University of Bergen

Department of Mathematics

Count Time Series

With application to corporate defaults

Author:
Elyas Dawud

Supervisor:
Bård Støve

Master's Thesis in Statistics

June 12, 2020

Abstract

Time series models are widely used and studied. However time series of counts have only recently garnered much interest. In this thesis we look at both univariate and multivariate count time series that are approached thorough the generalized linear models framework. Specifically we look at copula based multivariate models with Poisson marginals, and models including covariates time series. We examine large sample properties of the models by simulation. Furthermore, we use the models to study corporate defaults in the US, where we find evidence for contagion effects. By contagion effect is meant, one firms defaulting impact on other firms defaulting. We find that contagion effects are present even after controlling for financial covariats that explain the systematic risks that the firms face.

Acknowledgements

First of all I want to thank my supervisor Bård Støve for suggesting this interesting topic and for his support and guidance . I will also like thank my fellow students at the 4Th floor for many discussions and engaging lunch breaks.

It is said that it takes a village to raise a child, and i was raised by a number of people who would constitute a village, without whom i wouldn't have been here today. For that, I want to thank my mother, my grandmother, and my many aunts and cousins for the role they played in my upbringing.

Contents

1. Introduction

Counts in time occur frequently in nature and therefore count time series is an interesting topic of research. Models of count time series have been applied in different fields both in social and natural sciences. See for example, Agosto, Cavaliere, Kristensen, and Rahbek, 2016 for a financial application, A. M. Schmidt and Pereira, 2011 for a medical application and Ravishanker, Venkatesan, and Hu, 2015 for a marketing application. However, despite count time series being interesting topic of research, the theory and methods to study them is not as very well established as for continuous-valued time series. Focusing on autoregressive models, the more obvious challenge lies in obtaining a model that results in a non-negative integer-valued process. Consider the autoregressive model

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_i \tag{1.1}$$

where both $\{Y_t\}$ and $\{\epsilon_i\}$ are assumed to be continuous-valued and both possibly taking negative and positive values. An approach that carries over the autoregressive structure to discrete valued time series in the so called integer-valued autoregressive process, INAR, introduced by McKenzie, 1985 and Alzaid and Al-Osh, 1988. Since the above recursion cannot be applied to the integer-valued case, Because the multiplication of an integer by a real number usually results in a non-integer value. Thus, one needs to replace the scalar multiplication in the recursion (1.1) by a different operation with similar properties, such operations are thinning operations. A simple example is the INAR(1) model, that is based on the Binomial thinning operation process, defined by the recursion

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t \tag{1.2}$$

where

$$\alpha \circ Y_{t-1} = \sum_{i=1}^{Y_{t-1}} X_i.$$

The variables $X_i, i \in \mathbb{N}$ are independent and identically distributed (IID) Bernoulli variables with success probability $\alpha \in [0, 1]$, independent of the innovations $\{\epsilon_t\}$. Thus, conditionally on Y being given, $\alpha \circ Y$ is a binomial variable. A through review of thinning operators is given in Weiß, 2008. The Binomial thinning operator is the most common amongst others, because it is particularly suited for Poisson marginal and has an intuitive interpretation. Consider a population of size Y at a some time t. If the population have shrunk when observed at a letter point in time, say $t + 1$, which means that some individuals have died in the interval $[t, t + 1]$. If the probability of death is independent for individuals in the

population, then the probability of dying in the interval $[t, t+1]$ for each individual is $1-\alpha$. Then the number of survivors is given by $\alpha \circ Y$.

However the INAR approach is still restrictive in comparison to (1.1). For example the correlation is always positive and multivariate analogues of the models are not straight forward. In this thesis we work with an alternative approach of modeling count time series, namely generalized linear models. These models belong the class of *observation driven* models. We describe the approach in more detain in section 2. With this approach it is possible to construct models that allow for negative correlation and allow covariates to enter the model in a straight forward manner. See Kedem and Fokianos, 2005 for more on time series that follow generalized linear models.

The rest of the thesis is structured as follows; In sections 3 and 4 we discuss the linear and the log-linear models respectively, Key references are Fokianos, Rahbek, and Tjøstheim, 2009 and Fokianos and Tjøstheim, 2011. Sections 5 - 6 are preliminaries, where we introduce copulas, automatic differentiation, and local Gaussian correlation. A reader familiar with these concepts may skip these sections. In section 8 we study the multivariate analogues of the latter models by simulation, including extensions of the models that include covariates. In section 9 we provide application examples, specifically applications to counts of stocks Coca-cola and corporate defaults in the US. In section 10 an attempt is made to estimate copula parameters using a novel approach based on local Gaussian correlation suggested by Fokianos, Støve, Tjøstheim, Doukhan, et al., 2020.

2. Generalized Linear Models

Generalized linear models first introduced by Nelder and Wedderburn (1972) relaxes many of the assumptions in linear regression models. Particulary the assumption that the response variable is normally distributed. There are also cases where the response variable is categorical or counts, in which the simple linear regression is not suitable. In the generalized linear models one assumes that the response variable y is independently distributed with a distribution that belongs to the exponential family, i.e the distribution can be represented in the following form;

$$f(y) = c(y, \phi)\exp(\frac{y\theta - b(\theta)}{\phi}) \tag{2.1}$$

Generalized linear models have generally the following theree componenets

- Random Component.Specifies the distribution of the response variable, that belongs to the exponential family.
- Systematic Componenet. This is the same as the right hand side in linear regression. $\eta_i = \mathbf{x}_i^T\beta$, where $\mathbf{x}_i$ is a vector of covariates and β is a vector of model parameters.
- Link Function. As the name suggests this is a function g that connects the random component to the systematic component.

Thus a generalized linear model has the form,

$$g(\mu_i) = \mathbf{x}_i^T\beta, \text{ where } \mu_i = \mathrm{E}(\mathrm{y}_i) \tag{2.2}$$

2.1. Generalized Linear Models and Time Series

As mentioned above generalized linear models do away some of the restrictions encountered in simple linear regression models. Although generalized linear models were originally meant for independent data, the approach can still be extended to dependent data under some appropriate assumptions. There is a comprehensive treatment of generalized linear models for time series data in Kedem and Fokianos, 2005.

In order to extend generalized linear models ideas to time series, let us start by defining some components. let $\{Y_t\}$ be a time series of interest and let

$$\mathbb{Z}_{t-1} = (Z_{(t-1)1}, ..., Z_{(t-1)p})^{\mathsf{T}}$$

be the corresponding p-dimensional vector of past explanatory variables or covariates, $t = 1, ..., N$. In addition let

$$\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, Y_{t-2}, ..., \mathbb{Z}_{t-1}, \mathbb{Z}_{t-2}, ...\}$$

denote all information available at time $t-1$. Then the aim is to relate the conditional expectation of the response variable given past information,i.e $\mu = \mathrm{E}(Y_t|\mathcal{F}_{t-1})$ to the covariates.

We proceed by extending the key concepts of generalized linear models, exponential family of distributions and monotone link functions to time series data as we did in the case of independent data. Thus time series following generalized linear models have, using the same terminology a *random component*,that is the distribution of the response given past information belongs to the exponential family. That is, it can be represented in the following form

$$f(y_t; \theta_t, \phi|\mathcal{F}_{t-1}) = \exp\Big\{\frac{y_t\theta_t - b(\theta_t)}{\alpha_t(\phi)} + c(y_t; \phi)\Big\} \tag{2.3}$$

Furthermore it has a *systematic component* and a *link function* that relates the conditional mean to the systematic component as follows (see ch.1,2 Kedem and Fokianos, 2005)

$$g(\mu_t) = \eta_t = \sum_{i=1}^{p} \beta_i Z_{(t-1)i} = \mathbb{Z}_{t-1}^{\mathsf{T}}\boldsymbol{\beta} \tag{2.4}$$

2.2. Modified Likelihood functions

In extending the generalized linear models framework to dependent data, the main challenge can be obtaining a convenient likelihood function. Likelihood function defined as the joint distribution of the observed data given unknown parameters can be difficult to obtain, particularly when the data are dependent, and the nature of dependence is not understood. To resolve this problem one seeks to modify the likelihood function by means of some clever conditioning(Kedem and Fokianos, 2005). An example of such conditioning is partial likelihood, which is advanced by Cox, 1975. Here we follow Kedem and Fokianos, 2005 to motivate the partial likelihood. Suppose we have a time series data $\{Y_t, t = 1, ..., N\}$ with a joint density $f_{\boldsymbol{\theta}}(y_1, ..., y_n)$, where $\boldsymbol{\theta}$ is a vector of parameters. Let also $\mathbb{I}$ be auxiliary information known to the observer in the period of observation. Then the likelihood function is defined by

$$f_{\boldsymbol{\theta}}(y_1, ..., y_N|\mathbb{I}) = f_{\boldsymbol{\theta}}(y_1|\mathbb{I}) \prod_{t=2}^{N} f_{\boldsymbol{\theta}}(y_t|y_1, ..., y_{t-1}, \mathbb{I}) \tag{2.5}$$

When auxiliary information is not available or is not useful, the likelihood function reduces to the following eqation.

$$f_{\boldsymbol{\theta}}(y_1, ..., y_N) = f_{\boldsymbol{\theta}}(y_1) \prod_{t=2}^{N} f_{\boldsymbol{\theta}}(y_t|y_1, ..., y_{t-1}) \tag{2.6}$$

The main challenge with (2.6) is that if no other additional assumptions are made, as the number of observations increases so does the number of parameters. This raises the issue of consistency, as well as modelling issues. This however is resolved if the conditional dependence in the time series is limited and we have a fixed number of parameters.

Some assumptions and modifications are needed to make (2.6) also suitable for dependent data. A Markov dependence of some order would result in limiting the conditional dependence. suppose we have a first order markov process $Y_t, t = 1, 2, ..., N$ with joint density $f_{\boldsymbol{\theta}}(y_1, ..., y_N)$ where $\boldsymbol{\theta}$ is a fixed parameter vector. The density reduces to

$$f_{\boldsymbol{\theta}}(y_1, ..., y_N) = f_{\boldsymbol{\theta}}(y_1) \prod_{t=2}^{N} f_{\boldsymbol{\theta}}(y_t|y_{t-1}) \tag{2.7}$$

by the markov assumption. If we we ignore the fist term of (2.7) since it does not depend on N, inference about $\boldsymbol{\theta}$ can be made from the product term. This term is a product of conditional densities, thus constitutes a *conditional likelihood*.

Now consider *partial likelihood* due to Cox et al., 1981, an idea based on only using part of (2.6), that is a factorization that consists of only on odd numbered conditional denstities (Kedem and Fokianos, 2005). consider a pair of jointly observed time series (X_t, Y_t) where X_t is a covarite series. Then similar to (2.6) the likelihood function is given by

$$f_{\boldsymbol{\theta}}(x_1, y_1, ..., x_N, y_N) = f_{\boldsymbol{\theta}}(x_1)\Big[\prod_{t=2}^{N} f_{\boldsymbol{\theta}}(x_t|d_t)\Big]\Big[\prod_{t=1}^{N} f_{\boldsymbol{\theta}}(y_t|c_t)\Big] \tag{2.8}$$

where $d_t = (y_1, x_1, ..., y_{t-1}, x_{t-1})$ and $c_t = (y_1, x_1, ..., y_{t-1}, x_{t-1}, x_t)$. Then inference concerning $\boldsymbol{\theta}$ can be made just from the second product term of (2.8). There is loss of information from the ignored factor, however this loss is small according to Kedem and Fokianos, 2005. This leads to a form of a likelihood function that does not require specification of the entire joint distribution of the data and of eventually the covariates. Below is a definition of such likelihood that has fixed number of parameters as a product of conditional densities .

Definition 2.2.1. Let $\mathcal{F}_t, t = 0, 1, ...$ be an increasing sequence of σ-fields, $\mathcal{F}_0$ $subset\mathcal{F}_1 \subset \mathcal{F}_2...$, and let $Y_1, Y_2, ...$ be a sequence of random variables on some common probability space such that Y_t is $\mathcal{F}_t$ measurable. The partial likelihood function relative to $\boldsymbol{\theta}$, $\mathcal{F}_t$, and the data $Y_1, ..., Y_N$ is given by the product

$$L(\boldsymbol{\theta}; y_1, ..., y_N) = \prod_{t=1}^{N} f_{\boldsymbol{\theta}}(Y_t|\mathcal{F}_{t-1}) \tag{2.9}$$

3. Linear Model

let $\{Y_t\}$ be a time series of count and consider the model given by

$$Y|\mathcal{F}_{t-1}^{Y,\lambda} \sim \text{Poisson}(\lambda_t) \qquad \text{where } \lambda_t = d + a\lambda_{t-1} + bY_{t-1}, \qquad t \geq 1 \tag{3.1}$$

The parameters d,a and b are assumed to be positive and satisfy $0 < a + b < 1$. $\mathcal{F}_{t-1}^{Y,\lambda}$ is the σ-field generated by $\{Y_s, s \leq t-1; \lambda_s, s \leq t\}$.

While Poisson distribution is the most prevalent in modeling counts, it is sometime unsuitable because of its property that the conditional mean and variance are equal, i.e $E(Y_t|\mathcal{F}_{t-1}^{Y,\lambda}) = \text{Var}(Y_t|\mathcal{F}_{t-1}^{Y,\lambda}) = \lambda_t$. When this is the case one may want to try other distributions than the Poisson, this is more easily achieved in the GLM framework for time series as introduced above. Bosowski, Ingle, and Manolakis, 2017 models US-monthly nuclear tests using the Negative binomial distribution.

In order to see some properties of the linear model, we can represent (3.1) as follows

$$\begin{aligned} Y_t &= \lambda_t + (Y_t - \lambda_t) \\ &= \lambda_t + \epsilon_t \\ &= d + a\lambda_{t-1} + bY_{t-1} + \epsilon_{t-1} \\ &= d + a(\epsilon_{t-1} - Y_{t-1}) + bY_{t-1} + \epsilon_{t-1} \\ &= d + (a+b)Y_{t-1} + (\epsilon_{t-1} - a\epsilon_{t-1}) \end{aligned}$$

Where ϵ_t is a white noise process with $\text{E}(\epsilon_t) = 0$ and $\text{cov}(\epsilon_t, \epsilon_{h-1})$ for all h. From which we obtain an ARMA type innovations representaion

$$Y_t - \mu = (a+b)Y_{t-1} + \epsilon_t - a\epsilon_{t-1} \tag{3.2}$$

Where

$$\mu = \text{E}(Y_t) = \frac{d}{1-(a+b)} \tag{3.3}$$

It has been shown that the Model (3.1) is stationary if $0 < a + b < 1$ (Fokianos et al., 2009 and references therein). It has also been shown that the autocovariance function for Y_t, with mean $E(Y_t) = E(\lambda_t) \equiv \mu = d/1 - (a+b)$, is given by

$$\text{Cov}[Y_t, Y_{t+h}] = \begin{cases} \frac{(1-(a+b)^2+b^2)\mu}{1-(a+b)^2}, & h = 0 \\ \frac{b(1-a(a+b))(a+b)^{h-1}\mu}{1-(a+b)^2}, & h = 1 \end{cases} \tag{3.4}$$

It is also important to point out as Tjøstheim, 2012 does, There is a conceptual distinction between model (3.1) and an AR model, while the later can be characterized by second order

properties, the former is essentially a probabilistic model where the autocovariance function is of less importance.

Following the classification method of Non-Gaussian time series by Cox et al., 1981, where one has a general exponential family formulation, such models can either be ***observation driven*** or ***parameter driven***. Let y_t be a time series and $Y^{(t-1)} = (Y_{t-1}, Y_{t-2}, ...)$. Suppose Θ_t is a parameter at time t. In an observation driven model

$$\Theta_t = \Theta(Y^{t-1}, \eta_t)$$

where η_t is a random innovation at time t. On the other hand in a parameter driven model

$$\Theta_t = \Theta(\Theta^{t-1}, \eta_t)$$

where η_t is a pure noise innovations process and Θ_t forms a markov process. After repeated substitution we can rewrite model (3.1)

$$\begin{aligned}
\lambda_t &= d + a\lambda_t + bY_{t-1} \\
&= d + ad + a^2\lambda_{t-2} + abY_{t-2} + bY_{t-1} \\
&= d + ad + a^2d + a^3\lambda_{t-3} + a^2bY_{t-3} + abY_{t-2} + bY_{t-1} \\
&\vdots \\
\lambda_t &= d\frac{1-a^t}{1-a} + a^t\lambda_0 + b\sum_{i=0}^{t-1} a^i \log(Y_{t-j-1})
\end{aligned}$$

We can see from the latter equation that the unobserved process λ_t can be expressed in terms of past values of the observed process Y_t. Therefore the linear model belongs to observation driven models by Cox et al., 1981 terminology.

The model (3.1) is rephrased slightly differently in Fokianos et al., 2009 to make it more convenient for the proofs of the results shown therein, such as, ergodicity of the process and asymptotic normality of maximum likelihood estimates of the model. In particular they consider the first part of (3.1) which is a sequence of independent Poisson drawings more explicitly in terms of random variables. To that end, for each time point t, it is introduced a Poisson process $N_t(\cdot)$ of unit intensity. Then the first part of (3.1) is expressed in terms of these Poisson processes by assuming that $Y_t = N_t(\lambda_t)$ is equal to the number of events in the time interval $[0, \lambda_t]$. Thus, for a sequence of independent random Poisson processes of unit intensity,$\{N_t(\cdot), t = 1, 2, ...\}$, (3.1) can be restated as

$$\mathrm{Y}_t = N_t(\lambda_t) \qquad \lambda_t = d + a\lambda_{t-1} + bY_{t-1} \tag{3.5}$$

for $t \geq 0$ and Y_0 and λ_0 fixed.

The model (3.1 or 3.5) fits in the GLM framework discussed in the previous section. In particular, the random component is the Poisson distribution and belong to the exponential family. The systematic component is given by $\eta_t = \mathbb{Z}_{t-1}^{\mathsf{T}}\boldsymbol{\beta}$ where $\mathbb{Z}_{t-1} = (1, \lambda_{t-1}, Y_{t-1})^{\mathsf{T}}$

and $\boldsymbol{\beta}$ is a vector of parameters. The functions that connects these to components is the identity link, in the next section, we discuss the model with the canonical link.

In order to prove geometric ergodicity of the bivariate process $\{(Y_t, \lambda_t)\}$ Fokianos et al., 2009 considers the perturbed version of the chain $\{Y_t^m, \lambda_t^m\}$ defined by

$$Y_t^m = N_t(\lambda_t^m), \qquad \lambda_t^m = d + a\lambda_{t-1}^m + bY_{t-1}^m + \epsilon_{t,m} \tag{3.6}$$

with λ_0^m,Y_0^m fixed and

$$\epsilon_{t,m} = C_m 1(Y_{t-1} = 1)U_t, \quad C_m \geq 0, \quad C_m \longrightarrow 0, \quad \text{as } m \longrightarrow \infty$$

where $1(\cdot)$ is an indicator function and where $\{\mathrm{U}_t\}$ is a sequence of i.i.d uniform random variables on $(0, 1)$ such that $\{\mathrm{U}_t\}$ is independent $\{\mathrm{N}_t(\cdot)\}$. Without going into any detail, (3.1) obtain the following result

1. Consider model (3.6) and suppose that $0 < a+b < 1$. Then the process $\{(Y_t^m, \lambda_t^m, U_t), t \geq 0\}$ is a $V_{(Y,U,\lambda)}$ -geometrically ergodic Markov chain with $V_{Y,U,\lambda}(Y, U, \lambda) = 1 + Y^k + \lambda^k + U^k$

2. If $0 < a + b < 1$, then as $m \longrightarrow \infty$such that $c_m \longrightarrow 0$, the perturbed model can be made arbitrarily close to the unperturbed model.

