High order conditional quantile estimation: the case of returns on future contracts on agricultural commodities

Carlos Martins-Filho, Feng Yao and Maximo Torero

University of Colorado and IFPRI, West Virginia University and IFPRI

September 28, 2011
Motivation

In empirical finance there is often an interest in stochastic models for log returns

\[ r_t = \log \frac{P_t}{P_{t-1}} \text{ where } t \in \{0, \pm 1, \cdots \}. \]

A few popular models are:

a) ARCH (q)

\[ r_t = E(r_t|r_{t-1}, \cdots) + V(r_t|r_{t-1}, \cdots)^{1/2} U_t \]

where \( E(r_t|r_{t-1}, \cdots) = 0 \) and

\[ V(r_t|r_{t-1}, \cdots) = \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \cdots + \alpha_p r_{t-q}^2 \]

with \( E(U_t|r_{t-1}, \cdots) = 0 \), and \( E(U_t^2|r_{t-1}, \cdots) = 1 \)
b) GARCH(p,q)

\[ r_t = E(r_t | r_{t-1}, \cdots) + V(r_t | r_{t-1}, \cdots)^{1/2} U_t \]

where \( E(r_t | r_{t-1}, \cdots) = 0 \) and

\[ V(r_t | r_{t-1}, \cdots) = \alpha_0 + \alpha_1 r_{t-1}^2 + \cdots + \alpha_p r_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 \]

with \( E(U_t | r_{t-1}, \cdots) = 0 \), and \( E(U_t^2 | r_{t-1}, \cdots) = 1 \).

These models impose very specific functional structure on conditional means and variances. Martins-Filho and Yao (2006) show that this can be very costly.
Motivation

A more flexible modeling strategy is to consider a nonparametric model

\[ E(r_t|r_{t-1}, \cdots) = m(r_{t-1}, \cdots, r_{t-H}, w_t.) \]  \hspace{1cm} (1)

and

\[ V(r_t|r_{t-1}, \cdots) = h(r_{t-1}, \cdots, r_{t-H}, w_t.) \]  \hspace{1cm} (2)

where \( m, h \) belong to suitably defined classes of functions. Whatever model is used, their location-scale structure allows us to write, for \( a \in (0, 1) \)

\[
q_{r_t|r_{t-1}}(a) = E(r_t|r_{t-1}, \cdots, r_{t-H}, w_t.) + V(r_t|r_{t-1}, \cdots, r_{t-H}, w_t.)^{1/2} q(a)
\]
Model

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

$$m(X_t.) = m_0 + \sum_{a=1}^{d} m_a(X_{ta})$$

(3)

and

$$h(X_t.) = h_0 + \sum_{a=1}^{d} h_a(X_{ta})$$

(4)

and write

$$r_t = m_0 + \sum_{a=1}^{d} m_a(X_{ta}) + \left( h_0 + \sum_{a=1}^{d} h_a(X_{ta}) \right)^{1/2} U_t$$

(5)
Model

- $U_t$ has distribution $F(u)$ which is strictly increasing and belongs to the domain of attraction of an extremal distribution [Resnick(1987)]
- There are $F$’s that are not in the domain of attraction of $E$ [see Leadbetter et al. (1983)] but they constitute rather pathological examples.

We can write

$$q_{rt|X_t}(a) = m_0 + \sum_{a=1}^{d} m_a(X_{ta}) + \left( h_0 + \sum_{a=1}^{d} h_a(X_{ta}) \right)^{1/2} q(a) \quad (6)$$
The model

There are three unknown functionals in (6):

- If \( U_t \) were observed \( q(a) \) could be estimated from a random sample \( \{ U_t \}_{t=1}^n \).

Since we do not observe \( U_t \), a natural alternative is to produce an estimator for \( q(a) \) based on

\[
\hat{U}_t = \frac{Y_t - \hat{m}(X_t.)}{\hat{h}(X_t.)} \quad \text{for } i = 1, \ldots, n. \tag{7}
\]

where \( \hat{m} \) and \( \hat{h} \) are estimators of \( m \) and \( h \).
An interesting case

- We are particularly interested in the case where $a$ is very large (in the vicinity of 1), called high order (conditional) quantiles.
- These conditional quantiles have become particularly important in empirical finance where they are called conditional Value-at-Risk (CVaR) [see, inter alia, McNeil and Frey (2000), Martins-Filho and Yao (2006), Cai (2008)].
- Interestingly, the restriction that $a$ is in a neighborhood of 1 is useful in estimation. The result is due to Pickands (1975).

