Università
Ca'Foscari
Venezia

Master's Degree Programme – Second Cycle
(*D.M. 270/2004*) in Economics

Erasmus Mundus QEM: Models and Methods
of Quantitative Economics

Final Thesis

Ca' Foscari
Dorsoduro 3246
30123 Venezia

Dynamic Generalized Poisson Panel-Data Models : An Application to Cybersecurity Risk Modeling

Supervisor
Ch. Prof. Roberto Casarin

Graduand
Yashika Ranjan
Matriculation Number: 866295

Academic Year
2017 / 2018

Abstract

Cybersecurity risk modeling is a relatively new topic that has attracted the attention of companies seeking to provide insurance coverage against cyberattacks. In this study I introduce dynamic generalized Poisson panel-data models for cybersecurity risk modeling. Following Zhu (2012), I extend the univariate generalized Poisson INGARCH model to the case of panel data. I consider different parsimonious specifications of the model such as partial pooling and complete pooling of the model parameters, in addition to modeling the whole panel as one or otherwise grouping the units of the panel into clusters based on certain characteristic similarities. The stationarity conditions of the proposed models are studied. As an application, I use cyberattack data on 491 consecutive victim IP addresses which exhibit intrinsic spatiotemporal attack patterns (as analyzed by Chen et al. (2015)). After the models are estimated I compare them according to their likelihood value, AIC and BIC criteria. Finally, I provide a forecast comparison of the proposed models. The results of this study can be further used in the cyber-insurance industry for example, for the pricing of insurance products.

Contents

List of Tables

List of Figures

1 Introduction

Recent years have seen the advent of many new technologies and scientific breakthroughs that are revolutionizing business models in every sector of the economy. New technologies not only transform daily lives but also expose the global economy to a number of vulnerabilities including cyber-attacks. In this document I will focus on the analysis of cyber-attacks by introducing new models for cybersecurity risk measuring and forecasting.

Nowadays, cyber-security has become a major concern for organizations in every sector of the economy. From internet banking to government infrastructure, the whole world is networked together, thus calling for cyber-protection. These days almost all banks provide online services and are therefore exposed to the risk of information and identity theft of their customers. Data security breaches not only affect banking sector but also governments who maintain enormous amount of personal data and records of their citizens as well as confidential government information. Even businesses are prone to data breaches that target their customer and employee information. The technological developments in every sector and increasing volumes of available data are providing increasing grounds for attacks by malicious hackers The growth of the Internet of Things (IoT) has introduced new vulnerabilities. IoT refers to a system of interrelated computing devices such as mobiles, home appliances, vehicles etc embedded with sensors which allows these things to connect and transfer data over a network without the involvement of humans. With increase in the number of IoT devices cyberattacks have become an increasingly physical rather than virtual threat because not all of these connected devices are currently designed keeping security in mind (Camillo (2017)). Specific threats that pose risks to cyber security range from user surveillance algorithms, identity cloning and phishing to creating and distributing viruses.

Cyber-attacks not only have a negative impact on the performance of organizations but also direct economic consequences. In 2016, the once dominant internet giant, Yahoo reported two major data breaches that had occurred in late 2013 and 2014. Both breaches are considered the largest reported in the history of the Internet which affected all of Yahoo's 3 billion user accounts resulting in a $350 million drop in Yahoo's sale price. Another data breach was suffered by retail giant Target store in 2013 in which personally identifiable information (PII) of about 110 million customers was compromised amounting to a loss of $162 million. This list of victims is ever increasing. In a study entitled Cost of Cyber Crime, (Ponemon Institute, 2017) reveals that the global average cost of cyber-crime has been steadily increasing over the last five years. The study reported a 23% increase in cost of cyber-crime as compared to 2016 with average annualized cost of cyber-crime being 11.7 million USD in 2017. With this increase in economic loss due to cyber-attacks, organizations started making tremendous efforts to mitigate and cope with increasing cybersecurity threats. One such effort was the new privacy regulation in the US as a consequence of increase in the number of cases involving identity theft and data breaches. In particular, the first data breach notification law was enacted in California in 2002 which was eventually adopted by other states as well. This law requires companies to notify individuals when their personal information is exposed as a result of a data breach. In the EU, E-privacy Regulations were introduced in 2009 for telecommunication companies and internet service providers. Moreover, banks and insurers had been imposed with fines by financial regulators for data-failings. These new notification requirements, as a step towards insuring against the costs associated with a major data breach, led to a major shift in the demand for cyber insurance products. In particular in US, the cyber insurance market saw a rapid growth as the information about major data breaches and costs associated with them became increasingly available (Camillo (2017)).

In a survey conducted by Ponemon Institute in 2017, 87% of risk management professionals viewed cyber liability as one of their top ten business risks, however only 24% of them reported that their companies had cyber insurance. Inadequate risk coverage came out to be the main reason for not purchasing cybersecurity insurance. Nevertheless according to Aon (2018) Cybersecurity predictions report the situation is likely to change in 2018 when companies will demand coverage for the full impact of cyber risk. Moreover, the demand will come not only from the traditional buyers of cyber insurance, such as the retail, financial, and health care sectors but also from other sectors prone to cyber-related business disruption like airlines, power grids etc. The insurance industry can therefore capitalize on the growing cyber market by designing enterprise cyber insurance policies covering a wide range of cyber-related exposures and pricing them effectively by accurately quantifying the risks to which their clients are exposed.

Insurers understanding of cyber liability and risk aggregation is an evolving process as their experience on cyber-attacks increases. In case of natural calamities, such as flood or earthquake, the expected losses for a given insured asset can be estimated using years of historical data on occurrence and impact, scientific research into understanding the underlying drivers, engineering studies on structural vulnerabilities and past insurance claims data (OECD and Marsh & McLennan Companies (2018)). However, cybersecurity insurance is quite different from other insurance products because it has no standard scoring systems or actuarial tables constructed from historical records which can be used for pricing it (Xu and Hua (2017)). Underwriting insurance coverage for cyberattacks involves a lot of uncertainty due to various factors such as limited historical data, changing nature of risk and impact and limited understanding of and access to security information. The emergence of cyber risk as a matter of concern happened very recently due to which there is limited time-series data on incident frequency. The dearth is also due to a low share of cyber incidents being discovered and actually reported. Another special characteristic of cyber risk is that its dynamic because attacks are driven by human behavior rather than natural forces. Attack methods keep evolving alongside of improvements in defense strategies thereby challenging the effectiveness of different security practices over time (OECD and Marsh & McLennan Companies (2018)).

The modeling of cyber risk for the purposes of pricing insurance coverage, transferring risks to reinsurance and capital markets and calculating capital requirements is just emerging. Most of the models currently used are scenario-based rather than probabilistic and focused on extreme incidents for the purpose of managing accumulation risk. Many modeling firms are also developing new approaches focused on the security practices of companies instead of the attack nature (OECD and Marsh & McLennan Companies (2018)). Due to concerns over risk aggregation, many insurers are currently reluctant to offer substantial limits for cyberattacks. There is thus a growing need for better data and analytics surrounding the underlying risk which will assist the reinsurance market to increase its support towards the cyber insurers. The better cyber insurers are able to measure and monitor their accumulation risk, the more reinsurance coverage they will be able to access which will in turn allow them to provide broader insurance cover, resulting in the overall growth of the market (Camillo (2017)).

Therefore, keeping in mind these striking characteristics and dynamic nature of cyber-risk underwriting cyber insurance requires a profound understanding of both the likelihood of a cyber incident that would demand coverage along with expected financial implications of such incidents (OECD and Marsh & McLennan Companies (2018)). For this, cyber risk data needs to be standardized and probabilistic models need to be developed for higher frequency incident types. In order to have a better understanding of the evolving risk the data needs to be regularly updated and analyzed to identify potential attack patterns.

Indeed, analyzing and modeling the frequency of cyberattacks, plays an important role in assessing and quantifying the cyber liability and risk exposure of a company. Chen et al. (2015) analyzed an extensive dataset of time-dependent frequencies of attacks over 491 consecutive IP addresses, collected for a period of 18 days. These IP addresses were simulated by Honeypot. Honeypot is a research organization that develops open source security tools to investigate cyberattacks. Since these simulated IP addresses did not correspond to any legitimate services, any traffic that arrived at these was considered to be attacks. Due to the complexity of the modern cyberworld one might think that the distribution of cyberattacks is random and so it is difficult to predict them. However, their study revealed that cyberattacks are not so random and that there are spatiotemporal patterns inherent in them (here the term spatio has been used for the IP address space). The authors identified two types of attack patterns - deterministic and stochastic, and used various quantitative measures to characterize them in a comprehensive manner.

They identified blocks of IP addresses sharing similar attack patterns, where the time series of different IP addresses within each block were approximately synchronized and their amplitudes were of the same order of magnitude. Thus, each block exhibited unique spatiotemporal features which points towards the existence of a set of intimately correlated attackers or a single main attacker . Another important inference that can be drawn from regularity in attack patterns of consecutive IP addresses is the way attackers choose their targets, that is targeting each IP within a consecutive IP sector instead of distantly separated IP addresses. These deterministic rules followed by the attacks could be very helpful in predicting the location and time of next attack. In addition to this the stochastic component of cyberattacks was characterized using the flux-fluctuation relation where flux of a victim IP address refers to its attack frequency and fluctuation refers to the standard deviation of flux when the traffic system evolves in the presence of an external drive which in this case was the total number of attacks on a certain IP region. They used this relation to identify and differentiate the patterns of the external drives. For instance, in case of heavily attacked IP region flux fluctuation relation indicated a non-Poisson type of external drive with strong fluctuations. Another way of monitoring the attack behavior of a particular IP address block could be using inference probability of a consecutive IP region which measures how sufficient the information corresponding to one IP address is in capturing the key features of the attack patterns in that whole IP region. A reliable measure for this was constructed using Markov state transition probability matrix for each IP. IP addresses sharing similar attack patterns were found to have high correlation coefficients which points towards the fact that the attack frequencies of a group of IP addresses having the same attack pattern type can be inferred from one of the IPs.

Another important finding of this research was that the attack time series of all IP addresses possessed high degrees of predictability, which was quantified using information entropy. Information entropy gives a measure of the uncertainty associated with the state transitions in the time series of attack frequencies taking into account both the heterogeneous probability distribution in different states and the temporal correlations among the states. The authors found out that while deterministic attack patterns showed higher predictability than stochastic attacks for time resolution of 1000 seconds [1], the opposite was true in case of $\Delta t = 100s$. This finding could be a motivation to develop inference algorithms which can be very helpful in providing global insights into cyber security based on limited information sources. Prediction algorithms can also be designed to analyze current cyberattack data and forecast future cyberattacks.

The observations and results of Chen et al. (2015) can be used to anticipate and consequently mitigate large-scale cyberattacks based on macroscopic approaches. The revelations about the existence of intrinsic patterns of cyberattacks and correlation between the time series within each block could be further exploited to better understand the risk profile of a single or a group of IP addresses in general or in particular risk profile of a company or entity to which these IP addresses correspond.

Böhme and Kataria (2006) introduced a new classification of correlation properties of cyber-risks. The authors discussed two types of cyber risk correlation, one is the correlation of cyber-risks within a firm i.e. correlated failure of multiple systems on its internal network and the other is the correlation in risk at a global level i.e. across independent firms in an insurers portfolio. While internal risk correlation (within a firm) influences firms decision to seek insurance, the global correlation influences insurers decision in setting the premium. Taking into account this potential correlation among time series of attacks on different IP addresses gives better insights when modeling the attack frequency for each of them. This idea, together with the findings of Chen et al. (2015)motivate the present research.

Similar to the case of cyberattacks, time series of counts occur in a wide-variety of real world contexts such as incidence of certain disease, insurance claim counts across time, counts of accidents etc. Given the fame of count data in various research fields, a number of time series models for count data have been proposed previously, accounting for different types of marginal distribution and autocorrelation structure. Due to limitations of usual linear time series processes (ARMA processes) related to multiplication of a real constant by a non-integer random variable, discrete analogues of the usual ones have been proposed where the scalar multiplication is replaced by an

[1] time resolution Δt means the time unit of each time series

integer-valued operator with similar properties, called the thinning operation (Steutel and van Harn (1979)).

Out of the various integer-valued models that have been introduced, the conditionally heteroskedastic models are very useful for modeling count time series due to the presence of heteroskedasticity in most of the count series cases. Ferland et al. (2006)proposed an integer valued analogue of the generalized autoregressive conditional heteroskedastic (GARCH) (p,q) model with Poisson deviates. As an application of this model they used time series data of the number of cases of campylobacterosis infections in the north of the Province of Quebec for a period from January 1990 until the end of October 2000. The authors show that the data exhibits higher variance as the level increases justifying the use of INGARCH model in such cases. Count time series are often overdispersed which means that their variance is larger than the mean. In such cases Poisson distribution has some limitations since its conditional mean is equal to its conditional variance due to which the Poisson INGARCH model can lead to inaccurate results in case of potential extreme observations.

