Two-step conditional α-quantile estimation via additive models of location and scale[1]

Carlos Martins-Filho

Department of Economics
University of Colorado
Boulder, CO 80309-0256, USA
email: carlos.martins@colorado.edu
Voice: + 1 303 492 4599

&

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: c.martins-filho@cgiar.org
Voice: + 1 202 862 8144

Maximo Torero

IFPRI
2033 K Street
Washington, DC 20006-1002, USA
email: m.torero@cgiar.org
Voice: + 1 202 862 8144

and

Feng Yao

Department of Economics
West Virginia University
Morgantown, WV 26505, USA
email: feng.yao@mail.wvu.edu
Voice: +1 304 2937867

May, 2010

Abstract.

Keywords and phrases.

JEL Classifications. C14, C21

AMS-MS Classification. 62G05, 62G08, 62G20.

[1]

1 Introduction

Let P_t denote the price of an asset (commodity) of interest in time period t where $t \in T = \{0, \pm 1, \pm 2, \cdots\}$ We denote the net returns over the most recent period by $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$ and the log-returns by $r_t = log(1 + R_t) = logP_t - logP_{t-1}$. We assume that

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

where H is a finite number in $\{0, 1, 2, \cdots\}$, $w_{t.}$ is a $1 \times K$ dimensional vector of random variables which may include lagged variables of its components. The functions $m(\cdot) : \Re^d \to \Re$ and $h(\cdot) : \Re^d \to (0, \infty)$ belong to a to a suitably restricted class to be defined below but we specifically avoid the assumption that these functions can be parametrically indexed. ε_t are components of an independent and identically distributed process with marginal distribution given by F_ϵ which does not depend on $(r_{t-1}, r_{t-2}, \cdots, r_{t-H}, w_{t.})$, $E(\epsilon_t) = 0$ and $V(\epsilon_t) = 1$. 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

$$m(X_{t.}) = m_0 + \sum_{a=1}^{d} m_a(X_{ta}), \quad \text{and } h(X_{t.}) = h_0 + \sum_{a=1}^{d} h_a(X_{ta}). \tag{2}$$

Hence we write,[1]

$$r_t = m_0 + \sum_{a=1}^{d} m_a(X_{ta}) + \left(h_0 + \sum_{a=1}^{d} h_a(X_{ta})\right)^{1/2} \varepsilon_t. \tag{3}$$

There exists a sample of size n denoted by $\{(r_t, X_{t1}, \cdots, X_{td})\}_{t=1}^{n}$ which are taken to be realizations from an α-mixing process following (3) and for identification purposes we assume that $E(m_a(X_{ta})) = E(h_a(X_{ta})) = 0$ for all a.

Under the assumption that F_ϵ is strictly increasing in its domain we define for $\alpha \in (0, 1)$ the α-quantile $q(\alpha) = F_\epsilon^{-1}(\alpha)$. Then, the α-quantile for the conditional distribution of r_t given $X_{t.}$, denoted by $q(\alpha|X_{t.})$ is given by

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

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 ($w_{t.}$). Clearly, large (positive) log-returns indicate large changes in prices from periods $t-1$ to t and by considering α to be sufficiently large we

[1] We note that the set of random variables appearing as arguments in m and h need not coincide. We keep them the same to facilitate notation and accommodate the most general setting.

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.[2] In the next section we outline an estimation strategy for $q(\alpha|X_{t.})$.

2 Estimation

Estimation of $q(\alpha|X_{t.})$ will be conducted in two stages. 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 $q(\alpha)$. Conceptually, the estimation strategy follows Martins-Filho and Yao (2006) but the the set of allowable conditioning variables $(X_{t.})$ here is much richer than the set they considered. This added generality requires more involved steps in the estimation of m and h and motivated the additive structure described in (2).

2.1 Estimation of m and h

We estimate m by the spline backfitted kernel (SBK) proposed by Wang and Yang (2007). We assume that every component of $X_{t.}$ takes values in a compact interval $[l_a, u_a] \subset \Re$ for $a = 1, \cdots, d$. For each interval we select a collection of equally spaced *knots* $l_a = k_0 < k_1 < k_2 < \cdots < k_{N_n} < u_a = k_{N_n+1}$. $\{k_i\}_{i=}^{N_n}$ is the collection of *interior knots* and N_n, the number of interior knots, is proportional to n, specifically $N_n \propto n^{2/5} log\ n$ but does not dependent on a. The interior knots divide the interval $[l_a, u_a]$ in $N_n + 1$ subintervals $[k_j, k_{j+1})$ for $j = 0, 1, \cdots, N_n$ each of length $g_n = (u_a - l_a)/(N_n + 1)$. Let $I_{j,a}(x_a) = \begin{cases} 1 & \text{if } x_a \in [k_j, k_{j+1}) \\ 0 & \text{otherwise} \end{cases}$ for $j = 0, 1, \cdots, N_n$ and for all a. We define the B-spline estimator for m evaluated at $x = (x_1, \cdots, x_d)$ as