3.1. Partial Likelihood Inference and Estimation

let $\boldsymbol{\theta}$ be a three dimensional vector of parameters, $\boldsymbol{\theta} = (d, a, b)^T$ and the true values of the parameter be $\boldsymbol{\theta}_0 = (d_0, a_0, b_0)$. Then the conditional likelihood function for $\boldsymbol{\theta}$ given the time series data $Y_1, ..., Y_n$ is

$$L(\theta) = \prod_{t=1}^{n} \frac{\exp(-\lambda_t(\boldsymbol{\theta})_t^{Y_t}(\boldsymbol{\theta})}{Y_t!} \tag{3.7}$$

Where the intensity process $\lambda_t(\boldsymbol{\theta}) = d + a\lambda_{t-1}(\boldsymbol{\theta}) + bY_{t-1}$ is defined as in model(3.1). Hence the corresponding log-likelihood function is given by

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t(\boldsymbol{\theta}) = \sum_{t=1}^{n} (Y_t \log(\lambda_t(\boldsymbol{\theta})) - \lambda_t(\boldsymbol{\theta})) \tag{3.8}$$

We find the score function by differentiating the log-likelihood function, ans we obtain,

$$\mathbf{S}_n(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} (\frac{Y_t}{\lambda_t(\boldsymbol{\theta})} - 1) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{3.9}$$

Where the last term $\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}$ is differentiation with respect to each component of the parameter vector, given by

$$\frac{\partial \lambda_t}{\partial d} = 1 + a\frac{\partial \lambda_{t-1}}{\partial d}, \qquad \frac{\partial \lambda_t}{\partial a} = \lambda_{t-1} + a\frac{\partial \lambda_{t-1}}{\partial a}, \qquad \frac{\partial \lambda_t}{\partial b} = Y_{t-1} + a\frac{\partial \lambda_{t-1}}{\partial b} \tag{3.10}$$

The solution of to the equation $\boldsymbol{S}_n = 0$ gives the the parameter $\boldsymbol{\theta}$ denoted by $\hat{\boldsymbol{\theta}}$ that maximizes the conditional maximum likelihood, provided that it exists. we can also obtain the Hessian matrix by differentiating the log-likelihood function one more time

$$\begin{aligned}\mathbf{H}_n(\boldsymbol{\theta}) &= -\sum_{t=1}^{n} \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \\ &= \sum_{t=1}^{n} \frac{Y_t}{\lambda_t^2(\boldsymbol{\theta})} \Big(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)\Big(\frac{\partial \lambda_t(\theta)}{\partial \boldsymbol{\theta}}\Big)^T - \sum_{t=1}^{n} \Big(\frac{Y_t}{\lambda - t()\boldsymbol{\theta}} - 1\Big) \frac{\partial^2 \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\end{aligned} \tag{3.11}$$

Similarly the likelihood function for the perturbed model,L^m is given by

$$L^m = \prod_{t=1}^{n} \frac{exp(-\lambda_t^m(\theta))(\lambda_t^m(\theta))^{Y_t^m}}{Y_t^m!} \prod_{t=1}^{n} f_u(U_t) \tag{3.12}$$

where $f_u(\cdot)$ is the uniform density and λ_{t-1}^m is defined as in (3.6). Here, the product is due to the Independence assumption between $\{U_t\}$ and $N_{\lambda_{t-1}^m}$. The corresponding log likelihood function is given by,

$$l^m(\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t^m(\boldsymbol{\theta}) = \sum_{t=1}^{n} (Y_t^m \log \lambda_t^m(\boldsymbol{\theta}) - \lambda_t^m(\boldsymbol{\theta})) + \sum_{t=1}^{n} \log f_u(U_t) \tag{3.13}$$

Furthermore, the score function is almost identical to the score function of the unperturbed model given by the following

$$\boldsymbol{S}_n^m(\boldsymbol{\theta}) = \frac{\partial l^m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial l_t^m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \left(\frac{Y_y^m}{\lambda_t^m(\boldsymbol{\theta})} - 1\right) \frac{\partial \lambda_t^m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{3.14}$$

The last term of the latter equation is a vector consisting of the following components

$$\frac{\partial \lambda_t}{\partial d} = 1 + a\frac{\partial \lambda_{t-1}^m}{\partial a}, \quad \frac{\partial \lambda_t^m}{\partial a} = \lambda_{t-1}^m + a\frac{\partial \lambda_{t-1}^m}{\partial a}, \quad \frac{\partial \lambda_t^m}{\partial b} = Y_{t-1}^m + a\frac{\partial \lambda_{t-1}^m}{\partial b}.$$

The solution to the equation $\boldsymbol{S}_n^m = 0$, if it exists is denoted by $\hat{\boldsymbol{\theta}}^m$. By differentiating the log likelihood function once more, the Hessian matrix is obtained,

$$\begin{aligned}\boldsymbol{H}_n^m(\boldsymbol{\theta}) &= -\sum_{t=1}^{n} \frac{\partial^2 l_t^m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathsf{T}}} \\ &= \sum_{t=1}^{n} \frac{Y_t^m}{(\lambda_t^m(\boldsymbol{\theta}))^2} \left(\frac{\partial \lambda_t^m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \lambda_t^m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{\mathsf{T}} - \sum_{t=1}^{n} \left(\frac{Y_t^m}{\lambda_t^m(\boldsymbol{\theta})} - 1\right) \frac{\partial^2 \lambda_t^m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathsf{T}}}\end{aligned} \tag{3.15}$$

Asymptotic properties of the maximum likelihood estimator $\hat{\theta}$ of the linear model are obtained by Fokianos et al., 2009. To state the result obtained, define lower and upper bounds for the components of the parameter θ,

$$O(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} | 0 < \delta_L \leq d \leq \delta_U, 0 < \alpha_L \leq a \leq \alpha_U, \text{ and } 0 < \beta_L \leq b \leq \beta_U\}$$

Theorem 3.1.1. Consider model (3.1) and suppose that at the true value $\boldsymbol{\theta}_0$, $0 < \mathrm{a}_0 + \mathrm{b}_0 < 1$. Then, there exists a fixed open neighborhood $O = O(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$-see (3.1)-such that with probability tending to one, as $n \to \infty$, the log likelihood function (3.8) has a unique maximum point $\hat{\boldsymbol{\theta}}$. Furthermore, $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normal:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{G^{-1}}),$$

where the matrix $\boldsymbol{G}$ is defined by

$$\boldsymbol{G}(\boldsymbol{\theta}) = \mathrm{E}\left(\frac{1}{\lambda_t}\left(\frac{\partial \lambda_t}{\partial \theta}\right)\left(\frac{\partial \lambda_t}{\partial \theta}\right)^{\mathsf{T}}\right)$$

A consistent estimator of $\boldsymbol{G}$ is given by $\frac{\boldsymbol{G}}{n}$,where

$$\boldsymbol{G}(\boldsymbol{\theta}) = \sum_{t=1}^{n} \mathrm{var}[\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}] = \sum_{t=1}^{n} \frac{1}{\lambda_t}\left(\frac{\partial \lambda_t}{\partial \theta}\right)\left(\frac{\partial \lambda_t}{\partial \theta}\right)^{\mathsf{T}}$$

The proof the the above theorem and other details (such as, lemmas that show how the score function, the information matrix and the third derivative of the likelihood function of the perturbed model tend to their count part of the unperturbed model, and the respective proofs) are found in Fokianos et al., 2009.

4. Log-linear Model

It is argued in Fokianos and Tjøstheim, 2011 that the linear model discussed in the previous section has two main disadvantages. Firstly, since $\text{Cov}[Y_t, Y_{t+h}] > 0$ because $0 < a + b < 1$, the model is not suitable for modeling negatively correlated time series. The second is the difficulty of incorporating explanatory variables into the model. The linear model can only admit covariates that result in a positive regression term, Otherwise the poisson intensity process,λ_t, will be negative. However, count time series are often observed along side a covariate series, for instance the number of claims that occur of car crushes for an insurance company could be correlated with the number of rainy days in a month. Then it beneficial to be able to put all the information together in a model.

Moreover, the log linear model fits with the generalized linear model approach as described above, with the canonical link function. Let $\nu_t \equiv \log(\lambda_t)$, and $\mathcal{F}_{t-1}^{Y,\lambda}$ be the σ-field generated by $\{Y_0, ..., Y_t, \nu_0\}$. The model is given by

$$Y_t|\mathcal{F}_{t-1}^{Y,\lambda} \sim \text{Poisson}(\lambda_t), \qquad \nu_t = d + a\nu_{t-1} + b\log(Y_{t-1} + 1), \qquad t \geq 1 \tag{4.1}$$

Equivalently, by repeated substitution, we can write model(4.1) as

$$\begin{aligned}
\nu_t &= d + a\nu_{t-1} + b\log(Y_{t-1} + 1) \\
&= d + a(d + a\nu_{t-2} + b\log(Y_{t-2} + 1)) + b\log(Y_{t-1} + 1) \\
&= d(1 + a) + a^2\nu_{t-2} + ab\log(Y_{t-2} + 1) + b\log(Y_{t-1} + 1) \\
&= d(1 + a + a^2) + a^3\nu_{t-3} + b(a^2\log(Y_{t-3} + 1) + a\log(Y_{t-2} + 1) + \log(Y_{t-1} + 1)) \\
&\vdots \\
&= d\frac{1 - a^t}{1 - a} + a^t\nu_0 + b\sum_{i=0}^{t-1} a^i\log(Y_{t-i-1} + 1)
\end{aligned} \tag{4.2}$$

Similar to the intensity process of the linear model, ν_t can also be expressed by past value of the observed process Y_t, therefore the log linear model belongs to the class of observation driven models as well.

Following a similar approach as Fokianos et al., 2009, that is by first studying the ergodic properties of the perturbed model, Fokianos and Tjøstheim, 2011 show that the maximum likelihood estimator of the model is consistent and asymptotically normal.

The perturbed model is defined by

$$Y_t^m = N_t(\lambda_t^m) = N_t(\exp(\nu_t^m)), \qquad \nu_t^m = d + a\nu_{t-1}^m + b\log(Y_{t-1}^m + 1) + \epsilon_{t,m} \tag{4.3}$$

$$\epsilon_{t,m} = c_m 1(Y_{t-1}^m = 1)U_t, \qquad c_m > 0, c_m \to 0, \text{a s } m \to \infty$$

with ν_0^m, Y_0^m fixed, where $\{N_t(\cdot)\}$ is as defined in the previous section, $1(\cdot)$ is an indicator function and where $\{U_t\}$ is a sequence of i.i.d uniform random variables in $(0,1)$ which are indepent from $\{N_t(\cdot)\}$.

Proposition 4.0.1. Assume model (4.3) and suppose that $|a| < 1$. In addition, assume that when $b > 0$ then $|a + b| < 1$, and when $b < 0$ then $|a||a + b| < 1$. Then, the following conclusions hold:

1. The process $\{\nu_t^m, t \geq 0\}$ is a geometrically ergodic Markov chain with finite moments of order k, for an arbitrary k.

2. The process $\{(Y_t^m, U_t, \nu_t^m, t \geq 0)\}$ is a $V_{Y,U,\nu}$-geometrically ergodic Markov chain with $V_{Y,U,\lambda}(Y, U, \nu) = 1 + \log^{2k}(1 + Y) + \nu^{2k} + U^{2k}$, k being a positive integer.

4.1. Partial Likelihood Inference and Estimation

Once more suppose that $\boldsymbol{\theta}$ is a three dimensional vector of unknown parameters,that is, $\boldsymbol{\theta} = (d, a, b)^T$. let $\boldsymbol{\theta}_0$ be the true value of the unknown parameter. Then we define the conditional likelihood function for $\boldsymbol{\theta}$ for the log-linear model, where the intensity of the Poisson process is given by, $\nu_{\boldsymbol{\theta}} = d + a\nu_{t-1}(\boldsymbol{\theta}) + bY_{t-1}$, given the data $Y_{1,\ldots,Y_n}$

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{n} \frac{\exp(-e^{\nu_t(\boldsymbol{\theta})})e^{\nu(\boldsymbol{\theta})Y_t}}{Y_t!} \tag{4.4}$$

Furthermore, the log likelihood is defined by

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t(\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t(Y_t\nu_t(\boldsymbol{\theta}) - \exp(\nu_t(\boldsymbol{\theta}))) \tag{4.5}$$

Differentiating the log likelihood function with respect ot the parameter vector, we obtain the score function, given by

$$\mathbf{S}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} (Y_t - \exp(\nu_t(\boldsymbol{\theta})) \frac{\partial \nu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{4.6}$$

where $\frac{\partial \nu_t(\boldsymbol{\theta})}{\partial \theta \boldsymbol{\theta}}$ is differentiation of the log intensity process with respect to each component of the parameters vector, given by

$$\frac{\partial \lambda_t}{\partial d} = 1 + a\frac{\partial \lambda_{t-1}}{\partial d}, \qquad \frac{\partial \lambda_t}{\partial a} = \lambda_{t-1} + a\frac{\partial \lambda_{t-1}}{\partial a}, \qquad \frac{\partial \lambda_t}{\partial b} = \log(Y_{t-1} + 1) + a\frac{\partial \lambda_{t-1}}{\partial b} \tag{4.7}$$

The solution of to the equation $\boldsymbol{S}_n = 0$ gives the the parameter $\boldsymbol{\theta}$ denoted by $\hat{\boldsymbol{\theta}}$ that maximizes the conditional maximum likelihood, provided that it exists. we can also obtain

the Hessian matrix by differentiating the log-likelihood function one more time, such that

$$\begin{aligned}\mathbf{H}_n(\boldsymbol{\theta}) &= -\sum_{t=1}^{n} \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \\ &= \sum_{t=1}^{n} \exp(\nu_t(\boldsymbol{\theta}))\Big(\frac{\partial \nu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)\Big(\frac{\partial \nu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)^T - \sum_{t=1}^{n}(Y_t - \exp(\nu_t(\boldsymbol{\theta}))\frac{\partial^2 \nu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \end{aligned} \tag{4.8}$$

Similarly, the likelihood function for the perturbed log linear model is defined by (3.12), keeping in mined that $\nu_t \equiv \log(\lambda_t)$. The score function is then defined by

$$\boldsymbol{S}_n^m = \frac{\partial l^m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial l_t^m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n}(Y_t^m - \exp(\nu_t^m(\boldsymbol{\theta})))\frac{\partial \nu_t^m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{4.9}$$

The solution to the equation $\boldsymbol{S}_n^m = 0$ denoted by $\boldsymbol{\theta}^m$. The connection between between $\boldsymbol{\theta}^m$ and $\boldsymbol{\theta}$ is made through proposition **??**. The following result is obtained about the maximum likelihood estimator of (4.1), where the neighborhood $O(\boldsymbol{\theta_0})$ is as defined in the paragraph before 3.1.1.

Theorem 4.1.1. Consider model (4.1) and suppose that at the true value $\boldsymbol{\theta}_0$, $|a_0 + b_0| < 1$ if a_0 and b_0 have the same sign, and $a_0^2 + b_0^2 < 1$ if a_0 and b_0 have different sign. Then, there exists a fixed open neighborhood $O = O(\boldsymbol{\theta}_0)$ -such that with probability tending to 1, as $n \to \infty$, the log-likelihood function (4.5) has a unique maximum point $\boldsymbol{\theta}$. Furthermore, $\boldsymbol{\theta}$ is a consistent and asymptotically normal,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{G^{-1}}),$$

where the matrix $\boldsymbol{G}$ is defined by

$$\boldsymbol{G}(\boldsymbol{\theta}) = \mathrm{E}\left(\exp(\nu_t(\boldsymbol{\theta}))\left(\frac{\partial \nu_t}{\partial \theta}\right)\left(\frac{\partial \nu_t}{\partial \theta}\right)^{\mathsf{T}}\right)$$

A consistent estimator of $\boldsymbol{G}$ is given by $\frac{\boldsymbol{G}}{n}$,where

$$\boldsymbol{G}(\boldsymbol{\theta}) = \sum_{t=1}^{n} \mathrm{var}[\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}] = \sum_{t=1}^{n} \exp(\nu_t(\boldsymbol{\theta}))\left(\frac{\partial \nu_t}{\partial \theta}\right)\left(\frac{\partial \nu_t}{\partial \theta}\right)^{\mathsf{T}}$$

5. Introduction to Copulas

Copulas have wide range of application in statistics, finance, insurance and other areas. Intuitively copulas are functions that connect multivariate distributions to their marginal distributions. Alternatively copulas can be viewed as multivariate distribution functions whose one-dimensional margins are uniform in the interval (0,1). In Jaworski, Durante, Hardle, and Rychlik, 2010 copulas are characterized formally in the following definition and conditions.

Definition 5.0.1. A d-dimensional copula $C : \mathbf{I}^d \to \mathbf{d}$ is a function which is a cumulative distribution function with uniform marginals.

Any distribution function satisfies the following conditions and since a copula is defined to be a cumulative distribution function it satisfies these as well.

1. for every $j \in \{1, 2, ..., d\}$, $C(\mathbf{u}) = u_j$ when all the components of u are equal to 1 with the exception of the j th one that is equal to $u_j \in \mathbf{I}$;
2. C is isotonic, i.e. $C(\mathbf{u}) \leq C(\mathbf{v})$ for all $u, v \in \mathbf{I}^d, \mathbf{u} \leq \mathbf{v}$;
3. C is d-increasing.

The opposite is also true, any function that satisfies these properties is a copula.

We recall, the inverse transform method, a well known procedure for generating random variables of an arbitrary distribution function from the uniform distribution. Which is based on the following result. Let the *generalized inverse* of a distribution function F, be defind as

$$F^{\leftarrow}(x) := inf\{v : F(v) \geq x\}$$

Proposition 1. If $U \sim U[0, 1]$ and F_X is a CDF, then

$$P(F^{\leftarrow}(U) \leq x) = F_X(x)$$

conversely , if X has a continious CDF , then $F_X(X) \sim U[0, 1]$

Now suppose we have $X = (X_1, .., X_d)$ a multivariate random vector with CDF F_X. Then by the definition of copula and Proposition 1, the joint distribution of $F_{X_1(X_1)}, ..., F_{X_d}(X_d)$ is a copula. An expression for the copula can be found by noting that

$$\begin{aligned} C_X(u_1, ..., u_d) &= P(F_{X_1} \leq u_1, ..., F_{X_d} \leq u_d) \\ &= P(X_1 \leq F_{X_1}^{-1}(u_1), ..., X_d \leq F_{X_d}^{-1}(u_d)) \\ &= F_{\mathbf{X}}(F_{X_1}^{-1}(u_1), ..., F_{X_d}^{-1}(u_d)) \end{aligned}$$

If we as in Proposition 1 let $u_j = F_{X_j}(x)$, then

$$F_X(x_1, ..., x_d) = C(F_{X_1}(x_1), ..., F_{X_d}(x_d))$$

This is part of the following important theorem by Sklar(1959), which establishes a fundamental step in modeling dependence structures in multivariate distribution.

Theorem 5.0.1 (Sklar's theorem)**.** Let F be a d-dimensional distribution function. with univariate margins $F_1, F_2, ..., F_d$. Let A_j denote the range of $F_j, A_j := F_j(\overline{\mathbb{R}})j = 1, 2, ..., d$. Then there exists a copula C such that for all $(x_1, x_2, ..., x_d) \in \overline{\mathbb{R}}^d$,

$$F(x_1, x_2, ..., x_d) = C(F_1(x_1), F_2(x_2), ..., F_d(x_d)) \tag{5.1}$$

Such a C is uniquely determined on $A_1 \times A_2 \times \cdots \times A_d$ and, hence, it is unique when $F_1, F_2, ..., F_d$ are all continuous.

Hoeffding and Fréchet discovered independently that any copula lies between two extreme cases of dependence, lower and upper bounds.

Theorem 5.0.2 (Fréchet-Hoeffding bounds)**.** [section] Conside a copula $C(\boldsymbol{u}) = C(u_1, ..., u_d)$, then

$$\max\{\sum_{i=1}^{d} u_i + 1 - d, 0\} \leq C(\boldsymbol{u}) \leq \min\{u_1, ..., u_d\} \tag{5.2}$$

These upper and lower bounds are themselves copulas in the bivariate case. However whereas the *comonotonic copula*(upper bound) has a d-dimensional extension,there is no such extension of countermonotonicity(lowe bound) copula in the case of dimensions greater than two. The bounds still hold in a higher dimension whether or not the lower bound is also copula, T. Schmidt, 2007.

5.1. Some Special Copulas

5.1.1. Independence copula

A d-dimensional independence copula is given by

$$C(u_1, ..., u_d) = \prod_{i=1}^{d} u_i$$

Which means the copula function C is just a constant. An immediate consequence of Sklar's theorem is that, Any two or more random variables are independent if and only if the copula that relates them to each other is an independence copula,i.e a constant. If a copula is sufficiently differentiable, its density function can be computed by

$$c(u) := \frac{\partial^d C(U_1, ..., U_d)}{\partial u_1 \cdots \partial u_d}$$

5.1.2. Gaussian copula

As the name suggests a gaussian copula is a copula derived from the gaussian distribution. According to T. Schmidt, 2007 for elliptical distributions, such as multivariatete normal distribution, the dependence structure is fully described by the correlation between the random variables. Thus for non-elliptical distribution other measures of dependence may be required. Thus if we evaluate (5.1) at $u_j = F_i^{\leftarrow}(x_i)$ we obtain

$$c(u_1, ..., u_d) = F(F_1^{\leftarrow}(u_1), ..., F_d^{\leftarrow}(u_d)) \tag{5.3}$$

Then for $X \sim N_d(0, \sigma)$, where Σ is the correlation matrix, The corresponding copula is the Gaussian copula given by

$$C_\Sigma^G = \Phi_\Sigma(\Phi^{-1}(u_1), ..., \Phi_d^{-1}(u_d)) \tag{5.4}$$

where Φ denotes the cummulative distribution function of the standard normal distribution. The Gaussian copula can be used to model the three fundamental dependence structures by varying copula parameter. If $\rho = 0$ the Gaussian copula is the dependence copula, monotonicity if ρ, and countermonotonicity if $\rho = -1$.

5.1.3. Clayton copula

In contrary to the Gaussian copula and other parametric copulas that are derived from a certain distribution, the Clayton copula is stated directly by;

$$C_\theta^{Clyton} = (max\{u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1, 0\})^{-\frac{1}{\theta}} \tag{5.5}$$

It can be shown that the Clyaton copula can be used to model different forms of dependence, from both extremes, monotonicity in the limit as $\theta \to \infty$, independence as $\theta \to 0$, and comonotonicity if $\theta = -1$.

Figure 5.1 shows the density functions of the copulas that are mentioned in this section.