To account for this limitation, Zhu (2011)proposed a negative binomial INGARCH model that can deal with both overdispersion and potential extreme observations. They applied the model to a dataset of incidences of poliomyelitis in the US from 1970 to 1983 which displayed significant overdispersion. On the basis of AIC and BIC the authors found out that both INGARCH(1,1) and INARCH(1) model well describe overdispersion but the more sparsely parameterized INARCH(1) model should be preferred. However, amongst the different extensions of INARCH(1) model the one with negative binomial distribution provided the best fit as compared to Poisson and Double Poisson. Apart from overdispersion in counts time series some of them instead have underdispersion, that is, cases in which the variance is smaller than the mean.

To this end, Zhu (2012) introduced a generalized Poisson INGARCH (GP-INGARCH) model which can be used in cases of overdispersion as well as underdispersion. Binomial distribution b(n,p) is another alternative to model underdispersion but it has a limitation in terms of parameter estimation because the joint MLE of the discrete valued parameter n and other parameters is not possible to obtain. For a similar reason, GP-INGARCH model could perform better than the Negative binomial model as well. The authors demonstrate the usefulness of the proposed model for the case of overdispersion by applying it to a series of annual counts of major earthquakes during the years 1900-2006 and underdispersion by applying it to a dataset of number of different IP addresses registered every 2 minutes at the server of the Department of Statistics of the University of Würzburg on November 29th 2005, between 10 am and 6 pm, a time series of length 241. In the former case, based on AIC and BIC, INGARCH(1,1) models give a better fit as compared to INARCH(1) models while out of the 3 INGARCH(1,1) models GP model performs best as compared to double Poisson and Poisson models. In the latter case, Poisson INAR(1), GP-INARCH(1) and Poisson INARCH(1) perform better than the other specifications and while these three give a good fit of mean and FOAC, only the GP-INARCH(1) model is able to capture the variance dimension of the under-dispersed data. Thus, GP model is able to capture peculiar characteristics of these data and this motivates the modeling framework developed in this thesis.

I introduce new dynamic generalized Poisson panel-data models to analyze jointly the cyberattack series on a set of victim IP addresses. As mentioned before, the generalized Poisson distribution (GP) is very useful in regression count models because of the greater flexibility it provides as compared to Poisson distribution for modeling both under dispersed as well as over dispersed data. Dynamic models of time series, can be very useful in the presence of autocorrelation. Given the nature of the cyberattack data, combining GP distribution with INGARCH model can help account for heteroskedasticity and serial correlation to provide better forecasts. Moreover, in case of panel data, GARCH estimator proves to be parsimonious as it reduces the number of parameters to be estimated owing to the pooling of GARCH parameters.

In this study, I extend the theoretical results on the generalized Poisson INGARCH model introduced by Zhu (2012) to the case of panel data models with partial and complete pooling. Partial pooling refers to the case in which pooling is done only on the model parameters while complete pooling includes pooling on parameters as well as in the intensity of the attacks. In partial pooling, I simply extend GP-INGARCH model of Zhu (2012) to each unit (IP address) in the panel whereas in complete pooling I assume that attacks are correlated across all IP addresses so the number of attacks in each time period is given by the sample mean of attacks taken over all IP addresses for the given period. For each of these two models I consider further two cases. In the first case I model time series of all IP addresses as one whole panel. In the second case I segregate the 491 IP addresses into 3 blocks as identified by Chen et al. (2015) and then model the time series of all IPs falling in a particular block. In this case I believe that modeling IPs which share similar attack patterns together could give better estimates. After the models are estimated I compare them according to their likelihood value, AIC and BIC criteria. Finally, I provide a forecast comparison of the proposed models.

This document is organized as follows. In **Section 2**, a brief review of the GP-INGARCH model of Zhu (2012) is done followed by its extension to the case of panel data with 4 variations. The conditions for the existence and stationarity of the proposed processes are given. **Section 3** discusses Maximum likelihood estimation of the 4 models proposed. **Section 4** gives a real world application of the 4 models. In particular, the 4 models are applied to a dataset of cyber-attacks on 491 consecutive victim IP addresses. The estimated models are compared on the basis of negative log-likelihood, AIC and BIC values. The proposed models are further compared based on their forecasting accuracy using mean absolute error, mean squared error and mean root error. Finally, **Section 5** concludes.

2 Models

2.1 The generalized Poisson INGARCH Model

Let us begin by looking at the generalized Poisson INGARCH model proposed by Zhu (2012). Let X = $(X_t,\ t \in \mathbb{Z})$ be a time series with values in $\mathbb{N}_0$, where $\mathbb{Z}$ is the set of integers and $\mathbb{N}_0$ is the set of natural numbers including 0. For any t let $\mathcal{F}_{t-1}$ be the σ-field generated by $X_{t-j}, j \geq 1$. Suppose that the random variables X_1, X_2 ..., X_n are independent conditional on the past information. The process X is said to satisfy a GP-INGARCH(p, q) model if $\forall t \in \mathbb{Z}$,

$$X_t|\mathcal{F}_{t-1} \sim \mathcal{GP}(\lambda_t^*, \mathcal{K})$$
$$\frac{\lambda_t^*}{1-\mathcal{K}} = \lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j} \tag{2.1}$$

where
$\alpha_0 > 0,\ \alpha_i \geq 0,\ \beta_j \geq 0,\ i = 1, ..., p,\ j = 1, ..., q,\ p \geq 1,\ q \geq 0$
and max(-1,-$\lambda_t^*/4) < \mathcal{K} < 1$.

Here, $\mathcal{GP}(\lambda_t^*, \mathcal{K})$ denotes the generalized Poisson distribution with parameters λ_t^* and $\mathcal{K}$ whose probability mass function is given by

$$P(X = x) = \begin{cases} \lambda_t^*(\lambda_t^* + \mathcal{K}x)^{x-1} e^{-(\lambda_t^* + \mathcal{K}x)}/x! & x = 0, 1, 2, ... \\ 0 & for\ x > m\ if\ \mathcal{K} < 0 \end{cases}$$

where $\lambda_t^* > 0$, max(-1,-$\lambda_t^*/m) < \mathcal{K} < 1$ and m is the largest positive integer for which $\lambda_t^* + \mathcal{K}m > 0$ when $\mathcal{K} < 0$.

The conditional mean and conditional variance of X_t are given as follows:

$$E(X_t|\mathcal{F}_{t-1}) = \frac{\lambda_t^*}{1-\mathcal{K}} = \lambda_t, \quad Var(X_t|\mathcal{F}_{t-1}) = \frac{\lambda_t^*}{(1-\mathcal{K})^3} = \phi^2 \lambda_t \tag{2.2}$$

where $\phi = 1/(1-\mathcal{K})$.

The unconditional mean and variance of X_t are given as follows:

$$E(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}, \quad Var(X_t) = \phi^2 E(X_t) + Var(\lambda_t) \tag{2.3}$$

2.2 Dynamic generalized Poisson panel-data models

In this section I propose 4 models extending (2.1) to the case of panel data. The first two models involve pooling only on the GARCH parameters whereas in the other two models pooling is also done on the X_i's. The former allows the coefficients to vary only across time while the latter also allows interaction among the different units of the panel. Additionally, in the second and fourth model the panel is divided into 2 clusters and then each cluster is modeled separately. Therefore, in these 2 cases the coefficients also vary across clusters.

Let $X = \{X_{it}, i \in 1, ..., N, t \in \mathbb{Z}\}$ be a panel of time series of counts where N is the number of units in the panel. For any $i \in \{1, ..., N\}$ and for any t let $\mathcal{F}_{it-1}$ be the σ-field generated by $\{X_{it-j}, j \geq 1\}$ and $\mathcal{F}_{t-1} = \bigcup_{i=1}^{N} \mathcal{F}_{it-1}$. Suppose that the random variables X_{i1}, X_{i2} ..., X_{in} are independent conditional on the past information.

2.2.1 Partial Pooling

We say that $\forall i \in \{1, ..., N\}$ the process $X_i = \{X_{it}\}_{t\in\mathbb{Z}}$ satisfies a GP-INGARCH(p, q) model with partial pooling if $\forall t \in \mathbb{Z}$,

$$\begin{aligned} X_{it}|\mathcal{F}_{it-1} &\sim \mathcal{GP}(\lambda^*_{it}, \mathcal{K}) \\ \frac{\lambda^*_{it}}{1-\mathcal{K}} = \lambda_{it} &= \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{it-j} + \sum_{k=1}^{q} \beta_k \lambda_{it-k} \end{aligned} \tag{2.4}$$

where
$\alpha_0 > 0$, $\alpha_j \geq 0$, $\beta_k \geq 0$, $j = 1, ..., p$, $k = 1, ..., q$, $p \geq 1$, $q \geq 0$
and max(-1,-$\lambda^*_{it}/4$) $< \mathcal{K} < 1$.

The model (2.4) will be denoted as M1 from now onwards in the document.

2.2.2 Partial pooling with clustering

Here the $X_i's$ $\{i = 1, ..., N\}$ are grouped into different clusters. The clustering can be done on the basis of certain characteristics of the time series. For instance while modeling a panel data of cyberattacks on different IP addresses we can group the IP's into separate clusters on the basis of similarities in attack patterns such as amplitudes of time series. Suppose $A, B \subset \{1, ..., N\}$ such that $A \cap B = \phi$ and $A \cup B = \{1, ..., N\}$. Let $X_A = \{X_{it}, i \in A, t \in \mathbb{Z}\}$ and $X_B = \{X_{it}, i \in B, t \in \mathbb{Z}\}$ be two clusters of X. In this case, $\forall i \in A$ the process $X_i = \{X_{it}\}_{t\in\mathbb{Z}}$ satisfies a GP-INGARCH(p_A, q_A) model with partial pooling if $\forall t \in \mathbb{Z}$,

$$
\begin{aligned}
X_{it}|\mathcal{F}_{it-1} &\sim \mathcal{GP}(\lambda_{it}^{*A}, \mathcal{K}) \\
\frac{\lambda_{it}^{*A}}{1-\mathcal{K}} = \lambda_{it}^{A} &= \alpha_0^A + \sum_{j=1}^{p_A} \alpha_j^A X_{it-j} + \sum_{k=1}^{q_A} \beta_k^A \lambda_{it-k}^A
\end{aligned}
\tag{2.5}
$$

where
$\alpha_0^A > 0$, $\alpha_j^A \geq 0$, $\beta_k^A \geq 0$, $j = 1, ..., p_A$, $k = 1, ..., q_A$, $p_A \geq 1$, $q_A \geq 0$
and max(-1,-$\lambda_t^{*A}/4$) $< \mathcal{K} < 1$.

$\forall i \in B$ the process $X_i = \{X_{it}\}_{t\in\mathbb{Z}}$ can be defined in a similar manner. This model (2.5) will be denoted as M2 henceforth.

2.2.3 Complete pooling

This model assumes that there is interaction between different units of the panel and therefore involves complete pooling which includes pooling on GARCH parameters as well as $X_i's$. In this case, $\forall i \in \{1, ..., N\}$ the process $X_i = \{X_{it}\}_{t\in\mathbb{Z}}$ satisfies a GP-INGARCH(p, q) model with complete pooling if $\forall t \in \mathbb{Z}$,

$$
\begin{aligned}
X_{it}|\mathcal{F}_{t-1} &\sim \mathcal{GP}(\lambda_t^*, \mathcal{K}) \\
\frac{\lambda_t^*}{1-\mathcal{K}} = \lambda_t &= \alpha_0 + \sum_{j=1}^{p} \alpha_j \bar{X}_{t-j} + \sum_{k=1}^{q} \beta_k \lambda_{t-k} \\
\bar{X}_{t-j} &= \frac{1}{N}\sum_{i=1}^{N} X_{it-j}
\end{aligned}
\tag{2.6}
$$

where
$\alpha_0 > 0$, $\alpha_j \geq 0$, $\beta_k \geq 0$, $j = 1, ..., p$, $k = 1, ..., q$, $p \geq 1$, $q \geq 0$
and max(-1,-$\lambda_t^*/4$) $< \mathcal{K} < 1$.

This model (2.6) will be denoted as M3.

2.2.4 Complete pooling with clustering

Similar to M2 (2.5) this model also considers clustering and assumes that there is interaction only among those time series which belong to a particular cluster. In this case $\forall i \in A$ the process $X_i = \{X_{it}\}_{t\in\mathbb{Z}}$ satisfies a GP-INGARCH(p_A, q_A) model with complete pooling if $\forall t \in \mathbb{Z}$,

$$
\begin{aligned}
X_{it}|\mathcal{F}_{t-1} &\sim \mathcal{GP}(\lambda_t^{*A}, \mathcal{K}) \\
\frac{\lambda_t^{*A}}{1-\mathcal{K}} = \lambda_t^A &= \alpha_0^A + \sum_{j=1}^{p_A} \alpha_j^A \bar{X}_{it-j} + \sum_{k=1}^{q_A} \beta_k^A \lambda_{t-k}^A \\
\bar{X}_{t-j} &= \frac{1}{n(A)} \sum_{i \in A} X_{it-j}
\end{aligned}
\tag{2.7}
$$

where
$\alpha_0^A > 0$, $\alpha_j^A \geq 0$, $\beta_k^A \geq 0$, $j = 1, ..., p_A$, $k = 1, ..., q_A$, $p_A \geq 1$, $q_A \geq 0$
and max(-1,-$\lambda_t^{*A}/4$) $< \mathcal{K} < 1$. Here $n(A)$ denotes the cardinality of A.

