A note on lag selection in time series using multi-step estimation

Lars Forsberg*
Uppsala University, Uppsala, Sweden - lars.forsberg@statistics.uu.se

Johan Lyhagen
Uppsala University, Uppsala, Sweden - johan.lyhagen@statistics.uu.se

Abstract

In applied time series analysis there is a great need to decide upon the number of lags. This is due to the fact that the theories at hand most often do not supply information regarding the number of lags, so we need procedures to let the data decide for us. Commonly used method includes information criteria, such as AIC and BIC, the bottom up and top down methods (start with a minimum/maximum number of lags and then test up/down) and testing for autocorrelation in the residuals. In this note we propose a new approach which is based on comparing the parameter estimates gained from minimizing one step ahead forecast error with the ones from minimizing the $h-$ step ahead forecast errors. This evolves to a non-nested type of test. The proposed test statistic has an asymptotic χ^2-distribution. A small Monte Carlo Simulation is performed which shows that the size is close to the nominal and that it has good power. Compared to AIC, BIC, top-down and the Breusch-Godfrey test for autocorrelation in the residuals our test is the only one with both good size and power properties.

Keywords: AR(1); Size and power; Non-nested test.

1. Introduction One of the major challanges in statistics is how to chose amongst competing models. In a perfect world, the theory behind the application would completely describe the model. In reality, most theories are only partial in the sense that they only make statements about some aspects. The remaining aspects need to be decided by some data driven procedure. As an example, consider the Purchasing Power Parity, PPP, theory from economics which was introduced in its modern form by Cassel (1918). PPP states that the price level in one country should be the same as in another country, when adjusted for exchange rate. This is something that is expected to hold in the long run, but there are many reasons for short term deviations from this relationship.

One way to model this is by using a Vector Autoregressive Model, VECM, on the price levels of the two countries and the corresponding exchange rates. PPP then states that there is cointegration between the three variables (i.e. an long term equilibrium) but the theory is silent about the short run dynamics. In the VECM the short run dynamics are modeled as a linear function of the previous values of the variables, but how should we chose the number of lags? One way is to use information criteria such as AIC (Akaike, 1973) or the Bayesian Information Criteria (Schwarz, 1978). Another approach is to use the top down approach, i.e. select a large value, say p_{max} and then test if the largest lag is significantly different from zero, if not, reduce the lag order by one and test again, and continue until you reject the null, and that would be the model. The bottom up is to start with one lag and test if it is significant, if so, increase the lag order by one and proceed with testing if that lag is significant, and so on.

Yet another way would be to test if there is autocorrelation in the residuals from the model, start with a small number of lags and increase the lag order until the residuals are white noise. All of these approaches are used in applied sciences and are thoroughly studied in the statistics and econometrics literature.

An another idea, dating back at least to Haavelmo (1944), is the idea of multi-step estimation and forecasting, and this will be the basis for our test. Let's say that we have quarterly data, and we want to forecast one year out, i.e. four quarters. In a world that is linear, and we have a model that is correctly specified, it would be optimal to estimate the parameters of the model by minimizing the one step ahead forecast error, and then use those estimates to forecast four steps ahead. However, in a world that is not linear and/or the model is not correctly specified, it may be better to estimate the parameters of the model by minimizing the four step ahead forecast error, see e.g. Granger (1969). The idea of multi-step estimation has been investigated mainly in the forecasting literature, see e.g. Cox (1961) and Johnston (1974). Our test is based on the difference between the parameter estimates from the two ways of estimating the parameters.

The formal tests outlined above are nested. A non-nested test is when the parameter space in one case is not a special case of the other, see e.g. Pesaran and Weeks (1999) for an overview. The purpose of this paper is to propose a non-nested procedure for estimating the maximum lags in a time series model, and exemplify with the AR(p) class of models. The model in question is a model minimizing one step ahead forecast error and comparing it to a $h-$step ahead.

A small Monte Carlo simulation shows that the proposed method is superior compared to the standard methods.

The paper is organized as follows. The next section introduces the model and the proposed test procedure. Section 3 analyze the performance of the proposed procedure by Monte Carlo methods Section 4 concludes.

2. The model and test procedure

Let $\hat{y}_{t,h}$ be the estimated value of y at time $t+h$ conditional on the information set at time t. The standard least squares estimator is given by

$$\hat{\boldsymbol{\phi}}_1 = \arg\min \sum_{i=1}^{T} \left(y_{t+1} - \hat{y}_{t,1}\right)^2, \tag{1}$$

where $\boldsymbol{\phi}_1$ is a vector of parameters and where $\hat{y}_{t,1}$ depends on the model, e.g. for an AR(1) $\hat{y}_{t,1} = \hat{\phi} y_t$. In general minimizing the $h-$step ahead forecast error gives the estimator

$$\hat{\boldsymbol{\phi}}_h = \arg\min \sum_{i=1}^{T} \left(y_{t+h} - \hat{y}_{t,h}\right)^2, \tag{2}$$

