E C O N O M I C S   B U L L E T I N

Searching for the DGP when forecasting - Is it always meaningful for small samples?

Jonas Andersson
Norwegian School of Economics and Business Administration

Abstract

In this paper the problem of choosing a univariate forecasting model for small samples is investigated. It is shown that, a model with few parameters, frequently, is better than a model which coincides with the data generating process (DGP) (with estimated parameter values). The exponential smoothing algorithms are, once more, shown to perform remarkably well for some types of data generating processes, in particular for short-term forecasts. All this is shown by means of Monte Carlo simulations and a time series of realized volatility from the CAC40 index. The results speaks in favour of a negative answer to the question posed in the title of this paper.

The author thanks Jostein Lillestøl and an anonymous refereefor valuable comments.
Citation: Andersson, Jonas, (2006) "Searching for the DGP when forecasting - Is it always meaningful for small samples?." *Economics Bulletin,* Vol. 3, No. 28 pp. 1-9
Submitted: September 20, 2006. **Accepted:** November 6, 2006.
URL: http://economicsbulletin.vanderbilt.edu/2006/volume3/EB-06C20070A.pdf

Searching for the DGP when forecasting - Is it always meaningful for small samples?

Jonas Andersson

November 6, 2006

Abstract

In this paper the problem of choosing a univariate forecasting model for small samples is investigated. It is shown that a model with few parameters frequently is better than a model which coincides with the data generating process (DGP) with estimated parameter values. The exponential smoothing algorithms are again shown to perform remarkably well for some types of data generating processes, in particular for short-term forecasts. All this is shown by means of Monte Carlo simulations and a time series of realized volatility from the CAC40 index. The results speaks in favour of a negative answer to the question posed in the title of this paper.

1 Introduction

In Kim (2003), the author investigates the fact that bias of estimators of autoregressive parameters in small samples impair the forecast performance of such models. A related problem is that it is difficult to find an appropriate forecasting model when the sample size is small. This is the topic of the present paper. In Section 2 the problem is presented in general and then narrowed down for illustration purposes to a specific data generating process (DGP). Section 3 presents the forecast methods investigated, Section 4 the simulation study, Section 5 an example with realized volatility data from the CAC40 index and Section 6 summarizes the results.

2 Forecasts based on small samples

Imagine a realization of a time series $\{x_t\}_{t=1}^T$ that is generated by the data generating process P. Furthermore, assume that the model M is used to forecast the value of x_{t+k}. It is well known, and intuitively reasonable, that the best forecast of x_{t+k}, made at time t, in the sense of minimizing mean squared forecast errors, is given when M is chosen to be the expectation of x_{t+k} conditional on the realization $\{x_t\}_{t=1}^T$ under P. In the context of a small sample it is worth mentioning that the parameters in the model M have to be estimated. This means that even in a situation when we have correctly identified the data generating process, something not likely to be done in real world applications, the success of the forecasts produced hinges on the success in estimating the parameters of the model. This is the topic of Kim (2003) who studies the problem in autoregressive models. In small samples it might be difficult to identify P based on data, even if P is one of the models investigated in the model search. This problem is the focus of interest in this paper. I will, for the most part, follow Kim (2003) and study autoregressive DGP's. One simulation experiment will, however, be done with an MA-process.

3 Four forecasting methods

In the simulation study in next section, four different strategies for forecasting will be used. P will always be an $AR(6)$ model and M will be chosen in the following four different ways.

