HACK4GOOD: INTERPRETABLE INDICATORS

Peshal Agarwal, Jelena Čuklina, Andrei Kolar, Anna Maddux

November 30th, 2019

Abstract: *Using the commodity price data collected by IMPACT Initiatives' Market Monitoring exercise in Syria, we developed higher-level indicators that characterize price changes over time. The framework we present here is both flexible in which facet of the data to analyze, in addition to being generalizable to similar data sets of other crisis-affected countries. The output is kept in a form that is easy to interpret so that it can clearly inform humanitarian action. To demonstrate its usefulness, we show examples of insights that can be uncovered using the framework. Finally, we highlight the limitations of the current product and offer suggestions for continued improvements going forward.*

1. Introduction

Cash-based assistance is a recent reform in the humanitarian sector. Even given its relative novelty, established humanitarian actors have by now accepted the benefits of adding cash transfers to their existing humanitarian programming. However, there is still much work to be done to determine the most effective methods for implementing such programs. In general, humanitarian actors lack quantitative tools that can improve the decision-making processes surrounding cash transfers.

IMPACT Initiatives is an example of one of the humanitarian actors looking to develop such tools. Currently, they collect price data on an assortment of commodities in crisis-affected countries. The goal of this project was to help IMPACT harness the potential of these data sets by summarizing the data into a set of interpretable indicators. In particular, the framework developed seeks to characterize price changes as either stable, volatile, or a trend.

2. Methodology

The framework was developed using the data collected by IMPACT's Market Monitoring exercise in Syria. The Syria Market Monitoring team collects price data on commodities covering 140 markets across 10 governorates. The data has been recorded monthly, starting in November 2017. These commodity prices are used to calculate the Survival Minimum Expenditure Basket (SMEB), which is the 18 key consumable (food and non-food) commodities that an average six-person Syrian household needs to purchase for survival each month

We used the derivative of the price changes (the difference in price from one month to the next) as the base for building our framework. We decided to focus on the derivative instead of a scaled measured (i.e., the percentage price change from month-to-month) as it provides important information about the magnitude of the prices. An equal percentage price increase in 500g of salt and 1L of petrol would not imply setting aside the same amount of cash. Therefore, any framework should reflect these differences in magnitudes.

The user sets numerous inputs to determine which data is analyzed. The user specifies the geographical level (region, governorate, district, subdistrict, or town), the commodity type (a SMEB value or a specific commodity), the time window (in months), and the final month (the last month of the time window to be taken). The data is filtered based on these criteria and the derivatives are calculated by taking the difference in price between two consecutive months. Volatility is characterized by calculating the standard deviation of the derivative, and the

direction of price changes by taking the mean of the derivative.

To illustrate these characterizations, one of the outputs of the framework is a chart with the scaled standard deviation on the vertical axis and the scaled mean on the horizontal axis (see Figure 1). Scaling is done by dividing by the mean price in that time window so that all areas (i.e., districts, subdistricts) can be compared. The higher up on the chart a certain data point, the more volatile the price changes of the corresponding area. Similarly, a data point close to zero on the horizontal axis implies that prices were more stable.

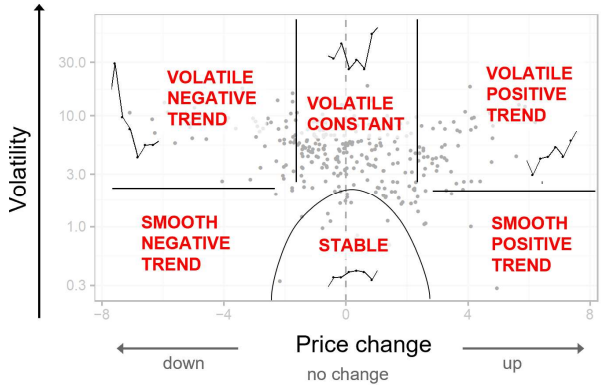

FIGURE 1: Example of model output (arrows and labels included for illustration)

The framework adheres to the goal of being generalizable by not defining the boundaries that separate the various indicators. The chart described provides the user with a country-wide overview of the direction of price changes and the volatility of those changes, without making any assumptions regarding the setting of boundaries. However, the user has the option to manually set the level of these boundaries. In that case, the framework will also output a table categorizing each price as either stable, volatile constant, a stable trend, or a volatile trend.

