Time Series of Count Data using Multifractal Process

Jean-Philippe Boucher[‡] Donatien Hainaut[†]

April 22, 2015

[‡] *Quantact / Département de mathématiques, UQAM. Montréal, Québec, Canada.*

Email: `boucher.jean-philippe@uqam.ca`

[†] *ESC Rennes Business school - CREST. France.*

Email: `donatien.hainaut@esc-rennes.fr`

Abstract

This work generalizes the multifractal process of Calvet & Fisher (2008), initially used to model the volatility component of autoregressive data, to counting processes. We include the multifractal process in the mean structure of the Poisson distribution to obtain a distribution that can be applied to time series of count data. We show that this kind of model is very flexible even if it only needs a few parameters. It is used with the Zeger's classic polio data, where we found an unobserved periodic time dependence structure that cannot be captured by a Poisson distribution with an autoregressive component.

Key Words: Seasonality, Time Series, Count Data, Multifractal process, Maximum Likelihood, Poisson.

1 Introduction

Time series of counts is a vast research area in statistics (see McKenzie(2003) or Jung & Tremayne (2006) for excellent reviews). Models for count data observed over time can be classified in many areas; a non-exhaustive list covers Markov Chain models (Raftery, 1985), DARMA models (Jacob & Lewis 1978a, 1978b), models based on thinning operators (Al-Osh & Alzaid, 1987), state space models, and hidden-Markov models (MacDonald & Zucchini, 1997).

In this paper, we are interested in the modeling of time series of count. We propose using a generalization of Calvet & Fisher's (2008) multifractal process to model the unobserved latent factor of a time series of counts. The interpretation of the model is interesting because its underlying process can represent latent unobserved factors (e.g. climatic, economic or social).

The word *fractal* emerged on the scientific scene with the work of Mandelbrot (1982) in the 1960s and 1970s. Subsequently, multifractal processes became a popular mean of modeling financial time series. We refer the interested reader to the numerous publications of Mandelbrot (e.g. 1997 and 2001), for applications of these processes to finance. To our knowledge, this approach has not been exploited in a non-Gaussian framework.

We illustrate our work using polio data, a classic data set for time series of counts, first used by Zeger (1988), and later by Chan & Ledolter (1995), Kuk & Cheng (1997), Oh & Lin (2001), Jung & Liesenfeld (2001), and Farrell et al. (2007). We show that the fit of this new model is interesting and can be advantageously compared to the Poisson-AR(1) model. For example, unlike the Poisson-AR(1) model, the multifractal count model of this paper can be estimated directly, without requiring simulations. A formal comparison of our approach with the Poisson-AR(1), illustrates major differences between models, and shows that the multifractal count distribution captures an unobserved time dependence structure, not present in the other model.

In Section 2, the multifractal process is described and used with a Poisson distribution. The inference technique using the Hamilton filter is then explained, and the basic properties of the model are explored. A numerical application with Zeger's dataset is presented in Section 3. The last section concludes the paper.

2 Multifractal Process Modeling

Models of time series of counts can be classified in many ways. Cox (1981) proposed separating the models into two families: observation-driven models and parameter-driven models. More recently, Jung & Tremayne (2011) carried out a detailed overview of time series models for counts, identifying several categories of models, such as static regression models, autoregressive conditional mean models, integer autoregressive models, and generalized linear autoregressive models (GLARMA). In their classification, a Poisson process with an autoregressive intensity is a parameter-driven model called the Poisson Stochastic Autoregressive Mean (Poisson-SAM).

We propose a counting process that can model overdispersion and time dependence exhibited by some time series. For this reason, we consider a Poisson random process, N_t, which has the following stochastic intensity for the period of time $[t, t+1)$:

$$\lambda(t) = d_t \exp(\beta^\top x_t) F_t = \tau(t) F_t, \tag{2.1}$$

where the vector x_t is used to include covariates, F_t is a multifractal process, and d_t is the exposure. An example of covariates is presented in the numerical application of Section 3. The random process F_t is used here to introduce overdispersion and time dependence. This approach is directly inspired by the binomial multifractal process used to model volatility in a Gaussian time series, as studied by Calvet & Fisher (2008). In our framework, the number of events observed on the time interval $[t, t+1)$ has the following probability function:

$$\Pr\left(N_t = n\right) = \frac{\lambda(t)^n e^{-\lambda(t)}}{n!}. \tag{2.2}$$

The process F_t is the product of m random factors that may be climatic, economic, or social, for example. Those factors are unobservable and are modeled by a Markov state vector, $\overline{M}_t$, of m components:

$$\overline{M}_t = (M_{1,t}\,, M_{2,t}\ \ldots M_{m,t}) \in \mathbb{R}_+^m.$$

The process F_t is the product of those factors:

$$F_t \;=\; \prod_{j=1}^{m} M_{j,t} \tag{2.3}$$

Each process $M_{j,t}$, $j = \{1, \ldots, m\}$, is built in a recursive manner. Let us assume that the vector of $M_{j,t-1}$ exists up to period $t-1$. For each $j = \{1, \ldots, m\}$, the next period $M_{j,t}$ is drawn from a fixed distribution $M(j)$ with probability γ_j, and is otherwise equal to its previous value $M_{j,t} = M_{j,t-1}$. It can be expressed as:

$$M_{j,t} = \begin{cases} M_{j,t-1} & \text{with probability } 1-\gamma_j \\ M(j) & \text{with probability } \gamma_j \end{cases}. \tag{2.4}$$

The random variable $M(j)$ is a simple binomial variable that is worth $m_{0,j}$ with probability p_0 and $2 - m_{0,j}$ with probability $1 - p_0$. Formally, we have:

$$M(j) = \begin{cases} m_{0,j} & p_0 = \frac{1}{2} \\ 2 - m_{0,j} \equiv m_{1,j} & 1 - p_0 = \frac{1}{2} \end{cases}. \tag{2.5}$$

The $m_{0,j=1\ldots m} \in (0, 1)$ are parameters to be estimated and p_0 is set to $\frac{1}{2}$. If the underlying Markov process $M_{j,t}$ equals $2 - m_{0,j}$, it will increase the intensity rate. Conversely, if $M_{j,t} = m_{0,j}$, the intensity of the process will be reduced.

Basic properties of the model can be computed. Using (2.5), we have:

$$\mathbb{E}(M_{j,t}) = 1, \tag{2.6}$$

and:

$$\mathbb{E}(M_{j,t}^2) = \frac{m_{0,j}^2 + m_{1,j}^2}{2}. \tag{2.7}$$

More generally, we can easily show that:

$$\mathbb{E}(M_{j,t}^q) = \frac{m_{0,j}^q + m_{1,j}^q}{2}. \tag{2.8}$$

By construction, the parameter γ_j represents the probability that the factor $M_{j,t}$ will change its value. If γ_j is inversely proportional to j (we will impose this relation in the inference), the last factor $M_{m,t}$ changes its value less frequently than the first factor $M_{1,t}$. This approach allows us to capture low-valued regime shifts and long volatility cycles of the counting process. We will revisit this point below.

2.1 Inference

As mentioned earlier, the number of events, N_t, on the time interval $[t, t+1]$ is distributed as a Poisson random variable of intensity $\lambda_t = \tau(t)F_t$ where $\tau(t) = d_t \exp(\beta^\top x_t)$ and $F_t = \prod_{i=1}^m M_{i,t}$. The parameters x_t, d_t and β are respectively a vector of covariates, the exposure and the vector of coefficients coupled with the covariates. The underlying multifractal process F_t can take $d = 2^m$ values that are noted $s_1, ...s_d$. Each of these values corresponds to a combination of processes $M_{j,t}$, $t = 1, ..., T$. Consequently, the vector F_t is a Markov chain with a transition matrix $A = (a_{i,j})_{1\leq i,j\leq d}$. The matrix A is fully determined by the $\gamma_{j=1...m}$ and has the following components:

$$\begin{aligned} a_{x,y} &= Pr(F_t = s_x \,|\, F_{t-1} = s_y) \\ &= \prod_{j=1}^m \left(\gamma_j \frac{1}{2} + (1-\gamma_j)\mathbb{I}_{(M_{j,t}=M_{j,t-1})} \right). \end{aligned} \tag{2.9}$$

The factors $(M_{j,t})_{j=1,\ldots,m}$ are not directly observable, but the filtering technique developed by Hamilton (1989) and inspired by Kalman (1960) filter allows us to retrieve the probabilities of being in a state given all the previous observations.

