FORECASTING WITH A TIME SERIES MODEL

by

Peter Welch

CENTRAL BANK OF BARBADOS
RESEARCH DEPARTMENT

DECEMBER 1985

Forecasting With A Time Series Model

This paper presents an "Autoregressive Integrated Moving Average" model (Arima) as a univariate forecasting model for the net foreign position of the commercial banks. These models have been pioneered in the time series literature by Box and Jenkins (1976).

The problem considered in this paper is that of finding the parameters (b_i, c_j), where i and j range from zero to p and q respectively, of an arima (p, d, q) process and using that to estimate a forecast function of the form.

$$(1)\ NFP_t\ (1) = 2NFP_t - NFP_{(t-1)} - 0.98a_t + 0.27a_{(t-6)}$$

where $c_1 = 0.98$ and $c_6 = 0.27$ are the moving average parameters. NFP_t is the net foreign position of the commercial banks at time t and a_t is the shock which drives the mechanism.

Identification

The first and second differences of the variable to be forecast were inspected for stationarity. The second difference looked like a good candidate. The second difference tailed out visually to zero after a few lags. Moreover, there was a cut off in the autocorrelation function after lag 6. For statistical confirmation of this hypothesis, a statistical test devised by Bartlett (1946) was used.

Table of the Autocorrelation Function of the Second Difference of NFP at Six Lags

	Value of the Autocorrelation at lag K	Square of the Autocorrelation at lag k
1.	-0.5807	0.3372
2.	0.0682	0.6047
3.	-0.0140	0.0017
4.	0.1930	0.0372
5.	-0.3086	0.0952
6.	0.2646	0.0700

The value of the sum of squared autocorrelations is 0.5460

If we denote the estimated autocorrelation function at lag k by r_k, then Bartlett's approximation gives

$$\text{Var}\,[r_k] = \frac{1}{54} \quad (1 + 24) \times (0.5460)$$

$$= 0.0387$$

This approximation holds for all lags greater than k = 6

The corresponding standard error

$$= 0.5285$$

The 'T' statistics have a maximum value of 0.5285 after lag 6. This invalidates the significance of any autocorrelation above six other than zero.

A visual inspection of the partial autocorrelation function suggested a cut off after lag three. The statistical test by Quenouille (1949) was used to make this observation rigorous. If we let O_{kk} be the estimated partial autocorrelation function at lag k then var $[O_{kk}]$ is approximately 1/n for lags greater than or equal to p + 1 = 4, where p is the cut off. Thus the standard error (SE) of the estimated partial autocorrelation O_{kk} is approximately the reciprocal of the square root of n over lags greater than or equal to p + 1.

Here n = 54, SO SE $[O_{kk}]$ is approximately .1361

The largest estimated partial autocorrelation value after lag 3 is at lag 5 with a value of -0.0948. The value of the 'T' statistic is 0.6956, hence all lags after three can be statistically assumed to be zero.

For our Arima model then the most we can infer is that $P \not> q$. Greenberg and Webster (1983) corroborate this conclusion.

The moving average process of order six was chosen on the basis of standard error. However, there are several methods other than standard error to decide on the correct specification for univariate time series models. One can use Lagrange multipliers, Wald statistics or Likelihood ratios. Harvey (1981) decribes these test statistics. At a higher level of complication one can resort to locally equivalent alternative models. A good discussion of these is given in Chow (1982). Engle (1984) describes the latter.

A pure significance test due to Box and Pierce (1970) was applied to the residuals from the regression of this specification. Twenty-five observations were used. Basically this test says that on the hypothesis that the fitted model is appropriate the statistic

$$Q = \sum_{K=1}^{25} r_k^2(\hat{a})$$

where $\hat{a}$ is the vector of estimated one period ahead forecast errors, is approximately distributed as a central chi-squared with (25 - p - q) degrees of freedom.