In addition to the trend analysis, our team investigated whether significant correlations exist between different SMEB values. The goal was to determine whether certain commodities tend to move together, or whether they exhibit negative correlation (which could be evidence of substitute goods). We examined the correlations between all pairs of items for each of the 22 months in the data set. Our analysis revealed that there are certain products whose prices tend to correlate significantly. This information can be particularly useful in situations where data for one of the items is missing and we would still like to draw some inference (say, volatility) of that item.

It is important to note that in our correlation analysis we aggregated the commodity prices by month and took their mean value, disregarding their location. A next step would be to investigate whether the observed correlations exist across all levels (governorate, districts, subdistricts), or whether they break in certain areas.

3. Results

Using the framework described, the user can vary the inputs and explore what insights can be uncovered. An example is shown in Figure 2. The chart is generated by running the framework using a time window of 6 months and analyzing the SMEB values at the district level. This is repeated for numerous endpoints. This shows that there are temporal and geographical relationships in the prices. For example, by colouring the points by month we observe that in some months a negative volatile trend exists in most areas, whereas in other months prices stabilize across most areas. This demonstrates that price changes are not completely random, and that these temporal and geographical relations can be used in improving price estimation.

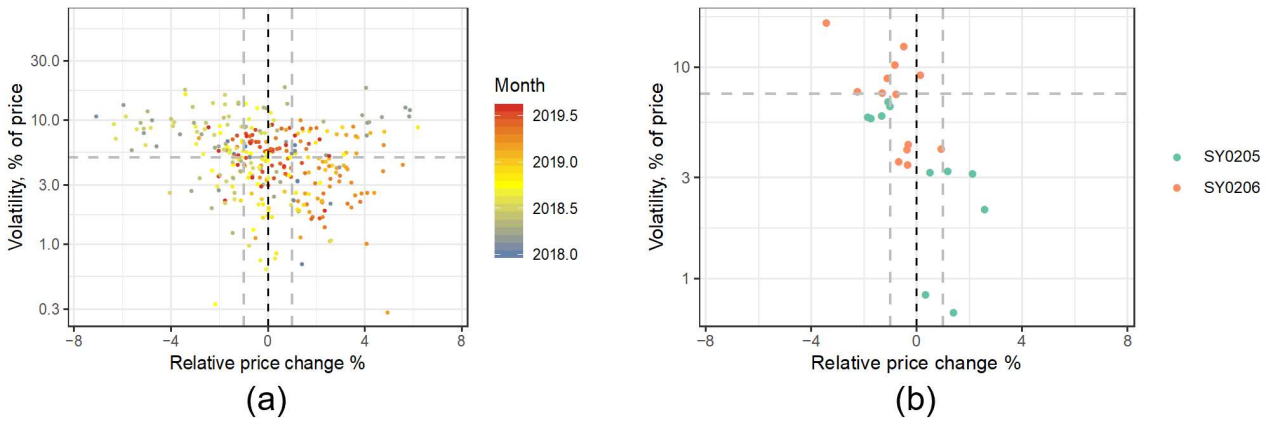

FIGURE 2: Example insight showing that volatility and trend are both months (a) and location (b) specific

4. Limitations

This framework can be limited in its analysis of excessively incomplete data sets. To calculate the volatility and trend, one needs at least 3 data points within the specified time window. This can become problematic when analyzing more granular facets of the data set, such as short time windows (e.g. 3 months) and lower geographical levels (e.g. subdistricts or towns).

