https://doi.org/10.26219/heal.aueb.6542

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
DEPARTMENT OF STATISTICS

Some models for multivariate time series for counts

By
Christina Xeni

Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
February 2022

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervising Prof.Konstantinos Fokianos and Prof.Dimitris Karlis for the valuable help and guidance in writing this work. But also those who supported me for the completion of this postgraduate program.

https://doi.org/10.26219/heal.aueb.6542

CURRICULUM VITAE

I was born in Nicosia on 13/03/1997. I spent my childhood in Limassol where I finished high school with a general grade 18.5. In 2015, I was admitted through the Cypriot examinations in the Department of Mathematics and Applied Mathematics in the field Mathematics of University of Crete. I graduated from the university in 2020 with an overall grade of 6.71. In the year 2021 I started my postgraduate course in Department of Statistics of the Athens University of Economics and Business.

https://doi.org/10.26219/heal.aueb.6542

ABSTRACT

In many fields data are presented that evolve together over time. Such data can be the prices of some shares on the stock exchange, the murders in different regions for a certain period of time or the arrivals at the different airports of a specific country. In the literature there are categories of models capable of describing such data as parameter driven models and observation driven models. Observation driven models are very popular for describing such data due to their ease in estimating parameters which is not true for parameter driven models. In this thesis, to emphasizing the advantages of parameter driven models, we present some of them that are flexible to describe data that evolve over time and describe cross-correlation, autocorrelation and overdispersion. Specifically, we will describe five parameter driven models, the State Space Multivariate Poisson model (SSMP), a doubly stochastic model with latent factors, multivariate Poisson scaled beta (MPSB) models, a dynamic factor model and the hierarchical Markov switching model (HMSM). All models to be presented are models that use modern numerical methods for parameter estimation and the suitability of these methods has been documented with examples.

Keywords: Multivariate time series of count, Autocorrelation, Cross correlation, Overdispersion, Parameter driven model, State space model.

https://doi.org/10.26219/heal.aueb.6542

Περίληψη

Σε πολλά επιστημονικά πεδία παρουσιάζονται δεδομένα που εξελίσσονται μαζί με την πάροδο του χρόνου. Τέτοια δεδομένα μπορεί να είναι οι τιμές κάποιων μετοχών στο χρηματιστήριο, οι δολοφονίες σε διαφορετικές περιοχές για μια συγκεκριμένη χρονική περίοδο ή οι αφίξεις στα διαφορετικά αεροδρόμια μιας συγκεκριμένης χώρας. Στη βιβλιογραφία υπάρχουν κατηγορίες μοντέλων ικανών να περιγράψουν τέτοια δεδομένα όπως μοντέλα που βασίζονται σε παραμέτρους και μοντέλα με γνώμονα την παρατήρηση. Τα μοντέλα που βασίζονται στην παρατήρηση (observation driven models) είναι πολύ δημοφιλή για την περιγραφή τέτοιων δεδομένων λόγω της ευκολίας τους στην εκτίμηση των παραμέτρων, κάτι που δεν ισχύει για τα μοντέλα που βασίζονται σε παραμέτρους (parameter driven models).

Σε αυτή τη διατριβή, για να τονίσουμε τα πλεονεκτήματα των μοντέλων που βασίζονται σε παραμέτρους, παρουσιάζοντας μερικά από αυτά, τα οποία είναι ευέλικτα και ικανα να περιγράψουν δεδομένα που εξελίσσονται με την πάροδο του χρόνου και περιγράφουν τη συσχέτιση μεταξύ των χρονολογικών σειρών (cross-correlation), την αυτοσυσχέτιση και την υπερδιασπορά των δεδομένων. Συγκεκριμένα, θα περιγράψουμε πέντε μοντέλα που βασίζονται σε παραμέτρους, το State Space Multivariate Poisson Model (SSMP), ένα διπλά στοχαστικό μοντέλο με λανθάνοντες παράγοντες, πολυμεταβλητά μοντέλα Poisson σε κλίμακα βήτα (MPSB), ένα μοντέλο δυναμικού παράγοντα και το ιεραρχικό λανθάνων μοντέλο μεταγωγής Markov (HMSM).

Όλα τα μοντέλα που θα παρουσιαστούν είναι μοντέλα που χρησιμοποιούν σύγχρονες αριθμητικές μεθόδους για την εκτίμηση των παραμέτρων και η καταλληλότητα αυτών των μεθόδων έχει τεκμηριωθεί με παραδείγματα.

Λέξεις κλειδία: Πολυμεταβλητές χρονοσειρές μέτρησης, Αυτοσυσχέτιση, Διασταυρούμενη συσχέτιση, Υπερδιασπορά, Μοντέλο βάσει παραμέτρων, Μοντέλο χώρου κατάστασης.

https://doi.org/10.26219/heal.aueb.6542

Contents

https://doi.org/10.26219/heal.aueb.6542

Chapter 1

Introduction

1.1 Multivariate time series of counts

Multivariate count data that evolve over time are data that correspond to multiple measurements observed together for time t, with t=$1, ..., T$. Such data occur in many fields, such as economics, epidemiology, biology, etc. A key feature of this type of data is that usually exhibits cross-correlation. By using the term cross-correlation we mean the measure of similarity of two series as a function of the displacement of one relative to the other, see Shumway et al. (2000). Alongside, quite often we see that such data have serial correlations with their previous observations i.e autocorrelation. By using the term autocorrelation, we mean the Pearson correlation between values of time series in different time points. For example for the time series X_t the autocorrelation is defined to be $corr(X_t, X_{t'})$ for $t \neq t^{'}$, where,

$$corr(X_t, X_{t'}) = \frac{Cov(X_t, X_{t'})}{\sqrt{Var(X_t)Var(X_t)}}, \quad \text{for } t = t^{'}.$$

Using autocorrelation matrix or autocovariance matrix term we mean a $n \times n$ matrix giving the autocorrellation between each pair of elements of a given time series. The autocorrelation of a time series X_t is defined in Brockwell

https://doi.org/10.26219/heal.aueb.6542

and Davis (2009). Note that, the serial correlation and cross-correlation can be both negative and positive because the correlation coefficient is betwenn -1 and 1. Furthermore, count data may have overdispersion that is the observation of a greater variation in the data than the mean. The overdispersion is often observed in Poisson regression models but also in time series models. More details on what exactly overdispersion is and what problems it can cause are presented both in Cox and Isham (1980) and Cox (1983).

All the above are the key features of multivariate count data that evolve over time that one must keep in mind when proceeding to the modeling of such data.

With term that derives from time series, the multivariate count data that evolve over time, is called multivariate time series of counts; Most models that describe the multivariate time series of counts are classified into two major categories named parameter-driven models and observation driven models respectively Cox et al. (1981). In general, these two types of models have a common observation equation

$$P(Y_t|\alpha_t) := P(Y_t|\alpha_t, \alpha_{t-1}, Y_{t-1})$$

where, Y_t=$\{Y_{t1}, .., Y_{td}\}$ are observations and α_t=$\{\alpha_{t1}, ..., \alpha_{td}\}$ are state variables. But for parameter driven models are state equation $P(\alpha_t)$ does not depend on the past observations in contrast to observation driven models Davis et al. (2004). One reason that observation driven models are quite popular is that due to the state parameters the likelihood function of them is given in closed-form. It is a significant important property because it gives models the advantage of simple estimation procedures. For parameter driven models, the opposite is true, i.e the likelihood function is not available in closed-form. However, the parameter-driven models have the advantage to being flexible and easily applied in new settings while the first category of

https://doi.org/10.26219/heal.aueb.6542

models contains models that are not so flexible. More and detailed comparisons between the two categories can be found in Creal et al. (2008).

1.2 Observation Driven models

The key feature of observation driven models is that there are deterministic predictions for the current state variables using the previous observations. In other words, if Y_t is a time series of counts (observed) and α_t the state variables then we can define an observation driven model as:

$$Y_t|\alpha_t \sim P(Y_t|\alpha_t)$$

where α_t at time $t-1$ and $P(\alpha_t|Y_{t-1})$ are known. There are several articles that deal extensively with observation driven models and their properties and estimation of their parameters, some of them are, Davis et al. (2003), Creal et al. (2008), Blasques et al. (2018), Fokianos and Fried (2010).

Some well recognized observation driven models are INGARCH (ARCH/-GARCH) and INAR models. Certainly, there are many popular observation driven models but we will present these two and some very specific properties of them.

The Autoregressive conditional heteroscedastic (ARCH) models for Y_t time series were introduced for the first time in the Engle (1982) in order to generalize the assumption that there is constant one-period forecast variance assumed by traditional econometric models. The form of ARCH(1) model in Engle (1982) was presented as follows:

$$\begin{aligned} Y_t &= \epsilon_t h_t^{1/2} \\ h_t &= \alpha_0 + \alpha_1 Y_{t-1}^2 \end{aligned}$$

https://doi.org/10.26219/heal.aueb.6542

where $Var(\epsilon_t) = 1$ so $Var(Y_t)$=h_t and α is a vector of unknown parameters. That is, the variance of Y_t depends on the previous observation and on an unknown parameter.
Assuming that:

$$\begin{aligned}
&Y_t|\psi_{t-1} \sim N(X_t\beta, h_t)\\
&h_t = h(\epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_{t-p}, \alpha)\\
&\epsilon_t = Y_t - X_t\beta.
\end{aligned} \tag{1.1}$$

Where X_t is a vector of exogenous variables and β is a vector of regression parameters.
We get ARCH regression model order p. The model (1.1) has been used extensively in econometric as it provides a plethora of features that make it very attractive in the space in question. See Engle (1982) for more information. Moreover, in the literature one can also find information about the generalization of ARCH, ie the GARCH model proposed by Bollerslev (1986) and Taylor (1986). Note that the difference between the two models is in the conditional variance of Y_t. Since for the GARCH model holds that:

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}.$$

More information on the models mentioned above can be found in Teräsvirta (2009).

In order to describe time series of counts a generalization of the above models was introduced. This generalization is names integer-valued autoregressive conditional heteroskedasticity time series model (INGARCH) and is defined as shown below.

$$Y_t|F_{t-1} \sim Poisson(\lambda_t) \tag{1.2}$$

where, λ_t=$\gamma_0 + \sum_{i=1}^{q} \gamma_i Y_{t-i} + \sum_{j=1}^{p} \delta_j Y_{t-j}$, $\gamma_0 > 0, \gamma_i \geq 0, i = 1, ..., q, \delta_j \geq 0, j = 1, ..., p$.

https://doi.org/10.26219/heal.aueb.6542

The INGARCH model can only support positive crosscorrelation amd autocorrelation structures.

McKenzie (1985) and independently Al-Osh and Alzaid (1987) introduced the Integer-valued AutoRegressive (INAR(1)) models. A kind of model for describing a Y_t time series of counts. The form of these kind of models is given as

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

where Y_t and ϵ_t are non-negative integer values and α is an unknown parameter. For ϵ_t we assume that is independent and identically distributed random variables (iid) and stochastically independent of the previous observations. Also, we assume that Y be non-negative integer-valued random variable and $\alpha \in [0, 1]$ such that,

$$\alpha \circ Y = \sum_{i=1}^{Y} X_i,$$

where X_i is a sequence of independent and identically distributed random variables (iid) which are independent of Y, with

$$P(X_i = 1) = 1 - P(X_i = 0) = \alpha.$$

In Al-Osh and Alzaid (1987) one can find three analytic ways for the parameter estimation of INAR(1) model, The Yule-Walker estimators, the conditional least squares estimators and the maximum likelihood estimators.

Since then, several variants of this model have emerged adapted to multivariate time series for counts. Some of them are the bivariate integer valued autoregressive model of order 1 (BINAR(1)) or the full multivariate INAR(1) (MINAR).

https://doi.org/10.26219/heal.aueb.6542

For a bivariate time series of counts $Y_t = \{Y_{t1}, Y_{t2}\}$ the BINAR(1) process has the following form:

$$Y_t = A \circ Y_{t-1} + \epsilon_t \quad \text{for t=1,...,T}$$

where Y_t and ϵ_t are non-negative integer vector and the matrix of autocorrelation parameters A is a 2×2 diagonal matrix with independent elements $\{\alpha_{j,j}\}_{j=1,2}$. It is noticeable, that in the bivariate case, the $\circ$ is the usual matrix multiplication that also keeps the properties of the binomial thinning operation. Some key results that come from how this model is defined are listed below.

