Basic concepts and how to measure price volatility

Carlos Martins-Filho

University of Colorado, Boulder and IFPRI

AGRODEP - Dakar, Senegal

Some fundamental assumptions

- Producers of agricultural commodities do not have market power.
- Observed prices are realizations of stochastic variables.
- Hence, at any time period t, a collection of n observed past prices can be represented by

$$\{P_{t-n}, \cdots P_{t-2}, P_{t-1}\}$$

Choosing, without loss of generality, $t = n + 1$ we have

$$\{P_1, \cdots, P_{n-1}, P_n\}$$

A simple theoretical model

- Let $c(y;w)$ be a producer's cost function, where y denotes output and w denotes a vector of input prices
- Assume P has distribution given by F_P with expected value $\mu_P = \int p dF_P(p)$ and variance $\sigma_P^2 = \int (p - \mu_P)^2 dF_P(p)$
- σ_P is called the "volatility" of P (it is just the square root of the variance)
- Producers maximize short run profits by choosing output y^* such that
$$\mu_P = c'(y^*;w) \tag{1}$$
- Output cannot be instantaneously adjusted to price levels

Why do we care about volatility of prices?

- Suppose w.l.o.g. that for price P we have $y > y^*$
- lack of output adjustment produces a loss in profit given by

$$L = -Pdy + \int_{y^*}^{y} c'(\alpha; w) d\alpha \text{ where } dy = y - y^* \qquad (2)$$

- For algebraic convenience take $c'(y; w) = b(w) + 2c(w)y$ where $b(w)$ and $c(w)$ are constants depending on input prices
- Expected losses are given by

$$E(L) = \frac{1}{4c(w)} E(P - \mu_P)^2 = \frac{1}{4c(w)} \sigma_P^2 \qquad (3)$$

Equation (3) suggests the following benefits from reduced price volatility:

1. Smaller price volatility reduces expected loss
2. Since choosing output to maximize profit equates marginal cost to price, there is optimal allocation of inputs into the agricultural sector. Hence, misallocation is reduced by reducing price volatility. Large values of σ_P^2 may imply increased misallocation of resources.

Some basic questions regarding statistical modeling of prices

- Do the stochastic variables $P_1, P_2, \cdots, P_n$ have the same distribution?
- Does knowledge of the value of P_{n-1} help us forecast or predict P_n?
- Can we construct statistical models that help us understand the evolution of prices through time?

Returns

- Because prices can be measured in different units, it will be convenient to adopt an unit free measure.
- Assuming that $P_t \in (0, \infty)$ the returns over 1 period are normally measured in one of the following ways:
  1. Net returns: $R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \in (-1, \infty)$
  2. Gross returns: $A_t = 1 + R_t \in (0, \infty)$
  3. log returns: $r_t = log(1 + R_t) = logP_t - logP_{t-1} \in (-\infty, \infty)$.

  Note that:
  1. if prices do not change from period $t-1$ to period t all of these measures of return take the value zero.
  2. if there is a large change from period $t-1$ to period t then all of these measures of return take on large values.
  3. "High prices" and "high price variability" are very different concepts that do not preclude each other.

A simple model

- Suppose $\{r_t\}_{t=1,2,\dots}$ is an independent and identically distributed sequence of returns and assume

$$r_t = log \frac{P_t}{P_{t-1}} \sim N(\mu, \sigma^2)$$

or

$$r_t = \mu + \sigma u_t$$

where $\{u_t\}_{t=1,2,\dots}$ forms an independent sequence with $u_t \sim N(0,1)$.

- μ is called the expected value of r_t and σ^2 is called the variance of r_t. Take the square root to obtain volatility.
- It is easy to show that under this assumption on log-returns

$$E(log\ P_t) = log\ P_0 + t\mu \text{ and } V(log\ P_t) = t\sigma^2$$

where P_0 is an "initial" value for price.

