

Statistical Analysis of Electricity Prices

Estáte Khmaladze
Victoria University of Wellington

Abstract: The paper presents a statistical analysis of electricity spot prices in a deregulated market in New South Wales, Australia, in the period 10 May, 1996 – 7 March, 1998. It is unusual that a single set of data, such as this, allows one to consider a relatively systematic sequence of statistical problems, each resulting in clear, although not always obvious, solutions. This is the reason why these data and their analysis can be used as a relatively good base for training in practical statistical analysis. Existing formerly as a report, the material has been used in lecture courses in several universities in Australia and New Zealand.

Key words: Key words: Abrupt changes in market conditions, AR-processes, change-point problem, daily average prices, deregulated electricity market, mixtures of normal distributions, net prices on options, options on electricity prices, periodic trends.

1. Introduction

There are many years since in New South Wales, Australia, the market on electric energy was de-regulated: generators produce the electric energy and sell it to consumers basically on demand and supply principle. In between generators and consumers there are distributing companies. These companies buy the energy from generators and sell it “in retail” to us, customers.

The prices can change every half hour - as often as whether forecast is released. One can be surprised to see how sharp these changes can be – around 30% and 50% and more in just half an hour.

Although there is enough political interference, as it is unavoidable and necessary when commodity of this importance is traded, the actual mechanism of forming the eventual price from requested prices (so called “bid-stacks”) is very interesting. It deserves separate description and analysis in game-theoretic terms. However, we will not go into this in the present paper.

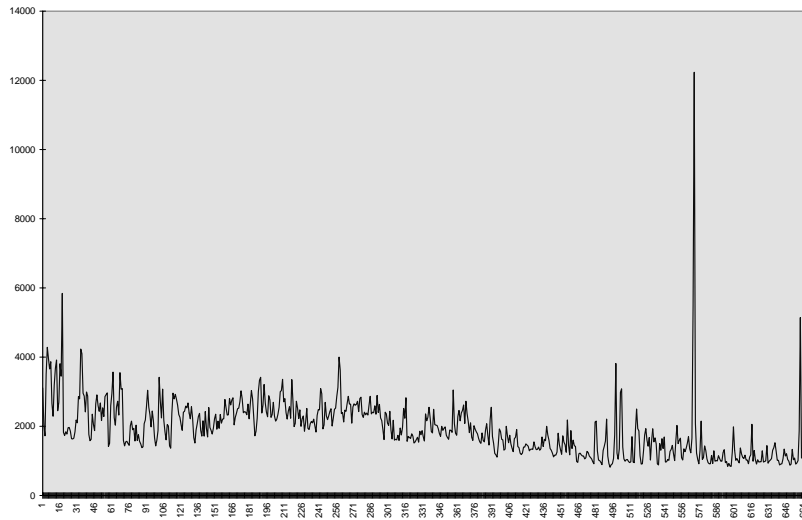


Figure 1: Daily averaaages

In this paper we consider sequence of daily average prices, that is, averages of 48 half-hourly prices on 1 Mgw of electric energy, as a time series. The series starts on 10 May 1996 and goes to 7 March 1998. The analysis was undertaken by the request from one of leading Australian banks and one of the major distributing companies in Sydney. So, although the sequence is relatively long the data now is also old enough, they believe, to make the publication not sensitive from the commercial point of view. The whole analysis was based only on the publicly available data on prices.

At first we will consider the dynamics of daly average prices. However, the distributor's eventual interest was the accurate evaluations of the net prices on various put options. Evaluation of net prices under "physical" and not transformed "martingale" measure (cf. Musiela and Rutkovski, 2001;. or Shiryaev, 1999) is of very clear interest in this market.

The author was using this work as a material for practical training for students in quantitative analysis of financial and economic data at UNSW and at VUW. It is used for the same purpose in some other universities. We submit it now for publication in a hope that it will be useful also for other students.

Once Antoine de Saint Exupery remarked that the best airplane is not the one to which you can not add anything, but the one which you can not remove anything from. The major difficulty for us in the present analysis was to follow this principle. It would be very easy, and at times tempting, to resort to more complicated models. But we tried not to allow ourselves to introduce any sort

of complications, unless there was very clear reason and unless the complications were unavoidable.

2. Daily Averages

The sequence of 667 average daily prices – daily averages for short – from 10 May 1996 until 7 March 1998, naturally forms a time series. The graph of these averages is given on Figure 1.

It can be noticed that in the first bit the behaviour of daily averages is more volatile than in the rest of the data, but near the end of data there are extremely high outliers present. Perhaps, some further speculations of heuristic nature are possible but without a formal probabilistic model they may be quite misleading.

Denote by $X(t)$ the daily average price at the day t . The first basic questions are

- is $X(t), t = 1, 2, \dots, n$ ($n = 667$ days in this study), a stationary time series? and if not
- does it have any systematic or periodic components?
- with these components filtered out, what is the behaviour of the rest?

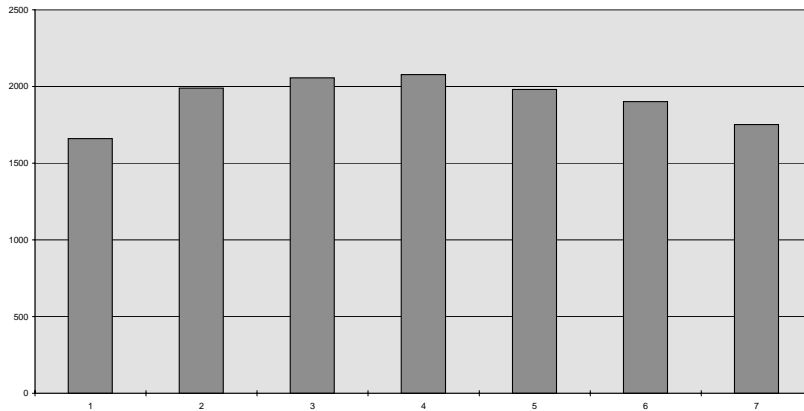


Figure 2: Weekly cycle (1-Sunday, 7-Saturday)

The obvious periodic component of $X(t)$ must be formed by weekly cycle of human activity. Hence we first assume that an expected value $EX(t) = m(t)$ of daily average $X(t)$ is a periodic function of t with period 7 days:

$$m(t) = m(t + 7)$$

An obvious estimator of $m(t)$ is provided by average prices for all Mondays, all

Tuesdays, etc., that is

$$\hat{m}(t) = \frac{1}{\ell} \sum_{k=0}^{\ell} X(t + 7k), \text{ for } t = 1, 2, \dots, 7$$

and for all other t it is defined as a periodic function: $\hat{m}(t) = \hat{m}(t + 7)$. Here ℓ is the number of weeks available. For weekdays Monday-Friday there are $\ell = 95$ weeks while for Saturday and Sunday there are $\ell = 96$ weeks. The graph of \hat{m} is shown on Figure 2.

We see that $\hat{m}(t)$ is of extremely regular and clear behaviour, which certainly suggests the same behaviour of $m(t)$.

Remark 1. Further refinement in the estimation of weekly cycle is possible, if necessary, although we did not do it here. Some of working days from 10 May 1996 to 7 March 1998 may actually be holidays which, most probably, followed the pattern of a weekend. So, these days may be identified and used to estimate $m(1)$ (Sunday) and $m(7)$ (Saturday) rather than $m(t)$ at nominal weekday.

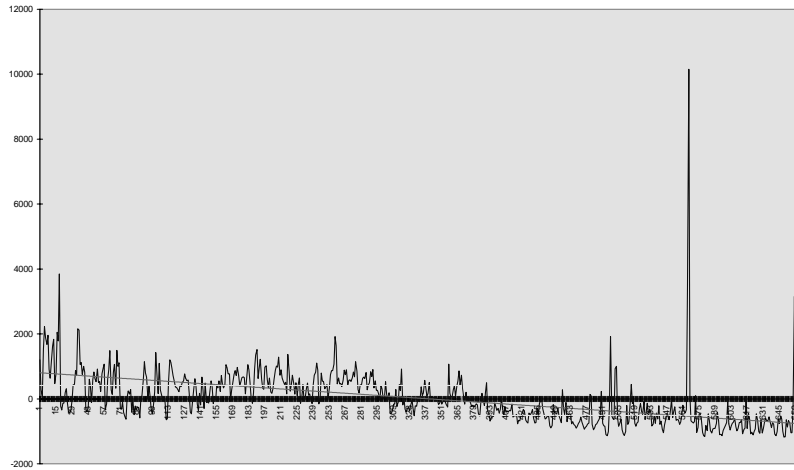


Figure 3: $\hat{Y}(t)$ = daily averages–weekday’s averages, linear regression and change-point regression.

