Economic Computation and Economic Cybernetics Studies and Research, Issue 3/2019; Vol. 53

Senior Lecturer Ng KOK-HAUR, PhD E-mail: kokhaur@um.edu.my Institute of Mathematical Sciences, University of Malaya, Malaysia Senior Lecturer Koh YOU-BENG, PhD E-mail: kohyoubeng@um.edu.my Institute of Mathematical Sciences, University of Malaya, Malaysia Professor Pooi AH-HIN, PhD E-mail: ahhinp@sunway.edu.my School of Mathematical Sciences, Sunway University, Malaysia

# MODELLING AND FORECASTING COUNT DATA WITH A MODEL BASED ON MULTIVARIATE POWER-NORMAL DISTRIBUTION: A COMPARATIVE STUDY WITH AN APPLICATION

Abstract. Time series count data are frequently encountered in many realworld applications. This paper uses a model, called the MPN model, which is based on the multivariate power-normal (MPN) distribution, for modelling time series of counts. Compared with the integer-valued GARCH (INGARCH) models, it is easier to implement the MPN model because it involves a simpler process of distribution and model selection. Furthermore it provides readily a prediction interval for the future observation. A set of count data is used to illustrate the capability of the MPN model. The fitted MPN model is found to be comparable to the INGARCH models with Poisson and negative binomial distributions in terms of root mean square error and mean absolute deviation. Diebold and Mariano, and Mincer-Zarnowitz tests are carried out to confirm the unbiasedness of the predicted values for all the models.

*Keywords*: Integer-valued, Count data, Multivariate power-normal, INGARCH, Over-dispersed.

#### JEL Classification: C1, C5, C22

#### 1. Introduction

Integer-valued time series models play an important role in modelling the time series of count data. There are two types of integer-valued time series models: thinning models and state space models. The thinning models are based on binomial or geometric thinning operator which is commonly used to ensure the realisations are positive, see Mckenzie (1985), Al-Osh and Alzaid (1987) and etc.

On the other hand, the state space model can be divided into two categories according to whether it is observation-driven or parameter-driven.

In this paper, we begin with the observation-driven state space models. Heinen (2003) introduced the autoregressive conditional Poisson (ACP) model with Poisson and double Poisson distributions to model count data. He showed that the model with double Poisson distribution gives a better fit to the data than the model with Poisson distribution. Ferland et al. (2006) proposed integer-valued generalised autoregressive conditional heteroscedasticity GARCH model of order p and q (INGARCH(p,q)) with conditional Poisson distribution and showed that under some regularly conditions all moments of the INGARCH(1,1) model exist. In fact the INGARCH model, also known as the ACP model share the same mean structure as autoregressive conditional duration and GARCH models except that in the former models, the count data follow a non-negative support discrete distribution. Weiß (2009) derived the variance and autocorrelation functions for the general INGARCH(p,q) model. His findings cover the results given by Heinen (2003) and Ferland et al. (2006) which is a special case of the INGARCH(1,1) model. In the empirical study on claims counts data, Weiß (2009) showed that INGARCH(1,0) model gives better fit than the integer-valued autoregressive model of order 1 (INAR(1)) in terms of Akaike information criterion (AIC) and Bayesian information criterion (BIC). Studies related to a special case of INGARCH(p,q) given by INGARCH(p,0) or integer-valued autoregressive conditional heteroscedasticity model of order p (INARCH(p)) have been carried out by Weiß (2010) and Zhu and Wang (2010, 2011).

Conditional Poisson distribution has equal mean and variance (equidispersion). However, empirical evidence indicates that many count data are overdispersed. To deal with the over-dispersion issue, many alternative models have been put forward. The negative binomial regression model which is a generalisation of Poisson log-linear regression models was proposed by Davis and Wu (2009) for count time series. Zhu et al. (2010) proposed a mixture integervalued ARCH model which is able to handle over-dispersion, multimodality and nonstationary components. Zhu (2011) proposed a negative binomial INGARCH model that can deal with both over-dispersion and potential extreme observations simultaneously. Empirical results by Zhu et al. (2010) show that negative binomial INARCH(1) model outperformance than the double Poisson and Poisson INARCH(1) model in modelling the over-dispersed data. Further extensions of INGARCH models included zero-inflated Poisson and negative binomial INGARCH models of Zhu (2012), long memory ACP model of Groß-Klubmann and Hautsch (2013), and INGARCH model with exogenous covariates of Agosto et al. (2016).