- Forecasting with knowledge of the DGP. As a benchmark, it will be assumed that the DGP is known except for its parameters which will be estimated. The estimation of the parameters is thus the only disturbance to an otherwise obvious best method.
- Using an AR(1)-model. The purpose of using this forecasting method is to investigate if the benefit of this naive reduction of number of parameters will outweigh the bias induced by the model misspecification.
- Determining lag length of an AR-model with the Akaike information criterion (AIC) (Akaike, 1974). This is an attempt of identifying the DGP and then use it for forecasting. While intuitively the most appealing method to this author, the small sample properties of the forecasts are not clear. On the one hand, the DGP might not be identified for small samples. If the DGP is identified, on the other hand, there are many parameters to estimate, inducing forecast inaccuracy.
- Exponential smoothing algorithms. These methods are used as representatives for "brute force" forecasts. Both the simple exponential smoothing algorithm and Holt's linear exponential smoothing algorithm will be investigated. The latter method (Holt, 1957) is included mainly to see what harm the uncertainty the trend estimation is doing to the precision of the forecasts in situations when no trend is present. The smoothing parameters will be estimated by minimizing the in-sample mean squared errors.

4 Simulation study

The data generating processes that are simulated are $AR(6)$ processes. For all models, one-step-ahead and five-step-ahead forecasts are evaluated. All comparisons are made out-of-sample, i.e. for each replicate, T_{obs} plus one (five) observations are generated but only the first T_{obs} are used to estimate parameters. The number of replicates is 1000 for each model and sample size and the maximum number of lags used in the AIC criterion based forecasts is 10. The measure of comparison is the mean squared error of forecasts (MSEF). All simulations are made using the R programming language (R Development Core Team, 2005) and the R-package fSeries (Wuertz et al., 2006).

4.1 The stationary autoregressive case

I start by considering a case where the DGP is a weakly stationary $AR(6)$ process, i.e. when all roots to the AR-polynomial lies outside the unit circle. First, a case which is clearly stationary is studied, then a case where one of the roots to the AR-polynomial is close to the unit circle.

4.1.1 Clearly stationary case

The first case has the AR-parameters $\phi = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1)$. The modulus to the roots to the corresponding AR-polynomial are $(1.476, 1.578, 1.578, 1.148, 1.476, 1.607)$, all well outside the unit circle. As can be seen from Table 1 using the DGP with estimated parameters actually give the next largest MSEF of all models for the smallest sample, 20 observations. Only Holt's linear exponentional smoothing algorithm is worse. The simple exponential smoothing algorithm outperforms this "true" model also for the sample sizes 50. For the studied sample sizes, the AIC criterion underestimate the lag length of the DGP. Eventually, of course, for large samples the "true" model will outperform all others.

The five-step-ahead forecasts are then investigated for the same DGP. The results here are more in favour of the DGP with estimated parameters or its more realistic version for applications, the AR process with lag length determined by the AIC criterion. A notable difference with the one-step-ahead case is that the exponential smoothing algorithms perform worse than for the one-step-ahead forecasts.

4.1.2 Close to unit root case

The AR-coefficients in the process that will now be studied are $\phi = (0.2, 0.2, 0.1, 0, 0.1, 0.35)'$ with roots to the corresponding AR-polynomial $(1.213, 1.228, 1.228, 1.014, 1.213, 1.270)$. Table 3 shows that simple exponential smoothing outperforms the DGP with estimated parameters and the AR-AIC forecasts up to sample size 50.

4.2 The non-stationary autoregressive case

In this subsection I will study an example of a unit root process. When forecasting a non-stationary process of the unit root type, the Holt's linear algorithm turns out to out-perform the DGP with estimated parameters even for the sample size 100. This is so even though less information have been used than for the other forecasts in the sense that a decision whether to difference the data or not was not needed to be made. This results, however, does not persist for the five-step-ahead forecasts.

4.3 The moving average case

Also for the MA(1) process studied here, for sample sizes up to 100, the results indicate that the DGP with estimated parameters is not the best way of producing good forecasts. The somewhat surprising observation here is that the best forecasts are made by using an AR model with lag-length determined by the AIC criterion. This is true both for the one-step-ahead and five-step-ahead forecasts.

