Statistics and Its Interface Volume 1 (2008) 321–326

A nonparametric threshold model with application to zero returns

Oliver Linton*

We propose a nonparametric censoring model for time series data. We propose an estimator of the censoring function based on extreme value regression. We obtain the pointwise distribution theory and suggest confidence intervals based on this theory. We use our model to explain the evolution of the frequency of zeros in stock index returns.

Keywords and phrases: Censoring, Extreme value theory, GARCH, Index Returns.
JEL classification: C12, C13, C22, G11, G32.

1. INTRODUCTION

High frequency stock returns are well known to possess discreteness, meaning that their marginal distributions and conditional distributions can possess atoms. In particular, there are atoms at zero and other places. There are a number of economic reasons for this including the actual discreteness of prices (until quite recently US stock prices could only vary in multiples of 1/8th of a dollar) and nontrading. A variety of models have been proposed to take account of that including rounding and barrier models where some underlying continuous time and continuous state price process is censored. Early approaches to this are reviewed in Campbell, Lo, and MacKinlay (1997). To obtain tractable results quite simple processes have to be assumed for the latent price process. More recent work includes Delattre and Jacod (1997), Zheng (2003), and Li and Mykland (2006).

This phenomena is quite common for individual stocks but it also occurs in large index series, like the S&P500 as shown in Table 1 below. One possible explanation is that the prices of individual stocks are discrete. However, in an index that is the average of 500 stocks the effects of discrete prices should wash out, and in any case the discreteness is only at zero, there is no other value of returns that has positive mass. So individual level price discreteness does not seem to be a plausible explanation. Also, non trading of the component stocks does not seem relevant because zeros are found even in monthly data where there has clearly been a lot of trading of these large capitalization stocks.

*Thanks for financial support to the ESRC and the Leverhulme foundation. Thanks to Ilze Kalnina for research assistance.

We propose a flexible model for censoring of returns in discrete time. Our framework allows for quite general nonlinear dynamic processes and censoring function. We propose an estimator of the censoring function and some features of the model for the latent return. We obtain some pointwise asymptotic distribution theory, which can be used to justify inference procedures. We apply our methods to the S&P500 index data. We find that the pattern of censoring is consistent with the fact that the S&P500 index is only computed to two significant digits. There may be other reasons for this discreteness but we did not find convincing evidence for them.

What are the consequences of exact zeros in the data? It will typically lead to (downward) biased estimates of volatility. Furthermore, it can cause some problems when logarithmic transforms are used, and usually some artificial device is adopted to avoid this effect. In the dataset we consider the consequences of censoring for volatility modelling seem rather limited, and any effect has evidently declined over time. In other cases, robust methods for measuring volatility can be used to mitigate the effects of censoring, Peng and Yao (2003).

2. MODEL AND ESTIMATOR

The simplest model is that there is some process for latent returns Y_t^* but that this is censored at some fixed level δ so we only observe a censored version of returns, Y_t, where $Y_t = Y_t^* 1(|Y_t^*| > \delta)$. Thus we will observe some exact zeros along with positive and negative returns, which is consistent with the facts concerning index returns. Evidently, this will mean estimates of the parameters of a model for Y^* based on a sample of Y will be inconsistent. Given a sample of observations $\{Y_t, t = 1, \ldots, T\}$, we can estimate the quantity δ by

$$\widehat{\delta} = \min_{t:Y_t \neq 0} |Y_t| = \min_{t:|Y_t^*|>\delta} |Y_t^*| 1\left(|Y_t^*| > \delta\right). \qquad (1)$$

Suppose that Y_t^* is a strongly stationary mixing process whose marginal density has strictly positive density at δ, $\widehat{\delta}$ will converge to δ in probability at rate T. The limiting distribution should be exponential following standard extreme value theory, see for example Embrechts, Klüppelberg and Mikosch (1998).

However, with regard to the S&P500 data, the above model appears inadequate because the frequency of zeros appears to be declining over time, so that most of the zeros occur early in the time period. This suggests that the basic censoring model is not appropriate. We propose instead a more general model that can allow the frequency of observed zeros to change over time. We suppose that latent returns Y_t^* are as before but we only observe Y_t, where

$$Y_t = Y_t^* 1\left(|Y_t^*| > g(t/T)\right), \tag{2}$$

where $g(u)$, $u \in [0,1]$ is an unknown function. We shall suppose that this function is smooth but one can allow also a finite number of jumps. The objective is to estimate the function g given a sample of observations $\{Y_t, t = 1, \ldots, T\}$. One may also be interested in features of the model for latent returns.