Indeed, if we note as $n_{i=0,\ldots,t-1}$ the total number of events observed in the previous time period, for a given observed number of events n_t, the likelihood vector is defined as:

$$p(t, n_t, x_t) \;=\; \begin{pmatrix} Pr(N_t = n_t \,|\, F_t = s_1,\; x_t, d_t) \\ \vdots \\ Pr(N_t = n_t \,|\, F_t = s_d,\; x_t, d_t). \end{pmatrix}$$

We then define the probabilities of being in a certain state j as:

$$\Pi_t^{(j)} \;=\; Pr\left(F_t = s_j \,|\, n_1, \ldots, n_t,\; x_t, d_t\right)$$

The Hamilton filter allows us to calculate recursively the vector $\Pi_t = \left(\Pi_t^{(j)}\right)_{j=1,\ldots,d}$ as a function of the probabilities of being during the previous period:

$$\Pi_t = \frac{p(t, n_t, x_t) * (\Pi_{t-1} A)}{\langle p(t, n_t, x_t) * (\Pi_{t-1} A)\;,\; \mathbf{1} \rangle}, \tag{2.10}$$

where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^d$ and $x * y$ is the Hadamard product $(x_1 y_1, \ldots, x_d y_d)$.

To start the recursion, we assume that the Markov processes have reached their stable distribution. The vector Π_0 is then set to the ergodic distribution, which is the eigenvector of the matrix A, coupled with the eigenvalue equal to 1. If we observed the count process on time T, the log-likelihood is:

$$\ln L(n_1 \ldots n_T \,| m_{0,j=1\ldots m}, \gamma_{j=1\ldots m}, \beta) = \sum_{t=1}^{T} \ln \left\langle p(t, n_t, x_t), (\Pi_{t-1} A) \right\rangle . \tag{2.11}$$

The MLE of the parameters $m_{0,1}, ..., m_{0,m}, \gamma_1, ..., \gamma_m$, and β are obtained through numerical maximization of (2.11):

$$\max_{m_{0,j=1\dots m}, \gamma_{j=1\dots m}, \beta} \ln L(n_1 \dots n_T \,|\, m_{0,j=1\dots m}, \gamma_{j=1\dots m}, \beta)$$

To reduce the number of parameters to be estimated, we parametrize the $m_{0,j=1\dots m}$ and $\gamma_{i=1\dots m}$ in a similar way to Calvet & Fisher (2008). In particular, we assume that the probability parameters γ_j are given by the function:

$$\gamma_j \equiv \gamma_1^{b^{j-1}} \quad j = 1, \dots, m. \tag{2.12}$$

where $\gamma_1 \in (0,1)$ and $b > 1$. When dealing with the modeling of the volatility component of autoregressive Gaussian data, Calvet & Fisher (2008) assumed that all $m_{0,j}$ are identical. However, empirical analyses show that differentiating $m_{0,j=1\dots m}$ often leads to a better fit of the count data. We therefore generalize the approach by supposing that $m_{0,j}$ can be different. We then adopt the following parametrization of $m_{0,j}$:

$$m_{0,j} = (m_0)^{j^c}, \tag{2.13}$$

where m_0 and c are such that $m_{0,j} \in (0,1)$. The $2m$ parameters $m_{0,1}, ..., m_{0,m}$ and $\gamma_1, ..., \gamma_m$ are then replaced by four parameters: γ_1, b, m_0 and c. Of course, other parametrizations can be used.

2.2 Moments

The multifractal process presented in this paper has several properties. Proofs of the properties listed in this section can be found in the appendix.

Proposition 2.1. *If (2.5) holds and if $M_{j,0}$ if distributed as $M(j)$, we have:*

$$\mathbb{E}(M_{j,t+1}|M_{j,t}) \quad = \quad (1-\gamma_j)M_{j,t} + \gamma_j$$

Corollary 2.2. *If $M_{j,0}$ are distributed as $M(j)$, we have:*

$$\mathbb{E}(M_{j,t+h}|M_{j,t}) \quad = \quad (1-\gamma_j)^h(M_{j,t}-1)+1$$

Proposition 2.3. *If $M_{j,0}$ are distributed as $M(j)$, we have that:*

$$\mathbb{E}(M_{j,t}M_{j,t+h}) \quad = \quad (1-\gamma_j)^h(E[M_{j,t}^2]-1)+1$$

Given that the process N_t is Poisson-distributed, we get the following proposition:

Proposition 2.4. *The expected number of events observed at time t is equal to:*

$$\mathbb{E}(N_t) \quad = \quad \lambda(t).$$

Its variance is given by :

$$\mathbb{V}(N_t) \quad = \quad \lambda(t) + (\lambda(t))^2\left(\prod_{j=1}^{m}\mathbb{E}(M_{j,t}^2) - 1\right),$$

where $\mathbb{E}(M_{j,t}^2)$ is given by equation (2.8).The covariance between N_t and N_{t+h} is as follows:

$$Cov(N_t, N_{t+h}) = \lambda(t)\lambda(t+h)\prod_{j=1}^{m}\mathbb{E}\left(M_{j,t}, M_{j,t+h}\right) - 1$$

where $\mathbb{E}(M_{j,t}, M_{j,t+h})$ is defined by equation (2.3).

Parameters	m=5 Est.	(std.err.)	m=6 Est.	(std.err.)	m=7 Est.	(std.err.)	m=8 Est.	(std.err.)
γ_1	0.074	(0.054)	0.075	(0.054)	0.075	(0.054)	0.076	(0.054)
b	5.602	(8.980)	5.241	(8.590)	5.023	(8.263)	4.873	(7.816)
m_0	0.529	(0.104)	0.529	(0.104)	0.528	(0.103)	0.528	(0.103)
β_0	0.337	(0.230)	0.328	(0.231)	0.325	(0.233)	0.323	(0.234)
β_1	-0.841	(3.102)	-0.870	(2.964)	-0.884	(2.951)	-0.895	(2.918)
β_2	0.127	(0.128)	0.123	(0.127)	0.121	(0.126)	0.121	(0.126)
β_3	-0.476	(0.152)	-0.475	(0.151)	-0.475	(0.150)	-0.475	(0.150)
β_4	0.427	(0.126)	0.424	(0.126)	0.423	(0.126)	0.423	(0.125)
β_5	-0.028	(0.123)	-0.024	(0.123)	-0.022	(0.123)	-0.022	(0.123)
c	-0.589	(0.402)	-0.661	(0.371)	-0.709	(0.349)	-0.740	(0.335)
Loglikelihood	-246.789		-246.767		-246.760		-246.755	

Table 3.1: Parameter estimates for different models of the Poisson-Multifractal

3 Empirical Illustrations Using Zeger's Data.

To apply the Poisson-Multifractal model, we use Zeger's dataset, first introduced by Zeger (1988), and used later by Chan & Ledolter (1995), Kuk & Cheng (1997), Oh & Lin (2001), Jung & Liesenfeld (2001), Benjamin et al. (2003) and Farrell et al. (2007). The data contain the monthly number of cases of poliomyelitis in the U.S from January 1970 to December 1983. The dataset comprises 168 observations. The main purpose of the analysis of the data is to determine whether the polio counts follow a decreasing time trend. To account for trend and seasonality, the following covariate vector is introduced

$$x_t = \left(1, \frac{t}{1000}, \cos\left(\frac{2\pi}{12}t\right), \sin\left(\frac{2\pi}{12}t\right), \cos\left(\frac{2\pi}{6}t\right), \sin\left(\frac{2\pi}{6}t\right)\right), \tag{3.1}$$

and the intensity function is given by $\lambda(t) = \exp(\beta^\top x_t)$.

Fitted parameters for the Poisson-Multifractal models are presented in Table 3.1 for different values of m. As we can observe, the value of m corresponding to the number of fractal components does not influence the estimates of covariates. Indeed, for $m = 5, ..., 8$, similar values of $\boldsymbol{\beta}$ are found.

3.1 Goodness-of-fit and Diagnostics

To verify the goodness-of-fit of the Poisson-Multifractal model, we first used the graphical tool developed by Davis et al. (2003). The tool is based on adapted residuals, and the probability integral transformation (PIT). Because discrete data are used, the classic PIT has to be modified. Davis et al. (2003) proposed a modification that involves adding a perturbation to the PIT to obtain a new tool called the randomized PIT. Formally, for $t = 1, ..., 168$, it can be computed as:

$$u_t = \sum_{i=1}^{n_t - 1} \Pr(i|\boldsymbol{N_{T-1}}) + U_t \Pr(n_t|\boldsymbol{N_{T-1}}) \tag{3.2}$$

where U_t is a sequence of i.i.d. uniform (0,1) random variables. If the fit of the Poisson-Multifractal model is correct, it can be shown that the u_t will be a sequence of i.i.d. uniform random variables. We can also choose $z_t = \Phi^{-1}(u_t)$ to verify the autocorrelation between random variables.