$$\hat{m}(x) = \hat{\lambda}_0 + \sum_{a=1}^{d} \sum_{j=1}^{N_n} \lambda_{j,a} I_{j,a}(x_a) \tag{5}$$

where

$$(\hat{\lambda}_0, \hat{\lambda}_{11}, \cdots, \hat{\lambda}_{N_n d}) = \underset{\Re^{dN_n+1}}{argmin} \sum_{t=1}^{n} \left(r_t - \lambda_0 - \sum_{a=1}^{d} \sum_{j=1}^{N_n} \lambda_{j,a} I_{j,a}(X_{ta}) \right)^2. \tag{6}$$

The $\hat{\lambda}_{ja}$ are used to construct pilot estimators for each component $m_a(x_a)$ in equation (3), which are defined as

$$\hat{m}_a(x_a) = \sum_{j=1}^{N_n} \hat{\lambda}_{j,a} I_{j,a}(x_a) - \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{N_n} \hat{\lambda}_{j,a} I_{j,a}(X_{ta}) \text{ and } \hat{m}_0 = \hat{\lambda}_0 + \frac{1}{n} \sum_{a=1}^{d} \sum_{t=1}^{n} \sum_{j=1}^{N_n} \hat{\lambda}_{j,a} I_{j,a}(X_{ta}). \tag{7}$$

[2] Unusual price changes may be indicative of speculative behavior on the market of market agents.

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

$$\hat{r}_{ta} = r_t - \hat{c} - \sum_{\alpha=1,\alpha\neq a}^{d} \hat{m}_\alpha(X_{t\alpha}). \tag{8}$$

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. There are various convenient kernel based choices. The simplest is a Nadaraya-Watson kernel estimator, i.e.,

$$\hat{m}_a^*(x_a) = \frac{\sum_{t=1}^{n} K\left(\frac{X_{ta}-x_a}{h_n}\right)\hat{r}_{ta}}{\sum_{t=1}^{n} K\left(\frac{X_{ta}-x_a}{h_n}\right)} \tag{9}$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth such that $h_n \propto n^{-1/5}$. Wang and Yang (2007) prove that for any $x_a \in [l_a + h_n, u_a - h_n]$

$$\sqrt{nh_n}(\hat{m}_a^*(x_a) - m_a(x_a) - h_n^2 b_a(x_a)) \xrightarrow{d} N(0, v_a^2(x_a) = E(h(X_1, \cdots, X_d)|X_a = x_a)(f_a(x_a))^{-1}\int K^2(u)du)$$

where $b_a(x_a) = \left((1/2)m_a^{(2)}(x_a)f_a(x_a) + m_a^{(1)}(x_a)f_a^{(1)}(x_a)\right)(f_a(x_a))^{-1}\int u^2K(u)du$, $f_a(x_a)$ is the marginal density of the random variable X_a, and for an arbitrary function g, $g^{(\delta)}$ indicates the δ-th derivative. The estimator for $m(x_1, \cdots, x_d)$ is naturally given by $\hat{m}^*(x_1, \cdots, x_d) = \hat{c} + \sum_{a=1}^{d}\hat{m}_a^*(x_a)$.

To estimate h we follow the same procedure outlined in the estimation of m with r_t substituted with the squared residulas $\hat{u}_t^2 = (r_t - \hat{m}^*(X_{t1}, \cdots, X_{td}))^2$. The resulting estimator for $h(x_1, \cdots, x_d)$ is denoted by $\hat{h}^*(x_1, \cdots, x_d)$. The estimators $\hat{m}^*$ and $\hat{h}^*$ are used to construct a sequence of estimated standardized residuals $\hat{\varepsilon}_t = \frac{r_t - \hat{m}^*(X_{t.})}{(\hat{h}^*(X_{t.}))^{1/2}}$ which will be used in the estimation of $q(\alpha)$.