$\forall i \in B$ the process $X_i = \{X_{it}\}_{t \in \mathbb{Z}}$ can be defined in a similar manner. Model (2.7) will be denoted as M4 in the following document.

2.3 Stationarity properties

The following paragraphs will discuss the stationarity of the models proposed in **Section 2.2**. It is very important to study the stationarity properties of time series models in order to evaluate their stability over time and hence achieve a good forecast.The following two theorems give a sufficient condition for the existence of a stationary GP-INGARCH(p, q) process with partial and complete pooling, i.e., satisfying M1 and M3 respectively. The two theorems are also applicable in case of clustering,i.e., models M2 and M4 because the only difference in these two cases is that we are modeling specific clusters instead of the complete panel as a whole.

2.3.1 Partial pooling

Theorem 1. *If* $\sum\limits_{j=1}^{p} \alpha_j + \sum\limits_{k=1}^{q} \beta_k < 1$, *then there exists a unique stationary process* $\{X_{it}\}_{t \in \mathbb{Z}}$ *that satisfies M1 such that its first two moments are finite.*

Proof. This theorem can be proved using arguments similar to those in Ferland et al. (2006). Let us begin by defining two polynomials C(B) $= 1 - \beta_1 B - \beta_2 B^2 - ... - \beta q B^q$ and $D(B) = \alpha_1 B - \alpha_2 B^2 - ... - \alpha_p B^p$ where B is the back-shift operator.
Let

$$
\begin{aligned}
\lambda_{it} = C^{-1}(B)(\alpha_0 + D(B)X_{it}) &= \alpha_0 C^{-1}(1) + G(B)X_{it} \\
&\textit{where} \\
G(B) = C^{-1}(B)D(B) &= \sum_{j=1}^{\infty} \psi_j B^j
\end{aligned}
\tag{2.8}
$$

Let $\{U_{it}\}_{t \in \mathbb{Z}}$ be a sequence of independent GP random variables with parameters $(\psi_0 = \alpha_0/C(1), \mathcal{K})$. For each $t \in \mathbb{Z}$ and $j \in \mathbb{Z}^+$, let $\{Z_{it,j,k}\}_{k \in \mathbb{Z}^+}$ represent a sequence of independent GP random variables having parameters $(\psi_j, \mathcal{K})$. We assume that all the random variables $U_s, Z_{it,j,k}$ $(s \in \mathbb{Z}, t \in \mathbb{Z}, j \in \mathbb{Z}^+, k \in \mathbb{Z}^+)$ are mutually independent. Define

$$X_{it}^{(n)} = \begin{cases} 0 & n < 0 \\ (1-\mathcal{K})U_{it} & n = 0 \\ (1-\mathcal{K})U_{it} + (1-\mathcal{K}) \sum\limits_{j=1}^{n} \sum\limits_{k=1}^{X_{it-j}^{(n-j)}} Z_{it-j,j,k} & n > 0 \end{cases}$$

Using the thinning operation [2], $X_{it}^{(n)}$ can be written as:

$$X_{it}^{(n)} = (1-\mathcal{K})U_{it} + (1-\mathcal{K}) \sum_{j=1}^{n} \varphi_j^{(t-j)} \circ X_{it-j}^{(n-j)} \quad n > 0$$

where $\varphi_j = \psi j/(1-\mathcal{K})$ is the common mean of the GP random variables $Z_{it,j,k}$. In the above notation $\varphi_j^{(\tau)}\circ$ signifies that the thinning operation involves the sequence $\{Z_{it,j,k}\}_{k\in\mathbb{N}}$ of GP random variables that correspond to time $t = \tau$.We can see that $X_{it}^{(n)}$ is a finite sum of independent GP random variables and therefore its expectation and variance are well-defined.

Using induction we can show that $\mathrm{E}(X_{it}^{(n)})$ doesn't depend on t, it only depends on n and we can denote it by μ_n. For the proof refer to Appendix section A.1.2. Since $\mu_j = 0$ for $j < 0$,
for $n > 0$,

$$\begin{aligned} \mu_n = \psi_0 + \sum_{r=1}^{\infty} \psi_r \mu_{n-r} = C^{-1}(B)(\alpha_0 + D(B)\mu_n) \\ \iff [C(B) - D(B)]\mu_n = \alpha_0 \iff K(B)\mu_n = \alpha_0 \end{aligned} \tag{2.9}$$

Therefore, $\forall i \in \{1, ..., N\}, \{(X_{it}^{(n)} : t \in \mathbb{Z}, n \in \mathbb{Z}^+\}$ is a sequence of first order stationary processes because the characteristic polynomial $K(z)$ has all its roots outside the unit circle using the fact $\sum\limits_{j=1}^{p} \alpha_j + \sum\limits_{k=1}^{q} \beta_k < 1$.
From this stationarity we get

$$\begin{aligned} \lim_{n\to\infty} \mu_n = \frac{\psi_0}{1 - \sum_{r=1}^{\infty} \psi_r} = \frac{\alpha_0 C^{-1}(1)}{1 - G(1)} = \frac{\alpha_0}{C(1) - D(1)} \\ = \frac{\alpha_0}{K(1)} = \frac{\alpha_0}{1 - \sum_{j=1}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k} = \mu \end{aligned} \tag{2.10}$$

Note that for a fixed value of t, the sequence $\{X_{it}^{(n)}\}_{n\in\mathbb{Z}^+}$ is a non-decreasing sequence of non-negative integer valued random variables. This can be shown using

[2] For definition see Appendix section A.1.1

induction with respect to n as

$$X_{it}^{(n+1)} - X_{it}^{(n)} = (1-\mathcal{K})\left[\sum_{k=1}^{U_{t-n-1}} Z_{it-n-1,n+1,k} + \sum_{j=1}^{n} \sum_{k=X_{it-j}^{(n-j)}+1}^{X_{it-j}^{(n+1-j)}} Z_{it-j,j,k}\right] \geq 0$$

Using this property of the sequence $\{X_{it}^{(n)}\}_{n\in\mathbb{Z}}$ and following Propositon 2 of Ferland et al. (2006) it can be shown that the sequence $\{X_{it}^{(n)}\}_{n\in\mathbb{N}}$ has an almost sure limit X_{it}. Please refer to Appendix section A.1.3 for proof.
Moreover, using arguments similar to those in proposition 3 of Ferland et al. (2006) the sequence $\{X_{it}^{(n)} : t \in \mathbb{Z}, n \in \mathbb{Z}^+\}$ is a strictly stationary process for each n. Therefore, we can conclude that X_{it} is a strictly stationary process.
Using 2.10 and Beppo Levi's theorem we get that

$$\mu = \lim_{n\to\infty} \mu_n = \lim_{n\to\infty} E(X_{it}^{(n)}) = E(X_{it}) \tag{2.11}$$

and we conclude that the first moment of X_{it} is finite.

To show that the second moment of X_{it} is finite it is enough to show that $E[(X_{it}^{(n)})^2]$ is bounded.

$$E\left[\left(X_{it}^{(n)}\right)^2\right] \leq (1-\mathcal{K})^2\left[E(U_{it}^2) + (2E(U_{it})+1)\sum_{j=1}^{n}\varphi_j E(X_{it}) + \left(\sum_{j=1}^{n}\varphi_j\right)^2 E\left[\left(X_{it}^{(n)}\right)^2\right]\right]$$

$$\begin{aligned} E\left[\left(X_{it}^{(n)}\right)^2\right] &\leq \frac{(1-\mathcal{K})^2 E(U_{it}^2) + (1-\mathcal{K})E(X_{it})(2E(U_{it})+1)\sum_{j=1}^{n}\psi_j}{1-(\sum_{j=1}^{n}\psi_j)^2} \\ &\leq \frac{(1-\mathcal{K})^2 E(U_{it}^2) + (1-\mathcal{K})E(X_{it})(2E(U_{it})+1)\sum_{j=1}^{\infty}\psi_j}{1-(\sum_{j=1}^{\infty}\psi_j)^2} \equiv A \end{aligned}$$

Using Lebesgue's dominated convergence theorem, we can conclude that $E(X_{it}^2) \leq A$. With this we get that the second moment of X_{it} is also finite.

Next step is to verify the distributional properties of X_{it}. The proof is available in Appendix section A.1.4. Using arguments similar to those in Section 2.6 of Ferland et al. (2006) and defining

$$r_{it}^{(n)} = (1-\mathcal{K})U_{it} + (1-\mathcal{K})\sum_{j=1}^{n}\sum_{k=1}^{X_{it-j}} Z_{it-j,j,k}$$

we can show that

$$X_{it}|\mathcal{F}_{it-1} \sim \mathcal{GP}(\lambda^*_{it}, \mathcal{K})$$
$$\text{where } \lambda^*_{it} = (\alpha_0 C^{-1}(1) + \sum_{j=1}^{n} \psi_j X_{it-j})(1-\mathcal{K})$$

■

The conditional expectation and conditional variance of X_{it} are given by

$$E(X_{it}|\mathcal{F}_{it-1}) = \frac{\lambda^*_{it}}{1-\mathcal{K}} = \lambda_{it}$$
$$Var(X_{it}|\mathcal{F}_{it-1}) = \frac{\lambda^*_{it}}{(1-\mathcal{K})^3} = \phi^2\lambda_{it} \text{ where } \phi = 1/(1-\mathcal{K})$$

Also, from (2.10) and (2.11) we get the unconditional mean and variance as follows,

$$\begin{aligned} E(X_{it}) = \mu &= \frac{\alpha_0}{1-\sum_{j=1}^{p}\alpha_j + \sum_{k=1}^{q}\beta_k} \\ Var(X_{it}) &= E(Var(X_{it}|\mathcal{F}_{it-1})) + Var(E(X_{it})|\mathcal{F}_{it-1}) \\ &= E(\phi^2\lambda_{it}) + Var(\lambda_{it}) \\ &= \phi^2\mu + Var(\lambda_{it}) \end{aligned}$$

2.3.2 Complete pooling

Theorem 2. *If* $\sum_{j=1}^{p}\alpha_j + \sum_{k=1}^{q}\beta_k < 1$, *then there exists a unique stationary process* $\{X_{it}\}_{t\in\mathbb{Z}}$ *that satisfies M3 such that its first two moments are finite.*

Proof. This theorem can be proved by following the same steps as in the proof of Theorem 1. I begin by constructing the INGARCH process with successive approximations. Define two polynomials C(B) $= 1 - \beta_1 B - \beta_2 B^2 - ... - \beta q B^q$ and $D(B) = \alpha_1 B - \alpha_2 B^2 - ... - \alpha_p B^p$ where B is the back-shift operator.
Let

$$\begin{aligned} &\lambda_t = C^{-1}(B)(\alpha_0 + D(B)\bar{X}_t) = \alpha_0 C^{-1}(1) + G(B)\bar{X}_t \\ &\text{where} \\ &G(B) = C^{-1}(B)D(B) = \sum_{j=1}^{\infty}\psi_j B^j \end{aligned} \tag{2.12}$$

Let the sequences $\{U_{it}\}_{t\in\mathbb{Z}}$ and $\{Z_{it,j,k}\}_{k\in\mathbb{Z}^+}$be defined as in Theorem 1 such that all the random variables $U_s, Z_{it,j,k}$ $(s \in \mathbb{Z}, t \in \mathbb{Z}, j \in \mathbb{Z}^+, k \in \mathbb{Z}^+)$ are mutually independent. Consider the case when $N = 2$, i.e., there are only two units in the panel. The proof can be easily extended to $N > 2$. Define a bivariate sequence $\{X_t^{(n)}\}_{n\in\mathbb{Z}^+} = \{(X_{1t}^{(n)}, X_{2t}^{(n)})\}_{n\in\mathbb{Z}^+}$ where

$$X_{1t}^{(n)} = \begin{cases} 0 & n < 0 \\ (1-\mathcal{K})U_{1t} & n = 0 \\ (1-\mathcal{K})U_{1t} + (1-\mathcal{K})\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{X_{1t-j}^{(n-j)}+X_{2t-j}^{(n-j)}} Z_{1t-j,j,k} & n > 0 \end{cases}$$

$$X_{2t}^{(n)} = \begin{cases} 0 & n < 0 \\ (1-\mathcal{K})U_{2t} & n = 0 \\ (1-\mathcal{K})U_{2t} + (1-\mathcal{K})\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{X_{1t-j}^{(n-j)}+X_{2t-j}^{(n-j)}} Z_{2t-j,j,k} & n > 0 \end{cases}$$

Using the thinning operation , $X_{it}^{(n)}(i = 1, 2)$ can be written as:

$$X_{it}^{(n)} = (1-\mathcal{K})U_{it} + (1-\mathcal{K})\frac{1}{2}\sum_{j=1}^{n}\varphi_j^{(t-j)} \circ (X_{1t-j}^{(n-j)} + X_{2t-j}^{(n-j)}) \quad n > 0$$

where $\varphi_j = \psi j/(1-\mathcal{K})$ is the common mean of the GP random variables $Z_{it,j,k}$.

