

TESTS D'ADÉQUATION DE MODÈLES ARMA FAIBLES: UNE APPROCHE BASÉE SUR L'AUTO-NORMALISATION

Yacouba Boubacar Mainassara ¹ & Bruno Saussereau ^{1,2}

¹ *Université de Franche-Comté, UFR Sciences et Techniques
Laboratoire de Mathématiques de Besançon - UMR 6623 CNRS-UFC
16 route de Gray 25030 Besançon Cedex, France.*

E-mail : yacouba.boubacar_mainassara@univ-fcomte.fr

² *E-mail : bruno.saussereau@univ-fcomte.fr*

Résumé. Dans ce travail, nous considérons les tests portmanteau, aussi appelés tests d'autocorrélation, pour tester l'adéquation de modèles autoregressive moving-average (ARMA) avec innovations linéaires non corrélées mais non nécessairement indépendantes (*i.e.* ARMA faibles). Dans un premier temps, nous étudions la distribution asymptotique jointe, auto-normalisée, de l'estimateur des moindres carrés et des autocovariances empiriques du bruit. Ceci nous permet ensuite d'obtenir les distributions asymptotiques des autocovariances et autocorrélations résiduelles auto-normalisées. Enfin, nous déduisons le comportement asymptotique des statistiques portmanteau modifiées de Ljung-Box (ou Box-Pierce) de modèles ARMA faibles.

Mots-clés. Autocorrélations résiduelles, auto-normalisation, modèles ARMA faibles, tests portmanteau de Ljung-Box et Box-Pierce.

Abstract. We consider modified portmanteau tests for testing the adequacy of ARMA models under the assumption that the errors are uncorrelated but not necessarily independent (*i.e.* weak ARMA). We first study the self-normalization joint distribution of the least squared estimator and the noise empirical autocovariances. We then derive the asymptotic distribution of residual empirical autocovariances and autocorrélations self-normalized. We deduce the asymptotic distribution of the Ljung-Box (or Box-Pierce) modified portmanteau statistics for weak ARMA models.

Keywords. Box-Pierce and Ljung-Box portmanteau tests, residual autocorrelation, self-normalization, weak ARMA models.

1 Introduction

On dit qu'un processus stationnaire au second ordre $(X_t)_{t \in \mathbb{Z}}$, à valeurs réelles, admet une représentation ARMA(p, q) faible, où p et q sont des entiers (souvent appelés ordres), s'il existe des réels a_1, \dots, a_p et b_1, \dots, b_q tels que :

$$X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} = \epsilon_t - b_1 \epsilon_{t-1} - \dots - b_q \epsilon_{t-q}, \quad (1)$$

où ϵ_t est un bruit blanc faible, c'est-à-dire une suite de variables aléatoires centrées, non corrélées et de même variance σ^2 . Par opposition, nous appelons ARMA forts les modèles standard dans lesquels le terme d'erreur est supposé être une suite indépendante et identiquement distribuée (iid). Dans la modélisation des séries temporelles, la validité des différentes étapes de la méthodologie traditionnelle de Box et Jenkins, à savoir les étapes d'identification, d'estimation et de validation, dépend des propriétés du bruit blanc. Après l'identification et l'estimation du processus ARMA, la prochaine étape importante dans la modélisation consiste à vérifier si le modèle estimé est compatible avec les données. Cette étape d'adéquation permet de valider ou d'invalidier le choix des ordres du modèle. Ce choix est important pour la précision des prévisions linéaires et pour une bonne interprétation du modèle.

La sélection d'ordres trop grands a pour effet d'introduire des termes qui ne sont pas forcément pertinents dans le modèle et aussi d'engendrer des difficultés statistiques comme par exemple un trop grand nombre de paramètres à estimer, ce qui est, aussi, susceptible d'engendrer une perte de précision de l'estimation des paramètres. Le praticien peut aussi choisir des ordres trop petits qui entraînent la perte d'une information qui peut être détectée par une corrélation des résidus ou encore une estimation non convergente des paramètres. L'objectif de ce travail est d'étudier le comportement asymptotique du test portmanteau dans le cadre de modèles ARMA (faibles) dont les termes d'erreur sont non corrélés mais peuvent contenir des dépendances non linéaires par une approche d'auto-normalisation.

