Mining Science and Technology, 3 (1986) 301–307
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

TIME-SERIES ANALYSIS OF KENTUCKY COAL PRODUCTION

Manouchehr Vaziri

Department of Civil Engineering, University of Kentucky, Lexington, KY 40506 (U.S.A.)

(Received December, 1985; accepted January 14, 1986)

ABSTRACT

This paper focuses on short-range modeling and forecasting of monthly coal production in the state of Kentucky. The 1976–82 time-series data suggest an autoregressive moving average (ARMA) model to replicate state level of monthly coal productions. The identified ARMA model has autoregressive component of lag 1 and lag 12, and a moving average component of lag 1. It satisfies all estimation and diagnostic requirements. Model predictions for 1983 were very reasonable when compared with actual 1983 monthly coal production data: cyclical patterns were correctly replicated. Incorporation of additional data for 1983 enhanced the estimated model. Similar time-series models could be integrated into state-level planning programs for short-range forecasting of other coal industry activities. The simplicity of the ARMA model, the reliability of its predictions and the ease of updating make it very appealing when compared with large-scale econometric models which are complex and impractical for short-term coal production forecasting.

INTRODUCTION

The future of the coal industry should be a major national concern as long as the United States endures long-term dependence on foreign sources of energy. Kentucky, as the leading coal producer, had a 16.7% share of total U.S. coal production in 1983 [1]. The future of Kentucky's coal industry is of utmost interest not only at the state level; its national impact is of great importance to federal energy policies. The energy crises of the 1970's therefore gave rise to numerous efforts to model the relationships between consumption and production of energy from different sources with a host of explanatory variables derived from economic, institutional and technological factors [2–5]. The models developed to date have often been based on econometric analyses and have utilized time-series data primarily to capture structural relationships.

Although econometric models have proved useful in national forecasting of structural changes in economic conditions, they have not been very practical in short-range fore-

0167-9031/86/$03.50

casting [6–8]. This is mainly because the explanatory variables themselves have to be predicted and then employed in rather complex econometric models for the short-range forecasting task. Furthermore, the effort needed to identify, measure, estimate and update a set of econometric equations is often too extensive to serve the limited purpose of short-range forecasting at regional and state levels. A class of models particularly well suited to short-term forecasting is often referred to as ARMA, autoregressive moving average [9–11]. ARMA models, free of the problems which large-scale econometric models suffer from, replicate past behavior of a univariate time-series rather than determine direct multivariate structural relationships. Such models are particularly useful for short-term forecasting when it is expected that the underlying factors determining the level of the variable of interest in the past, such as coal production, will behave much the same in the near future.

This paper presents an ARMA model for forecasting short-term Kentucky production of coal, using monthly data from the period 1976–1983. Planning institutions, especially state and private agencies, can incorporate the methodology and findings of this study to enhance forecasting endeavors.

KENTUCKY COAL PRODUCTION TREND

Monthly coal production in the state of Kentucky between 1976 and 1983 is shown in Fig. 1 [12]. Approximately 75% of the coal produced is from Eastern Kentucky mines and the remainder is from Western Kentucky. Figure 1 reveals no general secular decline or increase over the period 1976–1983, suggesting a rather stationary behavior. The figure further suggests seasonal variation involving periodicity over 12-month cycles, with a minimum most often occurring during mid-winter. The combination of distinct seasonality and non-secular behavior suggests that the monthly time-series is a good candidate for ARMA modelling. Compared with the past 23 years coal production, as reflected in Fig. 2, short-term trends now appear more stable, suggesting a maturity stage [13] which is more tractable, at least from a short-term modelling and forecasting point of view. Seasonal variations are clearly important when coal production is viewed within a 1-year time context. Nonetheless, shocks such as unanti-

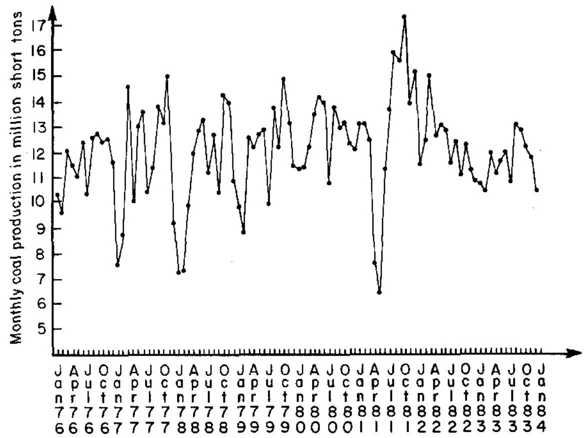

Fig. 1. Monthly Kentucky coal production for the period of 1976–83.