First of all the BINAR (1) can describe only data with positive correlations. Since the following relationships apply:

$$E(Y_{tj}) = \frac{\lambda_j}{1 - \alpha_{jj}} \tag{1.3}$$

$$Var(Y_{tj}) = \frac{(\alpha_{jj} + \nu_j)\lambda_j}{1 - \alpha_{jj}^2} \tag{1.4}$$

$$Cov(Y_{tj}, Y_{(t+i)j}) = \alpha_{jj}^i Var(x_{tj}) \quad \text{for i=1,2,3,...} \tag{1.5}$$

$$Corr(Y_{tj}, Y_{(t+i)j}) = \alpha_{jj}^i \quad \text{for i=1,2,3,...} \tag{1.6}$$

where λ_j is the mean of ϵ_{tj} and variance of ϵ_{tj} is the $\sigma_j = \nu_j \lambda_j$ for $\nu_j > 0$. Also, for the serial cross correlation of Y_{t1} where $i = 0, 1, 2, 3, ...$ the relation

$$Corr(Y_{t1}, Y_{(t+i)2}) = \frac{\alpha_{11}^i \sqrt{(1 - \alpha_{11}^2)(1 - \alpha_{22}^2)}}{(1 - \alpha_{11}\alpha_{22})\sqrt{(\alpha_{11} - \nu_1)(\alpha_{22} - \nu_2)\lambda_1\lambda_2}} Cov(\epsilon_{t1}, \epsilon_{t2}) \tag{1.7}$$

is valid. Consequently, the correlation between the Y_{tj} and Y_{ti} arises only through the correlation of ϵ_{tj} and ϵ_{ti}, so we end up with simple models. Finally, we can obtain the parameter estimation of the BINAR models using

https://doi.org/10.26219/heal.aueb.6542

maximum likelihood (ML). More information about bivariate integer valued autoregressive model of order 1 (BINAR(1)) one can find in Pedeli and Karlis (2011),Pedeli and Karlis (2013a),Mamode Khan et al. (2016).

The full multivariate INAR(1) (MINAR) was developed to relax the assumption that the matrix A is diagonal, see Pedeli and Karlis (2013b). In the special case of the full multivariate INAR(1) (MINAR), the form of the model is the same as the form of BINAR(1) model but now we have A is a $d \times d$ matrix with independent elements $\{\alpha_{ij}\}_{i,j=1}^{d}$. So, we obtain the equation

$$Y_{tj} = \sum_{i=1}^{d} \alpha_{ji} \circ Y_{(t-1)i} + \epsilon_{tj} \tag{1.8}$$

$for j = 1, .., d$. From Pedeli and Karlis (2013b) immediately follows that MINAR models can only support positive cross-correlations and autocorrelations of count data.

1.3 Parameter driven models

As it has been reported what makes the parameters driven models different from the observation driven models is that in the parameters driven models the state variables are not observed. Description of the parameter driven model can be seen among others in Liu (2012), Fokianos (2021). On the one hand, due to this feature, the estimation of parameters for the models that belong to this category proves to be difficult. On the other hand, due to different types of structure of the unobserved state variables there are many variants of this kind of models.

It can be said that parameter driven models are divided into two major categories, the linear state-space models with Gaussian innovations and the non-linear non-Gaussian state-space models Vermeer (2020). As for linear

https://doi.org/10.26219/heal.aueb.6542

state space model in Kalman (1960) one can find analytical methods for estimating state variables as in the case of observation driven models. Obviously, if we estimate the state variable, we can easily proceed to the one-step-ahead forecasts for Y_t. This is not the case with non-linear non-Gaussian state-space models. A review of non-Linear non-Gaussian State Space Models is presented in Triantafyllopoulos (2021). There are several articles that have studied the estimation problem such as Chan and Ledolter (1995). Therefore, numerical methods were developed to implement the parameter estimation for these models. We briefly mention some methods, Particle Filtering (Djuric et al. (2003)), Online Variation Filtering and Parameter Learning (Campbell et al. (2021)) and Bellman filter Lange (2020).

1.4 The structure of the thesis

The rest of the thesis is organized as follows. Section 2 introduces the five parameter driven models and and some conclusions for each model are presented. Section 3 introduces an algorithm for parameter estimation of a model. Section 4 provides concluding remarks.

https://doi.org/10.26219/heal.aueb.6542

Chapter 2

Some Parameter driven models for multivariate time series of counts

Parameter driven models are very common due to their flexible nature. The detriment of these models lies in the difficulty that exists during the estimation of their parameters. Then you can find five parameter driven models in which the parameter estimation is done with modern numerical means. That is, parameter estimation is not as difficult as one might expect. Therefore, their disadvantage seems to be smoothed out and thus it becomes easier to focus on the positives of these models.

More specifically, the state space multivariate Poisson model (SSMP) (Zhang et al. (2017)) is introduced first. This model describes autocorrelated multivariate counts and has the advantage that can describe both the cross-correlation and autocorrelation structure of count data. Also, overdispersion is accommodated. One can say that the disadvantage of this model lies in the fact that it can be applied only to stationary multivariate time series of counts. Prediction and parameter estimation are performed using the parti-

https://doi.org/10.26219/heal.aueb.6542

cle filtering (PF, also called Sequential Monte Carlo) and the Monte Carlo EM algorithm (MCEM).

Then a doubly stochastic model with latent factors is presented which is developed in Wang and Wang (2018) and can describe correlation and autocorrelation of multivariate time series of counts. This model has the advantage that can be used by both stationary and non-stationary time series. Here we used pseudo-maximum-likelihood for parameter estimation and eigenanalysis.

Next the Multivariate Poisson-Scaled Beta (MPSB) model is introduced (Aktekin et al. (2018)) which allows for serial and cross correlation. Through the development of this model, the creation of a new marginal distributions is achieved and is named dynamic multivariate negative binomial (DMNB) distribution. In general, this model is ideal for fast online updating but can only describe positive correlations. The Forward Filtering and Backward Sampling (FFBS) algorithm is used for parameter estimation and Particle Learning (PL) is used for the update of parameters.

The next model to be presented is a dynamic factor model for multivariate measurement data and which introduced in Jung et al. (2011). Like the above, this model describes serial correlation. One feature of this model that makes it stand out from the rest is that the conditional distribution of counts is not necessarily Poisson as is the case with most parameter driven models. To be more precise, the conditional distribution of counts when applying this model will be negative binomial (Negbin) which can take the overdispersion into account. The efficient importance sampling (EIS) is used for parameter estimation. Note that this model applies to specific data (The numbers of trades in 5-min intervals for five stocks traded at the New York Stock Ex-

https://doi.org/10.26219/heal.aueb.6542

change (NYSE)).

The last model presented in this thesis is called hierarchical Markov switching model (HMSM). Its properties and its special cases are presented in detail in Catania and Di Mari (2021). For parameter estimation of this model the EM algorithm is used.

In each model that follows, we will use $Y_t = [Y_{t1}, .., Y_{td}]$ for time $t = 1, .., T$ as the multivariate time series of counts that we want to model. To be more precise, we will use the term state space model instead of parameter driven model whenever necessary.

2.1 State Space Modeling of Autocorrelated Multivariate Poisson Counts

In this section we will introduce the "State Space Multivariate Poisson Model". This is a state space model which describes autocorrelated multivariate counts and is described in detail in Zhang et al. (2017). It is useful as it flexibly illustrates the structure of cross-correlation and autocorrelation of measurement data and also takes overdispersion into account.

We will start by making the following assumptions for the time series Y_{ti} which apply for all $t = 1, .., T$:

Assumption 1.1. Y_{ti}, conditional on its mean λ_{ti} are independent.

Assumption 1.2. Y_{ti} follows a Poisson distribution with rate λ_{ti} and with probability mass function (pmf) given by

https://doi.org/10.26219/heal.aueb.6542

$$P(Y_{ti}|\lambda_{ti}) = \frac{\exp(-\lambda_{ti})\lambda_{ti}^{Y_{ti}}}{Y_{ti}!}, \lambda_{ti} > 0. \tag{2.1}$$

The function (2.1) gives the probability that Y_{tj} get a specific value for a given λ_{tj}.

From Assumption 1.1 and Assumption 1.2, it follows that the joint conditional distribution of Y_t is given by

$$P(Y_t|\lambda_t) = \prod_{i=1}^{d} P(Y_{ti}|\lambda_{ti})$$

where λ_t= $[\lambda_{t1},..,\lambda_{td}]^T$. Next, we assume additionally that:

Assumption 1.3. $X_t = \log(\lambda_t) = [\log(\lambda_{t1}) ..., \log(\ \lambda_{td})\ ]^T$ is a latent random vector following multivariate normal distribution. Basically, evolution of X_t is described by a first order vector autoregressive VAR(1) model, that is the conditional probability density function (p.d.f) of X_t given X_{t-1} say $P_\theta(X_t|X_{t-1})$ is given.

$$P_\theta(X_t|X_{t-1}) : X_t - \mu = \Phi(X_{t-1} - \mu) + \epsilon_t \tag{2.2}$$

$$\epsilon_t \sim N(0, \Sigma)$$

where (ϵ_t) are independent and identically distributed (iid) sequence of a d-dimensional multivariate normal random vector with mean vector 0 and covariance matrix Σ, μ is the mean of X_t for $t = 1, .., T$. Note that VAR models are adequately described in Hansen and Johansen (1999), Lin and Tsay (1996).

https://doi.org/10.26219/heal.aueb.6542

It is worth mentioning that using X_t we manage to introduce cross-correlation and autocorrelation in Y_t.

For the matrix Φ we assume that it satisfies

$$\det(I - z\Phi) \neq 0, \quad \text{for all} \ \ |z| \leq 1, \ z \in C$$

where C is the set of complex numbers.

It turns out that X_t is a stationary and the marginal distribution of X_t is multivariate normal with mean vector μ and covariance matrix Γ, where Γ is the solution of

$$\Gamma = \Phi\Gamma\Phi^{-1} + \Sigma \tag{2.3}$$

according to the Yule-Walker relationship.

Finally, it is noted that by using assumptions 1.1-1.3, we are able to put all the parameters $\Theta = \{\mu, \Phi, \Sigma\}$ of the proposed model which is named State Space Multivariate Poisson model (SSMP).

2.1.1 Properties of the observed Y_t

In this section we will study some basic results for Y_t. First of all we note that marginalizing out X_t, the unconditional distribution of Y_t is expressed by:

$$\begin{aligned} P_\Theta(Y_t) &= \int P_\Theta(Y_t|X_t)P_\theta(X_t)dX_t \\ &= \int_{R_+^d} \prod_{i=1}^{d} \frac{\exp\left(-\exp\left(X_{ti}\right)\right)\exp\left(X_{ti}\right)^{Y_{ti}}}{Y_{ti}!} P_\theta(X_t)dX_t \end{aligned}$$

https://doi.org/10.26219/heal.aueb.6542

where $P_\theta(X_t)$ is the probability density function (pdf) of X_t i.e is a probability density function of multivariate Gaussian $N(\mu, \Gamma)$, where μ is the mean vector of X_t and Γ is the covariance matrix of Y_t, see equation (2.3).

It is revealed that the distribution of Y_t has not a close form. Nonetheless, the moment properties of Y_t have a closed form and are given below:

Proposition 1.1: The unconditional mean of Y_t is given by

$$E[Y_t] = [E(Y_{t1}), ..., E(Y_{td})]^T$$

where,

$$E(Y_{ti}) \equiv \alpha_i = \exp{(\mu_i + \frac{1}{2}\Gamma_{ii})}$$

Proof:
X_{ti} follows normal distribution with μ_i as mean and Γ_{ii} as variance, where Γ_{ii} is the ith diagonal element of Γ. Then the exponential function of X_{ti}, i.e $\exp{(X_{ti})}$=λ_{ti}, has a log-normal distribution with mean

$$E(\exp(X_{ti})) = \exp{(\mu_i + \frac{1}{2}\Gamma_{ii})} \quad \text{for } i = 1, .., d$$

Therefore,

$$E(Y_{ti}) = E(E(Y_{ti}|X_{ti})) = E(\exp{(X_{ti})}) \quad \text{for } i = 1, .., d.$$

It turns out that

$$E(Y_{ti}) = \exp{(\mu_i + \frac{1}{2}\Gamma_{ii})} \quad \text{for } i = 1, .., d.$$

Proposition 1.2: The unconditional covariance matrix of Y_t is denoted by

https://doi.org/10.26219/heal.aueb.6542

$Cov(Y_t)$ and its elements are given by

$$Var(Y_{ti}) = \alpha_i[1 + \alpha_i(\exp(\Gamma_{ii}) - 1)] \quad \text{for } i = 1, .., d,$$

$$Cov(Y_{ti}, Y_{tj}) = \alpha_i\alpha_j(\exp(\Gamma_{ij}) - 1) \quad \text{for } i \neq j \ i = 1, .., d, \ j = 1, .., d.$$