For the thinning models, as pointed out by Wei $\beta$  (2009) and Wei $\beta$  (2010), the models based on an appropriate type of thinning operation and with negative binomial or generalised Poisson marginal distribution have disadvantages

compared to INGARCH models. For example, the joint distribution of the thinning operators leads to a quite complex model structure. For the parameter-driven model, although the model is more flexible than the observation-driven model in term of modelling the over-dispersed count, the parameter-driven model is less popular because it is computationally burdensome.

The aim of this paper is to investigate the use of a time series model based on the multivariate power-normal (MPN) distribution for modelling the integervalued count data. This model will hereafter be referred to as the MPN model. The MPN distribution is able to capture the four major characteristics, namely mean, variance, skewness and kurtosis, exhibited by the data. Thus, it is able to fit a lot of real continuous datasets. As for the discrete datasets, the MPN distribution can provide a good approximation to the underlying multivariate discrete distribution provided that we can find the parameters of the MPN distribution of which the four major characteristics match with those exhibited by the data. The cases in which the matching of the four major characteristics is not satisfactory may arise. This may happen if the number of values in the support of an underlying univariate discrete distribution is very small (for example, two or three). Thus an obvious advantage of the model based on MPN distribution is that when the numbers of values in the supports of the underlying variables of the given discrete dataset are not too small, the MPN distribution is not bothered much by the issue of overdispersion of the count data.

The existing methods for modelling the integer-valued count data usually require careful selection of the basic distribution from the list of discrete distributions. However, the model based on MPN distribution does not need to bother further about the choice of distribution. Another advantage of the model based on MPN distribution is that a prediction interval for the future observation can be obtained easily without going through other procedures such as the bootstrapping method as has been used in Ferland et al. (2006).

The MPN model together with the INARCH and INGARCH models with Poisson and negative binomial distributions are fitted to one set of count data. The predictive performance of these models are measured and compared using root mean square error (RMSE) and mean absolute deviation (MAD). The prediction intervals based on the MPN distribution are computed along with their coverage probability and average length. Further, two tests, namely Mincer-Zarnowitz regression approach test (Mincer and Zarnowitz, 1969) and Diebold-Mariano test (Diebold and Mariano, 1995) are carried out to assess the comparative forecast accuracy between models.

The remainder of this paper is organised as follows: Section 2 reviews the general INGARCH models with Poisson and negative binomial distributions. Section 3 discusses the setup of the MPN model. Section 4 discusses an application

of the MPN model and carries out the model comparison. Concluding remarks are given in Section 5.

#### 2. A review of INGARCH model

In this section, we consider different types of INARCH and INGARCH models with Poisson and negative binomial distributions.

Let  $\{X_t\}$  be the integer-valued process.

(i) The INGARCH(p,q) with Poisson distribution is given as

$$X_{t} | F_{t-1} \sim P(\lambda_{t}),$$
  
$$\lambda_{t} = \omega + \sum_{i=1}^{p} \alpha_{i} X_{t-i} + \sum_{j=1}^{q} \beta_{j} \lambda_{t-j},$$
 (1)

where  $F_{t-1}$  is the information available on the series  $\{X_t\}$  up to time t-1 and  $\omega > 0$ ,  $\alpha_i \ge 0$ ,  $\beta_j \ge 0$  are parameters to be estimated. When q = 0, the INGARCH(p,q) reduces to the integer-valued autoregressive model of order p (INARCH(p)).

The probability mass function of  $X_t$  has the form

$$P(X_t = k | F_{t-1}) = \frac{\lambda_t^k e^{-\lambda_t}}{k!}, \text{ for } k = 0, 1, 2, \cdots,$$

with the conditional mean  $E(X_t | F_{t-1}) = \lambda_t$  and conditional variance  $Var(X_t | F_{t-1}) = \lambda_t$ . Note that the conditional distribution of  $X_t$  is equi-dispersed. However, it is easy to show that the unconditional distribution of  $X_t$  is overdispersed

$$\operatorname{Var}(X_{t}) = \operatorname{E}[\operatorname{Var}(X_{t} | F_{t-1})] + \operatorname{Var}[\operatorname{E}(X_{t} | F_{t-1})]$$
$$= \operatorname{E}(\lambda_{t}) + \operatorname{Var}(\lambda_{t})$$
$$> \operatorname{E}(\lambda_{t}).$$

(ii) The INGARCH(p,q)with negative binomial distribution (also known as NBINGARCH) is given as

$$X_{t} | F_{t-1} \sim NB(r, p_{t}),$$

$$\frac{(1-p_{t})}{p_{t}} = \lambda_{t} = \omega + \sum_{i=1}^{p} \alpha_{i} X_{t-i} + \sum_{j=1}^{q} \beta_{j} \lambda_{t-j},$$
(2)

where r is a positive number and  $p_t$  is the probability of a successful trial.