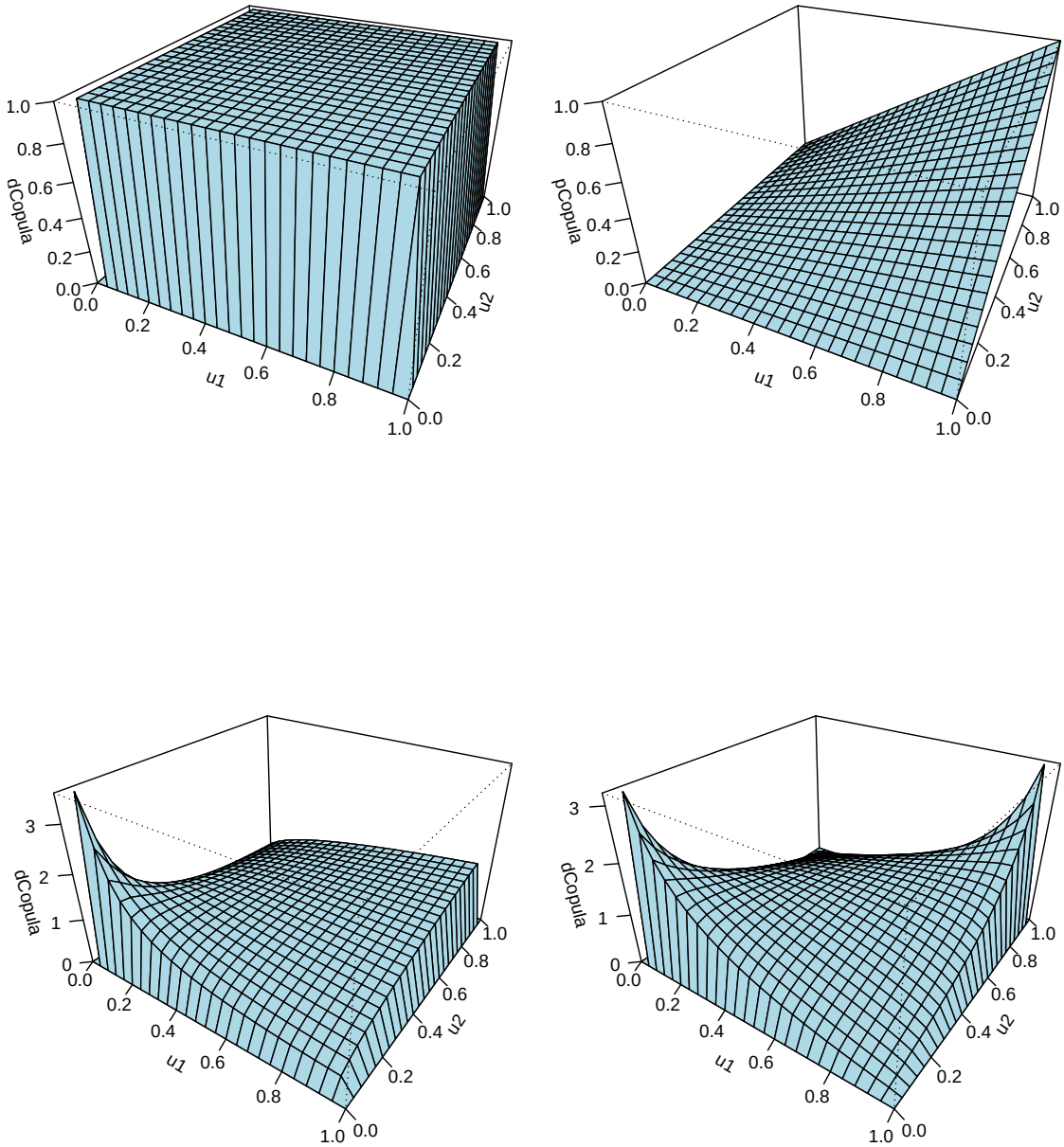

Figure 5.1.: Top, left and right show PDF and CDF og independent copula respectively, Bottom, left and right show PDF's of Gaussian(with correlation coefficient $\rho = 0.5$) and Clayton(with parameter $\theta = 0.5$) copulas respectively

6. Template Model Builder

In the preceding sections, we will study the two models discussed in the previous sections and their multivariate extensions by simulation. The simulation of the processes and estimation of parameters is done in the free and open source R (Team et al., 2013). Particularly we will use the template model builder, TMB (Kristensen, Nielsen, Berg, Skaug, and Bell, 2015) package. The package allows for quick implementation of complex non-linear random effects models. Although, Here we are only working with observation driven models.

TMB calculates derivatives of a differentiable objective function by *automatic differentiation*. This eliminates the need for writing and maintaining derivative codes. AD should not be confused with symbolic differentiation and numerical differentiation. AD decomposes expressions into elementary operations, on which differentiation rules from calculus can be applied, then the derivatives are merged together using the chain rule. TMB does not implement AD from first principles rather it uses the general purpose AD package CppAD (Bell, 2012) available from C++. For good example with AD applied to linear regression model and general discussion around AD see, sections (2.1) and (2.2) of Fournier et al., 2012.

In TMB the user defines the objective function as a c++ template, for our purpose the negative quasi likelihood function given by (8.5), denoted by $\log(L^*(\theta))$. Then TMB provides the exact gradient and Hessian matrix which are calculated by automatic differentiation. All other operations are performed in R, such as reading data and optimizing the objective function, possibly by gradient-based optimization algorithm. By reporting the parameters (using *ADREPORT()* in the C++ template) the parameter estimates and their corresponding standard deviations are obtained. For non-random effects models TMB reports the standard deviations of θ or any differentiable function of $\phi(\theta)$, that are calculated by the standard delta-method.

$$V(\phi(\hat{\theta})) = -\nabla_\theta \phi(\theta)[\mathbb{H}(\theta)]^{-1}\nabla_\theta \phi(\theta)^\top|_{\theta=\hat{\theta}} \tag{6.1}$$

where $\mathbb{H}(\theta) = \nabla^2_\theta \log L^*(\theta)$ is the Hessian matrix. The above covariance matrix can be reprted from a TMB object, It can also be calculated by inverting the Hessian matrix with respect to the MLE $\hat{\theta}$. For examples of C++ templates that are used in this thesis see A

7. Local Gaussian Correlation

Here we distinguish between local dependence measures and global dependence measure. The latter one describes the correlation between two random variables in terms of a scalar value. Pearson correlation is a widely used example of such a measure, which gives a single scalar value ρ meant to characterize the entire joint distribution of a two dimensional random variable $X = (X_1, X_2)$. On the other hand, the local Gaussian correlation is an example of a local measure, where, in principle, the dependence is characterized in terms of correlation at each point, i.e by $\rho_{X_1,X_2}(x_1.x_2)$. Here we briefly describe what local Gaussian correlation is, a detailed account is found in Tjøstheim and Hufthammer, 2013.

The idea is to approximate the density $f_{X_1,X_2}(x_1, x_2)$ locally in a neighbourhood of each point $x = (x_1, x_2)$ by a Gaussian bivariate density of the

$$\psi(v, \mu(x), \Sigma(x)) = \frac{1}{2\pi|\Sigma|^{1/2}}\exp\Big[-\frac{1}{2}(v-\mu(x)^\top\Sigma^{-1}(v-\mu(x))\Big] \tag{7.1}$$

where $v = (v_1, v_2)^\top$ is a running variable, $\mu(x) = (\mu_1(x), \mu_2(x))^\top$ is the local mean vector and $\Sigma(x) = (\sigma_{ij}(x))$ is the local cavariance matrix. With $\sigma_i^2 = \sigma_{ii}(x)$, the local Gaussian correlation at the point x is defined by $\rho(x) = \frac{\sigma_{12}(x)}{\sigma_1(x)\sigma_2(x)}$. Then the above equation (7.1) becomes

$$\begin{aligned}\psi(v, \mu_1(x), \mu_2(x),&\sigma_1^2, \sigma_2^2, \rho(x)) = \\ &\frac{1}{2\pi\sigma_1(x)\sigma_2(x)\sqrt{1-\rho^2(x)}}\exp\Big\{-\frac{1}{2(1-\rho^2(x))}\times\Big[\Big(\frac{v_1-\mu_1(x)}{\sigma_1(x)}\Big)^2 \\ &-2\rho(x)\Big(\frac{v_1-\mu_1(x)}{\sigma_1(x)}\Big(\frac{v_2-\mu_2(x)}{\sigma_2(x)}+\Big(\frac{\nu_2-\mu_2(x)}{\sigma_2(x)}\Big)^2\Big]\Big\}\end{aligned} \tag{7.2}$$

However (7.2) is not very well defined unless some conditions are imposed. The Gaussian approximation needed is the one that approximates $f(x)$ in the neighborhood x such that (7.2) holds at x. It is shown in Tjøstheim and Hufthammer, 2013 for a fixed bandwidth parameter b, the local population parameter $\theta(x) = (\mu_1(x), \mu_2(x), \sigma_1^2(x), \sigma_2^2(x), \rho(x))$ can be obtained by minimizing the penalty function

$$q = \int K_b(v-x)[\psi(v, \theta(x)) - \log\psi(v, \theta(x))f(v)]dv \tag{7.3}$$

The minimizer $\theta_b(x)$ should then satisfy the integral equation

$$\int K_b(v-x)\frac{\partial}{\partial\theta_j}\log\psi(v, \theta(x))[f(v)-\psi(v, \theta(x))]dv = 0, \qquad j = 1, ..., 5 \tag{7.4}$$

Furthermore, Tjøstheim and Hufthammer, 2013 show that if a unique population vector $\theta_b(x)$ exists then, for a shrinking bandwidth parameter, i.e letting $b \to 0$, population vector θ_x can be estimated by using a local log-likelihood function defined by

$$L(X_1, ..., X_n) = n^{-1} \sum_i K_b(X_i - x) \log \psi(X_i, \theta_b(x)) - \int K_b(v - x)\psi(v, \theta_b(x))dv \qquad (7.5)$$

for a given observations $X_1, ..., X_n$. They also show that $\theta_{n,b}(x) \to \theta_b(x)$ for b-fixed, and $\theta_{n,b} \to \theta(x)$ almost surely as $b = b_n$ tending to zero.

Local Gaussian correlation have been applied to study financial contagion , that is, whether the cross-market linkages in financial markets increase after a shock to a country, see Støve, Tjøstheim, and Hufthammer, 2014. Here, In section 10, we use local Gaussian correlation to capture the nonlinear dependence between marginals of a bivariate count time series. We use *localgauss* (Geir Drage Berentsen, Kleppe, and Tjøstheim, 2014) an R package in order to calculate local Gaussian correlation between the marginal time series.

8. Multivariate Autoregressive Models

In this section we look at multivariate count time series, and the multivariate extensions of the linear and log-linear models discussed in previous sections. The main reference for this section is Fokianos et al., 2020, which in many ways builds up on Fokianos et al., 2009. A challenging aspect of modelling multivariate count time series is the specification of the joint distribution. A natural idea is to generalize the univariate Poisson probability mass fucntion (PMF), However the resulting distribution of the multivariate Poisson discrete random vector is of a complicated functional form. Therefore, maximum likelihood inference can be difficult both theoretically and numerically. This is circumvented by Fokianos et al., 2020 using a copula based construction of the multivariate distribution. The copula function is imposed on a vector of continuous random variables, and the resulting multivariate count time series retains Poisson properties marginally.

8.1. Multivariate Linear Model

In order to extend the univariate linear model to the corresponding multivariate linear model , suppose $\{\boldsymbol{Y}_t = Y_{i,t}, i = 1, 2, ..., p, t = 1, 2, ...\}$ denotes a p-dimensional count time series. Let the p-dimensional intensity process corresponding to each univariate marginal of $\boldsymbol{Y}$ be denoted by $\{\boldsymbol{\lambda}_t = \lambda_{i,t}, i = 1, 2, ..., p, t = 1, 2, ...\}$. Let also $\mathcal{F}_t^{\boldsymbol{Y},\boldsymbol{\lambda}}$ be the σ-field generated by $\{\boldsymbol{Y}_0, ..., \boldsymbol{Y}_t, \boldsymbol{\lambda}_0\}$, then the multivariate linear model can be stated as

$$Y_{i,t}|\mathcal{F}_{t-1}^{\boldsymbol{Y},\boldsymbol{\lambda}} \quad \text{is marignally Poisson}(\lambda_{i,t}), \qquad \boldsymbol{\lambda}_t = \boldsymbol{b} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}Y_{t-1} \tag{8.1}$$

where $\boldsymbol{d}$-is a p-dimensional parameter vector, and $\boldsymbol{A}$, and $\boldsymbol{B}$ are $p \times p$ unknown parameter matrices. As in the case of the univariate linear model all components of $\boldsymbol{d}, \boldsymbol{A}$, and $\boldsymbol{B}$ are assumed to be positive, to ensure that $\lambda_{i,t} \geq 0$.

Looking at (8.1), it suggests that each $Y_{i,t}$ is a Poisson process, however $\{\boldsymbol{Y}_t\}$ itself is not necessarily distributed with multivariate Poisson distribution. Fokianos et al., 2020 suggest the following data generating scheme, see Algorithm [1], that keeps the Poisson properties marginally, assuming there are some starting values $\boldsymbol{\lambda}_0 = \lambda_{1,0}, ..., \lambda_{p,0}$ available.

Algorithm 1: Algorithm for generating marginally Poisson distributed multivariate random variables

//Initialize
$\boldsymbol{Y} \leftarrow$ matrix(N, p)
$\boldsymbol{\lambda} \leftarrow$ matrix(N, p)
// since we have the starting values, insert to the matrix
$\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda}_0$
$\boldsymbol{d} \leftarrow$ vector(p))
$\boldsymbol{A} \leftarrow$ matrix(p,p)
$\boldsymbol{B} \leftarrow$ matrix(p,p)
$\boldsymbol{U} \leftarrow$ Generate a sample from a p-dimensional copula, matrix$[k, p]$
$\boldsymbol{X} \leftarrow$ Combine columns : $(-\text{logU}[, 1]/\boldsymbol{\lambda}(1, t), ..., -\text{logU}(, 1)/\boldsymbol{\lambda}(p, t))$
$\boldsymbol{Y}(, 0) \leftarrow \max_{1 \leq k \leq K}\{\sum_{l=1}^{k} X_{i,l} \leq 1\}, i = 1, 2, ..., p$
for $t \leftarrow 1$ *to* N **do**
  1) $\boldsymbol{\lambda}[, t] \leftarrow \boldsymbol{d} + A * \boldsymbol{\lambda}(, t-1) + \boldsymbol{B} * \boldsymbol{Y}(, t-1)$
  2) $\boldsymbol{U} \leftarrow$ Generate a sample from a p-dimensional copula, matrix(k, p)
  $\boldsymbol{X} \leftarrow$ Combine columns : $(-\text{logU}[, 1]/\boldsymbol{\lambda}(1, t), ..., -\text{logU}(, 1)/\boldsymbol{\lambda}(p, t))$
  3) $\boldsymbol{Y}[, t] \leftarrow \max_{1 \leq k \leq K}\{\sum_{l=1}^{k} X_{i,l} \leq 1\}, i = 1, 2, ..., p$
end

The above algorithm utilizes a fundamental property of the Poisson process at step 2, that is the inter-arrival times of events in the Poisson process are exponentially distributed. Thus the marginal distribution of say $X_{i,l} = -\log(U_{i,l})/\lambda_{i,0}, l = 1, 2, ..., K$ is exponential with parameter $\lambda_{i,0}$. It is from these exponential inter-arrival times the Poisson processes are obtained. It is also on these variables the dependence structure is imposed, more specifically, the copula is imposed on the continuous uniform random variables that are used to generate the exponential variables. Therefore it is important to note that the instantaneous correlation between vectors of counts is not equal to the correlation induced by the copula imposed on the waiting times.

we can then rephrase (8.1) as it was done for the univariate linear model

$$\mathbf{Y}_t = \mathbf{N}_t(\boldsymbol{\lambda}_t), \qquad \boldsymbol{\lambda_t} = \mathbf{d} + \mathbf{A}\boldsymbol{\lambda}_{t-1} + \mathbf{B}\mathbf{Y}_{t-1} \tag{8.2}$$

where $\{\mathbf{N}_t\}$ is a sequence of independent p-variate copula-Poisson process which counts the number of events in $[0, \lambda_{1,t}] \times ... \times [0, \lambda_{p,t}]$.

$$\mathbf{Y}_t^m = \mathbf{N}_t(\boldsymbol{\lambda}_t^m), \qquad \boldsymbol{\lambda}_t^m = \mathbf{d} + \mathbf{A}\boldsymbol{\lambda}_{t-1}^m + \mathbf{B}\mathbf{Y}_{t-1}^m + \boldsymbol{\epsilon}_t^m \tag{8.3}$$

In the following we denote by $\|\mathbf{x}\|_d = (\sum_{i=1}^{p} |x_i|^d)^{1/d}$ the l^d-norm of a p-dimensional vector $\mathbf{x}$. For a $q \times p$ matrix $\mathbf{A} = a_{ij}, i = 1, ..., q, j = 1, ..., p$, we let $|||\mathbf{A}|||_d$ denote the generalized matrix norm $= |||\mathbf{A}|||_d = \max_{\|\mathbf{x}\|_d=1} \|\mathbf{A}\mathbf{x}\|_d$. If $d = 1$, then $|||A|||_1 = \max_{1 \leq j \leq p} \sum_{i=1}^{q} |a_{ij}|$, and when $d = 2$, $|||A|||_2 = \rho^{1/2}(\mathbf{A}^T\mathbf{A})$ where $\rho(\cdot)$ denotes the spectral radius. Employing the perturbation approach they obtain the following results

1. Consider model (8.3) and suppose that $|||\mathbf{A}+\mathbf{B}|||_2 < 1$. Then the process $\{\lambda_t, t>0\}$ is geometrically ergodic Markov chain with finite r'the moments, for any $r>0$. where $\boldsymbol{\epsilon}_t^m c_m \mathbf{V}_t$. Here, the sequence c_m is strictly positive and tends to zero, as $m \to \infty$, and $\mathbf{V}_t$ is a p-dimensional vector, where each component are positive random variables and have a bounded support of the form $[0, M]$ for $M>0$.

2. If $|||\mathbf{A}+\mathbf{B}||| < 1$, then $c_m \to 0$ as $m \to \infty$, then the difference between the perturbed model (8.3) and the unperturbed model (8.1) can be made arbitrarily small.

8.1.1. Quasi Likelihood Inference

Let $\{\mathbf{Y}_t\}$ be a given multivariate time series of counts, and denote a vector of unknown parameters by $\boldsymbol{\theta}$ such that $\boldsymbol{\theta}^\top = (\mathbf{d}, \text{vec}^\top(\mathbf{A}), \text{vec}^\top(\mathbf{B}))$, where vec$(\cdot)$ denotes the vectorization of a matrix,i.e a linear transformation which converts the matrix into a column vector. Then consider the following conditional qusi-likelihood function, given $\boldsymbol{\lambda}_0$, for the parameter vector $\boldsymbol{\theta}$,

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{n}\prod_{i=1}^{p}\Big[\frac{\exp(-\lambda_{i,t}(\boldsymbol{\theta}))\lambda_{i,t}^{y_{i,t}}(\boldsymbol{\theta})}{y_{i,t}!}\Big] \tag{8.4}$$

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{n}\sum_{i=1}^{p}(y_{i,t}\log\lambda_{i,t}(\boldsymbol{\theta}) - \lambda_{i,t}(\boldsymbol{\theta})) \tag{8.5}$$

and the score function is given by

$$S_n(\boldsymbol{\theta}) = \sum_{t=1}^{n}\sum_{i=1}^{p}\Big(\frac{y_{i,t}}{\lambda_{i,t}(\boldsymbol{\theta})}-1\Big)\frac{\partial\lambda_{i,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}) = \sum_{t=1}^{n}\frac{\partial\boldsymbol{\lambda}_t^\top(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{D}_t^{-1}(\boldsymbol{\theta})(\mathbf{Y}_t-\boldsymbol{\lambda}_t(\boldsymbol{\theta})) \equiv \sum_{t=1}^{n}s_t(\boldsymbol{\theta}) \tag{8.6}$$

The differentiation term in the score function leads to these recursions which are similar to the ones we have seen in the uni variate case.

$$\begin{aligned}
\frac{\partial\boldsymbol{\lambda}_t}{\partial\mathbf{d}^\top} &= \mathbf{I}_p + \mathbf{A}\frac{\partial\boldsymbol{\lambda}_{t-1}}{\partial\mathbf{d}^\top} \\
\frac{\partial\boldsymbol{\lambda}_t}{\partial\text{vec}^\top(\mathbf{A})} &= (\boldsymbol{\lambda}_{t-1}\otimes\mathbf{I}_p)^\top + \mathbf{A}\frac{\partial\boldsymbol{\lambda}_{t-1}}{\partial\text{vec}^\top(\mathbf{A})} \\
\frac{\partial\boldsymbol{\lambda}_t}{\partial\text{vec}^\top(\mathbf{B})} &= (\boldsymbol{\lambda}_{t-1}\otimes\mathbf{I}_p)^\top + \mathbf{A}\frac{\partial\boldsymbol{\lambda}_{t-1}}{\partial\text{vec}^\top(\mathbf{B})}
\end{aligned}$$

where $\otimes$ denotes the Kronecker's product. By further differentiation we obtain the Hessian matrix, which is given by

$$\mathbf{H}_n = \sum_{t=1}^{n}\sum_{i=1}^{p}\frac{y_{i,t}}{\lambda_{i,t}^2(\boldsymbol{\theta})}\frac{\partial\lambda_{i,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\lambda_{i,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} - \sum_{t=1}^{n}\sum_{i=1}^{p}\Big(\frac{y_{i,t}}{\lambda_{i,t}(\boldsymbol{\theta})}-1\Big)\frac{\partial^2\lambda_{i,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \tag{8.7}$$

Toward obtaining asymptotic results Fokianos et al., 2020 follow Fokianos et al., 2009. They use the sufficient conditions for ergodicity and stationarity for the perturbed model and then

show that the unperturbed and perturbed versions are "close"(see Lemma 3.1,Fokianos et al., 2020). Furthermore, they show that the score function, the Hessian information matrix and the third derivative of the log likelihood function, of the perturbed model tend to their corresponding counterpart of the unperturbed model. Thereby, obtaining the following result concerning the quasi maximum likelihood estimator $\hat{\boldsymbol{\theta}}$.

