Food and Agriculture Organization of the United Nations

FAO Statistics Division

Working Paper Series

ESS / 14-07

NOWCASTING REGIONAL CONSUMER FOOD INFLATION

September 2014

NOWCASTING REGIONAL CONSUMER FOOD INFLATION

Franck Cachia

Food and Agriculture Organization of the United Nations
Rome, 2014

Now-casting Regional Consumer Food Inflation

Franck Cachia
Associate Statistician, Statistics Division, FAO

Abstract

Consumer price indices (CPI) are disseminated by countries with a lag that typically varies from 1 to 4 months. Global CPI datasets, such as those maintained by the International Labour Organization (ILO), the United Nations' Statistics Division (UNSD) or the International Monetary Fund (IMF), have a longer average lag because of the time needed to collect, compile and publish the data provided by countries. In order to monitor current trends in food inflation, forecasting (or nowcasting) price changes to the current period is therefore necessary. This paper presents the methodological framework used by FAO's Statistics Division to now-cast consumer food inflation at regional level. Hybrid ARIMA-GARCH models are estimated for each region, with additional explanatory variables constructed from a large and high-frequency dataset. The out-of-sample analysis indicates a satisfactory performance of the models at predicting the overall variability in prices as well as the sign and direction of price changes.

Key words: Nowcasting; Regional food consumer prices; ARIMA-GARCH models

JEL codes: C53, Q11

1. Introduction

Real-time data is required for policy makers to anticipate or react in a timely manner to possible tensions on retail food markets. One of the only sources of near real-time information on food prices are price quotations of major agricultural commodities traded on international spot and futures markets. These price quotations are summarized in indices such as the FAO Food Price Indexes (FPIs)[1] or other commodity price indices produced by international organizations such as the World Bank or the International Monetary Fund.

These indices are a useful source of information to monitor current trends in food inflation. However, relying exclusively on them is both insufficient and, in certain circumstances, flawed. First, while a certain degree of transmission exists between price signals on international agricultural commodity markets and retail food markets, the pass-through from one to the other is incomplete, lagged and highly variable across regions (Cachia, 2014). Price trends in certain regions might even be completely decorrelated from international markets and depend only on internal drivers. For example, prices at country or local level may be affected by the sudden release of massive public food stocks, leading to a fall in local prices, while leaving international prices unchanged because the country or region is neither a major exporter nor importer of the commodity released. An absence of or a very low transmission may also reflect an economy which is structurally isolated from international price shocks because of buffer mechanisms provided by governments.

Up-to-date information on food prices at consumer-level is therefore necessary in order to monitor real-time developments of food security in countries and regions. Since August 2013, FAO's Statistics Division is compiling and disseminating estimates of consumer food inflation for different regions of the world and at the global level[2]. They complete the country Consumer Price Indices (CPIs), also published on FAOSTAT, based on data from the International Labour Organization (ILO).

The publication lag at country-level and the additional time needed by international organizations such as the ILO or the United Nations' Statistics Division (UNSD) to compile and harmonize country data inevitably reduces the timeliness. Currently, given these constraints, regional and global estimates are disseminated on FAOSTAT with a lag of 3 months. For example, for the data release of July 2014, CPI indices were published up to April 2014. This working paper presents a possible approach to estimate these 3 months of lacking information using an econometrically sound and flexible methodology.

The remaining sections of this paper are organized as follows: the second section presents the econometric approach used to construct the regional forecasting models and defines the

[1] Details at http://www.fao.org/worldfoodsituation/foodpricesindex/en/
[2] The analysis and underlying data are available at: www.fao.org/economic/ess/ess-economic/cpi/en/

statistics employed to test their performance out-of sample; section 3 presents the data and explanatory variables used; An illustration for one region, North Africa, is provided in section 4. The final section concludes and discusses the possible future improvements of the approach. Annexes provide additional details on the data and results.

2. Forecasting strategy

a. Econometric modeling

Main model

Monthly changes in food prices for each of the sub-region[3], measured by the corresponding CPIs, are predicted using linear regressions with ARMA/GARCH disturbances (also referred to as hybrid ARIMA-GARCH models) [4]. The equations are given below.

Let P_t be the food CPI for a given region measured in t, P_t^* a measure of international agricultural commodity prices, such as the FPIs, $\boldsymbol{X}_t$ a set of other explanatory variables (exchange rates, economic activity data, etc.) assumed to be exogenous and ε_t an independently and identically distributed random error term. Variables in low-cases represent natural logarithms, and growth rates or first log-differences when dotted. Vectors are in bold. The regression equation is:

$$[Reg]\colon\ \dot{p}_t = \mathrm{a} + \sum_{\mathrm{i=1}}^{\mathrm{p}} \varphi_{\mathrm{i}}\, \dot{p}_{t-i} + \sum_{\mathrm{j=0}}^{\mathrm{k}} \beta_{\mathrm{j}}\, \dot{p}^*_{t-j} + \sum_{\mathrm{l=0}}^{\mathrm{m}} \boldsymbol{\gamma}_{\mathbf{l}} \dot{\boldsymbol{x}}_{t-l} + \ \varepsilon_t$$

The presence of autocorrelation in the residuals and of "volatility clustering" of the residuals, when large changes tend to follow large changes and small changes follow small changes, is a distinctive feature of commodity prices in general and food prices in particular, even for highly aggregated indices such as food CPIs. This was well evidenced in the food price crisis of 2008-2009, with several episodes of price spikes followed by a period of easing.