He showed that $F(x) \in D(E)$ is equivalent, for some fixed $k$ and function $\sigma(\xi)$, to

$$\lim_{\xi \to u_\infty} \sup_{0 < u < u_\infty - \xi} |F_\mu(u) - G(u; \sigma(\xi), k)| = 0 \quad (8)$$

where
An interesting case

\[ F_\xi(u) = \frac{F(u+\xi)-F(\xi)}{1-F(\xi)} \]

\[ u_\infty = \text{l.u.b}\{x : F(x) < 1\} \leq \infty \text{ with } u_\infty > \mu \in \mathbb{R} \]

\[ G \text{ is a generalized Pareto distribution, i.e.,} \]

\[
G(y; \sigma, k) = \begin{cases} 
1 - (1 - ky/\sigma)^{1/k} & \text{if } k \neq 0, \sigma > 0 \\
1 - \exp(-y/\sigma) & \text{if } k = 0, \sigma > 0
\end{cases}
\]

(9)

with \(0 < y < \infty\) if \(k < 0\) and \(0 < y < \sigma/k\) if \(k > 0\).

Comments:

\[ \text{If } F \text{ belongs to the domain of attraction of a Fréchet distribution } (\Phi_\alpha) \text{ with parameter } \alpha, \text{ then } k = -\frac{1}{\alpha} \text{ and } \sigma(\xi) = \xi/\alpha. \]

\[ \text{By (8) } G \text{ is a suitable parametric approximation for the upper tail of } F, \text{ an estimator for } q(a) \text{ can be obtained from the estimation of the parameters } k \text{ and } \sigma(\xi)!! \]
Estimation

Let \( \{ \hat{U}_t \}^n_{t=1} \) and \( Z_j = \hat{U}_{(n-N+j)} - \hat{U}_{(n-N)} \) for \( j = 1, \cdots, N \) where

\[
\hat{U}_t = \frac{Y_t - \hat{m}(X_t.)}{\hat{h}(X_t.)}
\]

for \( t = 1, \cdots, n \). We define the B-spline estimator for \( m \) evaluated at \( x = (x_1, \cdots, x_d) \) as

\[
\hat{m}(x) = \hat{\lambda}_0 + \sum_{a=1}^{d} \sum_{j=1}^{N_n} \hat{\lambda}_{j,a} I_{j,a}(x_a)
\]

(10)

where

\[
(\hat{\lambda}_0, \hat{\lambda}_{11}, \cdots, \hat{\lambda}_{N_n d}) = \text{argmin}_{\mathbb{R}^{dN_n + 1}} \sum_{t=1}^{n} \left( r_t - \lambda_0 - \sum_{a=1}^{d} \sum_{j=1}^{N_n} \lambda_{j,a} I_{j,a}(X_{ta}) \right)^2.
\]

(11)
The $\hat{\lambda}_{ja}$ are used to construct pilot estimators for each component $m_a(x_a)$, which are defined as

$$
\hat{m}_a(x_a) = \sum_{j=1}^{N_n} \hat{\lambda}_{j,a} l_{j,a}(x_a) - \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{N_n} \hat{\lambda}_{j,a} l_{j,a}(X_{ta})
$$

(12)

and

$$
\hat{m}_0 = \hat{\lambda}_0 + \frac{1}{n} \sum_{a=1}^{d} \sum_{t=1}^{n} \sum_{j=1}^{N_n} \hat{\lambda}_{j,a} l_{j,a}(X_{ta}).
$$

(13)
Estimation

These pilot estimators, together with $\hat{c} = \frac{1}{n} \sum_{t=1}^{n} r_t$ are used to construct pseudo-responses

$$\hat{r}_{ta} = r_t - \hat{c} - \sum_{\alpha=1, \alpha \neq a}^{d} \hat{m}_\alpha(X_{t\alpha}).$$

(14)

We then form $d$ sequences $\{(\hat{r}_{ta}, X_{ta})\}_{t=1}^{n}$ which are used to estimate $m_a$ via an univariate nonparametric regression smoother. The simplest is a Nadaraya-Watson kernel estimator, i.e.,

$$\hat{m}^*_a(x_a) = \frac{\sum_{t=1}^{n} K \left( \frac{X_{ta} - x_a}{h_n} \right) \hat{r}_{ta}}{\sum_{t=1}^{n} K \left( \frac{X_{ta} - x_a}{h_n} \right)}$$