Figure 3.1 illustrates a histogram of the randomized PIT, and a QQ plot against the uniform distribution. Kolmogorov-Smirnov tests have done made to verify this uniformity hypothesis, and we concluded that we cannot reject the uniformity of the u_t for all Poisson-Multifractal models (from $m = 5$ to $m = 8$).

Figure 3.2 illustrates the autocorrelogram of the z_t for the Poisson-Multifractal ($m = 8$), with dashed lines corresponding to the upper and lower 95% bounds. All Poisson-Multifractal models obtain approximately the same form for the autocorrelagram, from which we cannot reject the independence assumption.

3.2 Model Comparison

To work with this data, previous authors often assumed a conditional Poisson distribution, with an autoregressive lognormal process in the mean function. The model, denoted

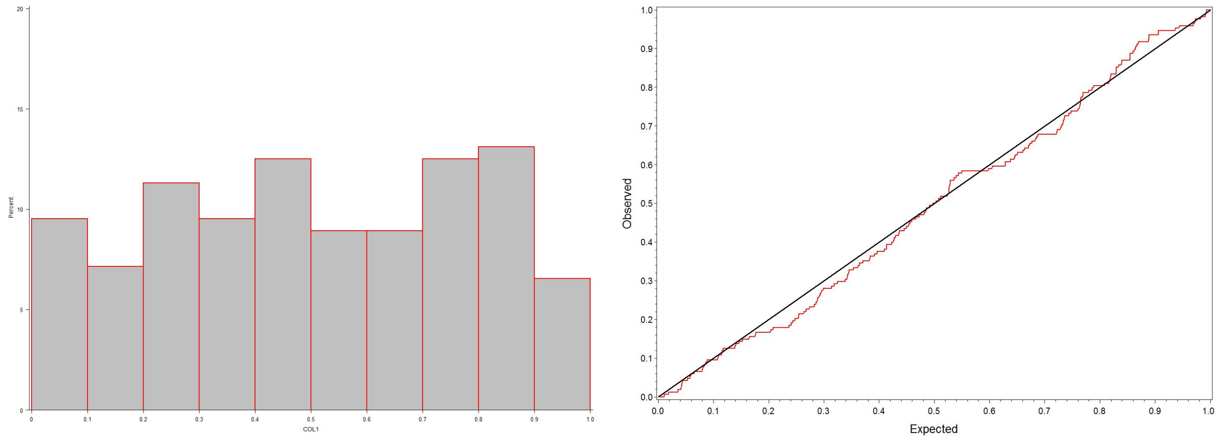

Figure 3.1: Randomized PIT and QQ plot

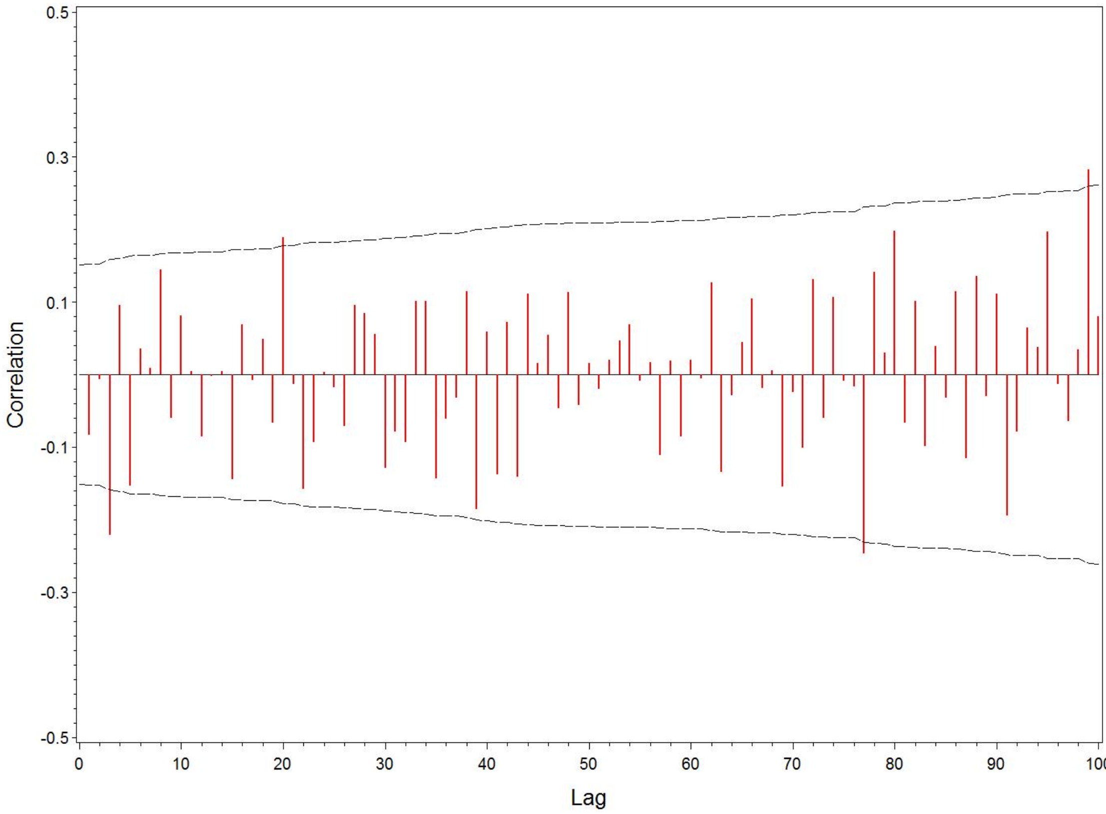

Figure 3.2: Autocorrelation function for the randomized PIT. The dashed lines represent approximate 99% confidence interval

Poisson-AR(1) or noted Poisson-SAM (stochastic autoregressive mean) (Jung et al., 2011), is described as:

$$N_t|\theta_t \sim Poisson(\lambda(t)\theta_t),\ \theta_t = \exp(W_t), \tag{3.3}$$

with:

$$W_t = \delta_1 W_{t-1} + \nu^2 \epsilon_t, , \tag{3.4}$$

where $\epsilon_t \sim N(0,1)$. To account for trend and seasonality, the same covariate vector as the Poisson-Multifractal models is used; see equation (3.1).

Many techniques for estimating the Poisson-SAM have been proposed. Zeger (1988) suggests to use the first two moments of the model by generalizing the GLM first-order conditions to add dependence between random variables. Frühwirth-Schnatter (1994) adapted a Kalman filter method to estimate the parameters. Chan & Ledolter (1995) used Monte-Carlo Expectation Maximization (MCEM), where the expectation step is estimated through Monte-Carlo simulations. Kuk & Cheng (1997) used the Monte-Carlo Newton-Raphson (MCNR) method, where classic Newton-Raphson algorithm is used with the derivatives of the log-likelihood being estimated with Monte-Carlo simulations. Durbin & Koopman (1997), using the approximated distribution of Frühwirth-Schnatter (1994), estimated the parameters through importance sampling. Jung & Liesenfeld (2001) also used an approximated distribution as part of importance sampling (a technique called efficient importance sampling or EIS), based on the minimization of the MC sampling variance. Oh & Lin (2001) and Farrell et al. (2007) adopted Bayesian techniques to estimate a Poisson-SAM model. Consequently, note that contrary to the Poisson-Multifractal model, the Poisson-AR(1) model is fitted by simulations. Procedures based on simulations are more time-consuming and involve loss of precision due to the variations in the estimates.

Parameters	Estimate	(std.err.)
δ_1	0.679	(0.149)
ν	0.507	(0.113)
β_0	0.222	(0.255)
β_1	-3.631	(2.572)
β_2	0.161	(0.132)
β_3	-0.484	(0.161)
β_4	0.415	(0.117)
β_5	-0.013	(0.113)
Loglikelihood	-248.630	

Table 3.2: Parameter estimates for the Poisson-AR(1) model of Jung & Liesenfield (2003)

Table 3.2 shows the value obtained by Jung & Liesenfeld (2001) for the Poisson-AR(1) model. We can see that the loglikelihood obtained by the Poisson-Multifractal model for $m = 8$ is better than the one obtained with the Poisson-AR(1) model. However, the Poisson-Multifractal has two more parameters than the Poisson-AR(1) model.

The randomized PIT and the corresponding Kolmogorov-Smirnov test were also used for the Poisson-AR(1). The model cannot be rejected using these tests. We obtain a similar conclusion regarding the autocorrelation between the z_t. Because the Poisson-Multifractal and the Poisson-AR(1) are non-nested, we cannot test the models directly against each other. Instead, we used various scoring rules (see Czado et al., 2009) to compare them.