We propose an estimator for the function g. Let h be some small number and let $\mathcal{N}_u = \{t : t/T \in [u-h, u+h]\}$ for all $u \in [h, 1-h]$. Then let

$$\widehat{g}(u) = \min_{t \in \mathcal{N}_u : Y_t \neq 0} |Y_t|. \tag{3}$$

This is a localized extreme value, see Chernozhukhov (1998).[1]

Provided Y_t^* is stationary and mixing and has marginal density that is strictly positive at $g(u)$ this estimate should converge to the true $g(u)$ at rate Th provided $h \to 0$ and the function g is smooth. We consider a more general data generating processes where

$$Y_t^* = \mu(t/T) + \sigma(t/T)\varepsilon_t \tag{4}$$

with ε_t a stationary mixing process with both ε_t and $\varepsilon_t^2 - 1$ martingale difference sequences. This model is consistent with the multiplicative components model of Engle and Rangel (2006) where $\varepsilon_t = v_t \eta_t$ and η_t is i.i.d., while v_t is a unit GARCH process. See also Dahlhaus (1997). Then it suffices for Th consistency of $\widehat{g}(u)$ that the marginal density of ε_t is strictly positive in some relevant (depending on $\mu(u), \sigma(u)$) neighborhood of zero. This assumption seems quite reasonable in the application; indeed stock index returns have very high density around zero.

3. DISTRIBUTION THEORY

Here we give the pointwise distribution theory of our estimator under specific assumptions. Let $N(T) = \sum_{t=1}^T 1\left(|Y_t| > 0\right)$ be the number of uncensored observations in a given sample and let

$$N_u(T,h) = \sum_{t=1}^T 1\left(|Y_t| > 0\right) 1\left(u - h \leq t/T \leq u + h\right)$$

be the number of uncensored observations in the time interval $[u-h, u+h]$. Let $F_{t/T}(y)$, $f_{t/T}(y)$ be the c.d.f., p.d.f. of $|Y_t|$ (in sample of size T) respectively, then for any $u \in [0,1]$

$$F_u(y) = F_\varepsilon\left(\frac{y - \mu(u)}{\sigma(u)}\right) - F_\varepsilon\left(\frac{-y - \mu(u)}{\sigma(u)}\right),$$

$$f_u(y) = \frac{1}{\sigma(u)}\left[f_\varepsilon\left(\frac{y - \mu(u)}{\sigma(u)}\right) + f_\varepsilon\left(\frac{-y - \mu(u)}{\sigma(u)}\right)\right],$$

where f_ε is the marginal density of ε_t, and let $\lambda_u(y) = f_u(y)/(1 - F_u(y))$. Let z_t^τ denote a random variable drawn from the truncated (from above) distribution of the random variable z_t at the point τ (so that $z_t^\tau \leq \tau$ with probability one).

Assumptions A.

1. *The process Y_t^* satisfies (4) where $\{\varepsilon_t\}$ is stationary; furthermore, ε_t and $\varepsilon_t^2 - 1$ are martingale difference sequences.*
2. *The marginal density of ε_t, f_ε, is finite, continuous, and strictly positive on a large enough neighborhood of the origin that includes $\tau_+(u) = (g(u) - \mu(u))/\sigma(u)$ and $\tau_-(u) = -(g(u) + \mu(u))/\sigma(u)$ for all $u \in (0,1)$.*
3. *The functions g, μ, σ^2 are twice continuously differentiable on $[0,1]$; $\inf_{u \in (0,1)} g(u), \sigma(u) > 0$.*
4. *For any integers p, q, T, $1 \leq i_1 < \cdots < i_p < j_1 < \cdots < j_q \leq T$, such that $j_1 - i_p \geq l$ we have for $z_t = \pm\varepsilon_t^\tau$ for any $\tau \in [\tau_-(u) - \epsilon, \tau_+(u) + \epsilon]$ with $\epsilon > 0$ and a sequence $u_T \to \tau$*

$$\left|\Pr\left(\max_{t \in A_1 \cup A_2} z_t \leq u_T\right) - \Pr\left(\max_{t \in A_1} z_t \leq u_T\right)\Pr\left(\max_{t \in A_2} z_t \leq u_T\right)\right| \leq \alpha_{T,l},$$

   where $A_1 = \{i_1, \ldots, i_p\}$, $A_2 = \{j_1, \ldots, j_q\}$ and $\alpha_{T,l} \to 0$ as $T \to \infty$ for some sequence $l = l(T)$.
5. *We have for $z_s = \pm\varepsilon_t^\tau$ for any $[\tau_-(u) - \epsilon, \tau_+(u) + \epsilon]$ with $\epsilon > 0$ and a sequence $u_T \to \tau$*

$$\lim_{k \to \infty} \limsup_{T \to \infty} T \sum_{s=2}^{[T/k]} \Pr\left(z_1 > u_T, z_s > u_T\right) = 0.$$