Theorem 8.1.1. Consider model (8.2). Let $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathcal{R}^d$. Suppose that $\boldsymbol{\Theta}$ is compact and assume that the true value $\boldsymbol{\theta}_0$ belongs to the interior of $\boldsymbol{\Theta}$. Suppose that at the true value $\boldsymbol{\theta}_0$ the condition $|||\mathbf{A} + \mathbf{B}||| < 1$ holds true. Then there exists a fixed open neighborhood, say $O(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0||_2 < \delta\}$,of $\boldsymbol{\theta}_0$ such that with probability tending 1 as $n \to \infty$, the equation $S_n(\boldsymbol{\theta}) = 0$ has a unique solution, say $\boldsymbol{\theta}$. Furthermore, $\hat{\boldsymbol{\theta}}$ is strongly consistent and asymptotically normal,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}) \tag{8.8}$$

where the matrices $\mathbf{G}(\boldsymbol{\theta})$ and $\mathbf{H}(\boldsymbol{\theta})$ are defined by

$$\mathbf{G}(\boldsymbol{\theta}) = \mathrm{E}[\frac{\partial \boldsymbol{\lambda}_t^\top(\boldsymbol{\theta})}{\partial \theta}\mathbf{D}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{D}_t^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}] \tag{8.9}$$

$$\mathbf{H}(\boldsymbol{\theta}) = \mathrm{E}[\frac{\partial \boldsymbol{\lambda}_t^\top(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{D}_t^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\lambda}_t(\theta)}{\partial \boldsymbol{\theta}^\top}] \tag{8.10}$$

and expectation is taken with respect to the stationary distribution of $\{\mathbf{Y}_t\}$

we investigate the above result by simulation where the gradient is calculated by automatic differentiating.

8.1.2. Simulation study for the linear model

Here we consider the model (8.2) with $p = 2$. The bivariate count series $(\mathbf{Y}_t, \boldsymbol{\lambda}_t)$ is generated according to Algorithm[1]. We generate 500 and 1000 realizations of the process by using the Clayton copula, which are simulated 1000 times. To obtain initial values for the maximization algorithm, first we fit each marginal series to an ARMA(1,1) (see 3.2) model and fit a univariate linear model. Secondly, using $(\hat{d}, \hat{a}, \hat{b})$ we predict the hidden intensity process. Finally, we fit a multivariate regression model, regressing the response (the bivariate count series) against lagged values of the predicted hidden process and the lagged values of the response itself.

Model parameter estimates averaged over the 1000 simulations is reported below in table (8.1). the table reports results, both where we have used automatic differentiation to obtain the score function and where the score function is calculated analytically . The true parameters used to generate the data are the following

$$\mathbf{A} = \begin{pmatrix} 0.3 & 0.05 \\ 0.1 & 0.25 \end{pmatrix} \quad , \mathbf{B} = \begin{pmatrix} 0.5 & 0.05 \\ 0.1 & 0.4 \end{pmatrix} \quad \text{and } \mathbf{d} = (0.5, 1) \tag{8.11}$$

Note that the above parameter satisfy the ergodicity conditions, i.e $|||\mathbf{A} + \mathbf{B}||| = 0.89$. The averaged parameter estimates are reasonably close to the true parameters for both values

Table 8.1.: Compare Simulation results for the multivariate linear model (8.1), where parameters are estimated in R and parameters estimated in TMB. Here n denotes the sample size of the simulations,and ϕ denotes the parameter of the Clayton copula used to generate marginally Poisson counts.

					R						
n	ϕ	$\hat{d}_1$	$\hat{d}_2$	$\hat{a}_{11}$	$\hat{a}_{22}$	$\hat{b}_{11}$	$\hat{b}_{22}$	$\hat{a}_{12}$	$\hat{a}_{21}$	$\hat{b}_{12}$	$\hat{b}_{21}$
500	0	0.871	1.421	0.289	0.222	0.493	0.396	0.087	0.167	0.051	0.098
		(0.205)	(0.349)	(0.071)	(0.084)	(0.049)	(0.050)	(0.082)	(0.077)	(0.045)	(0.049)
	0.5	0.772	1.116	0.279	0.200	0.494	0.395	0.083	0.161	0.051	0.099
		(0.170)	(0.264)	(0.074)	(0.087)	(0.051)	(0.085)	(0.081)	(0.050)	(0.052)	(0.0494)
					TMB						
500	0	0.784	1.406	0.288	0.158	0.504	0.407	-0.005	0.116	0.065	0.104
		(0.283)	(0.406)	(0.088)	(0.115)	(0.047)	(0.047)	(0.081)	(0.091)	(0.04)	(0.049)
	0.5	0.68	1.15	0.289	0.209	0.502	0.400	0.023	0.112	0.055	0.105
		(0.221)	(0.279)	(0.092)	(0.115)	(0.047)	(0.050)	(0.089)	(0.101)	(0.044)	(0.050)
					R						
1000	0	0.803	1.316	0.295	0.222	0.498	0.400	0.083	0.166	0.052	0.099
		(0.134)	(0.236)	(0.052)	(0.057)	(0.036)	(0.032)	(0.054)	(0.054)	(0.030)	(0.036)
	0.5	0.733	1.056	0.286	0.207	0.497	0.396	0.082	0.157	0.048	0.100
		(0.118)	(0.181)	(0.055)	(0.061)	(0.037)	(0.037)	(0.057)	(0.054)	(00.035)	(0.037)
					TMB						
1000	0	0.725	1.32	0.290	0.178	0.507	0.408	0.006	0.105	0.060	0.105
		(0.204)	(0.286)	(0.063)	(0.082)	(0.034)	(0.033)	(0.059)	(0.065)	(0.031)	0.034
	0.5	0.632	1.12	0.294	0.215	0.501	0.404	0.027	0.111	0.053	0.100
		(0.152)	(0.210)	(0.062)	(0.074)	(0.034)	(0.033)	(0.063)	(0.07)	(0.031)	(0.036)

of the copula parameter. Furthermore, we find from Figure (8.1) support for the result that the quasi maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$ is asymptotically normal. We also see that the parameter estimate, $\hat{\mathbf{d}}$ approaches normality unsatisfactorily. However, see Figure (8.2) with increasing sample sizes it tends closer to normality, which supports, that $\hat{\boldsymbol{\theta}}$ is also consistent. Similar situation is observed for $\hat{\mathbf{d}}$ also for different copula parameters.

8.2. Multivariate Log-linear Model

The multivariate analogue of the univariate Log-linear model is defined by

$$Y_{i,t}|\mathcal{F}_{t-1}^{\boldsymbol{Y},\boldsymbol{\lambda}} \quad \text{is marignally Poisson}(\lambda_{i,t}), \qquad \boldsymbol{\nu}_t = \boldsymbol{b} + \boldsymbol{A}\boldsymbol{\nu}_{t-1} + \boldsymbol{B}(\log(Y_{t-1}+1) \tag{8.12}$$

Where, as oppose to the linear model, we do not need to impose assumption passivity on elements of the unknown parameters. As previously mentioned in the case the univariate log-linear model, the model (8.12) is also expected to be a better model for count data than its linear counterpart. That is due to its ability to incorporate covaraite time series, and is suitable when negative correlation is observed. Including covariates into the model Fokianos et al., 2020 propose the following model, here the sigma field generated is expanded to incluse $\{x_t\}$. As in the case of the multivariate linear model it is useful to rephrase 8.12 as follows

$$\mathbf{Y}_t = \mathbf{N}_t(\boldsymbol{\nu}_t), \qquad \boldsymbol{\nu}_t = \mathbf{d} + \mathbf{A}\boldsymbol{\nu}_{t-1} + \mathbf{B}\log(\mathbf{Y}_{t-1} + \mathbf{1}_p) \tag{8.13}$$

where the process $\{\mathbf{N}_t\}$ is defined as before a sequence of independent p-variate copula-Poisson process which counts the number of events in $[0, \exp(\nu_{1,t})] \times \cdots \times [0, \exp(\nu_{p,t})]$, since $\boldsymbol{\nu}_t \equiv \log\boldsymbol{\lambda}_t$. And the perturbed version of the model is given by

$$\mathbf{Y}_t^m = \mathbf{N}_t(\boldsymbol{\nu}_t^m), \qquad \boldsymbol{\nu}_t = \mathbf{d} + \mathbf{A}\boldsymbol{\nu}_{t-1}^m + \mathbf{B}\log(\mathbf{Y}_{t-1}^m + \mathbf{1}_p) + \boldsymbol{\epsilon}_t^m \tag{8.14}$$

The log-likelihood function for the log-linear model is

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{n}\sum_{i=1}^{p}(y_{i,t}\nu_{i,t}(\boldsymbol{\theta}) - \exp(\nu_{i,t}(\boldsymbol{\theta}))) \tag{8.15}$$

we denote by $\hat{\boldsymbol{\theta}}$ the quasi maximum likelihood estimator that maximizes $l(\boldsymbol{\theta})$. The score function is by differentiating the log-likelihood function given by

$$S_n(\boldsymbol{\theta}) = \sum_{t=1}^{n}\sum_{i=1}^{p}(y_{i,t} - \exp(\nu_{i,t}(\boldsymbol{\theta})))\frac{\partial\nu_{i,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \sum_{t=1}^{n}\frac{\partial\nu_t^\top(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}(\mathbf{Y}_t - \exp(\nu_t(\boldsymbol{\theta}))) \tag{8.16}$$

the term $\partial\boldsymbol{\nu_t}(\boldsymbol{\theta})/\partial\boldsymbol{\nu}^\top$ leads to the following recursions used to calculate the QMLE

$$\begin{aligned}
\frac{\partial\boldsymbol{\nu}_t}{\partial\mathbf{d}^\top} &= \mathbf{I}_p + \mathbf{A}\frac{\partial\boldsymbol{\nu}_{t-1}}{\partial\mathbf{d}^\top} \\
\frac{\partial\boldsymbol{\nu}_t}{\partial\text{vec}^\top(\mathbf{A})} &= (\boldsymbol{\nu}_{t-1}\otimes\mathbf{I}_p)^\top + \mathbf{A}\frac{\partial\boldsymbol{\nu}_{t-1}}{\partial\text{vec}^\top(\mathbf{A})} \\
\frac{\partial\boldsymbol{\nu}_t}{\partial\text{vec}^\top(\mathbf{B})} &= (\log(\mathbf{Y}_{t-1}+\mathbf{1}_p)\otimes\mathbf{I}_p)^\top + \mathbf{A}\frac{\partial\boldsymbol{\lambda}_{t-1}}{\partial\text{vec}^\top(\mathbf{B})}
\end{aligned}$$

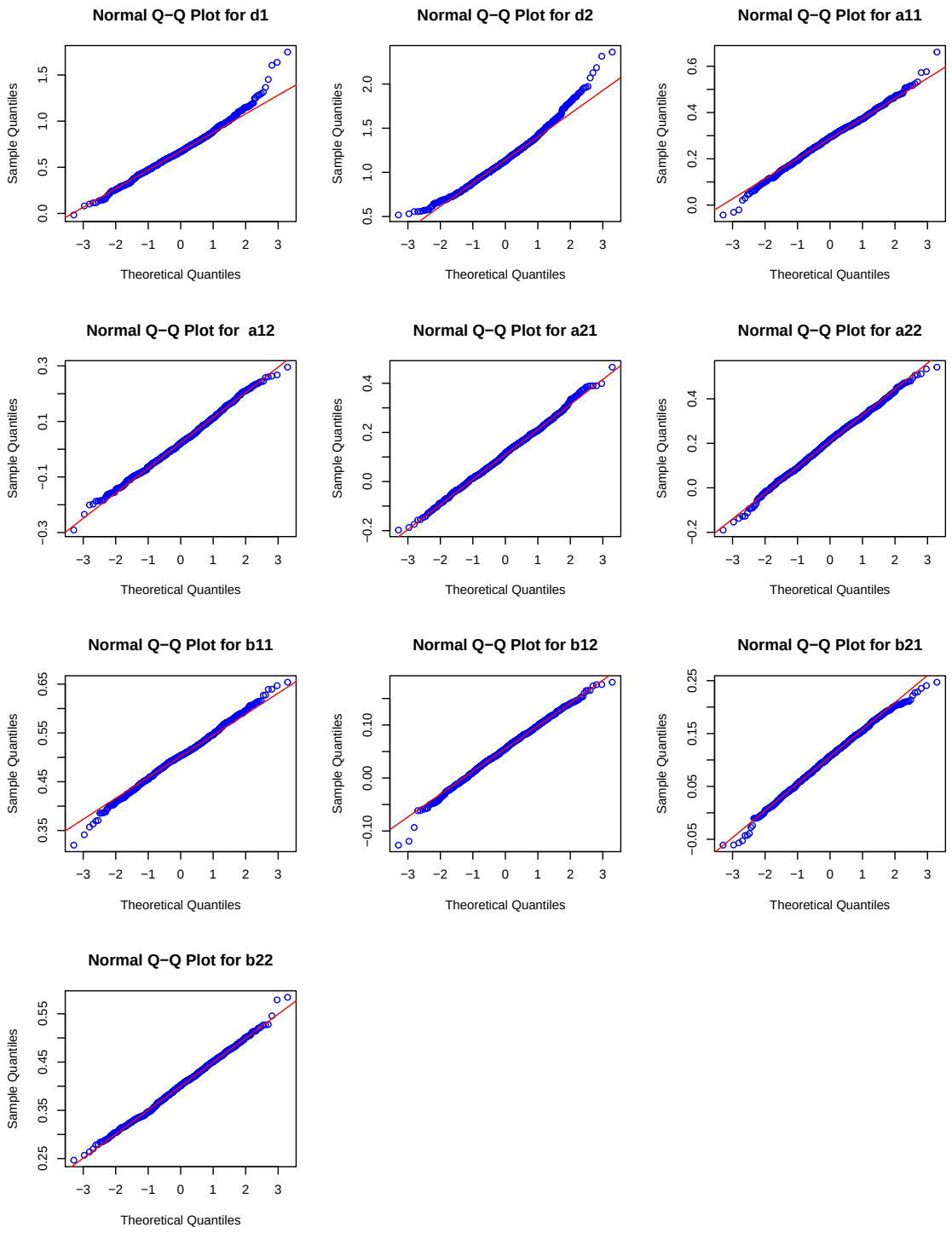

Figure 8.1.: QQ-plots of the standardized sampling distribution of $\hat{\boldsymbol{\theta}}$ for the multivariate linear model with true parameter values given by (8.11) with a sample size of 500. Data have been generated by a Clayton copula with $\phi = 0.5$. Result is based on 1000 simulations.

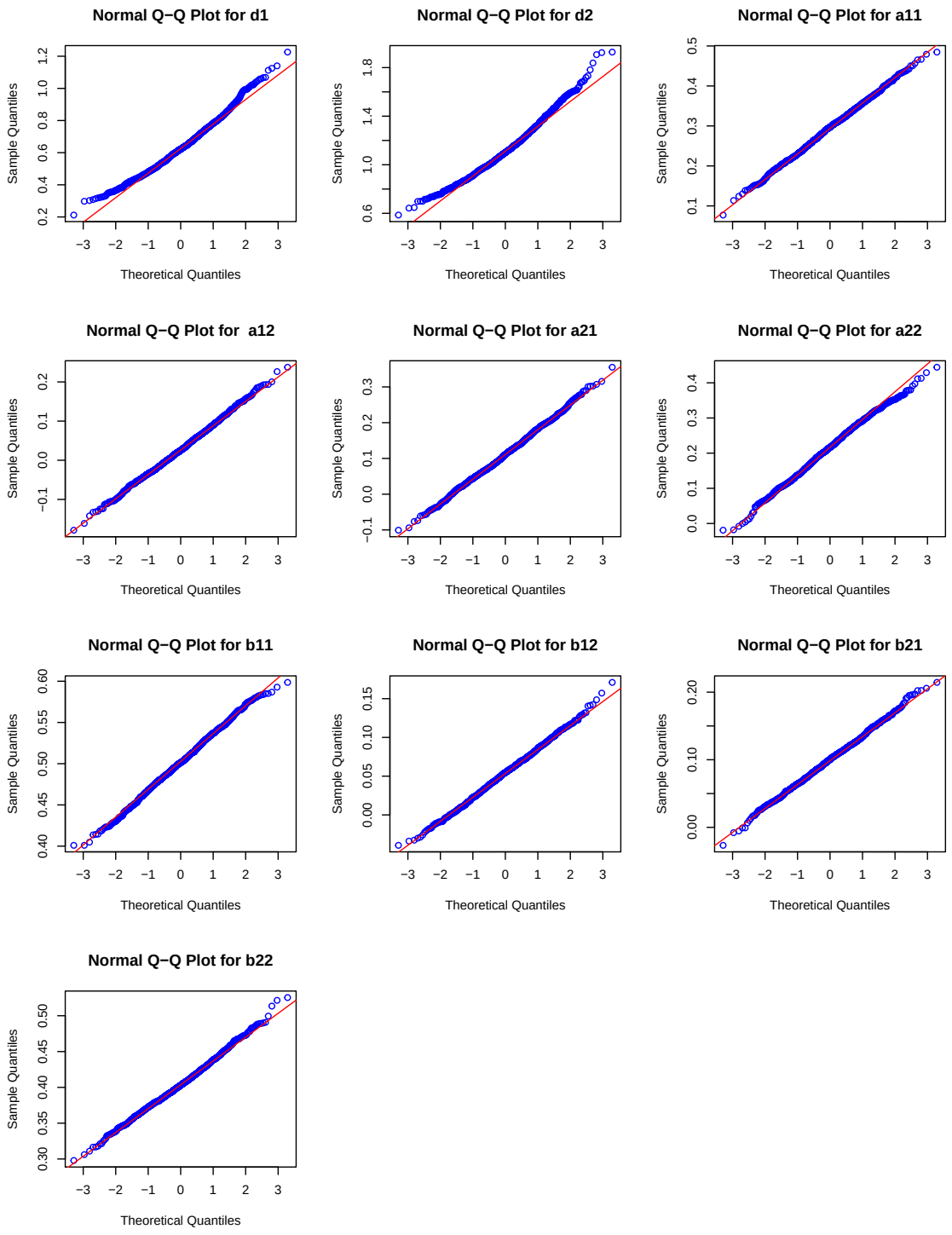

Figure 8.2.: QQ-plots of the standardized sampling distribution of $\hat{\boldsymbol{\theta}}$ for the multivariate linear model with true parameter values given by 8.11 with a sample size of 1000. Data have been generated by a Clayton copula with $\phi = 0.5$. Result is based on 1000 simulations.

and differentiating the log-likelihood once more the Hessian matrix is obtained

$$\mathbf{H}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n}\sum_{i=1}^{p}\exp(\nu_{i,t}(\boldsymbol{\theta}))\frac{\partial\nu_{i,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\nu_{i,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} - \sum_{t=1}^{n}\sum_{i=1}^{p}(y_{i,t}-\exp(\nu_{i,t}(\boldsymbol{\theta})))\frac{\partial^2\nu_{i,t}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \quad (8.17)$$

Analogous to the linear model, similar results about the properties of the log-linear model are also obtained. The approach taken to show these results is similar to that of the linear model. In summary the following results hold true for the log-linear model

1. Consider (8.14) and suppose that $|||A|||_2 + |||B|||_2 < 1$. Then the process $\{\boldsymbol{\nu}_t^m, t > 0\}$ is geometrically ergodic Markov chain with finite rth moments, for any $r > 0$

2. Consider the perturbed model (8.14) and the unperturbed model (8.13). If $|||A|||_2 + |||B|||_2 < 1$ the difference between (8.13) and (8.14) can be made arbitrarily close as $c_m \to 0$ as $m \to \infty$.

3. Consider model (8.13). Under the conditions of Theorem 8.1.1, the equation $S_n(\boldsymbol{\theta}) = 0$, where $S_n(\cdot)$ is defined by (8.16), has a unique solution, say $\hat{\boldsymbol{\theta}}$. And that $\hat{\boldsymbol{\theta}}$ is strongly consistent and asymptotically normal as in Theorem 8.1.1, where the matrices $\mathbf{G}(\boldsymbol{\theta})$ and $\mathbf{H}(\boldsymbol{\theta})$ are defined by

$$\mathbf{G}(\boldsymbol{\theta}) = \mathrm{E}\Big[\frac{\partial\boldsymbol{\nu}_t^\top(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\frac{\partial\boldsymbol{\nu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\Big], \quad \mathbf{H}(\boldsymbol{\theta}) = \mathrm{E}\Big[\frac{\partial\boldsymbol{\nu}_t^\top(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{D}_t(\boldsymbol{\theta})\frac{\partial\boldsymbol{\nu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\Big] \quad (8.18)$$

   and expectation is taken with respect to the stationary distribution of $\{\mathbf{Y}_t\}$.