A score function measures the power of prediction of the models. The scoring rule used to compare the Poisson-Multifractal and the Poisson-AR(1) models has the following form:

$$\text{Score} = \frac{1}{n}\sum_{t=1}^{168} s(n_t)$$

Many score functions $s(n_t)$ can be used. In this paper, we used three forms for the score function:

- The logarithm score (LS): $s(n_t) = -\log(\Pr(N_t = n_t))$
- The quadratic score (QS): $s(n_t) = -2\Pr(N_t = n_t) + \sum_{j=0}^{\infty}\Pr(N_t = j)^2$

Model	Log-likelihood	AIC	LS	QS	RPS
P.-Multifractal (m=5)	-246.789	513.578	1.4690	-0.2916	0.7316
P.-Multifractal (m=6)	-246.767	513.534	1.4689	-0.2918	0.7315
P.-Multifractal (m=7)	-246.760	513.520	1.4688	-0.2919	0.7315
P.-Multifractal (m=8)	-246.755	513.510	1.4688	-0.2920	0.7315
Poisson-AR(1)	-248.630	513.260	1.4762	-0.2877	0.7449

Table 3.3: LS, QS and RPS scores for all fitted model

- The ranked probability score (RPS): $s(n_t) = \sum_{j=0}^{\infty}(\Pr(N_t \leq j) - 1_{(N_t \leq j)})^2$

Table 3.3 presnts the scores for all the fitted models. We see that the scores obtained with each model are quite similar. Even if the Poisson-Multifractal models generate lower scores for LS, QS, and RPS, we cannot conclude that it is the best model to fit the data compared to the Poisson-AR(1) model. Consequently, instead of focussing solely on this scoring information, or other statistical tools, we chose to compare the models in detail in details by looking at what each model supposes and the differences between models.

3.2.1 Values of the Parameters

The β coefficients obtained with the Poisson-Multifractal differ from the β obtained by Jung & Liesenfeld (2001). In particular, the value of β_1 that corresponds to the time trend is clearly lower in the Poisson-Multifractal model. This trend is not statistically significant (other studies have also found a non-significant value for the time trend). Therefore, based on the dataset, we cannot infer a clear decrease in the number of poliomyelitis cases from 1970 to 1983. Other estimates of parameters $\boldsymbol{\beta}$ are quite similar to those reported in previous studies.

Using the results of the Poisson-AR(1) by Jung & Liesenfeld (2001), the evolution of the deterministic part of the intensity, $\lambda(t)$, of each model is compared graphically in Figure 3.3. The difference in the time trend is quite apparent, while the periodicity seems to be equivalent. The reason for this difference must be explained, for example, using smoothing techniques. We will revisit this point below.

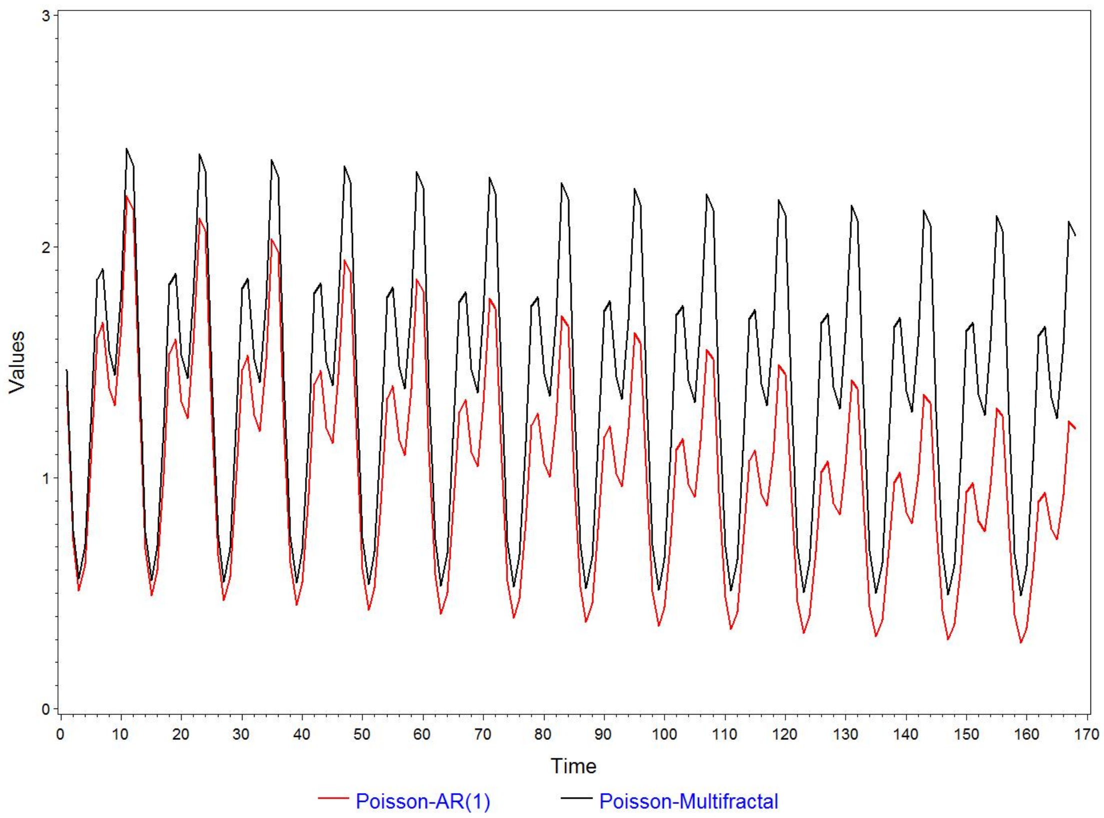

Figure 3.3: Evolution of $\lambda(t)$ for the Poisson-AR(1) model and the Poisson-Multifractal model

3.2.2 Analysis of Prior Distributions

In the Poisson-Multifractal model, the distribution of probabilities Π_t of F_t can be easily retrieved for $t = 1, ..., 168$. As recommended by Hamilton (1989), Π_0 is assumed to be equal to the ergodic distribution of the Markov processes ruling F_t. The vector Π_0 can be seen as a prior distribution of random effects that impact the counting process. In the Poisson-AR(1) model, the same type of assumption is made. In particular, Jung & Liesenfeld (2001) assume that $\theta_0 = \exp(W_0)$ is lognormal with parameters $\sigma = \sqrt{\nu^2/(1-\delta_1^2)} = \sqrt{0.507^2/(1-0.679^2)} = 0.6906$ and $\mu = -\sigma^2/2 = -0.2385$. The priors of the multifractal and Poisson-AR(1) processes have the same mean of 1. Furthermore, in the Poisson-AR(1) model, the variance of the prior $Var[\theta_0] = 0.6111$. Based on the results presented in Section 2.2, the variance of the multifractal prior has been calculated and is equal to 0.6291, which approaches that obtained for the lognormal distribution. To confirm our intuition that the prior distributions of both models are almost similar, we compare the density of priors graphically. As seen above, the multifractal process creates $d = 2^m$ possible values of F_t. For illustration, using $m = 8$ corresponds to 256 possible values of F_t. We have approached F_t through a continuous function using gamma kernel functions and plotted the result in Figure 3.4. This graph clearly shows that the shapes of the prior densities are almost identical. Because the variance and the shape of the priors are similar in the Poisson-AR(1) and multifractal models, we conclude that the differences seen in tables 3.1 and 3.2 should be explained by features other than prior distributions. As we will see in the following subsection, the difference between the Poisson-AR(1) and the Poisson-Multifractal models seems to be in the memory displayed by the Poisson-Multifractal process, which is higher than that of the Poisson-AR(1).