Theorem. *Suppose that assumptions A1–A5 hold and that $h = h(T) \to 0$ and $Th^2 \to \infty$ as $T \to \infty$. Then for all $u \in (0,1)$*

$$N_u(T,h)[\widehat{g}(u) - g(u)] \Longrightarrow \mathcal{E}_u,$$

where for any $z \geq 0$,

$$\Pr\left[\mathcal{E}_u \leq z\right] = 1 - \exp\left(-z\lambda_u(g(u))\right).$$

Furthermore, for any $u \neq u'$, $\widehat{g}(u), \widehat{g}(u')$ are asymptotically independent.

[1] A more general estimator (treated in Chernozhukhov (1998)) would be

$$\widehat{\theta} = \arg\min_\theta \sum_{t=1}^T K_h(u - t/T)\rho_{\alpha_T}(|Y_t| - \theta),$$

where ρ_α is the usual quantile check function and $\alpha_T \to 0$ as $T \to \infty$. Here, K is a kernel function with $K_h(.) = K(./h)/h$.

Suppose that $\widehat{f}_u(.)$ is a consistent estimator of $f_u(.)$. Then for $\alpha \in (0,1)$

$$\mathcal{C}_\alpha = \left[\widehat{g}(u) + \frac{\ln(1-\alpha)}{N_u(T,h)\widehat{\lambda}_u(\widehat{g}(u))}, \widehat{g}(u)\right] \tag{5}$$

is an asymptotic $1-\alpha$ confidence interval. We take $\widehat{\lambda}_u(y) = \widehat{f}_u(y)/(1-\widehat{F}_u(y))$ with

$$\widehat{F}_u(y) = \frac{1}{2Th}\sum_{t=1}^{T} 1\left(Y_t = 0\right) 1\left(u - h \le t/T \le u + h\right), \tag{6}$$

$$\widehat{f}_u(y) = \frac{1}{N_u(T,h)} \times \sum_{t=1}^{T} K_b(y - |Y_t|) 1\left(|Y_t| > 0\right) 1\left(u - h \le t/T \le u + h\right), \tag{7}$$

where $K_b(y) = K(y/b)/b$ for some one-sided kernel function K and positive number b. We need to use a one-sided kernel, that is $K : [0,1] \to \mathbb{R}$ with $\int_0^1 K(u)du = 1$ and $\int_0^1 K(u)udu = 0$,[2] because locally we only have observations on $|Y_t|$ greater than $\widehat{g}(u)$. Under some conditions we expect that $\widehat{f}_u(y)$ consistently estimates $f_u(y)$ for $y \ge g(u)$. See Chen (1999) for some discussion of similar estimation problems.

What are the additional consequences of censoring? It will typically lead to biased estimates of volatility. For simplicity suppose that Y_t^* is symmetric about zero, then

$$\begin{aligned}\operatorname{var}(Y_t) &= E(Y_t^2) = E[Y_t^{*2} 1\left(|Y_t^*| > \delta\right)] \\ &= \operatorname{var}(Y_t^*) - E[Y_t^{*2} 1\left(|Y_t^*| \le \delta\right)] < \operatorname{var}(Y_t^*),\end{aligned}$$

so that observed unconditional volatility is downward biased. However, note that

$$E[Y_t^{*2} 1\left(|Y_t^*| \le \delta\right)] \le \delta^2 \Pr[|Y_t^*| \le \delta]$$

so the magnitude of the downward bias on volatility is limited by $\delta^2 \Pr[Y_t = 0]$. In some cases this can be small.