2.2 Estimation of q_α

The estimation of q_α follows Martins-Filho and Yao (2006). The estimation is based on a fundamental result from extreme value theory, which states that the distribution of 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,

$$G(x; \beta, \psi) = 1 - \left(1 + \psi\frac{x}{\beta}\right)^{-1/\psi}, x \in D \tag{10}$$

where $D = [0, \infty)$ if $\psi \geq 0$ and $D = [0, -\beta/\psi]$ if $\psi < 0$. Estimated standardized residuals $\hat{\varepsilon}_t$ will be used to estimate the tails of the density f_ϵ. For this purpose we order the residuals such that $\hat{\varepsilon}_{j:n}$ is the j^{th} largest residual, i.e., $\hat{\varepsilon}_{1:n} \geq \hat{\varepsilon}_{2:n} \geq ... \geq \hat{\varepsilon}_{n:n}$ and obtain $k < n$ excesses over $\hat{\varepsilon}_{k+1:n}$ given by

$\{\hat{\varepsilon}_{j:n} - \hat{\varepsilon}_{k+1:n}\}_{j=1}^{k}$, which will be used for estimation of a GPD. By fixing k we in effect determine the residuals that are used for tail estimation and randomly select the threshold. It is easy to show that for $\alpha > 1 - k/n$ and estimates $\hat{\beta}$ and $\hat{\psi}$, $q(\alpha)$ can be estimated by,

$$\widehat{q(\alpha)} = \hat{\varepsilon}_{k+1:n} + \frac{\hat{\beta}}{\hat{\psi}} \left(\left(\frac{1-\alpha}{k/n} \right)^{-\hat{\psi}} - 1 \right). \tag{11}$$

Combining the estimator in (11) with first stage estimators, and using (4) gives estimators for $q(\alpha|X_{t.})$. We now discuss how we proceed with the estimation of β and ψ.

2.3 L-Moment Estimation of β and ψ

Given the results in Smith (1984, 1987), estimation of the GPD parameters has normally been conducted by constrained maximum likelihood (ML). Here we propose an alternative estimator based on L-Moment Theory (Hosking (1990); Hosking and Wallis (1997)). Traditionally, raw moments have been used to describe the location, scale, and shape of distribution functions. L-Moment Theory provides an alternative approach that exhibits a number of desirable properties.

Let F_ϵ be a distribution function associated with a random variable ϵ and $q(u) : (0,1) \rightarrow \Re$ its quantile. The r^{th} L-moment of ϵ is defined as,

$$\lambda_r = \int_0^1 q(u) P_{r-1}(u) du \text{ for } r = 1, 2, ... \tag{12}$$

where $P_r(u) = \sum_{k=0}^{r} p_{r,k} u^k$ and $p_{r,k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}$, which contrasts with the traditional raw moments $\mu_r = \int_0^1 q(u)^r du$. Theorem 1 in Hosking (1990) gives the following justification for using L-moments to describe distributions: a) μ_1 is finite if and only if λ_r exist for all r; b) a distribution F_ϵ with finite μ_1 is uniquely characterized by λ_r for all r. Thus, a distribution can be characterized by its L-moments even if raw moments of order greater than 1 do not exist, and most importantly, this characterization is unique, which is not true for raw moments.

It is easily verified that $\lambda_1 = \mu_1$, therefore the first L-moment when it exists provides the traditionally used measure of location for a distribution. As pointed out by Hosking (1990); Hosking and Wallis (1997), λ_2 is up to a scalar the expectation of Gini's mean difference statistic, therefore providing a measure of scale that differs from the traditional variance - $\mu_2 - \mu_1^2$ by placing smaller weights on differences between realizations of the random variable. Hosking (1989) shows that if μ_1 exists $-1 < \tau_3 \equiv \frac{\lambda_3}{\lambda_2} < 1$ with

$\tau_3 = 0$ for symmetric distributions, providing a bounded measure of skewness that is less sensitive to the extreme tails of the distribution than the traditional (unbounded) measure of skewness given by $\frac{\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}}$. Similarly, $-1 < \tau_4 \equiv \frac{\lambda_4}{\lambda_2} < 1$ can be interpreted as a bounded measure of kurtosis (Oja (1981)) that is less sensitive to the extreme tails of the distribution than the traditional (unbounded) measure given by $\frac{\mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2}$. Hence, contrary to traditional measures of location and shape, L-moment based measures of scale, skewness and kurtosis do not require the existence of higher order raw moments, allowing for synthetic measures of distribution shape even when higher order raw moments do not exist.

