Multivariate integer-valued autoregressive models applied to earthquake counts

Mathieu Boudreault, Arthur Charpentier

Multivariate integer-valued autoregressive models applied to earthquake counts

Mathieu Boudreault*
Département de mathématiques
Université du Québec à Montréal
201, avenue du Président-Kennedy
Montréal (Québec)
Canada H2X 3Y7
email: boudreault.mathieu@uqam.ca

Arthur Charpentier†
Département de mathématiques
Université du Québec à Montréal
201, avenue du Président-Kennedy
Montréal (Québec)
Canada H2X 3Y7
email: charpentier.arthur@uqam.ca

November 30, 2011

*Quantact Research Group, UQAM, Research supported by the NSERC.

†Quantact Research Group, UQAM, Research supported by the Research Chair AXA/FdR (*Large risk in insurance*).

Abstract

In various situations in the insurance industry, in finance, in epidemiology, etc., one needs to represent the joint evolution of the number of occurrences of an event. In this paper, we present a multivariate integer-valued autoregressive (MINAR) model, derive its properties and apply the model to earthquake occurrences across various pairs of tectonic plates. The model is an extension of (Pedeli and Karlis 2011*a*) where cross autocorrelation (spatial contagion in a seismic context) is considered. We fit various bivariate count models and find that for many contiguous tectonic plates, spatial contagion is significant in both directions. Furthermore, ignoring cross autocorrelation can underestimate the potential for high numbers of occurrences over the short-term. Our overall findings seem to further confirm (Parsons and Velasco 2011).

KEYWORDS: autoregressive; Granger causality; counts; earthquakes; INAR; multivariate INAR; Poisson process

Acknowledgment

The authors thank Lionel Truquet for pointing out interesting references on that topic, and François Bergeron for stimulating discussions on matrix based equations.

1. INTRODUCTION AND MOTIVATION

1.1 Motivation

Autoregression in the sense of ARIMA time series models cannot be directly applied to integer values for obvious reasons. Thus, most integer-valued autoregressive (INAR) time series models are based upon thinning operators such as (Steutel and van Harn 1979) (see also the excellent survey of thinning operators by (Weiβ 2008)). Such models have been mainly proposed and investigated by (McKenzie 1985) and (Al-Osh and Alzaid 1987) for first order autocorrelation, and by (Du and Li 1991) for autocorrelation of order p. (Gauthier and Latour 1994), (Dion, Gauthier and Latour 1995) and (Latour 1998) have also investigated a slightly more generalized type of thinning operator than (Steutel and van Harn 1979), in models known as generalized INAR (or GINAR). The statistical and actuarial literature has multiple successful applications of INAR-type of models (see for example (Gourieroux and Jasiak 2004) and (Boucher, Denuit and Guillen 2008) where both papers treat car insurance problems).

In a multivariate setting, the properties of a multivariate INAR (MINAR) model of order 1 (based upon independent binomial thinning operators) have been derived in (Franke and Subba Rao 1993) while the multivariate GINAR of order p is presented in (Latour 1997). However, there are very few attempts in the literature to estimate and use these types of models[1]. One notable exception is (Pedeli and Karlis 2011a) and (Pedeli and Karlis 2011b) who investigated the bivariate INAR model of order 1 with Poisson and negative binomial innovations with an application to the number of daytime and nighttime accidents. In their papers, the autoregression matrix is diagonal, meaning there is no cross-autocorrelation in the counts.

Insurance policies and earthquake catastrophe (cat) derivatives (such as cat-bonds and cat-options) offer protection against earthquake risk in exchange for periodic premiums. Thus, one important component in these contracts is the number of earthquakes at various locations. Earthquake count models are mostly based upon the Poisson process ((Utsu 1969), (Gardner and Knopoff 1974), (Lomnitz 1974), (Kagan and Jackson 1991)), Cox process (self-exciting, cluster or branching pro-

[1](Heinen and Rengifo 2007) use a Vector Autoregression model for the mean of two Poisson-type of random variables. Although the ultimate goal is to represent joint integer-valued random variables, the approach taken is very different from multivariate INAR-types of models.

cesses, stress-release models (see (Rathbun 2004) for a review), or Hidden Markov Models (HMM) (see (Zucchini and MacDonald 2009) and (Orfanogiannaki, Karlis and Papadopoulos 2010)[2]. However, these models are focused toward a single location whereas seismic risk can also be influenced by shocks that occurred at other locations (see e.g. space-time Poisson process in (Ogata 1988), (Zhuang, Y. and Vere-Jones 2002) or (Schoenberg 2003)). Thus, one of the purposes of this paper, is to propose a bivariate INAR model that accounts for cross-autocorrelation in the counts. From a seismological standpoint, that would mean the earthquake count at a given location can be function of the past earthquake counts at that site and at another site. These areas can be tectonic plates, regions, cities or points on a given geological fault.

The main objective of this paper is to investigate the effects of seismic space-time contagion (or clustering) on various risk management applications using seismological data and a specific model. Risk management considerations can be viewed over various time horizons. For example, prices of cat-derivatives will be influenced by short-term earthquake risk dynamics because a lack of an appropriate earthquake count prediction can mean arbitrage profits or losses may occur on the markets. Insurance and reinsurance contracts are managed over a much longer time horizon.

1.2 Outline of the paper

(Parsons and Velasco 2011) have confirmed that major earthquakes might have a significant impact on the number of earthquakes that occur during the hours following the main shock, but only in an area close to the main shock. They do also prove that there is no remote and large earthquakes beyond the main shock region. Figure 1 plots the number of quakes following a big one (magnitude exceeding 6.5), either within or outside a 2,000 km area from the main shock. One of the aims of our paper is to study the dynamics of the number of earthquakes, taking into account spatial contagion over tectonic plates. Using plates instead of distance (as in (Parsons and Velasco 2011)) allows us to work with multivariate counting processes.

[Figure 1 about here.]

The paper is structured as follows. In Section 2, we present the theoretical framework of the multivariate INAR of order 1 and the most important results. Some results have already been

[2]For a brief summary of statistical and stochastic models in seismology, see (Vere-Jones 2010).

derived in (Franke and Subba Rao 1993) but are presented here for the sake of completeness and consistency with our given notation. Additional (theoretical) results in terms of moments, auto-covariance functions and predictions are given in this section. In Section 3, we introduce Granger causality tests, derived in the given context of BINAR(1) processes, including an interpretation of each coefficient in terms of causal effect. Section 4 presents the specific application to the bivariate INAR model with cross autocorrelation and Poisson innovations. A Monte Carlo study will also illustrate how the maximum likelihood estimators behave (theoretical results are given in Section 2, but only in the context of a full cross-correlation matrix). Finally, Section 5 provides various applications of the model with earthquake counts. In subsection 5.3, several BINAR(1) processes are fitted over different tectonic plates and magnitudes. We confirm here the conclusions of (Parsons and Velasco 2011) claiming that the onset of a large earthquake does not cause other large ones at a very long distance. There might be contagion, but it will be between two close areas (e.g. contiguous tectonic plates), and over a short period of time (a few hours, perhaps a few days, but not much longer). In subsection 5.4, we have also observed that major earthquakes will generate several medium-size earthquakes on the same tectonic plate (so called aftershocks). Foreshocks were also observed, meaning that medium-size earthquakes might announce the arrival of more important earthquakes. To conclude, in subsection 5.5, we compare the sum of counts of earthquakes on two plates, assuming that there is - or not - cross correlation between consecutive days.

2. MULTIVARIATE INTEGER-VALUED AUTOREGRESSION OF ORDER 1, MINAR(1)

As mentioned in (Fokianos 2011), a natural way to define a linear model for counts might be to use the Poisson regression to derive an *autoregressive process*. Let (N_t) denote a count time series, and $(\mathcal{F}_t)$ the associated filtration. A GARCH-type model can be considered, as in (Ferland, Latour and Oraichi 2006)

$$N_t|\mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t), \text{ where } \lambda_t = \alpha_0 + \sum_{h=1}^{p} \alpha_h N_{t-h} + \sum_{k=1}^{q} \beta_k \lambda_{t-k}.$$

But one can easily imagine that it could be complicated (and not tractable) to extend such a process in higher dimension. An alternative can be to use a *thinning operator* as in (Al-Osh and Alzaid 1987) or (McKenzie 1985). The idea (introduced in (Steutel and van Harn 1979)) is to define

$\circ$ as

$$p \circ N = Y_1 + \cdots + Y_N \text{ if } N \neq 0, \text{ and 0 otherwise,}$$

where N is a random variable with values in $\mathbb{N}$, $p \in [0,1]$, and $Y_1, Y_2, \cdots$ are i.i.d. Bernoulli variables, independent of N, with $\mathbb{P}(Y_i = 1) = p$. Thus $p \circ N$ is a compound sum of i.i.d. Bernoulli variables. Hence, given N, $p \circ N$ has a binomial distribution with parameters N and p. Based on that thinning operator, the integer autoregressive process of order 1 is defined as

$$N_t = p \circ N_{t-1} + \varepsilon_t = \sum_{i=1}^{N_{t-1}} Y_i + \varepsilon_t,$$

where (ε_t) is a sequence of i.i.d. integer valued random variables. Such process will be called INAR(1). Note that such a process can be related to Galton-Watson process with immigration, and it is a Markov chain with integer states. As mentioned in (Al-Osh and Alzaid 1987), if (ε_t) are Poisson random variables, then (N_t) will also be a sequence of Poisson random variables, and the estimation can be done easily using a method of moments estimators or maximum likelihood techniques, for p and $\lambda = \mathbb{E}(\varepsilon_t)$. One of the main interest of the thinning operator approach is that it can be easily extended in higher dimension, as in (Franke and Subba Rao 1993) (we will also provide new results, as well as new interpretations, e.g. in terms of causality).

2.1 Thinning $\circ$ operator in dimension d

As in the univariate case, before defining a multivariate counting process $\boldsymbol{N_t} := (N_{1,t}, \cdots, N_{d,t})$, we need to define a multivariate thinning operator for a random vector $\boldsymbol{N} := (N_1, \cdots, N_d)$ with values in $\mathbb{N}^d$. Let $\boldsymbol{P} := [p_{i,j}]$ be a $d \times d$ matrix with entries in $[0,1]$. If $\boldsymbol{N} = (N_1, \cdots, N_d)$ is a random vector with values in $\mathbb{N}^d$, then $\boldsymbol{P} \circ \boldsymbol{N}$ is a d-dimensional random vector, with i-th component

$$[\boldsymbol{P} \circ \boldsymbol{N}]_i = \sum_{j=1}^{d} p_{i,j} \circ X_j,$$

for all $i = 1, \cdots, d$, where all counting variates Y in $p_{i,j} \circ X_j$'s are assumed to be independent.

Note that $\boldsymbol{P} \circ (\boldsymbol{Q} \circ \boldsymbol{N}) \stackrel{\mathcal{L}}{=} [\boldsymbol{PQ}] \circ \boldsymbol{N}$. Further, from Lemma 1 in (Franke and Subba Rao 1993), $\mathbb{E}(\boldsymbol{P} \circ \boldsymbol{N}) = \boldsymbol{P}\mathbb{E}(\boldsymbol{N})$, and

$$\mathbb{E}\left((\boldsymbol{P} \circ \boldsymbol{N})(\boldsymbol{P} \circ \boldsymbol{N})'\right) = \boldsymbol{P}\mathbb{E}(\boldsymbol{N}\boldsymbol{N}')\boldsymbol{P}' + \Delta,$$

with $\Delta := \text{diag}(\boldsymbol{V}\mathbb{E}(\boldsymbol{N}))$ where $\boldsymbol{V}$ is the $d \times d$ matrix with entries $p_{i,j}(1 - p_{i,j})$.

