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

Carlos Martins-Filho, Feng Yao and Maximo Torero<br>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)

$$
r_{t}=E\left(r_{t} \mid r_{t-1}, \cdots\right)+V\left(r_{t} \mid r_{t-1}, \cdots\right)^{1 / 2} U_{t}
$$

where $E\left(r_{t} \mid r_{t-1}, \cdots\right)=0$ and

$$
V\left(r_{t} \mid r_{t-1}, \cdots\right)=\sigma_{t}^{2}=\alpha_{0}+\alpha_{1} r_{t-1}^{2}+\cdots+\alpha_{p} r_{t-q}^{2}
$$

with $E\left(U_{t} \mid r_{t-1}, \cdots\right)=0$, and $E\left(U_{t}^{2} \mid r_{t-1}, \cdots\right)=1$
b) $\operatorname{GARCH}(p, q)$

$$
r_{t}=E\left(r_{t} \mid r_{t-1}, \cdots\right)+V\left(r_{t} \mid r_{t-1}, \cdots\right)^{1 / 2} U_{t}
$$

where $E\left(r_{t} \mid r_{t-1}, \cdots\right)=0$ and
$V\left(r_{t} \mid r_{t-1}, \cdots\right)=\alpha_{0}+\alpha_{1} r_{t-1}^{2}+\cdots+\alpha_{p} r_{t-q}^{2}+\beta_{1} \sigma_{t-1}^{2}+\cdots+\beta_{p} \sigma_{t-p}^{2}$
with $E\left(U_{t} \mid r_{t-1}, \cdots\right)=0$, and $E\left(U_{t}^{2} \mid r_{t-1}, \cdots\right)=1$.

- 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

$$
\begin{equation*}
E\left(r_{t} \mid r_{t-1}, \cdots\right)=m\left(r_{t-1}, \cdots, r_{t-H}, w_{t .}\right) \tag{1}
\end{equation*}
$$

and

$$
\begin{equation*}
V\left(r_{t} \mid r_{t-1}, \cdots\right)=h\left(r_{t-1}, \cdots, r_{t-H}, w_{t .}\right) \tag{2}
\end{equation*}
$$

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

$$
\begin{aligned}
& q_{r_{t} \mid r_{t-1}, \cdots}(a)=E\left(r_{t} \mid r_{t-1}, \cdots, r_{t-H}, w_{t .}\right) \\
& \quad+V\left(r_{t} \mid r_{t-1}, \cdots, r_{t-H}, w_{t .}\right)^{1 / 2} q(a)
\end{aligned}
$$

## Model

For simplicity, we put $X_{t .}=\left(r_{t-1}, r_{t-2}, \cdots, r_{t-H}, w_{t .}\right)$ a $d=H+K$-dimensional vector and assume that

$$
\begin{equation*}
m\left(X_{t .}\right)=m_{0}+\sum_{a=1}^{d} m_{a}\left(X_{t a}\right) \tag{3}
\end{equation*}
$$

and

$$
\begin{equation*}
h\left(X_{t .}\right)=h_{0}+\sum_{a=1}^{d} h_{a}\left(X_{t a}\right) \tag{4}
\end{equation*}
$$

and write

$$
\begin{equation*}
r_{t}=m_{0}+\sum_{a=1}^{d} m_{a}\left(X_{t a}\right)+\left(h_{0}+\sum_{a=1}^{d} h_{a}\left(X_{t a}\right)\right)^{1 / 2} U_{t} \tag{5}
\end{equation*}
$$

## 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

$$
\begin{equation*}
q_{r_{t} \mid X_{t}}(a)=m_{0}+\sum_{a=1}^{d} m_{a}\left(X_{t a}\right)+\left(h_{0}+\sum_{a=1}^{d} h_{a}\left(X_{t a}\right)\right)^{1 / 2} q(a) \tag{6}
\end{equation*}
$$

## The model

There are three unknown functionals in (6):