The next step is to show that $\{X_t^{(n)}\}_{n\in\mathbb{Z}^+}$ is a sequence of first order stationary processes. As before we can see that $X_{it}^{(n)}$ is a finite sum of independent GP random variables and therefore its expectation and variance are well-defined. Moreover, $\mathrm{E}(X_{it}^{(n)})$ doesn't depend on t, it only depends on n and we can denote it by μ_n. Since $\mu_j = 0$ for $j < 0$,
for $n > 0$,

$$\begin{aligned} \mu_n = \psi_0 + \frac{2}{2}\sum_{r=1}^{\infty}\psi_r\mu_{n-r} = C^{-1}(B)(\alpha_0 + D(B)\mu_n) \\ \Longleftrightarrow [C(B) - D(B)]\mu_n = \alpha_0 \iff K(B)\mu_n = \alpha_0 \end{aligned} \tag{2.13}$$

Therefore, $\{X_t^{(n)} : t \in \mathbb{Z}, n \in \mathbb{Z}^+\} = \{(X_{1t}^{(n)}, X_{2t}^{(n)})\}_{n\in\mathbb{Z}^+}$is a sequence of first order stationary processes because the characteristic polynomial $K(z)$ has all its roots outside the unit circle using the fact $\sum_{j=1}^{p}\alpha_j + \sum_{k=1}^{q}\beta_k < 1$.

From this stationarity we get

$$\begin{aligned}\lim_{n\to\infty}\mu_n &= \frac{\psi_0}{1-\sum_{r=1}^{\infty}\psi_r} = \frac{\alpha_0 C^{-1}(1)}{1-G(1)} = \frac{\alpha_0}{C(1)-D(1)} \\ &= \frac{\alpha_0}{K(1)} = \frac{\alpha_0}{1-\sum_{j=1}^{p}\alpha_j+\sum_{k=1}^{q}\beta_k} = \mu\end{aligned} \tag{2.14}$$

Using arguments similar to those in the proof of Theorem 1 we can show that $\{(X_{1t}^{(n)}, X_{2t}^{(n)})\}_{n\in\mathbb{Z}^+}$ has an almost sure limit $X_t, t \in \mathbb{Z} = (X_{1t}, X_{2t})_{t\in\mathbb{Z}}$ and is a strictly stationary process for each n. Therefore, $\{X_t\}_{t\in\mathbb{Z}}$ is a strictly stationary process.

Using (2.14) and Beppo Levi's theorem we get that

$$\mu = \lim_{n\to\infty}\mu_n = \lim_{n\to\infty} E(X_{it}^{(n)}) = E(X_{it}) \tag{2.15}$$

and we conclude that the first moment of X_{it} is finite and hence that of $X_t = (X_{1t}, X_{2t})$ is finite.

Next, to show that the second moment of X_t is also finite, I will show that $E[(X_{1t}^{(n)} + X_{2t}^{(n)})^2]$ is bounded.

$$\begin{aligned}E\left[(X_{1t}^{(n)}+X_{2t}^{(n)})^2\right] \le (1-\mathcal{K})^2 \Bigg[& E\left[(U_{1t}+U_{2t})^2\right] + (2E(U_{1t}+U_{2t})+1)\sum_{j=1}^{n}\varphi_j E(X_{1t}+X_{2t}) \\ &+ \left(\sum_{j=1}^{n}\varphi_j\right)^2 E\left[(X_{1t}^{(n)}+X_{2t}^{(n)})^2\right]\Bigg]\end{aligned}$$

$$\begin{aligned}&\Rightarrow E\left[(X_{1t}^{(n)}+X_{2t}^{(n)})^2\right] \\ &\le \frac{(1-\mathcal{K})^2 E\left[(U_{1t}+U_{2t})^2\right] + (1-\mathcal{K})E(X_{1t}+X_{2t})(2E(U_{1t}+U_{2t})+1)\sum_{j=1}^{n}\psi_j}{1-(\sum_{j=1}^{n}\psi_j)^2} \\ &\le \frac{(1-\mathcal{K})^2 E\left[(U_{1t}+U_{2t})^2\right] + (1-\mathcal{K})E(X_{1t}+X_{2t})(2E(U_{1t}+U_{2t})+1)\sum_{j=1}^{\infty}\psi_j}{1-(\sum_{j=1}^{\infty}\psi_j)^2} \\ &\equiv A\end{aligned}$$

Using Lebesgue's dominated convergence theorem, we can conclude that $E[(X_{1t} + X_{2t})^2] \le A$. With this we get that the second moment of X_t is also finite.

We now have to verify the distributional properties of $X_t = (X_1, X_2)$.The proof is given in Appendix section A.1.5. Following the steps as in the proof of Theorem 1 and defining the sequence $\{(r_{1t}^{(n)}, r_{2t}^{(n)})\}$ where

$$r_{it}^{(n)} = (1-\mathcal{K})U_{it} + (1-\mathcal{K})\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{X_{1t-j}+X_{2t-j}} Z_{it-j,j,k} \text{ for } i = 1, 2$$

we can show that

$$X_t|\mathcal{F}_{t-1} = (X_{1t}, X_{2t})|\mathcal{F}_{t-1} \sim \mathcal{GP}(\lambda_t^*, \mathcal{K})$$
$$\text{where } \lambda_t^* = \left(\alpha_0 C^{-1}(1) + \sum_{j=1}^{n}\psi_j \bar{X}_{t-j}\right)(1-\mathcal{K})$$

■

The conditional expectation and conditional variance of X_{it} are given by

$$E(X_{it}|\mathcal{F}_{t-1}) = \frac{\lambda_t^*}{1-\mathcal{K}} = \lambda_t$$
$$Var(X_{it}|\mathcal{F}_{t-1}) = \frac{\lambda_t^*}{(1-\mathcal{K})^3} = \phi^2\lambda_t \text{ where } \phi = 1/(1-\mathcal{K})$$

Also, from (2.14) and (2.15) we get that the unconditional mean and variance are given by,

$$\begin{aligned} E(X_{it}) = \mu &= \frac{\alpha_0}{1-\sum_{j=1}^{p}\alpha_j + \sum_{k=1}^{q}\beta_k} \\ Var(X_{it}) &= E(Var(X_{it}|\mathcal{F}_{t-1})) + Var(E(X_{it})|\mathcal{F}_{t-1}) \\ &= E(\phi^2\lambda_t) + Var(\lambda_t) \\ &= \phi^2\mu + Var(\lambda_t) \end{aligned}$$

3 Estimation

3.1 Maximum Likelihood Estimation

This section gives the MLE for models M1 and M3. The same formulas are applicable for M2 and M4 because the only difference is that in M1 and M3 we are taking into account panel as a whole whereas in M2 and M4 we consider each block of units within the panel separately. For models M2 and M4 the formula for the conditional likelihood function for X can be obtained by multiplying the conditional likelihood functions of all the blocks into which the panel has been divided.

3.1.1 Partial pooling

Let the parameter vector be $\Theta = (\alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, \phi)^T = (\theta, \phi)^T$ where $\phi = 1/1 - \mathcal{K}$
For any $i \in \{1, ..., N\}$ let $X_i = (X_{i1}, X_{i2}, ..., X_{iT})$ be generated by model M1 with the true parameter value Θ_0. Let X $= (X_1, X_2, ..., X_N)$ be the collection of $X_i's$. Then for any $i \in \{1, ..., N\}$, the conditional likelihood function is given by

$$L_i(\Theta) = \prod_{t=p+1}^{T} \frac{\lambda_{it}\left[\lambda_{it} + (\phi - 1)X_{it}\right]^{X_{it}-1} \phi^{-X_{it}} exp\{-\left[\lambda_{it} + (\phi - 1)X_{it}\right]/\phi\}}{X_{it}!}$$

where $\lambda_{it} = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{it-j} + \sum_{k=1}^{q} \beta_k \lambda_{it-k}$
and the conditional likelihood function for X is given by

$$L(\Theta) = \prod_{i=1}^{N} \prod_{t=p+1}^{T} \frac{\lambda_{it}\left[\lambda_{it} + (\phi - 1)X_{it}\right]^{X_{it}-1} \phi^{-X_{it}} exp\{-\left[\lambda_{it} + (\phi - 1)X_{it}\right]/\phi\}}{X_{it}!}$$

The log-likelihood is given by

$$\begin{aligned}\mathcal{L}(\Theta) = lnL(\Theta) &= \sum_{i=1}^{N} \sum_{t=p+1}^{T} \Big\{ ln\lambda_{it} + (X_{it} - 1)ln[\lambda_{it} + (\phi - 1)X_{it}] \\ &- X_{it}ln\phi - \frac{\lambda_{it} + (\phi - 1)X_{it}}{\phi} - ln(X_{it}!)\Big\} \\ &= \sum_{i=1}^{N} \sum_{t=p+1}^{T} l_{it}(\Theta)\end{aligned}$$

The score function is given by

$$
\begin{aligned}
\mathcal{S}_{iT}(\Theta) &= \frac{\partial \mathcal{L}(\Theta)}{\partial \Theta} = \sum_{i=1}^{N} \sum_{t=p+1}^{T} \frac{\partial l_{it}(\Theta)}{\partial \Theta} \\
\frac{\partial l_{it}(\Theta)}{\partial \phi} &= \frac{X_{it}(X_{it}-1)}{\lambda_{it}+(\phi-1)X_{it}} - \frac{X_{it}}{\phi} - \frac{X_{it}-\lambda_{it}}{\phi^2} \\
\frac{\partial l_{it}(\Theta)}{\partial \theta} &= \left(\frac{X_{it}-1}{\lambda_{it}+(\phi-1)X_{it}} - \frac{1}{\lambda_{it}} - \frac{1}{\phi} \right) \frac{\partial \lambda_{it}}{\partial \theta} \\
\frac{\partial \lambda_{it}}{\partial \alpha_0} &= 1 + \sum_{k=1}^{q} \beta_k \frac{\partial \lambda_{it-k}}{\partial \alpha_0} \\
\frac{\partial \lambda_{it}}{\partial \alpha_j} &= X_{it-j} + \sum_{k=1}^{q} \beta_k \frac{\partial \lambda_{it-k}}{\partial \alpha_j} \quad \text{for } j = 1, ..., p \\
\frac{\partial \lambda_{it}}{\partial \beta_k} &= \lambda_{it-k} + \sum_{r=1}^{q} \beta_r \frac{\partial \lambda_{it-r}}{\partial \beta_k} \quad \text{for } k = 1, ..., q
\end{aligned}
$$

The conditional maximum likelihood estimator of Θ, denoted my $\hat{\Theta}_{MLE}$is given by the solution of the equation $\mathcal{S}_{iT}(\Theta) = 0$.

The Hessian matrix is given by

$$
\mathcal{H}_T(\Theta) = -\sum_{i=1}^{N} \sum_{t=p+1}^{T} \frac{\partial^2 l_{it}(\Theta)}{\partial \Theta \partial \Theta^T} \tag{3.1}
$$

$$
\begin{aligned}
\frac{\partial^2 l_{it}(\Theta)}{\partial \phi^2} &= -\frac{X_{it}^2(X_{it}-1)}{[\lambda_{it}+(\phi-1)X_{it}]^2} + \frac{X_{it}}{\phi^2} - \frac{2(X_{it}-\lambda_{it})}{\phi^3} \\
\frac{\partial^2 l_{it}(\Theta)}{\partial \phi \partial \theta} &= -\left(\frac{(X_{it}-1)(X_{it})}{[\lambda_{it}+(\phi-1)X_{it}]^2} - \frac{1}{\phi^2} \right) \frac{\partial \lambda_{it}}{\partial \theta} \\
\frac{\partial^2 l_{it}(\Theta)}{\partial \theta \partial \theta^T} &= -\left(\frac{X_{it}-1}{[\lambda_{it}+(\phi-1)X_{it}]^2} + \frac{1}{\lambda_{it}^2} \right) \frac{\partial \lambda_{it}}{\partial \theta} \frac{\partial \lambda_{it}}{\partial \theta^T} \\
&\quad + \left(\frac{X_{it}-1}{\lambda_{it}+(\phi-1)X_{it}} + \frac{1}{\lambda_{it}} - \frac{1}{\phi} \right) \frac{\partial^2 \lambda_{it}}{\partial \theta \partial \theta^T} \\
\frac{\partial^2 \lambda_{it}}{\partial \alpha_0^2} &= 0 \qquad \frac{\partial^2 \lambda_{it}}{\partial \alpha_0 \partial \alpha_j} = 0 \quad \text{for } j = 1, ..., p \\
\frac{\partial^2 \lambda_{it}}{\partial \alpha_u \partial \alpha_j} &= 0 \quad \text{for } u, j = 1, ..., p \\
\frac{\partial^2 \lambda_{it}}{\partial \alpha_0 \partial \beta_k} &= \frac{\partial \lambda_{it-k}}{\partial \alpha_0} + \sum_{r=1}^{q} \beta_r \frac{\partial^2 \lambda_{it-r}}{\partial \alpha_0 \partial \beta_k} \quad \text{for } k = 1, ..., q
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2\lambda_{it}}{\partial\alpha_j\partial\beta_k} &= \frac{\partial\lambda_{it-k}}{\partial\alpha_j} + \sum_{r=1}^{q}\beta_r\frac{\partial^2\lambda_{it-r}}{\partial\alpha_j\partial\beta_k} \quad \text{for } j=1,...,p\,;\ k=1,...,q \\
\frac{\partial^2\lambda_{it}}{\partial^2\beta_k} &= \frac{2\partial\lambda_{it-k}}{\partial\beta_k} + \sum_{r=1}^{q}\beta_r\frac{\partial^2\lambda_{it-r}}{\partial\beta_k^2} \quad \text{for } k=1,...,q \\
\frac{\partial^2\lambda_{it}}{\partial\beta_v\partial\beta_k} &= \frac{\partial\lambda_{it-k}}{\partial\beta_v} + \frac{\partial\lambda_{it-v}}{\partial\beta_k} + \sum_{r=1}^{q}\beta_r\frac{\partial^2\lambda_{it-r}}{\partial\beta_v\partial\beta_k} \quad \text{for } v\neq k,\ \ v,k=1,...,q
\end{aligned}
$$