5 A real world data example

At first it seems difficult to investigate the results above on real world data. After all one cannot possible know the exact structure of the DGP that generated these data. What I will try to do is to convince the reader that the DGP of a reasonably long observed time series is well approximated by a particular process within the ARIMA-class of processes. Thereafter I will select a short subset of this dataset, estimate the parameters with this subset, and investigate the ability of the different forecasting methods to forecast observations outside of it. The data I have chosen for this exercise is a time series that comes as example data "cac40vol" in the package fSeries (Wuertz et al., 2006) to the R software (R Development Core Team, 2005). The series represents the realized volatility of the CAC40 index for 1249 days between 1995 to 1999. After inspection of the sample ACF and PACF of the data an AR(8) models was first hypothesized. This model passed the Ljung-Box test for the residuals up to 100 lags on the 5% level. After reducing the number of AR-parameters to 5 it turned out that the residuals still passed this test. This was not the case for an AR(4). Therefore, an AR(5) model was chosen to serve as the "approximated DGP" for these data. Alternatively, the AR(5) forecasts can, of course, be seen just as one of the forecasting methods that are compared.

The sample size was chosen to 100 as well as the number of forecasts to base the MSEF's on. The following procedure was used. First observations 51-150 was used to estimate the parameters of the different models. Thereafter, observation number 151 was forecasted and subtracted from the actual outcome, forming the first forecast error. Now observations 52-151 was used to estimate parameters and the 152'nd observation was forecasted. This procedure was repeated 100 times.

These 100 forecast errors was the basis for the MSEF's. This procedure was then in turn repeated for the subsample 151-250, and so on until the end of the sample.

The results clearly shows that the "approximate DPG" with estimated parameters is not the best way to forecast these series, at least not according to the MSEF criterion. In fact, the best forecasting method, would have been the simple exponential smoothing algorithm. In all but two of the 12 periods its performance is the best of the different forecast methods. In only one of the 12 periods its performance is worse than the "approximate DGP" with estimated parameters.

6 Conclusions

Forecasting with a model not identical to the data generating process turns out to be better in some cases for autoregressive models in the sense of minimizing mean squared forecast errors. These cases are when the number of parameters to estimate in the the DGP is large relative to the sample size. The intuition behind this is that even though a smaller model does not reflect the time dynamics of the process as well as the model identical to the DGP, the lower number of estimated parameters makes the precision of the forecasts higher.

Another result is that the more persistent the dynamics of the process, the more important it seems to be to capture the DGP correctly. Among the realistic methods, i.e. methods feasible to use in a real world situation, the AIC-determined forecasts are the best in this case. As in many cases, exponential smoothing algorithms turns out to be reliable also for the investigated autoregressive processes, in particular for small samples.

Furthermore, for $I(1)$ processes, Holt's linear exponential seems surprisingly able to capture the local trends occurring in such processes. Here however, the results is slightly more favourable to trying to find the DGP before forecasting. For MA-processes, again, knowledge of the DGP is no guarantee for good forecast performance. Using an AR-model where the lag-length is determined by the AIC criterion outperform the DGP with estimated parameters. The exponential smoothing algorithms are less useful here.

The empirical illustration, where the realized volatility of the CAC40 index is studied, strengthen the simulation findings. The forecast performance of the model identified by using the entire sample is by no means better than the other models. The simple exponential smoothing algorithm seems to be an good choice to forecast this series.

References

Akaike, H. (1974): "A new look at statistical model identification," *Transactions on Automatic Control*, 19, 716–723.

Holt, C. (1957): "Forecasting seasonal and trends by exponentially weighted moving averages," Memorandum 52, Office of Naval Research.

Kim, J. H. (2003): "Forecasting autoregressive time series with bias-corrected parameter estimators," *International Journal of Forecasting*, 19, 493–502.

R Development Core Team (2005): *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Wuertz, D., many others, and see the SOURCE file (2006): *fSeries: Rmetrics - The Dynamical Process Behind Markets*, r package version 221.10065.