Proof:

$$\begin{aligned}
Var(Y_{ti}) &= E(Y_{ti}^2) - E[Y_{ti}]^2 \\
&= E[Var(\ Y_{ti}|X_{ti})\ + E(\ Y_{ti}|X_{ti})^2] - [E[E(\ Y_{ti}|X_{ti})\ ]]^2 \\
&= E(Var(Y_{ti}|X_{ti})) + (\ E(E(\ Y_{ti}|X_{ti})^2) - E(E(Y_{ti}|X_{ti}))^2) \\
&= E(\ Var(Y_{ti}|X_{ti})\ )\ + Var(\ E(Y_{ti}|X_{ti})\ ) \\
&= E(\ \exp(X_{ti})\ )\ + Var(\ \exp(\ X_{ti})\ ) \\
&= E(\ \exp(X_{ti})\ )\ + E(\ (\ \exp(\ X_{ti})\ )^2) - E(\exp(X_{ti}))^2 \qquad (2.4) \\
&= \exp(\ \mu_i + \frac{1}{2}\Gamma_{ii})\ + \exp(\ 2\mu_i + 2\Gamma_{ii})\ - \exp(\ 2\mu_i + \Gamma_{ii}) \\
&= \exp(\ \mu_i + \frac{1}{2}\Gamma_{ii})\ + \exp(\ 2\mu_i + \Gamma_{ii})\ (\ \exp(\ \Gamma_{ii}) - 1) \\
&= \exp(\ \mu_i + \frac{1}{2}\Gamma_{ii})\ (1 + \exp(\ \mu_i + \frac{1}{2}\Gamma_{ii})\ (\exp(\ \Gamma_{ii})\ - 1)\ ) \\
&= \alpha_i[1 + \alpha_i(\ \exp(\Gamma_{ii})\ - 1)\ ]
\end{aligned}$$

$$\begin{aligned}
Cov(Y_{ti}, Y_{tj}) &= E(Y_{ti}Y_{tj}) - E(Y_{ti})E(Y_{tj}) \\
&= E(\ \exp(X_{ti} + X_{tj})) - E(exp(X_{ti}))E(exp(X_{tj})) \\
&= E(\exp(\mu_i + \mu_j + \frac{1}{2}\Gamma_{ii} + \frac{1}{2}\Gamma_{jj} + \Gamma_{ij})) - \exp(\mu_i + \mu_j + \frac{1}{2}\Gamma_{ii} + \frac{1}{2}\Gamma_{jj}) \\
&= \exp(\mu_i + \mu_j + \frac{1}{2}\Gamma_{ii} + \frac{1}{2}\Gamma_{jj})(\exp(\Gamma_{ij}) - 1) \\
&= \exp(\mu_i + \frac{1}{2}\Gamma_{ii})\exp(\mu_j + \frac{1}{2}\Gamma_{jj})(\exp(\Gamma_{ij}) - 1) \\
&= \alpha_i\alpha_j(\exp(\Gamma_{ij}) - 1).
\end{aligned}$$

https://doi.org/10.26219/heal.aueb.6542

From the above, immediately follows that, the correlation between Y_{ti} and Y_{tj}, for $i \neq j$, $i = 1, .., d, j = 1, ..., d$ is

$$Corr(Y_{ti}, Y_{tj}) = \frac{\exp(\Gamma_{ii}) - 1}{\sqrt{[\alpha_i^{-1} + \exp(\Gamma_{ii}) - 1]}\sqrt{[\alpha_j^{-1} + \exp(\Gamma_{jj}) - 1]}}$$

where Γ_{ij} is the (i, j) component of Γ.

The above formula shows that the cross-correlation structure of Y_t depends on Γ_{ij} as well, and so can take either positive or negative values. This is a very important feature of this model, as many models can only support limited positives correlation. However, as we said it is common for autocorrelated multivariate counts to show either a positive or negative correlation, so it is important for a model to be able to describe both negative and positive correlation.

Also, we take that, the unconditional variance of Y_{ti} for every dimension of SSMP exhibits overdispersion and the amount of this overdispersion increases with Γ_{ii} and α_{ii}. Because, with respect to the SSMP model the variance of Y_{ti} is λ_{ti}=$\exp(X_{ti})$ but, it has been shown in (2.4) that $Var(Y_{ti})$= $\alpha_i[1+\alpha_i(\exp(\Gamma_{ii})-1)]$ so, the larger the α_i and Γ_{ii}, the greater the difference in variance of Y_{ti} with respect to the model with the variance of Y_{ti} that we will observe.

Proposition 1.3: The unconditional autocorrelation of Y_t is

$$Corr(Y_{ti}, Y_{(t-\tau)j}) = \frac{E(\exp(X_{ti} + X_{(t-\tau)j})) - \alpha_i\alpha_j}{\alpha_i\alpha_j\sqrt{[\alpha_i^{-1} + \exp(\Gamma_{ii}) - 1][\alpha_j^{-1} + \exp(\Gamma_{jj}) - 1]}}$$

for τ= $0, .., i,\ i, j = 1, ..., d$.

https://doi.org/10.26219/heal.aueb.6542

Proof:

We calculate first

$$\begin{aligned} Cov(Y_{ti}, Y_{(t-\tau)j}) &= Cov(\exp{(X_{ti})}, \exp{(X_{(t-\tau)j})}) \\ &= E(\exp{(X_{ti} + X_{(t-\tau)j)})} \\ &\{E(\exp{(X_{ti} - \mu_i + X_{(t-\tau)j)-\mu_j})} - \exp{(\frac{1}{2}\Gamma_{ii} + \frac{1}{2}\Gamma_{jj})} \end{aligned} \tag{2.5}$$

Since,

$$\begin{aligned} X_t - \mu &= \Phi(\Phi(X_{t-2} - \mu) + \epsilon_{t-1}) + \epsilon_t \\ &= \Phi^t(X_{t-1} - \mu) + \Phi^{t-1}\epsilon_{t-\tau+1} + \Phi^1\epsilon_{t-1} + \epsilon_t \end{aligned}$$

Define $\Phi_i^{\tau-k}$=$[\Phi_{i,1}^{\tau-k}, ..., \Phi_{i,d}^{\tau-k}]$ as a row vector which represents the ith row of $\Phi^{\tau-k}$. Then

$$X_{ti} + X_{(t\tau)j} - \mu_i - \mu_j = \Phi_i^{\tau}(X_{t-\tau} - \mu) + \Phi_i^{\tau-1}\epsilon_{t-\tau+1} + ... + \Phi_i^1\epsilon_{t-1} + e_i\epsilon_t + X_{t-\tau}$$

Define $\Phi_i^{*\tau}$=$[\Phi_{i,1}^{\tau}, ..., \Phi_{i,j}^{\tau} + 1, ..., \Phi_{i,d}^{\tau}]$, Then,

$$X_{ti} + X_{(t\tau)j} - \mu_i - \mu_j = \Phi_i^{*\tau}(X_{t-\tau} - \mu) + \Phi_i^{*\tau-1}\epsilon_{t-\tau+1} + ... + \Phi_i^{*1}\epsilon_{t-1} + e_i\epsilon_t$$

Consequently,

$$\begin{aligned} &E[\exp{(X_{ti} + X_{(t\tau)j} - \mu_i - \mu_j)}] = \\ &E[\exp{(\Phi_i^{*\tau}(X_{t-\tau-\mu}))}]E[\exp{(\Phi_i^{*\tau-1}\epsilon_{t-\tau+1})}]...E[\exp{(\Phi_i^{*1}\epsilon_{t-1})}]E[\exp{(\epsilon_{i,t})}] \end{aligned}$$

According to the moment-generating function of the multivariate normal distribution, if X $\sim N(\mu, \Sigma)$, then the moment-generating function is given

https://doi.org/10.26219/heal.aueb.6542

by $m_x(t)$=$E[\exp(t'X)]$,

$$m_x(t) = E[\exp^{t^T x}] = \exp^{t^T \mu + \frac{1}{2} t^T \sum t}$$

$$E[\exp(\Phi_i^{\tau-1} \epsilon_{t-\tau+1})] = \exp(\frac{1}{2} \Phi_i^{\tau-1}{}' \Sigma \Phi_i^{\tau-1}) \quad \text{for all } t, \tau\text{=1,..,T.}$$

Then, because $X_t \sim$ N(μ,Γ), we obtain

$$E[\exp(\Phi_i^{*\tau}(X_{t-\tau} - \mu))] = \exp(\frac{1}{2} \Phi_i^{*\tau}{}' \Gamma \Phi_i^{*\tau}) \quad \text{for all } t, \tau\text{=1,..,T}$$

So,

$$\begin{aligned} E[\exp(X_{ti} + X_{(t\tau)j} - \mu_i - \mu_j)] = \exp(&\frac{1}{2}\Phi_i^{*\tau\prime}\Gamma(0) + \frac{1}{2}\Phi_i^{\tau-1\prime}\Sigma\Phi_i^{\tau-1} + ... \\ &+ \frac{1}{2}\Phi_i^{1\prime}\Sigma\Phi_i^1 + \frac{1}{2}\Sigma_{ii}) \end{aligned} \tag{2.6}$$

Plug (2.5) into (2.6). Finally, we have the covariance of Y_{ti} and $Y_{(t-\tau)j}$. Then the autocorrelation could be defined as

$$\begin{aligned} Corr(Y_{ti}, Y_{(t-\tau)j}) &= \frac{Cov(Y_{ti}, Y_{(t-\tau)j})}{\sqrt{Var(Y_{ti})Var(Y_{(t-\tau)j})}} \\ &= \frac{E(\exp(X_{ti} + X_{(t-\tau)j})) - \alpha_i\alpha_j}{\alpha_i\alpha_j\sqrt{[\alpha_i^{-1} + \exp(\Gamma_{ii}) - 1][\alpha_j^{-1} + \exp(\Gamma_{jj}) - 1]}} \\ &= \frac{E(\exp(X_{ti} + X_{(t-\tau)j}))\{E(\exp(X_{ti} - \mu_i + X_{(t-\tau)j)-\mu_j}) - \exp(\frac{1}{2}\Gamma_{ii} + \frac{1}{2}\Gamma_{jj})}{\alpha_i\alpha_j\sqrt{[\alpha_i^{-1} + \exp(\Gamma_{ii}) - 1][\alpha_j^{-1} + \exp(\Gamma_{jj}) - 1]}} \end{aligned}$$

where, $E(\exp(X_{ti} - \mu_i + X_{(t-\tau j)} - \mu_j)$ is defined in (2.6).

https://doi.org/10.26219/heal.aueb.6542

A very important feature of the model is that for small α_i,α_j, the model can hardly describe both situations with moderate or strong negative cross-correlations and autocorrelation. However, if α_i, α_j increase, the model is able to describe larger negative cross-correlation and autocorrelation.

Finally, from Propositions 1.1-1.3 it becomes clear that Y_{ti} is stationary, because the mean, the variance and the autocorrelation of Y_{ti} are independent of the time and consequently the Y_t is stationary.

Therefore, summarizing the above we conclude that the proposed model can be applied to stationary multivariate time series of count which have either positive or negative cross-correlation and autocorrelation and can take overdispersion into account.

2.1.2 Parameter Estimation

First, we want to estimate X_{t+1} based on the previous and current observations $\{Y_{1:t}\}$, i.e, $P_\theta(X_{t+1}|Y_{1:t})$ and make conclusions for X_t based on the total observations $\{Y_{1:T}\}$, i.e., $P_\theta(X_{tj}|Y_{1:T})$ for $t = 1, .., T$. The posterior distribution $P_\theta(X_t|Y_{1:t})$ is not available in closed form. Thus, we will turn to numerical methods of estimation and prediction. More specifically, to achieve the above we will assume that Θ is known, as defined the Θ are all parameters of model and we will use particle filtering (PF) or particle smoothing (PS) methods. A detailed description of particle filtering can be found in Djuric et al. (2003). Generally, the basic idea of PF method is to compute $P_\Theta(X_t|Y_{1:t})$ using a set with N_p samples that are called particles. Each particle has its associated weight (w_t^i), for $i = 1, .., N_p$ which is proportional to its probability of being sampled from the posterior distribution. Using the

https://doi.org/10.26219/heal.aueb.6542

above, the following relation emerges

$$P(X_t|Y_{1:t}) \approx \sum_{i=1}^{N_p} W_t^i \delta(X_t - x_t^i) \tag{2.7}$$

where δ is the Dirac delta function and W_t^i is the normalized weights.

Basically, we want to obtain the prediction distribution $P_\Theta(X_{t+1}|Y_{1:t})$. To be able to get this distribution, we will generate samples of X_{t+1} from $P_\Theta(X_{t+1}|Y_{1:t})$ using the equation (2.7). Distinctively, we will calculate $X_{t+1}^i = \mu + \Phi(X_t^i - \mu) + \epsilon_{t+1}^i$, see (2.2), where, $\epsilon_{t+1}^i \sim N(0, \Sigma)$ and then we will take the calculation

$$P(X_{t+1}|Y_{1:t}) \approx \sum_{i=1}^{N_p} W_t^i \delta(X_{t+1} - x_{t+1}^i).$$

Note, that the convergence of the approximated distribution by PF to the true $P(X_t|Y_{1:t})$ is guaranteed by the central limit theorem, see Li and Wang (2008). In addition, the PF method is computationally easy. Nevertheless, the distribution of the particles' weights becomes more and more skewed as t increases, this is a problem for PF method, but can be solved. Particularly, we calculate the effective sample size (ESS) which is given by

$$ESS = (\sum_{i=1} N_p (W_t^i)^2)^{-1}$$

and if ESS is smaller than a pre-specified threshold δ we resample. The algorithm of implement the PF method is presented below.

Algorithm for Particle Filtering (PF)

At time $t = 1$

 https://doi.org/10.26219/heal.aueb.6542

1.Initialization: Obtain sample $x_1^i \sim P_0(X_t)$, for i = 1,..,N_p.

2.Compute the weights $w_1^i = p(Y_1|x_1^i)$, for i = 1,..,N_p and normalize the weights $W_1^i = \frac{w_1^i}{\sum_{i=1}^{N_p} w_1^i}$, for i = 1,..,$N_p$.

