Application of Nonparametric Methods in Studying Energy Consumption

Levi Mbugua, Peter Mwita, Samuel Mwalili
Jomo Kenyatta University of Agriculture and Technology
Department of Statistics and Actuarial Sciences
P.O. Box 62000-00200, Nairobi. Kenya
Email: **mbugualevi@yahoo.co.uk** Telephone: (+254) 0721666777

Abstract

Consumer behaviour towards different forms of energy varies over time. The variance can be so large that the quality of the estimation functional relationship between the response variable and its associated explanatory variables is seriously affected. To attenuate this, kernel smoothing a nonparametric regression approach is proposed. This approach offers a powerful tool in modelling and adapts to various types of designs. The aim of this study is to produce a reasonable model that defines the structural change of a stationary time series which exhibits volatility over time. The explanatory variable used is the lagged values of the series. To study the effects at the tails, the quantiles are proposed. This model is functional in examining the characteristics of peak hour electricity consumption in Kenya. It is found that the mean peak consumption is a decreasing function of the lagged time and that the more extreme the peak consumption, the higher the volatility. This model provides insights on routine shift time energy consumption modelling.

Key words and phrases: *conditional quantiles, electricity consumption, kernel estimator, nonparametric methods*

1. Introduction

The statistical properties of regression smoothers have been mainly analyzed in the frame-work of an independent and identically distributed observation structure. The assumption that the pairs $\{(X_i, Y_i), i = 1, \dots, n\}$ are an independent sample from an unknown distribution can often be justified in practice and simplifies technical matters. However, there are many practical situations in which it is not appropriate to assume that the observations $\{X_i, Y_i\}_{i=1}^{n}$ or the errors $Y_i - m(X_i)$ are independent, it is very likely that the objects response will depend on its previous response.

When data means and variance are non-constant, typically skewed or contains some outliers, it is understood that the observations come from different distributions over time. Standard asymptotic distribution theory often does not apply to regression involving variables of this nature since inferences may be misleading. In such data, median regression a special case of quantile regression is more explicable and robust than the mean regression. More especially when the data pattern shows heteroscedasticity and asymmetries.

We discuss our methodologies and theory under the framework of nonparametric time series setting. We concentrate on the scenario, where there is a stationary sequence of random variables $\{(X_i, Y_i), i \geq 1, X \in \Re d, Y \in \Re$ observed. The observations may be dependent via the time index $i = 1,2, \ldots$ and it is desired to estimate a functional of the conditional distribution $L(Y|X)$ like the mean function or the median function $m(x)$. That is,

$$m(x) = E(Y|X = x) \qquad [1]$$

The second scenario is of a nonlinear autoregressive time series

$$Y_t = m(Y_{t-1}, \ldots, Y_{t-d}) + e_t, \quad t = 1,2, \ldots \qquad [2]$$

with independent innovation shocks $e_t = s_t . \xi_t$. One is interested in predicting new observations and in estimating the nonparametric autoregressive function m or the conditional variance function

$$v_t = s_t^2 = Var(\varepsilon_t | (Y_{t-1}, \ldots, Y_{t-d}) = x) \qquad [3]$$

2. Nonparametric Regression for Time Series

Let us assume a more general time series dependence, which is commonly used in the literature, described as follows.

2.1 Stationarity

A process is said to be strictly stationary if the statistical behaviour of $X_1, X_2, \ldots, X_k$ is identical to that of the shifted set $X_{t+1}, X_{t+2}, \ldots, X_{t+k}$ evaluated at the same set of points $x_1, x_2, \ldots, x_k$, for all t and for all k. A process is said to be weak stationary if $E(X_t) = \mu$ and $V(X_t) = \sigma^2$ and $Cov(X_t, X_{t+k}) = Cov(X_{t+1}, X_{t+1+k}) = Cov(X_{t+2}, X_{t+2+k}) = \cdots$ is a function of the time lag k only and does not depend on time t.

Let $\{X_t\}$ be a strictly stationary time series for $n \geq 1$. The stationary process is called strongly mixing (Rosenblatt, 1956b) if

$$\sup_{A \epsilon \mathcal{F}_1^n, B \epsilon \mathcal{F}_{n+k}^\infty} |P(A \cap B) - P(A)P(B)| \leq \alpha_k \qquad [4]$$