- If $U_{t}$ were observed $q(a)$ could be estimated from a random sample $\left\{U_{t}\right\}_{t=1}^{n}$
Since we do not observe $U_{t}$, a natural alternative is to produce an estimator for $q(a)$ based on

$$
\begin{equation*}
\hat{U}_{t}=\frac{Y_{t}-\hat{m}\left(X_{t .}\right)}{\hat{h}\left(X_{t .}\right)} \text { for } i=1, \cdots, n . \tag{7}
\end{equation*}
$$

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

$$
\begin{equation*}
\lim _{\xi \rightarrow u_{\infty}} \sup _{0<u<u_{\infty}-\xi}\left|F_{\mu}(u)-G(u ; \sigma(\xi), k)\right|=0 \tag{8}
\end{equation*}
$$

where

## An interesting case

- $F_{\xi}(u)=\frac{F(u+\xi)-F(\xi)}{1-F(\xi)}$
- $u_{\infty}=$ l.u.b $\{x: F(x)<1\} \leq \infty$ with $u_{\infty}>\mu \in \Re$
- $G$ is a generalized Pareto distribution, i.e.,

$$
G(y ; \sigma, k)=\left\{\begin{array}{cl}
1-(1-k y / \sigma)^{1 / k} & \text { if } k \neq 0, \sigma>0  \tag{9}\\
1-\exp (-y / \sigma) & \text { if } k=0, \sigma>0
\end{array}\right.
$$

with $0<y<\infty$ if $k<0$ and $0<y<\sigma / k$ if $k>0$.
Comments:

- If $F$ belongs to the domain of attraction of a Fréchet distribution $\left(\Phi_{\alpha}\right)$ with parameter $\alpha$, then $k=-\frac{1}{\alpha}$ and $\sigma(\xi)=\xi / \alpha$.
- By (8) $G$ is a suitable parametric approximation for the upper tail of $F$, an estimator for $q(a)$ can be obtained from the estimation of the parameters $k$ and $\sigma(\xi)!$ !


## Estimation

Let $\left\{\hat{U}_{(t)}\right\}_{t=1}^{n}$ and $Z_{j}=\hat{U}_{(n-N+j)}-\hat{U}_{(n-N)}$ for $j=1, \cdots, N$ where

$$
\hat{U}_{t}=\frac{Y_{t}-\hat{m}\left(X_{t .}\right)}{\hat{h}\left(X_{t .}\right)}
$$

for $t=1, \cdots, n$. We define the $B$-spline estimator for $m$ evaluated at $x=\left(x_{1}, \cdots, x_{d}\right)$ as

$$
\begin{equation*}
\hat{m}(x)=\hat{\lambda}_{0}+\sum_{a=1}^{d} \sum_{j=1}^{N_{n}} \hat{\lambda}_{j, a} l_{j, a}\left(x_{a}\right) \tag{10}
\end{equation*}
$$

where

$$
\begin{equation*}
\left(\hat{\lambda}_{0}, \hat{\lambda}_{11}, \cdots, \hat{\lambda}_{N_{n} d}\right)=\underset{\Re^{d N_{n}+1}}{\operatorname{argmin}} \sum_{t=1}^{n}\left(r_{t}-\lambda_{0}-\sum_{a=1}^{d} \sum_{j=1}^{N_{n}} \lambda_{j, a} l_{j, a}\left(X_{t a}\right)\right)^{2} \tag{11}
\end{equation*}
$$

## Estimation

The $\hat{\lambda}_{j a}$ are used to construct pilot estimators for each component $m_{a}\left(x_{a}\right)$, which are defined as

$$
\begin{equation*}
\hat{m}_{a}\left(x_{a}\right)=\sum_{j=1}^{N_{n}} \hat{\lambda}_{j, a} l_{j, a}\left(x_{a}\right)-\frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{N_{n}} \hat{\lambda}_{j, a} l_{j, a}\left(X_{t a}\right) \tag{12}
\end{equation*}
$$

and

$$
\begin{equation*}
\hat{m}_{0}=\hat{\lambda}_{0}+\frac{1}{n} \sum_{a=1}^{d} \sum_{t=1}^{n} \sum_{j=1}^{N_{n}} \hat{\lambda}_{j, a} I_{j, a}\left(X_{t a}\right) \tag{13}
\end{equation*}
$$