Using arguments in Bollerslev (1986), the maximum likelihood estimator $\hat{\Theta}_{MLE}$ is asymptotically normal with mean Θ_0 and as White (1982) suggested covariance matrix of $\hat{\Theta}_{MLE}$ is given by

$$
\begin{aligned}
&\mathcal{H}_T^{-1}(\hat{\Theta}_{MLE})S_T(\hat{\Theta}_{MLE})\mathcal{H}_T^{-1}(\hat{\Theta}_{MLE}) \\
&\text{where}\ \ S_T(\Theta) = \sum_{i=1}^{N}\sum_{t=p+1}^{T}\frac{\partial l_{it}}{\partial\Theta}\frac{\partial l_{it}}{\partial\Theta^T} \ \ \text{and } \mathcal{H}_T(\Theta) \text{ is given by (3.1)}
\end{aligned}
$$

3.1.2 Complete pooling

Let the parameter vector be defined as before $\Theta = (\alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, \phi)^T = (\theta, \phi)^T$ where $\phi = 1/1 - \mathcal{K}$

For any $i \in \{1, ..., N\}$ let $X_i = (X_{i1}, X_{i2}, ..., X_{iT})$ be generated by model M3 with the true parameter value Θ_0. Let X $= (X_1, X_2, ..., X_N)$ be the collection of $X_i's$. Then for any $i \in \{1, ..., N\}$, the conditional likelihood function is given by

$$
L_i(\Theta) = \prod_{t=p+1}^{T}\frac{\lambda_t\left[\lambda_t + (\phi-1)X_{it}\right]^{X_{it}-1}\phi^{-X_{it}}exp\{-\left[\lambda_t + (\phi-1)X_{it}\right]/\phi\}}{X_{it}!}
$$

where $\lambda_t = \alpha_0 + \sum\limits_{j=1}^{p}\alpha_j\bar{X}_{t-j} + \sum\limits_{k=1}^{q}\beta_k\lambda_{t-k}$

and the conditional likelihood function for X is given by

$$
L(\Theta) = \prod_{i=1}^{N}\prod_{t=p+1}^{T}\frac{\lambda_t\left[\lambda_t + (\phi-1)X_{it}\right]^{X_{it}-1}\phi^{-X_{it}}exp\{-\left[\lambda_t + (\phi-1)X_{it}\right]/\phi\}}{X_{it}!}
$$

The log-likelihood is given by

$$
\begin{aligned}
\mathcal{L}(\Theta) = lnL(\Theta) = \sum_{i=1}^{N}\sum_{t=p+1}^{T}\Big\{&ln\lambda_t + (X_{it}-1)ln[\lambda_t + (\phi-1)X_{it}] \\
&- X_{it}ln\phi - \frac{\lambda_t + (\phi-1)X_{it}}{\phi} - ln(X_{it}!)\Big\}
\end{aligned}
$$

$$= \sum_{i=1}^{N} \sum_{t=p+1}^{T} l_{it}(\Theta)$$

The score function is given by

$$\begin{aligned}
\mathcal{S}_{iT}(\Theta) &= \frac{\partial \mathcal{L}(\Theta)}{\partial \Theta} = \sum_{i=1}^{N} \sum_{t=p+1}^{T} \frac{\partial l_{it}(\Theta)}{\partial \Theta} \\
\frac{\partial l_{it}(\Theta)}{\partial \phi} &= \frac{X_{it}(X_{it}-1)}{\lambda_t + (\phi - 1)X_{it}} - \frac{X_{it}}{\phi} - \frac{X_{it} - \lambda_t}{\phi^2} \\
\frac{\partial l_{it}(\Theta)}{\partial \theta} &= \left(\frac{X_{it} - 1}{\lambda_t + (\phi - 1)X_{it}} - \frac{1}{\lambda_t} - \frac{1}{\phi} \right) \frac{\partial \lambda_t}{\partial \theta} \\
\frac{\partial \lambda_t}{\partial \alpha_0} &= 1 + \sum_{k=1}^{q} \beta_k \frac{\partial \lambda_{t-k}}{\partial \alpha_0} \\
\frac{\partial \lambda_t}{\partial \alpha_j} &= \bar{X}_{t-j} + \sum_{k=1}^{q} \beta_k \frac{\partial \lambda_{t-k}}{\partial \alpha_j} \quad \text{for } j = 1, ..., p \\
\frac{\partial \lambda_t}{\partial \beta_k} &= \lambda_{t-k} + \sum_{r=1}^{q} \beta_r \frac{\partial \lambda_{t-r}}{\partial \beta_k} \quad \text{for } k = 1, ..., q
\end{aligned}$$

The conditional maximum likelihood estimator of Θ, denoted my $\hat{\Theta}_{MLE}$is given by the solution of the equation $\mathcal{S}_{iT}(\Theta) = 0$.
The Hessian matrix is given by

$$\mathcal{H}_T(\Theta) = -\sum_{i=1}^{N} \sum_{t=p+1}^{T} \frac{\partial^2 l_{it}(\Theta)}{\partial \Theta \partial \Theta^T} \tag{3.2}$$

$$\begin{aligned}
\frac{\partial^2 l_{it}(\Theta)}{\partial \phi^2} &= -\frac{X_{it}^2(X_{it}-1)}{[\lambda_t + (\phi-1)X_{it}]^2} + \frac{X_{it}}{\phi^2} - \frac{2(X_{it} - \lambda_t)}{\phi^3} \\
\frac{\partial^2 l_{it}(\Theta)}{\partial \phi \partial \theta} &= -\left(\frac{(X_{it}-1)(X_{it})}{[\lambda_t + (\phi-1)X_{it}]^2} - \frac{1}{\phi^2} \right) \frac{\partial \lambda_t}{\partial \theta} \\
\frac{\partial^2 l_{it}(\Theta)}{\partial \theta \partial \theta^T} &= -\left(\frac{X_{it}-1}{[\lambda_t + (\phi-1)X_{it}]^2} + \frac{1}{\lambda_t^2} \right) \frac{\partial \lambda_t}{\partial \theta} \frac{\partial \lambda_t}{\partial \theta^T} \\
&\quad + \left(\frac{X_{it}-1}{\lambda_t + (\phi-1)X_{it}} - \frac{1}{\lambda_t} - \frac{1}{\phi} \right) \frac{\partial^2 \lambda_t}{\partial \theta \partial \theta^T} \\
\frac{\partial^2 \lambda_t}{\partial \alpha_0^2} &= 0 \qquad \frac{\partial^2 \lambda_t}{\partial \alpha_0 \partial \alpha_j} = 0 \quad \text{for } j = 1, ..., p \\
\frac{\partial^2 \lambda_t}{\partial \alpha_u \partial \alpha_j} &= 0 \qquad \text{for } u, j = 1, ..., p \\
\frac{\partial^2 \lambda_t}{\partial \alpha_0 \partial \beta_k} &= \frac{\partial \lambda_{t-k}}{\partial \alpha_0} + \sum_{r=1}^{q} \beta_r \frac{\partial^2 \lambda_{t-r}}{\partial \alpha_0 \partial \beta_k} \quad \text{for } k = 1, ..., q
\end{aligned}$$

$$\frac{\partial^2\lambda_t}{\partial\alpha_j\partial\beta_k} = \frac{\partial\lambda_{t-k}}{\partial\alpha_j} + \sum_{r=1}^{q}\beta_r\frac{\partial^2\lambda_{t-r}}{\partial\alpha_j\partial\beta_k} \quad \text{for } j=1,...,p\,; \quad k=1,...,q$$
$$\frac{\partial^2\lambda_t}{\partial^2\beta_k} = \frac{2\partial\lambda_{t-k}}{\partial\beta_k} + \sum_{r=1}^{q}\beta_r\frac{\partial^2\lambda_{t-r}}{\partial\beta_k^2} \quad \text{for } k=1,...,q$$
$$\frac{\partial^2\lambda_t}{\partial\beta_v\partial\beta_k} = \frac{\partial\lambda_{t-k}}{\partial\beta_v} + \frac{\partial\lambda_{t-v}}{\partial\beta_k} + \sum_{r=1}^{q}\beta_r\frac{\partial^2\lambda_{t-r}}{\partial\beta_v\partial\beta_k} \quad \text{for } v\neq k,\ v,k=1,...,q$$

Using arguments in Bollerslev (1986), the maximum likelihood estimator $\hat{\Theta}_{MLE}$ is asymptotically normal with mean Θ_0 and according to White (1982) covariance matrix of $\hat{\Theta}_{MLE}$ is given by

$$\mathcal{H}_T^{-1}(\hat{\Theta}_{MLE})S_T(\hat{\Theta}_{MLE})\mathcal{H}_T^{-1}(\hat{\Theta}_{MLE})$$
$$\text{where } \ S_T(\Theta) = \sum_{i=1}^{N}\sum_{t=p+1}^{T}\frac{\partial l_{it}}{\partial\Theta}\frac{\partial l_{it}}{\partial\Theta^T} \quad \text{and } \mathcal{H}_T(\Theta) \text{ is given by (3.2)}$$

3.2 Lag selection for GP-INGARCH panel-data models

Selecting appropriate lag length for the INGARCH(p,q) process, that is, the values of p and q, is an important part of estimation especially to get better forecasts. To do this, each of the four models proposed in **Section 2** can be estimated for different values of p and q and then the different estimates of a model can be compared using information criteria such as AIC and BIC. AIC (Akaike's information Criterion) and BIC (Bayesian information criterion) provide a good trade-off between fit and complexity while comparing two models.
They are defined as follows,

$$AIC = -2log(L(\Theta_{MLE})) + 2p$$
$$BIC = -2log(L(\Theta_{MLE})) + log(n)p$$

where p is the number of parameters and n is the sample size.
For both criteria, the preferred model is the one for which the corresponding criterion (AIC/BIC) value is minimum. The terms $2p$ and $log(n)p$ in the formula for AIC and BIC respectively are called complexity penalties since they penalize the model with a large number of parameters.

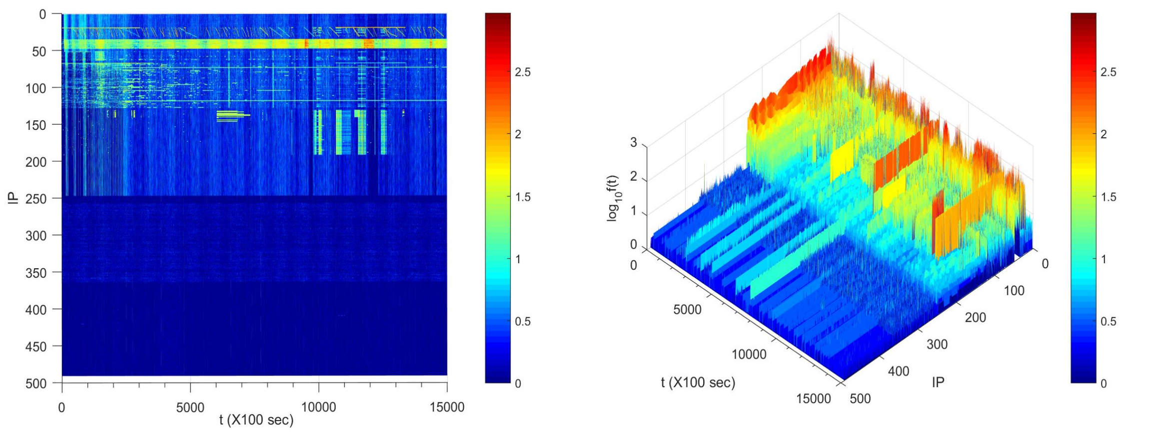
4 Application

Modelling Cyberattack Frequency

This section illustrates the applicability of the dynamic generalised Poisson panel-data models proposed in **Section 2** to real world data.

4.1 Brief description of the data

In particular I will use the dataset of cyberattacks on 491 consecutive victim IP addresses which has been previously analyzed by Chen et al. (2015) to identify different attack patterns across time. The raw dataset contained two columns - IP address and the time (in seconds) at which it was attacked. I aggregated and organized the raw data into a structure where each row represents an IP and each column represents a time period of 100 seconds, in total 15,155 periods accounting for approximately 18 days of data. Thus, each cell in the data structure corresponds to the total number of attacks received by the corresponding IP during the corresponding time period.



(a) **(b)**

Figure 1: Time series of attack frequency f(t) for all IP addresses.

