Humanitas

A prediction tool for volatile commodity prices in developing countries

Students: Alexander John Busser, Anton Ovchinnikov, Ching-Chia Wang, Duy Nguyen, Fabian Brix, Gabriel Grill, Julien Graisse, Joseph Boyd, Stefan Mihaila

Time Series data

Different sources Price categories: Retail prices, wholesale prices

Wholesale prices

Wholesale price index

(taken from investopedia.com)
An index that measures and tracks the changes in price of goods in the stages before the retail level. Wholesale price indexes (WPIs) report monthly to show the average price changes of goods sold in bulk, and they are a group of the indicators that follow growth in the economy.

Although some countries still use the WPIs as a measure of inflation, many countries, including the United States, use the producer price index (PPI) instead.

Price sequences

Other sources

distribution & production

exchange rate crude oil

Social Media data

Twitter

Historical tweets

Approach 1: Fetching "historical" tweets via user graph

able to fetch 4-5 million users a day

Approach 2: Filtering tweets provided by webarchive.org

https://archive.org/details/twitterstream

Daily? tweet aggregator

Issue of localization

Geolocalized tweets

Approximation: Mapping tweets to user location

Processing

Merging Series

Crafting indicators from tweets

Price Transmission Analysis

Interpretation

automate interpretation to a certain extent by learning about circumstances through online data.

Time Series Analysis

Time series data has a natural temporal relation between different data points. It is important in the analysis to extract significant temporal statistics out of data. We will focus on analyze stationarity, autocorrelation, trend, volatility change, and seasonality of our price datasets in R.

Stationarity of a series guarantees that the mean and variance of the data do not change over time. This is crucial for a meaningful analysis, since if the data is not stationary, we can not be sure that anything we derive from the present will be consistent in the future. We can transform our data into a stationary one by taking k-th difference to remove the underlying trend, and then apply standard test procedures such as KPSS test [1] to see if the differenced series is stationary.

Autocorrelation is another important trait in time series data. It suggests the degree of correlation between different time periods. By plotting correlograms (autocorrelation plots) of our data, we will be able to identify if the fluctuation of prices may be due to white noise or other hidden structures.

Seasonality is reasonably expected in our agricultural related time series. Several methods might help us to detect seasonality, such as common run charts, seasonal subseries plots, periodograms, and the correologrаms we mentioned before.

(trend and volatility change is straightforward and can be concluded once we have the datasets)

[1] Kwiatkowski, D.; Phillips, P. C. B.; Schmidt, P.; Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root". Journal of Econometrics 54 (1âĂŞ3): 159âĂŞ178.

Prediction Models

Time Series Forecasting

ARMA Model

The classical Time series forecasting approach is to use the ARMA (Auto-Regressive Moving Average) model to predict the target variable as a linear function which consists of the auto-regressive part (lag variables) and the moving average part (effects from recent random shocks).

The ARMA(p,q) model: (will refine math representations later)

$Phi(B) * Y_t = Theta(B) * eps_t$

The fitting of the model and the historical data can be accomplished by maximum likelihood estimation.

Regression

We can also apply ARMA to the linear regression model. It is formulated as such:

$Y = Beta * X + eps, eps\ ARMA(p, q)$

Through OLS (Ordinary Leasat Square) or GLS (General Least Square) processes, we can obtain an optimal Beta.

Echo State Networks (ESN)

Echo State Networks are a type of Recurrent Neural Network (RNN) applicable to many domains because unlike other RNNs they are easy to train.

[?] The third section explains how echo state networks can be trained in a supervised way. The natural approach here is to adapt only the weights of network-to-output connections. Essentially, this trains readout functions which transform the echo state into the desired output signal. Technically, this amounts to a linear regression task.

Echo States

For our task we need a discrete time neural network which is incidentally also the constraint in which Echo State Networks are defined. The ESN is assumed to have N input units, K internal network units and L output units. Direct connections from input to output units and from output to output units are allowed.

Training ESN