

FORECASTING MILK PRODUCTION IN PAKISTAN

Farhan Ahmed, Hassnain Shah, Irum Raza* and Abdul Saboor**

ABSTRACT:- This paper forecasts milk production in Pakistan using time series data from 1990 to 2010. ARIMA (p, d, q) model is considered for estimation where 'p' is the order of the autoregressive process; 'd' is the order of the homogeneity and 'q' the order of the moving average process. Presence of trend in data was checked through time series plots and stationarity through auto correlation and partial auto correlation functions. Using the ARIMA (p, d, q) forecasts from 2010 up to 2015 were made and production of milk in 2015 was estimated at 47492 t.

Key Words: Dairy Sector; Forecasting; Milk Production; ARIMA Model; Pakistan.

INTRODUCTION

Agriculture, the most important sectors of the Pakistan's GDP and provides employment to 45% population as well as input for agrobased industry. Moreover, the single most important subsector of agriculture is livestock which accounts for approximately 55.1% of the agriculture value added and 11.5% to GDP during 2010-11. One of the major products of livestock is milk and this sector produced 46.44 billion liters during the same year (GoP, 2011). Pakistan was second after India in buffalo milk production and 12th in cow milk production in the world in 2009 (FAOSTAT, 2011). Raising livestock for milk and other products is a key activity in Pakistan and it provides income support to most of the farmers, a portfolio for diversification in farming, employment to landless farmers and it is major source of income in rural areas (Lohano and Soomro, 2010).

As milk production has significant importance on both supply and demand side, the policy makers need accurate and advance information regarding future availability for short and long term planning from time to time. Many researchers have conducted studies on forecasting production of agriculture commodities according to their importance for certain regions of the world. Forecasts have traditionally been made using structural

econometric models. Concentration have been given on the univariate time series models known as auto regressive integrated moving average (ARIMA) models which are primarily due to world of Box and Jenkins (1970). These models have been extensively used in practice for forecasting economic time series, inventory and sales modeling (Brown, 1959; Holt et al., 1960) and are generalization of the exponentially weighted moving average process. Several methods for identifying special cases of ARIMA models have been suggested by Box and Jenkins (1970). Makridasis et al. (1982) and Meese and Geweke (1982) have discussed the methods of identifying univariate models. Among others Jenkins and Watts (1968), Yule (1926 and 1927), Barlett (1964), Quenouille (1949), Ljunge and Box (1978) and Pindyck and Rubinfeld (1981) have also emphasized the use of ARIMA models. Lohano and Soomro (2010), using historical time series data in the Random Walk Model with the drift trend-stationary auto-regressive model, forecast annual milk production to grow by 4.17% per annum. Their results indicate that shocks to production in a year have a permanent effect on the level of future production. Burki et al. (2004) forecasted the prices of fresh milk and UHT milk from 2004 to 2010 through ARIMA modeling by analyzing the data of nominal and real prices of milk from 1971 to 2004. They reported actual and projected price of fresh and UHT milk until

* Social Sciences Institute, National Agricultural Research Centre, Islamabad, Pakistan.

**PMAS- Arid Agriculture University, Rawalpindi, Pakistan.

2009-10 and find out that the gap between real price of fresh and unit milk 2009-10.

MATERIALS AND METHODS

Time series milk production data were taken from Agricultural Statistics of Pakistan (1990 - 2010). To forecast the production of future years a choice is often to be made as to which type of model should be developed. Before applying ARIMA model time series plot was obtained to check the presence of trend in data. Auto correlation and partial auto correlation functions were determined to check for the presence of stationary series. A choice was made from class of linear time series models introduced by Box and Jenkins (1970), which are now widely used and accepted. According to them the ARIMA model is denoted by ARIMA (p,d,q) where ‘p’ is the order of the autoregressive process; ‘d’ is the order of the homogeneity i.e. the number of difference to make the series stationary; and ‘q’ the order of the moving average process. The general form of the ARIMA (p,d,q) is

$$Z_t = C + (F_1 Z_{t-1} + \dots + F_p Z_{t-p}) - (\theta_1 a_{t-1} + \dots + \theta_q a_{t-q}) + a_t$$

where,

- C = constant,
- Z_{t-1}, ..., Z_{t-p} = past series values (lags)
- θ₁ = the coefficients,