3.2.3 Smoothing and Filtering

Using the efficient importance sampling (EIS) technique, Jung & Liesenfeld (2001) proposed to estimate a function of the sequence of the latent variable $(w_1, ..., w_T)$ numerically for

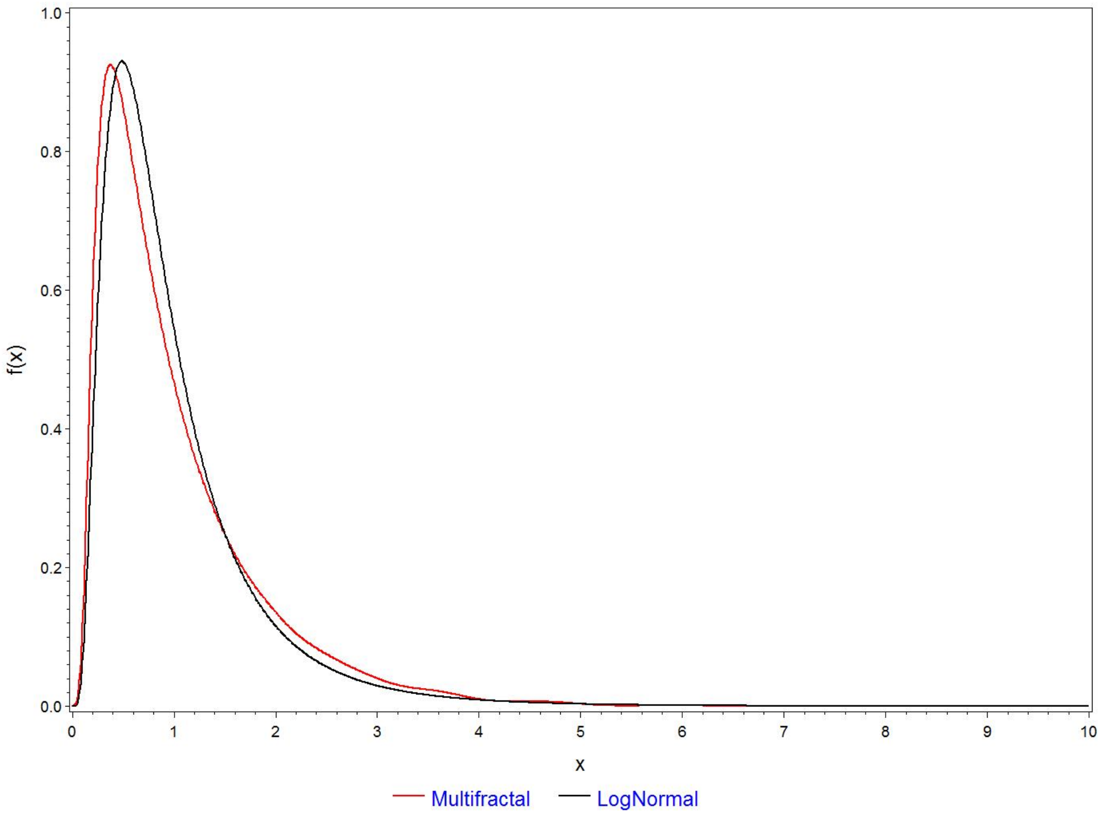

Figure 3.4: Prior distribution of the dynamic random effects for the Poisson model with polio data ($m = 8$ for the multifractal process)

the Poisson-AR(1) model. Using numerical integration techniques and efficient importance sampling, given all T observed counts $\boldsymbol{N_T} = (n_1, ..., n_T)$, they proposed a procedure to assess the expectation of any function $h(.)$ of w_t:

$$\mathbb{E}[h(w_t)|\boldsymbol{N_T}] = \frac{\int h(w) f(\boldsymbol{N_T}, w) dw}{\int f(\boldsymbol{N_T}, w) dw} \tag{3.5}$$

where $f(\mathbf{N_T}, w)$ is the joint density of N_T and w_t. The Poisson-Multifractal model also allows us to analyze the posterior distribution of the fractal components F_t at any time t, but is less time-consuming. In particular, the conditional distribution of F_t, given all T observed counts $\boldsymbol{N_T} = n_1, ..., n_T$, is noted $g(F_t|\boldsymbol{N_T})$ and is provided by the following relation:

$$\begin{aligned} g(F_t|\boldsymbol{N_T}) &= g(F_t|n_1, ..., n_T) \\ &= \frac{g(F_t, n_1, ..., n_T)}{\Pr(\boldsymbol{N_T})} \\ &= \frac{g(F_t, n_{t+1}, ..., n_T|n_1, ..., n_t) \Pr(n_1, ..., n_t)}{\Pr(\boldsymbol{N_T})} \\ &= \frac{\Pr(n_{t+1}, ..., n_T|F_t, n_1, ..., n_t) g(F_t|n_1, ..., n_t) \Pr(n_1, ..., n_t)}{\Pr(\boldsymbol{N_T})} \\ &= \frac{\Pr(n_{t+1}, ..., n_T|F_t) \Pr(\boldsymbol{N_t})}{\Pr(\boldsymbol{N_T})} g(F_t|\boldsymbol{N_t}), \end{aligned} \tag{3.6}$$

where all elements of this ratio that can be retrieved in the estimation procedure. $g(F_t|n_1, ..., n_t)$ is equal to the distribution of probabilities Π_t:

$$g(F_t = s_j \,|\, n_1, ..., n_t) = \Pi_t^{(j)}.$$

The joint density of all counts up to time t and T are noted $\Pr(n_1, ..., n_t)$ and $\Pr(\boldsymbol{N_T})$ respectively. They are approached by likelihood functions as defined in Section 2.1, e.g. :

$$
\begin{aligned}
\Pr(n_1, ..., n_t) &= L(n_1 \ldots n_t \,| \hat{m}_{0,j=1\ldots m}, \hat{\gamma}_{j=1\ldots m}, \hat{\beta}) \\
&= \prod_{u=1}^{t} \langle p(u, n_u, x_u), (\Pi_{u-1} A) \rangle
\end{aligned}
$$

where $\hat{m}_{0,j=1\ldots m}, \hat{\gamma}_{j=1\ldots m}, \hat{\beta}$ are estimators obtained by maximization of the loglikelihood. Similarly, $\Pr(n_{t+1}, ..., n_T | F_t = s_j)$ is the likelihood function given a prior distribution $\Pi_t = e_j$ (instead of the ergodic distribution). The smoothed values of $\mathbb{E}\left(F_t | \boldsymbol{N_T}\right)$ and of $\mathbb{E}(\exp(W_t) | \boldsymbol{N_T})$ are compared in Figure 3.5. Both series are similar for small values of t, but the difference becomes significant for larger values. Compared with the multifractal smoothed values, the smoothed lognormal values exhibit larger variations for $t > 100$, particularly for values of t between 110 and 120.

Figure 3.6 presents the evolution of the smoothed frequencies, $\lambda(t)\mathbb{E}[\exp(W_t)|\boldsymbol{N_T}]$ and $\lambda(t)\mathbb{E}[F_t|\boldsymbol{N_T}]$, for the Poisson-AR(1) and the Poisson-Multifractal models (using, as previously mentioned, estimates in Tables 2.1). Despite differences in $\lambda(t)$ (emphasized by Figure 3.3) and in smoothed values of processes ruling the intensity (Figure 3.5), we see that the values obtained for the *smoothed mean* of each model are almost identical. Thus, each model estimates the data differently: the first model supposes a steeper decreasing time trend with an autoregressive process with large variations for large t, and the second model supposes a weaker decreasing time trend with an underlying mean process that has more time dependence.

To better understand the time dependence of the underlying multifractal process of the Poisson distribution, we will use a specific property of the model. The multifractal component of the Poisson distribution, F_t, can be split into m elements, because F_t is equal to the product of $M_{j,t},\ j = 1..., m$. We can then compute the smoothed value of each $M_{j,t},\ j = 1..., m$, using a similar development as in (3.6):