## Estimation

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

$$
\begin{equation*}
\hat{r}_{t a}=r_{t}-\hat{c}-\sum_{\alpha=1, \alpha \neq a}^{d} \hat{m}_{\alpha}\left(X_{t \alpha}\right) \tag{14}
\end{equation*}
$$

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

$$
\begin{equation*}
\hat{m}_{a}^{*}\left(x_{a}\right)=\frac{\sum_{t=1}^{n} K\left(\frac{X_{t a}-X_{a}}{h_{n}}\right) \hat{r}_{t a}}{\sum_{t=1}^{n} K\left(\frac{X_{t a}-x_{a}}{h_{n}}\right)} \tag{15}
\end{equation*}
$$

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 $\left(r_{t}-\hat{m}\left(X_{t}\right)\right)^{2}$.

## Estimation

Order statistics are estimators for a quantiles associated with empirical distributions. That is,

$$
q_{n}(a)=\left\{\begin{array}{cc}
U_{(n a)} & \text { if } n a \in \mathbb{N} \\
U_{([n a]+1)} & \text { if } n a \notin \mathbb{N}
\end{array}\right.
$$

and for $a_{n}=1-\frac{N}{n}$ we can write
$\left\{Z_{j}\right\}_{j=1}^{N}=\left\{U_{(n-N+j)}-q_{n}\left(a_{n}\right)\right\}_{j=1}^{N}$.

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

Now we can define the observed sequence

$$
\left\{\tilde{Z}_{j}\right\}_{j=1}^{N}=\left\{\hat{U}_{(n-N+j)}-\tilde{q}\left(a_{n}\right)\right\}_{j=1}^{N}
$$

## Estimation of GPD parameters

2. Second stage:

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

$$
\begin{align*}
& \frac{\partial}{\partial \sigma} \frac{1}{N} \sum_{j=1}^{N} \log g\left(\tilde{Z}_{j} ; \tilde{\sigma}_{N}, \tilde{k}\right)=0  \tag{16}\\
& \frac{\partial}{\partial k} \frac{1}{N} \sum_{j=1}^{N} \log g\left(\tilde{Z}_{j} ; \tilde{\sigma}_{N}, \tilde{k}\right)=0 \tag{17}
\end{align*}
$$

where $g(z ; \sigma, k)=\frac{1}{\sigma}\left(1-\frac{k z}{\sigma}\right)^{1 / k-1}$

## Estimation of $q(a)$

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

$$
F_{U_{(n-N)}}(y)=\frac{F\left(y+U_{(n-N)}\right)-F\left(U_{(n-N)}\right)}{1-F\left(U_{(n-N)}\right)} \approx 1-\left(1-\frac{k y}{\sigma_{N}}\right)^{1 / k}
$$

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

$$
q(a)=U_{(n-N)}+y_{N, a}
$$

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

$$
\begin{equation*}
\frac{1-a}{N / n} \approx\left(1-\frac{k y}{\sigma_{N}}\right)^{1 / k} \tag{18}
\end{equation*}
$$

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

$$
\begin{equation*}
\hat{q}(a)=\tilde{q}\left(a_{n}\right)+\hat{y}_{N, a}=\tilde{q}\left(a_{n}\right)+\frac{\tilde{\sigma}_{N}}{\tilde{k}}\left(1-\left(\frac{(1-a) n}{N}\right)^{\tilde{k}}\right) \tag{19}
\end{equation*}
$$

## Estimation of $q_{r_{t} \mid X_{t} .}(a)$

Lastly, we combine the $\hat{q}(a)$ with $\hat{m}(x)$ to obtain,

$$
\hat{q}_{r_{t} \mid X_{t .}}(a)=\hat{m}\left(X_{t .}\right)+\hat{h}\left(X_{t .}\right)^{1 / 2} \hat{q}(a)
$$

the estimator for $q_{r t} \mid X=x(a)$.