The probability mass function of  $X_t$  has the form

$$P(X_{t} = k | F_{t-1}) = {\binom{k+r-1}{r-1}} p_{t}^{r} (1-p_{t})^{k}, \text{ for } k = 0,1,2,\cdots$$

224

DOI: 10.24818/18423264/53.3.19.13

with  $p_t = \frac{1}{1 + \lambda_t}$  and  $1 - p_t = \frac{\lambda_t}{1 + \lambda_t}$ . The corresponding conditional mean and variance are respectively  $E(X_t | F_{t-1}) = r\lambda_t$  and  $Var(X_t | F_{t-1}) = r\lambda_t(1 + \lambda_t)$ . It can be shown that the model given by Equation (2) is over-dispersed for both conditional and unconditional  $X_t$ , that is  $Var(X_t | F_{t-1}) > E(X_t | F_{t-1})$  and  $Var(X_t) > E(X_t)$  (see Zhu, 2011).

The parameters of these models with various distribution assumptions are estimated by maximum likelihood (ML)method. Let  $l(\theta)$  be the conditional log-likelihood function with parameter vector  $\theta$ :

$$l(\mathbf{\theta}) = \sum_{t=1}^{T} l_{t}(\mathbf{\theta}) = \sum_{t=1}^{T} \ln P(X_{t} = k \mid F_{t-1}).$$
(3)

Then the ML estimator,  $\hat{\theta}$ , of  $\theta$  is given by

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Lambda} l(\boldsymbol{\theta}),$$

where  $\Lambda$  is the parameter space for  $\theta$ . Note that the conditional log-likelihood functions in the models in (i) and(ii) are given by

1. INARCH or INGARCH model:

$$l(\mathbf{\theta}) = \sum_{t=1}^{T} \left( X_t \ln \lambda_t - \lambda_t - \ln(X_t!) \right),$$

2. NBINARCH or NBINGARCH model:

$$l(\mathbf{\theta}) = \sum_{t=1}^{T} \left( X_t \ln \lambda_t - (r + X_t) \ln(1 + \lambda_t) + \sum_{\nu=1}^{X_t} \ln(\nu + r - 1) - \ln(X_t!) \right).$$

# **3.** Time series model based on multivariate power-normal distribution

Let us begin with the non-normal distribution given in Yeo and Johnson (2000). These authors have introduced the following power transformation

$$\widetilde{\varepsilon} = \psi(\lambda^{+}, \lambda^{-}, z) = \begin{cases} [(z+1)^{\lambda^{+}} -1]/\lambda^{+} & , z \ge 0, \lambda^{+} \ne 0\\ \log(z+1) & , z \ge 0, \lambda^{+} = 0\\ -[(-z+1)^{\lambda^{-}} -1]/\lambda^{-} & , z < 0, \lambda^{-} \ne 0\\ -\log(-z+1) & , z < 0, \lambda^{-} = 0 \end{cases}$$
(4)

If z in Equation (4) has the standard normal distribution, then  $\tilde{\varepsilon}$  has a non-normal distribution which is derived by a type of power transformation of a random variable with normal distribution. We may say that  $\tilde{\varepsilon}$  has a power-normal distribution.

Let **y** be a vector consisting of *k* correlated random variables. The vector **y** is said to have a *k*-dimensional power-normal distribution with parameters  $\boldsymbol{\mu}, \mathbf{H}, \lambda_i^+, \lambda_i^-, \sigma_i, 1 \le i \le k$  if

$$V = \mathbf{\mu} + \mathbf{H}\mathbf{\epsilon},\tag{5}$$

where  $\boldsymbol{\mu} = E(\mathbf{y})$ , **H** is a  $(k \times k)$  orthogonal matrix, the components  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k$  of  $\boldsymbol{\varepsilon}$  are independent,

$$\varepsilon_i = \sigma_i [\widetilde{\varepsilon}_i - \mathcal{E}(\widetilde{\varepsilon}_i)] / [\operatorname{Var}(\widetilde{\varepsilon}_i)]^{1/2}, \quad i = 1, 2, \cdots, k,$$
(6)

 $\sigma_i > 0$  is a constant, and  $\tilde{\varepsilon}_i$  has a power-normal distribution with parameters  $\lambda_i^+$  and  $\lambda_i^-$ .