Let now $Y(t)$ denote the process of residuals

$$Y(t) = X(t) - m(t).$$

Is $Y(t), t = 1, 2, \dots, n$, already a stationary process? To answer this question consider estimated residuals

$$\hat{Y}(t) = X(t) - \hat{m}(t).$$

Apriori one could think that a seasonal periodic trend still can be present in $\hat{Y}(t), t = 1, \dots, n$. On the other hand, with commodity like electric energy it is not quite obvious what this periodic trend can be: if in cold seasons one needs more energy for heating, in hot seasons one needs more energy for refrigeration and cooling. Therefore, on the whole one can intuitively agree with no signs of the seasonal trend in $\hat{Y}(t), t = 1, \dots, n$.

At the same time, in difference to $X(t)$, the graph of $\hat{Y}(t)$ (Figure 3) reveals the presence of relatively sharp and quick change in the process of daily prices. This change probably occurred around 360th day of observation which is 5 May 1997.

One may also think about downward linear regression for $\hat{Y}(t)$ with consequent heuristic perception of the market as the one of regular and steady “bearish” tendency. In our view, however, the so called change-point regression should work in this case better. In other words, we assume that an expected value of $Y(t)$ depends on t in the following way:

$$a(t) = EY(t) = \begin{cases} a_1 & \text{for } 1 \leq t \leq t_0 \\ a_2 & \text{for } t_0 + 1 \leq t \leq n \end{cases}$$

As an estimator of the change-point one can adopt

$$\hat{t}_0 = 360,$$

or, rather, use the interval between $\hat{t}_1 = 305$ (10 March 1997) and $\hat{t}_2 = 370$ (14 May 1997) as the “short” period of “sharp” changes – see below. For the estimators of a_1 , and a_2 we used partial averages

$$\hat{a}_1 = \frac{1}{\hat{t}_0} \sum_{t=1}^{\hat{t}_0} Y(t), \quad \hat{a}_2 = \frac{1}{n - \hat{t}_0} \sum_{t=\hat{t}_0+1}^n Y(t).$$

The numerical values of \hat{a}_1 and \hat{a}_2 were

$$\hat{a}_1 = \$4.13 \text{ and } \hat{a}_2 = -\$5.07.$$

The choice of the change-point regression, as opposed to linear trend, was made based on visual impression and some analysis presented later in §4. To some extent, the statements like this is a dangerous thing to make because if one claims that there was a sharp and quick change in the market conditions one should be able also to identify what caused or instigated such change. The author, being an academic statistician, did not have any necessary knowledge to present any such cause. However, later the specialists in distributing company confirmed that in that period of time the new high voltage lines became operational between

Victoria and New South Wales. Electricity generators in Victoria use brown coals, which is cheap fuel. Hence the export possibilities for cheaper energy from Victoria to New South Wales were increased, and that the prices quickly lowered.

So, around 5 May 1997 the prices dropped on average by \$9.20.

Let now $Z(t)$ denote further residuals:

$$Z(t) = Y(t) - a(t) = X(t) - m(t) - a(t)$$

and consider estimated $\hat{Z}(t)$'s as $\hat{Z}(t) = \hat{Y}(t) - \hat{a}(t)$.

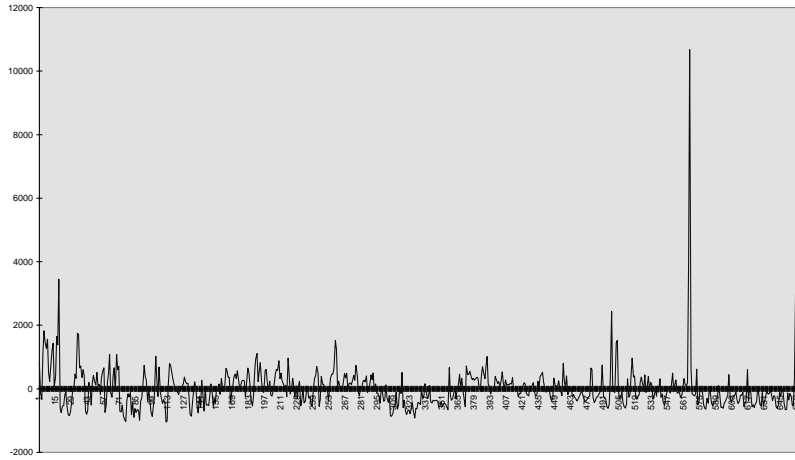


Figure 4: Graph of $Z(t)$'s with weekly components and change-point regression filtered out

As the graph of $\hat{Z}(t)$ suggests, the $Z(t)$ is already a good candidate for being modeled as a stationary time series. We consider this model in the next §2.

3. The Model for $Z(t), t = 1, 2, \dots, n$

We did not consider so far autoregression functions of $X(t), t = 1, 2, \dots, n$, or of $Y(t), t = 1, 2, \dots, n$. In the presence of components $m(t)$ and $a(t), t = 1, 2, \dots, n$, this will not be useful: the presence of systematic components will essentially deform the picture presented by this autoregression function. Indeed, as Figure 5 demonstrates, the sample autocorrelation function of $X(t), t = 1, 2, \dots, n$, denoted $\hat{\gamma}_x$, is much higher than that of $\hat{Z}(t)$, denoted $\hat{\gamma}_z$. The excessive manifestation of dependence in $X(t)$'s and certain periodic pattern is simply consequence of the presence of periodic trend $m(t)$ and the shift $a(t)$.

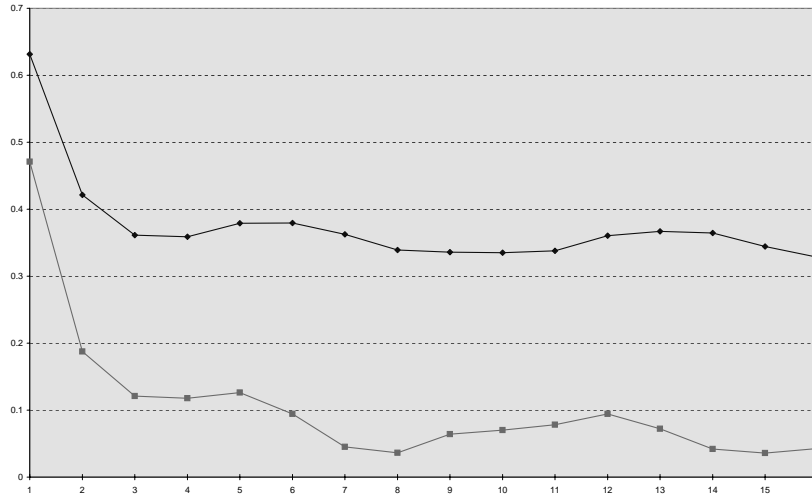


Figure 5: Empirical autocorrelation of daily averages $X(t)$ and of $Z(t)$

Analysis of empirical autocorrelation function of $\hat{Z}(t)$ suggests that $Z(t)$ can be reasonably assumed a stationary time series. Indeed, on Figure 6 empirical autocorrelation function $\hat{\gamma}_z(s)$ is shown for different starting points t , that is, for different parts of the whole sequence: for $t = 1$ (the whole sequence), $t = 200$ (the part of sequence starting with 23 November 1996), $t = 371$ (starting with 11 May 1997), $t = 400$ (starting with 11 June, 1997) and $t = 500$ (starting with 20 September 1997). These empirical autocorrelations are quite close to each other and follow identical pattern. We need, however, more specific model for $Z(t)$ than stationarity alone can suggest.

First of all we try autoregression step 1 models, or AR(1) in notations (see, e.g., Box and Jenkins (1976), Chan (1993) for the theory of autoregressive processes). According to this model

$$Z(t) = \theta Z(t-1) + \varepsilon(t), \quad t = 2, 3, \dots, n, \quad (3.1)$$

where θ is a constant autoregression coefficient and $\varepsilon(t), t = 2, 3, \dots, n$, is a sequence of independent and identically distributed random variables (i.i.d. sequence). These random variables are called “innovations” (of Z). Remark that AR(1) does not specify the distribution of $\varepsilon(t)$'s – in principle, it can be any.

