'FORECASTING CRIMES IN PAKISTAN' USING TRADITIONAL FORECASTING METHODS

Dr. Nawaz A. Hakro
Imdad Hussain Soomro
Aftab Ahmed Mangi

ABSTRACT

An important reason for formulating an econometric model or time series model is to generate a forecast of one or more variables. The categories for econometric and time series are very broad and the line between them is not yet clear. The model used for forecasting e.g. economic variables are not however fully applicable to non economic variables i.e. number of crimes reported in Pakistan. In this paper we have employed a new indicator for forecasting evaluation, that is coefficient of correlation between actual and fitted values. The traditionally employed methods which give some better forecast in sample period but their ability for out of sample forecasting is doubtful. Hence, need is to develop special type of models for particularly non economic data. Our results also fully supported the methods proposed by Granger and Ramanathan (1984).

INTRODUCTION

A forecast is a quantitative estimate (or set of estimates) about the likelihood of future events based on past and current information (Pindyck, 1981). Broadly speaking, usually, four techniques are used for forecasting of time series data, (1) Single equation regression model, (2) Simultaneous equation regression models, (3) Autoregressive moving average integrated models, and (4) Vector auto-regression models.

The four models were gradually employed due to critiques toward these models. Simultaneous equation models earn great popularity, but loose their reliability due to the Lucas critique (1976), and the thrust of this technique is that the parameters estimated from an econometric model are dependent on the policy prevailing at the time, the model was estimated and may be change if this particular policy is rejected in favor of new one. The time series model then get their place to econometric simultaneous and single equation methods. But their predictive power also proved doubtful due to their theoretic nature, then a mixture of econometric and time series model attract the policy makers and forecasting inclined personnel. The combining of econometric and time series models is also used with different methodologies i.e. with and without weights, with and without constant term. Different methods for forecasting performance evaluation techniques are also developed. The traditionally employed methods are average absolute error, root mean square error, Theil inequality coefficients. But new methodology employed by us is coefficient of correlation between fitted and actual values to have a clear-cut decision toward the accuracy of forecasted values. But work on non-economic data (number of accidents, number of crimes reported in Pak) is rare one. Hence, the objective of our study is to fulfill this thrust. We are interested to check either which one method is useful for forecasting of non-economic data for in sample and out sample.

METHODOLOGY OR FORECASTING EVALUATION

Since different models are candidate, we are employing simple linear model and those of proposed by Pindyck (1981). in first step, the following models are estimated.

i) Simple Linear model $\quad \gamma t = \alpha_0 + \alpha T + \mu_1 t$

ii) Logarithmic model $\quad \text{Log } \gamma t = \beta_0 + \beta_1 T + \mu_2 t$

iii) Autoregressive model $\quad \gamma t = \gamma_0 + \gamma_1 \gamma t - 1 + \mu_3 t$

iv) Logarithmic autoregressive model $\quad \text{Log } \gamma t = \Pi_0 + \Pi_1 \log \gamma t - 1 + \mu_4 t$

v) Quadratic trend model $\quad \gamma t = \delta_0 + \delta_1 T + \delta_2 T^2 + \mu_5 t$

Since time series models have achieved a greater popularity from their inception, hence we will use different order autoregressive and moving average models called (ARMA), with stationary checking of autocorrelation and partial autocorrelation. Our model in general form is:

$$\gamma t = C + AR(P) + MA(q) \quad (6)$$

Where P and q are order of moving average and autoregressive terms. Bates and Grander pointed out in 1969, however, that discarded models do still contain information about the underlying behavior of dependent variable and argued that combining forecast from several models would out perform those from a single model. The following three methods are used to combine forecasts (Ramnanathan, 1992).