3.Calculate the filtered distribution $P(X_1|Y_1) \approx \sum_{i=1}^{N_p} W_1^i \delta(X_1 - x_1^i)$.

At time t $\geq$ 2

4.Sample $X_t^i \sim P_\theta(X_t|x_{t-1}^i)$, for i = 1,..,$N_p$.

5.Compute the weights $w_t^i = W_{t-1}^i p(Y_t|x_t^i)$, for i = 1,..,$N_p$ and normalize the weights $W_t^i = W_{t-1}^i \frac{w_t{}^i}{\sum_{i=1}^{N_p} w_t^i}$.

6.Calculate the filtered distribution $P(X_t|Y_{1:t}) \approx \sum_{i=1}^{N_p} W_t^i \delta(X_t - x_t^i)$.

7.If the resample criterion is satisfied, i.e., ESS = $(\sum_{i=1} N_p (W_t^i)^2)^{-1} < \delta$, then we resample from the set $\{(W_t^i, x_t^i), i = 1, .., N_p\}$ with the probabilities $p(x_t^j = x_t^i) = W_t^i, i = 1, ..., N_p$, and replace the previous set $\{(W_t^i, x_t^i), i = 1, ..., N_p\}$ by $\{(\frac{1}{N_p}, x_t^j), j = 1, .., N_p\}$.

8.Terminate when $t = T$, otherwise $t = t + 1$, and go back to 4.

Algorithm for Particle smoothing (PS):

1. Start by setting $W_{T|T}^i$=W_T^i for $i = 1, .., N_p$.

https://doi.org/10.26219/heal.aueb.6542

2. For each $t = T-1, ..., 1$, compute the smoothed weights by

$$W^i_{t|T} = W^i_t \sum_{j=1}^{N_p} \frac{W^j_{t+1|T} P_\Theta(x^j_{t+1}|x^i_t)}{\sum_{l=1}^{N_p} W^l_t P_\Theta(x^j_{t+1}|x^l_t)} \quad i = 1, .., N_p.$$

3. Calculate the smoothed distribution $P_\theta(X_t|Y_{1:T}) \approx \sum_{i=1}^{N_p} W^i_{t|T} \delta(X_t - x^i_t)$ for $t = 1, .., T$.

Based on the above algorithms, we want to estimate the parameters Θ of SSMP. The estimation of these parameters will be performed using the Expectation-Maximization (EM) algorithm, see Laird (1993). This is a tool to find maximum-likelihood estimates for parameters of a model when the data are incomplete, i.e we have missing data points, or latent variables as in our data. The basic idea is, instead of maximizing the likelihood which contains only the observed data, to maximize a likelihood which contains both observed and unobserved data because it is easier to maximize. At E-step we estimate the unobserved data using the current likelihood which is denoted by $Q(\Theta, \Theta^{(q)})$ and contains observed data and at M-step we maximize a likelihood that contains the observed and the estimated data. A detailed description of the EM algorithm can be found in Ng et al. (2012). In our model the observed data are the Y_t and the unobserved are the X_t (latent variable). During the E-step we calculate

$$\begin{aligned} Q(\Theta, \Theta^{(q)}) &= \int P_{\Theta^{(q)}}(X_{1:\tau}|Y_{1:\tau}) \log P_\Theta(X_{1:\tau}, Y_{1:\tau}) dX_{1:\tau} \\ &= \sum_{t=1}^{\tau-1} \iint P_{\Theta^{(q)}}(X_t, X_{t+1}|Y_{1:}) \log_{P_\Theta}(X_{t+1}|X_t) dX_t dX_{t+1}. \end{aligned}$$

At M-step we want to find the revised parameter estimates $\Theta^{(q+1)}$ that maximize the function

$$\Theta^{(q+1)} = \arg\max\nolimits_\theta Q(\Theta, \Theta^{(q)}). \tag{2.8}$$

 https://doi.org/10.26219/heal.aueb.6542

The problem in our case is that we do not have the analytical form of

$$P_{\Theta^{(q)}}(X_t, X_{t+1}|Y_{1:\tau}),$$

but we can calculate it using PF and PC method. So the Monte Carlo EM (MCEM) algorithm is used for parameter estimation instead of the classic EM algorithm i.e a modification of the EM algorithm. See Neath (2013) for more information about the Monte Carlo EM algorithm. The Monte Carlo EM (MCEM) algorithm is given below.

Monte Carlo EM algorithm (MCEM):

1. Set $q = 0$ and initialize $\theta^{(q)} = \theta^{(0)}$ such that $\log_{P_{\Theta(q)}}(Y_{1:T})$ is finite.

2. Expectation (E) Step:
Run Algorithm PF and PC to obtain the filtered and smoothed distributions for $t = 1, ..., T$.

Calculate $Q(\theta, \Theta^{(q)})$

Where, $Q(\theta, \Theta^{(q)}) \approx \sum_{t=1}^{T-1} \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} W^{ij}_{t,t+1|T} \log(P_\theta(X^j_{t+1}|X^i_t))$,
and $W^{ij}_{t,t+1|T} = W^i_t \frac{W^i_{t+1|T} P_\Theta(q)}{i(x^j_{t+1}|x^i_t)} \sum_{i=1} N_p W^i_t P_\Theta(q)(x_{t+1^j}|x^i_t)$.

3.Compute $\Theta^{(q+1)}$=$\{\mu^{(q+1)}, \Phi^{(q+1)}, \Sigma^{q+1}\}$ using the following equations
$\Pi^{(q+1)}$=$[(I - \Phi^{(q+1)})\mu^{(q+1)}, \Phi^{(q+1)}]$=
$(\sum_{t=1}^{T-1} \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} W^{ij}_{t,t+1|T} x^{ij}_{t+1} Z^{i'}_t)$ $(\sum_{t=1}^{T-1} \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} W^{ij}_{t,t+1|T} Z^i_t Z^{i'}_t)^{-1}$
and,
$\Sigma^{(q+1)}$=$\frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} W^{ij}_{t,t+1|T}(x^j_{t+1} - \Pi^{(q+1)} Z^i_t)(x^j_{t+1} - \Pi^{(q+1)} Z^i_t)'$
where, $Z^i_t = [1, x^i_t]$.

https://doi.org/10.26219/heal.aueb.6542

4.Check the non-termination condition $Q(\Theta^{q+1}, \Theta^q) - Q(\Theta^q, \Theta^q) \geq \epsilon$ for some $\epsilon > 0$. If satisfied, update $q \to q + 1$, and return to 2, otherwise terminate.

More information for the above algorithms one can find in the original article Zhang et al. (2017).

2.1.3 Conclusions

As mentioned above, although autocorrelated multivariate counts data are frequently used, there are not enough models to describe them. In addition, several of the existing models are not flexible to describe cross-correlation and autocorrelation structures. In this section the model which was studied is a theoretical model which can be used from stationary multivariate count time series and not only flexible to describe cross-correlation (positive or negative) and autocorrelation structures but can also take overdispersion into account. Consequently, this model can be very useful in multivariate count series analysis.

2.2 Modelling non-stationary multivariate time series of counts via common factors

In this section, we will introduce a parameter-driven model to describe multivariate time series of counts so that we can study its cross correlation and autocorrellation. This model is presented and developed in Wang and Wang (2018).

First, we introduce $N_t = (N_{t1}, ..., N_{td})$ where $\{N_{tj}, t = 1, ..., T, j = 1, ..., d\}$ is a sequence of independent Poisson processes of unit intensity. For each Y_{tj} and N_{tj} we make the following assumption.

https://doi.org/10.26219/heal.aueb.6542

Assumption 2.1:
For each t and j, there is a positive (not necessarily stationary) process λ_{tj}, with finite variance and independent of N_t which counts the number of events of N_{tj} in the time interval $[0, \lambda_{tj}]$ i.e Y_{tj}=$N_{tj}(\lambda_{tj})$.

From Assumption 2.1 it turns out that Y_{tj} is conditionally Poisson distributed for given λ_{tj} i.e

$$Y_{tj}|\lambda_{tj} \sim Poisson(\lambda_{tj}) \tag{2.9}$$

and that Y_{tj} is conditionally independent of each other given $\Lambda_t = [\lambda_{t1}, ..., \lambda_{td}]^T$.

For the cross-covariance of the observed count Y_t for each t it holds that,

$$\begin{aligned}
Cov(Y_{ti}, Y_{t'j}) &= E[Y_{ti}Y_{t'j}] - E[Y_{ti}]E[Y_{t'j}] \\
&= E[E[Y_{ti}Y'_{tj}|\lambda_t]] - E[E[Y_{ti}|\lambda_t]]E[E[Y_{t'j}|\lambda_{t'}]] \\
&= E[Cov(Y_{ti}, Y_{t'j}|\lambda_t) - E[Y_{ti}|\lambda_t]E[Y_{t'j}|\lambda_t]] - E[E[Y_{ti}|\lambda_t]]E[E[Y_{t'J}|\lambda_{t'}]] \\
&= E[Cov(Y_{ti}, Y_{t'j}|\lambda_{t'}] + Cov(E[Y_{ti}|\lambda_t], E[Y_{t'j}|\lambda_{t'}]) \\
&= Cov(E[Y_{ti}|\lambda_t], E[Y_{t'j}|\lambda_{t'}]) \\
&= Cov(\lambda_{ti}, \lambda_{t'j}) \quad \text{ for any } t,\ t' \ \text{ and } \ i \neq j.
\end{aligned}$$

Because,

$$E(Y_{ti}) = E[E[Y_{ti}|\lambda_{ti}]] = E[\lambda_{ti}],$$

We obtain that,

$$Cov(Y_{ti}, Y_{t'j}) = Cov(\lambda_{ti}, \lambda_{t'j})$$

for any t, t' $i \neq j$. Basically, we prove that the unobserved d-dimensional process Λ_t preserves the cross-covariance of the observed count Y_t. So the autocorrelation $Corr(Y_{ti}, Y_{t'j})$ has the following form

 https://doi.org/10.26219/heal.aueb.6542

$$
\begin{aligned}
Corr(Y_{ti}, Y_{t'j}) &= \frac{Cov(Y_{ti}, Y_{t'j})}{\sqrt{Var(Y_{ti})Var(Y_{t'j})}} \\
&= \frac{Cov(\lambda_{ti}, \lambda_{t'j})}{\sqrt{Var(Y_{ti})Var(Y_{t'j})}},
\end{aligned}
\tag{2.10}
$$

where λ_{tj} is positive for all $j = 1, .., d$ and $t = 1, ..., T$.
Also, using the law of total variance it immediately appears that $Var(Y_{tj}) = E[\lambda_{tj}] + Var[\lambda_{tj}]$ and then, turn out that $Corr(Y_{ti}, Y_{t'j}) \leq Corr(\lambda_{ti}, \lambda_{t'j})$.

Next, we model the dependence structure of Y_t i.e the λ_t through latent positive stationary process ϵ_t=$[\epsilon_{t1}, ..., \epsilon_{td}]^T$. More specifically we assume that,

Assumption 2.2:
Conditioning on ϵ_t, λ_{tj}, for $j = 1, ..., d$ are independent of each other over time and across components.

Assumption 2.3:
The conditional mean of λ_{tj} given ϵ_{tj} is given by $E[\lambda_{tj}|\epsilon_t] \equiv \mu_{tj}$, i.e

$$\mu_{tj} \equiv E[\lambda_{tj}|\epsilon_t] = \alpha_{tj}\epsilon_{tj}$$

where,

$$\alpha_{tj} = E[\lambda_{tj}]$$

then $E[\epsilon_{tj}] = 1$.

By using the above we obtain that the covariance between λ_{ti} and λ_{tj} for

https://doi.org/10.26219/heal.aueb.6542

all $i \neq j$, $i, j = 1, ..., d$ is given by,

$$\begin{aligned} Cov(\lambda_{ti}, \lambda_{t'j}) &= E(Cov(\lambda_{ti}, \lambda_{t'j}|\epsilon_t)) + Cov(E(\lambda_{ti}|\epsilon_t), E(\lambda_{t'j}|\epsilon_{t'})) \\ &= Cov(E(\lambda_{ti}|\epsilon_t), E(\lambda_{t'j}|\epsilon_t)) \\ &= Cov(\alpha_{ti}\epsilon_{ti}, \alpha_{t'j}\epsilon_{t'j}) \\ &= \alpha_{ti}\alpha_{t'j}(cov(\epsilon_{ti}, \epsilon_{t'j})), \end{aligned} \tag{2.11}$$

for either:
(a) $t = t^{'}$ and $j, j^{'}$ or (b) $j = j^{'}$ and $t, t^{'}$.

Due to the (2.9) and (2.10) it seems that α_{tj}, $j = 1, ..., d$ can be designed to pick up time trends, seasonal effects, business cycles, day of the week effects, diurnal pattern and so forth. Also, from (2.11) we can obtain that ϵ_t accounts for the possible overdispersion and autocorrelation as well as the cross correlation. Hence, it makes sense to describe/model both α_{tj} and ϵ_{tj}. Regarding α_{tj}, we link them with q-dimensional covariates X_t in the way presented below.

$$\log(\alpha_{tj}) = \beta_j^T X_t \quad for\ j = 1, .., d, t = 1, ..., T, \tag{2.12}$$

where β_j^T is a is a $q \times 1$ vector.