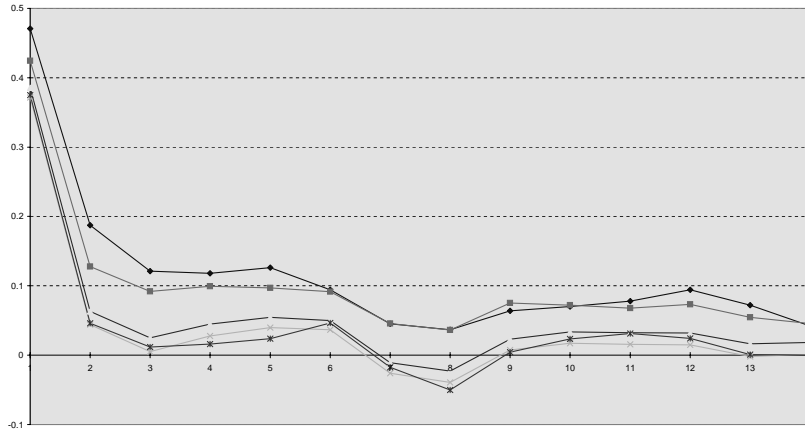


Figure 6: Empirical autocorrelation functions of $Z(t)$ starting at $t = 1, 200, 371, 400$ and 500

The model (3.1) specifies, however, autocorrelation function of $Z(t)$: if (3.1) is true then

$$\text{Corr}(Z(t), Z(t+s)) = \gamma_1(s) = \theta^s, \quad s = 0, 1, 2, \dots \quad (3.2)$$

The AR(1) model also implies that $Z(t)$'s have the following representation through $\varepsilon(t)$'s

$$Z(t) = \sum_{j=0}^{t-2} \theta^j \varepsilon(t-j) + \theta^{t-1} Z(1) \quad (3.3)$$

One can try higher order autoregression models as well and we will consider AR(2) model. According to this model

$$Z(t) = \theta_1 Z(t-1) + \theta_2 Z(t-2) + \varepsilon(t), \quad t = 3, 4, \dots, n, \quad (3.4)$$

where θ_1 and θ_2 are, again, autoregression coefficients and $\varepsilon(t), t = 3, 4, \dots, n$, forms an innovation i.i.d. sequence. This model too specifies the autocorrelation of $Z(t)$ completely: if (3.3) is true then

$$\begin{aligned} \gamma_2(1) &= \theta_1 / (1 - \theta_2) \\ \gamma_2(2) &= \theta_1 \gamma_2(1) + \theta_2 \\ \gamma_2(s) &= \theta_1 \gamma_2(s-1) + \theta_2 \gamma_2(s-2), \quad s = 3, 4, \dots \end{aligned} \quad (3.5)$$

and these recursive formulae determine $\gamma_2(s)$ uniquely.

Let us choose now the value $\hat{\theta}$ of θ , which gives the best fit of $\gamma_1(s)$ of (3.2) to the empirical autocorrelation function of $\hat{Z}(t), t = 1, 2, \dots, n$.

We choose $\hat{\theta} = 0.471$ and to what extent does $\hat{\gamma}_1(s) = \hat{\theta}^s$ agree with the empirical autocorrelation function can be seen in Figure 7. This fit for the first three values is quite good but for the rest part it seems perhaps not so good.

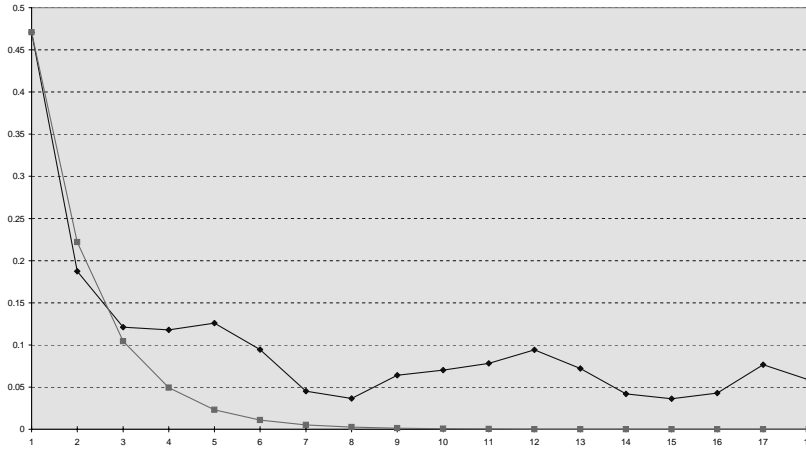


Figure 7: Empirical autocorrelation of $Z(t)$ and fitted AR(1) and AR(2)

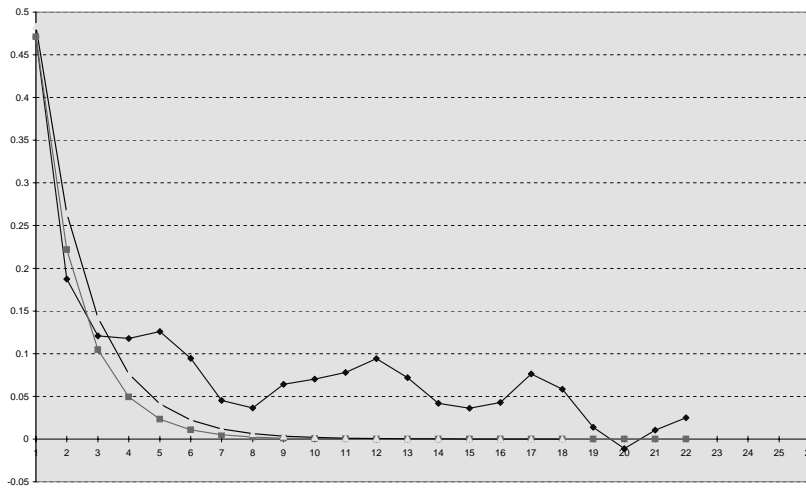


Figure 8: Empirical autocorrelation of $Z(t)$ and $s^{-\alpha}$ (long range dependence model)

Similarly we can choose values $\hat{\theta}_1$ and $\hat{\theta}_2$ of θ_1 and θ_2 such that they give the best fit of (3.4) to the empirical autocorrelation of $\hat{Z}(t)$. We choose these values as $\hat{\theta}_1 = 0.464$ and $\hat{\theta}_2 = 0.041$. All three autocorrelation functions are shown on Figure 7.

Several conclusions follow from what we have observed so far:

1. The fit of AR(2)-model is not much better than that of AR(1)-model. The value of coefficient $\hat{\theta}_2 = 0.041$ is quite small. Hence there is very little advantage, if any, in AR(2) model relative to more simple AR(1) model. Therefore, autoregression models of order higher than 2, in our view, will look artificial and unnecessary.
2. Both models do not provide very good approximation for the tail of empirical autocorrelation $\hat{\gamma}_z$ for $s \geq 4$ though this makes very little influence on what follows in 3. The tail of $\hat{\gamma}_z$ looks more like a power function $s^{-\alpha}$, and indeed the choice of $\hat{\alpha} = 1.2$ provides reasonable approximation for $\hat{\gamma}_z$. This may be an indication of very different nature of the process $Z(t), t = 1, \dots, n$, since power autocorrelation function is characteristic for the processes with so called long range dependence.