Ainsi pour des ordres p et q donnés, nous testons l'hypothèse nulle

(H0) : $(X_t)_{t \in \mathbb{Z}}$ satisfait une représentation ARMA(p, q) ;

contre l'alternative

(H1) : $(X_t)_{t \in \mathbb{Z}}$ n'admet pas une représentation ARMA(p, q) ou $(X_t)_{t \in \mathbb{Z}}$ satisfait une représentation ARMA(p', q') avec $p' > p$ or $q' > q$.

Le test portmanteau a été introduit par Box et Pierce (1970) (noté BP dans la suite) afin de mesurer la qualité d'ajustement d'un modèle ARMA fort univarié. Ce test est basé sur les résidus $\hat{\epsilon}_t$ résultant de l'estimation des paramètres du modèle (1). Le test portmanteau modifié que nous proposons est, en partie, une extension aux cas de modèles ARMA faibles des résultats de Lobato (2001). Pour l'estimation du paramètre, $\theta_0 = (a_1, \dots, a_p, b_1, \dots, b_q) \in \Theta^*$; où Θ^* est un espace compact de l'espace des paramètres

$$\Theta := \left\{ \theta = (\theta_1, \dots, \theta_p, \theta_{p+1}, \dots, \theta_{p+q})' \in \mathbb{R}^{k_0}, k_0 := p + q : \right.$$

$$a(z) = 1 - \sum_{i=1}^p a_i z^i \text{ and } b(z) = 1 + \sum_{i=1}^q b_i z^i$$

ont leurs racines à l'extérieur du disque unité et n'ont pas de racines communes $\left. \vphantom{\theta} \right\}$,

du modèle (1), nous utiliserons la méthode des moindres carrés définie dans Francq et Zakoïan (1998).

Ces auteurs ont établi la convergence forte et la normalité asymptotique de l'estimateur $\hat{\theta}_n$ des moindres carrés, sous les hypothèses essentielles d'ergodicité et de mélange suivantes

A1: Le processus (ϵ_t) est stationnaire et ergodique;

A2: Il existe un réel $\nu > 0$ tel que $\mathbb{E}\|\epsilon_t\|^{4+2\nu} < \infty$ et les coefficients de mélange du processus (ϵ_t) vérifient $\sum_{k=0}^{\infty} \{\alpha_{\epsilon}(k)\}^{\frac{\nu}{2+\nu}} < \infty$.

A3: Nous avons $\theta_0 \in \overset{\circ}{\Theta}$, où $\overset{\circ}{\Theta}$ est l'intérieur du sous espace compact Θ de l'espace des paramètres.

2 Comportement asymptotique des statistiques modifiées

Définissons, respectivement, les fonctions d'autocovariances "empiriques" du bruit et résiduelles $\gamma(h) = n^{-1} \sum_{t=h+1}^n \epsilon_t \epsilon'_{t-h}$ et $\hat{\gamma}_{\epsilon}(h) = n^{-1} \sum_{t=h+1}^n \hat{\epsilon}_t \hat{\epsilon}'_{t-h}$ pour $0 \leq h < n$. Nous considérons aussi des vecteurs des $m \geq 1$ premières autocovariances "empiriques" du bruit et résiduelles

$$\gamma_m = (\gamma(1), \dots, \gamma(m))', \text{ et } \hat{\gamma}_m = (\hat{\gamma}(1), \dots, \hat{\gamma}(m))'.$$

Soit $(B_k(r))_{r \geq 0}$ un mouvement brownien vectoriel de dimension k , commençant à 0. Notons, pour $k \geq 1$, la variable aléatoire U_k , dont les valeurs critiques sont tabulées dans Lobato (2001), définie par

$$U_k = B'_k(1) V_k^{-1} B_k(1) \text{ où } V_k = \int_0^1 (B_k(r) - r B_k(1)) (B_k(r) - r B_k(1))' dr. \quad (2)$$

Enfin, nous posons

$$\Upsilon_t = \left(-\epsilon_t(\theta_0) \frac{\partial \epsilon_t(\theta_0)}{\partial \theta}, \epsilon_t \epsilon_{t-1}, \dots, \epsilon_t \epsilon_{t-m} \right)' \text{ et } S_t = \frac{1}{n} \sum_{j=1}^t \left(\Upsilon_j - \frac{1}{n} \sum_{t=1}^n \Upsilon_t \right). \quad (3)$$

Le Théorème suivant, qui est une extension du résultat donné dans le lemme 1 de Lobato (2001), donne la distribution asymptotique jointe auto-normalisée de $\hat{\theta}_n$ et γ_m .