Definition 2.1 *A time series* $(\boldsymbol{N}_t)$ *with values in* $\mathbb{N}^d$ *is called a d-variate INAR(1) process if*

$$\boldsymbol{N}_t = \boldsymbol{P} \circ \boldsymbol{N}_{t-1} + \boldsymbol{\varepsilon}_t \tag{1}$$

for all t*, for some* $d \times d$ *matrix* $\boldsymbol{P}$ *with entries in* $[0,1]$*, and some i.i.d. random vectors* $\boldsymbol{\varepsilon}_t$ *with values in* $\mathbb{N}^d$*.*

Remark 2.2 *(Pedeli and Karlis 2011*a*) and (Pedeli and Karlis 2011*b*) defined bivariate INAR(1) processes where matrix* $\boldsymbol{P}$ *is a diagonal matrix.*

Remark 2.3 $(\boldsymbol{N}_t)$ *is a Markov chain with states in* $\mathbb{N}^d$ *with transition probabilities*

$$\pi(\boldsymbol{n}_t, \boldsymbol{n}_{t-1}) = \mathbb{P}(\boldsymbol{N}_t = \boldsymbol{n}_t | \boldsymbol{N}_{t-1} = \boldsymbol{n}_{t-1}) \tag{2}$$

satisfying

$$\pi(\boldsymbol{n}_t, \boldsymbol{n}_{t-1}) = \sum_{\boldsymbol{k}=0}^{\boldsymbol{n}_t} \mathbb{P}(\boldsymbol{P} \circ \boldsymbol{n}_{t-1} = \boldsymbol{n}_t - \boldsymbol{k}) \cdot \mathbb{P}(\boldsymbol{\varepsilon} = \boldsymbol{k}).$$

Remark 2.4 *Since* $\boldsymbol{P}$ *has entries in* $[0,1]$*, using a variant of Perron-Frobenius theorem, there exists an eigenvalue* κ_1 *of* $\boldsymbol{P}$ *such that* $\kappa_1 \geq |\kappa_i|$ *for all other eigenvalues of* $\boldsymbol{P}$*. And the associated eigenvector* $\boldsymbol{v}_1$ *satistfies* $\boldsymbol{v}_1 \geq \boldsymbol{0}$*. Further, if* $\boldsymbol{P}$ *has strictly positive entries,* $p_{i,j} \in (0,1]$*, then* $\kappa_1 > |\kappa_i|$ *and* $\boldsymbol{v}_1 > \boldsymbol{0}$*.*

From Remarks 2.3 and 2.4, we can derive sufficient conditions so that there exists a *stationary* MINAR(1) process (based on Theorem 1 and Lemma 2 in (Franke and Subba Rao 1993)). The proof is based on the fact that under those assumptions, the Markov chain is irreducible and aperiodic. One can prove that $\boldsymbol{0}$ is a positive recurrent state, and from Theorem 1.2.2 in (Rosenblatt 1971), there exists a strictly stationary solution.

Proposition 2.5 *Let* $\boldsymbol{P}$ *with entries in* $(0,1)$*, such that its largest eigenvalue is less than* 1*, and assume that* $\mathbb{P}(\boldsymbol{\varepsilon}_t = \boldsymbol{0}) \in (0,1)$ *with* $\mathbb{E}(\|\boldsymbol{\varepsilon}_t\|_\infty) < \infty$*, then there exists a strictly stationary d-variate INAR(1) process satisfying Equation* (1)*.*

Lemma 2.6 *Let* $(\boldsymbol{N}_t)$ *denote a stationary d-variate INAR(1) process, with autoregressive matrix* $\boldsymbol{P}$ *with entries in* $(0,1)$*, then* $(\boldsymbol{N}_t)$ *admits a d-variate INMA(*∞*) representation*

$$\boldsymbol{N}_t = \sum_{h=0}^{\infty} \boldsymbol{P}^h \circ \boldsymbol{\varepsilon}_{t-h}.$$

2.2 Maximum likelihood estimation in d-variate INAR(1) processes

Consider here a finite time series $\underline{\boldsymbol{N}} = (\boldsymbol{N}_0, \boldsymbol{N}_1, \cdots, \boldsymbol{N}_n)$, observed from time $t = 0$ until time $t = n$. The conditional log-likelihood is

$$\log \mathcal{L}(\underline{\boldsymbol{N}}, \boldsymbol{\theta} | \boldsymbol{N}_0) = \sum_{t=1}^{n} \log \pi(\boldsymbol{N}_{t-1}, \boldsymbol{N}_t) \tag{3}$$

where π is the transition probability of the Markov chain, given by Equation (2). Here parameter $\boldsymbol{\theta}$ is related to the autoregressive matrix $\boldsymbol{P}$ as well as parameters of the joint distribution of the noise process, denoted $\boldsymbol{\lambda}$. For convenience, assume that $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1)$ where $\boldsymbol{\lambda}_0$ are parameters related to the innovation process $(\boldsymbol{\varepsilon}_t)$ margins, and $\boldsymbol{\lambda}_1$ to the dependence among components of the innovation. Hence, here $\boldsymbol{\theta} = (\boldsymbol{P}, \boldsymbol{\lambda})$ on some open sets $(0,1)^{d^2} \times \ell$. From Theorem 2.2 in (Billingsley 1961), since $(\boldsymbol{N}_t)$ is a Markov chain, under standard assumptions, we can obtain asymptotic normality of parameters.

Proposition 2.7 *Let $(\boldsymbol{N}_t)$ be a d-variate INAR(1) process satisfying stationary conditions, as well as technical assumptions (called C1-C6 in (Franke and Subba Rao 1993)), then the conditional maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is asymptotically normal,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(\boldsymbol{0}, \Sigma^{-1}(\boldsymbol{\theta})), \text{ as } n \to \infty.$$

Further,

$$2[\log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{\theta}} | \boldsymbol{N}_0) - \log \mathcal{L}(\underline{\boldsymbol{N}}, \boldsymbol{\theta} | \boldsymbol{N}_0)] \xrightarrow{\mathcal{L}} \chi^2(d^2 + dim(\boldsymbol{\lambda})), \text{ as } n \to \infty.$$

2.3 Autocorrelation matrices for MINAR(1) processes

Based on the properties obtained in (Franke and Subba Rao 1993) it is possible to derive expressions for autocorrelation functions, which is a natural way to describe the dynamics of the process.

Theorem 2.8 *Consider a MINAR(1) process with representation $\boldsymbol{N}_t = \boldsymbol{P} \circ \boldsymbol{N}_{t-1} + \boldsymbol{\varepsilon}_t$, where $(\boldsymbol{\varepsilon}_t)$ is the innovation process, with $\boldsymbol{\lambda} := \mathbb{E}(\boldsymbol{\varepsilon}_t)$ and $\boldsymbol{\Lambda} := var(\boldsymbol{\varepsilon}_t)$. Let $\boldsymbol{\mu} := \mathbb{E}(\boldsymbol{N}_t)$ and $\boldsymbol{\gamma}(h) := cov(\boldsymbol{N}_t, \boldsymbol{N}_{t-h})$. Then $\boldsymbol{\mu} = [\mathbb{I} - \boldsymbol{P}]^{-1}\boldsymbol{\lambda}$ and for all $h \in \mathbb{Z}$, $\boldsymbol{\gamma}(h) = \boldsymbol{P}^h \boldsymbol{\gamma}(0)$ with $\boldsymbol{\gamma}(0)$ solution of $\boldsymbol{\gamma}(0) = \boldsymbol{P}\boldsymbol{\gamma}(0)\boldsymbol{P}' + (\boldsymbol{\Delta} + \boldsymbol{\Lambda})$, and where $\mathbb{I}$ is the $d \times d$ identity matrix.*

Proof 2.9 *Since $\mathbb{E}(\boldsymbol{P} \circ \boldsymbol{N}) = \boldsymbol{P}\mathbb{E}(\boldsymbol{N})$, then $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{N}_t)$ has to satisfy*

$$\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{N}_t) = \mathbb{E}(\boldsymbol{P} \circ \boldsymbol{N}_{t-1} + \boldsymbol{\varepsilon}_t) = \boldsymbol{P}\boldsymbol{\mu} + \boldsymbol{\lambda}, \text{ i.e. } [\mathbb{I} - \boldsymbol{P}]^{-1}\boldsymbol{\lambda}.$$

Further

$$\boldsymbol{\gamma}(0) = var(\boldsymbol{N}_t) = \mathbb{E}\left((\boldsymbol{P} \circ \boldsymbol{N}_{t-1} + \boldsymbol{\varepsilon}_t - \boldsymbol{\mu})(\boldsymbol{P} \circ \boldsymbol{N}_{t-1} + \boldsymbol{\varepsilon}_t - \boldsymbol{\mu})'\right)$$

Since $(\boldsymbol{\varepsilon}_t)$ *is the innovation process, and from the expression mentioned above (from (Franke and Subba Rao 1993))*

$$var(\boldsymbol{N}_t) = \boldsymbol{P} var(\boldsymbol{N}_{t-1})\boldsymbol{P}' + \boldsymbol{\Delta} + \boldsymbol{\Lambda}$$

thus $\boldsymbol{\gamma}(0)$ *satisfies*

$$\boldsymbol{\gamma}(0) = \boldsymbol{P}\boldsymbol{\gamma}(0)\boldsymbol{P}' + (\boldsymbol{\Delta} + \boldsymbol{\Lambda}).$$

Finally,

$$\boldsymbol{\gamma}(h) = cov(\boldsymbol{N}_t, \boldsymbol{N}_{t-h}) = cov\left((\boldsymbol{P} \circ \boldsymbol{N}_{t-1} + \boldsymbol{\varepsilon}_t), \boldsymbol{N}_{t-h}\right) = cov\left((\boldsymbol{P} \circ (\boldsymbol{P} \circ \boldsymbol{N}_{t-2} + \boldsymbol{\varepsilon}_{t-1}) + \boldsymbol{\varepsilon}_t), \boldsymbol{N}_{t-h}\right) \cdots$$

etc, so that

$$\boldsymbol{\gamma}(h) = cov\left(\boldsymbol{P}^h \circ \boldsymbol{N}_{t-h} + \sum_{i=0}^{h-1} \boldsymbol{P}^i \circ \boldsymbol{\varepsilon}_{t-i}, \boldsymbol{N}_{t-h}\right) = \boldsymbol{P}^h var(\boldsymbol{N}_{t-h}) = \boldsymbol{P}^h \boldsymbol{\gamma}(0),$$

since $(\boldsymbol{\varepsilon}_t)$ *is an innovation process.* Q.E.D.

Remark 2.10 $\boldsymbol{\gamma}(0)$ *is a covariance matrix (symmetric) solution of matrix expression*

$$\boldsymbol{Z} - \boldsymbol{P}\boldsymbol{Z}\boldsymbol{P}' = \boldsymbol{A}\ (= \boldsymbol{\Delta} + \boldsymbol{\Lambda}).$$

If $\boldsymbol{P}$ *was an* orthogonal *matrix, the term on the left could be related to Lie bracket* $[\boldsymbol{Z}, \boldsymbol{P}] = \boldsymbol{Z}\boldsymbol{P} - \boldsymbol{P}\boldsymbol{Z}$*. If* $\boldsymbol{P}$ *was diagonal, we would have obtained expression of (Pedeli and Karlis 2011*a*). Thus, assuming that* $\boldsymbol{P}$ *can either be orthogonalized or diagonalized will lead to tractable numerical algorithm. Another numerical strategy is to seek for a fixed point in equation* $\boldsymbol{Z}_n = \boldsymbol{P}\boldsymbol{Z}_{n-1}\boldsymbol{P}' + \boldsymbol{A}$ *with some starting value* $\boldsymbol{Z}_0$ *(e.g.* $\mathbb{I}$*). This numerical technique will be used in the applications (see Section* 5*).*

2.4 Forecasting with MINAR(1) processes

In order to derive the distribution (or moments) of $\boldsymbol{N}_{t+h}$ given $\boldsymbol{N}_t$ recall that

$$(\boldsymbol{N}_{t+h}, \boldsymbol{N}_t) = \left(\boldsymbol{P}^h \circ \boldsymbol{N}_t + \sum_{i=0}^{h-1} \boldsymbol{P}^i \circ \boldsymbol{\varepsilon}_{t+h-i}, \boldsymbol{N}_{t-h}\right)$$

Proposition 2.11 *Let* $\boldsymbol{\lambda} := \mathbb{E}(\boldsymbol{\varepsilon}_t)$ *and* $\boldsymbol{\Lambda} := var(\boldsymbol{\varepsilon}_t)$*, then*

$$\mathbb{E}(\boldsymbol{N}_{t+h}|\boldsymbol{N}_t) = \boldsymbol{P}^h \boldsymbol{N}_t + \left(\mathbb{I} + \boldsymbol{P} + \cdots + \boldsymbol{P}^{h-1}\right)\boldsymbol{\lambda}$$

Proof 2.12 *The conditional expectation is obtained by recurrence, since the one step ahead conditional expectation is*

$$\mathbb{E}(\boldsymbol{N}_{t+1}|\boldsymbol{N}_t) = \mathbb{E}(\boldsymbol{P} \circ \boldsymbol{N}_t + \boldsymbol{\varepsilon}_t|\boldsymbol{N}_t) = \boldsymbol{P}\boldsymbol{N}_t + \boldsymbol{\lambda}.$$

Q.E.D.

For the conditional variance, it is possible to derive iterative formulas, using recursions.