If one could consistently estimate the functions $\mu(.), \sigma(.)$ everywhere, then one could estimate $f_u(y)$ better but this does not seem to be generally possible in the presence of censoring. One can certainly estimate $\mu(.), \sigma(.)$ up to constants at points of little censoring using robust methods. For example, we can estimate $\sigma(\cdot)$ by taking the local interquartile range of Y_t divided by 1.31, that is

$$\widehat{\sigma}(u) = \frac{1}{1.31}\left[q_{0.75}(Y_t : t \in \mathcal{N}_u) - q_{0.25}(Y_t : t \in \mathcal{N}_u)\right]. \tag{8}$$

This estimator is robust to modest amounts of censoring and consistently estimates $\sigma(u)$ under normality. The short run dynamics of the volatility process are harder to estimate.

[2] For example $K(u) = 3 \times 1(u < 1/2) - 1(u \ge 1/2)$.

Table 1. Descriptive statistics by frequency

	Daily	Weekly	Monthly
Mean	0.000293	0.001391	0.00606
St. Deviation	0.00901	0.01999	0.04203
Skewness	-1.546	-0.375	-0.589
Excess Kurtosis	43.334	6.521	5.588
Minimum	-25.422	-6.577	-5.984
Maximum	9.623	6.534	3.450
Sample Size	11893	2475	568
Number zeros	81	15	3

Note: Descriptive statistics for the returns on the S&P500 index for the period 1955–2002 for three different data frequencies. Minimum and maximum are measured in standard deviations and from the mean.

4. APPLICATION

We apply our methodology to S&P500 index returns. This index is one of the main cited indexes on the NYSE. We first present the data we use and their descriptive statistics.

For the daily data the unconditional estimates are $\widehat{\delta} = \min_{t:Y_t \neq 0} |Y_t| = 9.436E-06$ (which is roughly $0.001s$) and $T^{-1}\sum_{t=1}^{T} 1(Y_t = 0) = 0.00681$, so that the effect of censoring on unconditional variance (for the daily data the sample variance is $8.12E-05$) is limited. It is possible though that this could have a bigger effect on estimation of conditional volatility although it is hard to believe that the consequences are that great, although it will make the usual estimators inconsistent.[3]

As mentioned in the introduction, the frequency of zeros in this dataset appears to decrease over time. This is evidenced in the following figure which displays the "c.d.f." of occurrence of zeros for the daily data plotted against time, i.e., $n^{-1}\sum_{i=1}^{n} 1(t_i \le t)$, where t_i is the time of the $i'th$ zero, with $i = 1, \ldots, n$. In the same curve is shown the uniform c.d.f., which would correspond to equally distributed zeros.

Over this time period, the tick size of individual stocks have decreased from one eighth of a dollar to one cent. Although this is relevant for discreteness of individual stocks, the properties of a large index like the S&P500 are likely to be little affected by this individual discreteness.

We now turn to our estimates for the model (2). Figure 1 below shows the estimated g function for the daily data computed with $Th = 500$ so that a total of 1000 observations are used to compute $\widehat{g}(u)$. The estimator shows a pronounced downward trend. Similar results are obtained for the weekly and monthly series.

The confidence intervals (5) were computed, but were rather narrow especially in the earlier period and so are not shown for clarity.

We now propose a simple obvious source for this effect and show that it is corroborated by the data. The S&P500

[3] Lee (1999) proposes a simulation based method to estimate a censored GARCH model by maximum likelihood under a distributional assumption on the error terms.

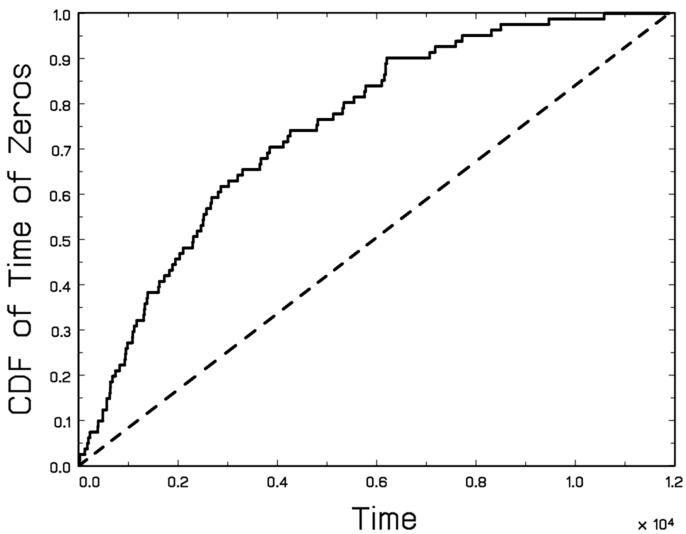

Figure 1.

