Markov Switching Model

Vladimir Pozdnyakov

University of Connecticut

2006

Based on joint work with Dipak Dey, Joseph Glaz and Nalini Ravishanker

Outline

- Model
- Hypothesis testing
- Estimation
- Scan connection
- Coalition Casualties Data

Introduction

A Markov switching model with a binary switch is proposed. We develop a likelihood ratio test when the underlying distributions are coming from general exponential family. Several examples are considered and the expressions for the likelihood ratio functions are obtained for each of those examples. Maximum likelihood estimation procedure is discussed in detail. Finally, our methods are illustrated with the modelling and analysis of data of coalition casualties in Iraq.

The null hypothesis

Consider a collection of (discrete or absolutely continuous) random variables $\{X_{ij}\}$ observed on a rectangle

$$[1, N] \times [1, M] = \{(i, j) | 1 \leq i \leq N, 1 \leq j \leq M\}.$$

Under H_0 the random variables $\{X_{ij}\}$ are i.i.d with (discrete or absolutely continuous) density f_0.

The alternative hypothesis: Definitions

Area of anomaly A is any subset of the rectangle $[1, N] \times [1, M]$.

Left neighborhood LN_{ij} of point (i, j) is any subset of $[1, i-1] \times [1, M]$.

For any left neighborhood of (i, j), LN_{ij}, we define *switch* s_{ij} as a binary (0 or 1) function of observations X_{kp} for $(k, p) \in LN_{ij}$. We say that $(k, p) < (i, j)$, if $k < i$ or $k = i \cap p < j$.

The alternative hypothesis: Definitions

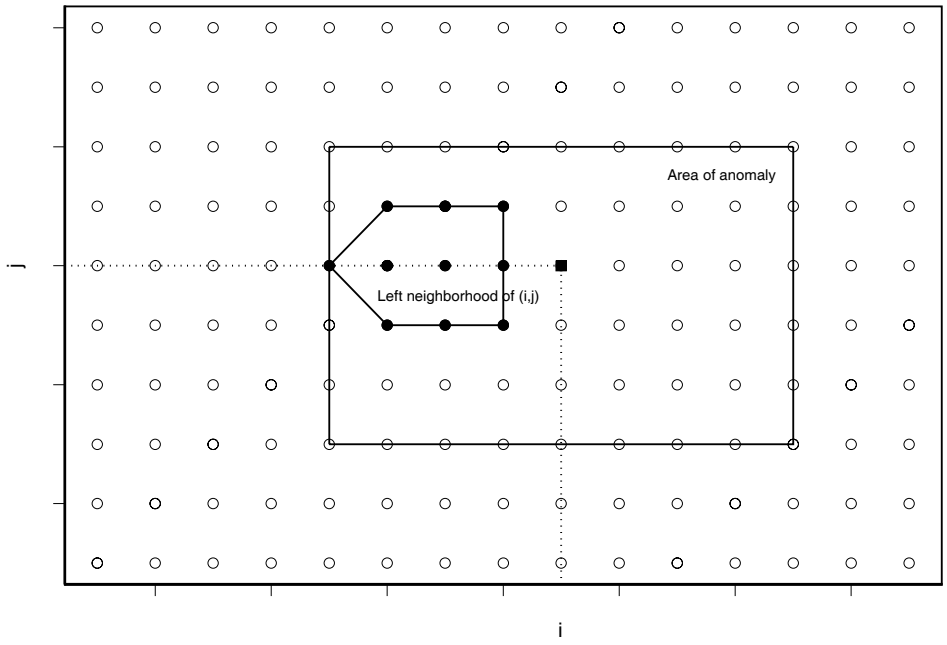

Figure 1: Area of anomaly with left neighborhood of (i, j)

The alternative hypothesis: Assumptions

Following assumptions are needed to develop the likelihood ratio test.