COMBINED FORECAST MODELS

(a) Unconstrained with Constant (Granger, Râmanathan, 1984).

$$\gamma t = \beta_0 + \beta_1 ft_1 =+ \ldots + \beta_k ft_k + \mu t$$

(b) Unconstrained with no constant.

$$\gamma t = ft + \mu t \Rightarrow \gamma t = \beta_1 ft_1 + \beta_2 ft_2 + \ldots + \beta_k ft_k + \mu t$$

(c) Constrained, with no constant and weights sum to one.

$$\gamma t = \beta_1 ft_1 + \beta_2 ft_2 + \ldots + \beta_k ft_k + \mu t$$

Where $\beta_1, \beta_2, \beta_k$ are Weights.

The underline procedure for model (a) is that if individual forecasts are biased, then their weights are likely to be biased. Hence, procedure is to add a constant term to the forecast and let the constant term adjust for the bias (Ramanathan, 1992), where γt in our model stands for the number of crimes reported in Pakistan, while At and ft are actual and fitted values respectively. All the econometric techniques and time series employed models are used for the forecast evaluation. Different criterion for in-sample and out-of sample forecast evaluation are used traditionally. Some of these criterion are:

Absolute average error (AAE) = $\sum_{t=1}^{m} \left| Ft+n - At+n \right| / m$

Root mean square error (RMSE) = $\left[\sum_{t=1}^{m} (Ft+n - At+n)^2 / m \right]^{1/2}$

Theil 'U' inequality coefficient =

$$\sqrt{\frac{\sum_{t=1}^{m} (Ft+n-At+n)^2}{\sum_{t=1}^{m} (At+n-At+n-1)^2}}$$

(Theil inequality coefficient)

(Theil inequality coefficient)

Absolute average error in percentage = $\left[\sum_{t=1}^{m} |Ft+n-At+n| /m\right]\%$

A new method for evaluation of forecast is by using simple correlation coefficient between actual and fitted values, i.e.

$$\gamma = \frac{\sum ft\ at}{\sqrt{(\sum ft^2)(\sum at^2)}}$$

$ft = (Ft - \bar{Ft})$

$at = (At - \bar{At})$

To test, if the forecasts are optimal. Mincer and Zarnowitz (1969) used the regression:

$$At = \alpha + \beta\ ft + \mu t$$

and test the joint hypothesis $\alpha = 0$ $\beta = 1$, rejection of the hypotheses indicates that the forecasts are not optimal and should be adjusted to remove the systematic error and may be called necessary condition for optimal forecast.

So forecast that gives the minimum (AAE, RMSE, AAE%) and maximum. γFT. AT will be preferred one. The evaluation criteria can be applied for both in-sample and out-sample forecasting. Total sample period is from 1990.01 to 1995.02, estimation period is 1990.01 to 1994.01, in-sample forecast period is 1993.01 to 1994.01 and out of sample forecasting period is 1994.02 to 1995.02.

RESULTS

The results of different models proposed in methodology reveal interesting features. As a first step we regress model (I to V) and obtain their RMSE AAE. AAE% and coefficient of correlation. The model (V) gave the satisfactory forecasting results in term of high coefficient of correlation and low RMSE. AAE% and Theil inequality coefficient. Hence we select this model as best models for econometric techniques employed representation. then time series model building procedure was applied. By analyzing partial and autocorrelation. different orders of autoregressive and moving average were selected. Particularly were (1.0.1), (2.0.2). (3.0.3). (5,0,5) and (6. 0. 6), but by analyzing coefficient of correlation, the order (6.0.6) was the best one. Here on explanation is necessary that by inspection of Q-statistic. no need for differentiation was noted. The results of ARMA and econometric techniques show that time series model is the best one in term of correlation and low forecast errors for in-sample forecasting (1993.03 to 1994.01) and out-sample (1994.02 to 1995.02).