8.2.1. Simulation study for the log-linear model

A simulation study for the log-linear model proceeds the same way as the case of the linear model. The aim is to obtain quasi maximum likelihood estimator of the model parameters and to investigate the asymptotic properties of the $\hat{\boldsymbol{\theta}}$. Towards that end, we generate data by Algorithm [1], making the necessary changes so that the the data generated follows (8.12). First, we look at the case where the parameters $\mathbf{A}$ and $\mathbf{B}$ are non-diagonal matrices. We obtain starting values for maximization of the log-likelihood function (8.15) as follows;

- Fit univariate log linear model including covariate time series to each series. We fit a Poisson regression to each series on the lagged log response, i.e $\log(Y_{t-1}+1)$ and the covariate series. Hence, obtain estimates of (d,b), say $(\tilde{d},\tilde{b})$, then initiate optimization algorithm by taking $(\tilde{d},0,\tilde{b})$ as the initial value for the unknown parameter vector.

- Calculate the predictions

$$\hat{\nu}_{1,t}(\boldsymbol{\theta}) = \nu_{1,t}(\hat{\boldsymbol{\theta}}) = \hat{d}_1 + \hat{a}_1\nu_{1,t-1}(\hat{\boldsymbol{\theta}}) + \hat{b}_1\log(Y_{1,t-1}+1)$$
$$\hat{\nu}_{2,t}(\boldsymbol{\theta}) = \nu_{2,t}(\hat{\boldsymbol{\theta}}) = \hat{d}_2 + \hat{a}_2\nu_{2,t-1}(\hat{\boldsymbol{\theta}}) + \hat{b}_2\log(Y_{2,t-1}+1)+$$

- Then run a vector generalized linear model regressing the response against its lagged log response, i.e $\log(\mathbf{Y}_{t-1}+1)$, the predicted intensity process, i.e $\hat{\boldsymbol{\nu}}_t(\boldsymbol{\theta}) = (\nu_{1,t}(\hat{\boldsymbol{\theta}}), \nu_{2,t}(\hat{\boldsymbol{\theta}}))$ and the covariate time series.

The estimated coefficient matrices from the later regression are used as starting values for the optimization.

The true parameter values used to generate the process are

$$\mathbf{A} = \begin{pmatrix} 0.3 & 0.05 \\ 0.1 & 0.25 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0.5 & 0.05 \\ 0.1 & 0.4 \end{pmatrix}, \quad \mathbf{d} = (0.3, 0.5) \tag{8.19}$$

Table (8.2) shows the result the simulation where we have used the Gaussian copula with $\phi = 0$ and $\phi = 0.5$. We can see that the results are reasonably close to the true parameters and improving for larger sample sizes. Furthermore, figure (8.3) and figure (8.4) support the asymptotic normality of the estimators, where once again $\hat{\mathbf{d}}$ deviates more from normality relative to the other parameter estimates.

8.2.2. Simulation study for the Log-linear model with covariates

As mentioned previously one of the arguments in favour of the log-linear model versus the linear model is its ability to include time dependent covariates with more ease. More specifically, let $\{X_t\}$ be some covariate time series, then a log-linear model that includes covariate time series is given by

$$Y_t|\mathcal{F}_{t-1}^{Y,X,\nu} \sim \text{Poisson}(\lambda_t), \qquad \nu_t = d + a\nu_{t-1} + b\log(Y_{t-1}+1) + cX_t \tag{8.20}$$

where c is a a real valued parameter and $\mathcal{F}_{t-1}^{Y,X,\nu}$ is the sigma field generated by $\sigma = (Y_s, X_s, \nu_0, s \leq t)$. It should be noted that a model such as the above one can only be cast within the framework developed by Fokianos et al., 2009 if $cX_t > 0$. However Agosto et al., 2016 looks at the following more general model that includes multiple lags and covariates,

$$\lambda_t = d + \sum_{i=1}^{p} a_i Y_{t-i} + \sum_{i=1}^{q} b_i \lambda_{t-1} + f(X_{t-1}, \gamma) \tag{8.21}$$

where X enters the intensity process thorough the link function $f(\cdot, \gamma)$. The link function allows for a function such that negative covariates can be included. In addition, f is required to be Lipschitz so as to exclude some functions such as the exponential one. They also provide conditions for stationarity and asymptotic results for the maximum likelihood estimators.

Here we look at the bivariate extension of the log-linear model that includes covariates. The model is given by

$$Y_{i,t}|\mathcal{F}_{t-1}^{\mathbf{Y},\mathbf{X},,\boldsymbol{\lambda}} \quad \text{is marginally Poisson}(\lambda_{i,t}), \quad \boldsymbol{\nu}_t = \mathbf{A}\boldsymbol{\nu}_{t-1} + \mathbf{B}\log(\mathbf{Y}_{t-1} + \mathbf{1}_p) + \mathbf{C}\mathbf{X}_t \tag{8.22}$$

Table 8.2.: Compare Simulation results for the log linear model (8.12) defined by (8.19), where parameters are estimated in R and parameters estimated in TMB. Here n denotes the sample size of each simulation,and ϕ denotes the parameter of the Gaussian copula used to generate the data.

R											
n	ϕ	$\hat{d}_1$	$\hat{d}_2$	$\hat{a}_{11}$	$\hat{a}_{22}$	$\hat{b}_{11}$	$\hat{b}_{22}$	$\hat{a}_{12}$	$\hat{a}_{21}$	$\hat{b}_{12}$	$\hat{b}_{21}$
500	0	0.319	0.391	0.304	0.220	0.491	0.401	0.164	0.039	0.100	0.059
		(0.199)	(0.198)	(0.076)	(0.079)	(0.049)	(0.046)	(0.069)	(0.070)	(0.044)	(0.046)
	0.5	0.294	0.491	0.290	0.192	0.492	0.405	0.159	0.059	0.098	0.060
		(0.162)	(0.157)	(0.075)	(0.084)	(0.048)	(0.052)	(0.071)	(0.082)	(0.049)	(0.050)
TMB											
500	0	0.279	0.457	0.162	0.168	0.573	0.478	0.050	0.050	0.117	0.166
		(0.232)	(0.230)	(0.110)	(0.100)	(0.060)	(0.057)	(0.107)	(0.091)	(0.060)	(0.051)
	0.5	0.319	0.550	0.163	0.159	0.557	0.465	0.060	0.057	0.109	0.152
		(0.185)	(0.185)	(0.108)	(0.103)	(0.056)	(0.059)	(0.114)	(0.094)	(0.059)	(0.054)
R											
1000	0	0.332	0.399	0.315	0.220	0.498	0.404	0.157	0.015	0.101	0.063
		(0.116)	(0.136)	(0.055)	(0.059)	(0.034)	(0.032)	(0.044)	(0.035)	(0.029)	(0.031)
	0.5	0.314	0.492	0.308	0.188	0.491	0.405	0.165	0.037	0.096	0.059
		(0.105)	(0.104)	(0.053)	(0.055)	(0.037)	(0.035)	(0.054)	(0.050)	(00.035)	(0.035)
TMB											
1000	0	0.246	0.439	0.205	0.2098	0.551	0.449	0.068	0.072	0.089	0.136
		(0.158)	(0.161)	(0.078)	(0.076)	(0.042)	(0.040)	(0.079)	(0.063)	(0.043)	0.037
	0.5	0.296	0.529	0.203	0.189	0.541	0.444	0.073	0.081	0.081	0.125
		(0.125)	(0.128)	(0.074)	(0.072)	(0.042)	(0.039)	(0.077)	(0.063)	(0.040)	(0.038)

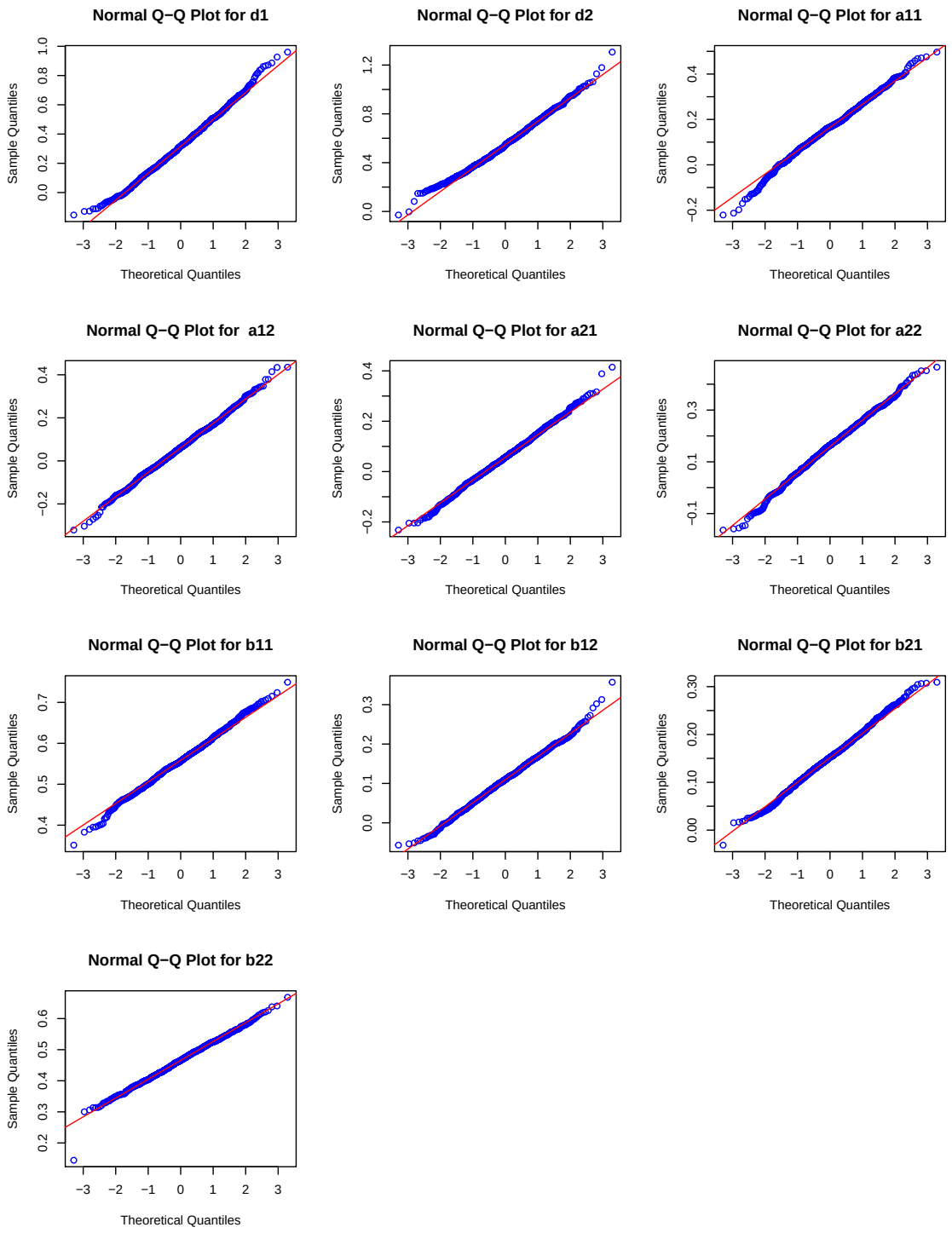

Figure 8.3.: QQ-plots of the standardized sampling distribution of $\hat{\boldsymbol{\theta}}$ for the multivariate log-linear model with true parameter values given by (8.19) with a sample size of 500. Data have been generated by a Gaussian copula with $\phi = 0.5$. The result is based on 1000 simulations.

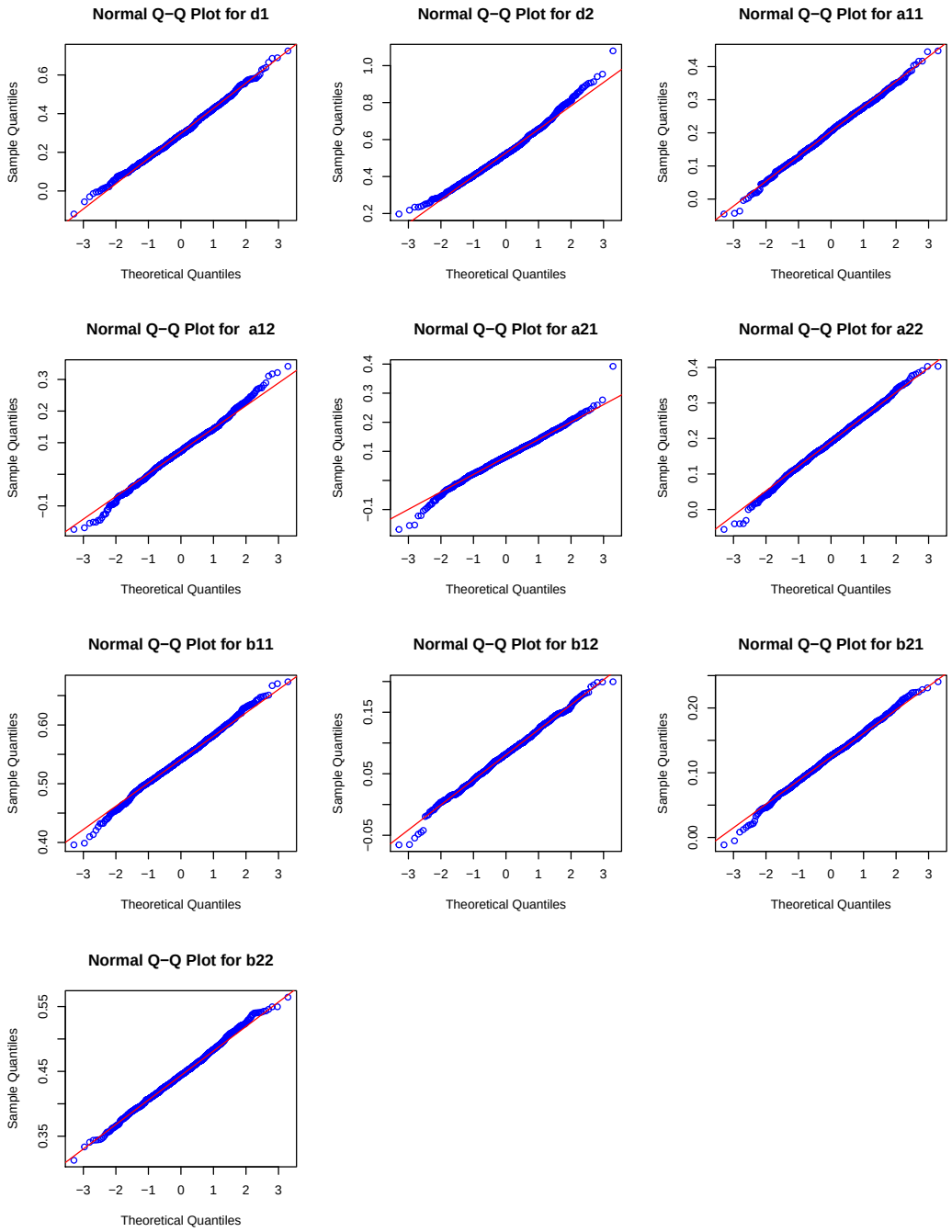

Figure 8.4.: QQ-plots of the standardized sampling distribution of $\hat{\boldsymbol{\theta}}$ for the multivariate log-linear model with true parameter values given by (8.19) with a sample size of 1000. Data have been generated by a Gaussian copula with $\phi = 0.5$. The result is based on 1000 simulations.

where the covariate time series $\{X_t\}$ is taken to be a two dimesional ccc-log-garch(1,1), see Sucarrat, Grønneberg, and Escribano, 2016, which here is allowed to be negative. Simulation from this model is performed in the same way as for the the models in the previous section. The bivariate count time series that depends on on $\{X_t\}$ is generated by Algorithm 1, that is, given $\{X_t\}$ and making appropriate changes at step 1 of the algorithm. Each of generated bivariate count time series depend on the same $\{\mathbf{X}_t\}$ covariate time series.

The true parameter values used to generate the process are

$$\mathbf{A} = \begin{pmatrix} 0.3 & 0.05 \\ 0.1 & 0.25 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0.5 & 0.05 \\ 0.1 & 0.4 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 0.2 & 0.05 \\ 0.1 & 0.2 \end{pmatrix} \quad \mathbf{d} = (0.3, 0.5) \tag{8.23}$$

The resulting estimates displayed in Table 8.3 are based on 500 runs and the reported estimates are obtained by averaging the result for all the simulations. Each of the 500 runs depend on the same covariate time seris $\{X_t\}$, and the standard errors correspond to the sampling standard errors of the estimates obtained by the simulation. Parameter estimates are close to the true parameter used to generete the process. Furthermore figure 8.5 appears to suggest that the MLE $\hat{\boldsymbol{\theta}}$ is asymptotically normal. Halliday and Boshnakov, 2018 study a similar multivariate Poisson autoregressive model with covariates, that is said to inherit similar properties as the univariate case studied by Agosto et al., 2016. Similar to (8.22), the model they look at is also copula based. However they impose the copula structure on the marginal time series, such that the resulting multivariate distribution is jointly Poisson. While here, copula structure is imposed on the waiting times, and the resulting joint distribution isn't necessarily Poisson. They obtain conditions for the process to be stationary and ergodic, but parameters are not estimated by method of inference function and not by maximum likelihood.

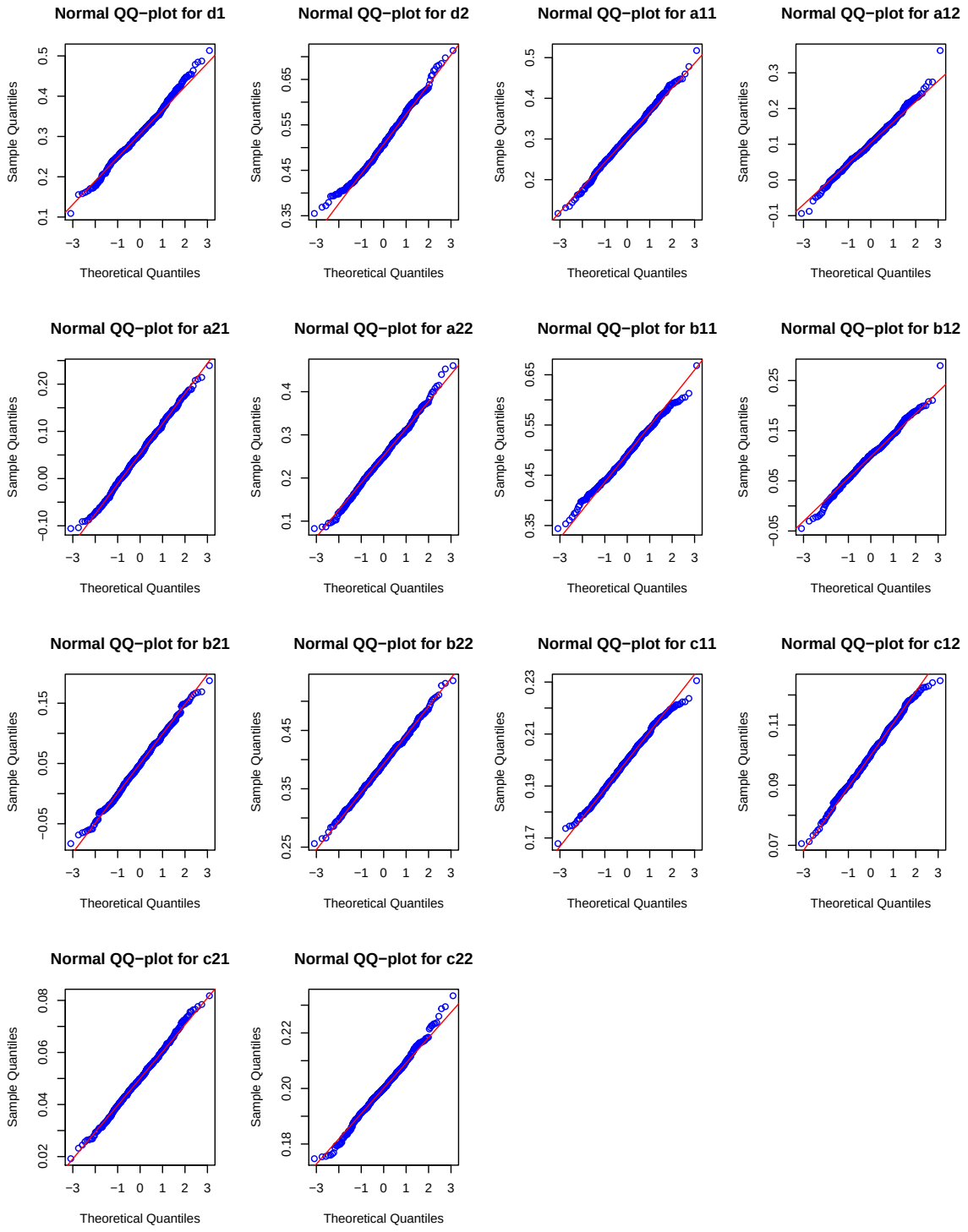

Figure 8.5.: QQ-plots of the standardized sampling distribution of $\hat{\boldsymbol{\theta}}$ for the multivariate log-linear model with true parameter values given by (8.23) with a sample size of 500. Data have been generated by a Gaussian copula with $\phi = 0.5$. The result is based on 500 simulations.