Where $\alpha_k \to 0$ and $\mathcal{F}_i^j$ is the α-field generated by $x_i, \dots, x_j$. The random variable X here may also stand for the pair (X, Y) so that the σ- fields are defined appropriately for the regression problem. Mixing dependence is commonly used to characterize the dependent structure and it is often referred to as short range dependence or weak dependence. This means that as the distance between two observations go farther and farther, the dependence becomes weaker and weaker very faster. The short term dependence does not have much effect on the local smoothing method since for any two given random variables X_i and X_j and a point x, the random variables $K_h(X_i - x)$ and $K_h(X_j - x)$ are nearly uncorrelated as $h \to 0$. $K_h(\cdot) = K(\cdot/h)/h$. K, being a kernel function assigning weights to each datum point and is supported on $[-1,1]$. K, determines the shape of the weights and satisfies the moment conditions, $\int_{-1}^{1} K(u)du = 1$, $\int_{-1}^{1} uK(u)du = 0$, $\int_{-1}^{1} u^2K(u)du \neq 0$ and $\int_{-1}^{1} K(u)^2 du < \infty$. The parameter h is the smoothing parameter which determines the size of the weights.

The dependent random variables can be approximated by a sequence of independent random variables having the same marginal distribution. This can be seen by taking $V_j = exp(it_j X_j)$ in the following lemma by Volkonskii and Rozanov (1959):

Lemma 1 Let $V_1, \dots, V_L$ be random variables measurable with respect to the σ-algebras $\mathcal{F}_{i1}^{ji}, \dots, \mathcal{F}_{iL}^{jl}$ respectively with $i_{l+1} - j_l \geq w \geq 1$ and $|V_j| \leq 1$ for $j = 1, \dots, L$. then

$$\left| E\prod_{j=1}^{L} V_j - \prod_{j=1}^{L} E(V_j) \right| \leq 16(L-1)\alpha(w)$$

This lemma becomes a statement about the characteristic function of the random variables.

2.2 Local Polynomial Fitting

Consider observations $(X_1, Y_1), \dots, (X_n, Y_n)$ that can be thought as a realization from a stationary process. Estimating $m(x)$ in [1] and its derivatives $m^{(j)}(x)$, can be done by fitting locally a polynomial by a weighted least squares regression problem. That is minimize

$$\sum_{i=1}^{n}\left\{Y_i - \sum_{j=0}^{p}\beta_j (X_i - x_0)^j\right\}^2 K_h(X_i - x_0) \qquad [5]$$

Under certain mixing conditions for local polynomial estimators lemma 1 holds. Let $f(x)$ be the density of X_1 and $\sigma^2(x) = Var(Y_1|X_1 = x)$. Let S, S^* and cp denote some Moment matrices and vector as in Fan and Gijbels (1996) then we have the following results proved by Masry and Fan (1993)

Theorem 1: If $h_n = O\left(n^{1/(2p+3)}\right)$, then as $n \to \infty$,

$$\sqrt{n}h_n\left[diag\left(1,\ldots,h_n^p\right)\left\{\hat{\beta}(x) - \beta(x)\right\} - \frac{h_n^{p+1}m^{(p+1)}(x)}{(p+1)!}S^{-1}cp\right] \xrightarrow{D} N\{0, \sigma^2(x)S^{-1}S^*S^{-1}/f(x)\} [6]$$

at x, a continuity point of $\sigma^2 f$, whenever $f(x) > 0$. An immediate consequence of theorem (1) is that derivative estimator $\widehat{m}_v(x)$ based on the local polynomial fitting is asymptotically normal;

$$\sqrt{nh_n^{2v+1}}\left\{\widehat{m}_v(x) - m^{(v)}(x)\int t^{p+1}K_v^* dt\,\frac{v!m^{(p+1)}(x)}{(p+1)!}h_n^{p+1-v}\right\} \xrightarrow{D} N\left\{0, \frac{(v!)^2\sigma^2(x)\int K_v^{*2}(t)dt}{f(x)}\right\} [7]$$

where K_v^* is the equivalent kernel. When $v = 0$, [7] gives the asymptotic normality of $\widehat{m}(x)$

2.3 Asymptotic Properties

We consider a simple case where $p = 1$ and the Nadaraya Watson estimate, then;

$$\widehat{m}_{nw}(x) = \underbrace{\frac{1}{n}\sum_{t=1}^{n} m(X_t)K_h(X_t - x)/f_n(x)}_{I_1} + \underbrace{\sum_{t=1}^{n} W_t\varepsilon_t}_{I_2} \qquad [8]$$