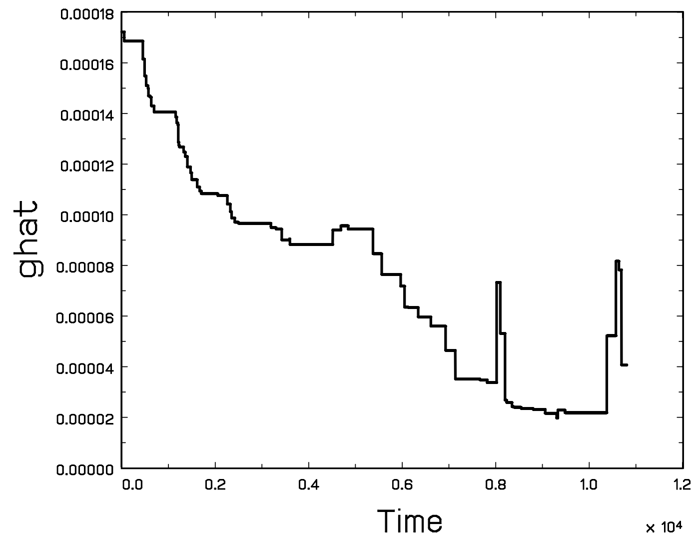

Figure 2.

index is only reported to two significant digits! Suppose that in fact the censoring rule arises from rounding of the index level, that is,

$$(9)\quad P_t = P_t^* 1(|P_t^* - P_{t-1}^*| > \delta) + P_{t-1}^* 1(|P_t^* - P_{t-1}^*| \leq \delta)$$

for some fixed δ, e.g., $\delta = 0.005$. Then the event that

$$|P_t^* - P_{t-1}^*| > \delta$$

is equivalent to the event that

$$|Y_t^*| \equiv \left| \frac{P_t^* - P_{t-1}^*}{P_{t-1}^*} \right| > \frac{\delta}{P_{t-1}^*}.$$

Now suppose that $P_t^* = P_{t-1}^* \exp(\alpha)$, then by continued substitution $P_{t-1}^* = P_0^* \exp(t\alpha)$. This would imply our censoring model with $g(t) = (\delta/P_0^*)\exp(-t\alpha)$.[4] This suggests that a regression of $\ln g$ on a constant and time should yield estimates of α, the logarithmic return. In fact, we find

$$\ln \widehat{g}(t/T) = -8.649 - 0.000188 * t$$

with an R^2 of 0.824. The t-statistic on the slope coefficient is over 200. Although we do not provide formal inference here, it seems likely that this is statistically significant at any conventional level. The slope parameter corresponds to an annualized return of 4.7%, which is a bit on the low side but in the right ballpark. Figure 2 shows the graph of $\ln \widehat{g}$ against time showing the expected linearity.

This result can be replicated for weekly ($\widehat{\alpha} = 0.000644$) and monthly data ($\widehat{\alpha} = 0.005571$) although the standard errors are larger and the R^2 lower (0.575 and 0.511). One can also bound the consequence of censoring on unconditional volatility estimation and find that it is a small effect.

Finally, we estimate the volatility function in model (4). Specifically, we estimate $\sigma(\cdot)$ by (8). We also computed the local standard deviation for comparison purposes. The local standard deviation is always above the local IQR even though under conditional normality and without censoring they should estimate the same thing. However, the censoring would be expected to downward bias the standard deviation so the main source of difference is the non-normality of the error.

If one wanted to estimate the parameters of a dynamic GARCH process say v_t then one would have to adopt some strategy like Lee (1999), which can be very time consuming and requires distributional assumptions for consistency.

5. CONCLUSION

We have proposed a nonparametric threshold model and developed an estimator for the threshold function. One can also allow the censoring function g to depend on covariates but we do not do that here. One can also allow different lower and upper censoring functions so that $Y_t = Y_t^* 1\,(Y_t^* \notin [g_L(t/T), g_U(t/T)])$ but we do not do that here. We applied our method to stock index returns. The empirical conclusion is that the source of zeros in the S&P500 index is due to the reporting of only two significant digits and that as the level of the index has risen the frequency of zeros has reduced. Furthermore, for many purposes the magnitudes of the biases caused by the censoring is small

[4]The assumption that latent prices grow deterministically is obviously unrealistic. If the latent prices were stochastic, this makes the calculations much more complicated. For example, suppose that $\ln(P_t^*) = \mu + \ln(P_{t-1}^*) + \eta_t$ with η_t a martingale difference sequence. In this case one does not obtain exactly model (2). But perhaps one can still expect something like what we obtain. However, since prices are (globally) nonstationary the statistical analysis of this model is more complicated.