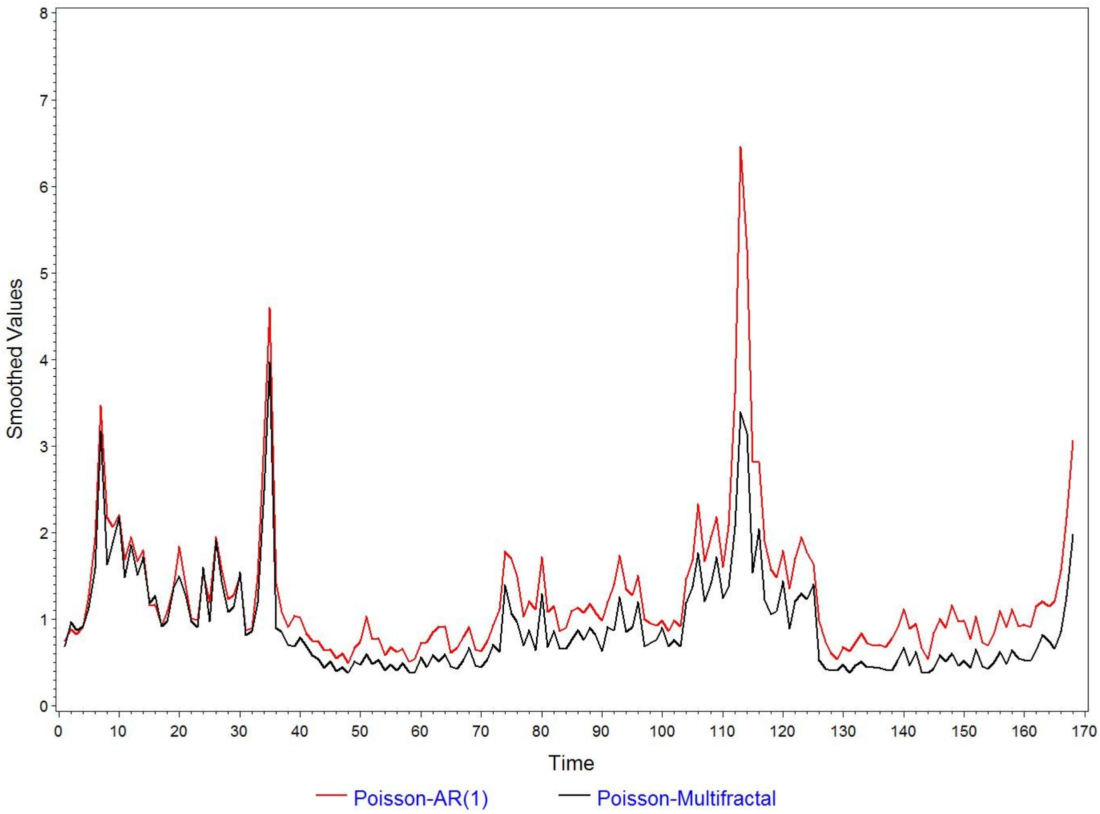

Figure 3.5: Smoothed values of $\exp(W_t)$ (for the Poisson-AR(1) model) and F_t (for the Poisson-Multifractal model)

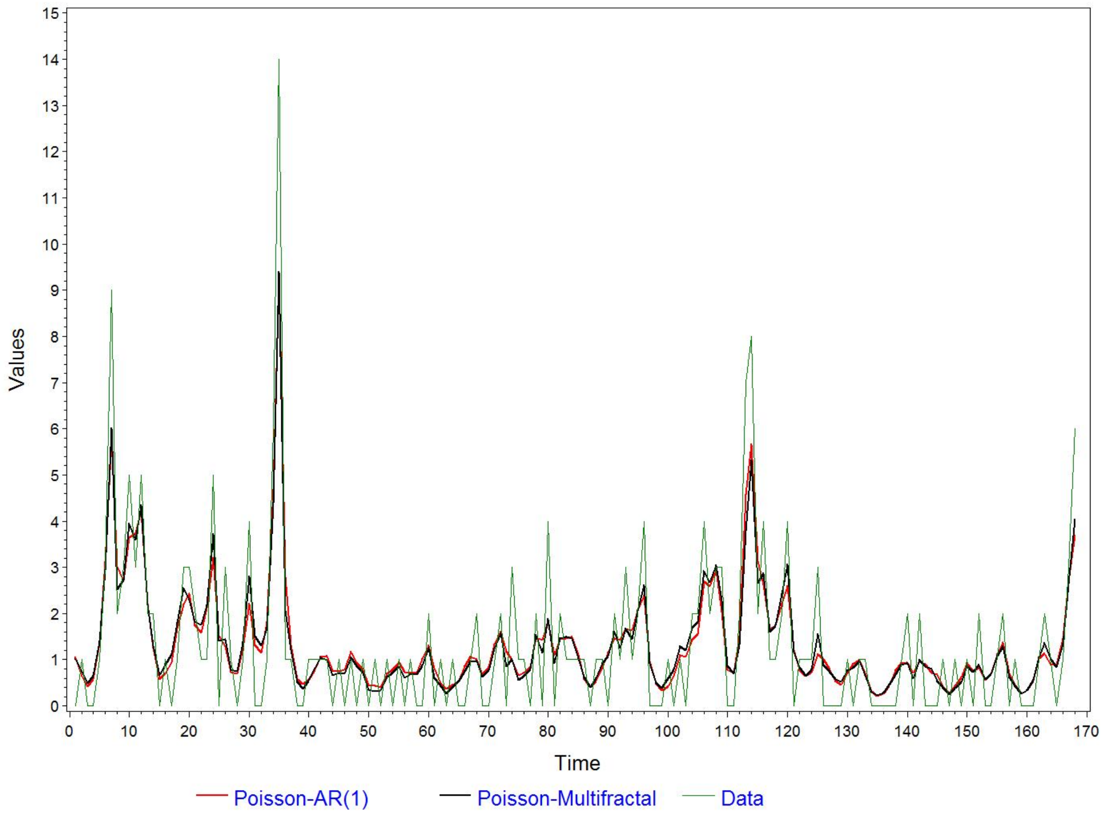

Figure 3.6: Smoothed values $\lambda(t)\mathbb{E}[\exp(W_t)|\boldsymbol{N_T}]$ (for the Poisson-AR(1) model) and $\lambda(t)\mathbb{E}[F_t|\boldsymbol{N_T}]$ (for the Poisson-Multifractal model)

$$g(M_{j,t}|\boldsymbol{N_T}) \quad = \quad \frac{\Pr(n_{t+1},...,n_T|M_{j,t})\Pr(\boldsymbol{N_t})}{\Pr(\boldsymbol{N_T})} g(M_{j,t}|\boldsymbol{N_t}) \quad j=1,...,m, \tag{3.7}$$

or using the decomposition of $g(F_t|\boldsymbol{N_T}) = \prod_{j=1}^{m} g(M_{j,t}|\boldsymbol{N_T})$. The evolution of the smoothed value of $M_{j,t}$, as defined in equation 3.7, and the distribution of $g(M_{j,t}|\boldsymbol{N_T})$ are shown in figures 2.5 and 2.6 for different values of m. For each values of $m = 5,6,7,8$, we observe a strong time dependence for the process $M_{1,t}$. This can be interpreted as an unobserved factor that has a specific random cycle affecting the counts. The cycles observed for $M_{1,t}$ are almost regularly periodic: a high value between $t = 7$ to $t = 33$, a low value for $t = 44$ to $t = 70$, a high value between $(105, 121)$ and a last time period between $(128, 162)$, with low values. Except for the period $(75, 100)$, which should be analyzed in more detail, the cycle of $M_{1,t}$ seems to be regular, which means that we can observe long-term cycles of about 24 to 30 months.

The cycle observed for the first process $M_{1,t}$ explains the differences between time trends, when compared with the Poisson-AR(1) model. Note that the other processes $M_{j,t}$, $j = 2,...,m$ also exhibit time dependence, but it is much lower than that observed for the first process. In fact, other processes can almost be seen as noise added to the count distribution. Indeed, each new process $M_{j,t}$ added to the model with $m = 4$ causes only slight variation in the model.

Consequently, even if the fit of the Poisson-AR(1) and the Poisson-Multifractal models are almost identical, each model proposes different interpretations in the number of poliomyelitis cases from 1970 to 1983.

4 Conclusion

This work introduces the Poisson-Multifractal process to the count distribution literature. It belongs to the family of autoregressive processes and combines the Poisson and fractal