In Pooi (2012),  $\mu_i$  is estimated by using the average of the observed values of  $y_i$ , **H** is estimated by using the eigenvectors of the variance-covariance matrix estimated from the observed values of **y**, while  $\lambda_i^+$ ,  $\lambda_i^-$  and  $\sigma_i$  are estimated by using the estimated second, third and fourth moments of the components of  $\mathbf{H}^{\mathrm{T}}(\mathbf{y}-\mathbf{\mu})$ . The details of the procedure for the above parameter estimation can be found in Section 4 of Pooi (2012).

When the values of  $y_1, y_2, \dots, y_{k-1}$  are given, an approximation for the conditional probability density function (pdf) of  $y_k$  may be found by the following numerical procedure given in Pooi (2012):

- (1) Select a large number  $N_p$  of equally spaced feasible values of  $y_k$ .
- (2) For the  $i_p$  th  $(1 \le i_p \le N_p)$  selected value  $y_k^{(i_p)}$  of  $y_k$ , form the vector  $\mathbf{y}^{(i_p)} = (y_1, y_2, \dots, y_{k-1}, y_k^{(i_p)})^{\mathrm{T}}$ .
- (3) Replace **y** by  $\mathbf{y}^{(i_p)}$ , and  $\boldsymbol{\mu}$  together with **H** in Equation (5) by their respective estimates and solve for  $\boldsymbol{\varepsilon}$ . Let the solution be denoted as  $\boldsymbol{\varepsilon}^{(i_p)}$ .
- (4) Use Equation (6), and Equation (4) with  $\tilde{\varepsilon}$ , z and  $(\lambda^+, \lambda^-)$  replaced respectively by  $\tilde{\varepsilon}_i$ ,  $z_i$  and  $(\lambda^+_i, \lambda^-_i)$  to find  $z_i = z_i^{(i_p)}$  such that  $\varepsilon_i = \varepsilon_i^{(i_p)}$ .
- (5) Find the joint pdf  $f_{i_p}$  of  $\mathbf{y}^{(i_p)}$  by multiplying the product of the pdf of  $z_i$  and the relevant Jacobian of transformation, over  $1 \le i \le k$ .
- (6) Estimate the conditional pdf (evaluated at  $y_k^{(i_p)}$ ) of  $y_k$  by  $f_{i_p} / \sum_{i_{p=1}}^{N_p} f_{i_p}$ .

Suppose we have a set of data which span over T evenly spaced time points. From the data, we can form a table of T - (k-1) rows with each row representing the observed values of  $y_1, y_2, \dots, y_k$  at k consecutive time points.

From the table, we can form the  $i_w$ -th moving window of size n from the  $i_w$ -th row till the  $(i_w + n - 1)$ -th row. We can form a total of N = T - (k - 1) - n such windows of size n. We next fit a k-dimensional power-normal distribution for  $(y_1, y_2, \dots, y_k)$  using the data in the  $i_w$ -th window.

We may find a conditional distribution for  $y_k$  when the values of  $y_1, y_2, \dots, y_{k-1}$  are given by the first k-1 values in the row which is immediately after the  $i_w$ -th window. Although the resulting conditional distribution for  $y_k$  is continuous, it can be converted to a discrete distribution by equating  $P(y_k = i)$  to the probability that the continuous version of  $y_k$  lies in  $(i^-, i+0.5]$  where  $i^- = i - 0.5$  if  $i \ge 1$  and  $i^- = -\infty$  if i = 0. This conditional distribution then specifies a time series model based on the MPN distribution for the future value  $y_k$  when the present value  $y_{k-1}$  and k-2 previous values  $y_1, y_2, \dots, y_{k-2}$  are given. We may refer to this model as the MPN model of order k-1. The mean  $\hat{y}_k^{(i_w)}$  of the conditional distribution is then a predicted value of  $y_k$ . The root mean square error (RMSE) and mean absolute deviation (MAD) of  $y_k$  are then given respectively by

RMSE = 
$$\left[\frac{1}{N}\sum_{i_{W}=1}^{N} (\hat{y}_{k}^{(i_{W})} - y_{k}^{(i_{W})})^{2}\right]^{1/2}$$

and

$$MAD = \frac{1}{N} \sum_{i_{W}=1}^{N} \left| \hat{y}_{k}^{(i_{W})} - y_{k}^{(i_{W})} \right|,$$

where  $y_k^{(i_w)}$  is the observed value of  $y_k$ .