Let f(t) denote the number of attacks received by a victim IP address per unit time $\Delta t = 100$ seconds, called the attack frequency. The time series f(t) for all IP addresses has been plotted in Figure 1 where the x-axis represents time t and y-axis represents IP address index from 1 to 491. Figures (1a) and (1b) show two-dimensional and three-dimensional representations of f(t) for the entire IP-space on a logarithmic scale, respectively. Based on similarities in attack patterns as shown by distinct colored blocks, the three IP regions - (1-246), (247-363) and (364-491) as identified by

Chen et al. (2015), can be easily spotted from the figure. The IP region 1-246 is identified by the acqua background, which is further overlaid by 4 particular patterns corresponding to IP regions (19-31,35-47,50-130 and 131-191). The IP region 247-363 is marked by horizontal blue lines which are overlaid on the dark-blue background lying under the entire IP-space. Due to logarithmic scaling to the base 10 the walls of length 1 are not visible in IP region 364-491. However, walls of higher length can be observed in Figure (1b) which correspond to the light-blue vertical lines overlaid on the dark-blue background in Figure (1a). From Figure (1b) it can be seen that the amplitudes of the time series within each block have approximately same magnitude, but the amplitudes from different blocks vary considerably. Having identified the three major IP regions, (1-246), (247-363) and (364-491), within the IP-space, I will consider these as 3 separate clusters of IPs while applying model M2 and M4 to this dataset. It would be very interesting to see how similarity in attack patterns within a block affects the parameter estimates and finally the forecast accuracy.

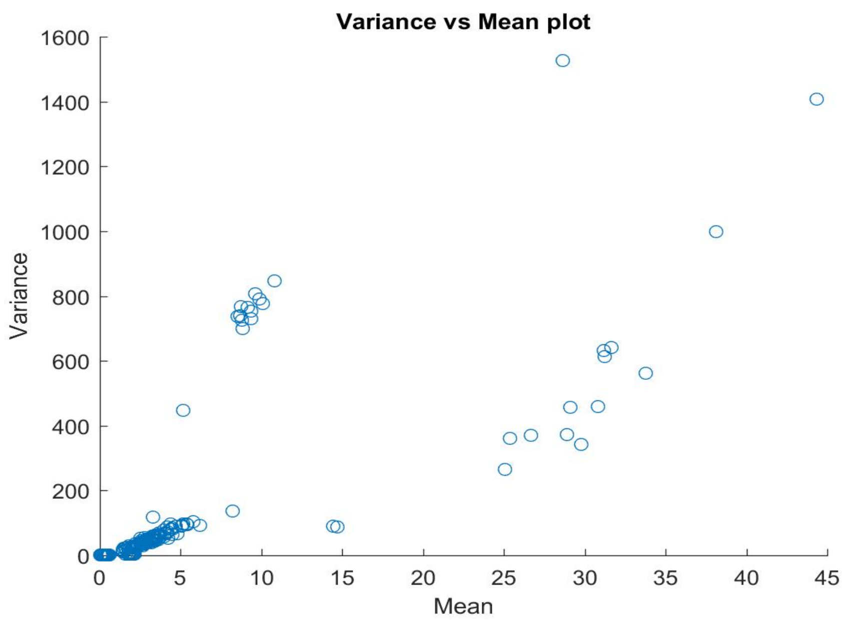
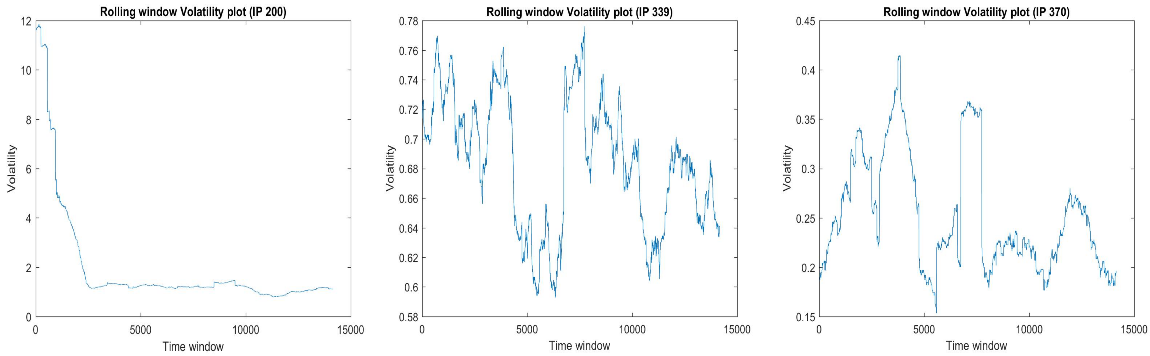


Figure 2: Mean/Variance plot for all IP addresses

Figure 2 shows a plot of variance vs mean for all IP addresses which indicates that the data is highly overdispersed. The data also exhibits heteroskedasticity as can be seen from figure 3.



(a) IP 200 (Cluster A) **(b) IP 339 (Cluster B)** **(c) IP 370 (Cluster C)**

Figure 3: Rolling window volatility plot.

Figure 3 shows volatility change over time for three IP addresses, one from each cluster respectively. The x-axis represents time windows ranging from 1 to 14156, each of size 1000 periods and the y-axis represents volatility corresponding to the time window. Figure 3 clearly displays that volatility of cyberattacks changes over time. Thus, combining GP distribution with INGARCH model is appropriate for modeling such a dataset.

4.2 Estimation Results

In this section I present the estimation results for the four models proposed in **Section 2**. For estimation purposes, I excluded data corresponding to the last one hour, that is data corresponding to last 36 time periods, from the training set. These 36 time periods will be used as the test set to compare the forecast accuracy of the 4 models.

I begin with lag selection for each of the 4 GP-INGARCH(p, q) panel data models proposed previously. For this each of the 4 models was estimated with different combinations of values of p and q, each taking value 1, 2 or 3. For values of p greater than 1, the models did not yield significant parameter estimates even at 10% significance level. Excluding such (p, q) combinations which yield insignificant estimates, we are left with only three (p, q) pairs - (1,1),(1,2) and (1,3). Table 1 shows negative log-likelihood, AIC and BIC values corresponding to each of the three (p, q) pairs for models M1(partial pooling without clustering) and M3(complete pooling without clustering).

Table 1: Lag selection (Without clustering)

		Model M1 (Partial pooling)		
		Criteria		
S. No.	**(p,q)**	**Negative log-lik**	**AIC**	**BIC**
1	(1,1)	9334781	18669570	18669625
2	(1,2)	9311388	18622786	18622855
3	(1,3)	9299161	18598335	18598418
		Model M3 (Complete pooling)		
		Criteria		
S. No.	**(p,q)**	**Negative log-lik**	**AIC**	**BIC**
1	(1,1)	11988102	23976212	23976267
2	(1,2)	11988070	23976150	23976219
3	(1,3)	11988070	23976152	23976234

From Table 1, it can be seen that based on AIC, BIC and negative log-likelihood values, GP-INGARCH(1,3) gives the best fit in case of Partial pooling without clustering (Model M1) whereas GP-INGARCH(1,2) gives the best fit in case of Complete pooling without clustering (M3).

In case of models M2 and M4, the estimation was performed separately for each cluster for each of the 3 (p, q) pairs and then the likelihood values of the three clusters were multiplied to obtain the likelihood and subsequently AIC and BIC for the corresponding model. Table 2 reports the negative log-likelihood, AIC and BIC values for models M2 (Partial pooling with clustering) and M4 (Complete pooling with clustering). In Table 2, I present estimates for 3 cases that give lower values for the 3 criteria out of the total 27 cases. For complete Table 2 refer to Appendix section A.2.

From Table 2 we can see that in case of Partial pooling with clustering (Model M2) GP-INGARCH(1,3) gives the best fit for all clusters with minimum values of AIC, BIC and negative log-likelihood. However, in case of Complete pooling with clustering (Model M4), GP-INGARCH process that gives the best fit for each of the blocks (clusters) on the basis of AIC and BIC has different lag-lengths. The lag-length for Block A should be (1,2), Block B should be (1,1) and Block C should be (1,3).

Table 2: Lag selection (With clustering)

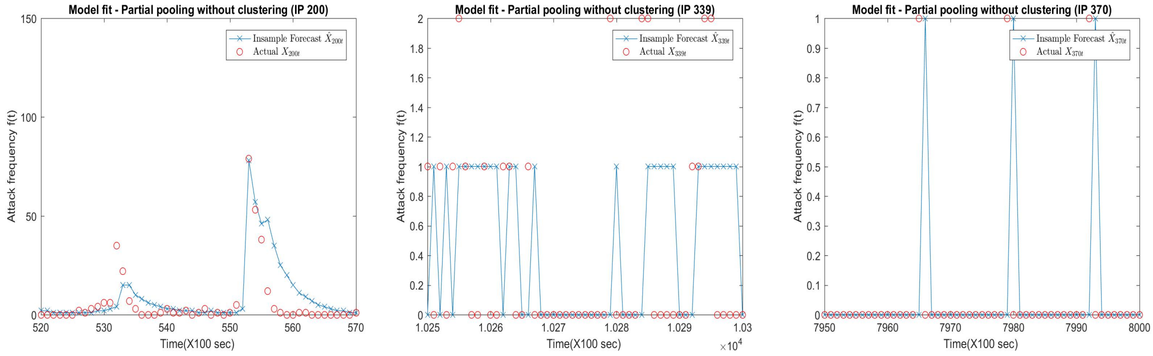
	(p,q)			Model M2 (Partial pooling) Criteria		
S. No.	**Block A**	**Block B**	**Block C**	**Negative log-lik**	**AIC**	**BIC**
1	(1,3)	(1,2)	(1,3)	8953898	17907830	17907912
2	(1,3)	(1,3)	(1,2)	8953897	17907828	17907911
3	(1,3)	(1,3)	(1,3)	8953864	17907764	17907852
	(p,q)			***Model M4 (Complete pooling) Criteria***		
S. No.	**Block A**	**Block B**	**Block C**	**Negative log-lik**	**AIC**	**BIC**
1	(1,2)	(1,1)	(1,3)	10187007	20374043	20374250
2	(1,2)	(1,2)	(1,3)	10187007	20374045	20374266
3	(1,3)	(1,1)	(1,3)	10187007	20374045	20374266

Table 3 summarizes the negative log-likelihood, AIC and BIC values for the 4 models after selecting the appropriate lag-length for each of them. The rows of Table 3 have been sorted in increasing order of negative log-likelihood, AIC and BIC with the model having minimum values for all the three being on the top.

Table 3: Model Comparison

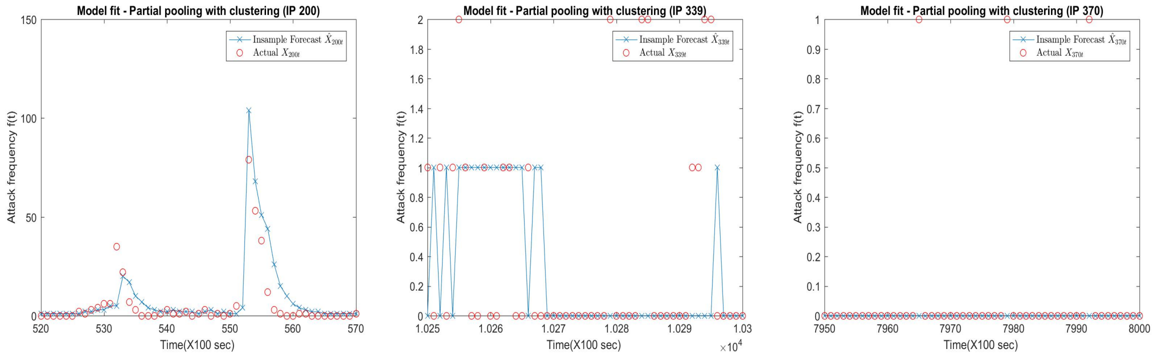
Model	Neg. log-likelihood	AIC	BIC
$M2$	8953864	17907764	17907852
$M1$	9299161	18598335	18598418
$M4$	10187007	20374043	20374250
$M3$	11988070	23976150	23976219

From Table 3 we can see that based on the values of the two information criteria, GP-INGARCH models with partial pooling (M1 and M2) give a better fit as compared to complete pooling (M3 and M4). For each of these two cases, a further observation is that the model with clustering performs better than the one without clustering. Figure 4 and figure 5 show the model fit for Partial pooling without clustering (M1) and partial pooling with clustering (M2) respectively for IP 200 from cluster A, IP 339 from cluster B and IP 370 from cluster C.



(a) IP 200 (Cluster A) (b) IP 339 (Cluster B) (c) IP 370 (Cluster C)

Figure 4: Fit of Model M1 (Partial pooling without clustering)



(a) IP 200 (Cluster A) (b) IP 339 (Cluster B) (c) IP 370 (Cluster C)

Figure 5: Fit of Model M2 (Partial pooling with clustering).