I_1, contributes only to the bias and I_2 gives the asymptotic normality. First we derive the asymptotic bias for the interior boundary points. By the Taylors expansion, when X_t is in$(x - h, x + h$, we have

$$m(X_t) = m(x) + m'(x)(X_t - x) + \frac{1}{2}m''(x_t)(X_t - x)^2$$

where, $x_t = x + \theta(X_t - x)$ with $-1 < \theta < 1$. $I_1 = \frac{1}{n}\sum_{t=1}^{n} m(X_t)K_h(X_t - x)$ is regarded as the asymptotic bias, denoted by $B_{nw}(\cdot)$. If $p > 1$ (multivariate case), $B_{nw}(x)$ becomes

$$B_{nw}(x) = \frac{h^2}{2} tr\left[\mu_2(K)\left\{m''(x) + 2f'(x)m'(x)^T/f(x)\right\}\right] \qquad [9]$$

where $\mu_2(K) = \int uu^T K(u)du$. Under some regularity conditions it can be shown that for x being an interior grid point,

$$nh^p Var\,(I_2) \to V_0(K)\sigma_\varepsilon^2(x)/f(x) = \sigma_\varepsilon^2(x) \qquad [10]$$

where $\sigma_\varepsilon^2(x) = Var(\varepsilon_t|X_t = x)$. Further we can establish the asymptotic normality

$$\sqrt{nh^p}\left[\widehat{m}_{nw}(x) - m(x) - B_{nw}(x) + o_p(h^2)\right] \to N\{0, \sigma_m^2(x)\} \qquad [11]$$

2.4 Model with exogenous variable

Let $\{Y_t, X_t, Z_t\}_{t=-\infty}^{\infty}$ be jointly stationary process, where X_t and Z_t take values in $\Re^p$ and $\Re^q$ with $p, q \geq 0$ respectively. The regression surface is defined by

$$m(x, z) = E\{Y_t|X_t = x, Z_t = z\} \qquad [12]$$

Here Y_t is measurable on the real line and it is assumed that $E|Y_t| < \infty$. The regression function $m(\cdot,\cdot)$ defined in [12] can be decomposed as the sum

$$m(x, z) = \mu + g_1(x) + g_2(z) \qquad [13]$$

Such decomposition hold for nonlinear additive autoregressive model with exogenous variables, such that

$$Z_t = \mu + g_1\left(X_{t-j1}, \dots, X_{t-jp}\right) + g_2\left(Z_{t-i1}, \dots, Z_{t-iq}\right) + \eta_t$$

$$X_{t-ji} = g_3\left(X_{t-j2}, \dots, X_{t-jp}\right) + \varepsilon_t$$

If we assume that $E\{g_1(X_t)\} = 0$ and $E\{g_2(Z_t)\} = 0$, then the projection of $m(x, z)$ on the $g_1(x)$ direction is defined by

$$E\{m(x, Z_t)\} = \mu + g_1(x) + E\{g_2(Z_t)\} = \mu + g_1(x) \qquad [14]$$

To get the estimate of $g_2(\cdot)$, we use a small bandwidth h_s so that the bias can be asymptotically negligible. Let the additional components have continuous second partial derivatives so that $m(u, v)$ can be locally approximated by a linear term in a neighbourhood of (x, z) namely, $m(u, v) \approx \beta_0 + \beta_1^T(u - x) + \beta_2^T(v - z)$, with $\{\beta_j\}$ depending on x and z, where β_1^T denotes the transpose of β_1. Let $K_0(\cdot)$ and $K_1(\cdot)$ be symmetrical kernel functions in $\Re^p$ and $\Re^q$ and $h_0(n) > 0$ and $h_s = h_s(n) > 0$ be the bandwidths. Let β_j be the minimiser of the following locally weighted least squares

$$\sum_{t=1}^{n}\{Y_t-\beta_0-\beta_1^T(X_t-x)-\beta_2^T(Z_t-z)\}^2K_{h_0}(X_t-x)K_{h_s}(Z_t-z)$$

where $K_{h_0}(\cdot)=K(\cdot\,|h)/h^p$ and $K_{h_1}(\cdot)=K(\cdot\,|h)/h^q$ then the local linear estimator of the regression surface $m(x,z)$ is $\widehat{m}(x,z)=\hat{\beta}_0$. Using [14] and computing the sample average of $\widehat{m}(\cdot,\cdot)$ the estimators of $g_1(\cdot)$ and $g_2(\cdot)$ becomes

$$\hat{g}_1(x)=\frac{1}{n}\sum_{t=1}^{n}\widehat{m}(x,Z_t)-\hat{\mu}\quad and\ \ \hat{g}_2(z)=\frac{1}{n}\sum_{t=1}^{n}\widehat{m}(X_t,z)-\hat{\mu}$$

where $\hat{\mu}=n^{-1}\sum_{t=1}^{n}Y_t$. Then using the partial residuals $Y_t^*=Y_t-\hat{\mu}-\hat{g}_2(Z_t)$ we apply the local linear regression technique to the regression model $Y_t^*=g_1(X_t)+\varepsilon_t^*$. To estimate $g_1(\cdot)$ we solve the weighted least squares problem

$$\sum_{t=1}^{n}\{Y_t^*-\beta_1-\beta_2^T(X_t-x)\}^2J_{h_2}(X_t-x)\qquad[15]$$

$J(\cdot)$, is the kernel function in $\Re^p$ and $h_2=h_2(n)>0$ is the bandwidth at the second stage. Maximising [15] with respect to β_1 and β_2 gives the estimate of $g_1(x)$ denoted by $\widetilde{g}_1(x)=\hat{\beta}_1$.