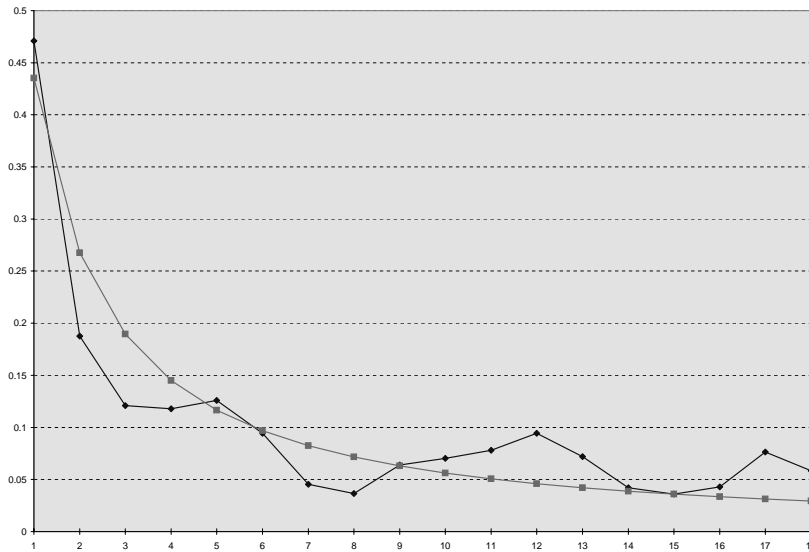


Figure 9: Empirical autocorrelation of AR(1) – innovations

The modern theory of the processes with long range dependence is extremely interesting and important (see, e.g., Beran (1994), Koul and Surgailis (2002), Qian, L. (1998)). However, for the purposes of the present investigation and description of marginal distribution of $Z(t)$'s and hence, of prices $X(t)$ themselves, it can be avoided.

3. Using AR(1) model with estimated $\hat{\theta}$ we construct very useful sequence of the estimated innovations

$$\hat{\varepsilon}(t) = \hat{Z}(t) - \hat{\theta}\hat{Z}(t-1), \quad t = 2, 3, \dots, n.$$

Empirical autocorrelation function $\hat{\gamma}_\varepsilon$ of these innovations is now very close to 0 throughout the whole sequence. Its first 10 values, $\hat{\gamma}_\varepsilon(s)$, $s = 1, 2, \dots, 10$, starting with different values of t , and namely with $t = 1, 200, 371, 400, 500$ are given on Figure 9. The luck of fit for $s \geq 4$ of γ_1 to γ_z makes here very little difference.

Remark 2. Further refinement in reconstruction of AR(1) innovations is possible. One, however, would need to consider the outliers first. The presence of outliers essentially influences the value of θ and variances of $Z(t)$ and $\varepsilon(t)$, especially for $t \geq t_o$.

4. Outliers

The prices which exceed certain high level η should be treated as “outliers”. What is this high level? The choice of its value can be a subject of a statistical theory and we could present it here. However, we believe that in the present case we do not have as much a problem of how to define outliers - basically prices above \$80-\$100 are already outliers - but more a problem of understanding what is the influence of outliers on the whole picture and how do they change inference concerning our ultimate object - determination of option prices.

The following Table 1 describes number of half-hourly intervals in the whole data set when the price on 1 Mgw exceeds the indicated threshold level. For example, there are 54 intervals when the price was over \$100/Mgw.

Table 1: Number of spot prices exceeding the indicated threshold level

| | | | | | | | | | |
|------|------|------|------|-------|-------|-------|-------|-------|-------|
| 886 | 327 | 139 | 76 | 54 | 26 | 21 | 18 | 13 | 0 |
| \$40 | \$50 | \$65 | \$80 | \$100 | \$150 | \$200 | \$250 | \$300 | \$350 |

The number of outliers may not seem large. For example, 54 prices exceeding \$100 barrier is not a big number relative to total number of 32000 observations. However, their influence is very noticeable.

There are several possible ways of treating outliers. We adopted the one when an outlier is replaced by a portion of the threshold level. Actually, we replaced an outlier by 1/3 of the nominal threshold level.

What we observed then can be summarized as follows:

1. The general structure of processes $X(t)$, $Z(t)$ and $\varepsilon(t)$ as described in §§1-2 remains unchanged and correct for all values of threshold starting from as small number as $\eta = \$80$. This means that the main body of observations essentially follows the models proposed above.

2. The estimated values of the parameters, however, may now be quite different and depend on the level of threshold. This includes, in particular, estimation of autoregression coefficient θ and variances σ_z^2 and σ_ε^2 of $Z(t)$ and $\varepsilon(t)$.
3. For smaller threshold level both variances σ_z^2 and σ_ε^2 are getting smaller quite significantly. At the same time the value of θ is getting considerably larger. This latter means that main body of $Z(t)$'s has heuristically speaking "stronger" dependence structure and outliers are masking it, to certain extent.

The following Table 2 shows values of estimations $\hat{\theta}$ of autoregression coefficient and of both standard deviations $\hat{\sigma}_z$ and $\hat{\sigma}_\varepsilon$ calculated for the whole sequence and for the part of the sequence after the change-point $\hat{t}_0 = 370$ (after 5 May 1997).

Table 2: Estimations of the autoregression coefficient θ and of the standard deviations σ_z and σ_ε

| Threshold | $\hat{\theta}$ | $\hat{\sigma}_z$ | $\hat{\sigma}_\varepsilon$ | after change-point | | |
|-------------------|----------------|------------------|----------------------------|--------------------|------------------|----------------------------|
| | | | | $\hat{\theta}$ | $\hat{\sigma}_z$ | $\hat{\sigma}_\varepsilon$ |
| $\eta = \infty^*$ | 0.47 | \$6.73 | \$5.93 | 0.36 | \$7.91 | \$7.30 |
| $\eta = \$100$ | 0.64 | \$5.03 | \$3.95 | 0.54 | \$3.80 | \$3.20 |
| $\eta = \$80$ | 0.65 | \$4.89 | \$3.74 | 0.57 | \$3.52 | \$2.87 |

* outlier unaltered.

Remark 3. Estimations shown in this table provide another indirect support for application of AR(1) model for $Z(t)$. According to this model equality

$$\sigma_z = \sigma_\varepsilon / \sqrt{1 - \theta^2} \quad (4.1)$$

must be true. But, for example, in the case of $\eta = \$100$ for the whole sequence we have $\hat{\sigma}_z = \$5.03$ while $\hat{\sigma}_\varepsilon / \sqrt{1 - \hat{\theta}^2} = \5.11 , and in the case of $\eta = \$80$ for the part of the sequence after change-point we have $\hat{\sigma}_z = \$3.52$ while $\hat{\sigma}_\varepsilon / \sqrt{1 - \hat{\theta}^2} = \3.49 . Even in the case of no truncation this relationship between variances seems remarkably stable: we have $\hat{\sigma}_\varepsilon = \$5.93.3$ and $\hat{\theta} = 0.47$, so that $\hat{\sigma}_\varepsilon / \sqrt{1 - \hat{\theta}^2} = \6.72 while actually $\hat{\sigma}_z = 6.83$ so that the difference is only 10c. We believe this is a good agreement with (6.1).

Another place where outliers influence the numerical values of the estimators is the marginal distribution of $Z(t)$ and $\varepsilon(t)$. We will see that it is very reasonable to assume that these distributions are normal (see §5). It is also clear that

the form of corresponding histograms is not affected by outliers. However, the values of estimated means and standard deviations are very much influenced by outliers. Therefore, either we would have to use truncated data or estimate these parameters not by their sample analogues but somewhat differently.

It would be, of course, very interesting to describe and model the whole process of outliers in more detail. However, we did not undertake this description here.

5. Change-point Problem: Detection of Abrupt Changes

The simplest formulation of the change-point problem is as follows. Let $\xi_1, \xi_2, \dots, \xi_n$ be a sequence of random variables, most often assumed independent. According to the null hypothesis the random variables are all identically distributed with some distribution F . Under the alternative hypothesis, however, they are not identically distributed – there is a moment n_0 such that for $i > n_0$ distribution of ξ_i has changed to some other distribution G . In particular, according to the null hypothesis expected value of all ξ_i is the same, say – some μ_1 , but under change-point alternative the expected value of ξ_i after some n_0 has changed to another quantity μ_2 . It is important to notice that under the alternative the value n_0 of a change-point is not specified and remains unknown. The problem is to detect whether a change-point exists at all, that is, if changes really occurred, and to estimate n_0 and also μ_1 and μ_2 or F and G .