As for the ϵ_t, we will not model it, but we will try to find lower dimensional factors that will contain the same information with it. That is, we will try to represent by,

$$\epsilon_t = Af_t \tag{2.13}$$

where $f_t = (f_{t1}, ..., f_{tr})$ is an $r \times 1$ vector of latent factors, and $A = (\alpha_1, ..., \alpha_r)$ is a $d \times r$ time invariant factor loading matrix of rank r $(r \times d)$. The reason we will not proceed with the modeling of the ϵ_t is the challenges involved in this task.

https://doi.org/10.26219/heal.aueb.6542

In conclusion, in this section we modeled each univariate time series Y_{tj} as a doubly stochastic process and we were able to describe correlation and autocorrelation of Y_t.

2.2.1 Parameter Estimation

From the study of the model defined in the previous section it appears that the parameters to be estimated are the β_j for $j = 1, .., d$ and the loading matrix A.

We start with the estimation of β_j for $j = 1, .., d$. First of all, we consider that $B = [\beta_1, ..., \beta_d]$ be a $q \times d$ matrix of coefficients, and b be its vectorization, i.e. $b = [\beta_1^T, ..., \beta_d^T]^T$. Also, let B_0 =$[\beta_{01}, ..., \beta_{0d}]$ is the true value of B with corresponding vectorization b_0. To obtain an estimator B' (b') of B_0 (b_0) we will use Pseudo-maximum-likelihood method (PML). This is a possible way of obtaining an estimator without resorting to non-parametric regression. Specifically, we obtain an estimator by maximizing a likelihood function that is associated with a family of probability distributions that does not necessarily contain the true distribution. In our case we maximize the log-likelihood function of b ($\log b$) which based on Poisson family and is given by

$$\log(b) = (dn)^{-1} \sum_{j=1}^{d} \sum_{t=1}^{T} (Y_{tj}\beta_j^T X_t - \exp{(\beta_j^T X_t)}). \tag{2.14}$$

It is observed that the (2.14) is the true rescaled log-likelihood function if Y_{tj} follow a Poisson distribution with mean $\exp{(\beta_j^T X_t)}$ and are independent within each time series and across a different time series. By maximizing the equation (2.13) we will get estimators for B_0. However, just calculating them is not enough, as we want them to meet certain conditions. Basically, we want to be consistent and converge in the normal distribution. To achieve this we should be make the following assumptions:

Assumption 2.4:
The latent factor f_t is stationary and strongly mixing with mixing coefficient $\alpha(m)$ satisfying $\sum_{m=1}^{\infty} \alpha(m)^{\lambda/(\lambda+2)} \leq \infty$ for some $\lambda \leq 0$ and $E[|f_{tj}|^{2(\lambda+2)}]/ < \infty$ for all j.

Assumption 2.5:
For each $j = 1, .., p$
(a) the limit of $T^{-1} \sum_{t=1}^{T} X_t X_t^T \exp(\beta_{0j}^T X_t)$ exists, and it is invertible, denoted by Ω_1^i
(b) the limit of $T^{-1} \sum_{t=1}^{T} X_t X_t^T E(Var(\lambda_{tj}|\epsilon_t))$ exists, and it is invertible, denoted by Ω_2^i
(c)$T^{-1} \sum_{t=1}^{T} X_{t+k} X_t^T exp(\beta_{0j}^T X_{t+k} + \beta_{0j'}^T C_t)$ converges to $W_k^{(j,j')}$ uniformly in $|k| < n$ as $n \to \infty$, and the limit $W_k^{(j,j')}$ fulfils the property that $\Omega_3^{(j,j')} = \sum_{k=-\infty}^{\infty} W_k^{(j,j')} \sum_{\epsilon}^{(j,j')(k)} < \infty$
(d)$sup_{n\geqslant 1} n^{-1} \sum_{t=1}^{n} |X_{ti} X_{tj} X_{tk}| \exp(\beta_j^T X_t) \leq \infty$ for any $i, j, k \in \{1, .., q\}$ and $b \in \Theta$ where X_{ti} is the ith elements of X_t

Assumption 2.6:
$\max_{1 \leq t \leq T} n^{-1/6} ||X_t|| = (1)$ and $\max_{1 \leq t \leq T, 1 \leq j \leq p} E[\lambda_{tj}^3 | \epsilon_t] = O_p(1)$

Theorem 1:
Suppose that Assumptions 2.4-2.6 hold. Then,
1. estimator which was obtained by PLM (b') is consistent estimator of b_0,
2. $n^{1/2}$ $(b' - b_0)$ converges to a normal distribution with mean equal to zero and covariance matrix equal to $\Omega_1^{-1} + \Omega_1^{-1}(\Omega_2 + \Omega_3)\Omega_1^{-1}$, and that
3. the mean-squared-error matrix $E[(b' - b_0)(b' - b_0)^T]$ to order T^{-1} is given

https://doi.org/10.26219/heal.aueb.6542

by,

$$E[(b^{'} - b_0)(b^{'} - b_0)^T] = n^{-2} H_1^{-1} \{ \sum_{t,t^{-1}}^{T} Cov(Y_t, Y_{t'} X_t X_{t'}^T) \} H_1^{-1} \tag{2.15}$$

where, H_1=$-T^{-1} \sum_{t=1}^{T} \text{diag}(\exp{(\beta_{01}^T X_t)}, ..., \exp{(\beta_{0d}^T X_t)})\ X_t X_t^T$.
If ν is a vector then $D = \text{diag}(\nu)$ returns a square diagonal matrix with the elements of vector v on the main diagonal.

Thus the estimator is, under certain assumptions, consistent and converges to a normal distribution.

Since we have managed to estimate b_0, we will move on to estimating the loading matrix A given a consistent estimator of b_0 or B_0. First we set that

$$\Sigma_\epsilon(k) = Cov(\epsilon_{t+k}, \epsilon_t) \tag{2.16}$$

$$\Sigma_f(k) = Cov(f_{t+k}, f_t). \tag{2.17}$$

Obviously, using (2.15),(2.16) turns out that

$$\Sigma_\epsilon(k) = A \Sigma_f(k) A^T \quad \text{for } k \in Z.$$

Instead of working with ϵ_t to estimate A we will work with η_t=$[\eta_{t1}, ..., \eta_{td}]$, where η_{ti} is such that

$$\eta_{ti} = \frac{Y_{ti}}{E[Y_{ti}]} = \frac{Y_{ti}}{E[\lambda_{ti}]} = \frac{Y_{ti}}{\exp{(\beta_{0i}^T X_t)}} = Y_{ti} \exp{(-\beta_{0i}^T X_t)}.$$

The covariance of η_{ti} and η_{tj} for $i = j$ is given by,

$$\begin{aligned} Cov(\eta_{ti}, \eta_{tj}) &= Cov(Y_{ti} \exp(-\beta_{0i}^T X_t), Y_{ti} \exp(-\beta_{0j}^T X_t)) \\ &= \exp(-\beta_{0i}^T X_t) \exp(-\beta_{0j}^T X_t) Cov(Y_{ti}, Y_{tj}) \\ &= (-\alpha_{ti})(-\alpha_{tj}) Cov(Y_{ti}, Y_{tj}) \\ &= \alpha_{ti}\alpha_{tj} Cov(Y_{ti}, Y_{tj}) \\ &= Cov(\epsilon_{ti}, \epsilon_{tj}). \end{aligned} \tag{2.18}$$

It also holds that,

$$\Sigma_\epsilon(k) = Cov(\eta_{t+k}, \eta_t).$$

So, from (2.17), (2.18) we obtain that

$$Cov(\eta_{t+k}, \eta_t) = A\Sigma_f(k)A^T \qquad k \neq 0. \tag{2.19}$$

The reason, we introduce the η_t is that the ϵ_t cannot be calibrated from the data, so we have to find something else that will be easier to use than ϵ_t but will have the same information with it. This is exactly what the η_t does. Also, η_t has the added property that it does not need to be stationary. For any H invertible matrix we have that,

$$\epsilon_t = Af_t = AH(H^{-1}f_t).$$

So we observed that the loading matrix A is not unique. Such a thing obviously does not apply to the linear space that is spanned by the columns of A. This space is defined as $M(A)$. If $\Sigma_f(k)$ is invertible, $M(A)$ is the orthogonal complement of the linear space that is spanned by the eigenvectors of $\Sigma_\eta(k)$ corresponding to the zero eigenvalues.

Define

$$L = \sum_{k=1}^{k_0} \Sigma_\eta(k)\Sigma_\eta(k)^T \tag{2.20}$$

https://doi.org/10.26219/heal.aueb.6542

where k_0 is a prescribed fixed positive integer. Under the following assumption $M(A)$ is also the orthogonal complement of the linear space that is spanned by the eigenvectors of L corresponding to the zero eigenvalues.

Assumption 2.7:
The $r \times r$ matrix $\Sigma_f(k)$ is invertible, for $k = 1, ..., k_0$.

Assumption 2.8:
The eigenvalues of the $d \times d$ matrix L, λ_1,..,λ_d satisfy λ_1 >...> $\lambda_r > \lambda_{r+1}$ =...=λ_d=0.
The loading matrix A is made up of the orthonormal eigenvectors corresponding to the positive eigenvalues of L. A brief description of the loading matrix estimate was introduced. All the assumptions for the estimation of the loading matrix A as well as additional information on exactly how A arises are in the original article Wang and Wang (2018).

2.2.2 Conclusions

In this section the model studied is a theoretical model which can be used from multivariate time series of counts. The time series from which it can be used does not need to be stationary. Finally, the conditional distribution of λ_{tj} is left unspecified. This makes the model quite general so it can be used by many multivariate time series of counts.

2.3 Sequential Bayesian Analysis of Multivariate Count Data

As above, in this section, a State Space model to describe multivariate time series of counts Y_t is introduced, so that we can study their cross correlation and autocorrelation. We obtain this model from Aktekin et al. (2018). Want-

 https://doi.org/10.26219/heal.aueb.6542

ing to develop this model, we will introduce the static parameters, $\lambda_j's$ and the dynamic environment θ_t, where, λ_j is the rate specific to the j'th series and θ_t represents the effects of the random common environment modulating λ_j. So we obtain a Bayesian hierarchical model of the form,

$$(Y_{tj}|\lambda_j, \theta_t) \sim Poisson(\lambda_j\theta_t) \quad \text{for} \quad j = 1, ..., d \text{ and } t = 1, ..., T. \tag{2.21}$$

Next we make the following assumptions:

Assumption 3.1:
Given $\theta_t's$ and λ_j, $Y_{tj}'s$ are conditionally independent over time.

Assumption 3.2:
At time t, given θ_t and $\lambda_j's$, $Y_{tj}'s$ are conditionally independent of each other.

Alongside, for θ_t, following Smith and Miller (1986) we assumed that

$$\theta_t = \frac{\theta_{t-1}}{\gamma}\epsilon_t \tag{2.22}$$

where, ϵ_t is the error term and γ is the discount factor which is common for all d series. For the ϵ_t we make the following assumption:

Assumption 3.3:
ϵ_t given D^{t-1}=$D^{t-2}, Y_{1,t-1}, .., Y_{j,t-1}$ represents the sequential arrival of data and λ=$(\lambda_1, ..., \lambda_d)$ follows a Beta distribution with parameters $\gamma\alpha_{t-1}$ and $(1-\gamma)\alpha_{t-1}$ i.e,

$$(\epsilon_t|D^{t-1}, \lambda_1, ..., \lambda_d) \sim Beta[\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}] \tag{2.23}$$

with, $\alpha_{t-1} > 0$, $0 < \gamma < 1$ and D^{t-1}=$\{D^{t-2}, Y_{t-1,1}, .., Y_{t-1,d}\}$, so D^{t-1} represents the sequential arrival of data.

https://doi.org/10.26219/heal.aueb.6542

From Assumption 3.3 immediately follows that,

$$E(\frac{\epsilon_t}{\gamma}) = \frac{1}{\gamma}\frac{\gamma\alpha_{t-1}}{(1-\gamma)\alpha_{t-1}+\gamma\alpha_{t-1}} = \frac{1}{\gamma}\frac{\gamma\alpha_{t-1}}{\alpha_{t-1}} = \frac{\gamma}{\gamma} = 1.$$

Studying equation (2.22) it becomes obvious that θ_t has a Markovian evolution, since to calculate θ_t we only need to know θ_{t-1} and no other previous values. Also, from (2.22) and (2.23) we can obtain the following scaled beta density for $(\theta_t|\theta_{t-1}, D^{t-1}, \lambda)$

$$\begin{aligned}(\theta_t|\theta_{t-1}, D^{t-1}, \lambda) &= \frac{\Gamma(\gamma\alpha_{t-1}+(1-\gamma)\alpha_{t-1})}{\Gamma(\gamma\alpha_{t-1})(1-\gamma)(\alpha_{t-1})}(\theta_t\frac{\gamma}{\theta_{t-1}})^{\gamma\alpha_{t-1}} \\ &(1-\theta_t\frac{\gamma}{\theta_{t-1}})^{(1-\gamma)\alpha_{t-1}} \\ &= \frac{\Gamma(\alpha_{t-1})}{\Gamma(\gamma\alpha_{t-1})(1-\gamma)(\alpha_{t-1})}(\frac{\gamma}{\theta_{t-1}})^{\gamma\alpha_{t-1}}\theta_t^{\gamma\alpha_{t-1}-1} \\ &(1-\frac{\gamma}{\theta_{t-1}}\theta_t)^{(1-\gamma)\alpha_{t-1}}\end{aligned} \tag{2.24}$$