5. Discussion

The generalizability of the framework developed has numerous immediate benefits to IMPACT. First, it is flexible in its analysis of a particular data set by allowing the user to choose what part of the data to zoom in to. A user of the framework can specify the commodity, geographical level, and time window to be analyzed - which is especially useful to users whose information of interest changes on a case-by-case basis. This structure ensures that no valuable insights are hidden, as they might be in a model with a predetermined output.

Secondly, by not including any Syria-specific assumptions, the framework can be readily applied to any crisis-affected population. This can facilitate consistency in the decision-making process for cash transfer programs in any country.

Finally, by focusing on a solution that is relatively simple and easy to understand, our framework ensures that all stakeholders can understand recommendations provided based on these indicators. As the goal of an interpretable indicator is to inform humanitarian action, it is important that any model developed is not a black box. Otherwise, not only does the model complicate the discussion around a certain decision, but it increases the likelihood of a wrong decision being made.

Nevertheless, there are numerous steps that can be taken to continue improving this tool. The nature of a country amid a crisis is that the situation can change at any moment. Therefore, if one of the goals of an interpretable indicator is to offer information on how a certain price is expected to behave in the future, characterizing historical price changes alone is insufficient.

6. Outlook

Adding labels

To define specific labels (such as stable, volatile constant, increasing trend), the user can set the thresholds on volatility and price change values. The intuition illustrated in Figure 1 could be used as a guide to the setting of the boundaries. These labels can simplify the data representation and be included in IMPACT reports.

Deeper analysis of framework results

Preliminary analysis indicated that there is a relationship at the month level between the

volatility and price change indicators. This is likely at least partially due to sliding window effect. Therefore, it must be tested with appropriate time-series analysis tools, such as autocorrelation analysis.

Furthermore, analysis was not done on lower price levels than the SMEB – for example, at the commodity level or even at the level of SMEB of certain groups (say, SMEB of non-food items). Repeating the described analysis at these levels is likely to yield interesting insights.

We would have additionally liked to explore the stability of the price ratios between the various commodities that constitute the SMEB. If such ratios are generally stable, we would have focused our analysis on the volatility of commodities that make up a larger portion of the SMEB. This would have ensured that our focus is not on items that have a relatively small contribution to the total SMEB value, and therefore are of less importance when informing future expectation of prices. This idea could likewise be used for a better assessment of the total SMEB if the price of the cheaper items is missing – these prices could be estimated using the ratio from previous periods or neighbouring geographical areas.

Integration with external factors team

This framework can be integrated with the solutions of the other teams involved in this Hack4Good. One of the teams explored connections between price variations and external factors. The output of our tool can be joined with an external factors event table (describing war activities and people relocation) to explore the relationships between these events and the dynamics of the prices. Insights found here could better inform the expectation of future price behaviour, compared to an expectation derived solely using historical data.

Integration with missing value imputation and prediction teams

Two teams sought to add values when SMEB value was not there: one of the teams sought to estimate missing values and another team worked to create a predictive model to understand future price trends. As a natural extension of our framework is using the characterization of historical price changes to inform expectations of future prices, we would expect consistency between their solution and the framework we developed. Thus, the next step could be to use all our solutions as complements to one another.

Other indicators

SMEB data collection in crisis areas is plagued by missing values. The pattern of missingness could be an indicator in and of itself - missing values most likely occur either in acute crisis areas or in relatively stable areas where cash transfers are no longer needed.

7. Conclusion

We see the framework that we have developed as one that immediately provides any humanitarian actors with insights that can inform the type and amount of aid to provide. While it is far from a finished product, by maintaining generalizability we have created a product that can continue to grow as cash-based programming continues to grow around the world.

A quantitative tool that can transform previously uncollected data into recommendations can innovate how humanitarian actors implement their cash transfer programs. An indicator that is built correctly can reach the same conclusion – and perhaps a more accurate one – that an organization would otherwise take more time to reach. This is crucial in a field where aiding crisis-affected populations is extremely time-sensitive. We hope that our work can be one tiny step towards increasing the amount of help that can be provided to people in need all around the world.