Proposition 2.13 *Let* $\boldsymbol{\lambda} := \mathbb{E}(\boldsymbol{\varepsilon}_t)$ *and* $\boldsymbol{\Lambda} := var(\boldsymbol{\varepsilon}_t)$*, then* $var(\boldsymbol{N}_{t+h}|\boldsymbol{N}_t) = V_h(\boldsymbol{N}_t)$ *where* $V_h(\boldsymbol{N})$ *is defined recursively by* $V_1(\boldsymbol{N}) = diag(\boldsymbol{V}\boldsymbol{N}) + \boldsymbol{\Lambda}$ *and*

$$V_h(\boldsymbol{N}) = \mathbb{E}[V_{h-1}(\boldsymbol{P} \circ \boldsymbol{N} + \boldsymbol{\varepsilon})|\boldsymbol{N}] + \boldsymbol{P}^{h-1}[diag(\boldsymbol{V}\boldsymbol{N}) + \boldsymbol{\Lambda}](\boldsymbol{P}^{h-1})'.$$

Proof 2.14 *The one step ahead conditional variance is*

$$var(\boldsymbol{N}_{t+1}|\boldsymbol{N}_t) = var(\boldsymbol{P} \circ \boldsymbol{N}_t + \boldsymbol{\varepsilon}_t|\boldsymbol{N}_t) = diag(\boldsymbol{V}\boldsymbol{N}_t) + \boldsymbol{\Lambda},$$

where $\boldsymbol{V}$ *is the* $d \times d$ *matrix with entries* $p_{i,j}(1 - p_{i,j})$*. Then, in order to use a recursive argument, we simply have to use the variance decomposition formula, and move one additional step ahead. At time* $t + h$*, we can write*

$$var(\boldsymbol{N}_{t+h}|\boldsymbol{N}_t) = \mathbb{E}[var(\boldsymbol{N}_{t+h}|\boldsymbol{N}_{t+1})|\boldsymbol{N}_t] + var[\mathbb{E}(\boldsymbol{N}_{t+h}|\boldsymbol{N}_{t+1})|\boldsymbol{N}_t],$$

i.e.

$$var(\boldsymbol{N}_{t+h}|\boldsymbol{N}_t) = \mathbb{E}[var(\boldsymbol{N}_{t+h}|\boldsymbol{N}_{t+1})|\boldsymbol{N}_t] + var[\boldsymbol{P}^{h-1}\boldsymbol{N}_{t+1} + \left(\mathbb{I} + \boldsymbol{P} + \cdots + \boldsymbol{P}^{h-2}\right)\boldsymbol{\lambda}|\boldsymbol{N}_t],$$

$$var(\boldsymbol{N}_{t+h}|\boldsymbol{N}_t) = \mathbb{E}[var(\boldsymbol{N}_{t+h}|\boldsymbol{N}_{t+1})|\boldsymbol{N}_t] + \boldsymbol{P}^{h-1}var[\boldsymbol{N}_{t+1}|\boldsymbol{N}_t](\boldsymbol{P}^{h-1})'.$$

Q.E.D.

In the case where $\boldsymbol{P}$ is diagonal, we obtain as particular case the expressions of (Pedeli and Karlis 2011*b*).

Corollary 2.15 *If $\boldsymbol{P}$ is a diagonal matrix, then*

$$\mathbb{E}(N_{i,t+h}|\boldsymbol{N}_t) = p_{i,i}^h N_{i,t} + [1 + p_{i,i} + \cdots + p_{i,i}^{h-1}]\lambda_i = p_{i,i}^h N_{i,t} + \left(\frac{1-p_{i,i}^h}{1-p_{i,i}}\right)\lambda_i$$

while

$$var(N_{i,t+h}|\boldsymbol{N}_t) = p_{i,i}^h[1-p_{i,i}^h]N_{i,t} + \left(\frac{1-p_{i,i}^{2h}}{1-p_{i,i}^2}\right)\Lambda_{i,i} + \left(\frac{1-p_{i,i}^h}{1-p_{i,i}} - \frac{1-p_{i,i}^{2h}}{1-p_{i,i}^2}\right)\lambda_i$$

Remark 2.16 *If $(\boldsymbol{N}_t)$ is a stationary MINAR(1) process, then, as $h \to \infty$, $\mathbb{E}(\boldsymbol{N}_{t+h}|\boldsymbol{N}_t)$ converges to $\sum_{k\geq 0} \boldsymbol{P}^k\boldsymbol{\mu} = [\mathbb{I} - \boldsymbol{P}]^{-1}\boldsymbol{\mu}$.*

3. NONDIAGONAL THINNING MATRICES AND GRANGER CAUSALITY

Based on the concepts introduced in (Franke and Subba Rao 1993), it is possible to get interpretations of parameters. Based on Granger terminology, N_2 causes N_1 at time t if and only if

$$\mathbb{E}\left(N_{1,t}|\underline{N}_{1,t-1}, \underline{N}_{2,t-1}\right) \neq \mathbb{E}\left(N_{1,t}|\underline{N}_{1,t-1}\right),$$

where $\underline{N}_{1,t-1} = (N_{1,0}, \cdots, N_{1,t-1})$ and $\underline{N}_{2,t-1} = (N_{2,0}, \cdots, N_{2,t-1})$.

Further, N_2 causes *instantaneously* N_1 at time t if

$$\mathbb{E}\left(N_{1,t}|\underline{N}_{1,t-1}, \underline{N}_{2,t-1}, N_{2,t}\right) \neq \mathbb{E}\left(N_{1,t}|\underline{N}_{1,t-1}, \underline{N}_{2,t-1}\right).$$

Thus, as for Gaussian VAR processes, the following interpretation holds (see Section 3.6. in (Lutkepohl 2005))

Lemma 3.1 *Let $\boldsymbol{N}_t = (N_{1,t}, N_{2,t})$ be a bivariate INAR(1) with representation* (1)

1. *$(N_{1,t})$ and $(N_{2,t})$ are instantaneously related if $\boldsymbol{\varepsilon}$ is a noncorrelated noise,*
2. *$(N_{1,t})$ and $(N_{2,t})$ are independent, which we denote $(N_{1,t}) \perp (N_{2,t})$, if $\boldsymbol{P}$ is diagonal, i.e. $p_{1,2} = p_{2,1} = 0$, and $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are independent,*
3. *$(N_{1,t})$ causes $(N_{2,t})$ but $(N_{2,t})$ does not cause $(N_{1,t})$, which we denote $(N_{1,t}) \to (N_{2,t})$, if $\boldsymbol{P}$ is a lower triangle matrix, i.e. $p_{2,1} = 0$ while $p_{1,2} \neq 0$,*
4. *$(N_{2,t})$ causes $(N_{1,t})$ but $(N_{1,t})$ does not cause $(N_{2,t})$, which we denote $(N_{1,t}) \leftarrow (N_{2,t})$, if $\boldsymbol{P}$ is a lower triangle matrix, i.e. $p_{1,2} = 0$ while $p_{2,1} \neq 0$,*

5. $(N_{1,t})$ *causes* $(N_{2,t})$ *and conversely, i.e. a feedback effect, which we denote* $(N_{1,t}) \leftrightarrow (N_{2,t})$*, if* $\boldsymbol{P}$ *is a full matrix, i.e.* $p_{1,2}, p_{2,1} \neq 0$

From those characterizations, and Proposition 2.7, it is possible to derive a simple testing procedure, based on a likelihood ratio test. For instantaneous causality, we test

$$H_0 : \boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_1^{\perp} \text{ against } H_1 : \boldsymbol{\lambda}_1 \neq \boldsymbol{\lambda}_1^{\perp},$$

where $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_1^{\perp}$ if and only if margins of the innovation process are independent.

Corollary 3.2 *Let* $\widehat{\boldsymbol{\lambda}}$ *denote the conditional maximum likelihood estimate of* $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1)$ *in the non-constrained MINAR(1) model, and* $\widehat{\boldsymbol{\lambda}}^{\perp}$ *denote the conditional maximum likelihood estimate of* $\boldsymbol{\lambda}^{\perp} = (\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1^{\perp})$ *in the constrained model (when innovation has independent margins), then under suitable conditions,*

$$2[\log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{\lambda}} | \boldsymbol{N}_0) - \log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{\lambda}}^{\perp} | \boldsymbol{N}_0)] \xrightarrow{\mathcal{L}} \chi^2(dim(\boldsymbol{\lambda}) - dim(\boldsymbol{\lambda}^{\perp})), \text{ as } n \to \infty, \text{ under } H_0.$$

For lagged causality, we test

$$H_0 : \boldsymbol{P} \in \mathcal{P} \text{ against } H_1 : \boldsymbol{P} \notin \mathcal{P},$$

where $\mathcal{P}$ is a set of constrained shaped matrix, e.g. $\mathcal{P}$ is the set of $d \times d$ diagonal matrices for lagged independence, or a set of block triangular matrices for lagged causality.

Corollary 3.3 *Let* $\widehat{\boldsymbol{P}}$ *denote the conditional maximum likelihood estimate of* $\boldsymbol{P}$ *in the non-constrained MINAR(1) model, and* $\widehat{\boldsymbol{P}}^c$ *denote the conditional maximum likelihood estimate of* $\boldsymbol{P}$ *in the constrained model, then under suitable conditions,*

$$2[\log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{P}} | \boldsymbol{N}_0) - \log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{P}}^c | \boldsymbol{N}_0)] \xrightarrow{\mathcal{L}} \chi^2(d^2 - dim(\mathcal{P})), \text{ as } n \to \infty, \text{ under } H_0.$$

4. BIVARIATE INAR(1) PROCESS WITH POISSON INNOVATION

MINAR(d) might appear as tractable models, but the number of parameters can be extremely large. In dimension d, the dynamics is characterized by $d^2 + \dim(\boldsymbol{\Lambda})$ parameters. The standard model for the innovation process would be the multivariate common shock Poisson random vector

(see (Mahamunulu 1967) or Section 37.2 in (Johnson, Kotz and Balakrishnan 1997)). Let $(u_I, I \subset \{1, \cdots, d\})$ be a collection of independent Poisson random variables, and define

$$\varepsilon_i := \sum_{I \subset \{1,\cdots,d\}, i \in I} u_I$$

then $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_d)$ has a multivariate Poisson distribution. In that case, $\dim(\boldsymbol{\Lambda}) = 2^d$. In moderate dimension (e.g. $d = 10$) $\dim(\boldsymbol{\Lambda})$ is larger than one thousand, which will not be tractable. Thus, for convenience, let us focus on the bivariate INAR(1) process.

4.1 The bivariate Poisson innovation process

A classical distribution for $\boldsymbol{\varepsilon}_t$ is the bivariate Poisson distribution, with one common shock, i.e.

$$\begin{cases} \varepsilon_{1,t} = M_{1,t} + M_{0,t} \\ \varepsilon_{2,t} = M_{2,t} + M_{0,t} \end{cases}$$

where $M_{1,t}$, $M_{2,t}$ and $M_{0,t}$ are independent Poisson variates, with parameters $\lambda_1 - \varphi$, $\lambda_2 - \varphi$ and φ, respectively. In that case, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t})$ has joint probability function

$$\mathbb{P}[(\varepsilon_{1,t}, \varepsilon_{2,t}) = (k_1, k_2)] = e^{-[\lambda_1+\lambda_2-\varphi]} \frac{(\lambda_1-\varphi)^{k_1}}{k_1!} \frac{(\lambda_2-\varphi)^{k_2}}{k_2!} \sum_{i=0}^{\min\{k_1,k_2\}} \binom{k_1}{i} \binom{k_2}{i} i! \left(\frac{\varphi}{[\lambda_1-\varphi][\lambda_2-\varphi]} \right)$$

with $\lambda_1, \lambda_2 > 0$, $\varphi \in [0, \min\{\lambda_1, \lambda_2\}]$. See e.g. (Kocherlakota and Kocherlakota 1992) for a comprehensive description of that joint distribution. Note that $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are both Poisson distributed, with parameter λ_1 and λ_2 respectively, and here $\text{cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) = \varphi$. Hence, parameter φ characterizes independence (or non-independence) of the innovation process. Hence

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \text{ and } \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

and most of the previous expressions can be derived explicitly.

4.2 BINAR(1) process with Poisson innovation

For univariate INAR(1) processes, if N_0 is assumed to have a Poisson distribution with mean $\lambda/(1-p)$, then N_t is also Poisson distributed, for all $t \geq 0$. But this result does not hold in higher dimensions ((Pedeli and Karlis 2011*a*) noticed that result with diagonal $\boldsymbol{P}$ matrices, and it is still true). Nevertheless, it is still possible to derive joint moments of the joint distributions ($\boldsymbol{\mu}$ and $\boldsymbol{\gamma}(0)$) as well as autocorrelations.

Example 4.1 $\boldsymbol{\mu} := \mathbb{E}(\boldsymbol{N}_t)$ *is given by*

$$\begin{cases} \mathbb{E}(N_{1,t}) = \mu_1 = \dfrac{(1-p_{2,2})\lambda_1 + p_{1,2}\lambda_2}{(1-p_{1,1})(1-p_{2,2}) - p_{2,1}p_{1,2}} \\ \mathbb{E}(N_{2,t}) = \mu_2 = \dfrac{(1-p_{1,1})\lambda_2 + p_{2,1}\lambda_1}{(1-p_{1,1})(1-p_{2,2}) - p_{2,1}p_{1,2}} \end{cases}$$

Expressions for $\boldsymbol{\gamma}(0)$ and $\boldsymbol{\gamma}(1)$ can be explicitly derived, but from Theorem 2.8 we do have matrices based expression that can be used numerically.

Example 4.2 *Auto and cross autocorrelations are given by*

$$corr(N_{1,t}, N_{2,t}) = \frac{\gamma_{1,2}(0)}{\sqrt{\gamma_{1,1}(0)\gamma_{2,2}(0)}},$$

$$corr(N_{1,t}, N_{1,t-1}) = \frac{\gamma_{1,1}(1)}{\gamma_{1,1}(0)} \text{ and } corr(N_{2,t}, N_{2,t-1}) = \frac{\gamma_{2,2}(1)}{\gamma_{2,2}(0)},$$

$$corr(N_{1,t}, N_{2,t-1}) = \frac{\gamma_{1,2}(1)}{\sqrt{\gamma_{1,1}(0)\gamma_{2,2}(0)}}.$$

Note that Poisson innovation satisfy technical assumption needed in Proposition 2.7 to insure that conditional maximum likelihood estimates converge to a normal distribution as n goes to infinity.