In addition, L-moments can be used to estimate a finite number of parameters $\theta \in \Theta$ that identify a member of a family of distributions. Suppose $\{F_\epsilon(\theta) : \theta \in \Theta \subset \Re^p\}$, p a natural number, is a family of distributions which is known up to θ. A sample $\{\epsilon_t\}_{t=1}^n$ is available and the objective is to estimate θ. Since, λ_r, $r = 1, 2, 3...$ uniquely characterizes F_ϵ, θ may be expressed as a function of λ_r. Hence, if estimators $\hat{\lambda}_r$ are available, we may obtain $\hat{\theta}(\hat{\lambda}_1, \hat{\lambda}_2, ...)$. From equation (12), $\lambda_{r+1} = \sum_{k=0}^r p_{r,k}\beta_k$ where $\beta_k = \int_0^1 q(u)u^k du$ for $r = 0, 1, 2, \cdots$. Given the sample, we define $\epsilon_{k,n}$ to be the k^{th} smallest element of the sample, such that $\epsilon_{1,n} \leq \epsilon_{2,n} \leq ... \leq \epsilon_{n,n}$. An unbiased estimator of β_k is

$$\hat{\beta}_k = n^{-1} \sum_{j=k+1}^{n} \frac{(j-1)(j-2)...(j-k)}{(n-1)(n-2)...(n-k)} \epsilon_{j,n}$$

and we define $\hat{\lambda}_{r+1} = \sum_{k=0}^r p_{r,k}\hat{\beta}_k$ for $r = 0, 1, \cdots, n-1$.

In particular, if F_ϵ is a generalized pareto distribution with $\theta = (\mu, \beta, \psi)$, it can be shown that $\mu = \lambda_1 - (2-\psi)\lambda_2$, $\beta = (1-\psi)(2-\psi)\lambda_2$, $\psi = -\frac{1-3(\lambda_3/\lambda_2)}{1+(\lambda_3/\lambda_2)}$. In our case, where $\mu = 0$, $\beta = (1-\psi)\lambda_1$, $\psi = 2 - \lambda_1/\lambda_2$ we define the following L-moment estimators for ψ and β,

$$\hat{\psi} = 2 - \frac{\hat{\lambda}_1}{\hat{\lambda}_2} \text{ and } \hat{\beta} = (1 - \hat{\psi})\hat{\lambda}_1.$$

Similar to ML estimators, these L-moment parameter estimators are $\sqrt{n}$-asymptotically normal for $\psi < 0.5$. However, they are much easier to compute than ML estimators as no numerical optimization or iterative procedure is necessary. Although asymptotically inefficient relative to ML estimators, L-moment based parameter estimators have reasonably high asymptotic efficiency (Hosking (1990)). For the GPD considered here, asymptoic efficiency is always higher than 70 percent when $0 < \psi < 0.3$.

More important, from a practical perspective, is that L-Moment based parameter estimators can

outperform ML (based on mean squared error) in finite samples as indicated by Hosking et al. (1985); Hosking (1987). The results are not entirely surprising as the efficiency of ML estimators is attained only asymptotically. In fact, as observed by Hosking and Wallis (1997), it may be necessary to deal very large samples before asymptotic distributions provide useful approximations to their finite sample equivalents. This seems to be especially true for GPD estimation, but it can also be verified in other more general contexts.

3 Empirical exercise

We have used the estimator described in the previous sections to estimate conditional quantiles for log returns of future prices (contracts expiring between one and three months) of hard wheat, soft wheat, corn and soybeans. For these empirical exercises we use the following model

$$r_t = m_0 + m_1(r_{t-1}) + m_2(r_{t-2}) + (h_0 + h_1(r_{t-1}) + h_2(r_{t-2}))^{1/2}\,\varepsilon_t. \tag{13}$$

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. Based on an asymptotic approximation of the binomial distribution by a Gaussian distribution, we calculate p-values to test the adequacy of our model in forecasting the conditional quantiles. The results for each price series are given below together with figures 1-4 that provide quantile forecasted values (blue line) and realized log returns (green line).

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.