Assuption 1 *The shape of left neighborhoods LN_{ij} is the same for all the points (i, j) from the area of anomaly A.*

Assuption 2 *The area of anomaly A is such that $LN_{ij} \subset [1, N] \times [1, M]$ for all $(i, j) \in A$.*

Assuption 3 *The switch s_{ij} as a function of random variables is the same for all (i, j).*

The alternative hypothesis: Markov dependence

Under H_1 we have Markov type dependence. The distribution of X_{ij} for $(i,j) \in A$ is fully determined by values of observations in the LN_{ij}. If the switch $s_{ij} = 0$ then X_{ij} has (discrete or absolutely continuous) density f_0, if $s_{ij} = 1$ then X_{ij} has density f_1.

More formally, for any $(i,j) \in A$ the conditional density of X_{ij} is given by

$$\begin{aligned} f\left(x_{ij} \Big| \bigcap_{(k,p)<(i,j)} X_{kp} = x_{kp}\right) &= f\left(X_{ij} = x_{ij} \Big| \bigcap\nolimits_{(k,p)\in LN_{ij}} X_{kp} = x_{kp}\right) \\ &= \begin{cases} f_0(x_{ij}), & \text{if } s_{ij} = 0, \\ f_1(x_{ij}), & \text{if } s_{ij} = 1 \end{cases} \\ &= f_0(x_{ij})(1 - s_{ij}) + f_1(x_{ij}) s_{ij} \end{aligned}$$

Likelihood Ratio

Mainly because the switch is a binary function, the expression for likelihood ratio is relatively simple and tractable.

Lemma 1 *The likelihood ratio for testing H_0 vs H_1 is given by*

$$R(x) = \prod_{(i,j)\in A} \left[\frac{f_1(x_{ij})}{f_0(x_{ij})}\right]^{s_{ij}}. \tag{1}$$

Likelihood Ratio: Exponential Family

This expression for the likelihood ratio is particularly convenient for distributions from the general exponential family (see Brown (1986)).

Assuption 4 *Assume that* $f_i(\cdot), k = 1, 2$ *are from a distribution that belongs to the general exponential family , i.e.,*

$$f_k(x) = \exp[\theta_k t(x) - \psi(\theta_k)],$$

where $\theta_k \in \Theta = \{\theta_k : \int \exp(\theta_k x)\nu(dx) < \infty\}$, *natural convex parameter space.*

Likelihood Ratio: Exponential Family

Lemma 2 *Under Assumptions 1–4 the loglikelihood ratio is given by*

$$\log R(x) = (\theta_1 - \theta_0)\langle S \cdot T\rangle - (\psi(\theta_1) - \psi(\theta_0))S, \tag{2}$$

where

$$S = \sum_{(i,j)\in A} s_{ij}, \tag{3}$$

$$T = \sum_{(i,j)\in A} t(x_{ij}), \tag{4}$$

and

$$\langle S \cdot T\rangle = \sum_{(i,j)\in A} s_{ij} \cdot t(x_{ij}). \tag{5}$$

The statistic S just counts how often the switch is activated. The statistic $\langle T \cdot S\rangle$ adds up $t(X_{ij})$ for (i, j) which left neighborhood triggers the switch.

Likelihood Ratio: Examples

- Geometric distribution:

$$f_k(x_{ij}) = p_k(1-p_k)^{x_{ij}}, \quad x_{ij} = 0,1,2,..., \text{ where } 0 < p_k < 1, \quad k = 1,2,$$

$$\log R(x) = (\log(1-p_1) - \log(1-p_0))\langle S \cdot T \rangle + (\log p_1 - \log p_0)S.$$

- Exponential distribution:

$$f_k(x_{ij}) = \lambda_k e^{-\lambda_k x_{ij}}, \quad x_{ij} > 0, \text{ where } \lambda_0, \lambda_1 > 0, \quad k = 1,2,$$

$$\log R(x) = (\lambda_0 - \lambda_1)\langle S \cdot T \rangle + (\log \lambda_1 - \log \lambda_0)S.$$

