Les erreurs-types dans les modèles non linéaires tels que les modèles de séries temporelles

Guy Mélard [1]

[1] *Université libre de Bruxelles, Belgique, gmelard@ulb.ac.be*

Résumé. L'estimation des paramètres de modèles statistiques non linéaires est généralement effectuée par une méthode de moindres carrés non linéaires ou de maximum de vraisemblance en employant de l'optimisation numérique. Les erreurs-types sont alors souvent déduites en employant des dérivées numériques. Une étude empirique dans le contexte de modélisation de séries temporelles au moyen de plusieurs logiciels statistiques révèle que ces erreurs-types ne sont pas très précises, correctes parfois à seulement 2 ou 3 chiffres significatifs. Une investigation complémentaire sur des modèles encore plus simples détermine la raison principale de ce manque de précision. On fournit plusieurs suggestions aux développeurs de logiciels statistiques dans le but d'améliorer leurs produits. Cette communication s'inscrit dans une suite de travaux de quelques auteurs qui mettent en gardent les utilisateurs trop confiants dans la pertinence de leurs résultats.

Mots-clés. Optimisation non linéaire, matrice d'information de Fisher, dérivées numériques, modélisation de séries temporelles.

Abstract. Estimation of the parameters in non-linear statistical models is usually performed by non-linear least-squares or maximum likelihood using numerical optimization. Then, standard errors are often derived by using numerical derivatives. An empirical study of time series modelling through several statistical packages reveals that these standard errors are not very accurate, sometimes with only 2 or 3 correct digits. A further investigation on much simpler models determines the main reason for such a lack of accuracy. Some suggestions are given to developers of statistical software in order to improve their products. The present communication is inline with a sequence of works from different authors who warn users that are too confident in the relevance of their results.

Keywords. Non-linear optimization, Fisher information matrix, numerical derivatives, time series modelling.

1 Introduction

Cette communication s'inscrit dans une suite de travaux de quelques auteurs qui mettent en gardent les utilisateurs trop confiants dans la pertinence de leurs résultats : McCullough (1998, 1999, 2004), McCullough et Renfro (2000), Yalta et Jenal (2009), et Hill et al. (2024), notamment.

Newbold et al. (1994) ont montré que différents logiciels conduisent, pour un modèle de séries temporelles spécifié, à des estimations et des erreurs-types différentes. Deux explications sont liées à la méthode d'estimation (moindres carrés conditionnels, pseudo-maximum de vraisemblance gaussien) et aux options spécifiques à l'optimisation. Les algorithmes utilisés pour les calculs sont aussi de première importance.

Nous avons d'abord voulu étendre cette vielle étude empirique à d'autres séries avec des modèles plus variés. On pourrait croire que ces différences ont disparu avec le temps mais ce n'est pas le cas, ou que des packages très employés comme R fournissent de meilleurs résultats mais c'est faux.

Notons que nous ne considérons ici que les erreurs-types déduites lors de l'estimation des paramètres d'un modèle. En supposant la normalité du processus générateur, il est aussi possible d'évaluer la matrice d'information de Fisher en tout point et, après inversion de la matrice en l'optimum, la matrice de variance-covariance des estimateurs des paramètres de modèles ARMA (et même VARMAX, c'est-à-dire ARMA multivariés et avec variables explicatives), voir Mélard et Klein (2023) et Klein et Mélard (2023). Nous ne considérons pas cette approche ici.

Partant d'une série $\{y_t, t = 1, ..., n\}$ de longueur n, et un modèle dépendant d'un vecteur de paramètres β, de vraie valeur inconnue β^0, on considère la log-vraisemblance $\ell_n(\beta)$ dont la maximisation conduit à un estimateur $\widehat{\beta}_n$. La matrice de variance-covariance de cet estimateur peut s'obtenir par une des deux espérances mathématiques suivantes

$$I(\beta^0) = \frac{1}{n} E_{\beta_0}\left(\frac{\partial \ell_n(\beta)}{\partial \beta^\top}\frac{\partial \ell_n(\beta)}{\partial \beta}\right) \quad \text{ou} \quad J(\beta^0) = -\frac{1}{n} E_{\beta_0}\left(\frac{\partial^2 \ell_n(\beta)}{\partial \beta^\top \partial \beta}\right), \tag{1}$$

où $\top$ indique la transposition. Mais β^0 est inconnu, les espérances sont pratiquement impossibles à évaluer et la forme analytique de $\ell_n(\beta)$ est trop complexe, de sorte qu'on est obligé d'employer une évaluation numérique en $\beta = \widehat{\beta}_n$. On n'a pas $I(\widehat{\beta}_n) = J(\widehat{\beta}_n)$ et il y a plusieurs manières d'évaluer le produit extérieur des gradients pour I ou la hessienne pour J.