The  $100(\alpha/2)\%$  and  $100(1-\alpha/2)\%$  points of the conditional distribution may be regarded as the lower and upper limits of the nominally  $100(1-\alpha)\%$  out-ofsample prediction interval for  $y_k$ . As the prediction interval contains a set of possible values of the observed  $y_k$ , it is more informative than the mean  $\hat{y}_k^{(i_w)}$ which is only one possible value of observed  $y_k$ .

The coverage probability of the prediction interval may be estimated by the proportion of prediction intervals (out of the N prediction intervals) which include the observed value of  $y_k$ . Meanwhile, the expected length of the prediction interval may be estimated by the average length of the prediction intervals. When the estimated coverage probability is close to the target value

#### DOI: 10.24818/18423264/53.3.19.13

 $1-\alpha$ , a small value of the average length is indicative of good predictive power of the model.

# 4. Application

Freeland (1998) and Wei $\beta$  (2009) analysed the data which consist of the monthly claims counts of workers collecting Short Term Wage Loss Benefits from the Richmond claims center between January 1987 and December 1994.Summary statistics of the monthly claims counts of workers are reported in Table 1. The dispersion index given by the ratio of variance to mean is 1.320, indicating that the data are over-dispersed. Figures 1 to 3show respectively the time series, sample autocorrelation function (ACF) and sample partial ACF (PACF) plots. The ACF plot shows decay of correlation while the PACF plot shows that the partial autocorrelation at the lag value of 1 is significant at the 0.05 level. Thus the autoregressive model of order 1 may fit the data well.



Figure 1. Time series plot for the monthly claims counts of workers

Table 1. Summary statistics of monthly claims counts of workers

Summary statistics			
Sample Size	96		
Mean	8.604		
Standard Deviation	3.370		
Skewness	0.074		
Excess Kurtosis	-0.252		
Dispersion Index	1.320		



Figure 2. ACF plot of the series





We first fit the data given by the 1<sup>st</sup>(i.e. when  $i_w = 1$ ) window of size n=50 with the MPN model of order one (i.e., when k=2). The 51 data points used to form the first window are also used to fit the INARCH and INGARCH models with Poisson and negative binomial distributions. The INARCH and INGARCH models of different orders are considered. These models include the

INARCH(1), INARCH(2) and INGARCH(1,1). The parameter estimates of the MPN model, INARCH and INGARCH models are respectively shown in Tables 2 and 3. A total of N = 45 one-step ahead out-of-sample rolling window monthly claims counts forecasts from April 1991 to December 1994 are obtained for all the models. The forecasting performances of these models are compared using two out-of-sample criteria, namely RMSE and MAD. The prediction intervals based on the MPN model for predicting the 45 monthly claims counts are shown in Figure 4 along with the corresponding predicted and observed values. The estimated coverage probability and average length of the prediction intervals is 11.182 and the estimated coverage probability of the prediction interval is 0.956 which is close to the target value 0.95.

Table 2.	Parameter	estimates,	coverage	probability	and av	erage l	ength o	of the
MPN me	odel of orde	<b>r</b> one $(T = 9)$	96, n = 50,	$k = 2, i_w = 1$ )				

MPN model of order one				
$\mu_1 = 9.340$	$\mu_2 = 9.340$			
$\sigma_1 = 2.654$	$\sigma_2 = 3.996$			
$\lambda_1^- = 1.150$	$\lambda_2^-=0.350$			
$\lambda_{1}^{+} = 0.720$	$\lambda_2^{\scriptscriptstyle +} = 0.010$			
$\mathbf{H} = \begin{bmatrix} 0.812 & -0.584 \\ 0.584 & 0.812 \end{bmatrix}$				
Estimated Coverage Probability = 0.956				
Average Length = 11.182				