Table 1: Mean squared errors of one-step-ahead forecasts. T_{obs} is the number of observations, p_{AIC} the average lag-length chosen by the AIC criterion, $AR(6)$ is the MSE's of when the DGP, with estimated parameters, is used for forecasting, $AR(1)$ when an AR(1) model is used, $AR(p_{AIC})$ when the AIC criterion is used and ES when and Holt when exponential smoothing is used. $\phi = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1)'$

T_{obs}	p_{AIC}	$AR(6)$	$AR(1)$	$AR(p_{AIC})$	ES	Holt
20	0.711	1.237	1.142	1.155	1.128	1.467
50	1.719	1.088	1.139	1.106	1.083	1.378
100	3.179	1.037	1.065	1.064	1.057	1.300
500	6.169	0.975	1.028	0.975	0.987	1.108

Table 2: Mean squared errors of five-step-ahead forecasts. T_{obs} is the number of observations, p_{AIC} the average lag-length chosen by the AIC criterion, $AR(6)$ is the MSE's of when the DGP, with estimated parameters, is used for forecasting, $AR(1)$ when an AR(1) model is used, $AR(p_{AIC})$ when the AIC criterion is used and ES and Holt when exponential smoothing is used. $\phi = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1)'$

T_{obs}	p_{AIC}	$AR(6)$	$AR(1)$	$AR(p_{AIC})$	ES	Holt
20	0.692	1.130	1.116	1.120	1.211	3.891
50	1.735	1.049	1.108	1.083	1.100	2.604
100	3.139	1.047	1.095	1.063	1.100	1.972
500	6.172	1.036	1.115	1.041	1.122	1.503

Table 3: Mean squared errors of one-step-ahead forecasts. T_{obs} is the number of observations, p_{AIC} the average lag-length chosen by the AIC criterion, $AR(6)$ is the MSE's of when the DGP, with estimated parameters, is used for forecasting, $AR(1)$ when an AR(1) model is used, $AR(p_{AIC})$ when the AIC criterion is used and ES and Holt when exponential smoothing is used. $\phi = (0.2, 0.2, 0.1, 0, 0.1, 0.35)'$

T_{obs}	p_{AIC}	$AR(6)$	$AR(1)$	$AR(p_{AIC})$	ES	Holt
20	1.394	1.297	1.412	1.316	1.185	1.647
50	3.594	1.176	1.491	1.298	1.116	1.386
100	5.937	1.011	1.539	1.033	1.084	1.240
500	6.504	0.918	1.481	0.925	1.028	1.162

Table 4: Mean squared errors of five-step-ahead forecasts. T_{obs} is the number of observations, p_{AIC} the average lag-length chosen by the AIC criterion, $AR(6)$ is the MSE's of when the DGP, with estimated parameters, is used for forecasting, $AR(1)$ when an AR(1) model is used, $AR(p_{AIC})$ when the AIC criterion is used and ES and Holt when exponential smoothing is used. $\phi = (0.2, 0.2, 0.1, 0, 0.1, 0.35)'$

T_{obs}	p_{AIC}	$AR(6)$	$AR(1)$	$AR(p_{AIC})$	ES	Holt
20	1.360	1.496	2.172	1.870	1.244	3.801
50	3.658	1.243	2.331	1.491	1.188	2.489
100	5.938	1.166	2.322	1.214	1.224	2.153
500	6.472	1.123	2.407	1.123	1.196	1.452

Table 5: Mean squared errors of one-step-ahead forecasts. T_{obs} is the number of observations, p_{AIC} the average lag-length chosen by the AIC criterion, $AR(6)$ is the MSE's of when the DGP, with estimated parameters, is used for forecasting, $AR(1)$ when an ARIMA(1,1,0) model is used, $AR(p_{AIC})$ when the AIC criterion is used and ES and Holt when exponential smoothing is used. $\phi = (0.1, 0.1, 0.1, 0, 0.1, 0.1)'$