- Poisson distribution:

$$f_k(x_{ij}) = e^{-\lambda_k}\frac{\lambda_k^{x_{ij}}}{x_{ij}!}, \quad x_{ij} = 0,1,2,... \text{ where } 0 < \lambda_0, \lambda_1 < 1, \quad k = 1,2,$$

$$\log R(x) = (\log \lambda_1 - \log \lambda_0)\langle S \cdot T \rangle + (\lambda_0 - \lambda_1)S.$$

Likelihood Ratio: Examples

- Bernoulli distribution:

$$f_k(x_{ij}) = p_k^{x_{ij}}(1-p_k)^{1-x_{ij}}, \quad x_{ij} = 0,1, \quad k = 1,2,$$

$$\log R(x) = \log\left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right]\langle S \cdot T\rangle + \log\left[\frac{1-p_1}{1-p_0}\right] S.$$

- Binomial distribution:

$$f_k(x_{ij}) = C_{x_{ij}}^n p_k^{x_{ij}}(1-p_k)^{1-x_{ij}}, \quad x_{ij} = 0,1,2,..., \quad k = 1,2,$$

$$\log R(x) = \log\left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right]\langle S \cdot T\rangle + \log\left[\frac{1-p_1}{1-p_0}\right] S.$$

- Normal distribution (the same σ):

$$f_k(x_{ij}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x_{ij}-\mu_k)^2}{2\sigma^2}\right], \quad \text{where } \sigma > 0, \quad k = 1,2,$$

$$\log R(x) = \left[\frac{\mu_1-\mu_2}{\sigma^2}\right]\langle S \cdot T\rangle + \left[\frac{\mu_0^2-\mu_1^2}{2\sigma^2}\right] S.$$

Asymptotic Normality under H_0

Lemma 3 *Let $X = \{X_{ij}\}_{(i,j)\in[1,N]\times[1,M]}$ and $E_0(\cdot)$ be the expected value under H_0 then*

$$E_0(\log R(X)) = |A|P_0(s_{ij} = 1)[\log A + E_0(t(X_{ij}))\log B], \tag{6}$$

where $|\cdot|$ is the cardinality of a set.

Normality is established here for the case $M = 1$, but it can easily be generalized for any fixed $M \geq 1$.

Proposition 1 *Let $M = 1$ and $A = [1, 1] \times [K + 1, N]$ where $K = |LN_{i1}|$. Then as $N \to \infty$*

$$\frac{\log R(X) - E_0(\log R(X))}{\sigma\sqrt{N}} \to \mathcal{N}(0, 1)$$

in distribution, where

$$\sigma^2 = Var_0(\xi_{K+1}) + 2\sum_{j=1}^{K+1} Cov_0(\xi_{K+1}, \xi_{K+1+j}),$$

and

$$\xi_j = s_{ij}[(\theta_1 - \theta_0)t(X_{j1}) - (\psi(\theta_1) - \psi(\theta_0))].$$

Maximum Likelihood Estimation

By multiplying the likelihood ratio (1) by the term $\prod_{(i,j)\in[1,N]\times[1,M]} f_0(x_{ij})$ and then taking logarithm we get the log-likelihood function:

$$l(\theta_0,\theta_1) = [\theta_1-\theta_0]\langle S\cdot T\rangle + [\psi(\theta_0)-\psi(\theta_1]S + \theta_0 \sum_{(i,j)\in[1,N]\times[1,M]} t(x_{ij}) - MN\psi(\theta_0). \qquad (7)$$

Taking derivatives we obtain the following expressions for the estimates $\widehat{\theta}_0$ and $\widehat{\theta}_1$:

$$\psi'(\widehat{\theta}_1) = \frac{\langle S\cdot T\rangle}{S}, \qquad (8)$$

and

$$\psi'(\widehat{\theta}_0) = \frac{\sum_{(i,j)\in[1,N]\times[1,M]} t(x_{ij}) - \langle S\cdot T\rangle}{NM-S}. \qquad (9)$$