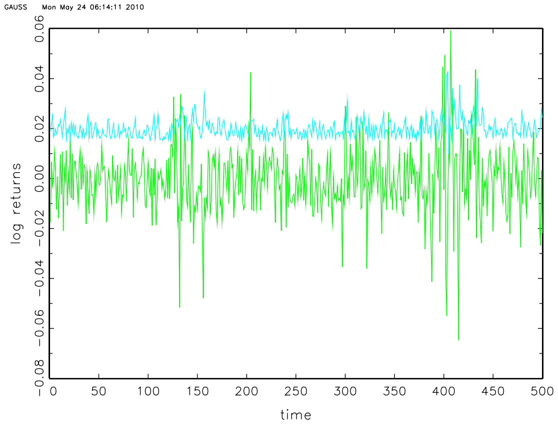

Figure: Estimated 95 % conditional quantile (blue) and realized log returns for soybeans (green)

Some questions

From the visual inspection of the graph, we ask:

- Does it seem that the expected value of returns is constant (μ)?
- Does it seem that variance (volatility) of returns (σ^2) is constant?
- Is the return distribution symmetric?
- Relative to a stochastic variable with a normal distribution, is the occurrence of extreme returns (too large or too small) more or less likely relative to the occurrence of "typical returns"?
- Is it possible to ascertain that some returns (or price variations from period $t-1$ to t) are in some sense too big or abnormally large?

A more sophisticated statistical model

We generalize the model described above as follows:

$$r_t = m(r_{t-1}, r_{t-2}, \cdots, r_{t-H}, w_{t.}) + h^{1/2}(r_{t-1}, r_{t-2}, \cdots, r_{t-H}, w_{t.})u_t$$

where

- $w_{t.}$ is a $1 \times K$ dimensional vector of random variables
- u_t are iid with distribution given by F_u, $E(u_t) = 0$ and $V(u_t) = 1$
- For simplicity, we put $X_{t.} = (r_{t-1}, r_{t-2}, \cdots, r_{t-H}, w_{t.})$ a $d = H + K$-dimensional vector and assume that

$$m(X_{t.}) = m_0 + \sum_{a=1}^{d} m_a(X_{ta}), \quad \text{and } h(X_{t.}) = h_0 + \sum_{a=1}^{d} h_a(X_{ta})$$

A more sophisticated statistical model

Note the following immediate consequences of the model:

- $E(r_t|X_{t.}) = m(r_{t-1}, r_{t-2}, \cdots, r_{t-H}, w_{t.}) \neq \mu$
- $V(r_t|X_{t.}) = h(r_{t-1}, r_{t-2}, \cdots, r_{t-H}, w_{t.}) \neq \sigma^2$
- The model structure permits a skewed conditional distribution of returns
- The model structure permits a leptokurtic or platykurtic conditional distribution.

A framework for identifying abnormally high returns

The α-quantile for the conditional distribution of r_t given $X_{t.}$, denoted by $q(\alpha|X_{t.})$ is given by

$$q(\alpha|X_{t.}) \equiv F^{-1}(\alpha|X_{t.}) = m(X_{t.}) + (h(X_{t.}))^{1/2} q(\alpha). \qquad (4)$$

- This conditional quantile is the value for returns that is exceeded with probability $1 - \alpha$ given past returns (down to period $t - H$) and other economic or market variables ($w_{t.}$)
- Large (positive) log-returns indicate large changes in prices from periods $t - 1$ to t and by considering α to be sufficiently large we can identify a threshold $q(\alpha|X_{t.})$ that is exceeded only with a small probability α.
- Realizations of r_t that are greater than $q(\alpha|X_{t.})$ are indicative of unusual price variations given the conditioning variables.

Estimation

- First, m and h are estimated by $\hat{m}(X_{t.})$ and $\hat{h}(X_{t.})$ given the sample $\{(r_t, X_{t1}, \cdots, X_{td})\}_{t=1}^{n}$
- Second, standardized residuals $\hat{\varepsilon}_t = \frac{r_t - \hat{m}(X_{t.})}{\hat{h}(X_{t.})^{1/2}}$ are used in conjunction with extreme value theory to estimate $q(\alpha)$.