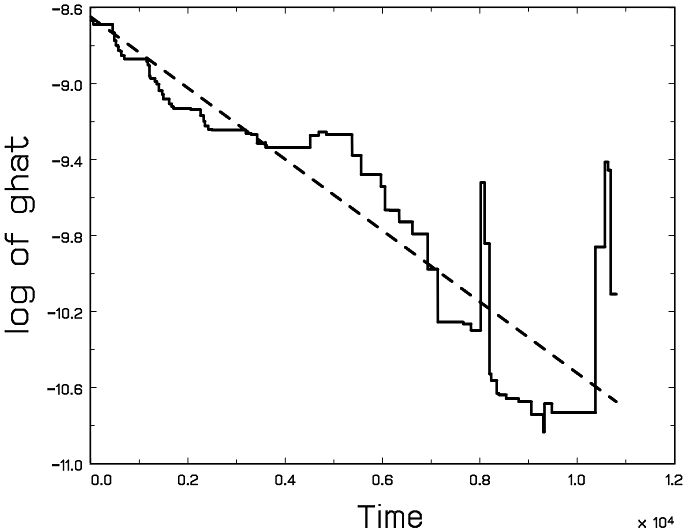

Figure 3.

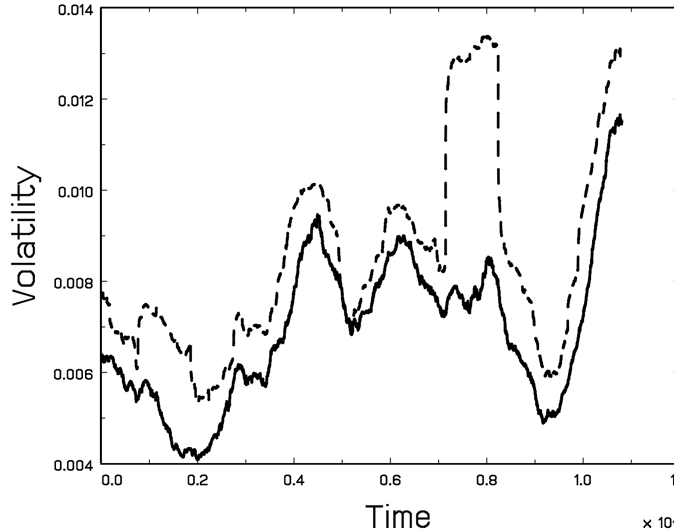

Figure 4. The estimated local IQR in solid line and the local standard deviation in dashed line.

enough to be ignored. However, for other datasets different conclusions might be reached.

APPENDIX A

Proof of Theorem. First note that the probability of censoring is

$$\Pr[Y_t = 0] = F_\varepsilon\left(\frac{g(t/T)-\mu(t/T)}{\sigma(t/T)}\right) - F_\varepsilon\left(\frac{-g(t/T)-\mu(t/T)}{\sigma(t/T)}\right).$$

Let $\widetilde{X}_t$ be the random variable with the truncated distribution of $-|Y_t|$. The estimator $\widehat{g}(u)$ can be rewritten as

$$\max_{t\in\mathcal{N}_u^+} \widetilde{X}_t, \tag{10}$$

where $\mathcal{N}_u^+ = \{t \in \mathcal{N}_u : Y_t \neq 0\}$ with (random) cardinality $N_u(T,h)$. Then for any $x \in (-\infty, -g(t/T)]$

$$\begin{aligned}\Pr\left(\widetilde{X}_t \leq x\right) &= \frac{\Pr(-|Y_t| \leq x)}{1-\Pr(Y_t = 0)} \\ &= \frac{1 - F_\varepsilon\left(\frac{-x-\mu(t/T)}{\sigma(t/T)}\right) + F_\varepsilon\left(\frac{x-\mu(t/T)}{\sigma(t/T)}\right)}{1 - F_\varepsilon\left(\frac{g(t/T)-\mu(t/T)}{\sigma(t/T)}\right) + F_\varepsilon\left(\frac{-g(t/T)-\mu(t/T)}{\sigma(t/T)}\right)} \\ &= F_{t/T}^X(x).\end{aligned}$$