= N - d   d = 2   n = 52   N = 54

So Q = 0.4918 *52

= 25.57

The 10% and 5% points for the chi-square with 19 degrees of freedom are 27.2 and 30.1 respectively. On the basis of the data we refuse to reject the null hypothesis of model adequacy. Table one attached presents in detail the output from the program which was used to estimate the parameters. The technique was based on the conditional sum of squares. It should be noted that the corresponding conditional log likelihood function is only an approximation to the unconditional log likelihood function under certain conditions. Moreover, it is a bad approximation when the roots of the autoregressive part of the Arima process are close to the boundary of the unit circle so that the process approaches non-stationarity. In the case of the model presented here, we

have used a pure moving average. These are always stationary. We, however, have to make the assumption of invertibility.

A finite order moving average process has an equivalen[t] autoregressive process of infinite order provided certain invertibility conditions are satisfied. If invertibility is satisfied the moving average part of an arma process can then [be] converted into a purely autoregressive process. The implicati[on] for forecasting is that future values of the time series can b[e] sensibly related to past values. In our case the violation of invertibility would invalidate the forecasting mechanism presented in equation one. For a detailed analysis of invertibility see Box and Jenkins (1976).

The Forecasting Model

Now the form of the model is

(2) $(1 - B)^2 \, NFP_t = (1 - c_1, B^1 - c_6, B^6) \, a_t$

where B is the lag operator.

This implies that

(3) $NFP_t = 2 \, NFP_{t-1} - NFP_{t-2} + a_t - c_1 \, a_{t-1} - c_6 \, a_{t-6}$

After taking conditional expectations the resulting mode[l] is equation one where $NFP_t(1)$ is the one step ahead forecast c[of] the dependent variable. One step ahead forecasts were calcula[ted] since the large standard error 11,099 implies that two step an[d] three step ahead forecast would have an error that would blow more than proportionately as lead time increases.

Confidence intervals are

$\pm$ 1.96 * 11099.66 for 90% confidence

$\pm$ 0.674 * 11099.66 for 50% confidence

The 50% confidence interval suggest that we would be within seven million dollars of our forecast 50% of the time.

Although the portmanteau lack of fit test was favourable for the null hypothesis its value was somewhat close to boundaries of the critical region of the chi-square distribution at the relevant number of the degrees of freedom creating a little uneasiness about the adequacy of model fit. Possible explanation for this might have been structural change in the parameters due to some change in policy either with the commercial banks or with the central bank over the period under consideration. Foreign interest rates may have had some implication for NFP. This hints at some approach using a more general transfer function. To see this the general class of transfer function models is presented here.

$$Y_t = \frac{A(L)}{B(L)} x_t + \frac{C(L)}{D(L)} e_t$$

The arima is a special case of this with A(L) equal to zero. A(L), B(L), C(L) and D(L) are polynomials and L is the lag operator. x_t is the input and Y_t the output series. Intervention analysis may also be relevant here. Gottman (1981) discusses this topic. Two forecasts were generated. The first a static and the second a dynamic forecast. The first forecast of

NFP versus the observed in an ex post presentation is attached. This historical comparison is through the second quarter in 1972 to the first quarter in 1984.

Conclusion

We assumed invertibility; this is a strong assumption. Moreover, Monte Carlo studies have shown that non-linear least squares outperforms maximum likelihood estimation in this case over a large part of the parameter space. Some authors point out though that non-linear least squares estimates frequently violate the invertibility conditions when maximum likelihood do not. For a sample size of one hundred there is not likely to be much difference. Judge (1980) has a good survey of the relevant issues. Despite these problems, time series model building will prove important for several reasons. These short-term forecasting properties have performed well relative to other methods. The forecast from these models can be combined with that from other sources optimally. Last, but not least time series univariate modelling is a bridge to more advanced dynamic simultaneous systems.

References

Wiley, George G Judge (1980) "The Theory and Practice of Econometrics" Publisher "Wiley Series in Probability and Mathematical Statistics".

Chow, Gregory C. (1982) "Evaluating the Reliability of Macro-Economic Models" Publisher "John Wiley and Sons".

Engle, Robert F. (1984) "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics" Handbooks in Econometrics, Edited by Kenneth J. Arrow and Michael D. Intrilligator Publisher "North Holland Amsterdam New York".