where $\hat{y}_{t,h}$ is the same as above. Now it is easily seen that the parameter space is not nested. For example consider the AR(2):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \tag{3}$$

then $\hat{y}_{t,1}$ is a restricted AR(2) with the restriction $\phi_2 = 0$. Note also that $\hat{y}_{t,2}$ is also nested within the AR(2) but with restriction $\phi_1 = 0$. Hence the parameter spaces of $\hat{y}_{t,1}$ and $\hat{y}_{t,2}$ are not nested. The relationship between the two set of parameters can be found in e.g. Wei (p.97, 2006) for the general AR(p). In this case the relationship is $\phi_1^2 = \phi_2$, so if our model is correctly specified then the relationship should hold. This is what we test for. In this case the null hypothesis would be one lag, and if that would be true $\phi_1^2 = \phi_2$. If the true lag length is more than one this relationship does not hold.

Before presenting our proposed test statistic, in general we define $\hat{\boldsymbol{\phi}}_1^h$ as the implied value of $\hat{\boldsymbol{\phi}}_h$ in the parametrization of minimizing the $h-$step ahead forecast error. Then the Wald type test is

$$W = \left(\hat{\boldsymbol{\phi}}_h - \hat{\boldsymbol{\phi}}_1^h\right)' \left(\hat{V}\left(\hat{\boldsymbol{\phi}}_h - \hat{\boldsymbol{\phi}}_1^h\right)\right)^{-1} \left(\hat{\boldsymbol{\phi}}_h - \hat{\boldsymbol{\phi}}_1^h\right) \tag{4}$$

Next, we need an expression for the variance. Fortunately, it can be shown that it has a Hausman-Wu type of structure, see e.g. Hausman (1978) and Wu (1973),

$$\hat{V}\left(\hat{\boldsymbol{\phi}}_h - \hat{\boldsymbol{\phi}}_1^h\right) = \hat{V}\left(\hat{\boldsymbol{\phi}}_h\right) - \hat{V}\left(\hat{\boldsymbol{\phi}}_1^h\right). \tag{5}$$

Further, $\hat{V}\left(\hat{\boldsymbol{\phi}}_1^h\right)$ can be derived from $\hat{V}\left(\hat{\boldsymbol{\phi}}_1\right)$ using the delta method. The asymptotic distribution of W is χ_p^2 where p here denotes the number of parameters. Note that we should exclude the constant when calculating W.

3. Monte Carlo simulations and some results

In this section we will focus only on the AR(1) model for ease of exposition, but the principle is more general than that and applies to other time series models, e.g. ARIMA, VAR etc. The first AR(1) minimize the one step ahead forecast error

$$y_{t+1} = \phi_1 y_t + \varepsilon_{t+1}. \tag{6}$$

The second model is also an AR(1) but minimize the two step ahead forecast error

$$y_{t+2} = \phi_2 y_t + \varepsilon_{t+1}, \tag{7}$$

where we have stated them such that the information set is at time t. The relationship between ϕ_1 and ϕ_2, see e.g. Wei (p. 97, 2006), is

$$\phi_2 = \phi_1^2. \tag{8}$$

Equation (8) is our null which is valid under our hypothesis that one lag is sufficient to model the dynamic structure of the data. In this simple example we can test the hypothesis using a $t-$test:

$$t = \frac{\hat{\beta}_1 - \hat{\phi}_1^2}{\sqrt{\hat{V}\left(\hat{\beta}_1\right) - \hat{V}\left(\hat{\phi}_1^2\right)}} \tag{9}$$

To analyze the small sample properties we generate data according to (6) with $\phi_1 = 0.5$ and ε distributed as $N(0,1)$. When analyzing the power an AR(2) is generated with $\phi_2 = 0.25$. The sample sizes considered are $T = 25, 50, 100, 150, 200, 250, 300, 350, 400$ and the number of replicates is 10,000. We compare our proposal with AIC, BIC, a top down procedure (with $p = 2$ as maximum number of lags), and the Breusch-Godfrey test on autocorrelation of the residuals from an AR(1). As AIC and BIC are not formal tests and that the Top-Down procedure and the Breusch-Godfrey has the opposite hypothesizes we only present rejection rates and do not call them power. Further, for the same reason we do not present size adjusted power. Although strictly wrong, for simplicity we denote rejection frequencies for the AR(1) as level and for the AR(2) as power. The nominal size, if applicable, is set to 5%.

In Figure 1 the results of the Monte Carlo simulation is displayed. For sample sizes above 50 it seems like that Our test, the Breusch-Godfrey and the top down procedure has rejection rates close to the nominal size. AIC performs bad, too often choosing a too large model. BIC is much better choosing the correct model, the AR(1), most of the time. But the drawback is that it has the lowest probability of selecting the AR(2) when that one is true. AIC and our test has the highest probability of choosing the AR(2) while Breusch-Godfrey and the top down procedure has lower power. Comparing Our test with AIC we note that the reason for AIC to have high power is that the size is very high.