If the area of anomaly A almost coincides with the rectangle $[1,N]\times[1,M]$ then

$$\sum_{(i,j)\in[1,N]\times[1,M]} t(x_{ij}) \approx T(X).$$

Maximum Likelihood Estimation

In the case of distributions from the exponential family we have separation of the parameters. It will simplify significantly the maximization of the likelihood function. For any given set of parameters that defines switch s_{ij} we have expressions for optimal values of θ_0 and θ_1. Therefore, quite often only discrete optimization with respect to switch parameters is needed.

Asymptotic normality of MLE can be established by similar tools, but instead of a stationary m-dependent sequence, here we have stationary α-mixing.

Scan Connection

Consider the following switch:

$$s_{ij} = \begin{cases} 1, & \text{if } \sum_{(k,p)\in LN_{ij}} X_{kp} \geq L, \\ 0, & \text{if } \sum_{(k,p)\in LN_{ij}} X_{kp} < L, \end{cases} \tag{10}$$

where $L > 0$ is a threshold parameter. This type of switch is closely related to the monitoring via scanning procedure. To make our decision—which distribution will be used to generate next observation—we have to *scan* the left neighborhood of the X_{ij} and basically compute a scan statistic for the neighborhood. Therefore, it is reasonable to expect that some tests based on scans can be used.

Scan Connection

To be more specific let us consider an one-dimensional Bernoulli model. That is, under H_1 for $i > k$ we have

$$P(X_{i1} = x_{i1}) = \begin{cases} p_1^{x_{i1}}(1-p_1)^{1-x_{i1}}, & \text{if } \sum_{i-K\le k\le i-1} X_{k1} \ge L, \\ p_0^{x_{i1}}(1-p_0)^{1-x_{i1}}, & \text{if } \sum_{i-K\le k\le i-1} X_{k1} < L, \end{cases} \tag{11}$$

where $0 < p_0 < p_1 < 1$ and p_1 is relatively small.

Thus under H_1 we deal with a Kth order two-state Markov chain, where K is the width of the window. In the beginning the Markov process is in a *common* regime when the probability of 1 (failure) p_0 is small. If by a pure accident the number of failures in the window of K consecutive observations exceeds the threshold level θ, the process self-excites and goes to an *excited* regime when probability of failure is higher. Since p_1 is still small after a while the number of failures in the window of length K will fall down below the threshold level, and the process will switch back to the common regime.

Scan Connection

Assume that our task is to determine whether we have an i.i.d. Bernoulli sequence or this self-exciting Markov process. The best way of doing this is to use the log-likelihood ratio test. However, the construction of such test requires the exact knowledge of the parameters of underling model under the alternative. To compute the log-likelihood ratio we need to know p_0, p_1, K, and θ. What if some part of this information is missing? What if we just know that it could be self-exciting Markov process, we have some idea about dependence order, and that is all?

It seems that scan statistics can be employed. Indeed, whenever we have a cluster of failures under the alternative there is a good chance that by switching to the excited regime this cluster will be *reinforced*, which will lead to a higher value of a scan statistic. Thus, the scan test might be a reasonable nonparametric alternative to the likelihood ratio test. However, the simulation study that we ran shows that the likelihood test is quite robust, and it outperforms or does as well as the scan test even when some parameters are misspecified.

Scan Connection: an Example

Here under H_0 we have 1000 observations of Bernoulli random variables with $p_0 = .02$. For the scan test we move a window of width $K = 50$. The rejection rule is: the scan statistic is more than 5. The significant level of this test is about .04 (.039556).

The likelihood ratio test assume that under H_1 the size of the left neighborhood (or the dependence order), K, is 20, the threshold level, θ, is 2, and $p_1 = .1$. The critical value of the test (.15) that corresponds $\alpha = .04(.0396)$ is obtained via 10000 simulations. The critical values of the "correct" likelihood tests are also obtained via simulations. The simulated power (10000 simulations) is presented in Table 1.