The exceedances of any random variable (ϵ) over a specified nonstochastic threshhold u, i.e, $Z = \epsilon - u$ can be suitably approximated by a generalized pareto distribution - GPD (with location parameter equal to zero) given by,

$$G(x; \beta, \psi) = 1 - \left(1 + \psi \frac{x}{\beta}\right)^{-1/\psi}, x \in D \tag{5}$$

where $D = [0, \infty)$ if $\psi \geq 0$ and $D = [0, -\beta/\psi]$ if $\psi < 0$.

Estimation

1. Using $\hat{\varepsilon}_{1:n} \geq \hat{\varepsilon}_{2:n} \geq ... \geq \hat{\varepsilon}_{n:n}$ and obtain $k < n$ excesses over $\hat{\varepsilon}_{k+1:n}$ given by $\{\hat{\varepsilon}_{j:n} - \hat{\varepsilon}_{k+1:n}\}_{j=1}^{k}$
2. It is easy to show that for $\alpha > 1 - k/n$ and estimates $\hat{\beta}$ and $\hat{\psi}$, $q(\alpha)$ can be estimated by,

$$\widehat{q(\alpha)} = \hat{\varepsilon}_{k+1:n} + \frac{\hat{\beta}}{\hat{\psi}}\left(\left(\frac{1-\alpha}{k/n}\right)^{-\hat{\psi}} - 1\right). \qquad (6)$$

Empirical exercise

For this empirical exercise we use the following model

$$r_t = m_0 + m_1(r_{t-1}) + m_2(r_{t-2}) + (h_0 + h_1(r_{t-1}) + h_2(r_{t-2}))^{1/2}\,\varepsilon_t. \tag{7}$$

- For each of the series of log returns we select the first $n = 1000$ realizations (starting January 3, 1994) and forecast the 95% conditional quantile for the log return on the following day. This value is then compared to realized log return.
- This is repeated for the next 500 days with forecasts always based on the previous 1000 daily log returns. We expect to observe 25 returns that exceed the 95% estimated quantile

Soybeans: We expect 25 violations, i.e., values of the returns that exceed the estimated quantiles. The actual number of forecasted violations is 21 and the the p-value is 0.41, significantly larger than 5 percent, therefore providing evidence of the adequacy of the model.

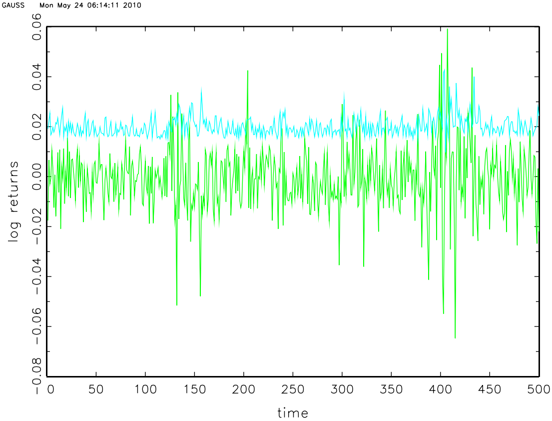

Figure: Estimated 95 % conditional quantile and realized log returns for soybeans

Hard wheat: We expect 25 violations, i.e., values of the returns that exceed the estimated quantiles. The actual number of forecasted violations is 21 and the the p-value is 0.41, significantly larger than 5 percent, therefore providing evidence of the adequacy of the model.

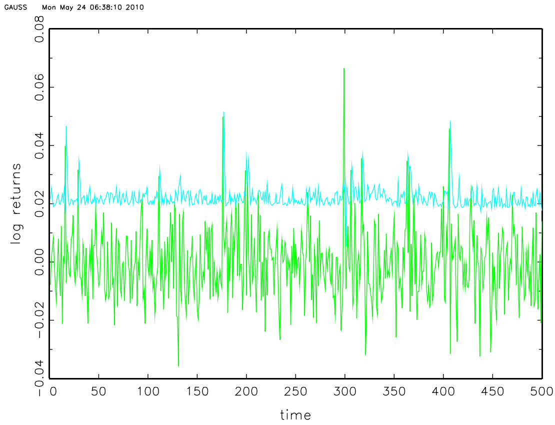

Figure: Estimated 95 % conditional quantile and realized log returns for hardwheat

Daily results are available at:

- http://www.foodsecurityportal.org/

In particular:

- Policy Analysis Tools