This can be approximated by

$$F_u^X(x) = \frac{1 - F_\varepsilon\left(\frac{-x-\mu(u)}{\sigma(u)}\right) + F_\varepsilon\left(\frac{x-\mu(u)}{\sigma(u)}\right)}{1 - F_\varepsilon\left(\frac{g(u)-\mu(u)}{\sigma(u)}\right) + F_\varepsilon\left(\frac{-g(u)-\mu(u)}{\sigma(u)}\right)}$$

in the region $\mathcal{N}_u = [u-h, u+h]$ in the sense that for some $C < \infty$,

$$\sup_{t\in\mathcal{N}_u}\ \sup_{-\infty<x\leq 0} \left|F_{t/T}^X(x) - F_u^X(x)\right| \leq Ch, \tag{11}$$

by Assumptions A2 and A3.

Consider the problem of having N i.i.d. observations from $\widetilde{X}_t^u \sim F_u^X$, that is, $Y_t^{*u} = \mu(u) + \sigma(u)\varepsilon_t'$, where ε_t' is i.i.d. with the same marginal distribution as ε_t, $Y_t^u = Y_t^{*u}1(|Y_t^{*u}| > g(u))$, and $\widetilde{X}_t$ be the random variable with the truncated distribution of $-|Y_t^u|$. Note that $F_u^X(x) \to 1$ as $x \to -g(u)$. In particular,

$$\begin{aligned}F_u^X(x) = 1 &+ \frac{(x+g(u))}{\sigma(u)} \\ &\times \frac{f_\varepsilon\left(\frac{g(u)-\mu(u)}{\sigma(u)}\right) + f_\varepsilon\left(\frac{-g(u)-\mu(u)}{\sigma(u)}\right)}{1 - F_\varepsilon\left(\frac{g(u)-\mu(u)}{\sigma(u)}\right) + F_\varepsilon\left(\frac{-g(u)-\mu(u)}{\sigma(u)}\right)} \\ &+ o(|x+g(u)|).\end{aligned}$$

Suppose that for some $\tau \in (0,\infty)$

$$N\overline{F}_u^X(x_N) \to \tau,$$

where $\overline{F}_u^X(x) = 1 - F_u^X(x)$. Then,

$$\begin{aligned}\Pr\left(\max_{1\leq t\leq N} \widetilde{X}_t^u \leq x_N\right) &= \left\{F_u^X(x_N)\right\}^N = \left\{1 - \overline{F}_u^X(x_N)\right\}^N \\ &= \left\{1 - \frac{\tau}{N} + o(1/N)\right\}^N \to \exp(-\tau).\end{aligned}$$

Specifically, letting $x_N = -g(u) + x/N$ we have

$$\begin{aligned} N\overline{F}_u^X(x_N) &= N\frac{(x_N + g(u))}{\sigma(u)} \\ &\quad \times \frac{f_\varepsilon\left(\frac{g(u)-\mu(u)}{\sigma(u)}\right) + f_\varepsilon\left(\frac{-g(u)-\mu(u)}{\sigma(u)}\right)}{1 - F_\varepsilon\left(\frac{g(u)-\mu(u)}{\sigma(u)}\right) + F_\varepsilon\left(\frac{-g(u)-\mu(u)}{\sigma(u)}\right)} \\ &\to \frac{x}{\sigma(u)} \frac{f_\varepsilon\left(\frac{g(u)-\mu(u)}{\sigma(u)}\right) + f_\varepsilon\left(\frac{-g(u)-\mu(u)}{\sigma(u)}\right)}{1 - F_\varepsilon\left(\frac{g(u)-\mu(u)}{\sigma(u)}\right) + F_\varepsilon\left(\frac{-g(u)-\mu(u)}{\sigma(u)}\right)} \\ &= x\lambda_u(g(u)) \equiv \tau. \end{aligned}$$

Now suppose that we have a random sample of size N from F_u^X with

$$\begin{aligned} N/2Th \xrightarrow{p} 1 - F_\varepsilon\left(\frac{g(u)-\mu(u)}{\sigma(u)}\right) \\ + F_\varepsilon\left(\frac{-g(u)-\mu(u)}{\sigma(u)}\right) \in (0,1). \end{aligned} \tag{12}$$