2 Description de la partie expérimentale

Nous sommes partis d'un ensemble de séries temporelles déjà considérées et de modèles ARIMA proposés dans la littérature, voir Mélard (1985). Nous avons estimé les paramètres de ces modèles et déterminé les erreurs-types au moyen d'un certain nombre de logiciels, parmi lesquels les packages stats (fonction arima) et forecast (fonction Arima) de R, SPSS, SAS, Stata et quelques autres. Notons que tous ces programmes utilisent le fonction de vraisemblance gaussienne exacte, contrairement à Newbold et al. (1994), donc les résultats sont plus proches. Nous avons choisi comme référence un programme personnel basé sur Mélard (1984), écrit spécialement et compilé en quadruple précision donc traitant des nombres avec plus 26 décimales. Le nombre de décimales correctes est calculées par la formule proposée par McCullough (1998), $-\log_{10}(|q - c|/|c|)$, où le résultat obtenu q est comparé à la vraie valeur c, sauf si $c = 0$, où on prend $-\log_{10}(|q|)$

Nous avons alors examiné les nombres de chiffres corrects, synthétisés sur les différents paramètres, pour les estimations et pour les erreurs-types. Il n'est pas possible de dćrire ici les 27 séries de longueur n entre 60 et 369, les spécification ARIMA utilisées avec entre 1 et 5 paramètres, les logiciels comparés avec les options appropriées choisies, ni de donner les résultats détaillés. En moyenne sur les séries, on observe que le nombre de chiffres corrects des erreurs-types 1 (pour les fonctions arima et Arima, respectivement dans les packages stats et forecast de R, et pour Stata), 3 (SAS), 4 (SPSS) sont bien inférieurs à ceux pour les estimations 3 (SPSS), 4, (la fonction arima du package stats de R et SAS), 5 (la fonction Arima du package forecast de R), et 6 (Stata). Notons que, contrairement aux autres, Stata inclut la variance des erreurs σ^2 parmi les paramètres d'intérêt ce qui a nécessité des ajustements. Dans le tableau 1, ils sont indiqués par la mention "ajust. pour variance".

3 Approche computationnelle

Comme nous l'indiquons dans Mélard (2024), les auteurs de logiciels ne documentent pas suffisamment les algorithmes employés, par exemple la méthode de calcul de la fonction objectif (parfois, mais pas toujours, la quasi-vraisemblance gaussienne exacte), l'algorithme d'optimisation, les critères d'arrêt, etc. On peut supposer que l'obtention des erreurs-types se fait par recours à des différences divisées, qui entraînent une perte de précision substantielle. Commme la plupart des logiciels (sauf Stata), on emploie la vraisemblance concentrée par rapport à la variance σ^2 des erreurs, donc on est ramené á minimiser une somme de carrés $S(\beta)$. Le minimum est alors divisé par $\widehat{\sigma}_n^2$, une estimation non biaisée de σ^2.

Au lieu de S, on peut aussi se baser sur les résidus e_t dont S est la somme des carrés. Il n'est pas possible de détailler ici les différentes approximations de I et de J qui ont été traitées. Pour examiner plus précisément ces aspects, nous avons traité des modèles non linéaires, puis des modèles linéaires pour terminer par l'estimation de la moyenne, c'est-à-dire d'une constante c dans un modèle de série temporelle $y_t = c + e_t$. En effet, déjà dans ce cas-là, on va voir que les logiciels donnent des estimations peu correctes des erreurs-types et expliquer pourquoi. L'estimation optimale de c est $\bar{y}$, la moyenne des y_t, donc $S(\bar{y})$ est n fois la variance des y_t.

Nous avons expérimenté avec notre propre programme et développé plusieurs approches, présentées succinctement dans le tableau 1 et détaillées dans Mélard (2024), de manière à améliorer assez fortement la précision des résultats. A titre d'exemple, nous avons pris une des séries de longueur $n = 106$. Des expressions de $n\sigma^2$ fois l'information de Fisher I (basée sur le produit extérieur des gradients ou OPG) ou J (basée sur le hessien) sont proches de n. On devrait retrouver que l'erreur-type sur la moyenne vaut $1/\sqrt{n}$ fois l'écart-type estimé $\hat{\sigma}$. Ces approches sont comparées dans le tableau 1 qui montre qu'il est possible de passer de 2 à 9 chiffres corrects pour les erreurs-types. Nous avons ajouté les résultats obtenus par quelques logiciels déjà considérés au paragraphe 2. C'est évidemment avec cette variante dans notre programme en quadruple précision que les résultats du paragraphe 2 ont été obtenus. L'amélioration inspirée par Goldberg (1991) dont il est question dans le tableau 1 consiste simplement à exploiter l'identité $(a^2 - b^2) = (a + b)(a - b)$ qui se révèle très utile pour réduire les pertes de précision lors de différences entre deux valeurs de e_t^2 en des valeurs proches des