3. Nonparametric Quantile Regression

Let us assume that $\{Y_t,X_t\}_{t=-\infty}^{\infty}$ is a stationary sequence as described in section 2. Denote, $F(Y|X)$, the conditional distribution of Y given $X=x$ where $X_t=(X_{t1},\dots,X_{td})'$ is the associated covariate vector in $\Re^d$ with $d\geq 1$ and might be a function of exogenous (covariate) variables or some lagged (endogenous) variables or a function of time t. Let $X_{(1)}\leq X_{(2)}\leq\cdots\leq X_{(n)}$ denote the order statistics of $\{X_t\}_{t=1}^{n}$. Define the inverse of $F(x)$ as

$$F^{-1}(\theta)=inf\{x\epsilon\Re;F(x)\geq\theta\}\qquad[16]$$

where $\Re$ is the real line. Our motivation is a convenient method of detecting conditional heteroscedasticity. So we define the quantile estimate as

$$q_\theta(x)=inf\{y\in\Re:F(y|x)\}=F^{-1}(\theta|x)\qquad[17]$$

According to Koenker and Bassett (1978), any $\theta-th$ quantile of a scalar random variable Y can be viewed as a solution to the problem

$$argmin_{a\in\Re}E\{p_\theta(Y_t-a)|X_t=x\}=q_\theta(x)\qquad[18]$$

To this end we assume that Y_t is related to X_t through the model

$$Y_t = \mu_\theta(X_t) + \sigma_\theta(X_t)Z_t \quad [19]$$

where $\mu_\theta(\cdot)$ is the mean function, $\sigma_\theta(\cdot)$ is the variance function and where X_t and Z_t are independent. The conditional quantile of Y_t given X_t is therefore

$$q_\theta(X_t) = \mu_\theta(X_t) + \sigma_\theta(X_t)F_{\varepsilon_t}^{-1}(\theta) \quad [20]$$

where $F_{\varepsilon_t}(\cdot)$ is the distribution of ε_t.

The kernel estimate of the mean function $\mu_\theta(X_t)$ at point x based on a sample (Y_t, X_t), $t = 1, \ldots, n$ from model [19] is obtained by estimating the conditional distribution function

$$F_X(y) = P(Y_t \leq y | X_t = x) = E\left(I_{\{Y_t \leq y\}} \middle| X_t = x\right) \quad [21]$$

where the conditional expectation of $I_{\{Y_t \leq y\}}$ may be estimated by the weighted version of Nadaraya-Watson kernel estimate of section 2.3. The weighted version is design adaptive and has high minimax efficiency (Cai, 2002). Thus we have

$$\hat{F}_X(y) = \frac{p_t(x)K_h(x-X_t)I_{\{Y_t \leq y\}}}{\sum_{t=1}^{n} p_t(x)K_h(x-X_t)} \quad [22]$$

Here $p_t(x)$ is the weight function at point x chosen such that $\sum_{t=1}^{n} log(p_t(x))$ is minimized subject to the constraints

$$p_t(x) > 0,\ \sum_{t=1}^{n} p_t(x) = 1 \ \ and \ \ \sum_{t=1}^{n} p_t(x)(x - X_t)K_h(x - X_t) = 0. \quad [23]$$

Through the Lagrange multiplier rule, then

$$p_t(x) = n^{-1}\{1 + \lambda(x - X_t)K_h(x - X_t)\}^{-1}$$

4. Bandwidth Selection

To attenuate the structure of time series data and the over-fitting or under-fitting tendency we opt for a nonparametric version of the Akaike Information Criterion(AIC) proposed by (Cai, 2002). That is select h by minimizing

$$AIC(h) = log\{\hat{\sigma}\} + \psi(tr(S_h), n) \quad [24]$$

Where $(\hat{\sigma}^2) = MASE(h) = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{m}_h(x_i)\}^2$, $\hat{m}_h$ is the fit (smoother) for n pairs of measurements and $\hat{m}_h^{(-i)}$ is the fit calculated by leaving out the $i-th$ data point. $(S_h)_{ii}$ is the $i-th$ diagonal element of the smoother matrix S_h, an $n \times n$ hat matrix depending on the X variate and bandwidth h. $\psi(\cdot)$ is a penalty function designed to decrease with increasing smoothness of $\hat{m}_h$ and