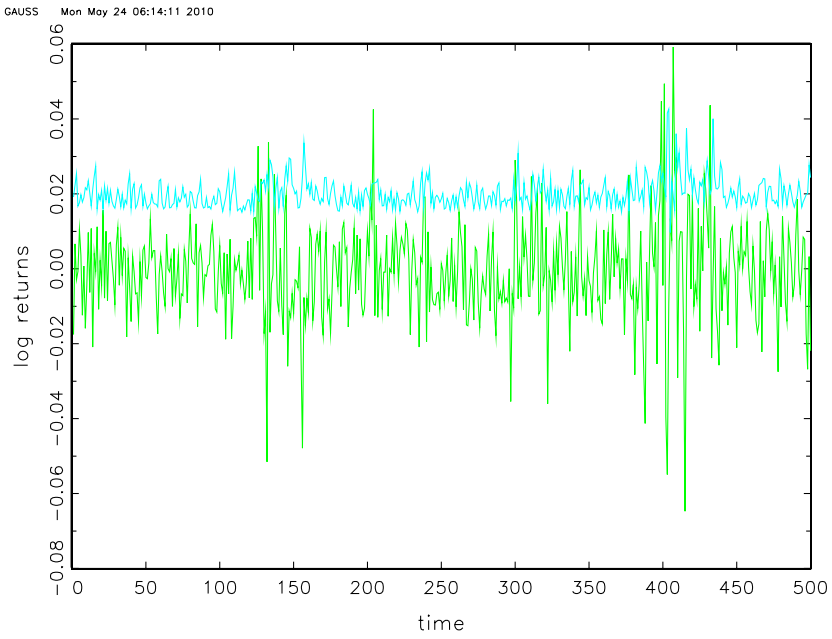

Figure 1: Estimated 95 % conditional quantile and realized log returns for soybeans

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.

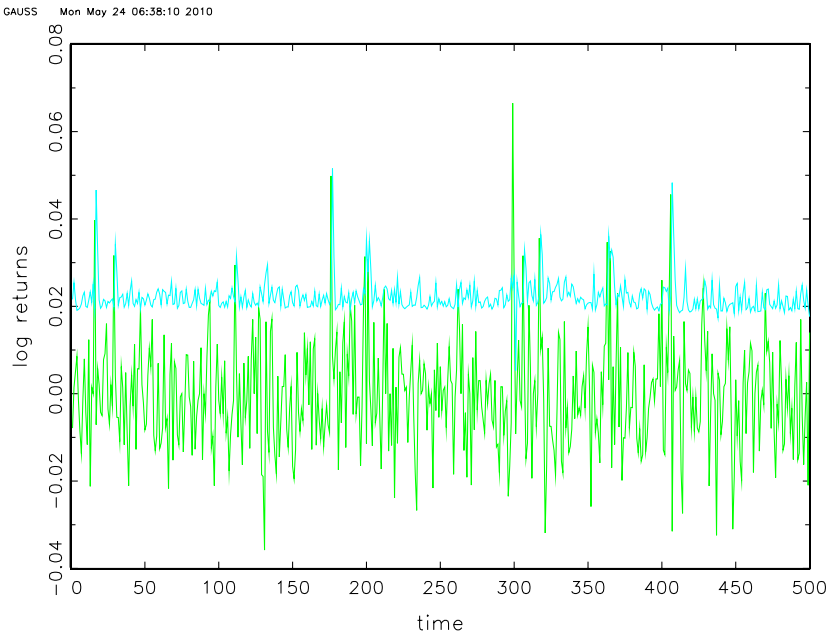

Figure 2: Estimated 95 % conditional quantile and realized log returns for hardwheat

Soft wheat: We expect 25 violations, i.e., values of the returns that exceed the estimated quantiles. The actual number of forecasted violations is 25 and the the p-value is 1, significantly larger than 5 percent, therefore providing evidence of the adequacy of the model.