Nom : méthode	Valeur	Précision	Err.-type	# décimales
J_1: J avec 1^e dérivée des e_t	106,00000000	0	0,352742733	
I	106,00000010	10^{-7}	0,352742733	9,3
J_2 (ancien): J avec 2^{de} dérivée de S	110,61729310	4,6	0,345302319	1,7
J_2 (nouveau) : idem, (somme avec DOT_PRODUCT)	106,00042515	$4\ 10^{-4}$	0,352742026	5,7
J_2 (amélioré): 2^{de} dérivée des e_t^2, (somme avec SUM)	106,00000212	$2\ 10^{-6}$	0,352742729	8,0
J_2 (amélioré bis): idem, (somme avec TwoSum, d'Ogita et al. (2005))	106,00000410	$4\ 10^{-6}$	0,352742726	7,7
J_2 (final): idem avec Goldberg (1991) (somme avec SUM)	106,00000015	$2\ 10^{-7}$	0,352742733	9,2
R forecast/Arima			0,351074987	2,3
SAS			0,352742733	10,4
Stata OIM ajust. pour variance			0,35107491	2,3
Stata OPG ajust. pour variance			0,35107491	2,3

Tableau 1: Résultats de $n\widehat{\sigma}^2 I$ ou de $n\widehat{\sigma}^2 J$ pour l'estimation d'une constante sur une série avec $n = 106$. SUM et DOT_PRODUCT calculent respectivement une somme et un produit scalaire avec une précision étendue au sens de la norme IEEE 754

paramètres.

4 Conclusions

Il est assez surprenant que des erreurs-types soient aussi peu précises dans les logiciels commerciaux ou même dans les logiciels libres utilisés partout.

On a aussi remarqué que les packages de R n'ont pas pu traiter tous les modèles, et que pour SAS et SPSS les options lors de l'optimisation doivent impérativement être modifiées (ce qui a été fait ici) sous peine d'avoir des résultats moins corrects.

On peut raisonnablement penser que nos constatations s'étendent à d'autres modèles statistiques et d'autres logiciels où l'optimisation numérique est employée, tout au moins quand la matrice de variance-covariance n'est pas obtenue en employant des dérivées analytiques de la log-vraisemblance.

Bibliographie

Goldberg, D. (1991), What every computer scientist should know about floating-point arithmetic, *ACM Computing Surveys* **23**(1), 5-48.

Hill, C., Du, L., Johnson, M., and McCullough, B. D. (2024), Comparing programming languages for data analytics: Accuracy of estimation in Python and R, *WIREs Data Mining*

Knowl Discov. **2024**, e1531.

Klein, A., and Mélard, G. (2023b), An algorithm for the Fisher information matrix of a VARMAX process, *Algorithms* **16**, 364. https://doi.org/10.3390/a16080364. 14 pp.

McCullough B. D (1998), Assessing the reliability of statistical software: part I, *The American Statistician* **52**, 358-366.

McCullough B. D. (1999), Assessing the reliability of statistical software: part II, *The American Statistician* **53**, 149-159.

McCullough B. D. (2004), Some details of nonlinear estimation, in Micah Altman, Jeff Gill and Michael P. McDonald (ed.) *Numerical Issues in Statistical Computing for the Social Scientist*, Chapter 8. Wiley, New York, pp 199-218.

McCullough B. D., and Renfro C. R. (2000), Some numerical aspects of nonlinear estimation, *J Economic and Social Measurement* **26**, 63-77.

Mélard, G. (1984), Algorithm AS197: A fast algorithm for the exact likelihood of autoregressive-moving average models. *Journal of the Royal Statistical Society Series C, Applied Statistics* **33**, 104-114.

Mélard, G. (1985), Exact derivatives of the likelihood of ARMA processes, *1985 Proceedings of the Statistical Computing Section*, American Statistical Association, Washington D.C., pp. 187-192.

Mélard, G. (2024), Standard errors in time series and non-linear models, submitted.

Mélard, G. et Klein, A. (2023), Sur les algorithmes pour l'information de Fisher de modèles vectoriels dynamiques, 54es Journées de Statistique de la SFdS, JdS2023, Bruxelles, 3-6 juillet 2023.

Newbold, P., Agiakloglou, C., and Miller, J. (1994), Adventures with ARIMA software. *International Journal of Forecasting* **10**, 573-581.

Ogita, T., Rump, S. M., and Oishi, S. (2005), Accurate sum and dot product, *SIAM J. Scientific Computing* **26**, 1955-1988.

Yalta, A. T., and Jenal O. (2009), On the importance of verifying forecasting results, *International Journal of Forecasting* **25**, 62-73.