Table 8.3.: Parameter estimates of the log-linear model including covariates, with the corresponding standard errors of the estimators. The sample size is 500 and the data is generated by Gaussian copula with copula parameter $\phi = 0.5$

parameter	true	estimated	$\widehat{\text{SE}}$
d_1	0.3	0.308	0.062
d_2	0.5	0.510	0.063
a_{11}	0.3	0.304	0.064
a_{12}	0.05	0.052	0.061
a_{21}	0.1	0.104	0.063
a_{22}	0.25	0.250	0.062
b_{11}	0.5	0.491	0.050
b_{12}	0.05	0.048	0.046
b_{21}	0.1	0.099	0.048
b_{22}	0.4	0.393	0.047
c_{11}	0.2	0.199	0.010
c_{12}	0.05	0.050	0.010
c_{21}	0.1	0.100	0.009
c_{22}	0.2	0.200	0.010

9. Application to Real Data

9.1. Application to counts of Stocks

In this section we look at application of the linear and log-linear models to real world data. Particularly we fit the models to a bivariate count time series which consists of number of transactions in 15 seconds of Coca-Cola Company (KO) and IBM stocks on September 19th 2005. Both of these are heavily traded stocks. It is therefore interesting to see how stocks from different sectors influence each other. The data has 1440 observations, registered from 09:30 to 16:30, and excluding observation if the first 15 minutes and the last 15 minutes of transaction. The reason for the exclusion is that there are usually more transaction at the opening and closing of a given trading day. This can then influence measurements of auto-and cross correlations. Table 9.1 below provides descriptive statistics for the data. Figure (9.2) shows the autocorrelation and cross-autocorrelation functions.

Table 9.1.: summary statistics

	min	max	1st Qu.	median	3rd Qu.	mean	variance
IBM	0	39	2	4	7	4.854	13.78
Coca-Cola	0	22	2	4	6	4.276	10.70

Which reveal that there is high correlation among observations of each transaction series and across the series. Furthermore, Figure (9.1) shows the time series plot of the data with the corresponding predictions from each model, the linear and the log-linear model. The predictions are calculated by $\hat{Y}_{i,t} = \lambda_{i,t}(\hat{\boldsymbol{\theta}})$ for $i = 1$ and 2. We compare the models using root mean square error (RMSE) of predictions, we find the RMSE for the linear model to be 180.01 and 183.25 for the log-linear model. This indicates that the linear model provides a better fit to the data than the log-linear model.

Table (9.2) shows parameter estimates and their corresponding standard errors. Here we can see that our estimates where we have used TMB, where the score function is calculated by automatic differentiation are close. However the RMSE that resulted from our estimates are slightly lower than those obtained by Fokianos et al., 2020 which are 190.06 and 193.25 for the linear and log-linear model respectively. This difference is likely due to the fact TMB calculates the exact gradient which allows for improved speed and accuracy of maximization of the quasi-likelihood function by means of gradient-based optimization methods. Here, we have opted for an unconstrained $BFGS$, an algorithm that belongs to the class of quasi-Newton methods. It is available in R from *optim* or *optimr*. The standard errors in Fokianos et al., 2020 are calculated by the robust estimator of the covariance matrix given

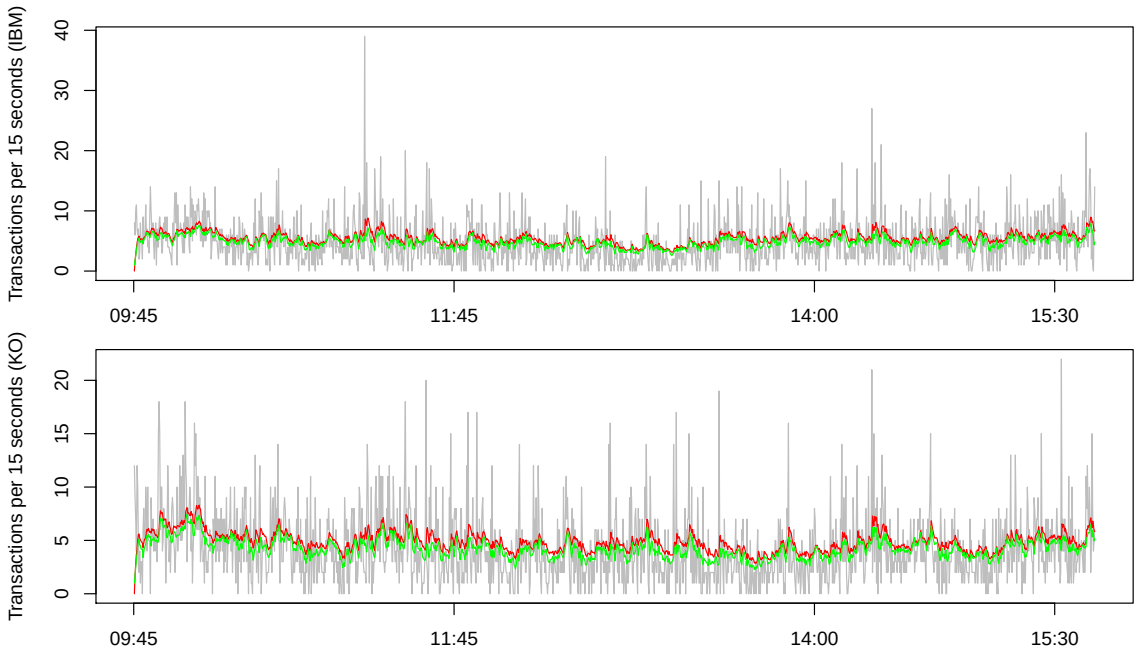

Figure 9.1.: Number of transactions per 15 seconds for IBM (top) and Coca-Cola (bottom), along side the predicted number of transaction by the linear model (red lines) and log-linear model (yellow lines)

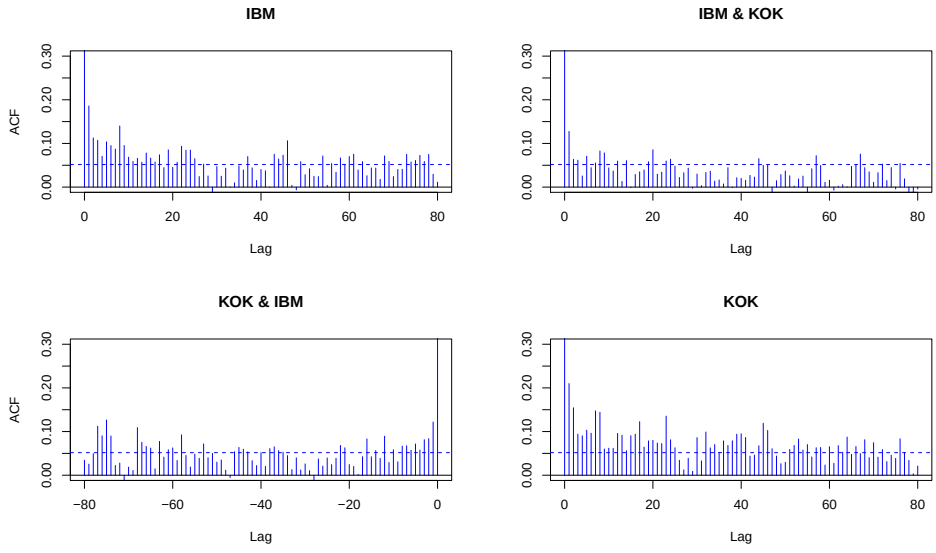

Figure 9.2.: Auto-and cross correlation function of the transaction data

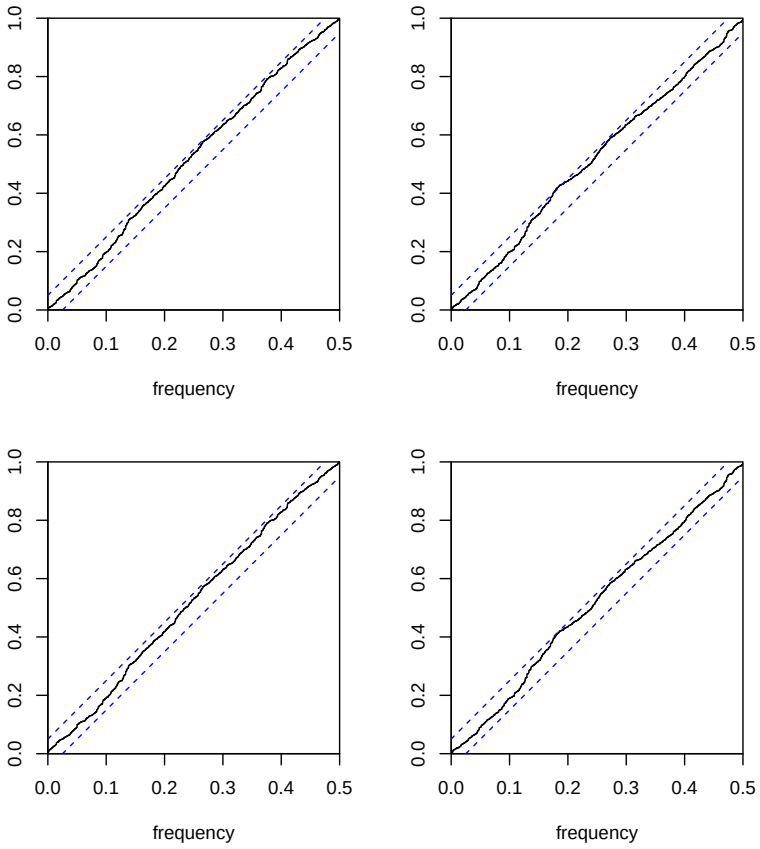

Figure 9.3.: Top: Cumulative periodogram plots of the Pearson residuals from from the linear fit of IBM (left) and Coca-Cola (right). Bottom: Cumulative periodogram plots of the Pearson residuals from the log-linear fit of IBM (left) and Coca-Cola (right)

Table 9.2.: Parameter estimates of the linear model and log-linear model in R and TMB, Standard errors are given in the parentheses

Fitted model	$\hat{d}_1$	$\hat{d}_2$	$\hat{a}_{11}$	$\hat{a}_{22}$	$\hat{b}_{11}$	$\hat{b}_{22}$	$\hat{a}_{12}$	$\hat{a}_{21}$	$\hat{b}_{12}$	$\hat{b}_{21}$
					R					
Linear	0.388	0.348	0.625	0.611	0.126	0.145	0.015	0.103	0.062	0.035
	(1.110)	(0.713)	(0.173)	(0.001)	(0.001)	(0.148)	(0.005)	(0.001)	(0.004)	(0.005)
					TMB					
Linear	0.371	0.329	0.864	0.874	0.086	0.081	-0.072	-0.068	0.042	0.040
	(0.085)	(0.080)	(0.024)	(0.024)	(0.011)	(0.011)	(0.023)	(0.018)	(0.011)	(0.009)
					R					
Log-linear	0.110	0.149	0.830	0.720	0.104	0.141	-0.008	-0.032	0.035	0.026
	(0.001)	(0.152)	(0.085)	(0.035)	(0.143)	(0.056)	(0.003)	(0.001)	(0.012)	(0.0005)
					TMB					
Log-linear	0.220	0.190	0.758	0.868	0.120	0.101	-0.078	-0.174	0.56	0.079
	(0.100)	(0.085)	(0.084)	(0.027)	(0.020)	(0.014)	(0.029)	(0.072)	(0.016)	(0.022)

by $\mathbf{H}_n(\hat{\boldsymbol{\theta}})^{-1}\mathbf{G}_n(\hat{\boldsymbol{\theta}})\mathbf{H}_n(\hat{\boldsymbol{\theta}})^{-1}$. Whereas in TMB, for non-random effect models the standard delta-method is used to calculate the covariance matrix of transformed parameters, see (6).

The models provide a reasonably good fit to the data. In order to access goodness of fit of the models, we use Pearson residuals defined by $e_{i,t} = (Y_{i,t} - \lambda_{i,t})/\sqrt{\lambda_{i,t}}$ for $i = 1$ and 2. We replace $\lambda_{i,t}(\boldsymbol{\theta})$ by $\lambda_{i,t}(\hat{\boldsymbol{\theta}})$ to obtain $\hat{e}_{i,t}$. Under the correct model the sequence $e_{i,t}$ is a white noise sequence with constant variance. This can be investigated by means of cumulative periodograms . If the residuals are white noise the periodogram is a set of iid exponential random variables, asymptotically. So, the cumulative periodogram should look like a straight line at a 45 degree angle. see Figure (9.3), which supports the marginal whiteness of the residuals.

9.2. Application to corporate defaults

A well known fact about corporate defaults is that they tend to cluster overtime. Which is a phenomena that many have sought to explain using different approaches. These have often attempted to distinguish between *contagion effects*, by which "one firm's default increases the probability of other firms defaulting" and *systematic risk*, where risk to the solvency of firms are caused by common underlying macroeconomic and financial factors. Both of these factors may affect the clustering of corporate defaults separately or jointly. Agosto et al., 2016 proposes 8.21 for describing and forecasting aggregate number of corporate defaults; that is, the number of defaults within a given time period. More recently, Geir D Berentsen, Bulla, Maruotti, and Støve, 2018 propose a Markov-switching model, where some of the model parameters depend on the state of an unobserved Markov chain.

9.2.1. Model Inteperatation

Here, we apply the bivariate model given by (8.22) for describing aggregate number of corporate defaults. Interpretation of the model is best approached by considering the following representation of the univariate model which was shown in (4.2) by repeated substitution. Here it is just augmented by a covariate term.

$$\nu_t = d\frac{1-a^t}{1-a} + a^t\nu_0 + b\sum_{i=0}^{t-1} a^i \log(Y_{t-i-1}+1) + c\sum_{i=0}^{t-1} a^i X_t \tag{9.1}$$

Then through this model the systematic risks that firms face that are rooted in common macroeconomic and financial factors are represented by $c\sum_{i=0}^{t-1} a^i X_t$, while $b\sum_{i=0}^{t-1} a^i \log(Y_{t-i-1}+1)$ captures possible feedback effects from past defaults, which is related to overall contagion effects. The parameter d can be thought as to fix the overall intensity. A similar interpretation is possible for the multivariate case as well. Consider the bivariate model which can be written as

$$\begin{aligned}
\nu_{1,t} &= d_1 + a_{11}\nu_{1,t-1} + a_{12}\nu_{2,t-1} + b_{11}\log(Y_{1,t-1}+1) + b_{12}\log(Y_{2,t-1}+1) + c_1 X_t \\
\nu_{2,t} &= d_2 + a_{21}\nu_{1,t-1} + a_{22}\nu_{2,t-1} + b_{21}\log(Y_{1,t-1}+1) + b_{22}\log(Y_{2,t-1}+1) + c_2 X_t
\end{aligned}$$

From the above, we can see for each marginal intensity there is feedback from past defaults coming from, first from its own past, and second from the other marginal , i.e for $\nu_{1,t}$ the feedback effects are captured through $b_{11}\log(Y_{1,t-1}+1)$ and $b_{12}\log(Y_{2,t-1}+1)$. While $a_{11}\nu_{1,t-1}$ and $a_{12}\nu_{2,t-1}$ parsimoniously capture the dependence on previous lags of Y_t as well as the overrates X_t. Supposing for example that 1 and 2 are two different regions, the model can capture possible dependence of defaults in one region on defaults in another region. In the case $a_{12} = b_{12} = 0$, $\nu_{1,t}$ depends only on its own past and covariates.

9.2.2. Model fitting and variable selection

We look at corporate default data from the US. The data contains all of the more than one-thousand large public companies that have filed bankruptcy cases since October 1, 1979. A company is consider "public" if it filed an Annual Report (form 10-K or form 10) with the Securities and Exchange Commission for a year ending not less than three years prior to the filing of the bankruptcy case. A company is considered "large" if that Annual Report reported assets worth $100 million or more, measured in 1980 dollars .

we obtain a bivariate count time series by counting the number of companies that filed for bankruptcy each month for each year from 1980 to 2017, where one series is the number of corporate defaults per month for the blue states, and similarly the other one for the red states. Red states are the states that traditionally vote republican and similarly blues stated are the states that traditionally vote democratic. The financial and macroeconomic covariates we consider are; the industrial production index (indpro), new housing permits (permit), civilian unemployment rate (unrate), Moody's seasoned baa corporate bond yield

(baa), 10-years treasury constant maturity rate (gs10), federal funds rate (fedfunds), producer price index by commodity for final demand: finished goods (ppifgs), and produce price index: fuels and related energy (ppieng). In addition, s&p500 annualized returns (sp500ret) and s&p500 annualized return volatility (sp500vol).

Corporate defaults in the US have been studied by autoregressive count time series models, notably by Agosto et al., 2016 and Geir D Berentsen et al., 2018, where however different the approaches are both fit univariate models. Here we fit bivariate models. The motivation is, Political and cultural polarization in the United States is widely discussed, but there may also be economic disconnection among states. Ishise and Matsuo, 2015, who studied the border effect between red states and blue states find that border effect is robustly confirmed for the 2000s, while not so robustly detected for the 1990s. Border effect can be defined as the additional reduction in the trade between different regions or countries, which is not explainable by simply the size and the distance between the regions or the considered countries. The Border effect is an important indicator for a potential dismantling of the economic connectivity in the United States. Table 9.3 below shows some descriptive statistics for corporate defaults in each group of states.

Table 9.3.: summary statistics

	min	max	1st Qu.	median	3rd Qu.	mean	variance
red states	0	7	0	1	1	0.9625	1.642
blue states	0	11	0	1	2	1.461	2.917

The mean and variance of the number of defaults in red and blue states states are 0.962 and 1.64 and in 1.46 and 2.91, respectively, displaying overdispersion, a classical property of count data. Figure 9.6 shows time series plot of defaults in red and blue states, and figure 9.4 displays the auto and cross correlation functions of defaults in red and blue states . The figures reveal some well known facts about defaults, first, the presence of default clustering over time, and second, the high temporal dependence in default counts.

We start by fitting univariate models, both with and without covariates. Table 9.6 shows the resulting parameter estimates for models without covariates and table 9.5 shows parameter estimates for models with covariates. For the univariate log-linear model without covariates, $\hat{b} = 0.431$, which suggest strong feedback from past defaults on current default counts, therefore strong contagion. The models with covariates allow us to control for financial and macroeconomic factors, representing systematic risk. Particularly, we find the model that includes the the financial market covarite, annualized return volatility (sp500vol), provides the better fit to the data. we find that when including financial covariates, specifically sp500vol and sp500ret, the contagion effect decreases. That is $\hat{b}$ decreases from 0.431 to $\hat{b} = 0.376$ and $\hat{b} = 0.262$, when including sp500vol and sp500ret in to the model, respectively. Therefore, the dependence over time in default counts can partially be explained by systematic risk present in the economy, However dependence of conditional intensity on past defaults counts is still present.

Furthermore, we fit bivariate linear and log-linear models without covariates, with these

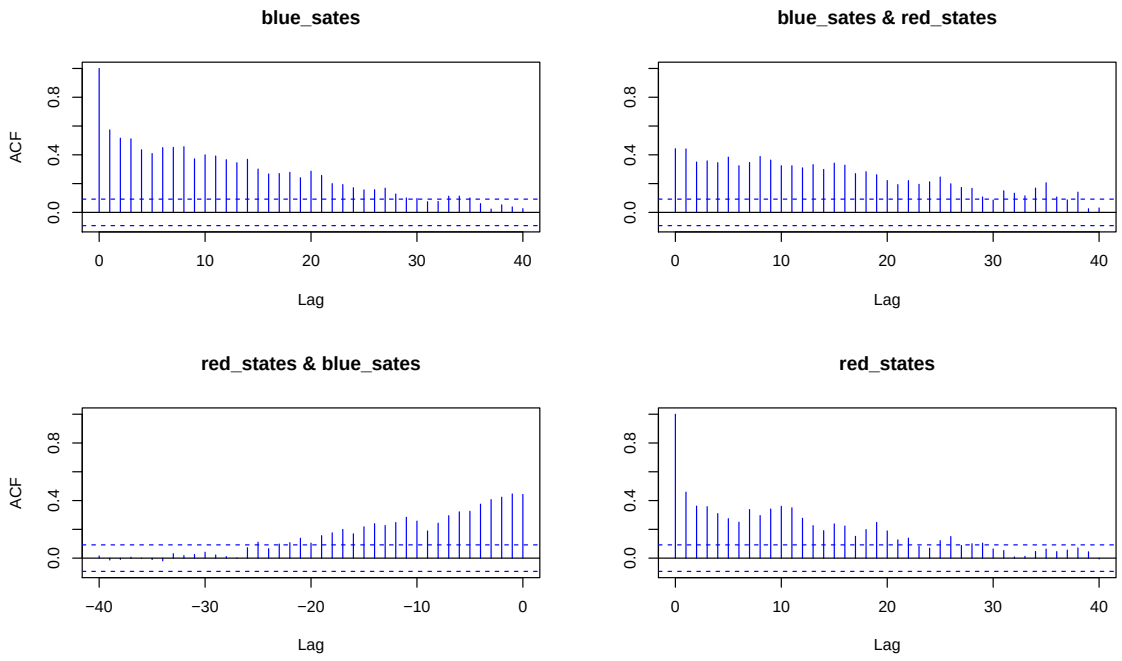

Figure 9.4.: auto and cross correlation functions of defaults in the US

models defaults in each state depends only on past defaults. The bivariate models without covariates perform better in terms of in-sample predictions than their univariate counterpart. The parameter estimates for these models with their corresponding standard errors is given in table 9.6. The assumption about elements of $\mathbf{d}$, $\mathbf{A}$, and $\mathbf{B}$ being positive is not met for the linear model, and the resulting estimates lead to some negative intensities. Thus, we further consider only the log-linear model and an extension of the model that includes covariates. Figure 9.5 shows cumulative periodogram and autocorrelation functions of the Pearson residuals resulting from the log-linear fit, both support the marginal whiteness of the residuals.