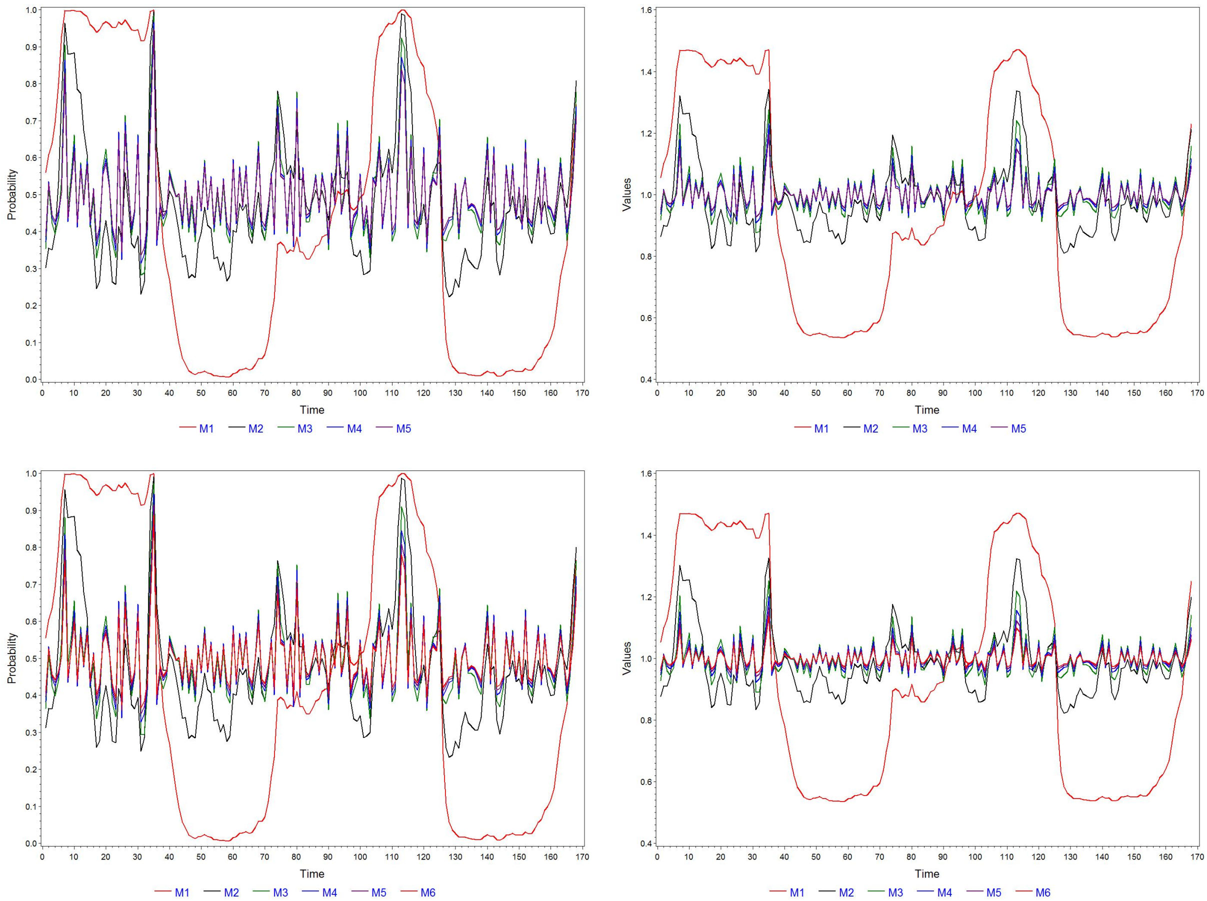

Figure 3.7: Evolution of the probability of $M_{j,t} = m_0(j)$ (left) and $\mathbb{E}[M_{j,t}|\boldsymbol{N_T}]$ (right), $j = 1, ..., m$, for $m = 5, 6$, for the Poisson-Multifractal model

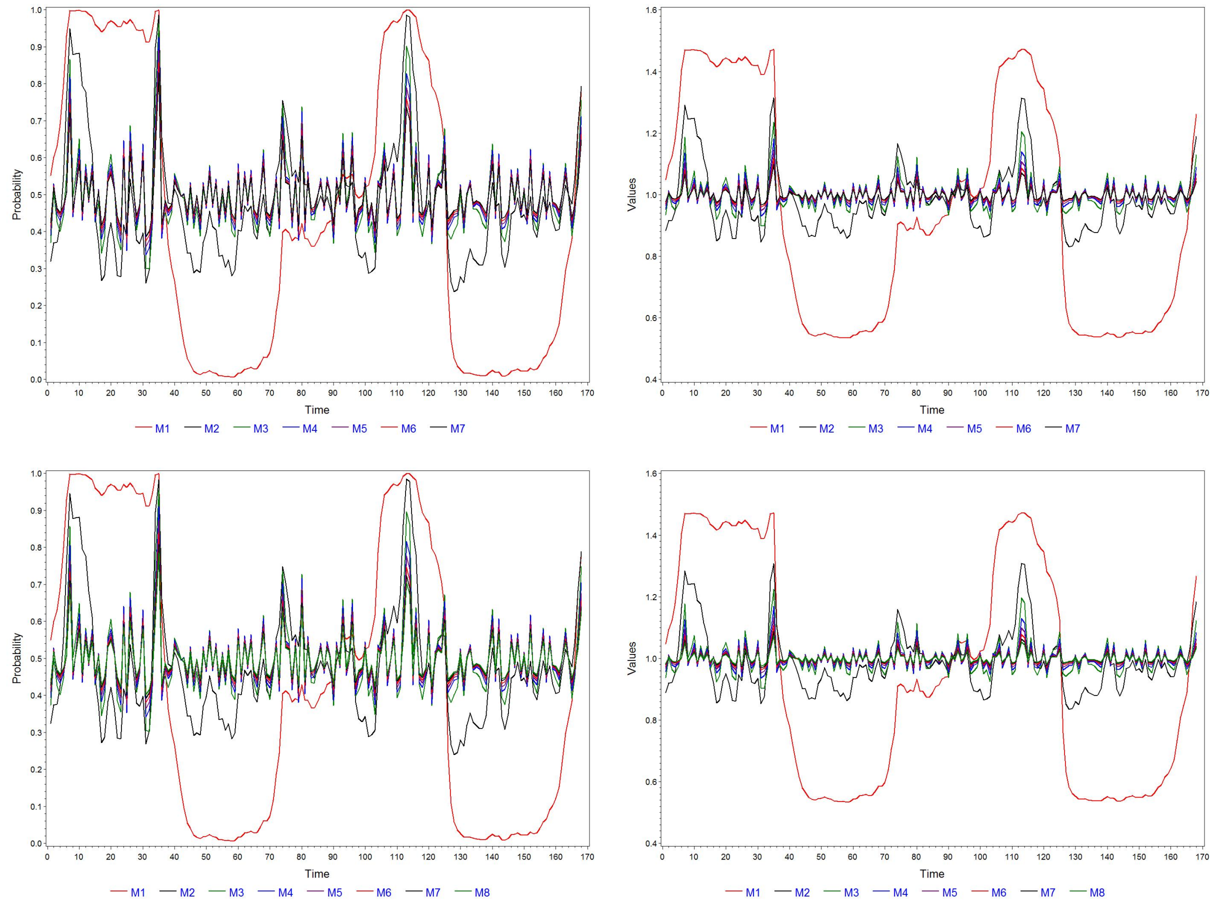

Figure 3.8: Evolution of the probability of $M_{j,t} = m_0(j)$ (left) and $\mathbb{E}[M_{j,t}|\boldsymbol{N_T}]$ (right), $j = 1, ..., m$, for $m = 7, 8$, for the Poisson-Multifractal model

processes. This is a flexible approach to time series of count, particularly when data exhibits overdispersion and slowly declining autocovariograms. Properties of the Poisson-Multifractal model have been shown, as well as is inference technique, which is less time consuming than inference techniques based on Monte Carlo simulations.

Using a reference dataset, we compare the Poisson-Multifractal and the classic Poisson-AR(1) models. Although the two models share similarities, we observe a difference in the time dependence implied by both approaches. In particular, the memory displayed by the Poisson-Multifractal process is more extensive than that of the Poisson-AR(1). Using filtering and smoothing techniques, we analyze each component of the multifractal process and show that they can be interpreted as unobserved climatic, economic, or social factors. For this example, the Poisson-Multifractal process points to an almost regular periodic cycle in Zeger's (1988) polio's data.

References

[1] Al-Osh M.A., Alzaid A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process. J.Time Ser. Anal. 8, 261-275.

[2] Benjamin M.A., Rigby R.A. Stasinopoulos D. M. (2003). Generalized autoregressive Moving Average Models. J. Am. Statist. Assoc. 98 (461), 214-223.

[3] Calvet L.E., Fisher A.J. (2008). Multifractal volatility. Theory, forecasting and pricing. Academic press. Elsevier.

[4] Chan K.S., Ledolter J. (1995). Monte Carlo EM estimation in time series models involving counts. J. Am. Statist. Assoc. 90 (429), 242-252.

[5] Cox D.R. (1981). Statistical Analysis of Time Series: Some Recent Developments, Scandinavian Journal of Statistics, 8, 93-115.

[6] Denuit M., Maréchal X., Pitrebois S., Walhin J.-F. (2007) Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Scales. Wiley, New York.

[7] Durbin J., Koopman S.J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models . Biometrika, 84: 669-684.

[8] Farrell P.J., MacGibbon B., Tomberlin T.J. (2007). A Hierarchical Bayes Approach to Estimation and Prediction for Time Series of Counts, Brazilian Journal of Probability and Statistics, 21, 187-202

[9] Frees E.W., Valdez E.A. (2008). Hierarchical Insurance Claims Modeling. J. Am. Statist. Assoc. 103 (484), 1457-1469.

[10] Frühwirth-Schnatter S. (1994). Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering. Statistics and Computing 4: 259-269.

[11] Hamilton J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357-384.

[12] Jacobs P.A., Lewis P.A.W. (1978a).Discrete Time Series generated by Mixtures I: Correlational and Runs Properties. J.R. Statist. Soc. (B) 40, 94-105.