Scan Connection: an Example

Misspecified parameters	Scan test $W = 50$	Power $L(X)$ based on $p_0 = .02$ $p_1 = .1$ $K = 20$, $\theta = 2$	Correct $L(X)$
None	.7607	.8674	.8674
$K = 10$	.2906	.3387	.4746
$K = 15$	.5569	.6602	.7017
$K = 25$	.8868	.9112	.9456
$K = 30$	.9486	.9480	.9815
$p_1 = .06$	.3994	.5217	.5428
$p_1 = .08$	.6123	.7475	.7325
$p_1 = .12$	.8524	.9191	.9220
$p_1 = .14$	.9116	.9527	.9673
$L = 3$	.2113	.1942	.2811
$L = 1$	.9989	.9966	1

Table 1: Power Comparison – Robustness

Coalition Casualties Data

Our goal is to model daily coalition casualty data from time period: January 2004 to March 2006. We did not include 2003 in our consideration because it was very "unusual" year. First it was active combat operations, then summer 2003 was very calm. But in the fall the insurgency started to pick up power. In view of that we decided to consider a time period when the "process" became stationary. Our goal is to develop a simple and tractable model incorporating the observed data which can also give some policy making recommendations.It is to be noted that the switching model can be translated into a two color code system. In practice, multiple color coded systems are used by the homeland security. However we have observed that often only two colors are used for warning the citizens.