Obviously $(\theta_t|\theta_{t-1}, D^{t-1}, \lambda)$ is defined over $(0, \frac{\gamma}{\theta_{t-1}})$. Next we assume that,

$$(\theta_0|D^0) \sim \Gamma(\alpha_0, \beta_0).$$

We consider that $(\theta_0|D^0) \sim \Gamma(\alpha_0, \beta_0)$ is prior of θ_1 and we also know that $(D^1|\theta_1, \lambda)$ follows Poisson distribution with parameters $\theta_1\lambda$ so we consider that θ_1 follows gamma distribution with parameters (α_1,β_1) and we get conjugate prior. Similar for any time t we obtain θ_{t-1} which follows gamma distribution with parameters (α_{t-1},β_{t-1}) is a prior of θ_t and we also know that $(D^t|\theta_t, \lambda)$ follows Poisson distribution with parameters $\theta_t\lambda$ so we consider that θ_t follows gamma distribution with parameters (α_t,β_t) and we get conjugate prior. Consequently, by induction we have

$$(\theta_{t-1}|D^{t-1}, \lambda) \sim Gamma(\alpha_{t-1}, \beta_{t-1}). \tag{2.25}$$

https://doi.org/10.26219/heal.aueb.6542

Additionally, using (2.24), (2.25) and the law of total probability theorem which says that

$$P(A) = \sum_n P(A|B_n)P(B_n)$$

we obtain that, $(\theta_t|D^{t-1}, \lambda) \sim Gamma(\gamma\alpha_{t-1}, \gamma\beta_{t-1})$ since,

$$\begin{aligned} P(\theta_t|D^{t-1}, \lambda) &= \int P(\theta_t|\theta_{t-1}, D^{t-1}, \lambda)P(\theta_{t-1}|D^{t-1}, \lambda)d\theta_{t-1} \\ &\sim Gamma(\gamma\alpha_{t-1}, \gamma\beta_{t-1}) \end{aligned} \tag{2.26}$$

In addition, using (2.21), (2.26) and Bayes rules we can obtain that,

$$\begin{aligned} P(\theta_t|D^t, \lambda) &\propto P(Y_{t1}, ..., Y_{td}|\theta_t, \lambda)P(\theta_t|D^{t-1}, \lambda) \\ &= (\prod_{j=a}^{d}(\theta_t\lambda_j)^{Y_{tj}} \exp(-\lambda_{tj}\theta_t))(\theta_t^{\gamma\alpha_{t-}} \exp(-\gamma\beta_{t-1}\theta_t)) \\ &= (\prod_{j=1}^{d}\lambda_j^{Y_{tj}})\theta_t^{\sum_{j=1}^{d}Y_{tj}} \exp(\theta_t - \sum_{j=1}^{d}\lambda_j)\theta_t^{\gamma\alpha_{t-1}} \exp(-\gamma\beta_{t-1}\theta_t) \\ &= (\prod_{j=1}^{d}\lambda_j^{Y_{tj}})\theta_t^{\sum_{j=1}^{d}Y_{tj}+\gamma\alpha_{t-1}} \exp(\theta_t - \sum_{j=1}^{d}\lambda_j - \gamma\beta_{t-1}) \\ &= (\prod_{j=1}^{d}\lambda_j^{Y_{tj}})Gamma(\alpha_t, \beta_t) \end{aligned}$$

So,

$$P(\theta_t|D^t, \lambda) \sim Gamma(\alpha_t, \beta_t)$$

where, $\alpha_t = \sum_{j=1}^{d} Y_{tj} + \gamma\alpha_{t-1}$ and $\beta_t = \sum_{j=1}^{d} \lambda_j - \gamma\beta_{t-1}$

Using all the above we can obtain a new model with name multivariate Poisson-scaled beta model (MPSB).

https://doi.org/10.26219/heal.aueb.6542

2.3.1 Dynamic Multivariate Negative Binomial (DMNB) Distribution

An important feature of the model we defined above is the availability of the marginal distribution of Y_{tj} conditional on $\lambda_j' s$ for $j = 1, .., d$ using the law of total probability theorem. This is given by,

$$\begin{aligned} P(Y_{tj}|\lambda, D^{t-1}) &= \int P(Y_{tj}|\theta_t, \lambda_j) P(\theta_t|D^t, \lambda) d\theta_t \\ &= \begin{pmatrix} \gamma\alpha_{t-1} + Y_{tj} - 1 \\ Y_{tj} \end{pmatrix} (1 - \frac{\lambda_j}{\gamma\beta_{t-1} + \lambda_j})^{\gamma\alpha_{t-1}} \\ &(\frac{\lambda_j}{\gamma\beta_{t-1} + \lambda_j})^{Y_{tj}}, \end{aligned} \tag{2.27}$$

which is a negative binomial model denoted as NB $[\gamma\alpha_{t-1}, \lambda_j/(\gamma\beta_{t-1} + \lambda_j)]$, where $\lambda_j/(\gamma\beta_{t-1} + \lambda_j)$ is the probability of success. From the conditional independence assumption, we can obtain the multivariate distribution of $Y_t = (Y_{t1}, .., Y_{td})$ conditional on $\lambda_j' s$ as,

$$P(Y_t|\lambda, D^{t-1}) = \frac{\Gamma(\gamma\alpha_{t-1} + \sum_j Y_{tj})}{\Gamma(\gamma_{t-1}) \prod_j \Gamma(Y_{tj} + 1))} \prod_j (1 - \frac{\lambda_j}{\gamma\beta_{t-1} + \lambda_j})^{\gamma\alpha_{t-1}} \prod_j (\frac{\lambda_j}{\gamma\beta_{t-1} + \lambda_j})^{Y_{tj}}$$

We refer to this distribution as the dynamic multivariate negative binomial distribution (DMNB). Using (2.27) the following proposition is immediately obtained,

Proposition 3.1:
The joint distribution, $P(Y_{ti}, Y_{tj}|\lambda, D^{t-1})$, for series i and j, is given by:

$$\begin{aligned} P(Y_{ti}, Y_{tj}|\lambda, D^{t-1}) =& \frac{\Gamma(\gamma\alpha_{t-1} + Y_{ti} + Y_{tj})}{\Gamma(\gamma\alpha_{t-1})\Gamma(Y_{ti} + 1)\Gamma(Y_{tj} + 1)} (\frac{\gamma\beta_{t-1}}{\lambda_i +_j + \gamma\beta_{t-1}})^{\gamma\alpha_{t-1}} \\ & (\frac{\lambda_i}{\lambda_i +_j + \gamma\beta_{t-1}})^{Y_{ti}} (\frac{\lambda_j}{\lambda_i +_j + \gamma\beta_{t-1}})^{Y_{tj}}. \end{aligned}$$

https://doi.org/10.26219/heal.aueb.6542

The conditional mean regression of Y_{tj} on Y_{ti} is given by

$$E[Y_{tj}|Y_{ti}, \lambda, D^{t-1}] = \frac{\lambda_j(\gamma\alpha_{t-1} + Y_{ti})}{\lambda_i + \gamma\beta_{t-1}}.$$

The correlation between Y_{ti}, Y_{tj} is given by:

$$Cor(Y_{ti}, Y_{tj}|\lambda, D^{t-1}) = \sqrt{\frac{\lambda_i\lambda_j}{(\lambda_i + \gamma\beta_{t-1})(\lambda_j + \gamma\beta_{t-1})}}.$$

Using the above is obvious that our model is suitable only for multivariate time series with positive correlation. For some this could be a drawback of the model. But, if we take into account that in many field such as finance there are often multivariate measurement time series that show a positive correlation, then we can say that this property of the model is just a limitation and not a disadvantage.

2.3.2 Parameters Estimation

Firstly, assuming that γ is known and that a priori $\lambda_j's$ are independent of each other as well as θ_0 and having Gamma priors

$$\lambda_j \sim Gamma(\alpha_j, b_j) \quad \text{for } \ j = 1, .., d.$$

We will use Forward Filtering and Backward Sampling (FFBS) to estimate the parameters of the model we defined above (in the previous subsections). The FFBS is an efficient block sampler that draws the states jointly given the parameters for linear Gaussian state-space models. The steps of this algorithm are presented in detail below.

Forward Filtering and Backward Sampling (FFBS) algorithm:

Step 1:

https://doi.org/10.26219/heal.aueb.6542

In the first step the conditional joint distribution of the state parameters is estimated. For the joint density $P(\theta_1, ..., \theta_t|\lambda, D^t)$ using the chain rule of probability for more than two random variables and the fact that θ_t has Markovian evaluation we have that

$$P(\theta_1, ..., \theta_t|\lambda, D^t) = P(\theta_t|\lambda, D^t)P(\theta_{t-1}|\theta_t, \lambda, D^t)....P(\theta_1|\theta_2, \lambda, D^1)$$

The implementation of FFBS would be straightforward in our model as we have the following shifted gamma densities

$$(\theta_{t-1}|\theta_t, \lambda, D^{t-1}) \sim Gamma[(1-\gamma)\alpha_{t-1}, \beta_{t-1}]$$

where $\gamma\theta_t < \theta_{t-1}$.

Step 2:
Using the Bayes' theorem we obtain that,

$$\begin{aligned} P(\lambda_j|\theta_t, D^t) &\propto P(Y_{1j}, ..., Y_{tj}|\theta_t, \lambda_j P(\lambda_j) \\ &\propto (\prod_t (\theta_t\lambda_j)^{Y_{tj}} \exp -\lambda_j\theta_t)(\lambda_j^{\alpha_j-1} \exp(-b_j\lambda_j)) \\ &= (\prod_t \theta_t^{Y_{tj}})\lambda_j^{\sum_t \alpha_{tj}-1} \exp(-\lambda_j(\sum_t(\theta_t - b_j)) \end{aligned} \tag{2.28}$$

So,

$$\lambda_j|\theta_t, D^t \sim Gamma(\alpha_{tj}, b_{tj})$$

where, α_{tj}=$\alpha_j + Y_{t1+...+Y_{td}} and$b$_{tj}$=$b_j + \sum_{j=1}^t \theta_j$.

Using the a priori independence of λ_j and the (2.28) it turns out that given the state parameters, θ_t and data, then the $\lambda_j's$ are conditionally independent. However, unconditionally they will not necessarily be independent.

Using the FFBS algorithm we were able to estimate both θ_t and λ_j. In

https://doi.org/10.26219/heal.aueb.6542

a next stage we want to update both the dynamic and the static parameters. To achieve this, we will use the Particle Learning method (PL). Before we introduce the method, we assume that γ is known and that z_t is the essential vector of parameters that we watch in time t. s_t is conditional sufficient statistics s_t = f(s_{t-1}, θ_t, Y_t).The PL method is presented briefly below.

Particle Learning method (PL):

1.(Resample) $\{z_t\}_{i=1}^{N}$ from $z_t^{(i)}= \{s_t, \lambda\}$ using weights $w_t^i \propto P(Y_{t+1}|z_t^i)$

2.(Propagate) θ_t^i to θ_{t+1}^i via $P(\theta_{t+1}|z_t^i, Y_{t+1})$

3.(Update) $s_{t+1}^i = f(s_t^i, \theta_{t+1}^i, Y_{t+1})$

4.(Sample) λ^i from $P(\lambda|s_{t+1}^j)$

More information on calculating steps 2, 3 of Particle Learning method (PL) and for updating the Discount Factor γ can be found in the original article Aktekin et al. (2018).

2.3.3 Conclusions

In this part a State Space model was studied to describe time series exposed in the same environment. This model was created by entering two parameters. The dynamic state parameter θ_t=$(\theta_1, ..., \theta_T)$ which refers to the common environment of time series at the time t and the static parameters $\lambda = (\lambda_1, ..., \lambda_d)$ which refers to specific characteristics of the $j-th$ time series. The model introduced here has the limitation that it can not deal with non-positively correlated time series.

https://doi.org/10.26219/heal.aueb.6542

2.4 Dynamic Factor Models for Multivariate Count Data: An Application to Stock-Market Trading Activity

I this subsection We will now introduce one more parameter-driven model to describe multivariate time series of counts. Specifically, we will introduce a model that can describe the joint behaviour of the number of trades for different stocks in the New York Stock Exchange (NYSE). A detailed description of the model that we will introduce here can be found in Jung et al. (2011).