Théorème 2.1 *Sous les hypothèses A1, A2 et A3, quand $n \rightarrow \infty$, nous avons*

$$n(\hat{\theta}_n - \theta_0, \gamma_m)' C_{k_0+m}^{-1} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \gamma_m \end{pmatrix} \xrightarrow{d} U_{k_0+m} := B'_{k_0+m}(1) V_{k_0+m}^{-1} B_{k_0+m}(1),$$

où $C_{k_0+m} = n^{-2} \sum_{t=1}^n S_t S_t'$ est une matrice de taille $(k_0 + m) \times (k_0 + m)$.

Soit Γ la matrice bloc de $\mathbb{R}^{m \times (p+q+m)}$ définie par

$$\Gamma = \begin{pmatrix} -2\Phi_m J^{-1} | I_m \end{pmatrix}, \quad (4)$$

où I_m est la matrice identité d'ordre m et où la matrice d'information de Fisher

$$J = \lim_{n \rightarrow \infty} \frac{\partial^2 O_n(\theta_0)}{\partial \theta \partial \theta'} \text{ a.s. où } O_n(\theta) = \frac{1}{2n} \sum_{t=1}^n \epsilon_t^2(\theta), \quad (5)$$

et la matrice Φ_m est donnée par

$$\Phi_m = \mathbb{E} \left\{ (\epsilon'_{t-1}, \dots, \epsilon'_{t-m})' \partial \epsilon_t(\theta_0) / \partial \theta' \right\}. \quad (6)$$

Définissons, respectivement, les autocorrélations théoriques et résiduelles de retard ℓ par $\rho(\ell) = \gamma(\ell)/\gamma(0)$ and $\hat{\rho}(\ell) = \hat{\gamma}(\ell)/\hat{\gamma}(0)$, avec $\gamma(0) := \sigma^2$. Nous considérons des vecteurs des $m \geq 1$ premières autocorrélations résiduelles $\hat{\rho}_m = (\hat{\rho}(1), \dots, \hat{\rho}(m))'$.

Le résultat suivant donne le comportement asymptotique des autocovariances et autocorrélations résiduelles normalisées.

Théorème 2.2 *Sous les hypothèses du Théorème 2.1, nous avons*

$$n\hat{\gamma}'_m \tilde{C}_m^{-1} \hat{\gamma}_m \xrightarrow{d} U_m := B'_m(1) V_m^{-1} B_m(1) \text{ et } n\sigma^4 \hat{\rho}'_m \tilde{C}_m^{-1} \hat{\rho}_m \xrightarrow{d} U_m,$$

où

$$\tilde{C}_m = \frac{1}{n^2} \sum_{t=1}^n \tilde{S}_t \tilde{S}'_t = \frac{1}{n^2} \sum_{t=1}^n \Gamma S_t S'_t \Gamma' = \Gamma C_{k_0+m} \Gamma'$$

avec

$$V_m = \int_0^1 (B_m(r) - rB_m(1)) (B_m(r) - rB_m(1))' dr.$$

Pour un niveau asymptotique de risque α , le test modifié de BP que nous proposons consiste à rejeter l'hypothèse d'adéquation de modèles ARMA(p, q) faibles quand la statistique modifiée est supérieure au quantile d'ordre α de la variable aléatoire U_m .

Ainsi quand le processus d'erreurs est un bruit blanc faible, la distribution asymptotique de la statistique de BP normalisée ne dépend plus des observations, contrairement à l'approche de Francq, Roy, et Zakoïan (2005) où la distribution asymptotique de la statistique de BP est une somme pondérée de chi-deux. Ce qui est, d'un point de vue, non négligeable en pratique.