[13] Jacobs, P.A., Lewis P.A.W. (1978b). Discrete Time Series generated by Mixtures II: Asymptotic Properties. J.R. Statist. Soc. (B) 40, 222-228.

[14] Jung R.C., Liesenfeld R. (2001). Estimating Time Series Models for Count Data Using Efficient Importance Sampling. Allgemeines Statistisches Archiv, 85: 387-407.

[15] Jung R.C., Tremayne A.R. (2006). Coherent forecasting in integer time series models. Internat. J. of Forecast. 22, 223-238.

[16] Jung R.C., Tremayne A.R. (2011), Useful models for time series of counts or simply wrong ones? AStA Advances in Statistical Analysis, 95(1), 59-91

[17] Jung R., Liesenfeld R., Richard J.-F. (2011). “Dynamic Factor Models for Multivariate Count Data: An Application to Stock-market Trading Activity Journal of Business and Economics Statistics, 29, 73-85.

[18] Kalman R.E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82(1), 35-45.

[19] Kuk A.Y.C, Cheng Y. W. (1997). The monte carlo newton-raphson algorithm. Journal of Statistical Computation and Simulation, 59(3), 233-250

[20] MacDonald I.A., Zucchini W. (1997). Hidden Markov and Other Models for Discrete valued Time Series. Chapman & Hall, London.

[21] Mandelbrot B. (1982).The Fractal Geometry of Nature, Freeman.

[22] Mandelbrot B. (1997). Fractals and Scaling in Finance: Discontinuity, Concentration, Risk, Springer Verlag, Berlin.

[23] Mandelbrot B. (2001). Scaling in financial prices: Multifractals and the star equation. Quantitative finance, 1, 124-130

[24] McKenzie E. (2003).Discrete variate time series. In: Shanbhag, D.N., Rao, C.R. (Eds.), Handbook of Statistics, vol. 21. Elsevier, Amsterdam, pp. 573-606.

[25] Oh M.S., Lin Y.B. (2001) Bayesian Analysis of Time Series Poisson Data, Journal of Applied Statistics, 28, 259-271

[26] Raftery A.E. (1985) A new model for discrete-valued time series: autocorrelations and extensions. Rassegna di Metodi Statistici ed Applicazioni, 3-4 149-162.

[27] Weiss C. (2008). Thinning operations for modeling time series of counts-a survey. Adv. Stat. Anal. 92:319- 341

[28] Zeger S.L. (1988). A regression-model for time-series of counts, Biometrika 75, 621-629.

Appendix

Proof of Proposition 2.1. This result can be obtained directly using the facts that the $M_{j,t} = M_{j,t-1}$ with probability $1-\gamma_j$, while $M_{j,t}$ is drawn from $M(j)$ (having a mean of 1) with probability γ_j:

$$\mathbb{E}(M_{j,t+1}|M_{j,t}) = (1-\gamma_j)M_{j,t} + \gamma_j$$

□

Proof of Corollary 2.2. We have

$$\begin{aligned}
\mathbb{E}(M_{j,t+h}|M_{j,t}) &= \mathbb{E}(\mathbb{E}(M_{j,t+h}|M_{j,t+h-1})|M_{j,t}) \\
&= \mathbb{E}((1-\gamma_j)M_{j,t+h-1} + \gamma_j|M_{j,t}) \\
&= (1-\gamma_j)\mathbb{E}(M_{j,t+h-1}|M_{j,t}) + \gamma_j \\
&= (1-\gamma_j)\mathbb{E}(E(M_{j,t+h-1}|M_{j,t+h-2})|M_{j,t}) + \gamma_j \\
&= (1-\gamma_j)\mathbb{E}((1-\gamma_j)M_{j,t+h-2} + \gamma_j|M_{j,t}) + \gamma_j \\
&= (1-\gamma_j)^2\mathbb{E}(M_{j,t+h-2}|M_{j,t}) + (1-\gamma_j)\gamma_j + \gamma_j \\
&= (1-\gamma_j)^2\mathbb{E}(M_{j,t+h-2}|M_{j,t}) + 1 - (1-\gamma_j)^2 \\
&= (1-\gamma_j)^2(\mathbb{E}(M_{j,t+h-2}|M_{j,t}) - 1) + 1
\end{aligned}$$

Continuing this development leads to the result. □

Proof of Proposition 2.3. As in previous propositions, we condition on $M_{j,t}$ to obtain:

$$\begin{aligned}
\mathbb{E}(M_{j,t}M_{j,t+h}) &= \mathbb{E}(\mathbb{E}(M_{j,t}M_{j,t+h}|M_{j,t})) \\
&= \mathbb{E}(M_{j,t}\mathbb{E}(M_{j,t+h}|M_{j,t})) \\
&= \mathbb{E}(M_{j,t}((1-\gamma_j)^h M_{j,t} + 1 - (1-\gamma_j)^h)) \\
&= (1-\gamma_j)^h \mathbb{E}(M_{j,t}^2) + \mathbb{E}(M_{j,t}) - (1-\gamma_j)^h \mathbb{E}(M_{j,t}) \\
&= (1-\gamma_j)^h (\mathbb{E}(M_{j,t}^2) - \mathbb{E}(M_{j,t})) + \mathbb{E}(M_{j,t}) \\
&= (1-\gamma_j)^h (\mathbb{E}(M_{j,t}^2) - 1) + 1,
\end{aligned}$$

where we can use the equation (2.7) to conclude.

□

Proof of Proposition 2.4.

$$\begin{aligned}
\mathbb{E}(N_t) &= \mathbb{E}\left(\mathbb{E}\left(N_t \,|\, F_t\right)\right) \\
&= \lambda(t) \prod_{j=1}^{m} \mathbb{E}\left(M_{j,t}\right) \\
&= \lambda(t)
\end{aligned}$$

The variance is given by the following expression:

$$\begin{aligned}
\mathbb{V}(N_t) &= \mathbb{E}\left(\mathbb{V}\left(N_t \,|\, F_t\right)\right) + \mathbb{V}\left(\mathbb{E}\left(N_t \,|\, F_t\right)\right) \\
&= \lambda(t) + (\lambda(t))^2\, \mathbb{V}\left(F_t\right).
\end{aligned}$$

We have:

$$
\begin{aligned}
\mathbb{V}\left(F_t\right) &= \mathbb{V}\left(F_t\right) \\
&= \mathbb{E}\left(F_t^2\right) - \left(\mathbb{E}\left(F_t\right)\right)^2 \\
&= \mathbb{E}\left(\prod_{j=1}^{m} M_{j,t}^2\right) - \left(\mathbb{E}\left(\prod_{j=1}^{m} M_{j,t}\right)\right)^2 \\
&= \prod_{j=1}^{m} \mathbb{E}\left(M_{j,t}^2\right) - 1,
\end{aligned}
$$

where $\mathbb{E}(M_{j,t}^2)$ is given by equation (2.7). The covariance is, by definition, equal to:

$$
\begin{aligned}
Cov(N_t, N_{t+h}) &= \mathbb{E}\left(\, Cov\left(N_t\, N_{t+h}\,|F_{t+h}\right)\right) + Cov\left(\mathbb{E}\left(N_t\,|F_t\right)\mathbb{E}\left(N_{t+h}\,|F_{t+h}\right)\right) \\
&= 0 + \lambda(t)\lambda(t+h) Cov\left(F_t, F_{t+h}\right)
\end{aligned}
$$

Multipliers are statistically independent, so we can conclude that:

$$
\begin{aligned}
Cov\left(F_t, F_{t+h}\right) &= Cov\left(F_t, F_{t+h}\right) \\
&= Cov\left(\prod_{j=1}^{m} M_{j,t}, \prod_{j=1}^{m} M_{j,t+h}\right) \\
&= \mathbb{E}\left(\prod_{j=1}^{m} M_{j,t} \prod_{j=1}^{m} M_{j,t+h}\right) - \mathbb{E}\left(\prod_{j=1}^{m} M_{j,t}\right)\mathbb{E}\left(\prod_{j=1}^{m} M_{j,t+h}\right) \\
&= \mathbb{E}\left(\prod_{j=1}^{m} M_{j,t} M_{j,t+h}\right) - \left(\prod_{j=1}^{m} \mathbb{E}\left(M_{j,t}\right)\right)\left(\prod_{j=1}^{m} \mathbb{E}\left(M_{j,t+h}\right)\right) \\
&= \prod_{j=1}^{m} \mathbb{E}\left(M_{j,t}, M_{j,t+h}\right) - 1.
\end{aligned}
$$

□