HOTSPOTS

5.1 INTRODUCTION

- **Heat maps and area-based tabulations will be the main output.** For quick and simple analytics, we propose to generate heat maps which would be in a raster format. GIS tools would allow extraction of data by geographic area (administrative unit, livelihood).
- **Heat maps would represent anomalies or outliers** to provide a clear indication of where things might be going wrong, i.e. locations where prices appear to be unusually high or low.
- **Current prices are compared with adjusted averages over time and space**. The basic technique is to compare price indices over time (time series comparison of current values with historic values for that location) and over space: current price values in a location compared to average current values.
- **Additional information may be required to ensure that the outliers are correctly identified**. Techniques are needed to understand whether spatial and temporal variation in prices genuinely reflect short term anomalies or are just caused by underlying trend, seasonality or local characteristics (such as low demand or poor road quality).
- **The tool** should provide a basic explanation of why prices are anomalous in order to identify the policy implication.

5.2 MAPPING AND REPORTING

- **Difference image of current prices compared to seasonal norm**: normalised difference between the inflation-adjusted current value of the index and the long-term average after allowing for seasonality probably using a z-score.
- **Double difference mapping**: comparison of current seasonal variation with "normal" price rise for a given location and month, i.e. increase in prices between August and September 2012 with the "normal" seasonal increase for this time of year.
- **Difference image of current price compared to spatial mean for the current period**. How does the current price in a given location compare with the mean price for the area as a whole? This is useful as a snapshot of where prices are particularly high. Note however that some markets have habitually higher prices than others.
- **Difference with IPP image**: current local market price compared to Import Parity Price (IPP)
- **Double difference image**: current price difference (from spatial mean) compared to historic difference for that pixel/market. This captures the markets that are currently anomalous, i.e. those where current price differentials are abnormal.
- **Mark up estimation and mapping comparing current price differentials with efficiency price:** it should be possible to identify locations where current spatial price differentials are significantly different from the efficiency price - i.e. the price that would prevail if food were transported to the location from the nearest supply source at competitive transport rates. This method would use cost surface models. There is generally enough data to run such models in the FEWSNET countries.

5.3 ANALYSIS

- **Econometric techniques can be used to explain what is driving the temporal and spatial price anomalies** (for the surface as a whole, not for a specific point). A regression model can be used to assess how far the variation is explained by known domestic and international determinants of local prices including:
  - Current storage rates – i.e. rental rates on storage space
  - Current and recent world food prices and IPP
  - Current and recent oil prices
  - National supply and demand from the Food Balance Sheets
  - Performance of most recent season using NDVI-based proxies compared to historic norm
  - Transport rates and distances to supply sources
- **The regression model uses explanatory variables that have not been used in the interpolation.** Note that it is important not to use the same variables that are used in the interpolation function otherwise the model will return false correlations (because the explanatory variables are correlated with the prices by defined, not because of underlying causality)
- **The methods should measure unexplained variation or "residual"**: how much variation is not explained by the model? The unexplained error could actually be mapped. It may be explained by market inefficiencies.

5.4 WHAT ARE THE POLICY IMPLICATIONS?

- **Emergency needs assessment.** The key objective is to assess the threat of famine or food insecurity for people in a given geographic location. Price analysis alone cannot determine how many people might require assistance (only household or individual level data can determine this with any reliability). By mapping price anomalies, we are able to indicate areas where in-depth field assessments should be focused.
- **Cash versus kind.** Market assessment of this kind is crucial for identifying the optimal response to food insecurity and specifically whether programmes should focus on cash, in-kind or some combination. In particular, the tools can be used to determine where using cash would be high risk (anomalous high prices with the risk of further price inflation to the detriment of consumers) or where using food would be high risk (anomalous low prices with the risk of deflating food prices to the detriment of farmers).
- **International trade and food aid**: large unexplained discrepancies between domestic and international prices may provide a justification for interventions to improve the efficiency of external food trade (e.g. reduction in tariff and non-tariff barriers) and/or consider international food aid. The analysis could support a modern version of the "Bellmon Assessment" to look at the possible effects of subsidized food imports on domestic markets.
- **Stabilisation policies and stocks:** unexplained local price hikes may reflect inefficiencies or low capacity in the local storage and wholesale markets. This may provide a justification for investment in the storage infrastructure.

6. HOW TO GET THERE

The approach needs to be practical and replicable. It is important to recognize that FEWS NET is starting from a weak base. While it is tempting to begin with a pilot activity, it is more important to start with a model that works at scale. A stepwise development process is recommended, starting with simple functions with low data requirements. We suggest a series of deliverables, each independently useful and in growing complexity.

1. Phase I: Get the raw data in order
    a. Develop market point mapping tool with consistent market codes and geography
    b. Create and apply commodity code library
    c. Ensure consistent time definition
    d. Move raw data to consistent format
    e. Develop point and shoot raw data retrieval /query tool to allow data extraction and dumping for a given market
2. Phase II: Build interpolation capability
    a. Develop simple temporal interpolation algorithms
    b. Experiment with increasing accuracy of temporal interpolation
    c. Develop online graphics for interpolated series (D)
3. Phase III: Development of indexing

   a. Develop indexing algorithms
   b. Development of sub-national weights datasets
   c. Development of IPP indexes
   d. Implementation of monthly food price indices
   e. Production of online food price index (D)
4. Phase IV: Spatial Interpolation of indices
   a. Development of simple interpolation algorithms
   b. Development of data structures
   c. Development of mapping tool
   d. Testing complex algorithms
   e. Generation of monthly FPI maps (D)
5. Development of basic heat maps
   a. Development of differencing algorithms
   b. Processing of FPI time series
   c. Development of low resolution mapping functionality
   d. Production of monthly heat maps (D)
6. Development of analytical products
   a. Design and testing of econometric models
   b. Data acquisition for explanatory variables
   c. Integration of datasets of explanatory variables
   d. Development of API/backend for automated integration
   e. Implementation of tool for analysis of anomalies (D)