4.3 Forecast Comparison

For all 4 models, let $\{X_{it} : i \in \{1, ..., N\} \ \ and \ \ t \in \{1, ..., T\}\}$ be the training set. Considering data corresponding to the last 36 time periods out of a total of 15,155 let the test set be $\{X_{it} : i \in \{1, ..., N\} \ \ and \ \ t \in \{T+1, ..., T+36\}\}$. Then one-step-ahead forecast of the cyber-attack frequency can be done using the following formulas

Partial pooling without clustering

$$\widehat{\lambda}_{iT+m} = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{iT+m-j} + \sum_{k=1}^{q} \beta k \widehat{\lambda}_{iT+m-j} \tag{4.1}$$
$$where \ \ m \in \{1, 2, ..., 36\}, \ \ i \in \{1, ..., N\}$$

Complete pooling without clustering

$$\widehat{\lambda}_{T+m} = \alpha_0 + \sum_{j=1}^{p} \alpha_j \bar{X}_{T+m-j} + \sum_{k=1}^{q} \beta_k \widehat{\lambda}_{T+m-j} \tag{4.2}$$
$$where \;\; m \in \{1, 2, ..., 36\}, \;\; i \in \{1, ..., N\}$$

For models M2 and M4 the same formulas can be applied to each cluster separately and the results can be combined to get the forecast for all IPs.

Forecasts of the 4 models can be compared on the basis of their Mean absolute error (MAE), Mean squared error (MSE) and Mean Root error (MRE) which was proposed by Kourentzes et al. (2014). Let T^* be the total number of forecasting periods, $\widehat{Y}_{it}$ be the prediction and Y_{it} be the true value of the variable being predicted then these errors can be calculated using the following formulas:

Mean Absolute Error

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^*} \sum_{t=1}^{T^*} |Y_{it} - \widehat{Y}_{it}|$$

Mean Squared Error

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^*} \sum_{t=1}^{T^*} (Y_{it} - \widehat{Y}_{it})^2$$

Mean Root Error

$$\begin{aligned} e_{it} &= \sqrt{(Y_{it} - \widehat{Y}_{it})} = a_{it} + ib_{it} \\ MRE &= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^*} \sum_{t=1}^{T^*} e_{it} \\ &= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^*} \sum_{t=1}^{T^*} (a_{it} + ib_{it}) \end{aligned}$$

The Mean root error can be expressed in polar form to get the magnitude r which is a measure of accuracy of the forecast and angle γ which is a measure of its bias. $\gamma = \pi/4$ is the unbiased behavior using which γ can be normalised to a scale and unit free bias coefficient $\mathcal{B} = 1 - \frac{\gamma}{\pi/4}$. $\mathcal{B} \in [-1, 1]$ where negative bias coefficient signifies over-forecasting, positive coefficient signifies under-forecasting and 0 is unbiased. The lower the value of r the more accurate the forecast is and the closer the $\mathcal{B}$ is to 0 the less(weakly) biased is the forecast.

The forecasting errors for each model have been summarised in Table 4.

Table 4: Forecast Comparison

Error	Model	All IPs	Block A	Block B	Block C
MAE	M1*	0.88	1.61	0.28	0.02
	M2*^{c}	0.96	1.83	0.18	0.01
	M3**	2.83	3.74	1.83	1.99
	M4**^{c}	2.55	5.0	0.18	0.01
MSE	M1	6.05	11.92	0.33	0.02
	M2	6.48	12.82	0.23	0.01
	M3	56.2	108.4	3.53	3.97
	M4	52.13	103.9	0.23	0.01
MRE r ($\mathcal{B}$)	M1	0.39 (-1.3%)	0.69 (-0.7%)	0.19 (-5.7%)	0.01 (0.1%)
	M2	0.42 (-20.9%)	0.80 (-31%)	0.16 (93.7%)	0.01 (100%)
	M3	1.09 (-59.7%)	0.97 (-2.2%)	1.33 (-99.7%)	1.41 (-100%)
	M4	0.77 (-54.1%)	1.52 (-60.5%)	0.16 (100%)	0.01 (100%)

* Partial Pooling, ** Complete Pooling, ^{c} Clustering

Looking at the forecast errors it can be seen that GP-INGARCH panel-data models with partial pooling (M1 and M2) give better forecast as compared to complete pooling models (M3 and M4). In the former case looking at the MAE and MSE, the model without clustering (M1) performs better while forecasting cyber attacks for the whole panel and for Block A taken separately whereas for Block B and C the model with clustering (M2) gives low error values. On the basis of mean root error, even though the error magnitude shows a similar trend, however, M1 gives much lower forecast bias overall as compared to M2. Models with clustering (M2 and M4) are over-forecasting for all IPs taken together and for cluster A as can be seen from their negative bias coefficient whereas in case of cluster B and C clustering models M2 and M4 are under-forecasting more than 90% of the time as shown by their highly positive bias coefficient. However, model M3 which is complete pooling without clustering has a high error magnitude as compared to other models and is over-forecasting in all cases.

5 Conclusion

In this document, four new dynamic generalized Poisson panel-data models were proposed for modeling cyber-risk. The models can also be applied to other types of count data which exhibit under-dispersion or over-dispersion and heteroskedasticity. In particular, the GP-INGARCH model of Zhu (2012) has been extended to the case of panel data with 4 variations. First (M1), involved partial pooling where only the GARCH parameters were pooled across all units of the panel for each time period, therefore allowing them to vary only across time. Second (M2), introduced clustering in model M1 on the basis of similarities between certain units of the panel. Third (M3), involved complete pooling where along with GARCH parameters $X_i's$ were also pooled by taking their sample mean for each time period. This allowed interaction among different units of the panel. Fourth (M4), introduced clustering in M3.

Stationarity properties of the proposed models were studied. Theorems establishing sufficient conditions for existence and stationarity of the processes for the case of partial pooling and complete pooling have been proved using arguments similar to those in Ferland et al. (2006). In both the cases I started with constructing the GP INGARCH process by successive approximations and then proving that it is a strictly stationary process given the hypothesis on the model parameters. The expressions for the unconditional as well as conditional mean and variance were also derived.

Maximum likelihood estimation of the models has been discussed. Expressions for the score function and Hessian matrix have been given in each case. Procedure for selecting appropriate lag length for the INGARCH process in each of the 4 models has been described.

The proposed models were applied to a dataset of cyber-attacks on consecutive victim IP addresses and then compared on the basis of negative log-likelihood, AIC and BIC values. It was observed that GP-INGARCH models with partial pooling, that is M1 and M2, give a better fit than complete pooling models (M3 and M4). Additionally, the model with clustering which means M2 performed better than the one without clustering (M1). A further comparison of the 4 models was done based on their forecast accuracy using 3 types of errors namely - Mean absolute error, Mean squared error and Mean root error. Similar to the case of model fit comparison, here as well it was found that partial pooling models (M1 and M2) provide better forecast than complete pooling models (M3 and M4). The forecast errors were significantly lower for M1 and M2. While M1 gives lower magnitude of error on forecasting cyberattacks only for the whole panel and Block A taken separately, the bias coefficient of Mean root error shows that overall the forecast of M1 is weakly biased as compared to that of M2 and therefore it can be concluded that the partial pooling model without clustering (M1) provides a better forecast as compared to the model with clustering (M2).

A Appendix

A.1 Detailed proofs for section 2.3

A.1.1

Definition A.1. *Thinning operation Let X be a non-negative integer-valued random variable and $\phi \geq 0$. The thinning operation is defined as $\phi \circ X = \sum_{j=1}^{X} Z_j$ where the counting series $\{Z_j\}$ is a sequence of i.i.d. non-negative integer-valued random variables, independent of X and $E(Z_j) = \phi$.*

A.1.2

To prove: $E(X_{it}^{(n)})$ does not depend on t, it only depends on n.

Proof. This is trivial for $n < 0$.
For $n = 0$, $E(X_{it}^{(0)}) = (1 - \mathcal{K})E(U_{it}) = \psi_0$ which is independent of t.
As an induction hypothesis, suppose for any arbitrary fixed value of t and until $n > 0$, $E(X_{it}^{(n)})$ is independent of t. For $n + 1$, consider

$$E(X_{it}^{(n+1)}) = (1 - \mathcal{K})[\varphi_0 + \sum_{r=1}^{n+1} \varphi_r E(X_{it-r}^{(n+1-r)}) = g(E(X_{it-n-1}^{(0)}), ..., E(X_{it-1}^{(n)}))$$

which is independent of t using induction hypothesis.

■

A.1.3

To prove: The sequence $\{X_{it}^{(n)}\}_{n\in\mathbb{N}}$ has an almost sure limit.

Proof. Following propositon 2 of Ferland et al. (2006) let $(\Omega, \mathbb{F}, \mathbb{P})$ be the common probability space on which the random variables are defined. We know that $X_{it}^{(n)}$ is a non-decreasing sequence of non-negative integers. Therefore, we just have to show that

$$\forall \omega \in \Omega \lim_{n\to\infty} X_{it}^{(n)}(\omega) = X_{it} \text{ is finite.}$$

Define $A_n = \{\omega : X_{it}^{(n)}(\omega) - X_{it}^{(n-1)}(\omega) > 0\}$, for $n > 1$
Consider

$$A_\infty = \{\omega : X_{it}(\omega) = \infty\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \limsup_n A_n$$

$$E(X_{it}^{(n)} - X_{it}^{(n-1)}) \geq \sum_{k=1}^{\infty} Pr\{\omega : X_{it}^{(n)}(\omega) - X_{it}^{(n-1)}(\omega) = k\} = Pr\{A_n\}$$

Also, $E(X_{it}^{(n)} - X_{it}^{(n-1)}) = \mu_n - \mu_{n-1} = \upsilon_n \ (say)$
The sequence υ_n satisfies a finite difference equation with characteristic polynomial K(z) that has all its roots outside the unit circle. Also, from section 3.6 of Brockwell and Davis(1991) $\exists M \geq 0$ and a constant $0 < c < 1$ $s.t.$ $\upsilon_n \leq Mc^n$.
Since $Pr\{A_n\} \leq \upsilon_n$ we get,

$$\sum_{n=1}^{\infty} Pr\{A_n\} \leq M \sum_{n=1}^{\infty} c^n < \infty$$

$\therefore$ using Borel-Cantelli lemma, $Pr\{A_\infty\} = 0$.
Hence, we have proved that $\{X_{it}^{(n)} : t \in \mathbb{Z}, n \in \mathbb{Z}\}$ converges almost surely to a process $(X_{it} : t \in \mathbb{Z})$ that is almost surely finite.

■

A.1.4

To prove: Distributional properties of X_{it}

Proof. Given $\mathcal{F}_{it-1} = \sigma(X_{it-1}, X_{it-2}, ...)$ for $t \in \mathbb{Z}$ let

$$\lambda_{it} = \alpha_0 C^{-1}(1) + \sum_{j=1}^{\infty} \psi_j X_{it-j}$$

The sequence $\{\lambda_{it}\}$ satisfies,

$$\lambda_{it} = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{it-j} + \sum_{k=1}^{q} \beta_k \lambda_{it-k}$$

For a fixed t, consider the sequence $\{r_{it}^{(n)}\}_{n \in \mathbb{N}}$ defined as

$$r_{it}^{(n)} = (1-\mathcal{K})U_{it} + (1-\mathcal{K}) \sum_{j=1}^{n} \sum_{k=1}^{X_{it-j}} Z_{it-j,j,k}$$

We claim there is a subsequence $\{n_k\}$ s.t. $r_{it}^{(n_k)}$ converges almost surely to X_{it}.
Consider

$$X_{it} - r_{it}^{(n)} = (X_{it} - X_{it}^{(n)}) + (X_{it}^{(n)} - r_{it}^{(n)})$$

Let

$$Y_{it}^{(n)} = r_{it}^{(n)} - X_{it}^{(n)} = (1-\mathcal{K}) \sum_{j=1}^{n} \sum_{k=X_{it-j}^{n-j}+1}^{X_{it-j}} Z_{it-j,j,k}$$

I will now show that there exists a subsequence of $Y_{it}^{(n)}$ which converges almost surely to 0.

$$\begin{aligned}
\lim_{n\to\infty} E(Y_{it}^{(n)}) &= (1-\mathcal{K}) \lim_{n\to\infty} \sum_{j=1}^{n} E\left[\sum_{k=X_{it-j}^{n-j}+1}^{X_{it-j}} Z_{it-j,j,k}\right] \\
&= (1-\mathcal{K}) \lim_{n\to\infty} \sum_{j=1}^{n} E\left[X_{it-j} - X_{it-j}^{(n-j)}\right] \varphi_j \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} E\left[X_{it-j} - X_{it-j}^{(n-j)}\right] \psi_j \\
&= \lim_{n\to\infty} \left[\mu \sum_{j=1}^{n} \psi_j - \sum_{j=1}^{n} \mu_{n-j}\psi_j\right] \\
&= \frac{\alpha_0 G(1)}{C(1)-D(1)} - \lim_{n\to\infty} \sum_{j=1}^{n} \mu_{n-j}\psi_j \quad \text{from 2.10 and 2.8} \\
&= \frac{\alpha_0 D(1)}{C(1)[C(1)-D(1)]} - \lim_{n\to\infty} \mu_n + \psi_0 \quad \text{from 2.9} \\
&= \frac{\alpha_0 D(1)}{C(1)[C(1)-D(1)]} - \lim_{n\to\infty} \mu_n + \frac{\alpha_0}{C(1)} \\
&= \frac{\alpha_0 D(1)}{C(1)[C(1)-D(1)]} - \frac{\alpha_0}{C(1)-D(1)} + \frac{\alpha_0}{C(1)} \quad \text{from 2.10} \\
&= 0
\end{aligned}$$

Since $Y_{it}^{(n)}$ is non-negative we get that $Y_{it}^{(n)}$ converges to 0 and hence there is a subsequence $Y_{it}^{(n_k)}$ converging almost surely to the same limit.