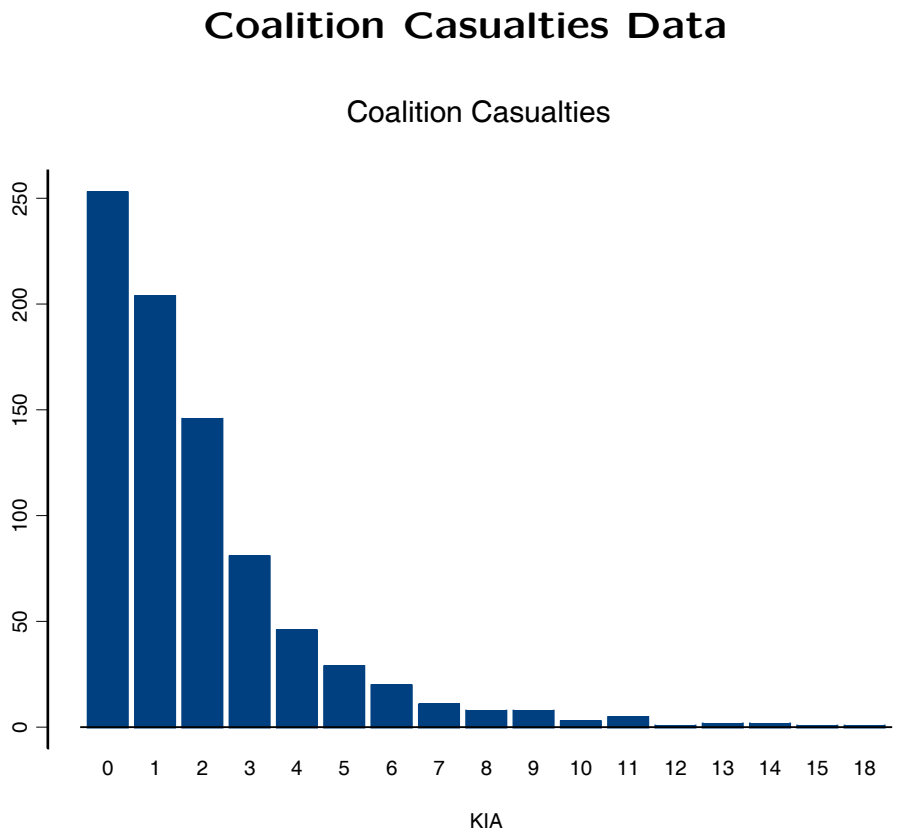

Figure 2. Coalition casualties, KIAs from Jan 1, 2004 to Mar 31, 2006

N=821, daily average number of KIAs is 1.93

Coalition Casualties Data

The total number of observations N is 821. The total number of KIAs (Killed in Actions) is 1585. Note that i.i.d. observations from the geometric, Poisson or exponential distribution with mean 1.93 is not a good fit for that data. Simple χ^2 test rules out both Poisson and exponential distributions. The shape of the histogram in Figure 2 is consistent with the geometric distribution with mean 1.93. However, it cannot be the i.i.d. sequence, because the time series is clearly not "stationary" in mean. For instance, one-way ANOVA for groups of consecutive quarters produces extremely low p-value. The geometric distribution to explain the shape of the histogram is also supported by the power law which governs many natural phenomenon. See for reference, Johnson et.al. (2006).

Coalition Casualties Data: Model

In order to explain the general pattern of the daily casualties, it seems natural to introduce some form of dependency in our model. But fitting a general Markov process is not a good idea because we do not want to have too many parameters. In view of that we consider fitting the data with a Markov process that switches back and forth between two regimes: low and high terrorist activities. That is, we impose some structure on transition probabilities which allows us to decrease greatly a number of parameters in the model.

Coalition Casualties Data: Model

The exact mechanism of switching from one regime to another is very complex, so we model it by a moving sum type switch:

$$s_i = \begin{cases} 1, & \text{if } \sum_{k=i-K}^{i-1} Y_k \geq L, \\ 0, & \text{if } \sum_{k=i-K}^{i-1} Y_k < L, \end{cases} \tag{12}$$

where L is a threshold parameter. After preliminary analysis of daily coalition casualties data we decided to use geometric densities $f_i(\cdot)$.

Coalition Casualties Data: Analysis

For any given window width K and threshold level L formulas (8) and (9) give us optimal values of $\widehat{p}_0$ and $\widehat{p}_1$. Therefore, all we need is a discrete optimization of log-likelihood function for these values of $\widehat{p}_0$ and $\widehat{p}_1$ with respect to K and L which is easy to implement. Based on the smoothed log-likelihood function (Figure 3 gives the original log-likelihood function) we found that

$$\widehat{K} = 29, \quad \widehat{L} = 51, \quad \frac{1-\widehat{p}_0}{\widehat{p}_0} = 1.503, \quad \frac{1-\widehat{p}_1}{\widehat{p}_1} = 2.318, \quad S = 431, \quad \langle S \cdot T \rangle = 999.$$