Harvey, A.C. (1981) "Time Series Models" Publisher "Philip Allan Publishers Limited".

Jenkins G. M. and Box G.E.P. "(1976) Times Series Analysis" Publisher "Holdenday INC".

Bartlett, M.S. "On the Theoretical Specification of Sampling Properties of Autocorrelation Time Series" Journal Royal Statistical Society, B8, 27, 1946.

Quenouille, M.H. "Approximate Tests of Correlation in Time Series" Journal Royal Statistical Society, B11, 68, 1949.

Table 1

Summary of Statistical Output From Estimation

SMPL 1970.3 - 1983.4

54 Observations

LS - Dependent Variable is ddnfp

Convergence Achieved After 23 Iterations

	COEFFICIENT	STANDARD ERROR	T-STATISTICS
C	-158.89557	1517.7592	-0.1046909
MA (1)	-0.9757501	0.1415428	-6.8936770
MA (6)	0.2684031	0.1403579	1.9122768

R-squared	0.531141	Mean of dependent var	-314.5000
Adjusted R-squared	0.512754	S.D. of dependent var	15901.41
S.E. of regression	11099.66	Sum of squared resid.	6.28D+09
Durbin-Watson Stat.	2.007250	Log Likelihood	-578.0716

Table 2

Listing of Principal Data Used in the Studies

		nfp	dnfp	ddnfp
1970	Mar.	18165	1955	
	Jun.	20120	-9585	11540
	Sept.	10535	-8885	700
	Dec.	1650	3080	11965
1971	Mar.	4730	4972	1892
	Jun.	9702	-5973	-10945
	Sept.	3729	-11566	-5593
	Dec.	-7837	-7475	4091
1972	Mar.	-15312	-735	6740
	Jun.	-16047	1998	2733
	Sept.	-14049	-3388	-5386
	Dec.	-17437	4594	7982
1973	Mar.	-12843	3835	-759
	Jun.	-9008	-29689	-33524
	Sept.	-38697	-5522	24167
	Dec.	44219	11316	16838
1974	Mar.	-32903	-3372	-14688
	Jun.	-36275	-1680	1692
	Sept.	-37955	4496	6176
	Dec.	-33459	15925	11429
1975	Mar.	-17534	-11228	-27153
	Jun.	-28763	10836	22064
	Sept.	-17926	-1065	-11901
	Dec.	-18991	-1174	-109
1976	Mar.	-20165	-5374	-4200
	Jun.	-25539	-3957	1417
	Sept.	-29496	-3419	538
	Dec.	-32915	-578	2841
1977	Mar.	-33493	704	1282
	Jun.	-32789	-3886	-4590
	Sept.	-36675	1801	5687
	Dec.	-34874	2385	584
1978	Mar.	-32489	4166	1781
	Jun.	-28323	-393	-4559
	Sept.	-28716	4095	4488
	Dec.	-24621	11744	7649
1979	Mar.	-12877	3326	-8418