First of all we introduce θ_t=$(\theta_{t1}, ..., \theta_{td})$, where θ_{tj} is latent random variables for all t=$1, .., T$, j=$1, .., d$ and we make the following assumption:

Assumption 4.1:
Y_{tj} are conditionally independent with Poisson distributions for given θ_{tj}, i.e,

$$P(Y_{tj}|\theta_{tj}) = \frac{\exp{(-\theta_{tj})}\theta_{tj}^{Y_{tj}}}{Y_{tj}!} \quad \text{for } t = 1, ..., T\ , j = 1, .., d.$$

Our aim is to describe the Y_{tj}. By the Assumption 4.1 immediately follows that in order to study the Y_{tj} it is enough to study-describe the θ_{tj}. Therefore, we will enter a link function $b(.)$ which will be such that,

$$b(\theta_t) = \mu + \Gamma f_t$$

where, μ denotes a vector of fixed intercepts and Γ is a $(d \times P)$ matrix of factor loadings. For the f_t we make the following assumption:

Assumption 4.2:
f_t is a P-dimensional vector of latent random factors, where the P latent

https://doi.org/10.26219/heal.aueb.6542

factors in f_t is independent of each other.

Essentially, using $b(.)$ we managed to express the mean of Y_t i.e the θ_t as a linear function of f_t.

It is obvious that it is very useful to set $b(\theta_t)$=log (θ_t), because, using logarithm immediately follows that θ_t are positive and that we do not need any constraint for μ and Γ since the set of values of the logarithm function is the set of real numbers.

Using the Assumptions 4.1 and 4.2 we introduce a model for Y_t. But as we have mentioned above, we want our model to describe the joint behavior of the number of trades for different stocks in the NYSE. Therefore, we will describe f_t using the factors λ_t, τ_t, ω_t, i.e f_t=$(\lambda_t,\, \tau_t^{'},\, \omega_t^{'})^{'}$ where λ_t is a single common market factor, τ_t=$(\tau_{t1}, ..., \tau_{tS})$ is industry specific factor with $S < d$ and ω_t=$(\omega_{t1}, .., \omega_{td})$ is a stock specific factor. Obviously, it holds that P=$S + d + 1$. The matrix of factor loading is partitioned conformably with f_t into Γ=$(\Gamma_\lambda, \Gamma_\tau, \Gamma_\omega)$, where, Γ_λ=(γ^λ) is a d-dimensional vector, Γ_τ=$(\gamma_j^{\tau_s})$ is a $d \times P$ matrix with zero entries for any firm j that does not belong to a sector s and, Γ_ω=$diag(\gamma_j^\omega)$ is a $d \times d$ diagonal matrix.

Having defined all of the above, we obtain that the log link function for the θ_{tj}, i.e the log link function for the stock j that belong to industry s is given by

$$\begin{aligned} \log(\theta_{tj}) &= \mu_j + \Gamma f_{tj} \\ &= \mu_j + (\gamma_j, \gamma_j^{\tau_{sj}}, \gamma_j^{\omega_j})(\lambda_s, \tau_{tsj}, \omega_{tj}) \\ &= \mu_j + \gamma_j \lambda_s + \gamma_j^{\tau_{sj}} \tau_{tsj} + \gamma_j^{\omega_j} \omega_{tj}. \end{aligned} \tag{2.29}$$

Using the equation (2.29) we described the $\theta_{tj}^{'}s$ and therefore, the $Y_{tj}^{'}s$. However, our goal is not just to describe the Y_{tj} but to take into account both

 https://doi.org/10.26219/heal.aueb.6542

the serial and the cross correlation of the Y_{tj}. To achieve this we will make the following assumption:

Assumption 4.3:
The factors λ_t, τ_{ts} and ω_{tj} follow independent Gaussian AR (1) processes, i.e

$$\lambda_t|\lambda_{t-1} \sim N(k^{\lambda} + \delta^{\lambda}\lambda_{t-1}, (\nu^{\lambda})^2) \tag{2.30}$$

$$\tau_{t_s}|\tau_{t-1_s} \sim N(k^{\tau_s} + \delta^{\tau_s}\tau_{t-1s}, (\nu^{\tau_s})^2) \tag{2.31}$$

$$\omega_{tj}|\omega_{t-1j} \sim N(k^{\omega_j} + \delta^{\omega_j}\omega_{t-1_j}, (\nu^{\omega_j})^2). \tag{2.32}$$

2.4.1 Properties of observed process Y_t

Using the equations (2.29)-(2.32) we obtain that:

$$\log(\theta_{tj}) \sim N(\mu_j+k^{\lambda}+\delta^{\lambda}\lambda_{t-1}+k^{\tau_s}+\delta^{\tau_s}\tau_{t-1s}+k^{\tau_s}+\delta^{\tau_s}\tau_{t-1s}, (\nu^{\lambda})^2+(\nu^{\tau_s})^2+(\nu^{\omega_j})^2)$$

Therefore, $\exp(\theta_{tj})$ follows log-normal distribution.

We set α_j=$\mu_j + k^{\lambda} + \delta^{\lambda}\lambda_{t-1} + k^{\tau_s} + \delta^{\tau_s}\tau_{t-1s} + k^{\tau_s} + \delta^{\tau_s}\tau_{t-1s}$ and σ_j=$(\nu^{\lambda})^2 + (\nu^{\tau_s})^2 + (\nu^{\omega_j})^2$, so we can obtain the follow propositions.

Proposition 4.1: The unconditional mean of Y_t is given by $E[Y_t]$= [E(Y_{t1}), ..., $E(\mathrm{Y}_{td})]^T$, where,

$$E(\mathrm{Y}_{tj}) = \exp{(\alpha_j + \frac{1}{2}\sigma_j)}.$$

Proof:

$\log(\theta_{tj})$ follows normal distribution with α_j as mean and σ_j^2 as variance,it is easy to see that the θ_{tj} has a log-normal distribution with mean

$$E(\theta_{tj}) = \exp{(\alpha_j + \frac{1}{2}\sigma_j)} \quad \text{for } i = 1, .., d.$$

This is so because of

$$E(Y_{tj}) = E(E(Y_{tj}|\theta_{tj})) = E(\theta_{tj}) \quad \text{for } j = 1, .., d.$$

It turns out that

$$E(Y_{tj}) = \exp{(\alpha_j + \frac{1}{2}\sigma_j)} = k_j \quad \text{for } j = 1, .., d.$$

Proposition 4.2: The unconditional covariance matrix of Y_t is denoted by $Cov(Y_t)$ and its elements are given by

$$Var(Y_{ti}) = k_i[1 + k_i(\exp{(\sigma_i)} - 1)] \quad \text{for } i = 1, .., d,$$

$$Cov(Y_{ti}, Y_{tj}) = E(\theta_{ti}\theta_{tj}) - E(\theta_{ti})E(\theta_{tj}) \quad \text{for } i \neq j \ i = 1, .., d, \ j = 1, .., d.$$

https://doi.org/10.26219/heal.aueb.6542

Proof:

$$
\begin{aligned}
Var(Y_{ti}) &= E(Y_{ti}^2) - E[Y_{ti}]^2 \\
&= E[Var(\ Y_{ti}|\theta_{ti})\ + E(\ Y_{ti}|\theta_{ti})^2] - [E[E(\ Y_{ti}|\theta_{ti})\ ]]^2 \\
&= E(Var(Y_{ti}|\theta_{ti})) + (\ E(E(\ Y_{ti}|\theta_{ti})^2) - E(E(Y_{ti}|\theta_{ti}))^2) \\
&= E(\ Var(Y_{ti}|\theta_{ti})\ )\ + Var(\ E(Y_{ti}|\theta_{ti})) \\
&= E(\theta_i)\ + Var(\theta_i) \\
&= E(\theta_i)\ + E(\theta_i^2) - E(\theta_i)^2 \\
&= \exp(\ \alpha_i + \frac{1}{2}\sigma_i)\ + \exp(\ 2\alpha_i + 2\sigma_i)\ - \exp(\ 2\alpha_i + \sigma_i) \\
&= \exp(\ \alpha_i + \frac{1}{2}\sigma_i)\ + \exp(\ 2\alpha_i + \sigma_{ii})\ (\ \exp(\ \sigma_i) - 1) \\
&= \exp(\ \alpha_i + \frac{1}{2}\sigma_i)\ (1 + \exp(\ \alpha_i + \frac{1}{2}\sigma_i)\ (\exp(\ \sigma_i)\ - 1)\ ) \\
&= k_i[1 + k_i(\ \exp(\sigma_i)\ - 1)\ ]
\end{aligned}
$$

$$
\begin{aligned}
Cov(Y_{ti}, Y_{tj}) &= E(Y_{ti}Y_{tj}) - E(Y_{ti})E(Y_{tj}) \\
&= E(\theta_{ti}\theta_{tj}) - E(\theta_{ti})E(\theta_{tj}) \\
&= E(\exp(\mu_i + \gamma_i\lambda_s + \gamma_i^{\tau_{si}}\tau_{tsi} + \gamma_i^{\omega_i}\omega_{ti} + \mu_j + \gamma_j\lambda_s + \gamma_j^{\tau_{sj}}\tau_{tsj} + \gamma_j^{\omega_j}\omega_{tj})) - \\
&\exp(\alpha_i + \alpha_j + \frac{1}{2}\sigma_i + \frac{1}{2}\sigma_j) \\
&= E(\exp(\alpha_i + \alpha_j + \frac{1}{2}(\sigma_i + \sigma_j\sigma_{ij}))) - \exp(\alpha_i + \alpha_j + \frac{1}{2}\sigma_i + \frac{1}{2}\sigma_j) \\
&= \exp(\alpha_i + \alpha_j + \frac{1}{2}\sigma_i + \frac{1}{2}\sigma_j)(\exp(\sigma_{ij}) - 1) \\
&= \exp(\alpha_i + \frac{1}{2}\sigma_i)\exp(\alpha_j + \frac{1}{2}\sigma_j)(\exp(\sigma_{ij}) - 1) \\
&= k_ik_j(\exp(\sigma_{ij}) - 1)
\end{aligned}
$$

https://doi.org/10.26219/heal.aueb.6542

From the above, immediately follows that:

The correlation between Y_{ti} and Y_{tj}, for $i \neq j$, $i = 1, .., d, j = 1, ..., d$ is

$$Corr(Y_{ti}, Y_{tj}) = \frac{\exp(\sigma_{ij}) - 1}{\sqrt{[k_i^{-1} + \exp(\sigma_i) - 1]}\sqrt{[k_j^{-1} + \exp(\sigma_j) - 1]}}$$

where σ_{ij} is the correlation between $\log(\theta_{ti})$ and $\log(\theta_{tj})$.

The above formula shows that the cross-correlation structure of Y_t depends on σ_{ij}.

So far we have managed to describe Y_{tj} and take into account both serial and cross correlation. Also we take that, the unconditional variance of Y_{tj} exhibits overdispersion and the amount of this overdispersion increases with α_i and σ_i where α_j=$\mu_j + k^\lambda + \delta^\lambda \lambda_{t-1} + k^{\tau_s} + \delta^{\tau_s}\tau_{t-1s} + k^{\tau_s} + \delta^{\tau_s}\tau_{t-1s}$ and σ_j=$(\nu^\lambda)^2 + (\nu^\lambda)^2 + (\nu^{\omega_j})^2$. Using (2.30)-(2.31) we take α_i and σ_i are critically dependent on the persistence parameters $(\delta^\lambda, \delta^{\tau_s}, \delta^{\omega_j})$. In order to relax this close relationship between overdispersion and persistence, we will substitute the more flexible Negative Binomial (Negbin) distribution for the Poisson which is given by,

$$P(Y_{tj}|\theta_{tj}) = \frac{\Gamma(Y_{tj} + 1/\sigma_j^2)}{\Gamma(1/\sigma_j^2)\Gamma(Y_{tj}+1)}(\frac{1}{1+\sigma_j^2\theta_{tj}})^{1/\sigma_j^2}(\frac{\theta_{tj}}{\theta_{tj}+1/\sigma_j^2})^{Y_{tj}}$$

where,

$$E(Y_{tj}|\theta_{tj}) = \theta_{tj}$$

and

$$Var(Y_{tj}|\theta_{tj}) = \theta_{tj}(1 + \sigma_j^2\theta_{tj}). \tag{2.33}$$

https://doi.org/10.26219/heal.aueb.6542

We observe that,

$$\lim_{\sigma_j \to 0} Var(Y_{tj}|\theta_{tj}) = \lim_{\sigma_j \to 0} \theta_{tj}(1 + \sigma_j^2 \theta_{tj}) = \theta_{tj}$$

i.e the Poisson distribution in Assumption 4.1 is obtained as the limit for $\sigma_j \to 0$. Also, from Equation (2.33) seem that overdispersion is a monotone increasing function of $\sigma_j > 0$.

The parameters of models are estimated by Maximum Likelihood (ML) based upon efficient importance sampling (EIS). More information about efficient importance sampling (EIS) is found in the original article Jung et al. (2011) and in Richard and Zhang (2007).

2.4.2 Conclusions

In this section a State Space model was studied to describe time series exposed in the same environment. This model is developed for very specific data (NYS), which is limiting but, it allows the use of another conditional distribution for Y_t instead of Poisson distribution. The model introduced here has also the advantage that it can deal with both positive and negative correlated time series.

2.5 Hierarchical Markov-switching models for multivariate integer-valued time-series

In this part of the thesis, we will introduce a flexible with easy parameter estimation parameter-driven model to describe multivariate time series of counts. This model was first introduced in Catania and Di Mari (2021).