(15)

where $K(\cdot)$ is a kernel function and $h_n$ is a bandwidth such that $h_n \propto n^{-1/5}$. The same procedure is used for the estimation of $h$, using as regressand $(r_t - \hat{m}(X_t.)^2)$. 
Order statistics are estimators for a quantiles associated with empirical distributions. That is,

\[ q_n(a) = \begin{cases} 
U_{(na)} & \text{if } na \in \mathbb{N} \\
U_{([na]+1)} & \text{if } na \notin \mathbb{N}
\end{cases} \]

and for \( a_n = 1 - \frac{N}{n} \) we can write

\[ \{Z_j\}_{j=1}^{N} = \left\{ U_{(n-N+j)} - q_n(a_n) \right\}_{j=1}^{N}. \]
Estimation of GPD parameters

1. First stage

Inspired by Azzalini (1981), Falk (1985) and Martins-Filho and Yao (2007) we define \( \tilde{q}(z) \) as the solution for

\[
\tilde{F}(\tilde{q}(z)) = z
\]

where \( \tilde{F}(u) = \int_{-\infty}^{u} \frac{1}{nh_{2n}} \sum_{i=1}^{n} K_{2} \left( \frac{\hat{U}_{i} - y}{h_{2n}} \right) dy \), \( K_{2}(\cdot) \) is a kernel function and \( 0 < h_{2n} \) is a bandwidth.

Now we can define the observed sequence

\[
\{ \tilde{Z}_{j} \}_{j=1}^{N} = \left\{ \hat{U}_{(n-N+j)} - \tilde{q}(a_{n}) \right\}_{j=1}^{N}
\]
2. Second stage:

We consider a solution \( (\tilde{\sigma}_N, \tilde{k}) \) for the following likelihood equations:

\[
\frac{\partial}{\partial \sigma} \frac{1}{N} \sum_{j=1}^{N} \log g(\tilde{Z}_j; \tilde{\sigma}_N, \tilde{k}) = 0 \tag{16}
\]

\[
\frac{\partial}{\partial k} \frac{1}{N} \sum_{j=1}^{N} \log g(\tilde{Z}_j; \tilde{\sigma}_N, \tilde{k}) = 0. \tag{17}
\]

where \( g(z; \sigma, k) = \frac{1}{\sigma} \left(1 - \frac{kz}{\sigma}\right)^{1/k-1} \)
**Estimation of \( q(a) \)**

Given a threshold \( \xi = U_{(n-N)} \) we can write

\[
F_{U_{(n-N)}}(y) = \frac{F(y + U_{(n-N)}) - F(U_{(n-N)})}{1 - F(U_{(n-N)})} \approx 1 - \left( 1 - \frac{ky}{\sigma_N} \right)^{1/k}
\]

For \( a \in (0, 1) \) we can write that

\[
q(a) = U_{(n-N)} + y_{N,a}
\]

where \( F(U_{(n-N)} + y_{N,a}) = a \). If \( 1 - F(U_{(n-N)}) \) is estimated by \( N/n \), we have

\[
\frac{1 - a}{N/n} \approx \left( 1 - \frac{ky}{\sigma_N} \right)^{1/k}, \quad (18)
\]

which suggests \( y_{N,a} \approx \frac{\sigma_N}{k} \left( 1 - \left( \frac{(1-a)n}{N} \right)^k \right) \). We define

\[
\hat{q}(a) = \tilde{q}(a_n) + \hat{y}_{N,a} = \tilde{q}(a_n) + \frac{\tilde{\sigma}_N}{\tilde{k}} \left( 1 - \left( \frac{(1-a)n}{N} \right)^{\tilde{k}} \right). \quad (19)
\]
Lastly, we combine the $\hat{q}(a)$ with $\hat{m}(x)$ to obtain,

$$\hat{q}_{r_t|x_t}(a) = \hat{m}(X_t) + \hat{h}(X_t)^{1/2} \hat{q}(a)$$

the estimator for $q_{r_t|x=x}(a)$. 

Estimation of $q_{r_t|x_t}(a)$