Table 2 Cont'd

Listing of Principal Data Used in the Studies

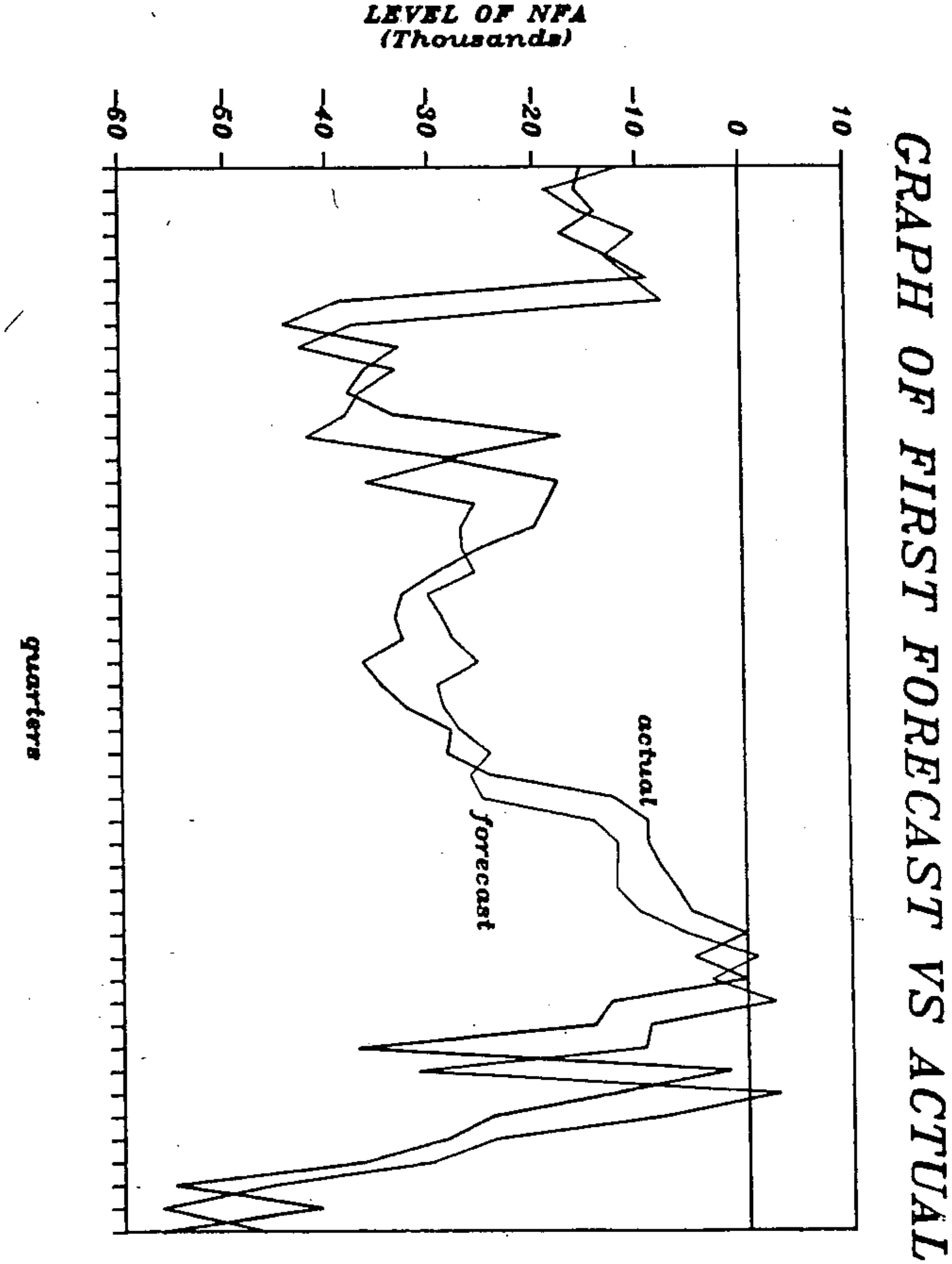
		nfp	dnfp	ddnfp
1979	Jun.	-9551	140	-3186
	Sept.	-9411	1178	1038
	Dec.	8233	1582	404
1980	Mar.	-6651	1334	-248
	Jun.	-5317	5269	3935
	Sept.	-48	-5004	-10273
	Dec.	-5052		
1981	Mar.	-104		
	Jun.	-13358		
	Sept.	-14980		
	Dec.	-37506		
1982	Mar.	-1746		
	Jun.	-13318		
	Sept.	-24778		
	Dec.	-29036		
1983	Mar.	-36897		
	Jun.	-55123		
	Sept.	-41248		
	Dec.	-56276		

Source: Annual Statistical Digest, Central Bank of Barbados - from publications between 1977 and 1984.

Note: nfp "The Net Foreign Position of the Commercial Banks"

dnfp "The First Difference of Net Foreign Position of the Commercial Banks".

ddnfp "The second difference of Net Foreign Position of the Commercial Banks.


GRAPH OF FIRST FORECAST VS ACTUAL
LEVEL OF NFA
(Thousands)
10
0
-10
-20
-30
-40
-50
-60
actual
forecast
quarters