Figure 4. Observed and predicted values together with the 95% prediction interval found by using the MPN model of order one

Table 3. Parameter estimates, out-of-sample forecast criteria for INARCH, INGARCH and MPN models. Values in italic are standard errors of the parameter estimates  $(i_w = 1)$ 

			Model		
Parameter	INARCH	INARCH	NBINARCH	NBINARCH	NBINARCH
	(1)	(2)	(1) with $r = 1$	(1) with $r = 2$	(2) with <i>r</i> =1
ω	5.0440	2.9642	4.6465	2.3380	2.7585
	1.2485	1.3475	3.7685	1.4099	3.8706
α.	0.4600	0.2879	0.5053	0.2510	0.2999
5.1	0.1330	0.1462	0.4292	0.1596	0.4710
$\alpha_{2}$		0.3904			0.4021
		0.1460			0.4916
RMSE	2.986	3.130	2.983	4.716	3.139
MAD	2.428	2.514	2.427	4.096	2.512

### Table 3- continued

			Model		
Parameter	NBINARCH	INGARCH	NBINGARCH	NBINGARCH	MPN
	(2) with $r =$	(1,1)	(1,1) with $r = 1$	(1,1) with $r=2$	model of
	2				order one
ω	1.3878	1.0107	0.8874	0.6732	-
	1.4522	0.8454	2.3930	1.0810	

Ng	Kok-Haur,	Koh	You-Beng,	Pooi	Ah-Hin
<u> </u>					

α.	0.1500	0.3246	0.3396	0.2039	-
<i>a</i> <sub>1</sub>	0.1750	0.1069	0.3287	0.1144	
$\alpha_2/\beta_1$	0.2000	0.5682	0.5673	0.4470	-
2.71	0.1811	0.1397	0.4148	0.2713	
RMSE	4.797	3.091	3.091	4.771	3.056
MAD	4.173	2.498	2.493	4.140	2.459

From Table 3, we observe that the NBINARCH(1) with r=1 gives the smallest RMSE among the fitted models. The models with slightly larger RMSE are the INARCH(1), MPN, INGARCH(1,1), NBINARCH(2) with r=1 and NBINGARCH(1,1) with r=1. By comparison, the remaining models in Table 3do not perform well. As expected, the NBINARCH(1) model with r=1 performs better than the INARCH(1) model which is based on the Poisson distribution because the dispersion index exhibited by the data is more than 1.Table 3 shows that similar forecasting performances of the models can be observed if we use the MAD criteria.

In using the MPN model, we do not have to take note of the dispersion index because the selection of the distribution is automatically taken care by the choice of the parameters  $\lambda^+$  and  $\lambda^-$ . In using the discrete models, we need to decide whether it is necessary to include lag values of the mean  $\lambda_i$  in the model. Thus the process of model selection is very much simplified if we use the MPN model. Another advantage of the MPN model is that it provides directly the prediction intervals without having to generate them using another method such as the bootstrap procedure. As the RMSE of the MPN model is only marginally larger than those of the NBINARCH(1) with r=1 and INARCH(1) models, and the prediction interval is available without further computation, it is worthwhile to consider using the MPN model to analyse the discrete time series.

To compare further the forecasting performances of these models, we carry out the Mincer-Zarnowitz (MZ) test based on regression approach and the Diebold-Mariano (DM) test which is a pair-wise test of equal forecast accuracy. Suppose that we have a time series  $X_t$ ;  $t = 1, 2, \dots, n+m$  of count data, where a model is fitted using the first *n* observations. Let  $\hat{X}_t^{(1)}$  and  $\hat{X}_t^{(2)}$  be the fitted values found by using the MPN model and the INGARCH models respectively.

(i) Mincer-Zarnowitz regression approach considers the following model

$$X_t = a + b \hat{X}_t^{(j)} + u_t, \qquad j = 1, 2,$$

where the coefficients a and b are parameters to be estimated and  $u_t$  represents the error term. An F-test on the joint hypothesis a=0 and b=1 is carried out. The insignificant result indicates that the forecasts are unbiased.