4.3 Maximum likelihood estimation for BINAR(1) with Poisson innovation

From Equation 3 the conditional likelihood of $(\boldsymbol{P}, \boldsymbol{\lambda}, \varphi)$ given a sample $\underline{N} = ((N_{1,t}, N_{2,t}), t = 1, 2, \cdots, n)$ is

$$\mathcal{L}((\boldsymbol{P}, \boldsymbol{\lambda}, \varphi); \underline{N}) = \prod_{t=1}^{n} \sum_{k_1=\underline{k}_1}^{N_{1,t}} \sum_{k_2=\underline{k}_2}^{N_{2,t}} \pi((N_{1,t}-k_1, N_{2,t}-k_2), \boldsymbol{N}_{t-1}) \cdot \underbrace{\mathbb{P}[(\varepsilon_{1,t}, \varepsilon_{2,t}) = (k_1, k_2)]}_{\text{bivariate Poisson}}$$

with $\underline{k}_1 = \max\{N_{1,t} - N_{1,t-1} - N_{2,t-1}, 0\}$ and $\underline{k}_2 = \max\{N_{2,t} - N_{1,t-1} - N_{2,t-1}, 0\}$. Here

$$\pi((n_1, n_2), \boldsymbol{N}_{t-1}) = \pi_1(n_1, \boldsymbol{N}_{t-1}) \cdot \pi_2(n_2, \boldsymbol{N}_{t-1})$$

since given with $(\varepsilon_t, \boldsymbol{N}_{t-1})$, components of $\boldsymbol{N}_t$ are assumed to be independent (from the definition of the multivariate thinning operator $\circ$), where $\pi_1(\cdot, \boldsymbol{N}_{t-1})$ and $\pi_2(\cdot, \boldsymbol{N}_{t-1})$ are convolutions of binomial distributions, i.e. for $n_1, n_2 = 0, 1, \cdots, N_{1,t-1} + N_{2,t-1}$

$$\pi_1(n_1, \boldsymbol{N}_{t-1}) = \sum_{m=0}^{N_{1,t-1}} \binom{N_{1,t-1}}{m} p_{1,1}^{m} (1-p_{1,1})^{N_{1,t-1}-m} \binom{N_{2,t-1}}{n_1-m} p_{1,2}^{n_1-m} (1-p_{1,2})^{N_{2,t-1}-(n_1-m)}$$

$$\pi_2(n_2, \boldsymbol{N}_{t-1}) = \sum_{m=0}^{N_{1,t-1}} \binom{N_{1,t-1}}{m} p_{2,1}^m (1-p_{2,1})^{N_{1,t-1}-m} \binom{N_{2,t-1}}{n_2-m} p_{2,2}^{n_2-m} (1-p_{2,2})^{N_{2,t-1}-(n_2-m)}$$

Using numerical optimization routines, it is possible to compute $(\widehat{\boldsymbol{P}}, \widehat{\boldsymbol{\lambda}}, \widehat{\varphi}) = \text{argmax}\{\mathcal{L}((\boldsymbol{P}, \boldsymbol{\lambda}, \varphi); \underline{N})\}$, and Proposition 2.7 insures convergence of that estimator: the conditional maximum likelihood estimates (CMLE) are asymptotically normal and unbiased.

4.4 Monte Carlo study

Based on the previous expression of the likelihood, it is possible to run Monte Carlo simulations to study the behavior of the estimators on simulated series. This numerical example illustrates with two sets of hypothetical parameters how fast is convergence. The two sets of parameters are:

- $p_{1,1} = 0.25$, $p_{1,2} = 0.05$, $p_{2,1} = 0.1$, $p_{2,2} = 0.4$ with $\lambda_1 = 5$, $\lambda_2 = 3$ and $\varphi = 1$;
- $p_{1,1} = 0.25$, $p_{1,2} = p_{2,1} = 0$, $p_{2,2} = 0.4$ with $\lambda_1 = 5$, $\lambda_2 = 3$ and $\varphi = 1$.

The second set of parameters is a special case of the proposed BINAR, which is the diagonal BINAR model of (Pedeli and Karlis 2011*b*) and (Pedeli and Karlis 2011*a*). This will illustrate that in some instances such as $p_{1,2} = p_{2,1} = 0$, the CMLE of the multivariate INAR still converges to true values (even if Proposition 2.7 insured only convergence in the interior of the support of parameters, i.e. $(0,1)^4 \times (0,\infty)^3$, not on borders).

To perform this experiment, 250 samples of different sizes have been generated. Sample sizes of 25, 50, 100, 250, 500, 1000, 5000 and 10000 observations are considered.

Figure 2 shows the kernel-smoothed[3] density function of the distribution of each parameter in the first set, over the various sample sizes. Tables 1 and 2 show the mean and standard deviation of the parameter values over different sample sizes. One can see that in the first set of parameters, the estimates converge quickly to a normal distribution and the bias goes steadily to 0. In the second set of parameters, the results are shown in Figure 3 and Tables 3 and 4 . One sees that even though $p_{1,2} = p_{2,1} = 0$, the distribution rapidly concentrates at 0. This indicates that the approach is valid.

[3]Although, kernel smoothed densities show some curves in a negative domain, none of estimated parameters was negative in the samples. Thus, the negative domain is only due to smoothing.

[Figure 2 about here.]

[Figure 3 about here.]

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

5. EMPIRICAL APPLICATION TO EARTHQUAKES

5.1 Data

To illustrate the potential of the model for various uses, we apply the proposed BINAR approach on earthquake counts of the Earth's tectonic plates. Since the proposed model accounts for serial correlation and cross-autocorrelation, earthquake counts is an interesting application for the following reasons. First, when a mainshock occurs, it provokes many aftershocks, thus creating serial correlation. Moreover, the seismic waves travel over a large distance, and may cross different tectonic plates, provoking other earthquakes on these other plates (within some time range). This is why earthquake counts on contiguous plates should show statistical dependence and cross autocorrelation. Given the purpose of the paper, the empirical application is by no means an exhaustive seismological analysis of earthquake risk across the planet. From a *seismological standpoint*, some of the results are indicative and further investigation would be required in some aspects of the application.

The data used in the example comes from two sources. First, the limits of each tectonic plate come from the Department of Geography of the University of Colorado at Boulder, who provide on their website, various shapefiles for use with ArcGIS[4]. Figure 4 shows the mapping of the tectonic plates in the latter reference. The tectonic plates are: North American, Eurasian, Okhotsk, Pacific (split in two, East and West), Amur, Indo-Australian, African, Indo-Chinese, Arabian, Philippine, Coca, Caribbean, Somali, South American, Nasca and Antarctic. We have

[4] `http://www.colorado.edu/geography/foote/maps/assign/hotspots/hotspots.html`

decided not to group together the West and East Pacific plates to keep the integrity of the input. Secondly, the listings of past earthquakes come from the Advanced National Seismic System (ANSS) Composite Earthquake Catalog. Each entry provides the date, time, longitude and latitude, depth and magnitude (and its type) of each earthquake. The database spans the time period from 1898 to 2011, but as mentioned on the ANSS website, many databases have been added between 1898 to the mid-1960s. Other factors may have affected the data as well. The addition of seismic stations and the technological improvement of seismological instruments may inflate the number of small earthquakes in the database. To twart this issue, we focus on earthquakes with a magnitude of at least 5 ($M \geq 5$), and we used a subset of the data. To find the most appropriate cutoff date in the data, we used statistical tests of changes in structures (F test (Chow test), from (Andrews 1993) or (Zeileis, Kleiber, Kramer and Hornik 2003) for implementation issues). Based on these tests, events between January 1st, 1965 up to March 30th, 2011 have been kept in the sample. The total number of earthquakes is approximately 70 000. Finally, for $M \geq 6$ earthquakes, the time period considered is January 1st, 1992 up to March 30th, 2011, which amount to approximately 3000 events.

The proposed BINAR model will be largely used to investigate first-order autocorrelation and cross-autocorrelation in earthquake counts, which is equivalent to measuring the degree of the first order type of space and time contagion. To do this, earthquake counts have been computed at several frequencies. Time ranges of 3, 6, 12, 24, 36 and 48 hours have been considered to count the number of earthquakes.

[Figure 4 about here.]

5.2 Quality of fit

The proposed BINAR model encompasses various models as well. When, $p_{i,j} = 0, i, j = 1, 2$ and $\varphi = 0$, there is no serial correlation, no cross autocorrelation, and both series are independent. Those are two independent Poisson noises. When $\varphi \neq 0$, the Poisson noises are dependent. When $p_{1,2} = p_{2,1} = 0$ and $\varphi = 0$, there is serial correlation but both series are independent. Those are equivalent to two univariate INAR processes. Finally, when $p_{1,2} = p_{2,1} = 0$ and $\varphi \neq 0$, we find the diagonal BINAR model of (Pedeli and Karlis 2011*a*). In the latter model, there is no cross

autocorrelation. Thus, in this section we compare the fit of those four models, along with the proposed BINAR approach, on each of the 136 possible pairs of tectonic plates.

Table 5 shows the results of the likelihood ratio test (LRT) of 2 dependent Poisson noises and of 2 independent INARs, both compared with 2 independent Poisson noises. Each column shows descriptive statistics for various sampling frequencies. The meaning of each row is as follows: mean LRT across the 136 combinations of tectonic plates, standard deviation of LRTs, along with various quantiles (50%, 75%, 90%, 95%, 97.5%) and the proportion of combinations that is statistically significant. Thus, the value of the LRT provides an idea of how useful the added feature really is. The upper part of Table 5 shows that dependence in the noises is important for about 10% of the combinations of plates (most of them are contiguous). However, serial correlation is a much more important feature, even though noises are independent. This should have been expected given earthquake mechanics. Thus, the LRT shows that autocorrelation is one important component when analyzing earthquake counts at these sampling frequencies.

Sampling frequency also influences the degree of autocorrelation and cross autocorrelation. When the sampling frequency is h hours, all earthquakes on two tectonic plates will count towards φ in the time interval $[0, h]$ hours. All earthquakes occurring in the time interval $[h, 2h]$ will help find first degree autocorrelation and cross autocorrelation i.e. it should appear in the $\boldsymbol{P}$ matrix. Finally, all earthquakes occurring after $2h$ hours, would be accounted for if a second or third degree BINAR was considered. Thus, when h increases, ρ should become more significant. This is what we observe in the upper panel of Table 5. Even though the percentage of combinations that are significant very slightly increase, the mean LRT and higher percentiles of LRT tend to grow more importantly. Note that only a few combinations are such that tectonic plates are contiguous.

Table 6 shows similar computations for the diagonal BINAR model over the independent INAR model (contribution of φ), and for the proposed BINAR model over the diagonal BINAR model (contribution of $p_{1,2}$ and $p_{2,1}$). One sees that 6-13% of combinations of tectonic plates show a significant dependence in the noise (upper part of Table 6). Those are in large part contiguous plates, which explains the rather low percentage. For other combinations, plates are too far apart and independent INAR models are often sufficient. When we compare the proposed BINAR model to the diagonal model (lower part of Table 6), we find that 6-9% of the combination of plates show

cross autocorrelation. In other words, the non-diagonal terms in the $\boldsymbol{P}$ matrix, i.e. $p_{1,2}$ and $p_{2,1}$, are both statistically different from zero. In most of the cases, the combination of plates that had a significant fit to both models, are contiguous. We further investigate some of those in the next subsection.

5.3 Analysis of pairs of tectonic plates

In this section, we take a look at specific pairs of tectonic plates to observe parameters and interpret them. We will look at four different combinations of plates, which are all closely related to Japan. The Japanese area is one of the most seismically active regions of the world, being at the limit of 4 tectonic plates. Table 7 shows the (CMLE) parameter estimates at four different sampling frequencies, for the four combinations of plates. The 8th line of each panel displays the LRT over the diagonal BINAR model. A value of more than 5.99 is significant at a level of 95%, meaning that both cross autocorrelation terms are significant. The last two lines show the unconditional mean number of earthquakes per period on each plate.

The first and second order moments estimators for the Okhotsk and West Pacific plates are given in Table 8.

Let us focus on the Okhotsk and West Pacific plates, where the results are shown at the bottom left part of Table 7, and assume the sampling frequency is 24 hours. Thus, the daily number of earthquakes on the Okhotsk plate is explained by three sources: the number of earthquakes on the previous day on both plates, and a random noise effect. When no earthquake was observed on both plates on a given day, a Poisson r.v. with mean 0.16 earthquake per day will generate seismicity on the next day. The probability of observing one or more earthquakes by noise only is 15% in this case.