In second step. combination of ARMA and econometric models were estimated with methods as (unconstrained with constant. unconstrained with constant and constrained with weights sum to one) for both in-sample and out-sample forecasting period. The results are presented in table 1, 2, and 3. Ranking of forecast mean their ranking in term of high correlation. low forecast errors, and satisfying efficiency condition. Efficiency condition was tested with hypothesis, intercept of model zero and slope coefficient one. The results guide us toward the conclusion that models proposed by Granger and Ramanathan (1984) gives the best forecast for both in-sample and out-sample forecast. But in case of out of sample. all these methods do not give satisfactory results as correlation between actual and fitted value is only 0.35. But an interesting debate starts that all three proposed methods (combined econometric and time series) give the approximately same results (correlation coefficient differs only 0.01 approximately). Hence. need is to develop other methods. although this coefficient is 0.55 for in-sample forecasting.

Conclusion

The main objective of the econometric and time series modeling is generally to get a best forecast in term of low forecast error. From their inception, econometric techniques were criticized on different grounds. The time series and simultaneous equation modeling also get their popularity. But at recent, the combination of time series and econometric modeling are performing well. We have tried to use these estimates and then use it for forecast. The new techniques we have used for having a clear-cut decision of coefficient of correlation between actual and fitted values, the ranking of models in term of low errors (RMSE, AAE, AAE% and Theil inequality coefficient). We have used the graphs of different models(actual and fitted values, and reached at the conclusion.

i) The work on economic data forecasting is common practice. But non-economic data forecasting is rare one.

ii) Traditionally employed methods although give some type of satisfactory forecast in case of in-sample forecast, but out-of-sample forecast power of these models are not good one and even with same number of forecast, root mean square error forecast (RMSE) is 11% time s higher in case of in-sample forecast, and coefficient of correlation in case of out of sample forecast is only 0.35, which is quite low. Hence, text book methods prescribed in undergraduate level type cannot be used for these types of data.

iii) So, future need is to develop the techniques which are equally good for forecasting in both cases(economic and non-economic data).

iv) Our results are although not acknowledging combined techniques in term of very high correlation, due to the fact that the data we have used is unpredictable (number of crimes reported in Pakistan), and our results may not hold in general for other data sets.

TABLE 1
IN-SAMPLE FORECAST
(ESTIMATION Period 1990.01 to 1994.01)

Forecast	AAE	RMSE	AAE %	Theil inequality Coefficient	Constant	Weights for		Coefficient of correlation	Ranks
						Econon	ARMA		
Original Econometric	1319.36	1894.65	5.59	0.04	-	1.00	-	0.14	4
ARMA	1397 53	1652.83	5.67	0.03	-	-	1.00	0.50	3
a) Combined (Unconstrained with constant									
Econometric and ARMA	1289.67*	1581.58*	5.32*	0.03*	23687.02*	0.94*	1.03*	0.55*	1
b)Combined (Unconstrained without constant)									
Econometric and ARMA	1361.56	1626.67	5.60	0.03	-	0.18	0.82	0.53	2
c) Combined (Constrained with No Constant and Weights Sum to 1)									
Econometric and ARMA	1359.46	1626.82	5.60	0.03	-	0.20	0.80	0.52	2

Notes:

Econometric model is $yt = \delta_0 + \delta_1 T + \delta_2 T^2$

ARMA model is yt = Constant + AR(6) + MA(6)

a) Combined model is $yt = \beta_0 + \beta_1$(Econometric forecast) + β_2(ARMA forecast)

b) Combined model is $yt = \beta_1$(Econometric forecast) + β_2(ARMA forecast)

c) Combined model is $yt = \beta_1$(Econometric forecast) + β_2(ARMA forecast)

Where in model (c) β_1 β_2 are weights (0.20,0.80) respectively.