3 Application numérique

Considérons le modèle ARMA(1, 1) suivant

$$X_t = aX_{t-1} + \epsilon_t + b\epsilon_{t-1}, \text{ where } \epsilon_t = \eta_t \sqrt{1 + \alpha_1 \epsilon_{t-1}^2}, \eta_t \sim \text{IID } \mathcal{N}(0, 1). \quad (7)$$

Les Tables 1 et 2 montrent que les résultats des tests modifiés de LB et de BP sont meilleurs que ceux des tests standard de LB et de BP et ceux de l’approche de Francq, Roy, et Zakoïan (2005). Il est donc préférable d’utiliser la distribution que nous proposons dans le théorème 2.2 plutôt que celle des sommes pondérées de chi-deux. Nous constatons aussi que, comme dans le cas où les erreurs sont gaussiennes, la statistique de LB a de meilleurs résultats pour des échantillons de petites tailles que celle de BP pour des erreurs dépendantes.

Table 1: Niveau empirique (en %) des tests standard et modifié dans le cas du modèle ARMA(0, 0) faible : modèle (7), avec $\theta_0 = (0, 0)'$ et $\alpha_1 = 0.4$. Le niveau asymptotique des tests est de $\alpha = 5\%$.

	$m = 2$		$m = 3$		$m = 4$	
Length n	300	1,000	300	1,000	300	1,000
LB _{BMS}	5.1	3.8	3.4	4.5	4.2	4.4
BP _{BMS}	5.1	3.9	3.4	4.5	4.0	4.4
LB _{FRZ}	3.2	4.0	3.0	3.9	3.1	4.0
BP _{FRZ}	3.1	4.0	2.5	3.9	3.1	4.0
LB _S	19.1	19.4	18.3	20.1	16.1	18.9
BP _S	18.8	19.4	17.5	20.0	15.6	18.7
	$m = 6$		$m = 12$		$m = 18$	
Length n	300	1,000	300	1,000	300	1,000
LB _{BMS}	4.8	4.6	1.9	4.7	1.3	4.9
BP _{BMS}	4.3	4.4	1.9	4.7	1.1	4.7
LB _{FRZ}	2.7	3.8	2.6	2.1	1.9	3.4
BP _{FRZ}	2.3	3.8	2.3	2.0	1.4	3.2
LB _S	14.3	16.2	11.3	13.7	11.7	13.1
BP _S	13.8	16.0	10.8	13.4	9.7	12.3

Bibliographie

[1] Box, G. E. P. and Pierce, D. A. (1970) Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–26.

Table 2: Niveau empirique (en %) des tests standard et modifié dans le cas du modèle ARMA(1,1) faible : modèle (7), avec $\theta_0 = (0.95, -0.6)'$ et $\alpha_1 = 0.4$. Le niveau asymptotique des tests est de $\alpha = 5\%$.

Length n	$m = 1$		$m = 2$		$m = 3$	
	1,000	2,000	1,000	2,000	1,000	2,000
LB_{BMS}	4.6	4.8	5.0	4.8	3.8	2.9
BP_{BMS}	4.5	4.7	5.0	4.8	3.8	2.9
LB_{FRZ}	4.3	4.8	3.9	4.7	2.1	2.7
BP_{FRZ}	4.3	4.8	3.8	4.7	2.1	2.7
LB_S	n.a.	n.a.	n.a.	n.a.	25.4	26.8
BP_S	n.a.	n.a.	n.a.	n.a.	25.4	26.7
Length n	$m = 6$		$m = 12$		$m = 18$	
	1,000	2,000	1,000	2,000	1,000	2,000
LB_{BMS}	6.2	5.7	5.7	4.1	3.6	4.7
BP_{BMS}	6.2	5.6	5.3	3.8	3.5	4.6
LB_{FRZ}	1.2	1.0	0.7	0.6	0.8	0.6
BP_{FRZ}	1.2	1.0	0.7	0.6	0.6	0.6
LB_S	13.4	15.2	9.6	10.7	7.7	9.6
BP_S	13.3	15.0	9.0	10.6	7.2	8.9

[2] 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.

[3] Francq, C. and Zakoïan, J-M. (1998) Estimating linear representations of nonlinear processes, *Journal of Statistical Planning and Inference* 68, 145–165.

[4] Lobato, I. N. (2001) Testing that a dependent process is uncorrelated, *Journal of the American Statistical Association* 96, 1066–1076.