Coalition Casualties Data: Analysis

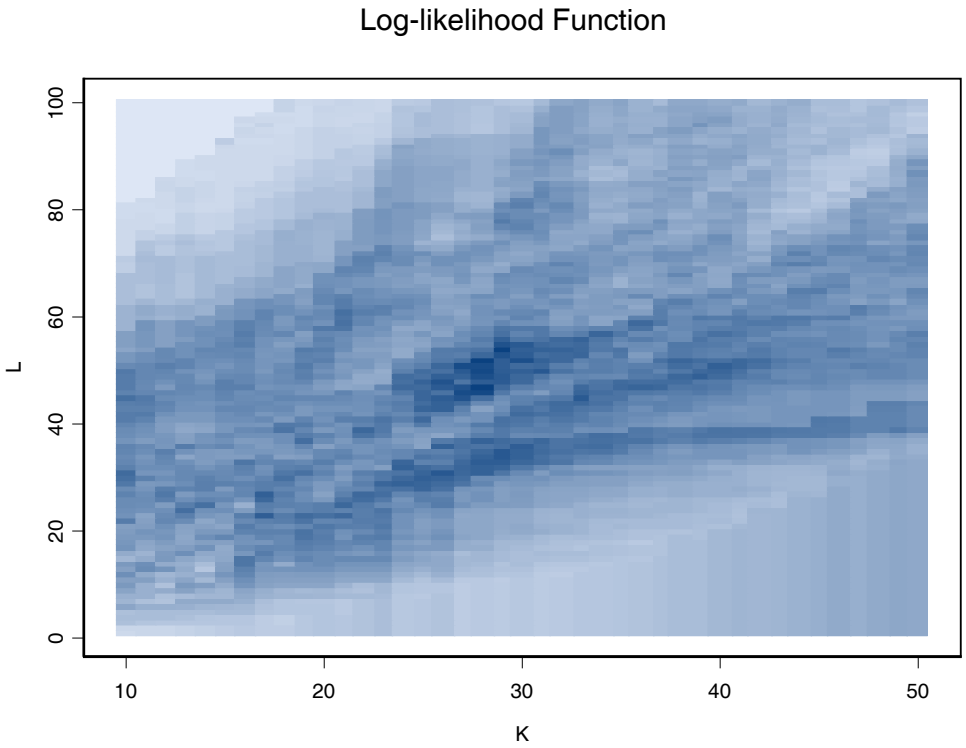

Figure 3. Log-likelihood for optimized λs

Coalition Casualties Data: Analysis

Thus if a number of KIAs for the last 29 days is above 50 we have a period of a high terrorist activity. During these periods the average number of KIAs is almost 55% higher than we are in a low terrorist activity regime. According to our model we had 431 days (which is S) of the high terrorist activity out of total 821 day (Figure 4). During this time number of KIAs—which is $\langle S \cdot T \rangle$—is 999 (out of total 1585)

Coalition Casualties Data: Analysis

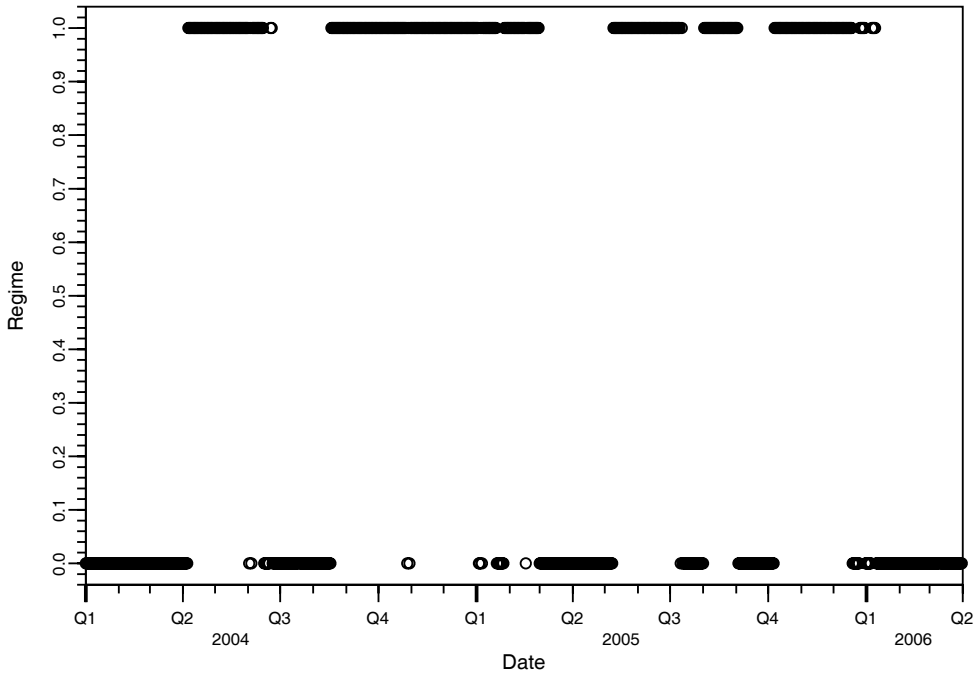

Figure 4. Regimes of high and low terrorist activities

THANK YOU