$\psi(tr(S_h), n)$ is chosen particularly to be the form of the bias-corrected version of the (AIC), due to Hurvich et.al. (1998) and

$$(tr(S_h), n) = \{tr(S_h) + 1/n\}/[1 - \{tr(S_h) + 2\}/n] \qquad [25]$$

$tr(S_h)$ is the trace of the smoothing matrix S_h regarded as the nonparametric version of degrees of freedom, called the effective number of parameters.

For fixed X, a nonparametric fit estimator is defined by setting the value $\hat{a}$ in [18] that minimize

$$\sum_{i=1}^{n} \rho_\theta (Y_i - a)^2 K\left(\frac{x - X_i}{h}\right) \qquad [26]$$

with a loss function $\rho_\theta(u) = \theta I_{\{u \geq 0\}}(u).u + (\theta - 1) I_{\{u<0\}}(u).u$. For the minimiser [24] to fit in the quantile regression estimation, we use $\frac{1}{n}\sum_{i=1}^{n} \rho_\theta\left(y_i - \hat{q}_\theta(x_i)\right)$ instead of $\hat{\sigma}$. The second modification concerns approximating the smoother matrix S_h by the iteratively reweighted least squares fit of the model. Thus we choose the bandwidth to the minimiser of

$$2log\left(\frac{1}{n}\sum_{i=1}^{n} \rho_\theta\left(y_i - \hat{q}_\theta(x_i)\right)\right) + \psi(S_h). \qquad [27]$$

Where, ψ (.) is the AIC penalizing function and (S_h) is the approximate smoother matrix, as the suitable smoothing parameter for the nonparametric quantile regression.

5. Empirical Results.

We consider the daily hourly peak electricity consumption data $\{D_t, t = 1, \ldots, 2250\}$ of Kenya Power and Lighting Company (hourly totals (in mega watts) from January 1st 2005 to 28th February 2011. To eliminate trend and seasonality, we consider transformations. We first use logarithmic transformation since $\{D_t\}$ is a series whose standard deviation increases linearly with the mean. The transformed series becomes $T_t = log(D_t)$ as shown in figure 1(a). To introduce stationarity in the series we apply the analysis of Brockwell and Davis (1991), by using the difference operator $(1 - B)(1 - B^7)$. We obtain a new series $Y_t = (1 - B)(1 - B^7)T_t$ which does not display any apparent deviations from stationarity. (See figure 1 (b))

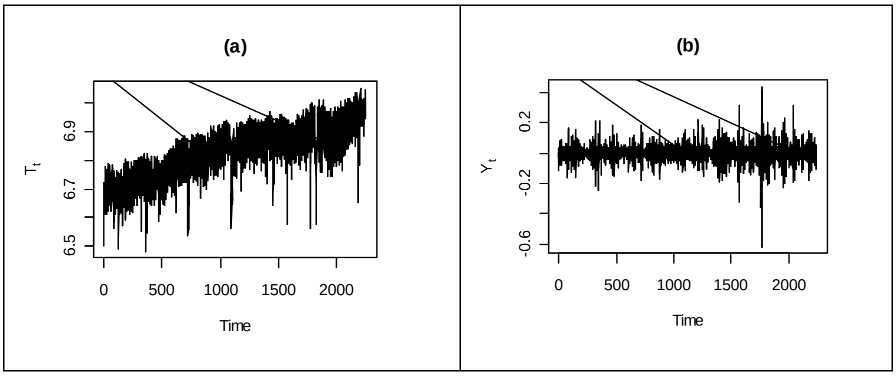

Figure 1 *(a) Time series plot for log transformed electricity consumption data. (b) Time series plot of twice-differenced log-transformed data.*

We consider a simple model $Y_t = \mu(Y_{t-1})dt + \sigma(Y_{t-1})dw_t$, where Y_t, is a stationary transformed consumption data, $\mu(\cdot)$ and $\sigma(\cdot)$ are smooth functions of Y_{t-1} and w_t is the standard Brownian motion. Our interest is to identify $\mu(\cdot)$ and $\sigma(\cdot)$. Since the time series sequence is observed at equally spaced time points, we use the infinitesimal generator, the first order approximation of moments of Y_{t-1}, a discretized version of the Ito's process $\partial Y_t = \mu(Y_{t-1})\partial + \sigma(Y_{t-1})\epsilon\sqrt{\partial}$. For simplicity we use $|Y_t|$ as a proxy of volatility.