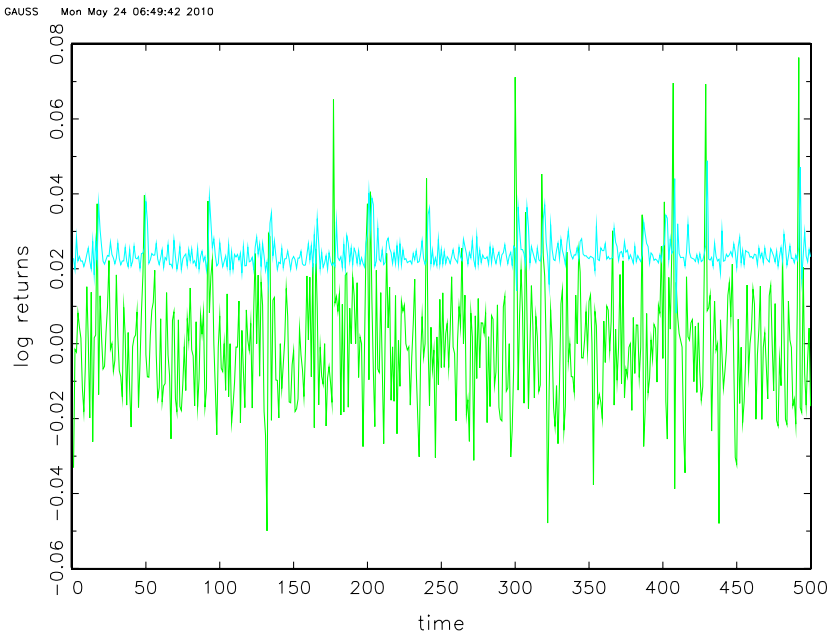

Figure 3: Estimated 95 % conditional quantile and realized log returns for softwheat

Corn: We expect 25 violations, i.e., values of the returns that exceed the estimated quantiles. The actual number of forecasted violations is 34 and the the p-value is 0.06, larger than 5 percent, therefore providing evidence of the adequacy of the model. However, in this case evidence is not as strong as in the case for soybeans, hard wheat or soft wheat.

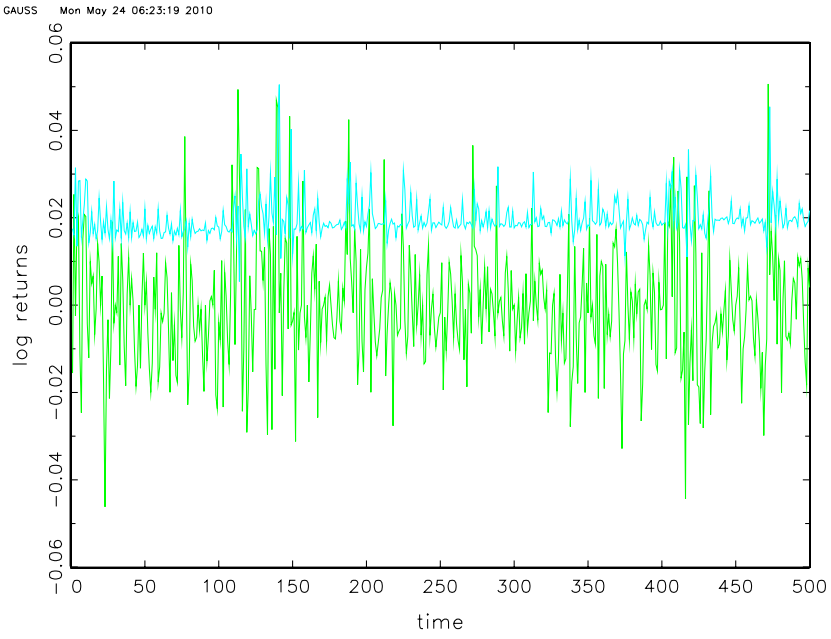

Figure 4: Estimated 95 % conditional quantile and realized log returns for corn

References

Hosking, J. R. M., 1987. Parameter and quantile estimation for the generalized pareto distribution. Technometrics 29, 339–349.

Hosking, J. R. M., 1989. Some theoretical results regarding L-moments.
URL http://www.research.ibm.com/people/h/hosking/lmoments.papers1.html

Hosking, J. R. M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society B 52, 105–124.

Hosking, J. R. M., Wallis, J. R., 1997. Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge, UK.

Hosking, J. R. M., Wallis, J. R., Wood, E. F., 1985. Estimation of the generalized extreme value distribution by the method of probability weighted moments. Technometrics 27, 251–261.

Martins-Filho, C., Yao, F., 2006. Estimation of value-at-risk and expected shortfall based on nonlinear models of return dynamics and extreme value theory. Studies in Nonlinear Dynamics & Econometrics 10, Article 4.

Oja, H., 1981. On location, scale, skewness and kurtosis of univariate distributions. Scandinavian Journal of Statistics 8, 154–168.

Smith, R. L., 1984. Thresholds methods for sample extremes, 1st Edition. D. Reidel, Dordrecht.

Smith, R. L., 1987. Estimating tails of probability distributions. Annals of Statistics 15, 1174–1207.

Wang, L., Yang, L., 2007. Spline-backfitted kernel smoothing of nonlinear additive autoregression model. Annals of Statistics 35, 2474–2503.