The literature on the change-point problem is very wide we refer to Brodski and Darkhovski (2000) for modern presentation of the topic). The methods available can be divided in two major groups – those designed for the quickest detection of changes (see, e.g., Chow, Robbins, and Siegmund (1971)) and those designed for so called posterior detection of changes (see, e.g., Brodski and Darkhovski (2000)). The former are based on the methods of sequential analysis while the later are based on the so-called posterior detection approach. These methods are mathematically different but have many features in common. To illustrate the latter approach consider so called partial sum process (or cumulative sums, or cusum process) based on ξ_i 's: let

$$S_k = \sum_{i=1}^k \xi_i, \quad k = 1, 2, \dots, n,$$

and let

$$V_k = S_k - \frac{k}{n}S_n = \sum_{i=1}^k (\xi_i - \bar{\xi}) \quad (5.1)$$

where $\bar{\xi}$ is sample average of ξ_1, \dots, ξ_n . It is not difficult to see that under the null hypothesis of no change expected value of V_k for all k is 0,

$$EV_k = 0, \quad k = 1, 2, \dots, n.$$

Under alternative it is different: since

$$ES_k = \begin{cases} k\mu_1, & k = 1, 2, \dots, n_0 \\ n_0\mu_1 + (k - n_0)\mu_2, & k = n_0 + 1, \dots, n, \end{cases}$$

then

$$EV_k = \begin{cases} k(\mu_1 - \bar{\mu}), & k = 1, 2, \dots, n_0 \\ n_0(\mu_1 - \bar{\mu}) + (k - n_0)(\mu_2 - \bar{\mu}), & k = n_0 + 1, \dots, n; \end{cases}$$

where

$$\bar{\mu} = \frac{n_0}{n}\mu_1 + \frac{n - n_0}{n}\mu_2.$$

It is clear that if an average price after a change point n_0 jumped up, $\mu_1 < \mu_2$, then $\mu_1 - \bar{\mu} < 0 < \mu_2 - \bar{\mu}$ and if it dropped down, $\mu_1 > \mu_2$, then $\mu_1 - \bar{\mu} > 0 > \mu_2 - \bar{\mu}$. It is also clear that $V_n = 0$. Hence under alternative the graph of EV_k is either

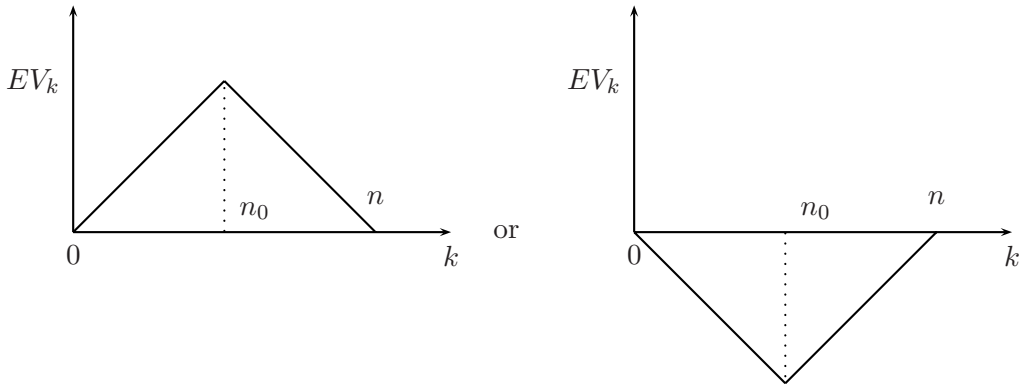


Figure 10: Shift of V_k under the change point alternative

and $\max |EV_k|$ is attained at the change-point n_0 , while the expected value of S_k at the change-point n_0 only bends. It is easier to estimate the point of maximum of the shift than the point of a bend. Besides, the expected value of V_k depends only on the difference $\mu_1 - \bar{\mu}$ (since $\mu_2 - \bar{\mu} = -(\mu_1 - \bar{\mu})$) and not on the “common level” $\bar{\mu}$, which is very convenient property. Hence to detect the presence of

changes it is advantageous to use statistics based on $V_k, k = 1, \dots, n$, instead of $S_k, k = 1, \dots, n$.

Below on Figure 11 we show the graph of V_k based on the whole period of $Y(t), t = 1, \dots, 667$.

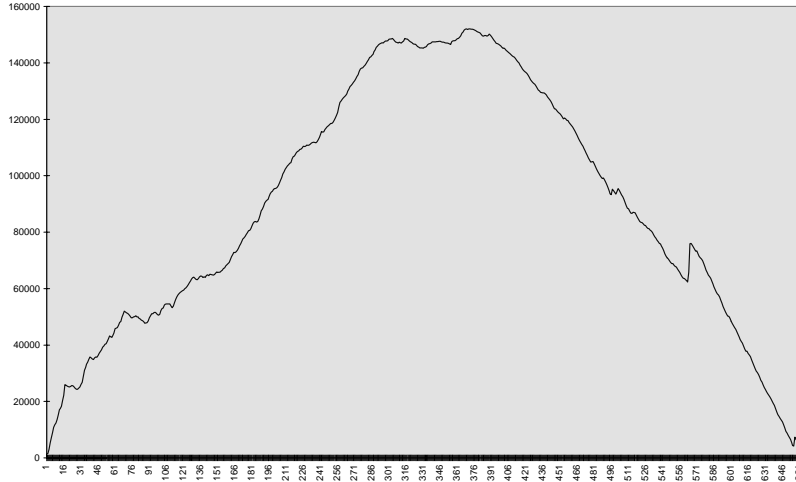


Figure 11: Cusums (5.1): graph shows indirect support the for piece-wise stationarity

The picture on Figure 11 is strikingly similar to the graph of EV_k above, and presence of a change is absolutely obvious.

One could say that there is not just one single change-point t_0 but rather there is a short transition period between, approximately, $\hat{t}_1 = 305$ and $\hat{t}_2 = 370$. Slightly better fit to $EY(t)$ then the function $a(t)$ of §1 is, therefore, possible. This is true but we will not engage ourselves with these improvement here, however.

The commonly used test for the presence of change is based on Kolmogorov–Smirnov statistic $\max |V_k|$: if it exceeds certain high level the presence of the change is accepted. However, what values of $\max_k |V_k|$ already indicate presence of changes? To answer this question it is better to standardize $V_k, k = 1, 2, \dots, n$ and consider the process on $[0, 1]$ defined as

$$u_n(t) = \frac{1}{\sqrt{ns_n}} V_k, \quad \frac{k-1}{n} \leq t < \frac{k}{n},$$

where s_n^2 is the sample variance of ξ_1, \dots, ξ_n . If there is no change in expected valued of our ξ_i -s, then, for large n , the distribution of the process $u_n(t), 0 \leq t \leq 1$,

is close to the distribution of the *standard Brownian bridge* and therefore the distribution of, say, Kolmogorov–Smirnov statistic $\max_{0 \leq t \leq 1} |u_n(t)|$ is close to the so-called *Kolmogorov distribution* (see, e.g., Shorack and Wellner (1986)). The limit distributions of many other goodness of fit statistics from $u_n(t), 0 \leq t \leq 1$, can also be found in the classical goodness of fit theory (cf., e.g., D’Agostino and Stephens (1986)).

What is the power of change-point tests? What is the accuracy in estimation of a change-point n_0 and of the value of changes occurred? How small changes can be detected? These and other questions also have exact answers in the statistical theory. For i.i.d. observations this theory, as we said, is well developed and is implemented within advanced data bases in the real practice.

The reader may like to conduct simulation experiment with the change-point present and see what would be the accuracy of the estimators \hat{a}_1, \hat{a}_2 and \hat{t}_0 . According to the theory of the change-point detection, $\hat{a}_k - a_k = O_P(1/\sqrt{n}), k = 1, 2$, while estimator of the change-point itself is much more accurate, $\hat{t}_0 - t_0 = O_P(1/n)$.