TABLE 2
OUT-SAMPLE FORECAST
(ESTIMATION Period 1990.01 to 1994.01)

Forecast	AAE	RMSE	AAE%	Theil inequality Coefficient	Constant	Weights for		Coefficient of correlation	Ranks
						Econon	ARMA		
i) Original Econometric	1410.94	1890.51	6.89	0.04	-	1.00	-	-0.24	6
ii) Original Econometric	1540.82	1731.97	5.69	0.03	-	1.00	-	0.18	5
ARMA	2458.63	2857.27	8.92	0.06	-	-	1.00	0.32	4
a) Combined (Unconstrained with constant									
i) Econometric & ARMA	1204.97*	1436.59*	4.45*	0.03*	-47156.8*	-1.78*	0.96*	0.35*	1*
ii) Econometric &ARMA	1216.95	1450.79	4.50	0.03	-23687	0.94	1.03	0.33	3
b)Combined (Unconstrained without constant)									
i)Econometric & ARMA	1223.94	1452.91	4.52	0.03	-	-0.14	1.24	0.33	3
ii) Econometric &ARMA	1217.54	1451.16	4.50	0.03	-	0.12	0.96	0.33	3
c) Combined (Constrained with No Constant and Weights Sum to 1)									
i)Econometric &ARMA	1314.56	1562.30	4.94	0.03	-	-1.16	2.16	0.34	2
ii) Econometric &ARMA	1217.54	1451.17	4.50	0.03	-	0.12	0.88	0.33	3

Notes:

Original model (I) Econometric $yt = \gamma_0 + \gamma_1 yt - 1$

Original model (ii) Econometric $yt = \delta_0 + \delta_1 T + \delta_2 T^2$

ARMA model is yt = Constant + AR(6) + MA(6)

a) (i) Combined model is $yt = \beta_0 + \beta_1$(Econometric[1] forecast) + β_2(ARMA forecast)

(ii) Combined model is $yt = \beta_0 + \beta_1$(Econometric[ii] forecast) + β_2(ARMA forecast)

b) (i)Combined model is $yt = \beta_1$(Econometric[I]forecast) + β_2(ARMA forecast)

(ii) Combined model is $yt = \beta_1$(Econometric[ii]forecast) + β_2(ARMA forecast)

c) (I)Combined model is $yt = \beta_1$(Econometric [I]forecast) + β_2(ARMA forecast)

(ii) Combined model is $yt = \beta_1$(Econometric [ii]forecast) + β_2(ARMA forecast)

TABLE 3
RESULTS OF FORECAST EFFICIENCY

MODEL	COMMENTS
In-Sample Forecast	
Econometric model	YES
ARMA model	YES
Combined	YES
a) Econometric and ARMA	YES
b) Econometric and ARMA	YES
c) Econometric and ARMA	YES
Out of Sample Forecast	
a) (I) Econometric and ARMA	YES
ii) Econometric and ARMA	YES
b) (I) Econometric and ARMA	YES
(ii) Econometric and ARMA	YES
c) (I) Econometric and ARMA	YES
(ii)Econometric and ARMA	YES

Notes: Comment 'YES' means the model obeys the forecast efficiency and 'No' otherwise.

Forecast efficiency model is:

$At = \alpha + \beta ft + et$

and joint hypothesis is $\alpha = 0, \beta = 1$

At = Actual values

ft = Fitted values

References

1. Granger, C.W.J. and Ramanathan Ramu, (1984), 'Improved Methods of combining forecasts", Journal of Forecasting, Vol.3, 197-204.

2. Gujrati, Damodar N.(1995), Basic Econometrics, New York, McGraw Hill.

3. Mincer, J. and Zarnowitz, V.(1969), 'The evaluation of economic Forecasts in J.Mincer(ed), Economic forecasts and Expectations, pp.3-46, New York, National Bureau of Economic Research.

4. Pindyck, R.S. and Rubinfeild, D.L. (1981), Economic Models and Economic Forecasts, New York: McGraw Hill.

5. Ramanathan, Ramu (1992), Introductory Econometrics with Applications, Harrcourt Brace Jovarovich International Edition.

6. Yaseen, Ghulam 1996), 'Time series Analysis of Number of crimes reported in Pakistan, A term paper submitted at Applied Economics Research Centre, University of Karachi.(Unpublished).