By the arguments of Barakat and El-Shandidy (1990), we have

$$\Pr\left(N_u(T,h)\left[\max_{t\in\mathcal{N}_u^+} \widetilde{X}_t^u - g(u)\right] \le x\right) \to \exp(-x\lambda_u(g(u)))$$

$$\Pr\left(2Th\left[\max_{t\in\mathcal{N}_u^+} \widetilde{X}_t^u - g(u)\right] \le x\right) \to \exp(-xf_u(g(u))).$$

We now consider the estimator itself drawn from a design $Y_t^* = \mu(t/T) + \sigma(t/T)\varepsilon_t'$, where ε_t' is i.i.d. with the same marginal distribution as ε_t. In that case we have under the i.i.d. error assumption

$$\begin{aligned} &\Pr\left(\max_{t\in\mathcal{N}_u} \widetilde{X}_t \le x_{2Th}\right) \\ &\quad = \left\{\prod_{t\in\mathcal{N}_u} F_{t/T}^X(x_{2Th})\right\} = \left\{\prod_{t\in\mathcal{N}_u}\left(1 - \overline{F}_{t/T}^X(x_{2Th})\right)\right\} \\ &\quad = \left\{1 - \frac{\tau}{2Th} + O(h) + o(1/Th)\right\}^{2Th} \to \exp(-\tau), \end{aligned}$$

provided $h = o(1/Th)$, which is implied by $Th^2 \to 0$. Likewise, when the sample size is random

$$\Pr\left(N_u(T,h)\left[\max_{t\in\mathcal{N}_u^+} \widetilde{X}_t - g(u)\right] \le x\right) \to \exp(-x\lambda_u(g(u)))$$

$$\Pr\left(2Th\left[\max_{t\in\mathcal{N}_u^+} \widetilde{X}_t - g(u)\right] \le x\right) \to \exp(-xf_u(g(u))).$$

Finally, we consider the case where the data is drawn from (4) where ε_t is stationary and mixing and satisfies conditions A1, A4 and A5. We can apply Proposition 4.4.3 of Embrechts, Klüppelberg and Mikosch (1998) to conclude that the asymptotic properties continue to hold. The asymptotic independence of $\widehat{g}(u)$ and $\widehat{g}(u')$ follows from the weak dependence assumptions and the assumption that $h \to 0$. $\square$

REFERENCES

Barakat, H. M. and M. A. El-Shandidy (1990). On the limit distribution of the extremes of a random number of independent random variables. *Journal of Statistical Planning and Inference* **26** 353–361. MR1086106

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton.

Chen, S. X. (1999). Beta kernel estimators for density estimation. *Computational Statistics and Data Analysis* **31** 131–145. MR1718494

Chernozhukov, V. (1998). Nonparametric Extreme Regression Quantiles. Unpublished manuscript, Stanford.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes, *Annals of Statistics* **25** 1–37. MR1429916

Dahlhaus, R., and S. Subba Rao (2006). Statistical inference for time-varying ARCH processes, *The Annals of Statistics* **34** 1075–1114. MR2278352

Delattre, S. and Jacod, J. (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors, *Bernoulli* **3** 1–28. MR1466543

Embrechts, P., C. Klüppellberg and T. Mikosch (1998). *Modelling Extremal Events for Insurance and Finance*, Springer, New York. MR1458613

Engle, R. F. and Rangel (2006). The spline GARCH model for unconditional volatility and its global macroeconomic causes, Mimeo, NYU.

Hill, J. B. (2006). On tail index estimation for dependent, heterogenous data, *Working Paper*, http://www.unc.edu/~jbhill/.

Lee, L-F. (1999). Estimation of dynamic and ARCH Tobit models. *Journal of Econometrics* **92** 355–390.

Li, Y., and P. A. Mykland (2006). Determining the Volatility of a Price Process in the presence of rounding errors. Working Paper, Department of Statistics, the University of Chicago.

Peng, L. and Yao, Q. (2003). Least absolute deviation estimation for ARCH and GARCH models. *Biometrika* **90** 967–975. MR2024770

Zeng, Y. (2003). A partially-observed model for micromovement of asset process with Bayes estimation via filtering. Math. Finance 13 411–444 available on: http://www.ams.org/mathscinet-getitem?mr=1995285. MR1995285

Oliver Linton
Department of Economics
London School of Economics
Houghton Street, London WC2A 2AE
United Kingdom
E-mail address: o.linton@lse.ac.uk