The interest of the proposed BINAR model lies in the representation of the spatial contagion effect between tectonic plates. Suppose that n earthquakes were observed on the Okhotsk plate on a given day, while m earthquakes were observed on the West Pacific plate on that same day. The number of earthquakes on the Okhotsk plate the next day will be the result of the convolution of a binomial(n, 0.0817) (autocorrelation of order 1), a binomial(m, 0.028) (cross autocorrelation of order 1) and a Poisson noise (mean 0.162). Thus, on average, the number of earthquakes on the

next day on the Okhotsk plate will be

$$0.0817n + 0.028m + 0.162.$$

Under a diagonal model estimated with CMLE (but not provided in Table 7), that quantity is

$$0.0922n + 0.1748$$

which ignores the contribution of the West Pacific plate's earthquakes. With $n \geq 0$ and $m \geq 1$, the diagonal model will understate the potential number of earthquakes on the Okhotsk plate. When m is relatively large, which is one of the cases we are interested in risk management, that understatement can be important. For example, if 3 earthquakes are observed on the West Pacific plate, and only one on the Okhotsk plate, then we have an average of 0.3277 earthquake on the latter plate with the proposed BINAR model, compared with 0.267 with the diagonal model.

Further, suppose a case where $n = 0$ and $m \geq 1$, in a context where we want to compare the mean number of earthquakes on the Okhotsk plate using the diagonal and proposed models. In that situation, the mean number of earthquakes in the proposed model is roughly 16% larger[5] than the mean number of earthquakes in the diagonal model, for each additional earthquake we observe on the West Pacific plate (i.e. 16% times m). Picking other sets of plates, even if the LRT is much smaller and still significant, will lead to similar analyses (Okhotsk and Amur at 12- or 24-hour frequency is one example among others).

One may be tempted to directly compare values of $p_{1,2}$ and $p_{2,1}$ and conclude that earthquakes on one plate provokes earthquakes on the other. However, the gross values of $p_{1,2}$ and $p_{2,1}$ are obviously influenced by the number of earthquakes on each plate. For example, at the 24-hour frequency, $p_{1,2} = 0.028$ and $p_{2,1} = 0.106$ so that one may mistakenly pretend that Okhotsk earthquakes generally provoke earthquakes on the West Pacific plate, and not the converse. However, there are approximately 3 times more earthquakes on the West Pacific plate than on the Okhotsk plate, meaning that $p_{1,2}$ has to be lower to compensate for the higher counts on the second plate. Thus, if one is interested in determining if earthquake counts on one plate determine the other, one should perform Granger causality tests.

[5] $\frac{0.028m+0.162}{0.1748} - 1$ is approximately $\frac{0.028m}{0.1748} = 0.16m$.

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

5.4 Foreshocks and aftershocks

As another application of the model, we illustrate the relationship between medium-size earthquakes (i.e. $5 \leq M \leq 6$) and large earthquakes ($M > 6$). Using the proposed BINAR model in this context will help understand how the size of a set of earthquakes at a given time period can help predict the size of future earthquakes. Most of the time, large earthquakes (mainshocks) are followed by aftershocks, which are usually smaller (medium-size or smaller). The inverse, in which case a medium-size earthquake may announce a larger earthquake, is usually less likely, but still regularly observed. Figure 5 illustrates this relationship between foreshocks, mainshocks and aftershocks.

[Figure 5 about here.]

As a first exercise, we have fitted the same five models, that is the proposed BINAR model, the diagonal BINAR model, independent INARs, dependent Poisson noises, and independent Poisson noises. According to the LRT, the fit of the diagonal BINAR model over independent Poisson noises is statistically significant for all tectonic plates, at all sampling frequencies. This is also the case when the diagonal BINAR is compared to independent INAR models. Finally, for all but a few tectonic plates and/or sampling frequencies, the diagonal BINAR model has a very significant fit over dependent Poisson noises.

Thus, for this application, we would like to measure if cross autocorrelation is important, i.e. if earthquake size on a given period helps explain future earthquake sizes. Table 9 shows the LRT for the proposed BINAR model over the diagonal model, for various sampling frequencies. A value larger than 5.99 means that $p_{1,2} \neq 0$ and $p_{2,1} \neq 0$, implying that large earthquakes are followed by medium-size earthquakes, and the opposite also holds. This is indeed the case in the large majority of tectonic plates, although this relationship clearly gets weaker when the sampling frequency goes

from 3-hours to 48-hours (last row of the table). This should have been expected given Omori's law, which explains the temporal decay of aftershock rates.

Table 10 shows the CMLE parameter estimates for all plates at the 12-hour frequency. The last two columns provide the unconditional mean number of medium ($5 \leq M \leq 6$) and large ($M > 6$) earthquakes. With a 12-hour sampling frequency, only the Coca and Somali plates have an unsignificant LRT at a level of 95%, meaning that for 15 plates, cross autocorrelation is important. Thus, one should not directly compare values of $p_{1,2}$ and $p_{2,1}$ since only Granger causality tests will provide the true significance of contagion between the two sets of data.

Let us illustrate the impact of cross autocorrelation for a given tectonic plate. Assume that on the Okhotsk plate, which seats beneath part of Japan, there is a large earthquake in the prior 12-hour period (and no medium-size earthquake). Then, cross autocorrelation will be the most important component of the mean number of earthquakes in the next period. Indeed, the expected number of medium earthquakes in the next period is $0.2444 + 0.0780 = 0.3224$ and cross autocorrelation will account for more than the two thirds of the total expectation. One can compare the size of $p_{1,2}$ and $p_{2,1}$ with the noise components (λs) and observe that the ratio is much larger in this section than in Section 5.3. Thus, cross sectional effects are a key element in this context. Finally, the ratio of the expected number of $M > 6$ earthquakes over $5 \leq M \leq 6$ earthquakes is on average (across plates) approximately 10 which is consistent with Gutenberg and Richter's law.

[Table 9 about here.]

[Table 10 about here.]

5.5 Risk management

In many risk management applications, such as the computation of premiums and reserves or the pricing of catastrophe derivatives, the total loss amount over a given area, region, or city is what matters most. One important driver of the total loss amount, is the total number of earthquakes over the area in question for various time horizons $[0, T]$. In this section, we compare the distribution of the sum of the number of earthquakes over a given area, for the diagonal and the proposed BINAR models. We do so for pairs of tectonic plates (see Section 5.3) where the LRT was statistically significant, otherwise the two models are too similar.

Two sets of tectonic plates are analyzed: (1) Okhotsk and West Pacific plates (Japan is at the limits of the West Pacific, Okhotsk, Philippine and Amur plates) and (2) South American and Nasca plates (which holds the South American continent and the West Coast of South America (Chile for example) is located at the limit of these two plates). Two extreme scenarios are generated. In the first set of plate, we assume that 23 earthquakes were observed on the Okhotsk tectonic plate and 46 were observed on the West Pacific plate (this is indeed what happened in the last 12 hours of March 10th, 2011). In the second set of plates, we assume that 24 earthquakes were observed on the South American plate, whereas only 3 were observed on the Nasca plate (this is what occurred on the second half of February 27th, 2010). Using 100 000 paths of a bivariate diagonal INAR and the proposed bivariate INAR models, we have computed the total number of earthquakes that occurred on both plates (of a given set), on the next T days ($T = 1, 3, 7, 14$ and 30). The results are shown in Table 11. The left (right) panel focuses on the first (second) set of tectonic plates. The numbers shown are $\mathbb{P}\left(\sum_{k=1}^{T}(N_{1,k}+N_{2,k}) \geq n \middle| \mathcal{F}_0\right)$ for various values of n.

One sees that the diagonal model really understates the number of earthquakes in the following days, especially in the tails. For example, in the first set of plates (Okhotsk and West Pacific), the probability of having a total of at least 20 earthquakes in the next day is 6.7% with the proposed model, whereas it is 0.7% with the diagonal model; it is a ten-fold increase. This increase is all due to the non-diagonal terms in the $\boldsymbol{P}$ matrix as it accounts for the cross auto-correlation between earthquake counts. A less dramatic increase is observed in the second set of plates (South America and Nasca). For example, the probability of having a total of at least 7 earthquakes over a week on both plates is 39.6% in the diagonal model whereas this probability is 44% in the proposed model. As expected, over the long-term, both processes converge to their equilibrium and the effect of the initial conditions seem to dissipate.

We now suppose that with both sets of plates, no earthquake occurred on a given day. Table 12 shows the results of $\mathbb{P}\left(\sum_{k=1}^{T}(N_{1,k}+N_{2,k}) \geq n \middle| \mathcal{F}_0\right)$ for $T = 14$ and 30 days. For smaller T values, the probabilities generated by the two models are very similar since it takes a lot of time to develop earthquakes and thus to observe cross-sectional effects. For the given T values, the probabilities are very similar for both models, with a slightly fatter tail for the proposed model in the first set of tectonic plates. In the second set of plates, the probabilities are too close to be able to conclude

of any difference.

In summary, we have also run different scenarios on other sets of plates and it confirms that the effect of the non-diagonal terms in the $\boldsymbol{P}$ matrix is to generate fatter tails in the sum of the number of earthquakes. This is very useful for short-term risk management applications such as the pricing of earthquake bonds and other derivatives. An underestimation of the number of earthquakes could mean arbitrage opportunities if the market model has a similar behavior to the diagonal model.

[Table 11 about here.]

[Table 12 about here.]

6. CONCLUSION

In this paper, we confirm the conclusion of (Parsons and Velasco 2011) claiming that very large earthquakes do not necessarily cause large ones at a very long distance. There might be contagion, but it will be within two close areas (e.g. contiguous tectonic plates), and over a short period of time (a few hours, perhaps a few days, but not much longer). Nevertheless, not taking into account possible spatial contagion between consecutive periods may lead to large underestimation of overall counts. In the context of foreshocks, mainshocks and aftershocks, we have also observed that major earthquakes might generate several medium-size earthquakes on the same tectonic plate, and also foreshocks might announce possible large earthquakes.

REFERENCES

Al-Osh, M., and Alzaid, A. (1987), "First-order integer-valued autoregressive process," *Journal of Time Series Analysis*, 8, 261–275.

Andrews, D. W. K. (1993), "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 61, 821–856.

Billingsley, A. (1961), *Statistical Inference for Markov Processes*, : University of Chicago Press.

Boucher, J., Denuit, M., and Guillen, M. (2008), "Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions," *Variance*, 2, 135–162.

Dion, J.-P., Gauthier, G., and Latour, A. (1995), "Branching processes with immigration and integer-valued time series," *Serdica Mathematical Journal*, 21, 123–136.

Du, J.-G., and Li, Y. (1991), "The integer-valued autoregressive (INAR(p)) model," *Journal of Time Series Analysis*, 12, 129–142.

Ferland, R., Latour, A., and Oraichi, D. . (2006), "Integer-valued GARCH process," *Journal of Time Series Analysis*, 27, 923–942.

Fokianos, K. (2011), "Count time series models," in *Handbook of Time Series Analysis*, : Wiley Interscience.

Franke, J., and Subba Rao, T. (1993), Multivariate first-order integer-valued autoregressions,, Technical report, Forschung Universitat Kaiserslautern.

Gardner, J., and Knopoff, I. (1974), "Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonean?," *Bulletin of the Seismological Society of America*, 64, 1363–1367.

Gauthier, G., and Latour, A. (1994), "Convergence forte des estimateurs des paramètres d'un processus GENAR(p)," *Annales des Sciences Mathématiques du Québec*, 18, 49–71.

Gourieroux, C., and Jasiak, J. (2004), "Heterogeneous INAR(1) model with application to car insurance," *Insurance: Mathematics & Economics*, 34, 177–192.

Heinen, A., and Rengifo, E. (2007), "Multivariate autoregressive modeling of time series count data using copulas," *Journal of Empirical Finance*, 14, 564–583.

Johnson, N., Kotz, S., and Balakrishnan, N. (1997), *Discrete Multivariate Distributions*, : Wiley Interscience.

Kagan, Y., and Jackson, D. (1991), "Long-term earthquake clustering," *Geophysical Journal International*, 104, 117–133.

Kocherlakota, S., and Kocherlakota, K. (1992), *Bivariate Discrete Distributions*, : CRC Press.

Latour, A. (1997), "The multivariate GINAR(p) process," *Advances in Applied Probability*, 29, 228–248.

Latour, A. (1998), "Existence and stochastic structure of a non-negative integer-valued autoregressive process," *Journal of Time Series Analysis*, 19, 439–455.

Lomnitz, C. (1974), *Global Tectonic and Earthquake Risk*, : Elsevier.

Lutkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, : Springer Verlag.

Mahamunulu, D. M. (1967), "A note on regression in the multivariate Poisson distribution," *Journal of the American Statistical Association*, 62, 251–258.

McKenzie, E. (1985), "Some simple models for discrete variate time series," *Water Resources Bulletin*, 21, 645–650.

Ogata, Y. (1988), "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Statistical Association*, 83, 9–27.

Orfanogiannaki, K., Karlis, D., and Papadopoulos, G. (2010), "Identifying Seismicity Levels via Poisson Hidden Markov Models," *Pure and Applied Geophysics*, 167, 919–931.