Similar to regression coefficients to be estimated of the autoregressive model where autoregressive (AR) model of order p, denoted by AR(p) is

$$Z_t = C + F_1 Z_{t-1} + F_2 Z_{t-2} + \dots + F_p a_{t-p} + a_t$$

a_t is a random variable with zero mean and constant variance. O’s are the coefficients in the moving average (MA) model, where moving average model of order q or MA (q) is

$$Z_t = a_t + \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

The major problem in ARIMA modeling technique is to choose the most appropriate values for the p, d and q. This problem can be partially resolved by looking at the Auto Correlation Function (ACF) and Partial Auto Correlation Functions (PACF)

for the series (Pindyck and Rubinfeld, 1981). The degree of the homogeneity, (d) i.e. the number of time series to be differenced to yield a stationary series was determined on the basis where the ACF approached zero.

After determining “Δ” a stationary series Δy_t its auto correlation function and partial auto correlation were examined to determine values of p and q, next step was to “estimate” the model. The model was estimated using computer package “Minitab”.

Diagnostic checks were applied to the so obtained results. The first diagnostic check was to draw a time series plot of residuals. When the plot made a rectangular scatter around a zero horizontal level with no trend, the applied model was declared as proper. Identification of normality served as the second diagnostic check. Therefore normal scores were plotted against residuals and it was declared a straight line. Secondly, a histogram of the residuals was plotted. Finding out the fitness of good served as the third check. Residuals were plotted against corresponding fitted values. Model was declared a good fit when the plot showed no pattern.

Using the results of ARIMA (p, d, q) forecasts from 2010 up to 2015 were made. These projections were based on the assumptions that agricultural price structure and policies will remain unchanged and consumer preference will remain the same.

RESULTS AND DISCUSSION

Annual production of milk for the year 1990-91 to 2009-10 has been used for modeling purposes. The modeling of the time series involved the steps of model specification, model estimation, diagnostic checking and forecasts. Auto correlation function (ACF) for the series was obtained using MINITAB software which showed a non stationary series due to large spike at lag 1 which was also beyond the 95 % confidence limits. To make the series stationary the parameters were specified as p=1 means the order of autoregressive, d=1 the difference and q =1 the moving average

FORECASTING MILK PRODUCTION IN PAKISTAN

length. ACF was plotted again which showed a stationary series because all lags were lying under the confidence limits. So, the ARIMA (1,1,1) was chosen to be an appropriate model for estimation.

Model Estimation

ARIMA (1, 1, 1) model was estimated using Minitab Computer package and estimates of model are given in Table 1.

Diagnostic Checking

For diagnostic checking of the estimated model, different diagnostic checks were applied for whether these were properly fitted or not.

Residual Analysis

The time series plot of residuals of milk production data showed normal probability plot, therefore, model was fitted properly by residual.

Modified Box-Pierce (Ljunge-Box) Chi-Square statistics

| | | | | |
|------------|-------|----|----|----|
| Lag | 12 | 24 | 36 | 48 |
| Chi-Square | 1.2 | DF | 9 | |
| P-Value | 0.999 | | | |

The modified Box and Pierce (1970) statistics for milk production, calculated above, for lag 12 is 1.2 which has the observed significance level 0.99. It indicated that it is non significant at SL = 0.05.

Forecast of Milk Production

ARIMA (1,1,1) was taken for esti-

mating 5 years ahead forecasts of milk production which are given in Table 2. Along with 95% confidence interval values. For 1990-91 forecasts of milk production was about 39650.1 thousand tonnes with lower and upper limits of 33888.1 thousand tonnes and 45412.1 thousand tonnes. A milk production forecast for the year 2015 was 47494.2 thousand tonnes with lower and upper limits 38826.3 thousand tonnes and 56162.1 thousand tonnes. Forecast of milk production showed an increasing trend.

In this study, ARIMA model was applied to forecast the production of milk in Pakistan. This would enable to predict expected milk production from 2010 onward. Such an exercise would facilitate the policy makers to foresee the requirements of milk storage, import and/or export of milk thereby allowing them to take appropriate measures in this regard. The forecasts would thus help save much of precious resources of our country which otherwise would have been wasted.