Figure 2(a) shows a local smooth estimate of $\mu(Y_{t-1})$.The estimate is essentially zero. However to better understand the estimate, figure 2(b) shows this estimate on a finer scale. This estimate suggest that when $\mu(Y_{t-1})$ is positive Y_{t-1} is negative and when $\mu(Y_{t-1})$ is negative, Y_{t-1} is positive. Thus we infer that the conditional mean is a decreasing function. Figure 3(a) shows an estimate of $\hat{\sigma}(Y_{t-1})$. The plot shows that the lower the demand the higher the volatility and the higher the demand the higher the volatility. Figure 3(b) shows the estimate $\hat{\sigma}(Y_{t-1})$ on a finer scale. It confirms that the extreme the demand the higher the volatility.

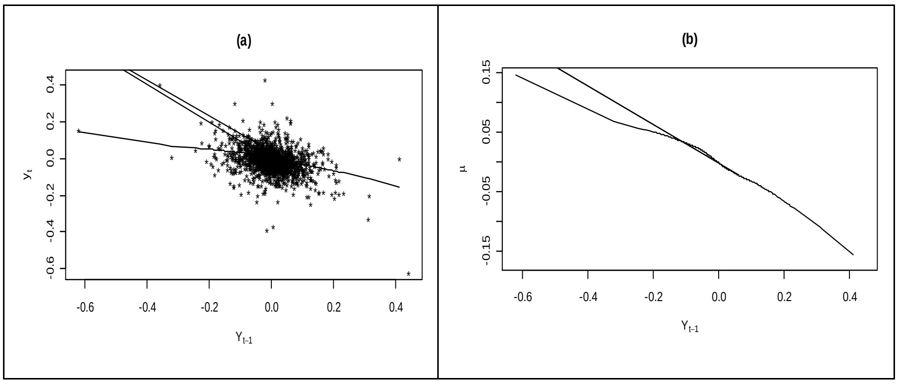

Figure 2 *(a) The smooth estimate of $\mu(Y_{t-1})$. (b) The estimate $\hat{\mu}(Y_{t-1})$ on a finer scale*

If we consider a time series model where gasoil price is the response variable and electricity peak hour consumption is an exogenous variable that is $P_t = \mu(P_{t-1}, D_{t-1}) + \varepsilon$ where, P_t is the stationary gasoil price at time t (data from Kenya National Oil Corporation) and D_t is the corresponding stationary electricity consumption. We see from figure 4(a) that consumption decreases from -0.4 to -0.2 then increases wiggly to 0.0 followed by a sharp decrease to 0.15 then a sharp increase. Figure 4(b) shows a residual plot of this model where electricity consumption is an exogenous variable. The residual plot shows no deviation from normality. To try to answer the question if the estimated quantile regression relationships confirm to the location shift hypothesis that assumes that all the conditional quantile functions have the same slope parameters we estimate the quantile fits. Figure 5(a) shows the estimated conditional quantile at some extreme quantiles; quantile 0.1 and quantile 0.9 respectively. In figure 5(b), two density estimates are presented, one for relatively lower consumption $(\theta = 0.1)$ and the other one for relatively higher consumption $(\theta = 0.9)$.

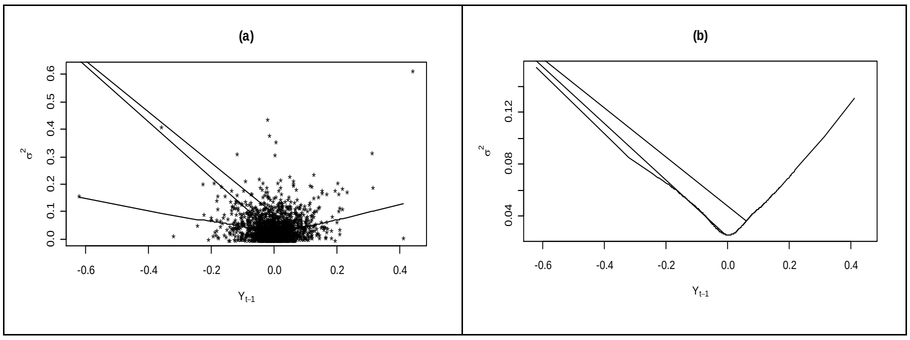

Figure 3 *(a) The estimate of volatility $\hat{\sigma}(Y_{t-1})$. (b) The estimate of volatility on a finer scale*