(ii) To apply the DM test, we first compute the forecast errors from the two models:

 $\varepsilon_{n+h}^{(1)} = X_{n+h} - \hat{X}_{n+h}^{(1)}, \qquad h = 1, 2, \cdots, m,$ 

$$\mathcal{E}_{n+h}^{(2)} = X_{n+h} - \hat{X}_{n+h}^{(2)}, \qquad h = 1, 2, \cdots, m.$$

and

The accuracy of each forecast is measured by a suitable loss function,  $L(\varepsilon_{n+h}^{(i)})$ , i = 1,2. Two popular loss functions are the absolute deviation loss and square error loss given by

- Absolute deviation loss:  $L(\varepsilon_{n+h}^{(i)}) = |\varepsilon_{n+h}^{(i)}|$ ,
- Square error loss:  $L(\varepsilon_{n+h}^{(i)}) = (\varepsilon_{n+h}^{(i)})^2$ .

The following DM test statistic evaluates the forecasts in terms of an arbitrary loss function  $L(\cdot)$ :

$$DM = \frac{\sum_{h=1}^{m} [L(\varepsilon_{n+h}^{(1)}) - L(\varepsilon_{n+h}^{(2)})]/m}{\sqrt{S^2/m}}$$

where  $S^2$  is an estimator of the variance of  $d_h = L(\varepsilon_{n+h}^{(1)}) - L(\varepsilon_{n+h}^{(2)})$ . Under the null hypothesis of equal forecast accuracy, the distribution of the DM statistic is approximately standard normal (see Diebold and Mariano, 1995).

Table 4 shows the results of MZ test. We observe that the p-values of joint F-test for the models NBINARCH(1) with r = 2 and NBINGARCH(1,1) with r = 2 exceed the significance level of 0.05 showing sufficient evidence to reject the null hypothesis. These findings are also reflected at RMSE and MAD of these two models. Table 5 shows the DM test statistics and their corresponding p-values. The results also show that the MPN model and NBINARCH(1) with r = 2 and NBINGARCH(1,1) with r = 2 are significance different in the forecasting ability at 0.05 significant level.

Table 4. Coefficients, standard errors in italic, p-values for the MZ test based on out-of-sample forecasts using the MPN model and various INARCH and INGARCH models

	MPN	INARCH(1)	NBINARCH(1)	NBINARCH(1)
			with $r = 1$	with $r = 2$
а	2.2161	1.4856	1.9650	1.9121
	2.3476	2.5078	2.2428	2.2689
ра	0.3504	0.5567	0.3858	0.4040
b	0.6657	0.7626	0.7088	1.4297
	0.2725	0.2960	0.2650	0.5362

Ng Kok-Haur, Koh You-Beng, Pooi Ah-Hin

	0.2266	0.4270	0.2779	0.4273
	0.2200	01.127.0	0.2779	01.1270
pb				
<b>r</b> -				
F	1.6783	0.9363	1.1196	34.997
	0.1007	0.0000	0.0055	0.0000
pF	0.1987	0.3999	0.3357	0.0000
r-				

#### Table 4 - continued

	INGARCH(1,1)	NBINGARCH(1,1)	NBINGARCH(1,1)
		with $r = 1$	with $r = 2$
а	2.9193	3.0659	2.7433
	2.5115	2.2970	2.2956
ра	0.2515	0.1890	0.2386
b	0.5983	0.5812	1.2399
	0.3000	0.2738	0.5470
pb	0.1876	0.1334	0.6631
F	1.2547	1.5115	34.045
pF	0.2954	0.2321	0.0000

\*F is refer to the value of F-test. pa, pb and pF are p-values for t-tests and F-test respectively.

## 5. Conclusion

This paper uses the MPN model to analyse a set of discrete time series data. The performance of the MPN model is found to be comparable to the best INGARCH model.

The other advantages of the MPN model include the ease in selecting distributions, a simpler process of model selection and the readily available prediction intervals.

It is also noted that the MPN model permits the investigation of the effects of the explanatory variables on the future observation if we form a new value of  $\mathbf{y}$  of which the initial components are the values of the explanatory variables followed by the desired number of y at consecutive time points.

