Portmanteau test for count time series models

Ali AHMAD

ali.ahmad@etu.univ-lille3.fr

Résumé. Nous proposons un test portmanteau robuste et général pour les modèles des séries temporelles à valeurs entières. Le test représente un outil fiable pour évaluer l'adéquation de l'ajustement pour des classes larges et importantes des modèles. Par exemple, les modèles INGARCH(p, q), les modèles log-linéaire(1,1), les modèles non-linéaires(1,1) avec diverses distributions conditionnelles et les modèles INAR(p) avec différentes distributions marginales des innovations. La distribution asymptotique de la statistique est dérivée et ses propriétés à distance finie sont étudiées par des simulations de Monte Carlo.

Mots-clés. Test Portmanteau, Test de validation, Modèles INGARCH(p,q), Modèles log-linéaire(1,1), Modèles non-linéaires(1,1), Modèles INAR(p), L'estimateur de quasi maximum de vraisemblance de Poisson, Séries temporelles à valeurs entières.

Abstract. We propose a robust and general goodness-of-fit test for the count time series models. The test represents a very useful tool for checking the adequacy of fit for wide and important classes of models. For example, the INGARCH(p,q), the log-linear(1,1), the non-linear(1,1) models with diverse conditional distributions and the INAR(p) models with several distributional cases of the innovations. The asymptotic distribution of the statistic is derived and its finite sample properties are studied through Monte Carlo simulations.

Keywords. Portmanteau, Goodness-of-Fit, INGARCH models, INAR models, Log-linear models, Non-linear models, Poisson quasi-maximum likelihood estimator, Time series of counts.

1 Introduction

The portmanteau test is considered as one of the most important tools for evaluating the goodness-of-fit in the context of the time series analysis. Firstly, this test based on the sum of the squared autocorrelation functions of the residuals, was limited to be used when the residuals are independent and identically distributed (iid) (see Box and Pierce, 1970 and Ljung and Box, 1978 for more details). Lot of modifications have been proposed in the literature to make this test able to treat with more general residual cases (see *e.g* Francq et al. (2005) and the references therein). For the time series of counts which found, in the recent years, a remarkable interest, the goodness-of-fit tests still less developed.

However, several important breakthroughs have been achieved in this field. For instance, Neumann et al. (2011) proposed to use the conditional equi-dispersion assumption of the Poisson INGARCH(p,q) model for testing the adequacy of the specification of the intensity process. In addition, Fokianos et al. (2013) have been studied a non parametric goodness-of-fit test for the Poisson INGARCH(p,q) models. Recently, Meintanis and Karlis (2014) propsed a goodness-of-fit test for the innovation distribution of the Poisson INAR(1) model. Schweer (2016) has been introduced a more general test in which the empirical joint probability generating function is considered for testing the adequacy of wide class of count time series models, but this test depends on arbitrary parameter. In addition, the generalization of the methodology of this test to be used for higher-order models seems quite challenging. In this contribution, the adopted approach is simpler and much more general than the previous mentioned tests. Where, under some conditions, the test can be used as a diagnostics checking tool for INGARCH(p,q), log-linear and non-linear models with large variety of exponential discrete conditional distributions having non negative integer-valued supports. Moreover, the test can also be used to evaluate the adequacy of fit for the INAR(p) model for several distributional cases of the innovations. This goodness-of-fit test is based on the residuals autocovariances obtained after estimating the model using Poisson quasi maximum likelihood estimator (PQMLE) studied by Ahmad and Francq (2015).

2 Model and assumptions

Assume that $\{X_t \in \mathbb{N}\}$ is a count time series, such that

$$E\left(X_t \mid \mathcal{F}_{t-1}\right) = \lambda_t(\theta_0) = \lambda(X_{t-1}, X_{t-2}, \ldots; \theta_0), \tag{2.1}$$

where $\mathcal{F}_{t-1}$ denotes the $\sigma-$field generated by $(X_u, u < t)$,

$$\lambda \text{ is a measurable function valued in } (\underline{\omega}, +\infty) \text{ for some } \underline{\omega} > 0 \tag{2.2}$$

and θ_0 is an unknown parameter belonging to some parameter space $\Theta \subset \mathbb{R}^d$. We assume also that the fourth-order moment of the marginal distribution of X_t exists

$$EX_t^4 < \infty. \tag{2.3}$$

We define the residual as follows

$$\epsilon_t(\theta_0) = X_t - \lambda_t(\theta_0). \tag{2.4}$$

If the conditional mean is correctly specified, under the stationarity assumption, one can show that $\epsilon_t(\theta_0)$ is uncorrelated white noise sequence, where

$$E(\epsilon_t(\theta_0)) = E\left(E\left(X_t - \lambda_t(\theta_0) | \mathcal{F}_{t-1}\right)\right) = 0, \qquad var(\epsilon_t(\theta_0)) = E\left(var\left(X_t | \mathcal{F}_{t-1}\right)\right)$$

and, for $h > 0$, we have

$$cov(\epsilon_t(\theta_0)\epsilon_{t+h}(\theta_0)) = E\left(\epsilon_t(\theta_0)E\left(\epsilon_{t+h}(\theta_0)|\mathcal{F}_{t+h-1}\right)\right) = 0.$$