Parsons, T., and Velasco, A. (2011), "Absence of remotely triggered large earthquakes beyond the mainshock region," *Nature Geoscience*, 4, 312–316.

Pedeli, X., and Karlis, D. (2011*a*), "A bivariate Poisson INAR(1) model with application," *Statistical Modelling*, 11, 325–349.

Pedeli, X., and Karlis, D. (2011*b*), "On estimation for the bivariate Poisson INAR process," *Communications in Statistics, Simulation and Computation*, .

Rathbun, S. (2004), "Seismological modeling," in *Encyclopedia of Environmetrics*, : Wiley Interscience.

Rosenblatt, M. (1971), *Markov processes, Structure and Asymptotic Behavior*, : Springer Verlag.

Schoenberg, P. (2003), "Multidimensional Residual Analysis of Point Process Models for Earthquake Occurrences," *Journal of the American Statistical Association*, 98, 789795.

Steutel, F., and van Harn, K. (1979), "Discrete analogues of self-decomposability and stability," *Annals of Probability*, 7, 893–899.

Utsu, T. (1969), "Aftershocks and earthquake statistics (I) - Some parameters which characterize an aftershock sequence and their interrelations," *Journal of the Faculty of Science of Hokkaido University*, pp. 121–195.

Vere-Jones, D. (2010), "Foundations of Statistical Seismology," *Pure and Applied Geophysics*, 167, 645–653.

Weiβ, C. (2008), "Thinning operations for modeling time series of counts: a survey," *Advances in Statistical Analysis*, 92, 319–341.

Zeileis, A., Kleiber, C., Kramer, W., and Hornik, K. (2003), "Testing and dating of structural changes in practice," *Computational Statistics & Data Analysis,*, 44, 109–123.

Zhuang, J., Y., Y. O., and Vere-Jones, D. (2002), "Stochastic declustering of space-time earthquake occurrences," *Journal of the American Statistical Association*, 97, 369–380.

Zucchini, W., and MacDonald, I. (2009), *Hidden Markov Models for Time Series: An Introduction Using R*, : Chapman & Hall.

List of Figures

1 Number of earthquakes (magnitude exceeding 2.0) per 15 seconds, following a large earthquake (of magnitude 6.5) either close to the main shock (less than 2,000 km) or far away (more than 2,000 km). Counts were normalized so that the expected number of earthquakes before is 100 in the two regions. Plain lines are spline regressions, either before or after the main shock). 29
2 Distribution of estimators $\widehat{p}_{1,1}$, $\widehat{p}_{1,2}$, $\widehat{p}_{2,1}$, $\widehat{p}_{2,2}$, $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ and $\widehat{\varphi}$, as a function of the sample size n, case of *non*-diagonal $\boldsymbol{P}$ matrix. 30
3 Distribution of estimators $\widehat{p}_{1,1}$, $\widehat{p}_{1,2}$, $\widehat{p}_{2,1}$, $\widehat{p}_{2,2}$, $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ and $\widehat{\varphi}$, as a function of the sample size n, case of diagonal $\boldsymbol{P}$ matrix. 31
4 The 17 tectonic plates (North American, Eurasian, Okhotsk, Pacific (split in two, East and West), Amur, Indo-Australian, African, Indo-Chinese, Arabian, Philippine, Coca, Caribbean, Somali, South American, Nasca and Antarctic). 32
5 Number of earthquakes (magnitude exceeding 2.0) per 15 seconds, following a large earthquake (of magnitude 6.5), normalized so that the expected number of earthquakes before is 100. Plain lines are spline regressions, either before or after the main shock). 33

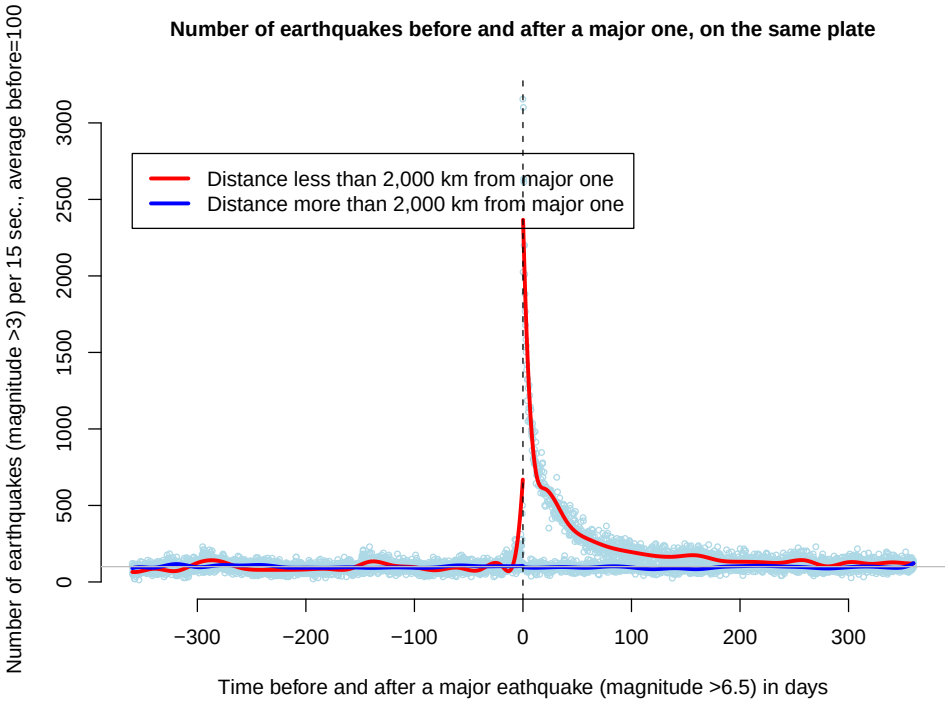

Figure 1: Number of earthquakes (magnitude exceeding 2.0) per 15 seconds, following a large earthquake (of magnitude 6.5) either close to the main shock (less than 2,000 km) or far away (more than 2,000 km). Counts were normalized so that the expected number of earthquakes before is 100 in the two regions. Plain lines are spline regressions, either before or after the main shock).

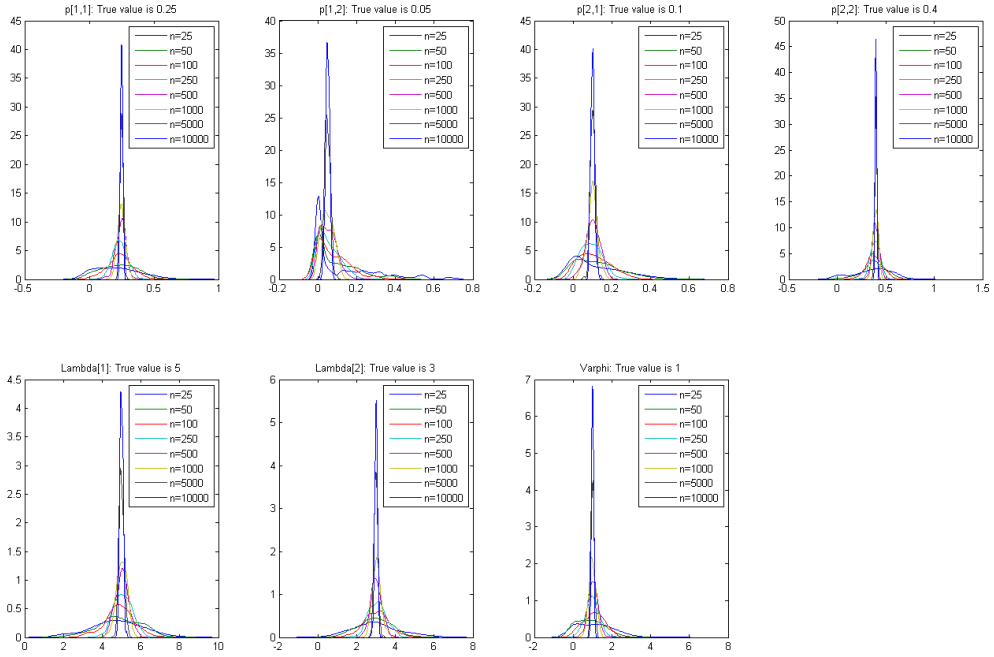

Figure 2: Distribution of estimators $\widehat{p}_{1,1}$, $\widehat{p}_{1,2}$, $\widehat{p}_{2,1}$, $\widehat{p}_{2,2}$, $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ and $\widehat{\varphi}$, as a function of the sample size n, case of *non*-diagonal $\boldsymbol{P}$ matrix.

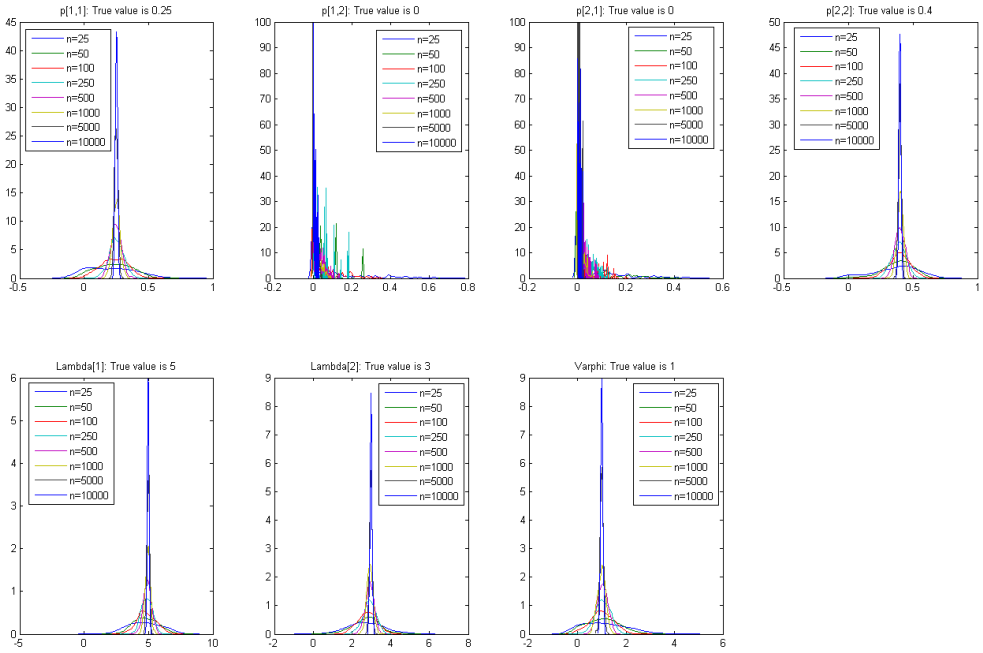

Figure 3: Distribution of estimators $\widehat{p}_{1,1}$, $\widehat{p}_{1,2}$, $\widehat{p}_{2,1}$, $\widehat{p}_{2,2}$, $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ and $\widehat{\varphi}$, as a function of the sample size n, case of diagonal $\boldsymbol{P}$ matrix.

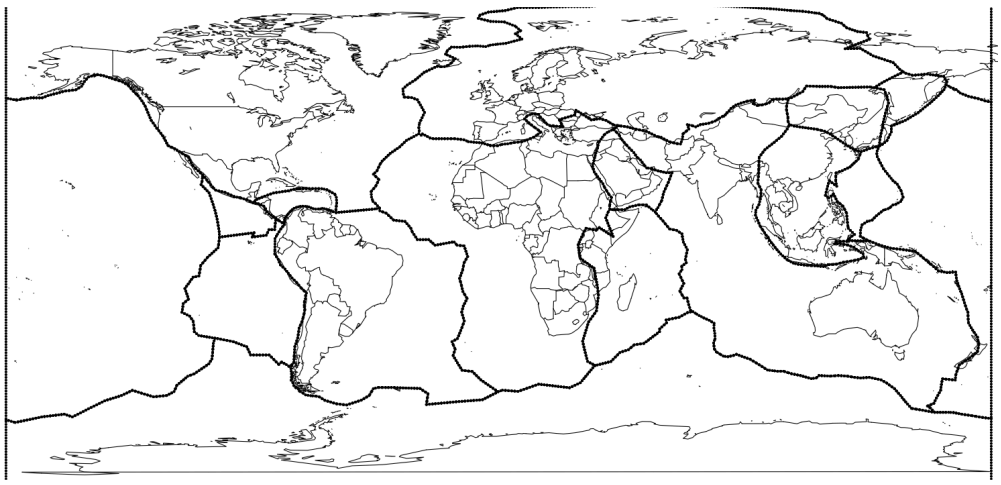

Figure 4: The 17 tectonic plates (North American, Eurasian, Okhotsk, Pacific (split in two, East and West), Amur, Indo-Australian, African, Indo-Chinese, Arabian, Philippine, Coca, Caribbean, Somali, South American, Nasca and Antarctic).