Table 5. The results of the Diebold-Mariano test for m=45 one-step ahead forecast, with p-values in parentheses. The benchmark is the MPN model

Loss function	INARCH(1)	NBINARCH(1)	NBINARCH(1)
based on		with $r = 1$	with $r = 2$
Absolute	0.60838	0.63581	-3.4878
Deviation	(0.5461)	(0.5282)	(0.0011)
Square Error	1.1397	1.3136	-3.2389
	(0.2606)	(0.1958)	(0.0023)

#### Table 5 - continued

Loss function	INGARCH(1,1)	NBINGARCH(1,1)	NBINGARCH(1,1)
based on		with $r = 1$	with $r = 2$
Absolute	-0.56274	-0.48761	-3.5543
Deviation	(0.5765)	(0.6282)	(0.0009)
Square Error	-0.46994	-0.52661	-3.3016
	(0.6407)	(0.6011)	(0.0019)

#### ACKNOWLEDGEMENTS

Authors would like to thank the editor and anonymous referees for their helpful comments. This work is partially supported by the Ministry of Higher Education (MOHE), Malaysia under Fundamental Research Grant Scheme (FRGS), no: FP041-2017A and the University of Malaya Special Research Fund Assistance (BKS), no: BKS076-2017.

#### REFERENCES

[1] Agosto, A., Cavaliere, G., Kristensen, D. and Rahbek, A. (2016), Modeling Corporate Defaults: Poisson Autoregressions with Exogenous Covariates (PARX); Journal of Empirical Finance, 38: 640-663;

[2] Al-Osh, M.A. and Alzaid, A.A. (1987), *First-order Integer-valued Autoregressive (INAR(1)) Process; Journal of Time Series Analysis*, 8 (3): 261-275;

[3] Davis, R.A. and Wu, R. (2009), *A Negative Binomial Model for Time Series* of Counts; *Biometrika*, 96 (3): 735-749;

[4] Diebold, F.X. and Mariano, R.S. (1995), *Comparing Predictive* 

Accuracy; Journal of Business & Economic Statistics, 13 (3): 253-263;

[5] **Freeland, R.K.** (1998), Statistical Analysis of Discrete Time Series with Application to the Analysis of Workers' Compensation Claims Data; PhDthesis, The University of British Columbia, Canada;

[6] Ferland, R., Latour, A. and Oraichi, D. (2006), *Integer-valued GARCH Process*; *Journal of Time Series Analysis*, 27 (6): 923-942;

[7] Groβ-Klubmann, A. and Hautsch, N. (2013),*Predicting Bid-ask Spreads Using Long-memory Autoregressive Conditional Poisson Models*;*Journal of Forecasting*, 32 (8): 724-742;

[8] Heinen, A. (2003), Modelling Time Series Count Data: An Autoregressive Conditional Poisson Model; Discussion paper, vol2003/62. Center for operations research and econometrics (CORE), Catholic University of Louvain, Belgium; [9] McKenzie, E. (1985), Some Simple Models for Discrete Variate Time Series; Water Resources Bulletin, 21 (4): 645-650;

[10] Mincer, J.A. and Zarnowitz, V. (1969), *The Evaluation of Economic Forecasts*; Chapter 1, Pages 3-46. National Bureau of Economic Research, New York;

[11] **Pooi, A.H. (2012)**, *A Model for Time Series Analysis*; *Applied Mathematical Sciences*, 6(115): 5735-5748;

[12] Weiβ, C.H. (2009), *Modelling Time Series of Counts with Overdispersion*; *Statistical Methods and Applications*, 18(4): 507-519;

[13] Weiβ, C.H. (2010), *The INARCH (1) Model for Overdispersed Time Series of Counts*; *Communications in Statistics: Simulation and Computation*, 39 (6): 1269-1291;

[14] Yeo, I.K. and Johnson, R.A. (2000), A New Family of Power Transformations to Improve Normality or Symmetry; Biometrika, 87 (4): 954-959;
[15] Zhu, F. (2011), A Negative Binomial Integer-valued GARCH Model; Journal of Time Series Analysis, 32 (1): 54-67;

[16] Zhu, F. (2012), Zero-inflated Poisson and Negative Binomial Integervalued GARCH Models; Journal of Statistical Planning and Inference, 142 (4): 826-839;

[17] Zhu, F., Li, Q. and Wang, D. (2010), A Mixture Integer-valued ARCH
Model; Journal of Statistical Planning and Inference, 140 (7): 2025-2036;
[18] Zhu, F. and Wang, D. (2010), Diagnostic Checking Integer-valued

*ARCH(p) Models Using Conditional Residual Autocorrelations*; *Computational Statistics and Data Analysis*, 54(2): 496-508;

[19] Zhu, F. and Wang, D. (2011), *Estimation and Testing for a Poisson Autoregressive Model*;*Metrika*, 73 (2): 211-230.