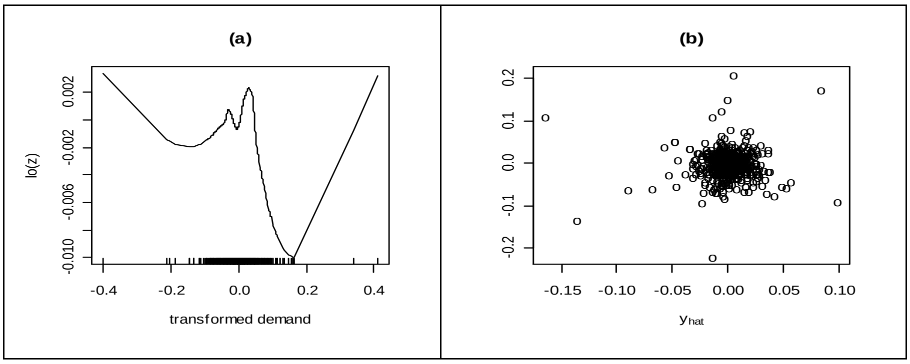

Figure 4 *(a) The smooth estimate of D_{t-1} in a model where consumption is an exogenous variable. (b) The residual plot of the model with electricity consumption as an exogenous variable*

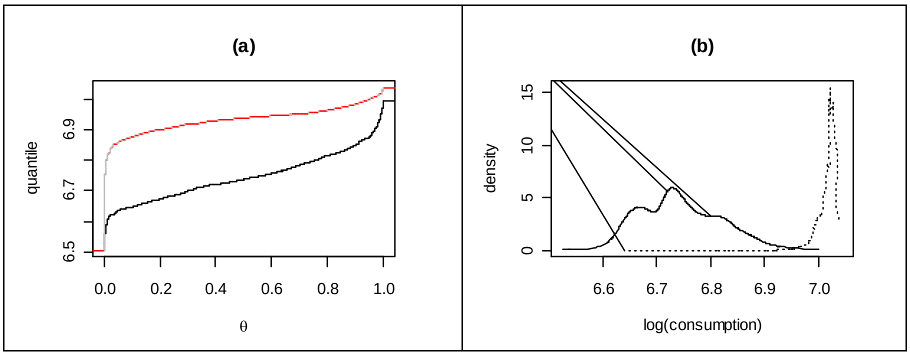

Figure 5 *(a) The estimated conditional quantile of electricity consumption at two extremes (0.9 quantile) and (0.1 quantile) respectively. (b) The probability density for lower demand and higher demand (0.1 and 0.9) respectively.*

We also find that of all competing seasonal first differenced parametric models, $ARIMA\ (1,1,2) \times (1,1,1)_7$ is the most parsimonious model with the smallest (AIC). On the other hand $GARCH(1,1)$ fits well the residuals, with the diagnostic tests revealing no deviations from normality. The 0.9 ,0.5 and 0.1 quantiles for the residuals plots are shown in figure 6. From these plots it is clear that the extreme quantiles; 0.1 and 0.9 are more volatile than the 0.5 quantile. These plots are in agreement with the results given in figure 3 which shows that volatility increases as you move towards the extremes.

Conclusion

We conclude that nonparametric kernel estimators possess some appealing properties when displaying the mean and volatility functions. Also we have shown that quantile regression provide a more complete picture on how the distribution of the response variable is conditioned on the previous outcomes. Also for nonlinear interactive covariates, nonparametric estimators are flexible enough to capture the underlying complex dependence structure. This model can be used to generate routine short time forecast and also as an analytical tool to avoid incorrect inferences.

Acknowledgement

We would like to thank The Kenya Power and Lighting Company and the Kenya National Oil Corporation for helping us obtain the data for this study. We would also like to thank the National Council for Science and Technology for their grant support towards this research. Finally we wish to thank the reviewers and the editors for their comments and suggestion.