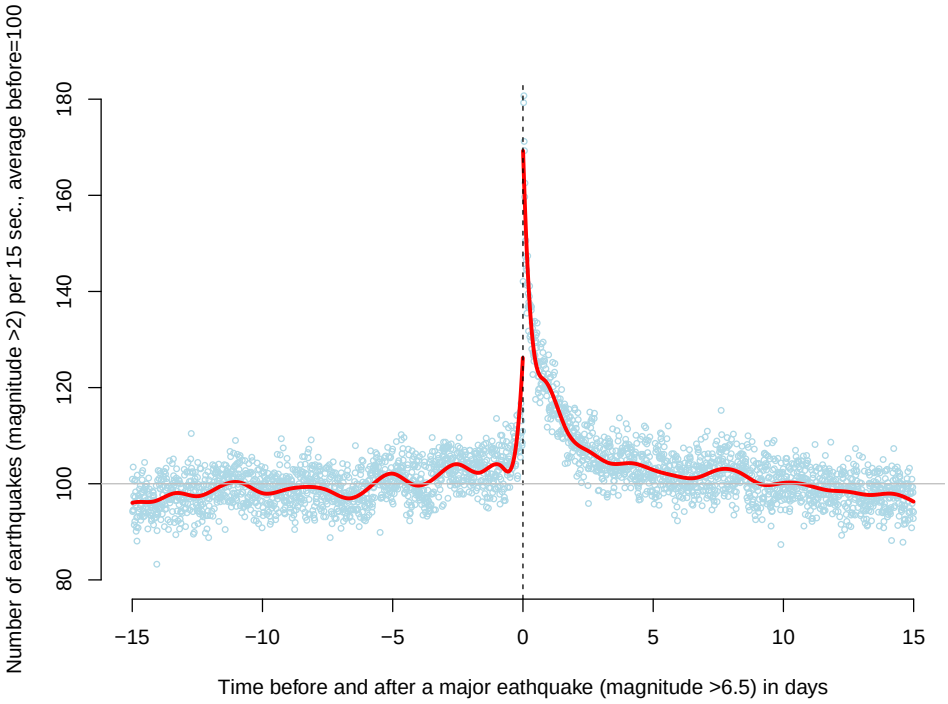

Figure 5: Number of earthquakes (magnitude exceeding 2.0) per 15 seconds, following a large earthquake (of magnitude 6.5), normalized so that the expected number of earthquakes before is 100. Plain lines are spline regressions, either before or after the main shock).

List of Tables

1 Mean parameter values - First parameter set 35
2 Standard deviation of parameter values - First parameter set 36
3 Mean parameter values - Second parameter set 37
4 Standard deviation parameter values - Second parameter set 38
5 Likelihood ratio test (1) independent Poisson vector (with $\lambda \neq 0$) over independent Poisson variables (2) two independent INAR processes versus two independent Poisson variables 39
6 Likelihood ratio test (1) diagonal BINAR over two independent INAR processes (2) proposed BINAR over diagonal BINAR, with Poisson innovation. 40
7 Estimation of parameters for counts of earthquakes on several tectonic plates, Okhotsk vs. Philippine; Okhotsk vs. Amur; Okhotsk vs. West Pacific; and Okhotsk vs. Indo-Chinese plates. Includes a likelihood ratio test (null: $\boldsymbol{P}$ is a diagonal matrix). . . . 41
8 First and second order moments, $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}(0)$ and cross-lagged correlations $\boldsymbol{\rho}(0)$ and $\boldsymbol{\rho}(1)$, for counts on two plates, Okhotsk vs. West Pacific. 42
9 Likelihood Ratio Test for the proposed BINAR model over the diagonal model, for various sampling frequencies, when $N_{1,t}$ denotes the number of medium size earthquakes (magnitude between 5 and 6) during period t, and $N_{1,t}$ denotes the number of large earthquakes (magnitude exceeding 6) during period t. 43
10 CMLE estimators for the proposed BINAR model, for 12-hour frequency, when $N_{1,t}$ denotes the number of medium size earthquakes (magnitude between 5 and 6) during period t, and $N_{1,t}$ denotes the number of large earthquakes (magnitude exceeding 6) during period t. 44
11 Empirical evolution of $\mathbb{P}\left(\sum_{k=1}^{T}(N_{1,k}+N_{2,k}) \geq n \middle| \mathcal{F}_0\right)$ for various values of n (per line) and T (per column), on two plates (Okhotsk vs. West Pacific and South American vs. Nasca), either for a diagonal $\boldsymbol{P}$ matrix - on top - or for a full matrix - below. 45
12 Empirical evolution of $\mathbb{P}\left(\sum_{k=1}^{T}(N_{1,k}+N_{2,k}) \geq n \middle| \mathcal{F}_0\right)$ for various values of n and two time horizon T, on two plates (Okhotsk vs. West Pacific and South American vs. Nasca), either for a diagonal $\boldsymbol{P}$ matrix, or for a full matrix. 46

Sample size n	$\widehat{p}_{1,1}$	$\widehat{p}_{1,2}$	$\widehat{p}_{2,1}$	$\widehat{p}_{2,2}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\varphi}$
25	21.27%	12.15%	14.54%	34.24%	4.8023	2.9854	1.1057
50	23.18%	9.97%	11.73%	38.62%	4.8538	2.9850	0.9704
100	23.29%	6.63%	10.11%	39.93%	5.0075	3.0081	1.0034
250	23.95%	5.52%	10.47%	39.56%	5.0531	2.9801	0.9774
500	24.57%	6.19%	10.18%	39.49%	4.9704	3.0318	1.0162
1000	24.93%	5.02%	10.09%	39.84%	5.0044	3.0040	0.9843
5000	24.88%	4.96%	9.95%	39.92%	5.0097	3.0086	0.9952
10000	25.06%	4.98%	10.08%	39.94%	4.9969	2.9972	1.0036
True value	25%	5%	10%	40%	5	3	1

Table 1: Mean parameter values - First parameter set

Sample size n	$\widehat{p}_{1,1}$	$\widehat{p}_{1,2}$	$\widehat{p}_{2,1}$	$\widehat{p}_{2,2}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\varphi}$
25	16.87%	15.43%	14.77%	19.42%	1.2599	1.1232	0.9243
50	13.28%	11.24%	10.58%	11.98%	1.0497	0.9038	0.7480
100	9.25%	7.89%	7.99%	8.35%	0.7336	0.6218	0.5344
250	6.04%	5.24%	5.52%	5.33%	0.5043	0.4063	0.3701
500	3.94%	4.44%	3.82%	3.72%	0.3601	0.3106	0.2516
1000	2.94%	3.22%	2.74%	2.55%	0.2587	0.2144	0.1813
5000	1.37%	1.54%	1.17%	1.19%	0.1148	0.1002	0.0766
10000	0.92%	1.00%	0.83%	0.84%	0.0841	0.0660	0.0568

Table 2: Standard deviation of parameter values - First parameter set

Sample size n	$\widehat{p}_{1,1}$	$\widehat{p}_{1,2}$	$\widehat{p}_{2,1}$	$\widehat{p}_{2,2}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\varphi}$
25	20.87%	11.18%	8.15%	36.00%	4.6948	2.7182	1.1375
50	23.09%	6.10%	5.29%	38.16%	4.8105	2.7352	1.0503
100	23.22%	5.47%	3.20%	39.66%	4.8545	2.7997	1.0095
250	25.08%	2.59%	2.09%	39.24%	4.8667	2.9008	1.0442
500	24.76%	1.98%	1.38%	39.91%	4.9131	2.9150	1.0333
1000	24.93%	1.42%	1.00%	40.22%	4.9382	2.9211	0.9906
5000	24.84%	0.72%	0.38%	40.00%	4.9740	2.9743	1.0017
10000	25.05%	0.44%	0.34%	39.92%	4.9738	2.9820	1.0018
True value	25%	0%	0%	40%	5	3	1

Table 3: Mean parameter values - Second parameter set

Sample size n	$\widehat{p}_{1,1}$	$\widehat{p}_{1,2}$	$\widehat{p}_{2,1}$	$\widehat{p}_{2,2}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\varphi}$
25	17.48%	16.80%	11.65%	17.46%	1.2515	0.9555	0.8950
50	13.42%	11.42%	7.75%	12.09%	0.9161	0.7196	0.6621
100	10.05%	7.83%	4.83%	8.07%	0.7286	0.4765	0.4751
250	5.71%	4.19%	3.08%	5.37%	0.4263	0.3293	0.3125
500	4.07%	2.73%	2.04%	3.62%	0.3101	0.2191	0.2048
1000	2.82%	2.00%	1.36%	2.48%	0.1981	0.1605	0.1624
5000	1.33%	0.97%	0.61%	1.14%	0.0992	0.0702	0.0696
10000	0.88%	0.68%	0.53%	0.78%	0.0635	0.0519	0.0446

Table 4: Standard deviation parameter values - Second parameter set

Likelihood ratio test - Dependent Poisson over independent Poisson

	3 hours	6 hours	12 hours	24 hours	36 hours	48 hours
Mean	6.9755	8.9167	9.7735	9.9896	10.6302	10.4286
Stdev	67.8029	76.2997	83.3619	85.4670	91.0044	88.9381
50%	0.0000	0.0000	0.0000	0.0000	0.0000	0.0087
75%	0.0315	0.4181	0.5951	1.0178	1.0533	1.2933
90%	1.6329	3.6041	4.3060	3.7918	4.1652	4.7956
95%	5.1601	6.6625	11.2303	11.7174	12.8974	10.5802
97.5%	9.9646	16.9966	18.8452	22.7211	24.8961	19.5306
% > 3.84	7.35%	10.29%	12.50%	10.29%	11.03%	12.50%

Likelihood ratio test - independent INARs over independent Poisson

	3 hours	6 hours	12 hours	24 hours	36 hours	48 hours
Mean	1215.16	1150.81	1036.95	864.04	557.15	542.15
Stdev	1084.90	1082.05	964.69	798.20	483.23	456.70
50%	851.01	781.32	735.33	561.29	399.63	416.24
75%	1630.64	1497.45	1415.05	1173.66	819.89	745.15
90%	2979.68	3030.71	2678.53	2147.04	1423.80	1319.45
95%	3227.93	3170.56	2837.07	2308.66	1515.82	1435.19
97.5%	3551.20	3589.28	3213.06	2580.33	1761.71	1650.87
% > 5.99	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Table 5: Likelihood ratio test (1) independent Poisson vector (with $\lambda \neq 0$) over independent Poisson variables (2) two independent INAR processes versus two independent Poisson variables

Likelihood ratio test - diagonal BINAR over independent INARs

	3 hours	6 hours	12 hours	24 hours	36 hours	48 hours
Mean	3.3744	4.8765	5.4843	6.3690	9.1950	8.0845
Stdev	26.9264	36.7864	42.7106	52.6144	72.6791	63.8500
50%	0.0048	0.0391	0.0955	0.0171	0.2013	0.0337
75%	0.5303	0.5109	0.7107	0.8735	1.1663	1.0702
90%	1.8276	2.7837	4.4992	5.0022	4.2161	4.3279
95%	4.9423	4.7359	8.3814	9.9875	10.7675	8.4055
97.5%	9.2657	15.0495	13.3836	13.4121	24.5204	16.5784
% > 3.84	6.62%	8.82%	12.50%	13.24%	10.83%	10.83%

Likelihood ratio test - proposed BINAR over diagonal BINAR

	3 hours	6 hours	12 hours	24 hours	36 hours	48 hours
Mean	4.9409	4.1814	3.7077	4.0927	2.8335	3.5242
Stdev	30.4265	25.5655	23.8860	22.6459	12.6545	15.1581
50%	0.9720	0.5379	0.3533	0.4504	0.3631	0.5931
75%	2.6904	2.3514	1.6824	2.4026	2.0943	2.3357
90%	5.2349	5.2194	4.0567	4.7533	4.4141	5.3640
95%	9.9654	7.8503	6.4292	8.5268	6.5271	8.7984
97.5%	15.7839	15.6885	12.0481	11.9986	11.9423	18.2802
% > 5.99	8.09%	8.82%	7.35%	8.82%	6.67%	9.02%

Table 6: Likelihood ratio test (1) diagonal BINAR over two independent INAR processes (2) proposed BINAR over diagonal BINAR, with Poisson innovation.