For more complex time-series, many important facts are also known – see, e.g., Berkes, Gombay, Horvath and Kokoszka (2004); Chan (1993); Horvath, Hšková, Kokoszka and Steinbach, J. (2004); Qian (1998) and references therein. However, the present case of AR(1) sequence can be easily transferred to the i.i.d.-case. Indeed, if $Y(t), t = 1, \dots, 667$, form an AR(1) sequence with constant mean, that is, if there is no abrupt change, then for the partial sums $S_k^Y = \sum_{t=1}^k Y(t)$ we obtain

$$S_k^Y = \theta S_{k-1}^Y + \sum_{t=1}^k \varepsilon(t)$$

or

$$S_k^Y = \frac{1}{1-\theta} \sum_{t=1}^k \varepsilon(t) - \frac{\theta}{1-\theta} Y(k).$$

Hence the normalized processes

$$u_n^Y(t) = \frac{1}{\sqrt{n} s_n^Y} V_k^Y \quad \text{and} \quad u_n^\varepsilon(t) = \frac{1}{\sqrt{n} s_n^\varepsilon} V_k^\varepsilon$$

are related as

$$u_n^Y(t) = \frac{1}{1-\theta} \frac{s_n^\varepsilon}{s_n^Y} u_n^\varepsilon(t) + o(1/\sqrt{n})$$

If we recall that $s_n^{\varepsilon,2}/s_n^{Y,2} \approx 1 - \hat{\theta}^2$, then, eventually, we see that the process

$$\frac{1 - \hat{\theta}}{\sqrt{1 - \hat{\theta}^2}} \sqrt{n} s_n^Y V_k^Y = u_n^\varepsilon(t) + o(1/\sqrt{n})$$

for large enough n behaves again as standard Brownian bridge. Hence, the asymptotic null distribution of many goodness of fit statistics from this process are known and can be used.

Wider presentation of methods of change-point detection will, however, distract us from the main subject of the present report.

6. Marginal Distribution of AR(1)-innovations

As we remarked in 3, autoregression model does not specify or require anything from the distribution of innovations $\varepsilon(t)$'s. However, as our initial data $X(t), t = 1, \dots, n$, are formed as averages of 48 not very dependent summands, one could expect normality in the distribution of its innovations. In Figure 12 we show histogram of estimated AR(1)-innovations $\hat{\varepsilon}(t)$ and for comparison, on Figure 13, we show histogram of average daily prices $X(t)$ themselves.

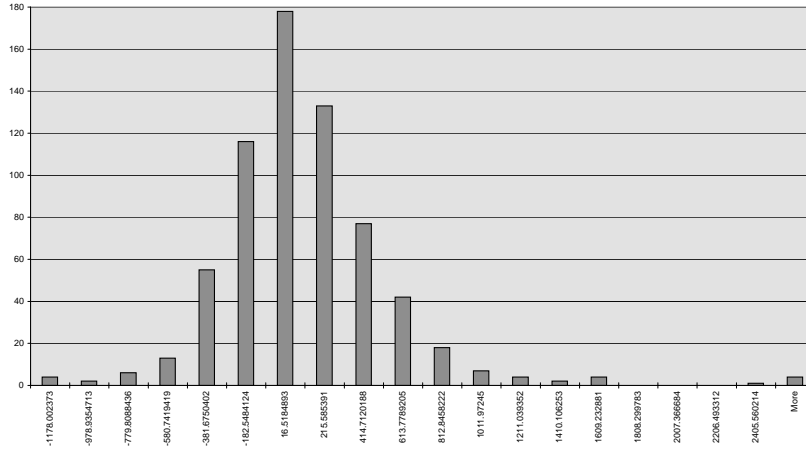


Figure 12: Histogram of AR(1) – innovations, $e(t)$, in scale of stand deviation.

Below it will become clear that this histogram should be close to a mixture of normal distributions, but we think it is not easy to recognize any such mixture behind the graph of Figure 13.

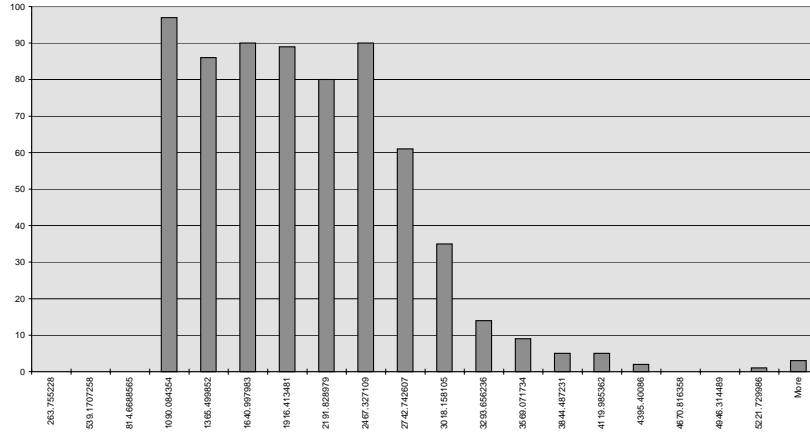


Figure 13: Histogram of daily averages, $X(t)$, in scale of stand deviation.

According to (3.3), the process $Z(t)$ can be represented as a linear transformation of innovation process $\varepsilon(t)$. After as small t as $t = 5$ or 6 the term $\theta^{t-1}Z(1)$ becomes negligible and we will drop it. Hence we consider

$$Z(t) = \sum_{j=0}^{t-2} \theta^j \varepsilon(t-j) \quad (6.1)$$

This represents Z as a so-called linear process (see, e.g., Box and Jenkins (1976)). For $\varepsilon(t)$'s independent and normal $N(0, \sigma_\varepsilon)$ we obtain that $Z(t)$ also has to be normal:

$$Z(t) \sim N(0, \sigma_z) \quad \text{with} \quad \sigma_z^2 = \sigma_\varepsilon^2 / (1 - \theta^2). \quad (6.2)$$

Histogram of $Z(t)$ is shown on Figure 14. It is again of quite regular nature. Indirectly this again supports assumptions of AR(1) nature of $Z(t)$ and independence of $\varepsilon(t)$.

As soon as we can positively state that marginal distributions of $\varepsilon(t)$'s and $Z(t)$'s are the normal ones we will be able to describe marginal distributions of $X(t)$'s. And indeed histograms of Figure 12 and Figure 14. agree well with a normal distributions $N(\mu, \sigma)$ with an expected value μ and standard deviation σ . Note that these histograms are based on the full data, outliers included.

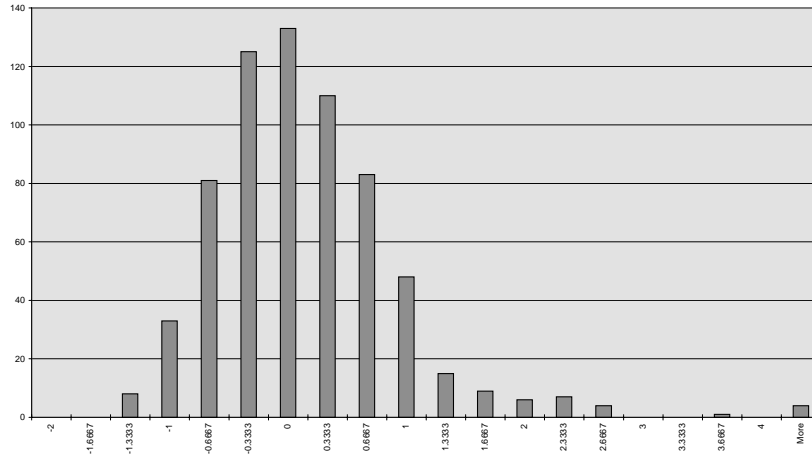


Figure 14: Histogram of $Z(t)$'s in scale of stand deviation.

This explains the irregularities on the right hand tail of the histogram on Figure 14. In other words, without truncation of outliers the marginal distribution of $Z(t)$ is a mixture of normal distribution and the distribution of outliers. Hence empirical variance $\hat{\sigma}_z^2$ is an estimator of variance of this mixture and not of the main normal component only. The variance of the latter is only about 43% of the total variance σ_z^2 and is remarkably small: $\sigma = \$2.84$.

So far all this may look consistent and logical. However, we will see in the next that we are ready to fall into a very interesting trap.