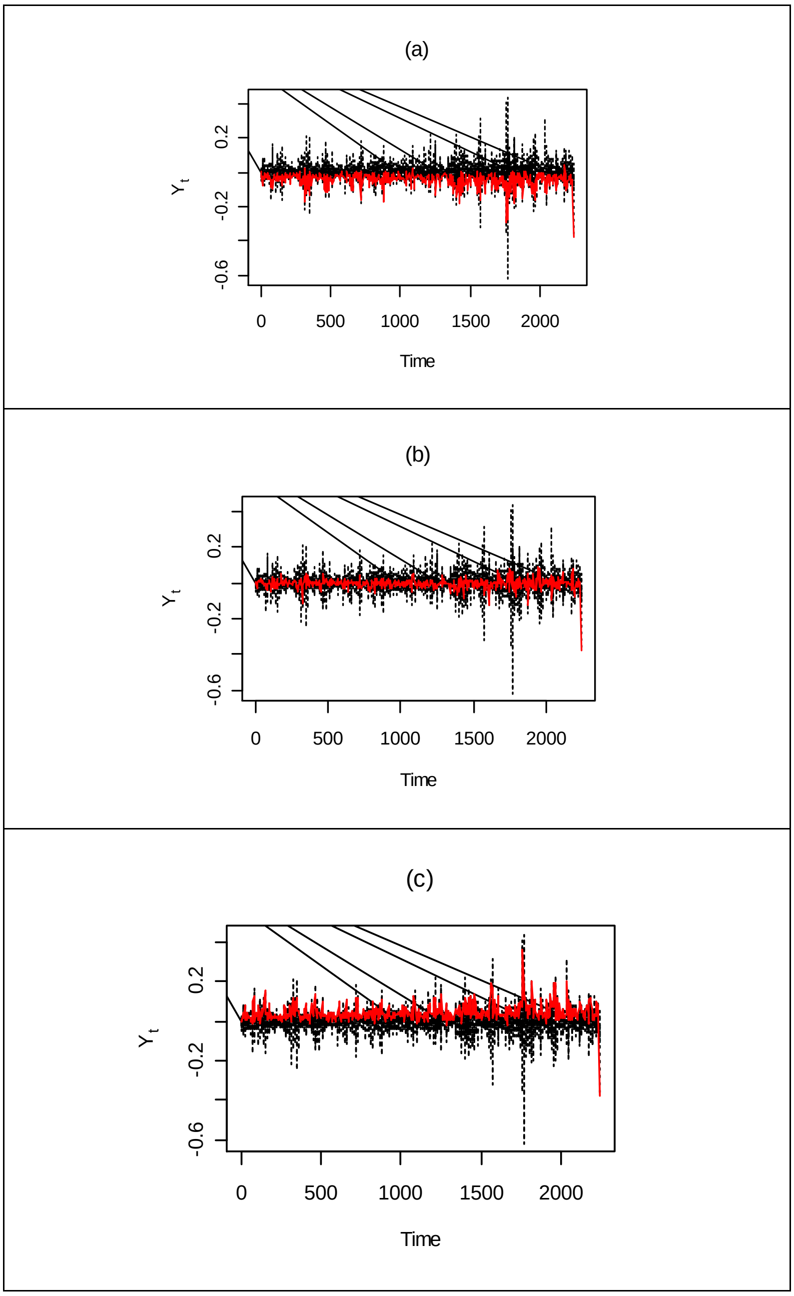

Figure 6 *(a) Residuals* 0.1 *quantiles for the electricity consumption data. (b) Residuals* 0.5 *quantile estimate and (c) The Residuals* 0.9 *quantile estimate.*

References

1. Engle, R.F. et al. (1986). Semi parametric estimates of the relation between weather and energy sales. Journal *of the American Statistical Association.* **81**, 310-320.
2. Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. Chapman and Hall, London.
3. Hall, P. et al (1999). Methods of estimating a conditional distribution function, *Journal of the American Statistical Association*. **94**,154-163
4. Hurvich, C.M. et al (1998). "Smoothing parameter selection in nonparametric regression using an improved Akaike Information Criterion," *Journal of the Royal Statistical Society*. Series B, **60**, 271-293.
5. Koenker, R., Basset, G.S. (1978). Regression Quantiles, *Econometrica*, **46**, 33-50.
6. Li, Q., Racine, J.S. (2004). Cross Validated Local Linear Nonparametric Regression, *Statistica Sinica*. **14** (2), 485-512.
7. Mwita, P.N., Otieno, R.O. (2005). Conditional Scale Function Estimate in the Presence of Unknown Conditional Quantile Function, *African Journal of Science and Technology (AJST)* Science and Engineering Series **6**, No. 1, pp. 50-80.
8. Nadaraya, E.A. (1964). On Estimating Regression, *Theory Prob. Appli.*, **10**, 186-1
9. Rosenblatt, M. (1956b.). A central limit theorem and strong mixing conditions, *Pro. Nat. Acad. Sci*. **4**, 43-47.
10. Volkonskii, V.A. and Rozanov, Y.A. (1959). Some limit theorems of random functions. *Theory Prob. Appl*. **4**, 178-197.
11. Watson, G.S. (1964). Smoothing Regression Analysis. *Sankhya, ser. A.* **26**, 101-116.
12. Yu, K., Jones, M.C. (1998). Local Linear Quantile Regression, *Journal of the American Statistical Association*. **93**, 228-237.
13. Zongwu, C. (2002). Regression Quantiles For Time Series. *Econometric theory*, 169-192.