Plates	**Okhotsk (#1) vs. Philippine (#2)**				**Okhotsk (#1) vs. Amur (#2)**			
Params/Frequency	3 hours	12 hours	24 hours	48 hours	3 hours	12 hours	24 hours	48 hours
$\widehat{p}_{1,1}$	7.44%	9.45%	10.38%	12.81%	7.44%	9.44%	10.30%	12.75%
$\widehat{p}_{1,2}$	0.61%	0.60%	1.15%	0.00%	0.35%	0.83%	3.06%	2.31%
$\widehat{p}_{2,1}$	0.00%	0.00%	0.00%	0.00%	0.16%	0.42%	0.44%	0.40%
$\widehat{p}_{2,2}$	3.87%	5.83%	8.52%	8.80%	4.68%	6.44%	8.67%	10.59%
$\widehat{\lambda}_1$	0.0222	0.0868	0.1711	0.3358	0.0223	0.0871	0.1720	0.3348
$\widehat{\lambda}_2$	0.0156	0.0612	0.1187	0.2368	0.0032	0.0122	0.0237	0.0466
$\widehat{\varphi}$	0.0000	0.0001	0.0000	0.0021	0.0000	0.0003	0.0009	0.0024
LRT (over diag.)	4.1106	1.6737	3.6011	0.0000	2.8874	9.4113	9.5405	4.0631
Uncond. mean (#1)	0.0241	0.0963	0.1926	0.3852	0.0241	0.0963	0.1927	0.3852
Uncond. mean (#2)	0.0162	0.0650	0.1298	0.2596	0.0034	0.0134	0.0269	0.0538
Plates	**Okhotsk (#1) vs. West Pacific (#2)**				**Okhotsk (#1) vs. IndoChinese (#2)**			
Params/Frequency	3 hours	12 hours	24 hours	48 hours	3 hours	12 hours	24 hours	48 hours
$\widehat{p}_{1,1}$	6.12%	7.18%	8.17%	10.13%	7.45%	9.46%	10.36%	12.83%
$\widehat{p}_{1,2}$	1.85%	2.85%	2.80%	3.13%	0.02%	0.28%	0.24%	0.10%
$\widehat{p}_{2,1}$	5.84%	7.56%	10.60%	9.74%	0.22%	0.40%	0.00%	0.75%
$\widehat{p}_{2,2}$	10.71%	13.52%	15.52%	15.67%	6.71%	10.29%	11.58%	13.68%
$\widehat{\lambda}_1$	0.0214	0.0818	0.1620	0.3132	0.0223	0.0863	0.1710	0.3344
$\widehat{\lambda}_2$	0.0576	0.2212	0.4261	0.8539	0.0767	0.2948	0.5818	1.1326
$\widehat{\varphi}$	0.0012	0.0098	0.0269	0.0739	0.0002	0.0015	0.0046	0.0099
LRT (over diag.)	352.5998	275.2342	257.0215	157.0995	0.2839	3.0150	1.2208	0.6136
Uncond. mean (#1)	0.0241	0.0963	0.1926	0.3852	0.0241	0.0963	0.1926	0.3852
Uncond. mean (#2)	0.0661	0.2643	0.5285	1.0570	0.0823	0.3290	0.6580	1.3155

Table 7: Estimation of parameters for counts of earthquakes on several tectonic plates, Okhotsk vs. Philippine; Okhotsk vs. Amur; Okhotsk vs. West Pacific; and Okhotsk vs. Indo-Chinese plates. Includes a likelihood ratio test (null: $\boldsymbol{P}$ is a diagonal matrix).

Plates	**Okhotsk (#1) vs. West Pacific (#2)**			
Params/Frequency	3 hours	12 hours	24 hours	48 hours
$\mathbb{E}(N_{1,t})$	0.024	0.096	0.192	0.385
$\mathbb{E}(N_{2,t})$	0.065	0.264	0.528	1.057
$\text{var}(N_{1,t})$	0.022	0.084	0.167	0.326
$\text{var}(N_{2,t})$	0.060	0.239	0.466	0.934
$\text{cor}(N_{1,t}, N_{2,t})$	0.038	0.079	0.110	0.150
$\text{cor}(N_{1,t}, N_{1,t-1})$	0.062	0.075	0.086	0.109
$\text{cor}(N_{2,t}, N_{2,t-1})$	0.108	0.138	0.162	0.165
$\text{cor}(N_{1,t}, N_{2,t-1})$	0.033	0.053	0.055	0.068

Table 8: First and second order moments, $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}(0)$ and cross-lagged correlations $\boldsymbol{\rho}(0)$ and $\boldsymbol{\rho}(1)$, for counts on two plates, Okhotsk vs. West Pacific.

Plate name	**3 hours**	**6 hours**	**12 hours**	**24 hours**	**36 hours**	**48 hours**
North American	60.9470	31.3554	21.6005	16.5618	15.9287	6.0150
Eurasian	3.4172	2.4860	17.5732	1.3990	8.4201	1.0743
Okhotsk	135.5948	109.1666	109.3060	113.5049	36.3703	52.2677
East Pacific	37.2827	50.2991	32.2566	19.9613	19.1339	4.4437
West Pacific	101.3846	96.3865	110.2205	109.0303	62.4744	81.3029
Amur	12.9162	17.2652	7.7396	4.0498	10.2767	15.0012
Indo-Australian	303.2257	233.9429	169.1037	124.9187	75.0355	48.7183
African	35.0197	11.1661	15.9194	12.7146	28.1233	9.0930
Indo-Chinese	63.2515	29.9391	49.5970	64.0781	29.8289	45.1555
Arabian	4.5921	4.5763	12.3768	3.1358	2.1765	0.1744
Philippine	9.2969	21.3144	20.1805	15.5858	18.9310	17.5329
Coca	12.6070	15.1147	3.1709	8.2198	5.3469	9.0246
Caribbean	20.4764	24.5509	21.2256	3.4367	7.6112	2.4771
Somali	0.2432	5.1726	3.2162	0.0039	0.1625	0.0392
South American	81.8145	58.0135	50.4781	55.8060	77.3867	58.4621
Nasca	76.8393	38.6514	20.2549	17.6382	8.6659	11.1903
Antarctic	2.9275	9.2410	9.1584	4.7911	1.8339	0.9290
Average LRT	56.5786	44.6260	39.6105	33.8139	23.9827	21.3471

Table 9: Likelihood Ratio Test for the proposed BINAR model over the diagonal model, for various sampling frequencies, when $N_{1,t}$ denotes the number of medium size earthquakes (magnitude between 5 and 6) during period t, and $N_{1,t}$ denotes the number of large earthquakes (magnitude exceeding 6) during period t.

Plate name	$\widehat{p}_{1,1}$	$\widehat{p}_{1,2}$	$\widehat{p}_{2,1}$	$\widehat{p}_{2,2}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\varphi}$	$\widehat{p}_{1,2}/\widehat{p}_{2,1}$	**Mean<6**	**Mean>6**
North American	0.05633	0.11372	0.00444	0.01027	0.0844	0.0091	0.0028	25.63	0.0906	0.0096
Eurasian	0.01348	0.14431	0.01082	0.00006	0.0181	0.0011	0.0002	13.34	0.0185	0.0013
Okhotsk	0.11224	0.24445	0.00995	0.01951	0.0780	0.0104	0.0033	24.57	0.0910	0.0115
East Pacific	0.07950	0.12959	0.00385	0.00025	0.2631	0.0285	0.0075	33.64	0.2900	0.0296
West Pacific	0.15688	0.21797	0.00642	0.01163	0.1995	0.0212	0.0084	33.93	0.2426	0.0231
Amur	0.00931	0.09470	0.00676	0.02041	0.0107	0.0024	0.0008	14.02	0.0110	0.0025
Indo-Australian	0.19749	0.24562	0.01079	0.03095	0.4039	0.0490	0.0225	22.76	0.5205	0.0564
African	0.03906	0.13683	0.00204	0.00716	0.0564	0.0054	0.0014	67.20	0.0595	0.0056
Indo-Chinese	0.09744	0.16198	0.00563	0.00956	0.2501	0.0236	0.0080	28.76	0.2816	0.0254
Arabian	0.04026	0.24457	0.00347	0.00009	0.0167	0.0007	0.0001	70.51	0.0176	0.0008
Philippine	0.03630	0.09681	0.00864	0.03512	0.0536	0.0055	0.0012	11.20	0.0563	0.0062
Coca	0.06228	0.04115	0.00155	0.00534	0.0439	0.0069	0.0020	26.61	0.0471	0.0070
Caribbean	0.03080	0.26310	0.00001	0.00009	0.0083	0.0008	0.0004	31969	0.0088	0.0008
Somali	0.02325	0.02809	0.00384	0.00000	0.0284	0.0012	0.0001	7.32	0.0291	0.0013
South American	0.13661	0.12043	0.01141	0.01507	0.1384	0.0160	0.0046	10.55	0.1628	0.0181
Nasca	0.11426	0.13361	0.00307	0.01442	0.0378	0.0034	0.0013	43.49	0.0433	0.0036
Antarctic	0.02875	0.03897	0.00879	0.00153	0.0548	0.0056	0.0010	4.43	0.0567	0.0061

Table 10: CMLE estimators for the proposed BINAR model, for 12-hour frequency, when $N_{1,t}$ denotes the number of medium size earthquakes (magnitude between 5 and 6) during period t, and $N_{1,t}$ denotes the number of large earthquakes (magnitude exceeding 6) during period t.

Diagonal model (Okhotsk and West Pacific)						**Diagonal model (South American and Nasca)**					
n / days	1 day	3 days	7 days	14 days	30 days	n / days	1 day	3 days	7 days	14 days	30 days
5	0.9680	0.9869	0.9978	0.9999	1.0000	2	0.8489	0.9166	0.9757	0.9981	1.0000
10	0.5650	0.7207	0.8972	0.9884	0.9999	5	0.2708	0.4321	0.6965	0.9277	0.9988
15	0.1027	0.2270	0.4978	0.8548	0.9985	7	0.0685	0.1628	0.3959	0.7655	0.9906
20	0.0067	0.0277	0.1308	0.4997	0.9752	10	0.0035	0.0192	0.1041	0.4108	0.9334
25	0.0003	0.0018	0.0170	0.1684	0.8588	15	0.0000	0.0002	0.0033	0.0547	0.5885
30	0.0000	0.0001	0.0014	0.0319	0.5965	20	0.0000	0.0000	0.0000	0.0031	0.1873
40	0.0000	0.0000	0.0000	0.0002	0.1034	25	0.0000	0.0000	0.0000	0.0001	0.0290
50	0.0000	0.0000	0.0000	0.0000	0.0041	30	0.0000	0.0000	0.0000	0.0000	0.0030
Proposed model (Okhotsk and West Pacific)						**Proposed model (South American and Nasca)**					
n / days	1 day	3 days	7 days	14 days	30 days	n / days	1 day	3 days	7 days	14 days	30 days
5	0.9946	0.9977	0.9997	1.0000	1.0000	2	0.8780	0.9321	0.9805	0.9979	1.0000
10	0.8344	0.9064	0.9712	0.9970	1.0000	5	0.3323	0.4888	0.7331	0.9362	0.9990
15	0.3638	0.5288	0.7548	0.9479	0.9995	7	0.0990	0.2034	0.4410	0.7913	0.9921
20	0.0671	0.1573	0.3616	0.7256	0.9917	10	0.0082	0.0309	0.1271	0.4435	0.9386
25	0.0053	0.0246	0.0970	0.3815	0.9357	15	0.0000	0.0004	0.0056	0.0688	0.6145
30	0.0002	0.0023	0.0151	0.1268	0.7646	20	0.0000	0.0000	0.0001	0.0039	0.2099
40	0.0000	0.0000	0.0001	0.0038	0.2335	25	0.0000	0.0000	0.0000	0.0001	0.0380
50	0.0000	0.0000	0.0000	0.0001	0.0221	30	0.0000	0.0000	0.0000	0.0000	0.0036

Table 11: Empirical evolution of $\mathbb{P}\left(\sum_{k=1}^{T}(N_{1,k}+N_{2,k}) \geq n \middle| \mathcal{F}_0\right)$ for various values of n (per line) and T (per column), on two plates (Okhotsk vs. West Pacific and South American vs. Nasca), either for a diagonal $\boldsymbol{P}$ matrix - on top - or for a full matrix - below.

Diagonal model					
(Okhotsk and West Pacific plates)			**(South American and Nasca plates)**		
n / days	14 days	30 days	n / days	14 days	30 days
5	0.9495	1.0000	3	0.9096	0.9991
10	0.5281	0.9943	5	0.6710	0.9904
15	0.1138	0.9125	7	0.3733	0.9506
20	0.0121	0.6302	10	0.0962	0.7770
25	0.0008	0.2791	12	0.0296	0.5883
30	0.0000	0.0774	15	0.0038	0.2996
35	0.0000	0.0137	20	0.0001	0.0498
40	0.0000	0.0018	25	0.0000	0.0036
45	0.0000	0.0002	30	0.0000	0.0002
50	0.0000	0.0000	35	0.0000	0.0000
Proposed model					
(Okhotsk and West Pacific plates)			**(South American and Nasca plates)**		
n / days	14 days	30 days	n / days	14 days	30 days
5	0.9444	0.9999	3	0.9061	0.9990
10	0.5211	0.9927	5	0.6662	0.9899
15	0.1261	0.9033	7	0.3754	0.9497
20	0.0139	0.6242	10	0.0990	0.7745
25	0.0007	0.2877	12	0.0317	0.5856
30	0.0000	0.0870	15	0.0049	0.2986
35	0.0000	0.0177	20	0.0001	0.0513
40	0.0000	0.0024	25	0.0000	0.0044
45	0.0000	0.0003	30	0.0000	0.0002
50	0.0000	0.0001	35	0.0000	0.0000

Table 12: Empirical evolution of $\mathbb{P}\left(\sum_{k=1}^{T}(N_{1,k}+N_{2,k}) \geq n \middle| \mathcal{F}_0\right)$ for various values of n and two time horizon T, on two plates (Okhotsk vs. West Pacific and South American vs. Nasca), either for a diagonal $\boldsymbol{P}$ matrix, or for a full matrix.