Moreover, we fit different bivariate models, each model including different and only one univariate covariate time series. Table 9.4 shows the resulting model parameter estimates for each covariate considered. The bivarite model including covariates, particularly annualized returns (sp500ret), annualized return volatility (sp500vol) and industrial production index (indpro) predict the observed process better than their univariate counterpart. Model selection in this case is difficult through the measures AIC and BIC, as models with lower AIC or BIC do not necessarily result in an improved prediction of the processes. Thus we further examine the models that achieve lower AIC, BIC and prediction error by looking at the resulting Pearson residuals. We find the model with the covariate annualized returns (sp500ret) results in residuals that best approach a white noice, as figure 9.8 shows, and is

Table 9.4.: Parameter estimates of the log-linear model with covariates from all covariates considered

	indpro	permit	ppifgs-up	piend	unrate	baa	fedfunds	gs10	sp500ret	sp500vol
d_1	-0.153	-0.997	-0.245	-0.290	-0.351	-0.312	-0.326	-0.182	-0.182	-0.205
d_2	0.055	-0.343	-0.168	0.164	-0.054	-0.204	0.266	-0.109	-0.109	-0.076
a_{11}	1.000	0.349	0.794	0.794	0.714	0.849	0.905	1.06	1.067	1.046
a_{12}	-0.226	0.631	-0.225	-0.208	-0.139	-0.308	-0.321	-0.480	-0.480	-0.281
a_{21}	0.105	-0.231	-0.107	-0.073	-0.024	-0.030	0.015	0.183	0.183	0.153
a_{22}	0.743	1.23	0.695	0.699	0.845	0.631	0.582	0.395	0.395	0.731
b_{11}	0.165	0.177	0.256	0.290	0.292	0.283	0.275	0.232	0.232	0.098
b_{12}	.162	0.317	0.286	0.266	0.287	0.315	0.311	0.270	0.270	0.169
b_{21}	0.033	0.069	0.054	0.045	-0.089	0.096	0.127	0.130	0.131	-0.070
b_{22}	0.165	0.0929	0.392	0.368	0.249	0.363	0.401	0.325	0.225	0.139
c_1	-0.007	-0.001	-0.016	-0.002	0.027	-0.017	-0.048	-0.010	-0.002	0.114
c_2	-0.009	-0.001	-0.011	-0.001	0.049	-0.022	-0.053	-0.013	-0.001	0.116
AIC	566.83	486.88	538.78	547.13	620.93	561.06	611.99	591.08	662.73	571.18
BIC	555.06	475.11	527.00	535.36	609.16	549.29	600.23	579.31	650.96	559.41
RMSE	35.10	10600	36.21	36.05	35.55	38.81	35.68	35.55	35.09	34.79

thus the preferred model.

With respect to disconnection between the two economies, total disconnection would result in independent marginal intensities, i.e the off-diagonal elements of $\hat{a}_{12}, \hat{a}_{21}, \hat{b}_{12}$ and $\hat{b}_{21}$ all being equal to zero. However we find that these are non-zero as Table 9.6 shows. There is rather strong dependence in defaults across red and blues states as Figure 9.4 also shows.

Moreover, We find that for the log-linear model without covariates, $\hat{b}_{12} = 0.283$, so that the feedback from past defaults on current defaults in red states, comes from past defaults from both red and blue states. Similarly $\hat{b}_{21} = 0.114$, which means current defaults in blue states get feedback from past defaults from red states. Moreover, this suggests that the feedback from past defaults in blue states on current counts of defaults in red states is more than twice stronger than the feedback from past defaults in red states on current counts of defaults in blue states, more simply $\hat{b}_{12} > 2\hat{b}_{21}$.

Furthermore, we look at how $(\hat{b}_{11} + \hat{b}_{12})$ and $(\hat{b}_{21} + \hat{b}_{22})$ change when covariates (sp500ret)

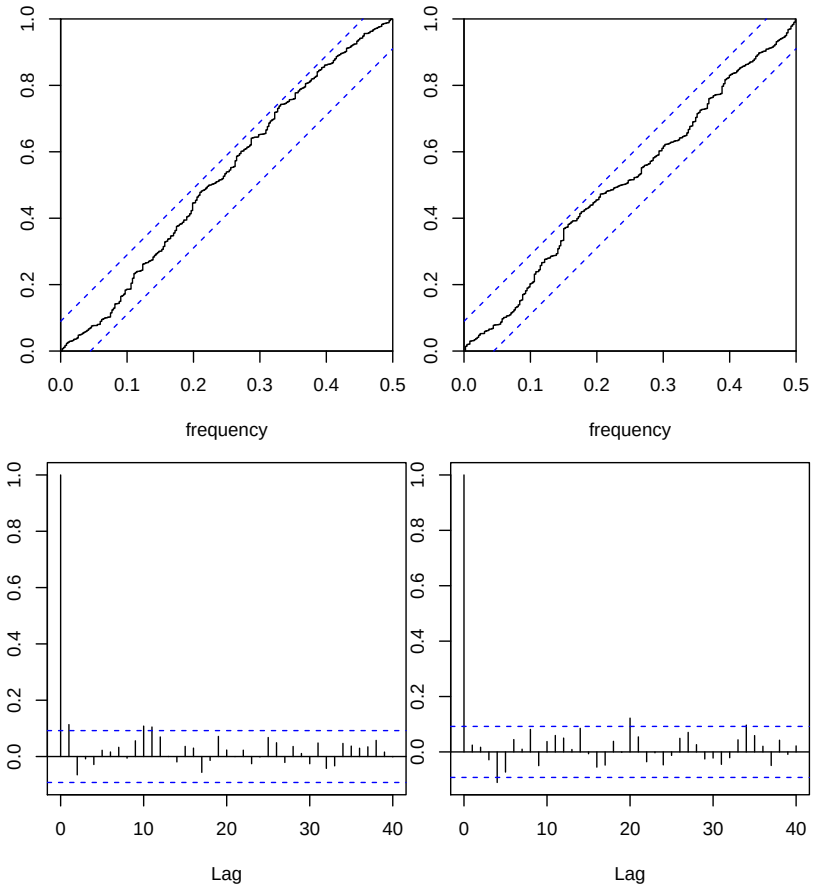

Figure 9.5.: Cumulative periodogram plots of the Pearson residuals from the log-linear fit defaults in red states (left) and blue states (right),and autocorrelation functions Pearson residuals resulting from the model fit, respectively top and bottom

Table 9.5.: Parameter estimates of the univariate log-linear model with covariates from all covariates considered

	indpro	permit	ppifgs-up	piend	unrate	baa	fedfunds	gs10	sp500ret	sp500vol
d	-0.299	-0.0727	-0.311	-0.084	-0.123	-0.058	-0.294	-0.108	-0.086	-0.165
a	0.098	0.681	0.121	0.611	0.543	0.714	0.239	0.587	0.617	0.722
b	0.902	0.312	0.878	0.374	0.450	0.281	0.758	0.406	0.376	0.262
c	-0.013	-0.001	-0.005	-0.001	0.027	0.031	-0.043	-0.022	-0.006	0.107
AIC	1571.85	1589.99	1589.06	1592.50	1595.75	1595.03	1596.04	1598.40	1582.61	1565.48
BIC	1576.08	1594.22	1593.29	1596.73	1599.97	1599.26	1600.27	1602.27	1586.84	1569.72
RMSE	36.45	37.22	36.87	37.04	37.67	37.58	37.52	37.70	36.96	36.02

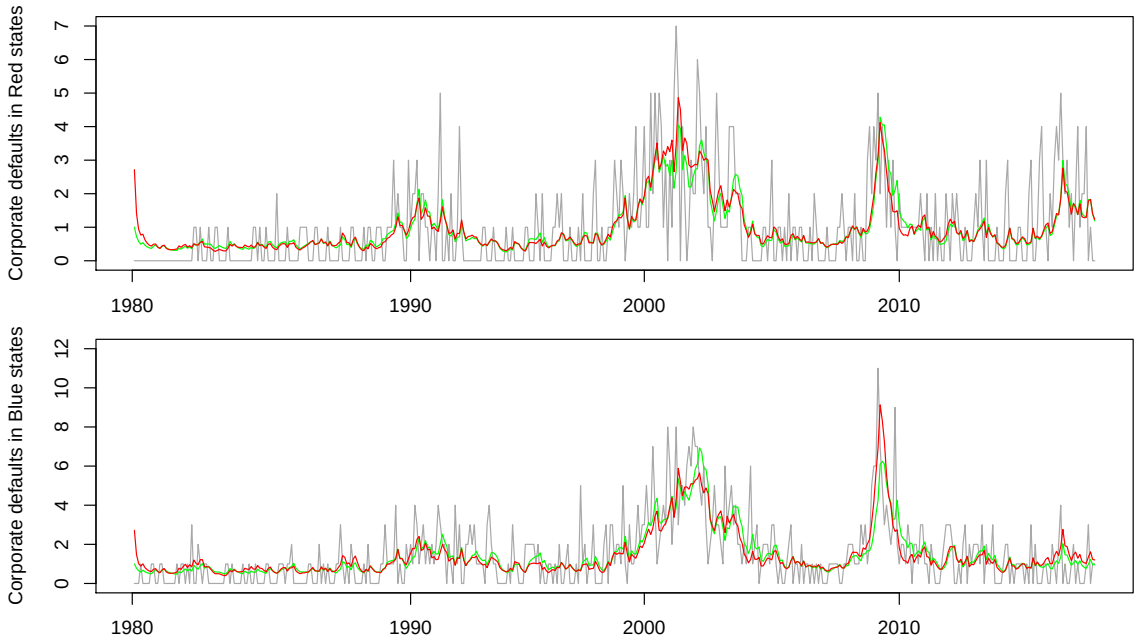

Figure 9.6.: Number of corporate defaults per month inn red states (Top) and blue states (Bottom). The grey line is time series plot of the data. The Green line from log-linear model (8.12) fit to the data, and the red line from log-linear model including covariates (sp500ret) (8.20) fit to the data

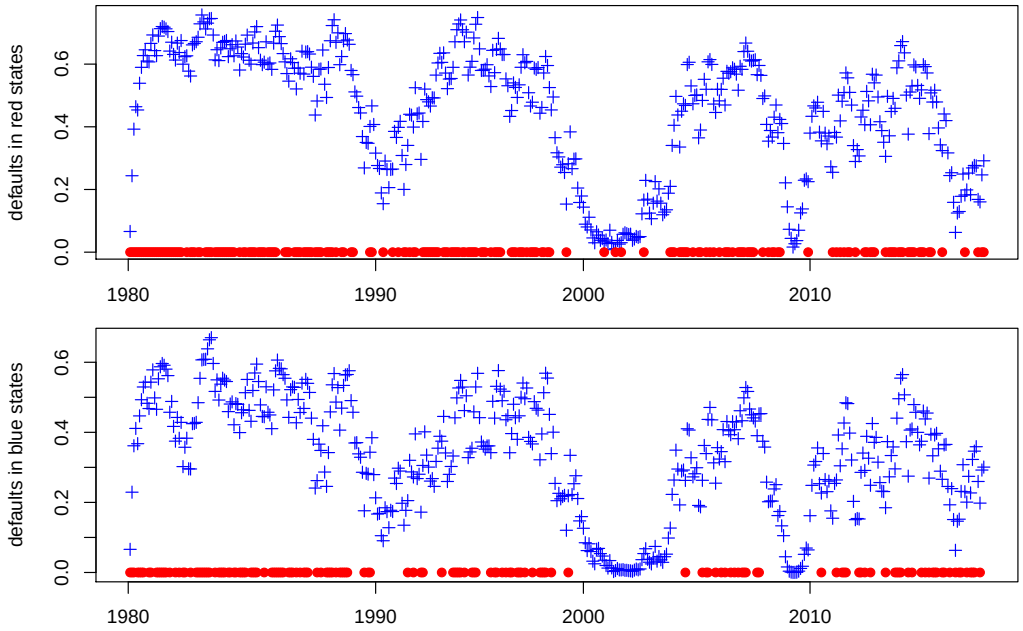

Figure 9.7.: Observed zeroes red and probability of having zeroes predicted by the model with sp500ret (blue crosses)

Table 9.6.: Univariate and bivariate fits without covariates

	linear	SE	log-linear	SE		linear	SE	log-linear	SE
d	0.094	(0.045)	-0.203	(0.035)	d_1	-0.225	(0.269)	-0.326	(0.079)
a	0.692	(0.059)	0.661	(0.048)	d_2	-1.70	(0.220)	-0.254	(0.077)
b	0.261	(0.046)	0.432	(0.059)	a_{11}	-1,90	(0.809)	0.876	(0.138)
					a_{12}	1.79	(0.703)	-0.240	(0.097)
					a_{21}	-2.47	(0.763)	-0.012	(0.239)
					a_{22}	2.43	(0.628)	0.695	(0.111)
					b_{11}	0.233	(0.041)	0.270	(0.069)
					b_{12}	0.117	(0.034)	0.283	(0.092)
					b_{21}	0.243	(0.038)	0.114	(0.066)
					b_{22}	0.145	(0.033)	0.346	(0.081)
AIC	1602.76		1624.39		AIC	1201.18		1215.21	
BIC	1609.00		1630.62		BIC	1193.41		1207.45	
RMSE	37.79		38.36		RMSE	35,49		35.79	

are included to the model. We find that the feedback from past defaults decrease for both red and blue states, $(\hat{b}_{11} + \hat{b}_{12}) = 0.502$, $(\hat{b}_{21} + \hat{b}_{22}) = 0.355$, respectively, when covariates are included, from what they were when leaving covariates out, $(\hat{b}_{11} + \hat{b}_{12}) = 0.553$ and $(\hat{b}_{21} + \hat{b}_{22} = 0.460)$. Decrease in the feedback from past defaults on current defaults is observed for some other covariates considered as well, see 9.4. Even though the feedback decreases when including covariates, it however does not vanish, as it was the case in the univariate model as well. This is in agreement with the results reported in Agosto et al., 2016. Thus conditional on the correct financial and macroeconomic covariates being included this result provides evidence for contagion effects.

To further assess the preferred model (sp500ret), we look at the marginal time series, i.e defaults in red and blue state, which appear to be zero inflated, particularly in the beginning of the 1980s, see figure 9.6. However the model assumes that the multivariate time series is marginally Poisson distributed. Therefore to assess if this assumption holds, we compare the the observed zero counts and the model implied probabilities of observing zeroes, i.e $\hat{P}(Y_{i,t} = 0|\mathcal{F}_{t-1}) = \hat{\lambda}_{i,t}$. Though there are many zero counts, figure 9.7 shows that there is clear correspondence between observed zero counts and model implied probabilities of

observing zero count.

9.2.3. Out of sample prediction

we found in the previous subsection that the preferred model provides a relatively good in-sample performance. In this subsection we further examine the model for out-of-sample performance. To that end, we perform a pseudo-out-of-sample forecasting exercise for the preferred model, and the simpler log-linear model, that is without covariates. We split the sample in to two, where we use the first part to obtain initial estimates, i.e we use $Y_t = Y_1, ..., Y_{T_0=269}$ to obtain initial estimates and the rest $Y_{T_0+1}, ..., Y_T$ are used for one step ahead prediction. Then given MLE $\hat{\boldsymbol{\theta}}_t$ for $t \geq T_0$ we compute the corresponding one-step- ahead prediction $\boldsymbol{\lambda}_{t+1}$ based on information only up to time t, that is $\hat{\boldsymbol{\lambda}}_{t+1} = \boldsymbol{\lambda}_{t+1}(\hat{\boldsymbol{\theta}})$. Practically we have done this by fitting the model repeatedly for ascending data, starting at T_0 from the first split.

we evaluate the out-of-sample performance by a mean-square forecasting error,defines by

$$MSFE_t = \frac{1}{t - T_0} \sum_{s=T_0}^{t} \sum_{i=1}^{p} (Y_{i,s+1} - \hat{\lambda}_{i,s+1}) \tag{9.2}$$

and the average logarithmic forcasting score, given by

$$FS_t = \frac{1}{t - T_0} \sum_{s=T_0}^{t} \sum_{i=1}^{p} (Y_{i,s+1} \log \hat{\lambda}_{i,s+1} - \hat{\lambda}_{i,s+1}) \tag{9.3}$$

For both measures smaller values indicates better forecasting ability for the model. The MSFE measures how well the model does in terms of forecasting defaults, while FS evaluates how well the model does in forecasting the distribution of defaults. Figure 9.9 shows the models out of sample performance, It shows that log-linear model (red line) lies below the lies consistently lower achieving better out of sample performance, despite the covariates model doing best in terms of in-sample prediction. The figure also shows that FS is is the measure that best displays the difference in forecasting performance of the models.

Structural instabilities of model parameters were found in Agosto et al., 2016, which are also present in the models we have fitted here. From figure 9.9 we can see that out-of-sample prediction errors are high in the early 2000s, and again right after the 2008 financial crises. A rolling window for the log-linear model is displayed by figure 9.10, which also supports the structural instabilities. we do not formally test if indeed the parameters vary over time, but it is apparent that the parameters react strongly to the 2008 financial crisis. Agosto et al., 2016 find that during 2007-2011 contagion effects vanish and that default counts are explained by systematic risk. We also fit the bivaraite models on such sub samples, however the models suffer from convergence problems, and the results are not trustworthy, therefore not reported here.

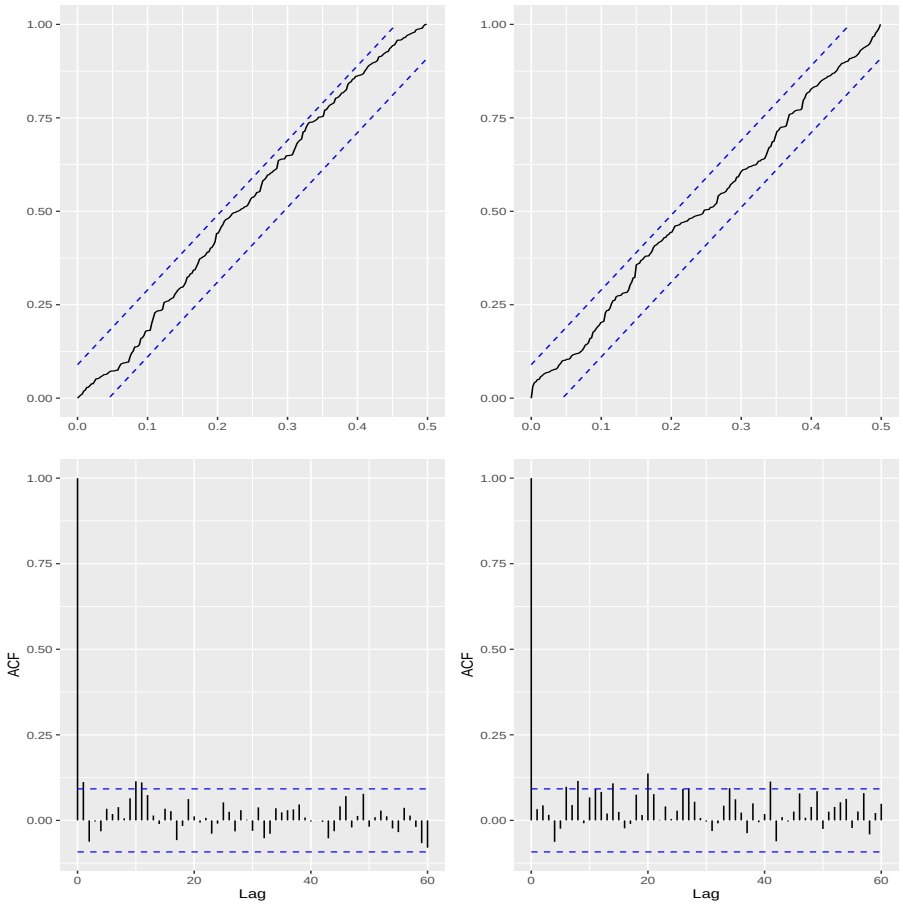

Figure 9.8.: Top: Cumulative periodogram plots of the Pearson residuals from from the log-linear fit including covarites (sp500ret), of defaults in red States (left) and defaulats blue states (right),Bottom: ACF of Pearson residuals resulting from the model fit of defaults in red and blue stated respectively.