4. Conclusions

In this paper we have proposed an alternative procedure to decide the maximum number of lags in a time series model. The idea is based on the comparison of the minimization of one step ahead forecast error with $h-$ step ahead leading to a non-nested test. The resulting asymptotic distribution is standard χ^2 but it can be written in an standard $F-$form, a route not pursued here as such test statistic is most often sensitive to departures from normality. A limited Monte Carlo simulation is performed to analyse the small sample properties and compare it to some commonly used alternatives. The result is that Our Test has good size and power properties and the only one with both.

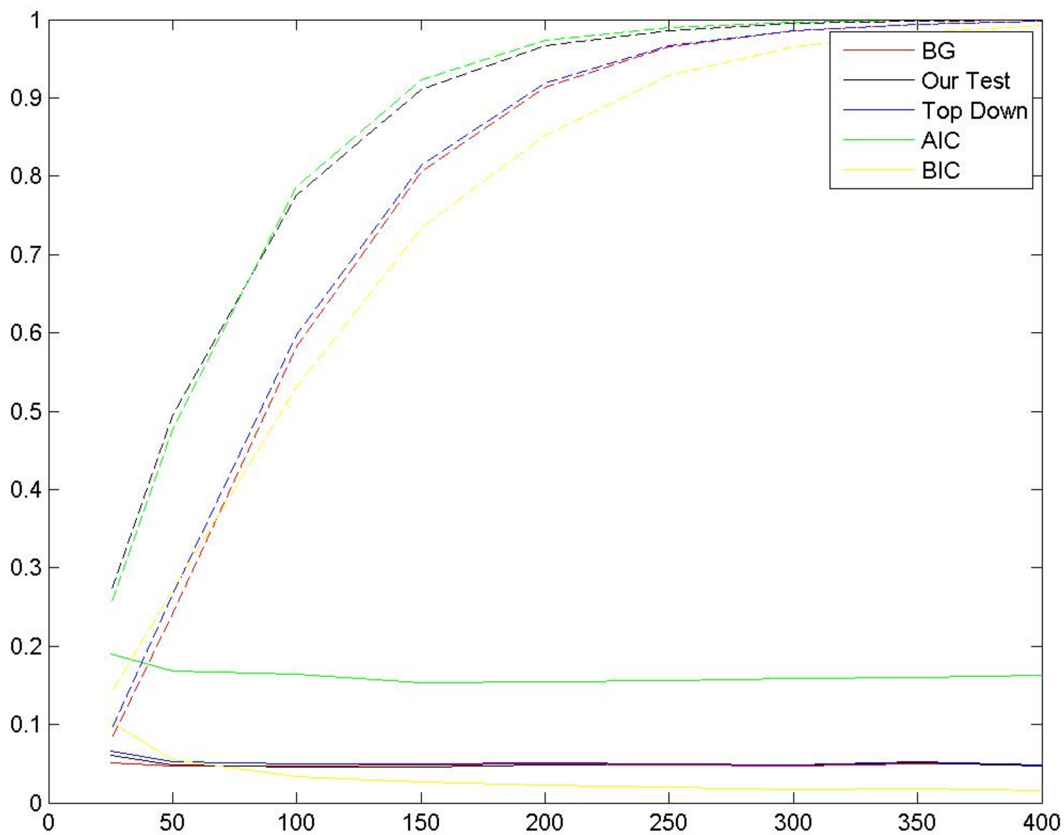

Figure 1: Rejection rates. The horizontal lines corresponds to data generated by an AR(1) while the upward sloping to an AR(2).

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In Proc. 2nd Int. Symp. Information Theory (eds B. N. Petrov and F. Cslki), pp. 267-281. Budapest: Akadkmiai Kiado.

Cassel, G. (1918) Abnormal Deviations in International Exchanges, The Economic Journal 28(112): 413-415.

Cox, D.R. (1961) Prediction by Exponentially Weighted Moving Averages and Related Methods, Journal of the Royal Statistical Society. Series B (Methodological) 23(2): 414-422.

Granger, C.W.J. (1969) Prediction with a Generalized Cost of Error Function, Operational Research Quarterly 20(2): 199-207.

Haavelmo, T. (1944) The Probability Approach in Econometrics, Econometrica 12(Supplement):iii-vi+1-115.

Hausman, J.A. (1978) Specification Tests in Econometrics, Econometrica 46(6): 1251-1271.

Johnston, H.N. (1974) UniversityA Note on the Estimation and Prediction Inefficiency of "Dynamic" Estimators, International Economic Review 15(1): 251-255.

Pesaran, M.H. and Weeks, M. (1999) Non-nested Hypothesis Testing: An Overview, Cambridge Working Papers in Economics 9918, Faculty of Economics, University of Cambridge.

Schwarz, G.E. (1978) Estimating the dimension of a model, Annals of Statistics 6(2): 461-464.

Wei, W.W.S. (2006) Time Series Analysis: univariate and multivariate methods, 2nd ed., Pearson.

Wu, D.-M. (1973) Alternative Tests of Independence between Stochastic Regressors and Disturbances, Econometrica 41(4): 733-750.