7. Marginal Distribution of Daily Averages $X(t)$

If we agree with the model

$$X(t) = m(t) + a(t) + Z(t) , t = 1, 2, \dots, n,$$

where $m(t)$ is (deterministic) periodic function and $a(t)$ is (deterministic) step-function with one jump, and if we also agree that the marginal distribution of $Z(t)$'s is normal, we must conclude that the marginal distribution of $X(t)$ is a mixture of normal distributions. If, in particular, we consider $X(t)$ for $t > \hat{t}_0$ only, we have to state that the marginal distribution of $X(t)$ must be close to the following mixture of normal distributions

$$\frac{41}{289} \sum_{j=1}^4 N(m(j) + a_2, \sigma_z) + \frac{42}{289} \sum_{j=5}^7 N(m(j) + a_2, \sigma_z) \quad (7.1)$$

Indeed, for $368 < t \leq 667$ there are 41 Sundays, Mondays, Tuesdays and Wednesdays and 42 Thursdays, Fridays and Saturdays. For the reader's convenience we recall the estimated values of the parameters (in \$)

Table 3: *****Please add caption ***

| $\hat{m}(1)$ | $\hat{m}(2)$ | $\hat{m}(3)$ | $\hat{m}(4)$ | $\hat{m}(5)$ | $\hat{m}(6)$ | $\hat{m}(7)$ | $\hat{a}(2)$ | $\hat{\sigma}_z$ |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|------------------|
| 16.60 | 19.90 | 20.56 | 20.78 | 19.81 | 19.01 | 17.52 | -5.07 | 2.84 |

The next Figure 15 shows the histogram of $X(t)$'s along with approximation (7.1).

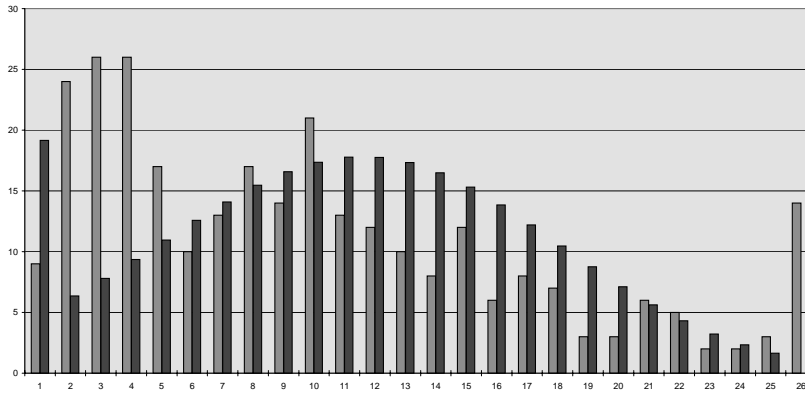


Figure 15: Histogram of $X(t)$'s after change-point and mixture (7.1) of Normal distributions

The first reaction of the author of this paper on this Figure was strong disappointment and the feeling that the whole harmonious and well-balanced theory above falls apart. Indeed, the agreement was very poor.

It was only after detailed analysis that it became clear where the problem lies. We will see that the model above is still basically correct apart from the statement that marginal distribution of $Z(t)$'s is normal and that $Z(t)$ is strictly

stationary.

To detect this, one needs to do tedious and somewhat careful analysis of the data. We do not want, however, to overload the present paper with all of details of this analysis, and instead will present the basic final results.

These are the following ones:

1. The distributions of daily averages for each day of a week before and after the change point agree with normal distribution. Most of the time the agreement is good although sometimes it is on a border.
2. However, not only expected values of these distributions are different (this was clear initially), but also the variances for different weekdays are different. Variances after the change-point are, in general, smaller.

This implies that centered process $Z(t)$ is not exactly stationary. Though it still is a Gaussian process, the empirical distribution of $Z(t)$'s, all put together, does not correspond to a Gaussian but to a mixture of 0-mean Gaussian distributions. Since variances of the admixtures are not dramatically far apart, for given number of observations to distinguish such a mixture from normal distribution statistically is extremely difficult.

3. Not simply variances are different for different days of a week, but the larger variances correspond to the larger expected values.

For mixture of normal distributions with the same variance, taken with almost equal weights, both left and right tails must have similar behavior. However, if to higher expected values it corresponds also higher value of variance, the right hand tail will become more flat and closer to 0. Exactly this is the case for the histogram of $X(t)$'s.

4. Too abrupt left tail of this histogram is connected with certain truncation of low prices and is of technical and not very important nature. Altogether the left tail of the marginal distribution is unimportant for the option pricing problem.
5. The estimates of variances we derived are very far from sample variances. The use of these sample variances would be completely inadequate because they are very strongly influenced by outliers.

Also note that these variances and individual histograms (for each day of a week) are not looking quite stable. It means that the number of 41 weeks of observation is not quite enough and also that, perhaps, the market had not been quite stabilized yet.

6. It is possible to assume variances for all working days equal with probably some reservation concerning Fridays, and variances for Saturdays and Sundays also equal. But the latter are much smaller than the former (see below).

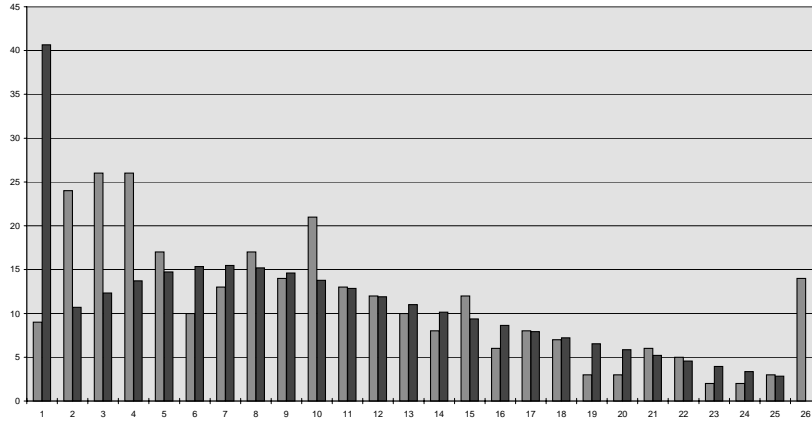


Figure 16: Histogram of $X(t)$'s after the change-point and mixture (6.2)

Now we are ready to suggest to approximate the marginal distribution of $X(t)$'s by the following mixture of normal distributions

$$\begin{aligned} & \frac{41}{290}N(\hat{\mu}_{(1)}, \hat{\sigma}_0) + \frac{41}{290}[N(\hat{\mu}_{(2)}, \hat{\sigma}_1) + \cdots + N(\hat{\mu}_{(4)}, \hat{\sigma}_1)] + \\ & + \frac{42}{290}[N(\hat{\mu}_{(5)}, \hat{\sigma}_1) + N(\hat{\mu}_{(6)}, \hat{\sigma}_1)] + \frac{42}{290}N(\hat{\mu}_{(7)}, \hat{\sigma}_0) \end{aligned} \tag{6.2}$$

where $\hat{\mu}_{(j)}$ is our estimator of average of the day j ($j = 1$ – Sunday, $j = 7$ – Saturday), while $\hat{\sigma}_0$ and $\hat{\sigma}_1$ are our estimators for standard deviations of working days and of weekends respectively.

Table 4: *****Please add caption ***

| $\hat{\mu}(1)$ | $\hat{\mu}(2)$ | $\hat{\mu}(3)$ | $\hat{\mu}(4)$ | $\hat{\mu}(5)$ | $\hat{\mu}(6)$ | $\hat{\mu}(7)$ | $\hat{\sigma}_0$ | $\hat{\sigma}_1$ |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|------------------|------------------|
| 10.76 | 13.70 | 15.69 | 16.80 | 15.17 | 13.69 | 11.77 | 1.73 | 4.20 |

The histogram of $X(t)$'s along with approximation derived from (6.2) is shown on Figure 16. It may look not much better than Figure 15 to some eye, but it would be quite wrong impression.

Indeed, the next Figure 17, which shows empirical distribution function of $X(t)$, instead of histogram, along with approximating distribution function of (6.2) exhibits very good agreement.

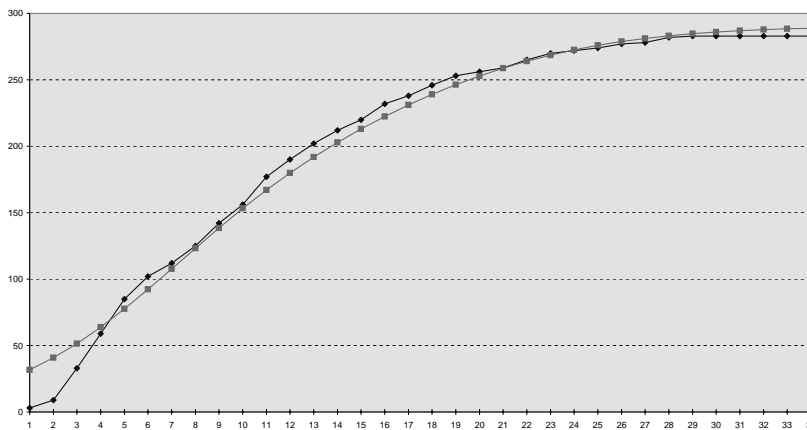


Figure 17: Empirical distribution of $X(t)$'s and mixture (6.2) of normal distributions

Remark 6. We are touching here very important question of how and in what terms should we measure a goodness of approximation of the distribution of $X(t)$'s. We will see in the next Section 7, that goodness of fit of distribution function is more natural criteria than anything else and, may be, the only natural criteria if we are interested in accurate estimation of net prices on options.