T_{obs}	p_{AIC}	$AR(6)$	$AR(1)$	$AR(p_{AIC})$	ES	Holt
20	0.680	1.525	1.113	1.124	1.166	1.246
50	1.608	1.175	1.143	1.147	1.206	1.132
100	3.174	1.065	1.106	1.064	1.146	1.052
500	6.219	0.939	1.011	0.945	1.030	0.994

Table 6: Mean squared errors of five-step-ahead forecasts. T_{obs} is the number of observations, p_{AIC} the average lag-length chosen by the AIC criterion, $AR(6)$ is the MSE's of when the DGP, with estimated parameters, is used for forecasting, $AR(1)$ when an ARIMA(1,1,0) model is used, $AR(p_{AIC})$ when the AIC criterion is used and ES and Holt when exponential smoothing is used. $\phi = (0.1, 0.1, 0.1, 0, 0.1, 0.1)'$

T_{obs}	p_{AIC}	$AR(6)$	$AR(1)$	$AR(p_{AIC})$	ES	Holt
20	0.660	9.683	8.903	9.054	10.110	10.381
50	1.787	8.656	9.720	9.177	10.522	9.934
100	3.149	7.966	9.597	8.656	10.139	9.534
500	6.089	7.760	10.111	7.760	10.653	8.852

Table 7: Mean squared errors of one-step-ahead forecasts. T_{obs} is the number of observations, p_{AIC} the average lag-length chosen by the AIC criterion, $MA(6)$ is the MSE's of when the DGP, with estimated parameters, is used for forecasting, ma1 when an MA(1) model is used, $AR(p_{AIC})$ when the AIC criterion is used and ES and Holt when exponential smoothing is used. $\theta = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1)'$

T_{obs}	p_{AIC}	ma6	ma1	$AR(p_{AIC})$	ES	Holt
20	0.555	1.413	1.023	1.008	1.073	1.572
50	1.173	1.271	1.042	1.074	1.113	1.406
100	2.025	1.114	1.077	1.079	1.097	1.352
500	5.502	1.070	1.076	1.074	1.122	1.258

Table 8: Mean squared errors of five-step-ahead forecasts. T_{obs} is the number of observations, p_{AIC} the average lag-length chosen by the AIC criterion, $AR(6)$ is the MSE's of when the DGP, with estimated parameters, is used for forecasting, $MA(1)$ when an MA(1) model is used, $AR(p_{AIC})$ when the AIC criterion is used and ES and Holt when exponential smoothing is used. $\theta = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1)'$

T_{obs}	p_{AIC}	ma6	ma1	$AR(p_{AIC})$	ES	Holt
20	0.511	1.107	1.008	0.996	1.241	4.388
50	1.047	1.139	1.100	1.099	1.255	2.928
100	1.916	1.082	1.081	1.076	1.130	2.198
500	5.475	1.056	1.071	1.052	1.099	1.583

Table 9: Out-of-sample one-step-ahead MSEF's for the different forecasts of the realized volatility of the CAC40 index. The sample is divided in 12 parts of 100 observations each.

n_{start}	n_{end}	n_f	p_{AIC}	$AR(5)$	$AR(1)$	$AR(p_{AIC})$	ES	Holt
150	249	100	5.060	1.287	1.799	1.194	1.153	1.318
250	349	100	4.930	0.298	0.381	0.339	0.270	0.289
350	449	100	5.780	0.688	0.821	0.574	0.557	0.630
450	549	100	5.490	1.533	1.925	1.505	1.390	2.091
550	649	100	5.040	7.221	9.968	7.728	7.283	8.900
650	749	100	3.160	284.467	185.375	131.361	214.509	322.965
750	849	100	2.280	3.770	4.210	4.044	3.547	3.584
850	949	100	4.020	29.300	30.586	28.261	25.644	27.999
950	1049	100	3.680	4.302	5.740	4.638	4.003	4.698
1050	1149	100	5.320	0.874	1.293	0.929	0.783	0.999
1150	1249	100	4.730	0.490	0.676	0.430	0.404	0.560