First, we will introduce two unobserved parameters S_t and Z_t. More specif-

https://doi.org/10.26219/heal.aueb.6542

ically, for S_t we assume that is an unobserved first order stationary and ergodic Markov chain with state space $S = \{1, ..., J\}$ and transition probability matrix $\Gamma = [\gamma_{j,h}]$, where $\gamma_{j,h} = P(S_t = h|S_{t-1} = j)$ and $\sum_{h\in S} \gamma_{j,h} = 1$ for all $j, h \in S$ and stationary distribution π, which satisfies $\Gamma\ \pi = \pi$. For Z_t we assume that is an unobserved conditionally independent distributed integer-valued random variables given S_t with support $K = 1, ..., K$. Furthermore, we assume that the probability of $Z_t = k$ given $S_t = j$ denoted by $\omega_{j,k}$, i.e;

$$P(Z_t = k|St = j) = \omega_{j,k} > 0$$

with $\sum_{k\in K} \omega_{j,k} = 1$ for all $\in S$. We collect all the $\omega_{j,k}$ in a $J \times K$ matrix Ω.

Having all of the above we make the following assumption.

Assumption 5.1:
Y_{tj} are conditionally independent with Poisson distributed with rate $\lambda_{j,k,i}$ given S_t and Z_t, i.e.

$$P_{\lambda_{j,k,i}}(Y_{ti} = q|S_t = j, Z_t = k) = \frac{\lambda_{j,k,i}^q \exp(-\lambda_{j,k,i})}{q!}$$

where $q = 0, 1, 2, 3, ...$
Due to independence of Y_{ti} the following equation arises immediately.

$$P(Y_t|S_t = j, Z_t = k) = \prod_{i=1}^{d} P_{\lambda_{j,k,i}}(Y_{ti}|S_t = j, Z_t = k)$$

Also, using marginalization we obtain the following factorization,

$$P(Y_t|S_t = j) = \sum_{k=1}^{K} \omega_{j,k} P_{\lambda_{j,k}}(Y_t|S_t = j, Z_t = k)$$

https://doi.org/10.26219/heal.aueb.6542

The joint density of a sequence of T variables, $Y_{1:t}$, is obtained after marginalization of the latent variables and using the law of total probability as follows:

$$P(Y_{1:T}:\theta) = \sum_{S_1,..,S_T \in S} \delta_{S_1} \prod_{t=2}^{T} \gamma_{S_{t-1},S_t} \prod_{t=1}^{T} \sum_{k=1}^{K} \omega_{j,k} P_{\lambda_{j,k}}(Y_t|S_t = j, Z_t = k) \tag{2.34}$$

or, in matrix notation,

$$P(Y_{1:T};\Theta) = \delta' P_1 \Gamma P_2 ... \Gamma P_T 1$$

where, 1 is a J-vector of ones and P_t is a $J \times J$ diagonal matrix with typical element $p_{j,j,t}$=$\omega_j' P_{j,t}$, with ω_j=$(\omega_{j,k}, k = 1,..,K)'$ and $\prod_{i=1}^{d} P_{\lambda_{j,k}}(Y_t|S_t = j, Z_t = k), k = 1,..,K)'$, in this paper, the writers set the initial distribution of the chain, δ, equal to the stationary distribution.

Essentially, the Equation 2.34 defines our model which has the name Hierarchical Markov switching model (HMSM).

2.5.1 Properties of observed process Y_t

First of all, we initiate the "Forward Probabilities" of S_t, α_t'=$\alpha_{t-1}'\Gamma P_t$ with α_1'=$\delta' P_1$, such that $\alpha_t = (\alpha_{j,t}, j = 1,..,J)'$, where $\alpha_{j,t} = P(Y_{1:t}, S_t = j)$. Exploiting the forward probabilities, the predictive distributions of the latent chain and the observables at time $t+h$, for $h > 0$, given observations up to time t are simply given by:

$$\pi_{t+h|t} = P(S_{t+h}|Y_{1:t}) = \alpha_t'\Gamma^h/(\alpha_t' 1)$$

and

$$P(Y_{t+h}|Y_{1:t}) = \sum_{j=1}^{J} \pi_{j,t+h|t} \sum_{k=1}^{K} \omega_{j,k} \prod_{i=1}^{d} \frac{\lambda_{j,k,i}^{q} \exp\left(-\lambda_{j,k,i}\right)}{q!}. \tag{2.35}$$

https://doi.org/10.26219/heal.aueb.6542

The parameter estimation of this model is done using EM algorithm. For more information refer to the original article Catania and Di Mari (2021).

2.5.2 Conclusions

In this section we presented the Hierarchical Markov switching model (HMSM). A model which is flexible to describe multivariate time series of counts. In the original article Catania and Di Mari (2021) one can find two variants related to the relationship between the unobserved parameters S_t and Z_t in addition to a complete description of the model.

https://doi.org/10.26219/heal.aueb.6542

https://doi.org/10.26219/heal.aueb.6542

Chapter 3

Conclusions

The topic of this thesis is the discription of some parameter driven models for multivariate time series of counts. Specifically, influenced by the popularity of observation driven models due to their ease of parameter estimation, this thesis emphasizes on the development and description of five parameter driven models. The aim is to focus on the positives of the models that belong to this category.

We first described the State space multivariate Poisson model which has the advantages that can describe both negative and positive cross correlation and autocorrelation, and it turn into account the overdispersion of a Y_t multivariate time series of counts. The parameter estimation for this model is performed through the Monte Carlo EM process which uses the Particle filtering (PF) and Particle smoothing (PS) methods. This model can be used only from stationary Y_t multivariate time series of counts, this can be considered a limitation of the SSMP model.

Later we described a doubly stochastic model with latent factors which can be used by both stationary and non-stationary multivariate time series of counts and it can describe the cross correlation the autocorrelation and take

https://doi.org/10.26219/heal.aueb.6542

into account the overdispersion too. In this model we do not make any assumptions about the distribution of unobserved d-dimensional process Λ_t that we introduce. Therefore, the model is very flexible and able to be used in many different situations. A discrepancy of this model could be considered the fact that part of the parameter estimation are made by eigendecomposition. A process probably not familiar to several people.

Third, a Bayesian approach was then given to the multivariate time series of counts description through the model multivariate Poisson scaled beta (MPSB) models. This model is the only one described in this thesis that has the limitation that it can only describe positive correlations. It is noteworthy that during the development of this model the dynamic multivariate negative binomial distribution (DMNB) is developed.

This was followed by a model which in order to describe specific data makes some specific assumptions which however can be changed and so the model can be used in general. An important feature of this model is that initially assumes that the multivariate time series of counts Y_t conditional on θ_t follows Poisson distribution like all others models which are developed here, but in order to take into account the overdispersion of a Y_t it assumes that the Y_t conditional on a latent factor follow Negative binomial distribution.

Lastly, a flexible dynamic model was introduced. The latest model has easy development and common parameter estimation since it uses the well-known EM algorithm.

In conclusion, in this thesis one can identify different parameter driven models that may have common points in their development, such as the case of distribution in Poisson terms, but each has its own characteristics which make it unique and easy to use in specific cases. Resulting we have a set of

https://doi.org/10.26219/heal.aueb.6542

parameter driven models that flexibly describe different types of multivariate time series of counts.

An interesting approach would be to run all the algorithms mentioned in this thesis. So one can easily use the models. At the same time one could focus on finding articles that evolve these methods so that they are even more computationally easy and efficient.

Finally, we would like to mention that the aim of thesis is to highlight features that parameter driven models have and make them ideal for explaining multivariate time series of counts. But, in any case, parameter driven models are not the only category that can be considered ideal. Observation driven models rightly have their own place in the literature and have their own characteristics that make them stand out. After all, the goal is to develop as many models as possible to describe the multivariate time series of counts since such time series are now widely used.

https://doi.org/10.26219/heal.aueb.6542

Bibliography

Aktekin, T., Polson, N., and Soyer, R. (2018). Sequential bayesian analysis of multivariate count data. *Bayesian Analysis*, 13(2):385–409.

Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive process. *Journal of Time Series Analysis*, 8(3):261–275.

Blasques, F., Gorgi, P., Koopman, S. J., and Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, 12(1):1019–1052.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Brockwell, P. J. and Davis, R. A. (2009). *Time series: theory and methods.* Springer Science & Business Media.

Campbell, A., Shi, Y., Rainforth, T., and Doucet, A. (2021). Online variational filtering and parameter learning. *Advances in Neural Information Processing Systems*, 34.

Catania, L. and Di Mari, R. (2021). Hierarchical markov-switching models for multivariate integer-valued time-series. *Journal of Econometrics*, 221(1):118–137.

https://doi.org/10.26219/heal.aueb.6542

Chan, K. and Ledolter, J. (1995). Monte carlo em estimation for time series models involving counts. *Journal of the American Statistical Association*, 90(429):242–252.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika*, 70(1):269–274.

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

Cox, D. R. and Isham, V. (1980). *Point processes*, volume 12. CRC Press.

Creal, D., Koopman, S. J., and Lucas, A. (2008). A general framework for observation driven time-varying parameter models. Technical report, Tinbergen Institute Discussion paper.

Davis, R., Rodriguez-Yam, G., Dunsmuir, W., and Streett, S. (2004). Parameter-and observation-driven state space models. *Lecture available at http://www. stat. columbia. edu/˜ rdavis/lectures/Cyprus1_ 04. pdf*.

Davis, R. A., Dunsmuir, W. T., and Streett, S. B. (2003). Observation-driven models for poisson counts. *Biometrika*, 90(4):777–790.

Djuric, P. M., Kotecha, J. H., Zhang, J., Huang, Y., Ghirmai, T., Bugallo, M. F., and Miguez, J. (2003). Particle filtering. *IEEE Signal Processing magazine*, 20(5):19–38.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation.

Fokianos, K. (2021). Multivariate count time series modelling. *arXiv preprint arXiv:2103.08028*.

https://doi.org/10.26219/heal.aueb.6542

Fokianos, K. and Fried, R. (2010). Interventions in ingarch processes. *Journal of Time Series Analysis*, 31(3):210–225.

Hansen, H. and Johansen, S. (1999). Some tests for parameter constancy in cointegrated var-models. *The Econometrics Journal*, 2(2):306–333.

Jung, R. C., Liesenfeld, R., and Richard, J.-F. (2011). Dynamic factor models for multivariate count data: An application to stock-market trading activity. *Journal of Business & Economic Statistics*, 29(1):73–85.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, pages 10–13.

Laird, N. (1993). 14 the em algorithm.

Lange, R.-J. (2020). Bellman filtering for state-space models. *arXiv preprint arXiv:2008.11477*.

Li, Y.-X. and Wang, J.-F. (2008). An almost sure central limit theorem for products of sums under association. *Statistics & Probability Letters*, 78(4):367–375.

Lin, J.-L. and Tsay, R. S. (1996). Co-integration constraint and forecasting: an empirical examination. *Journal of Applied Econometrics*, 11(5):519–538.

Liu, H. (2012). *Some models for time series of counts, PHD Thesi*. Columbia University.

Mamode Khan, N., Sunecher, Y., and Jowaheer, V. (2016). Modelling a non-stationary poisson process. *Journal of Statistical Computation and Simulation*, 86(15):3106–3126.

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

https://doi.org/10.26219/heal.aueb.6542

Neath, R. C. (2013). On convergence properties of the monte carlo em algorithm. In *Advances in modern statistical theory and applications: a Festschrift in Honor of Morris L. Eaton*, pages 43–62. Institute of Mathematical Statistics.

Ng, S. K., Krishnan, T., and McLachlan, G. J. (2012). The em algorithm. In *Handbook of Computational Statistics*, pages 139–172. Springer.

Pedeli, X. and Karlis, D. (2011). A bivariate process with application. *Statistical modelling*, 11(4):325–349.

Pedeli, X. and Karlis, D. (2013a). On estimation of the bivariate poisson inar process. *Communications in Statistics-Simulation and Computation*, 42(3):514–533.

Pedeli, X. and Karlis, D. (2013b). Some properties of multivariate processes. *Computational Statistics & Data Analysis*, 67:213–225.

Richard, J.-F. and Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2):1385–1411.

Shumway, R. H., Stoffer, D. S., and Stoffer, D. S. (2000). *Time series analysis and its applications*, volume 3. Springer.

Taylor, W. R. (1986). Identification of protein sequence homology by consensus template alignment. *Journal of Molecular Biology*, 188(2):233–258.

Teräsvirta, T. (2009). An introduction to univariate garch models. In *Handbook of Financial Time Series*, pages 17–42. Springer.

Triantafyllopoulos, K. (2021). Non-linear and non-gaussian state space models. In *Bayesian Inference of State Space Models*, pages 263–339. Springer.

Vermeer, N. (2020). Multivariate parameter-and observation-driven models with applications in risk management, thesis.

https://doi.org/10.26219/heal.aueb.6542

Wang, F. and Wang, H. (2018). Modelling non-stationary multivariate time series of counts via common factors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):769–791.

Zhang, C., Chen, N., and Li, Z. (2017). State space modeling of autocorrelated multivariate poisson counts. *IISE Transactions*, 49(5):518–531.