Therefore $r_{it}^{(n_k)}$ converges almost surely to X_{it} which implies $r_{it}^{(n)}|\mathcal{F}_{it-1}$ converges almost surely to $X_{it}|\mathcal{F}_{it-1}$.

But

$$r_{it}^{(n)}|\mathcal{F}_{it-1} \sim \mathcal{GP}((\alpha_0 C^{-1}(1) + \sum_{j=1}^{n} \psi_j X_{it-j})(1-\mathcal{K}), \mathcal{K})$$

Therefore, it follows that

$$\begin{aligned}
&X_{it}|\mathcal{F}_{it-1} \sim \mathcal{GP}(\lambda_{it}^{*}, \mathcal{K}) \\
&\text{where } \lambda_{it}^{*} = (\alpha_0 C^{-1}(1) + \sum_{j=1}^{n} \psi_j X_{it-j})(1-\mathcal{K})
\end{aligned}$$

■

A.1.5

To prove: Distributional properties of (X_{1t}, X_{2t})

Proof. Given $\mathcal{F}_{t-1} = \sigma((X_{1t-1}, X_{2t-1}), (X_{1t-2}, X_{2t-2}), ...)$ for $t \in \mathbb{Z}$ let

$$\lambda_{1t} = \lambda_{2t} = \lambda_t = \alpha_0 C^{-1}(1) + \sum_{j=1}^{\infty} \psi_j \bar{X}_{t-j}$$

The sequence $\{\lambda_t\}$ satisfies,

$$\lambda_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j \bar{X}_{t-j} + \sum_{k=1}^{q} \beta_k \lambda_{t-k}$$

For a fixed t, consider the sequence $\{(r_{1t}^{(n)}, r_{2t}^{(n)})\}_{n\in\mathbb{N}}$ defined as

$$r_{it}^{(n)} = (1-\mathcal{K})U_{it} + (1-\mathcal{K})\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{X_{1t-j}+X_{2t-j}} Z_{it-j,j,k}$$

We claim there is a subsequence $\{n_k\}$ s.t. $\{(r_{1t}^{(n_k)}, r_{2t}^{(n_k)})\}$ converges almost surely to $\{(X_{1t}, X_{2t})\}$.
For $i = 1, 2$ consider $X_{it} - r_{it}^{(n)} = (X_{it} - X_{it}^{(n)}) + (X_{it}^{(n)} - r_{it}^{(n)})$

Let $Y_t^{(n)} = (r_{1t}^{(n)} - X_{1t}^{(n)}) + (r_{2t}^{(n)} - X_{2t}^{(n)}) = \frac{1}{2}(1-\mathcal{K})\left[\sum_{j=1}^{n}\sum_{k=X_{1t-j}^{n-j}+X_{2t-j}^{n-j}+1}^{X_{1t-j}+X_{2t-j}} [Z_{1t-j,j,k} + Z_{2t-j,j,k}]\right]$

We now have to find a subsequence of $Y_t^{(n)}$ which converges almost surely to 0.

$$\begin{aligned}
\lim_{n\to\infty} E(Y_t^{(n)}) &= \frac{(1-\mathcal{K})}{2} \lim_{n\to\infty} \sum_{j=1}^{n} E\left[\sum_{k=X_{1t-j}^{n-j}+X_{2t-j}^{n-j}+1}^{X_{1t-j}+X_{2t-j}} (Z_{1t-j,j,k} + Z_{2t-j,j,k})\right] \\
&= \frac{(1-\mathcal{K})}{2} \lim_{n\to\infty} \sum_{j=1}^{n} 2E\left[X_{1t-j} + X_{2t-j} - X_{1t-j}^{(n-j)} - X_{2t-j}^{(n-j)}\right] \varphi_j \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} E\left[X_{1t-j} + X_{2t-j} - X_{1t-j}^{(n-j)} - X_{2t-j}^{(n-j)}\right] \psi_j \\
&= \lim_{n\to\infty} \left[2\mu \sum_{j=1}^{n} \psi_j - 2\sum_{j=1}^{n} \mu_{n-j}\psi_j\right] \\
&= 2\left[\frac{\alpha_0 G(1)}{C(1)-D(1)} - \lim_{n\to\infty} \sum_{j=1}^{n} \mu_{n-j}\psi_j\right] \quad \text{from 2.14 and 2.12} \\
&= 2\left[\frac{\alpha_0 D(1)}{C(1)[C(1)-D(1)]} - \lim_{n\to\infty} \mu_n + \psi_0\right] \quad \text{from 2.13} \\
&= 2\left[\frac{\alpha_0 D(1)}{C(1)[C(1)-D(1)]} - \lim_{n\to\infty} \mu_n + \frac{\alpha_0}{C(1)}\right] \\
&= 2\left[\frac{\alpha_0 D(1)}{C(1)[C(1)-D(1)]} - \frac{\alpha_0}{C(1)-D(1)} + \frac{\alpha_0}{C(1)}\right] \quad \text{from 2.14} \\
&= 0
\end{aligned}$$

This shows that $Y_{it}^{(n)}$ converges to 0 because $Y_t^{(n)}$ is non-negative. Hence, there is a subsequence $Y_t^{(n_k)}$ converging almost surely to the same limit.

Therefore, $(r_{1t}^{(n_k)}, r_{2t}^{(n_k)})$ converges almost surely to (X_{1t}, X_{2t}) which implies that $(r_{1t}^{(n)}, r_{2t}^{(n)})|\mathcal{F}_{t-1}$ converges almost surely to $(X_{1t}, X_{2t})|\mathcal{F}_{t-1}$.

But

$$(r_{1t}^{(n)}, r_{2t}^{(n)})|\mathcal{F}_{t-1} \sim \mathcal{GP}((\alpha_0 C^{-1}(1) + \sum_{j=1}^{n} \psi_j \bar{X}_{t-j})(1-\mathcal{K}), \mathcal{K})$$

Therefore, we get that

$$X_t = (X_{1t}, X_{2t})|\mathcal{F}_{t-1} \sim \mathcal{GP}(\lambda_t^*, \mathcal{K})$$

$$\text{where} \quad \lambda_t^* = (\alpha_0 C^{-1}(1) + \sum_{j=1}^{n} \psi_j \bar{X}_{t-j})(1-\mathcal{K})$$

■

A.2 Table 2

Lag selection (With clustering) - Model M2 (Partial Pooling)

	(p,q)			*Criteria*		
S. No.	**Block A**	**Block B**	**Block C**	**Negative log-lik**	**AIC**	**BIC**
1	(1,1)	(1,1)	(1,1)	8962780	17907830	17907912
2	(1,1)	(1,1)	(1,2)	8962655	17925336	17925399
3	(1,1)	(1,1)	(1,3)	8962622	17925272	17925340
4	(1,1)	(1,2)	(1,1)	8962726	17925478	17925542
5	(1,1)	(1,2)	(1,2)	8962601	17925229	17925298
6	(1,1)	(1,2)	(1,3)	8962568	17925165	17925238
7	(1,1)	(1,3)	(1,1)	8962692	17925413	17925481
8	(1,1)	(1,3)	(1,2)	8962567	17925164	17925237
9	(1,1)	(1,3)	(1,3)	8962534	17925100	17925178
10	(1,2)	(1,1)	(1,1)	8956427	17912881	17912944
11	(1,2)	(1,1)	(1,2)	8956302	17912632	17912700
12	(1,2)	(1,1)	(1,3)	8956269	17912568	17912641
13	(1,2)	(1,2)	(1,1)	8956373	17912774	17912843
14	(1,2)	(1,2)	(1,2)	8956248	17912526	17912599
15	(1,2)	(1,2)	(1,3)	8956215	17912462	17912540
16	(1,2)	(1,3)	(1,1)	8956340	17912709	17912782
17	(1,2)	(1,3)	(1,2)	8956214	17912460	17912538
18	(1,2)	(1,3)	(1,3)	8956181	17912396	17912479
19	(1,3)	(1,1)	(1,1)	8954110	17908249	17908317
20	(1,3)	(1,1)	(1,2)	8953985	17908000	17908073
21	(1,3)	(1,1)	(1,3)	8953952	17907936	17908014
22	(1,3)	(1,2)	(1,1)	8954056	17908142	17908215
23	(1,3)	(1,2)	(1,2)	8953931	17907893	17907971
24	(1,3)	(1,2)	(1,3)	8953898	17907830	17907912
25	(1,3)	(1,3)	(1,1)	8954022	17908077	17908155
26	(1,3)	(1,3)	(1,2)	8953897	17907828	17907911
27	(1,3)	(1,3)	(1,3)	8953864	17907764	17907852

Lag selection (With clustering) - Model M4 (Complete pooling)

	(p,q)			***Criteria***		
S. No.	**Block A**	**Block B**	**Block C**	**Negative log-lik**	**AIC**	**BIC**
1	(1,1)	(1,1)	(1,1)	10187231	20374486	20374652
2	(1,1)	(1,1)	(1,2)	10187098	20374222	20374402
3	(1,1)	(1,1)	(1,3)	10187054	20374137	20374330
4	(1,1)	(1,2)	(1,1)	10187231	20374488	20374667
5	(1,1)	(1,2)	(1,2)	10187098	20374224	20374418
6	(1,1)	(1,2)	(1,3)	10187054	20374139	20374346
7	(1,1)	(1,3)	(1,1)	10187231	20374490	20374683
8	(1,1)	(1,3)	(1,2)	10187098	20374226	20374434
9	(1,1)	(1,3)	(1,3)	10187054	20374141	20374362
10	(1,2)	(1,1)	(1,1)	10187183	20374392	20374572
11	(1,2)	(1,1)	(1,2)	10187050	20374129	20374322
12	(1,2)	(1,1)	(1,3)	10187007	20374043	20374250
13	(1,2)	(1,2)	(1,1)	10187183	20374394	20374588
14	(1,2)	(1,2)	(1,2)	10187050	20374131	20374338
15	(1,2)	(1,2)	(1,3)	10187007	20374045	20374266
16	(1,2)	(1,3)	(1,1)	10187183	20374396	20374603
17	(1,2)	(1,3)	(1,2)	10187050	20374133	20374354
18	(1,2)	(1,3)	(1,3)	10187007	20374047	20374282
19	(1,3)	(1,1)	(1,1)	10187183	20374394	20374588
20	(1,3)	(1,1)	(1,2)	10187050	20374131	20374338
21	(1,3)	(1,1)	(1,3)	10187007	20374045	20374266
22	(1,3)	(1,2)	(1,1)	10187183	20374396	20374603
23	(1,3)	(1,2)	(1,2)	10187050	20374133	20374354
24	(1,3)	(1,2)	(1,3)	10187007	20374047	20374282
25	(1,3)	(1,3)	(1,1)	10187183	20374398	20374619
26	(1,3)	(1,3)	(1,2)	10187050	20374135	20374370
27	(1,3)	(1,3)	(1,3)	10187007	20374049	20374298

References

Aon (2018). Cybersecurity predictions report - a shift to managing cyber as an enterprise risk. https://www.strozfriedberg.com/wp-content/uploads/2018/01/2018-Cybersecurity-Predictions-Report-Aon-Cyber-Solutions.pdf.

Böhme, R. and Kataria, G. (2006). Models and measures for correlation in cyber-insurance. In *Presented at Workshop on the Economics of Information security (WEIS)*, University of Cambridge, UK.

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

Camillo, M. (2017). Cyber risk and the changing role of insurance. *Journal of Cyber Policy*, 2(1):53 – 63.

Chen, Y.-Z., Huang, Z.-G., Xu, S., and Lai, Y.-C. (2015). Spatiotemporal patterns and predictability of cyberattacks. *PloS One*, 10(5).

Ferland, R., Latour, A., and Oraichi, D. (2006). Integervalued garch process. *Journal of Time Series Analysis*, 27(6):923–942.

Kourentzes, N., Trapero, J. R., and Svetunkov, I. (2014). Measuring the behaviour of experts on demand forecasting: a complex task. *Working Paper*. Department of Management Science, Lancaster University Management School. Lancaster, Lancashire, LA1 4YX, UK.

OECD and Marsh & McLennan Companies (2018). Unleashing the potential of the cyber insurance market. https://www.oecd.org/daf/fin/insurance/Unleashing-the-potential-of-the-cyber-insurance-market-final-programme.pdf.

Ponemon Institute (2017). 2017 cost of cyber crime study. https://www.accenture.com/t20171006T095146Z__w__/us-en/_acnmedia/PDF-62/Accenture-2017CostCybercrime-US-FINAL.pdf#zoom=50.

Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7(5):893–899.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Xu, M. and Hua, L. (2017). Cybersecurity insurance :modelling and pricing. `https://www.soa.org/Files/Research/Projects/cybersecurity-insurance-report.pdf`. *Society of Actuaries Report*.

Zhu, F. (2011). A negative binomial integervalued garch model. *Journal of Time Series Analysis*, 32(1):54–67.

Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *Journal of Mathematical Analysis and Applications*, 389(1):58 – 71.