To accommodate for residual autocorrelation and volatility clustering, $[Reg]$ can be estimated using a procedure that allows for the residuals to follow an ARMA-GARCH process. The ARMA component represents the autocorrelation structure of the residuals, while the GARCH process reproduces the structure of this autocorrelation in unexpected shocks. The resulting model is:

[3] FAO's Food Consumer Price Indices are available at country, sub-regional (e.g. South-Eastern Asia), regional (Asia) and global. Annex 3 provides the country composition of the different sub-regions.

[4] AR(I)MA stands for Auto-Regressive (Integrated) Moving Average and GARCH for Generalized AutoRegressive Conditional Heteroscedasticity.

$$[M]: \begin{cases} [Reg] \\ [ARMA]:\ \varepsilon_t = \mathrm{b} + \sum_{\mathrm{i}=1}^{\mathrm{P}'} \mu_{\mathrm{i}}\, \varepsilon_{t-i} + u_t + \sum_{\mathrm{i}=1}^{\mathrm{q}'} \theta_{\mathrm{i}}\, u_{t-i} \\ [GARCH]:\ \sigma_{\mathrm{t}}{}^2 = c + \sum_{i=1}^{Q} \tau_i u_{t-i}{}^2 + \sum_{j=1}^{P} \rho_j \sigma_{\mathrm{t-j}}{}^2 + \vartheta_{\mathrm{t}} \end{cases}$$

where ϑ is an independently and identically distributed random term and σ the conditional standard error of u. $[M]$ can be estimated using a four-step procedure well described in Ruppert (2011):

- Step 1: estimate $[Reg]$ using ordinary least squares and determine the structure of lags $\mathrm{S(p)} \subseteq \{1, \dots, \mathrm{p}\}, \mathrm{S(k)} \subseteq \{1, \dots, \mathrm{k}\}$ and $\mathrm{S(m)} \subseteq \{1, \dots, \mathrm{m}\}$;
- Step 2: estimate an ARMA for the residuals of $[Reg]$;
- Step 3: compute the conditional variance of the Step 2 residuals using a GARCH equation; and
- Step 4: re-estimate $[Reg]$ using weighted least squares, with the weights equal to the reciprocal of the conditional variances computed in step 3.

Benchmarking models

The forecasting accuracy of $[M]$ is assessed against two basic models. Failure of $[M]$ to outperform the benchmarking models indicates that the forecasting methodology is not appropriate or, in other words, that the information generated by the explanatory variables and the way it is used does not significantly improve the forecasting of food inflation compared to models with no additional information and with a simple structure. The following models are used for the benchmarking:

$$[AR1]:\ \dot{p}_t = \mathrm{c} + \varphi \dot{p}_{t-1} + \varepsilon_t$$
$$[AR0]:\ \dot{p}_t = \mathrm{c} + \dot{p}_{t-1} + \varepsilon_t$$

Where ε is an independently and identically distributed random term. $[AR1]$ is a simple autoregressive model of degree one and $[AR0]$ is generally referred to as a random walk process.

b. Measuring forecasting accuracy

The different models will be assessed on their capacity to accurately forecast monthly food price changes, using the following metrics:

Root Mean Square Error (RMSE)

The RMSE measures the average magnitude of the forecasting error. It is expressed in the same unit as the endogenous variable and is therefore directly interpretable. Its mathematical expression is the following:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\dot{p}_t - \hat{\dot{p}}_t\right)^2} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_t)^2}$$

Where $\hat{\dot{p}}_t$ is the out-of-sample prediction of $\dot{p}_t$. One of the drawbacks of this measure is that it gives equal weight to overestimation and underestimation. This is also a purely quantitative indicator, which does not inform on other dimensions of forecasting accuracy, such as the capacity to anticipate changes in the sign of the variation (inflation or deflation, in our case) and its direction.

Sign of variation (Sign)

The capacity to adequately predict increases or decreases should be one of the essential properties of any model attempting to forecast economic time-series such as food prices. The best models are those that minimize the risk of wrongly forecasting inflation or deflation. Two statistics, $Sign^{+}$ and $Sign^{-}$ measure, respectively, the share of episodes of inflation and deflation accurately predicted by the model. $Sign$, the weighted average of the two, measures the average share of inflation and deflation episodes accurately forecasted. These statistics are computed using the following formulae:

$$\begin{cases} Sign^{+} = \sum_{t=1}^{T} 1\left\{\left(\hat{\dot{p}}_t \geq 0\right) \cap \left(\dot{p}_t \geq 0\right)\right\} \Big/ \sum_{t=1}^{T} 1\{(\dot{p}_t \geq 0)\} \\ Sign^{-} = \sum_{t=1}^{T} 1\left\{\left(\hat{\dot{p}}_t < 0\right) \cap \left(\dot{p}_t < 0\right)\right\} \Big/ \sum_{t=1}^{T} 1\{(\dot{p}_t < 0)\} \\ Sign = \dfrac{\sum_{t=1}^{T} 1\{(\dot{p}_t \geq 0)\}}{T} Sign^{+} + \dfrac{\sum_{t=1}^{T} 1\{(\dot{p}_t < 0)\}}{T} Sign^{-} \end{cases}$$

Where $1(a) = \begin{cases} 1 \; if \; the \; condition \; a \; is \; met \\ 0 \; if \; else \end{cases}$