8. Net Prices on Options

An average loss associated with an option on random price X with strike k is equal to the expected value

$$v(k) = E(X - k)^+ = \int_k^\infty (x - k)f(x)dx \quad (8.1)$$

where $(X - k)^+$ denotes positive part of $X - k$ and $f(x)$ denotes the density of the random variable X . Using integration by parts we can rewrite (8.1) as

$$v(k) = \int_k^\infty [1 - F(x)]dx \quad (8.2)$$

which shows quite clearly that this loss is rather “smooth” functional of the distribution function $F(x)$. It is also clear that the value of $v(k)$ is just the net price on the option under the “physical” measure.

The density function $f(x), x \geq 0$, the distribution function $F(x), x \geq 0$ and the net price $v(k), k \geq 0$ are all in one-to-one relationship to each other: if we know one of them we also know two others.

Statistically, however, the problems of estimation of $f(x), x \geq 0$ and $F(x), x \geq 0$ are very different. It is difficult to estimate f , estimators \hat{f}_n are typically unstable and large sample sizes are required. One of the most common density estimators is a histogram. It is true that visually histograms can be very appealing. However, they are rather unstable and the way they look depends very much, in particular, on the choice of intervals of grouping. Estimation F_n of F is, in most cases, is straightforward and has very good statistical properties. As the expression (8.2) suggests, we do not need really to estimate the density but only need a good estimator of the distribution function. In our case, as in the most of other cases, this will be an empirical distribution function F_n :

$$F_n(x) = \frac{1}{n} \sum_{t=1}^n I_{\{X(t) \leq x\}} \quad \text{or} \quad 1 - F_n(x) = \frac{1}{n} \sum_{t=1}^n I_{\{X(t) \geq x\}} \quad (8.3)$$

In the previous section we have seen that this empirical distribution function agrees well with a mixture of normal distributions. Hence, the integrals (8.2) from this mixture should provide good statistical estimator of $v(k)$.

This statement should be made with certain reservations, however. The reason for this reservation lies in the fact that $v(k), k \geq 0$, is still not a continuous functional of $F(x), x \geq 0$, that is, very small changes of F on the right tail may change the values of $v(k)$ quite substantially. In other words, the presence of outliers will have very small influence on empirical distribution function, but may change the integral

$$v_n(k) = \int_k^\infty [1 - F_n(x)]dx \left(= \frac{1}{n} \sum_{t=1}^n (X(t) - k)^+ \right) \tag{8.4}$$

quite seriously.

Before we come to the last numerical illustration we want to make two remarks. The histogram of $X(t)$'s as we have seen in §6, is rather irregular in shape. It was, to some extent, the matter of sport and of eventual justification of the models and methods, that we undertake its approximation through simpler distributions. It is also important as soon as we consider options on a weekly average prices. However, having only options on daily averages in mind, it would be possible to consider distributions for each day of a week separately.

If we consider sequence of a given weekday, say, sequence of Mondays, it practically will be a sequence of i.i.d. random variables, because the daily average prices 7 day apart are practically independent. If, however, F_n is based on i.i.d. observations, then for all λ we know the probability

$$\mathbf{P}\{\max_x [F(x) - F_n(x)] \leq \lambda\}.$$

Hence we know the probability

$$\mathbf{P}\{1 - F(x) \geq 1 - F_n(x) - \lambda, \text{ for all } x > 0\}.$$

Therefore, the function $1 - F_n(x) - \lambda$, truncated at 0, forms a confidence boundary for $1 - F(x)$ for all x simultaneously. But then the integral

$$\int_k^\infty \max[1 - F_n(x) - \lambda, 0]dx$$

gives the lower confidence boundary for $v(k)$ for all k simultaneously. This is, of course, a very convenient object to have.

We already know, that each distribution F , in its main body, can be approximated by a normal distribution. But for the normal distribution the integral in (8.1) can be explicitly calculated

$$v(k) = v(k, \mu, \sigma) = \sigma^2 \varphi(k, \mu, \sigma) + (\mu - k)[1 - \Phi(k, \mu, \sigma)] \tag{8.5}$$

where $\varphi(k, \mu, \sigma)$ and $\Phi(k, \mu, \sigma)$ denote normal density and normal distribution function with expected value μ and standard deviation σ respectively.

The last graph shows empirical $v_n(k)$ for Mondays, the days when the number of outliers was small. The agreement between $v_n(k)$ and $v(k, \hat{\mu}, \hat{\sigma})$ is quite good.

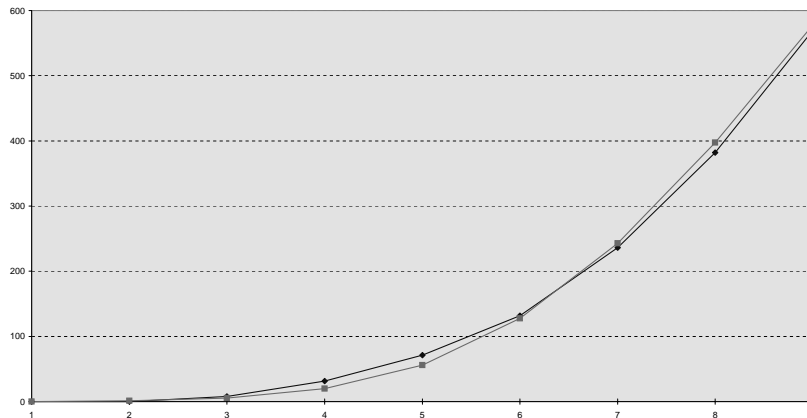


Figure 18: Mondays: empirical net price $v_n(k)$ and its normal approximation

References

- Beran, J. (1994). *Statistics for long-memory processes*. Monographs on Statistics and Applied Probability, **61**. Chapman and Hall.
- Berkes, I., Gombay, E., Horvath, L., Kokoszka, P. (2004). Sequential change-point detection in GARCH(p, q) models. *Econometric Theory* **20**, 1140-1167.
- Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis, Forecasting and Control*, 2nd edition. Holden-Day.
- Brodski, B. E. and Darkhovski, B. S. (2000). *Non-Parametric Statistical Diagnosis* Kluwer Academic.
- Chan, K. S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21**, 520-533.
- Chow, Y. S., Robbins, H., Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Hughton Mifflin.
- D'Agostino, R. B. and Stephens, M. A. (Eds.) (1986) *Goodness-of-Fit Techniques*, 421-459. Marcel Dekker. (***** Is this a chapter of a book? More detail please *****)
- Horvth, L., Huškova, M., Kokoszka, P. and Steinbach, J. (2004). Monitoring changes in linear models. *J. Statist. Plann. Inference* **126**, 225-251.
- Koul, H. L. and Surgailis, D. (2002). Asymptotic expansion of the empirical process of long memory moving averages. An invited review article for the book *Empirical Process Techniques for Dependent Data* (Edited by Dehling, H. G., Mikosch, T. and Sørensen, M.), 213-239, Birkhäuser.

-
- Musiela, M., Rutkowski, M. (2001). *Martingale Methods in Financial Modelling*. Springer Verlag.
- Qian, L. (1998). On maximum likelihood estimators for a threshold autoregression. *J. Statist. Plann. Inference* **75**, 21-46.
- Shiryaev, A. N. (1999). *Essentials of Stochastic Finance. Facts, Models, Theory*. World Scientific.
- Shorack, G., Wellner, J. A. (1986). *Empirical processes* J. Wiley and Sons.

Received November 17, 2005; accepted April 14, 2006.

Estate V. Khmaladze,
School of Mathematics, Statistics and Computer Science
Victoria University of Wellington
P.O. Box 600
Wellington, New Zealand
Estate.Khmaladze@mcs.vuw.ac.nz