The above formulation can accommodate large variety of count time series models. For example, the INGARCH(p,q) models (see *e.g* Heinen, 2003, Ferland et al., 2006, Zhu, 2012 and Christou and Fokianos, 2014), the log-linear model (see Fokianos and Tjøstheim, 2011), the non-linear model (see Fokianos and Tjøstheim, 2012) and the INAR(p) models (see *e.g* Alzaid and Al-Osh, 1990). According to Ahmad and Francq (2015), these models can be consistently estimated by Poisson Quasi Maximum Likelihood Estimator (PQMLE) which is defined as any measurable solution of

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \widetilde{L}_n(\theta), \quad \widetilde{L}_n(\theta) = \frac{1}{n}\sum_{t=s+1}^{n} \widetilde{\ell}_t(\theta), \tag{2.5}$$

where $\widetilde{\ell}_t(\theta) = -\widetilde{\lambda}_t(\theta) + X_t \log \widetilde{\lambda}_t(\theta)$, $\widetilde{\lambda}_t(\theta)$ is obtained by setting to some integer x_0 the unknown initial values $X_0, X_1, ...$ involved in $\lambda_t(\theta)$. This value x_0 can either be a fixed integer, for instance, $x_0 = 0$, or a value depending on θ, or a value depending on the observations. The integer s is asymptotically unimportant, but it can affect the finite sample behaviour of the PQMLE by reducing the impact of the initial value x_0. Under some regularity conditions, PQMLE is consistent and $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ is asymptotically normal with mean 0 and covariance matrix $\Sigma_\theta := J^{-1}IJ^{-1}$, where

$$J = E\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'}, \quad I = E\frac{var(X_t|\mathcal{F}_{t-1})}{\lambda_t^2(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'}. \tag{2.6}$$

3 Portmanteau test

As mentioned in the introduction, the goodness-of-fit test is based on the residuals autocovariances. Firstly, the residuals considered in the test are defined as follows

$$\widetilde{\epsilon}_t(\widehat{\theta}_n) = X_t - \widetilde{\lambda}_t(\widehat{\theta}_n).$$

We denote by $\widehat{\gamma}_m = (\widehat{\gamma}_{(1)}, .., \widehat{\gamma}_{(m)})'$ the vector of residuals autocovariance functions, where, for $m < n$ and $h \in \{1, .., m\}$, its elements are given by

$$\widehat{\gamma}_{(h)} = \frac{1}{n}\sum_{t=h+1}^{n} \widetilde{\epsilon}_t(\widehat{\theta}_n)\widetilde{\epsilon}_{t-h}(\widehat{\theta}_n).$$

Define the $m \times m$ matrix $\widehat{\Sigma}_{\gamma_m}$, whose the elements for $(h, l) \in \{1, .., m\}$ are given as follows

$$\widehat{\Sigma}_{\gamma_m}(h, l) = \frac{1}{n}\sum_{t=max(h,l)+1}^{n} \widetilde{\epsilon}_t^2(\widehat{\theta}_n)\widetilde{\epsilon}_{t-h}(\widehat{\theta}_n)\widetilde{\epsilon}_{t-l}(\widehat{\theta}_n).$$

Let $\widehat{\Sigma}_\theta$ a consistent estimator of the matrix Σ_θ defined in Section 2

$$\widehat{\Sigma}_\theta = \widehat{J}^{-1}\widehat{I}\widehat{J}^{-1},$$

where

$$\widehat{J} = \frac{1}{n}\sum_{t=s+1}^{n} \frac{1}{\widetilde{\lambda}_t(\widehat{\theta}_n)} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta'} \quad \text{and} \quad \widehat{I} = \frac{1}{n}\sum_{t=s+1}^{n} \left(\frac{X_t}{\widetilde{\lambda}_t(\widehat{\theta}_n)} - 1 \right)^2 \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta'}.$$

Let d denotes the number of parameters of the model, we define the $d \times m$ matrices $\widehat{C}_m$ and $\widehat{\Sigma}_{\widehat{\theta}_n,\gamma_m}$ whose the $(k,h)th$ elements, for $1 \leq k \leq d$ and $1 \leq h \leq m$, are obtained respectively by

$$\widehat{C}_m(k,h) = -\frac{1}{n}\sum_{t=h+1}^{n} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta_k} \widetilde{\epsilon}_{t-h}(\widehat{\theta}_n)$$

and

$$\widehat{\Sigma}_{\widehat{\theta}_n,\gamma_m}(k,h) = \widehat{J}^{-1}\frac{1}{n}\sum_{t=h+1}^{n} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta_k} \frac{\widetilde{\epsilon}_t^2(\widehat{\theta}_n)}{\widetilde{\lambda}_t(\widehat{\theta}_n)} \widetilde{\epsilon}_{t-h}(\widehat{\theta}_n).$$