RECOMMENDATIONS

The recent development in dairy sector has stimulated the milk production and high production is also forecasted through the model. However to keep pace with the demand and existing fodder, feed and natural resource base it is suggested that the future milk supply should be based on the increase in milk productivity. Therefore there is need to develop a comprehensive

Table 1. Final estimates of parameters

| Type | Coefficients | S.E. Coefficient | t-value | P-value |
|----------|--------------|------------------|---------|---------|
| AR 1 | 0.5987 | 0.5060 | 1.18 | 0.254 |
| MA 1 | 0.8668 | 0.3852 | 2.25 | 0.039 |
| Constant | 518.26 | 95.81 | 5.41 | 0.000 |

Table 2. Forecasts from period 20 (95% Limits)

| Period | Forecast | (000't) | |
|-----------|----------|---------|--------------|
| | | Lower | Upper Actual |
| 2010-2011 | 39650.1 | 33888.1 | 45412.1 |
| 2011-2012 | 42174.7 | 35034.1 | 49315.4 |
| 2012-2013 | 44204.5 | 36341.2 | 52067.8 |
| 2013-2014 | 45938.1 | 37611.3 | 54264.8 |
| 2014-2015 | 47494.2 | 38826.3 | 56162.1 |

policy through integration of research, extension and development activities. The adoption of best practices in the dairy sector would not only help in increasing the milk supply but would also help in entering the export market.

LITERATURE CITED

- Bartlett, M.S. 1964. On the Theoretical Specification of Sampling Properties of Auto Correlated Time Series. *J. Roy. Stat. Soc. B8*: 27-41.
- Burki, A. A. Khan, M.A. and Faisal, B. 2004. The state of Pakistan's dairy sector: An Assessment: *The Pakistan Develop. Rev* 43 (2): 149-174.
- Box, G.E.P. and Jenkins, G.M. 1970. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, California, USA.
- Box, G. E. P. and Pierce, D. A. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Stat. Assoc.* 65:1509-1526.
- Brown, R.G. 1959. *Statistical Forecasting for Inventory Control*. New York, McGraw-Hill
- GoP, 2011. *Economic Survey of Pakistan (2010-2011)*, Government of Pakistan, Economic Adviser's Wing, Finance Division, Islamabad.
- Holt, C.C. Modogiliani, F. Muth, J.F. and Simon, H. A. 1960. *Planning, Production, Inventories and Work Force*. Printice Hall, Englewood Cliffs, NJ, USA
- Jenkins, G.M. and Watts, D.G. 1968. *Spectral Analysis and its Application*, Day, San Francisco, California, USA.
- Ljung, G.M. and Box, G.E.P. 1978. On a measure of lack of fit in time series models. *Biometrika*, 65:67-72.
- Lohano, H.D. and Soomro, F.M. 2010. Unit root test and forecast of milk production in Pakistan. *Intern. J. Dairy Sci.* 5(3): 204-210.
- Meese, R. and Geweke, J. 1982. A Comparison of Autoregressive Univariate Forecasting Procedures for Macroeconomics Time Series. Unpublished Manuscript, University of California Berkeley, CA, USA.
- Makridakis, S. Anderson, A. Carbone, R. Fildes, H. R. Lewandowski, M. R. Newton, J. Parzen, E. and Winkler, R. 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *J. Forecasting*, 1: 111-153.
- Pindyck, R.S. and Rubinfeld, D.L. 1981. *Econometric Models and Economic Forecasts*, 2nd edn. McGraw-Hill, USA.
- Quenouille, M.H. 1949. Approximate Tests of Correlation in time Series. *J.Roy.Stat.Soc. B11*: 68-84
- Yule, G.U. 1927. On a method of investigation periodicities in disturbed series, with special reference to Wolfer's Sun-spot Number. *Phill. Trans. A226*:267-98
- Yule, G.U. 1926. Why do we sometime get nonsense correlations between time Series: A study in sampling and the nature of series. *J. Roy.Stat.Soc.* 89:1-69.