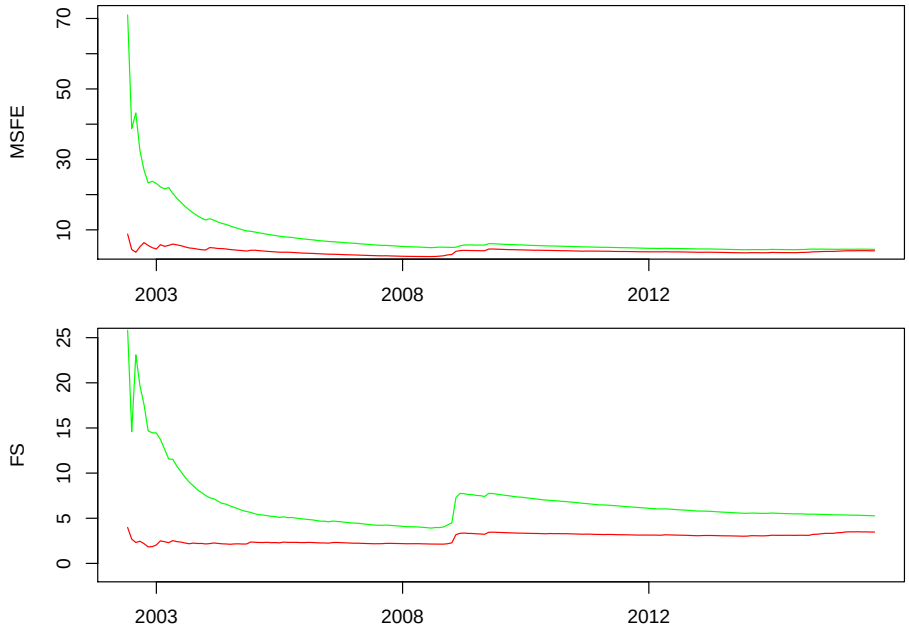

Figure 9.9.: MSFE ans Fs of out of sample one step ahead predictions, Green line resulting from the log-linear covaraite model, and Red line resulting from log-linear model with no cavariates.

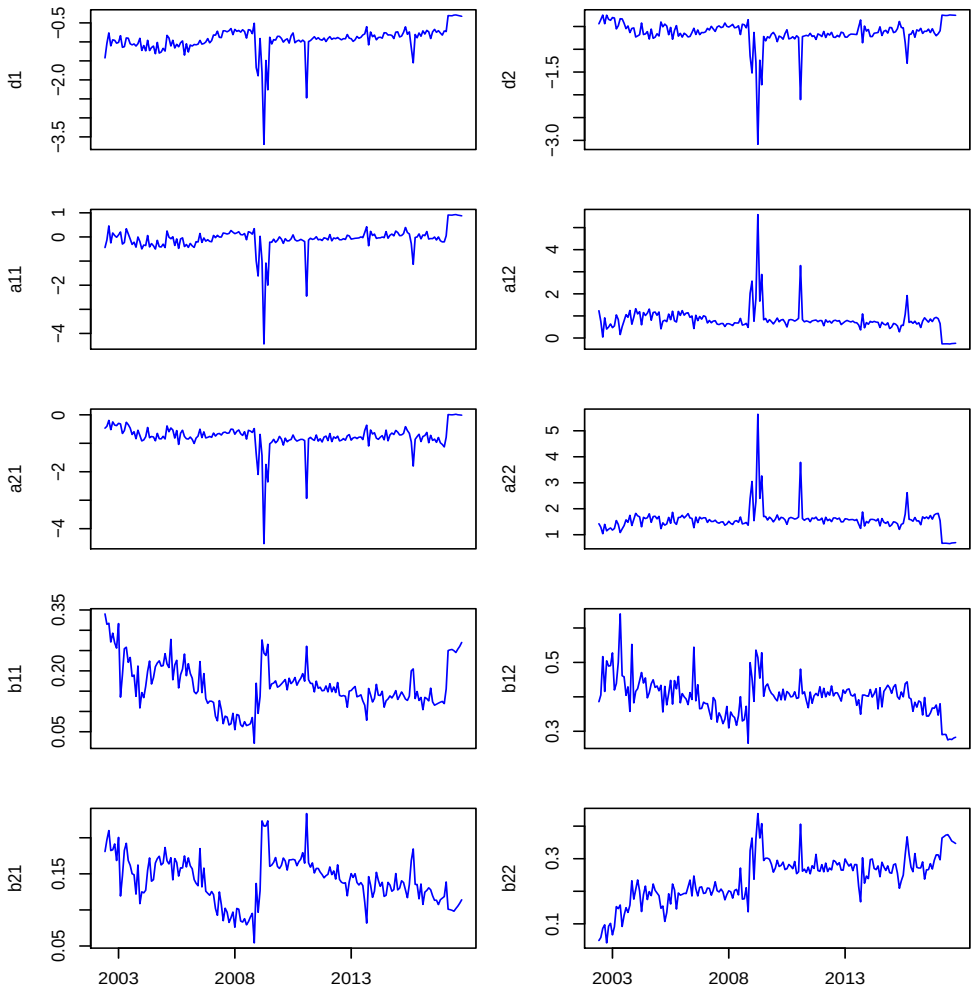

Figure 9.10.: rolling window of log-linear model parameter estimates

10. Copula Estimation

The log-likelihood function (8.5) seems to suggest independence between the marginal series. However dependence is captured through (8.2) and (8.13) on which (8.5) depends. The copula structure does not explicitly appear in (8.5), even though it does indirectly because of the conditional innovations $\mathbf{Y}_t|\boldsymbol{\lambda}_t$. In this section we aim to determine dependence among marginal series. Considering how (8.2) and (8.13) are defined the task reduces to identifying the copula structure and parameter. To that end, One has to compare the conditional distribution of $\mathbf{Y}_t|\boldsymbol{\lambda}_t$ with $\mathbf{Y}_t^*|\boldsymbol{\lambda}_t$, where the former is a parametric bootstrap sample for example generated by the data generating process described in Algorithm 1. Fokianos et al., 2020 suggest an approach based on local Gaussian correlation.

10.1. Copula estimation on simulated data

Fokianos et al., 2020 (see, supplement) suggest the following parametric bootstrap procedure to simulate bivariate count time series and retrieve the copula structure and parameter used to generate the process.

1. Given the observations $Y_{1,t}$ and $Y_{2,t}, t = 1, ..., n$ estimate $\hat{\boldsymbol{\theta}}$ and compute $\hat{\boldsymbol{\lambda}}_{\boldsymbol{t}}$.

2. For a given copula structure and copula parameter generate $Y_{1,t}^*, Y_{2,t}^*$, based on $\hat{\boldsymbol{\theta}}$ from step 1 by the data generating process described in Algorithm 1.

3. Compute the local Gaussian correlation $\rho_{n,b}(\cdot)$ between $Y_{1,t}$and$Y_{2,t}$ and the local Gaussian correlation $\rho_{n,b}^*(\cdot)$ between $Y_{1,t}^*$and$Y_{2,t}^*$ on a pre-defined grid $(u_j, v_j), j = 1, ..., m$.

4. Compute the distance measure $D_m = \frac{1}{m}\sum_{j=1}^{m}[\rho_{n,b}(\cdot) - \rho_{n,b}^*(\cdot)]^2$

5. Repeat steps 2 to 4 for different copula structures and over a grid of values for the copula parameter. The estimate of the copula structure and corresponding copula parameter,$\hat{\phi}$, is the one that minimizes D_m.

Here we use this same approach to retrieve the copula structure and parameter for the log-linear model including covariate time series given by (8.23). Specifically, we generate 100 realizations of bivariate count time series and steps 2-5 are executed for each realization. we only choose between Clayton and Gaussian copulas, which constitutes a pair of simulations. The true parameters used to generate the process is 0.5 for in the Gaussian copula case and 4 for the Clayton copula. Step 5 i executed over a grid of copula parameter values ranging from -1 to 1 and from 0.5 to 8, for the the Gaussian copula and the Clayton copula

Table 10.1.: Simulation results for the parametric copula estimation and copula structure identification for the given model. Results are based on 100 runs, where the distance measure is based on local Gaussian correlation

copula	sample size	parameter	std.error	identified
log-linear model with covariates, where parameters are given by 8.23				
Guassian	500	0.361	0.274	73
with $\phi = 0.5$	1000	0.424	0.215	63
Clayton with	500	2.67	1.81	64
$\phi = 4$	1000	2.38	1.72	71
log-linear model, where parameters are given by 8.19				
Guassian	500	0.458	0.122	76
with $\phi = 0.5$	1000	0.458	0.082	79
Clayton with	500	3.65	0.978	78
$\phi = 4$	1000	3.77	0.772	83

respectively. The local Gaussian correlations are calculated over a diagonal, i.e $(u_j = v_j)$ grid starting from 1 and up to the maximum value in the generated Poisson process in step 1. The bandwidths are calculated as the standard deviation of the same process multiplied with 1.1, Table 10.1 shows the result of the simulation.

Furthermore, we here also look at the copula estimation utilizing the global dependence measure and the most frequently used dependence measure, the Pearson correlation. we make adjustment to the above method as follows; At step 3, Compute the Pearson correlation ρ_n between $Y_{1,t}$ and $Y_{2,t}$ and the Pearson correlation $\rho_n^*(\cdot)$ between $Y_{1,t}^*(\cdot)$ and $Y_{2,t}^*$. At step 4, Compute the distance measure $D = |\rho_n(\cdot) - \rho_n^*(\cdot)|$ and proceed to step 5 as before.

The result of the copula estimation based on Pearson's correlation are given in table 10.2. Copula parameter estimates based on local Gaussian correlation, er close to the true parameter values for the Gaussian copula and not as well for the Clayton copula. However both copula structures are well identified with this methodology. we also see that for increasing sample size copula parameter estimates improve. On the other hand, Copula parameter estimates based Pearson's correlation are more inconsistent. The number of times the copula structure is identified and the copula parameter estimates don't improve for increasing sample size. With both of these distance measures the Clayton copula parameter is not very well estimated. The procedure works well, however a more comprehensive measure of distance between the sample and the bootstrapped sample may improve its performance.

10.2. Copula estimation on real data

Here we estimate the copula parameter and identify the copula structure for the multivariate log-linear model with covarites fitted to the default data in section (9). The Pearson

Table 10.2.: Simulation results for the parametric copula estimation and copula structure identification for the given model. Results are based on 100 runs, where the distance measure is based on Pearson correlation

copula	sample size	parameter	std.error	identified
linear model, where parameters are given by 8.11				
Guassian	500	0.353	0.404	58
with $\phi = 0.5$	1000	0.489	0.206	58
Clayton with	500	6.18	1.52	82
$\phi = 4$	1000	5.76	1.74	79
log-linear model, where parameters are given by 8.19				
Guassian	500	0.468	0.065	73
with $\phi = 0.5$	1000	0.477	0.050	70
Clayton with	500	3.52	0.573	80
$\phi = 4$	1000	3.54	0.393	81

correlation between defaults in red and blue states in 0.443. The local Gaussian correlation is however between 0.259 annd -0.984, and highly nonlinear as figure (10.1) shows. For this model the Gaussian copula is selected 52 times out of 100 and the copula parameter is estimated to be 0.182. The standard error is of the estimates 0.222, which is not too high.

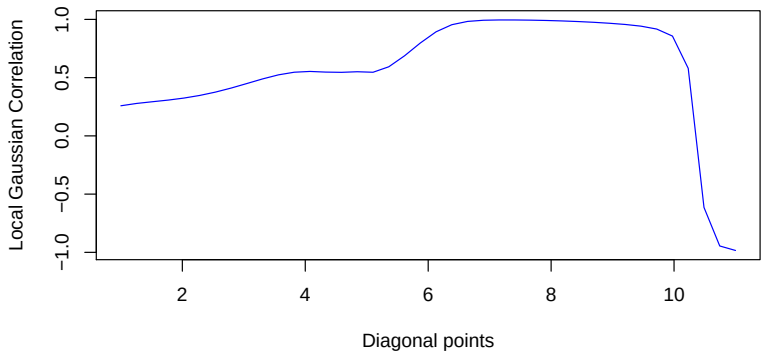

Figure 10.1.: Estimated local Gaussian correlation between defaults in red and blues states on a diagonal grid

11. Conclusion

In this thesis, we have simulated from bivariate linear, log-linear and log-linear with covarites count time series that are marginally Poisson. By simulation we find that the log-linear model with covariates appears to inherit the large sample properties of the log-linear model, i.e maximum likelihood estimators of the models parameters are asymptotically normal. Furthermore, we have used these models to study corporate defaults in the US. We find across region contagion effects are present in the entire sample period, from 1980 to 2017. More specifically, feedback effect coming from past default counts in blue states have stronger impact on current default counts in red states than the other way around. It is however important to note that an economic analysis of contagion and default clustering, such as ours, may sufferer from missspecification error due to the fact that our chosen covariates may not capture all the systematic risk present in the economy, that the firms face. Similarly we have ignored possible feedback from past default counts to the set of covariates.

The two real world application given in this thesis are in finance. For future work, interesting and relevant to current state of the world is to apply the models to COVID-19 data, that are often recorded as aggregated number of confirmed cases. The multivariate models we have discussed can be a great tool to study, say "infection contagion" across regions, with the possibility of including covariates, these covariates can be mitigation strategies taken in the different regions considered. Furthermore, Here we have assumed the conditional distribution $\mathbf{Y}_t|\mathcal{F}_{t-1}$ is marginally Poisson. A natural next step can be to try other marginal conditional distribution,such as negative binomial, that can be generated from continuous waiting times. For example Fokianos et al., 2020 suggest a way to generate count vector whose marginal distribution is mixed Poisson.

A. TMB examples

A.1. Template for a Poisson log-linear model with covariates

```
#include <TMB.hpp>
// Likelihood for a linear count model (model 1)
template<class Type>
Type objective_function<Type>::operator() ()
{

  // Data: univariate  time series and exogeneous covariate time series,
  DATA_VECTOR(y);
  DATA_VECTOR(x);
  // Model parameters

  PARAMETER(d);
  PARAMETER(a);
  PARAMETER(b);
  PARAMETER(c);

  // Set starting values to the expectation of the stationary distribution
  Type ny0 = d/(1 - a - b);
  Type y0 = exp(ny0);
  // define
  int n = y.size();
  vector<Type> ny(n);
  vector<Type> lambda(n);
  Type nll = 0.0;


  // ADREPORT on a, b, d, to obtain standard errors after optimization
  ADREPORT(d);
  ADREPORT(a);
  ADREPORT(b);
  ADREPORT(c)

  // t=1
```

```
  ny(0) = d + a*ny0 + b*log(y0 + 1);
  lambda(0) = exp(ny(0));
  nll -= dpois(y(0), Type(lambda(0)), true);

  // t = 2,...,n
  for(int i = 1; i < n; i++){
    ny(i) = d + a*ny(i - 1) + b*log(y(i - 1) + 1) + c*x(i);
    lambda(i) = exp(ny(i));
    nll -= dpois(y(i), Type(lambda(i)), true);
  } //end of i
  return nll;
}
```

A.2. Template for a bivariate Poisson log-linear model with covarites

```
#include <TMB.hpp>
// Likelihood for a bivariate poisson log linear model including covariates.
template<class Type>
Type objective_function<Type>::operator() ()
{
  // bivariate posisson distributed count time series,Nx2 matrix, with a bivariate covariate t.s
  DATA_MATRIX(Y);
  DATA_MATRIX(X);
  // parameters
  PARAMETER_VECTOR(p);

  // split the data into two column vectors,
  vector<Type> Y1 = Y.col(0);    // 1st col of Y
  vector<Type> Y2 = Y.col(1);    // 2dn col of Y
  vector<Type> X1 = X.col(0);    // 1st col of X
  vector<Type> X2 = X.col(1);    // 2dn col of X

  int n = Y1.size();              // Length of the data
  vector<Type> loglik1(n);
  vector<Type> loglik2(n);
  vector<Type> loglik(n);

  // report parameter estimates with standard errors
  ADREPORT(p);
```

```
25
26    vector<Type> nu1(n);
27    vector<Type> nu2(n);
28    vector<Type> lambda1(n);
29    vector<Type> lambda2(n);
30    //Some initial values
31    loglik1(0) = 0;
32    loglik2(0) = 0;
33    loglik(0) = 0;
34    nu1(0) = 0;
35    nu2(0) = 0;
36    lambda1(0) = 1;
37    lambda2(0) = 1;
38    for(int t=1;t<n;t++){
39      nu1(t) = p(0) + p(2)*nu1(t-1) + p(4)*nu2(t-1) +
40               p(6)*log(Y1(t-1)+1) + p(8)*log(Y2(t-1)+1)+ p(10)*X1(t)+p(12)*X2(t);
41      lambda1(t) = exp(nu1(t));
42      nu2(t) = p(1) + p(3)*nu1(t-1) + p(5)*nu2(t-1) +
43               p(7)*log(Y1(t-1)+1) + p(9)*log(Y2(t-1)+1)+p(11)*X1(t)+p(13)*X2(t);
44      lambda2(t) = exp(nu2(t));
45      if(lambda1(t)<=0){
46        loglik1(t)=0;
47      }
48      else if(lambda1(t)>0){
49        loglik1(t) = -Y1(t)*log(lambda1(t)) + lambda1(t);
50      }
51      if(lambda2(t)<=0){
52        loglik2(t) = 0;
53      }
54      else if(lambda2(t)>0){
55        loglik2(t) = -Y2(t)*log(lambda2(t)) + lambda2(t);
56      }
57      loglik(t) = loglik1(t) + loglik2(t);
58    }
59    return(loglik.sum());
60  }
```

Bibliography

Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62*(2), 269–276.

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., ... Lauritzen, S. L. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, 93–115.

McKenzie, E. (1985). Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association*, *21*(4), 645–650.

Alzaid, A., & Al-Osh, M. (1988). First-order integer-valued autoregressive (inar (1)) process: Distributional and regression properties. *Statistica Neerlandica*, *42*(1), 53–61.

Kedem, B., & Fokianos, K. (2005). *Regression models for time series analysis*. John Wiley & Sons.

Schmidt, T. (2007). Coping with copulas. *Copulas-From theory to application in finance*, 3–34.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts—a survey. *AStA Advances in Statistical Analysis*, *92*(3), 319.

Fokianos, K., Rahbek, A., & Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, *104*(488), 1430–1439.

Jaworski, P., Durante, F., Hardle, W. K., & Rychlik, T. (2010). *Copula theory and its applications*. Springer.

Fokianos, K., & Tjøstheim, D. (2011). Log-linear poisson autoregression. *Journal of Multivariate Analysis*, *102*(3), 563–578.

Schmidt, A. M., & Pereira, J. B. M. (2011). Modelling time series of counts in epidemiology. *International Statistical Review*, *79*(1), 48–69.

Bell, B. M. (2012). Cppad: A package for c++ algorithmic differentiation. *Computational Infrastructure for Operations Research*, *57*(10).

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., ... Sibert, J. (2012). Ad model builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, *27*(2), 233–249.

Tjøstheim, D. (2012). Some recent theory for autoregressive count time series. *Test*, *21*(3), 413–438.

Team, R. C. et al. (2013). R: A language and environment for statistical computing.

Tjøstheim, D., & Hufthammer, K. O. (2013). Local gaussian correlation: A new measure of dependence. *Journal of Econometrics*, *172*(1), 33–48.

Berentsen, G. D. [Geir Drage], Kleppe, T. S., & Tjøstheim, D. (2014). Introducing localgauss, an R package for estimating and visualizing local gaussian correlation. *Journal of Statistical Software*, *56*(12), 1–18.

Støve, B., Tjøstheim, D., & Hufthammer, K. O. (2014). Using local gaussian correlation in a nonlinear re-examination of financial contagion. *Journal of Empirical Finance*, *25*, 62–82.

Ishise, H., & Matsuo, M. (2015). Trade in polarized america: The border effect between red states and blue states. *Economic Inquiry*, *53*(3), 1647–1670.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. (2015). Tmb: Automatic differentiation and laplace approximation. *arXiv preprint arXiv:1509.00660*.

Ravishanker, N., Venkatesan, R., & Hu, S. (2015). Dynamic models for time series of counts with a marketing application. *Handbook of discrete-valued time series, RA Davis, SH Holan, R. Lund and N. Ravishanker (eds.)* 423–445.

Agosto, A., Cavaliere, G., Kristensen, D., & Rahbek, A. (2016). Modeling corporate defaults: Poisson autoregressions with exogenous covariates (parx). *Journal of Empirical Finance*, *38*, 640–663.

Sucarrat, G., Grønneberg, S., & Escribano, A. (2016). Estimation and inference in univariate and multivariate log-garch-x models when the conditional density is unknown. *Computational statistics & data analysis*, *100*, 582–594.

Bosowski, N., Ingle, V., & Manolakis, D. (2017). Generalized linear models for count time series. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 4272–4276). IEEE.

Berentsen, G. D. [Geir D], Bulla, J., Maruotti, A., & Støve, B. (2018). Modelling corporate defaults: A markov-switching poisson log-linear autoregressive model. *arXiv preprint arXiv:1804.09252*.

Halliday, J., & Boshnakov, G. N. (2018). Poarx modelling for multivariate count time series. *arXiv preprint arXiv:1806.04892*.

Fokianos, K., Støve, B., Tjøstheim, D., Doukhan, P., et al. (2020). Multivariate count autoregression. *Bernoulli*, *26*(1), 471–499.