Theorem 3.1 *Under the assumption* (2.3) *and the other regularity conditions required for the PQMLE (see Ahmad and Francq, 2015), we have*

$$n\widehat{\gamma}_m' \widehat{\Sigma}_{\widehat{\gamma}_m}^{-1} \widehat{\gamma}_m \xrightarrow{\mathcal{L}} \chi_m^2,$$

where

$$\widehat{\Sigma}_{\widehat{\gamma}_m} = \widehat{\Sigma}_{\gamma_m} + \widehat{C}_m' \Sigma_\theta \widehat{C}_m + \widehat{C}_m' \widehat{\Sigma}_{\widehat{\theta}_n,\gamma_m} + \widehat{\Sigma}_{\widehat{\theta}_n,\gamma_m}' \widehat{C}_m.$$

The adequacy of model is rejected at the asymptotic level $\underline{\alpha}$ when

$$n\widehat{\gamma}_m' \widehat{\Sigma}_{\widehat{\gamma}_m}^{-1} \widehat{\gamma}_m > \chi_m^2(1-\underline{\alpha}).$$

4 Example of the Monte Carlo simulation results

To examine the finite sample behaviour of the test defined in Theorem 3.1, we report a Monte Carlo simulation with 1000 independent replications. We evaluate the size and the power of the test for the INGARCH(p,q) with three conditional distributions: Poisson $\mathcal{P}(\lambda_t)$, negative binomial $\mathcal{NB}(p_t,\nu)$ and double-Poisson $\mathcal{DP}(\lambda_t,\gamma)$ distributions. The INGARCH(p,q) models are defined by assuming that the conditional mean take the following general linear representation

$$E\left(X_t \mid \mathcal{F}_{t-1}\right) = \lambda_t(\theta_0) = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} X_{t-i} + \sum_{j=1}^{p} \beta_{0j} \lambda_{t-j}(\theta_0), \tag{4.1}$$

where $\omega_0 > 0$, $0 \leq \alpha_{0i} < 1$ $(i = 1, .., q)$ and $0 \leq \beta_{0j} < 1$ $(j = 1, .., p)$. The models are estimated using PQMLE. For each replication, we carried out the portmanteau test for evaluate the adequacy of the fitted models at asymptotic level $\underline{\alpha} = 5\%$. In view of Table 1, we can note that the empirical sizes for all the models are satisfactorily close to the theoretical nominal level, especially when the sample size is large. Moreover, the results summarized in Table 2 show that the proposed test achieves good power even when the sample size is relatively small.

Table 1: Size of test for the INGARCH model

INGARCH(1,1) $\omega_0 = 2,\ \alpha_0 = 0.3,\ \beta_0 = 0.6 \quad df = m$									
	m=4			m=12			m=20		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$
					$\underline{\alpha} = 5\%$				
500	5.1	5.5	4.9	3.2	3.7	4.3	3.1	3.2	3.9
1000	4.2	5.2	4.2	4.3	4.4	5.5	4.3	3.5	4.4
4000	4.7	6.2	4.9	5.1	5.7	5.4	5	4.5	3.9

Table 2: Power of test for the INGARCH model

INGARCH(1,1) vs INGARCH(1,2) $\omega_0 = 2,\ \alpha_{01} = 0.4,\ \alpha_{02} = 0.3,\ \beta_0 = 0.2 \quad df = m$									
	m=4			m=12			m=20		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$
					$\underline{\alpha} = 5\%$				
500	69.3	31.6	74.3	48.4	18.3	50.6	33.9	12.1	37.3
1000	96.4	50.6	96.9	87	34.7	88.3	78.8	24.4	80.5
4000	100	90.9	100	100	85	100	100	80.4	100

References

Ahmad, A. and Francq, C. (2015). Poisson QMLE of count time series models. *Journal of Time Series Analysis*, forthcoming.

Alzaid, A. and Al-Osh, M. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability* 27, 314–324.

Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressiveintegrated moving average time series models. *Journal of the American statistical Association* 65, 1509–1526.

Christou, V. and Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis* 35, 55–78.

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

Fokianos, K., Neumann, M. H., et al. (2013). A goodness-of-fit test for Poisson count processes. *Electronic Journal of Statistics* 7, 793–819.

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

Fokianos, K. and Tjøstheim, D. (2012). Nonlinear Poisson autoregression. *Annals of the Institute of Statistical Mathematics* 64, 1205–1225.

Francq, C., Roy, R., and Zakoïan, J.-M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* 100, 532–544.

Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. *CORE Discussion Paper 2003/62, Université Catholique de Louvain*.

Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.

Meintanis, S. G. and Karlis, D. (2014). Validation tests for the innovation distribution in INAR time series models. *Computational Statistics* 29, 1221–1241.

Neumann, M. H. et al. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* 17, 1268–1284.

Schweer, S. (2016). A goodness-of-fit test for integer-valued autoregressive processes. *Journal of Time Series Analysis* 37, 77–98.

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