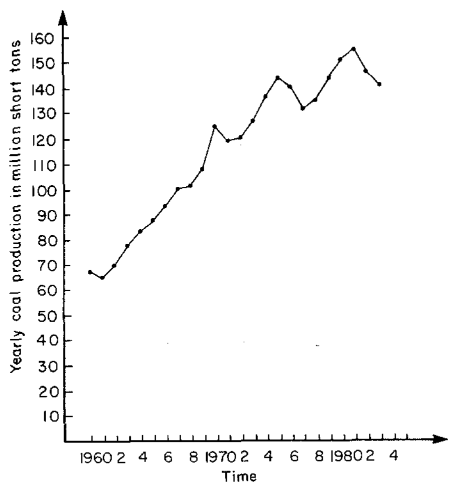

Fig. 2. Yearly Kentucky coal production for the period of 1960–83.

cipated economic and political events can easily perturb the short-term trend so as to complicate any forecasting effort.

MODEL STRUCTURE

The ARMA model and the method for accessing its parameters, as presented in the following section, were primarily developed by Box and Jenkins as a means of predicting and controlling a time series [9]. The ARMA model applied to the 1976–83 monthly time series for Kentucky coal production was of the following general functional form:

$$Y_t - \mu = \sum_{j=1}^{p} \left[\phi_j(Y_{t-j} - \mu)\right] - \sum_{j=1}^{q} \theta_j a_{t-j} + a_t \tag{1}$$

where Y_t is monthly coal production for month t; a_{t-j} and a_t are white noise variables, independent and normally distributed with mean = zero and variance σ^2; μ denotes the mean of time series; $\theta_1, \ldots, \theta_q$ and $\phi_1, \ldots, \phi_p$ are coefficients. If

$$y_t = Y_t - \mu, \tag{2}$$

$$\phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \cdots - \phi_p B^p, \tag{3}$$

and

$$\theta(B) = 1 - \theta_1 B^1 + \theta_2 B^2 - \cdots - \theta_q B^q, \tag{4}$$

where B is a backshift operator ($By_t = y_{t-1}$ and $B^d y_t = y_{t-d}$), eqn. (1) can be written as:

$$\phi(B) y_t = \theta(B) a_t. \tag{5}$$

It then follows that

$$y_t = \frac{\theta(B)}{\phi(B)} a_t = \Psi(B) a_t, \tag{6}$$

and the process y_t can be viewed as being generated from the independent random variable $\{a_t\}$, which is filtered using a linear filter with the transfer function $\Psi(B)$ and constrained by the bounds of stationarity and invertibility [10].

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + a_t \tag{7}$$

is an autoregressive model of the order p and is indicated by AR(p). In this model, y_t is expressed as a weighted average of p past y plus random noise a_t. The model

$$y_t = -\theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} + a_t \tag{8}$$

is a moving average of the order q and is indicated by MA(q). In this model, y_t is expressed as a weighted average of q prior to observed noises a, plus current noise a_t. The general model represented by eqn. (1) is an autoregressive moving average of order p and q, respectively, and is indicated by ARMA-(p, q). In this model, y_t is expressed as a function of both the preceding y and of the noise.

MODEL BUILDING

The selection of a model for any time-series data from the family of autoregressive moving average is largely a matter of judgment. Nonetheless, a generally accepted model-building strategy includes iterative identification, estimation and diagnosis stages [10]. Identification is usually based on inspection of autocorrelations (ACFs) to single out moving average (MA) components, and partial autocorrelations (PACFs) to identify autoregressive (AR) components. Furthermore, non-stationary behavior can be detected if autocorrelations fail to die out rapidly. Once a tentative model is identified, its parameters are estimated and tested for statistical significance. In addition, parameter estimates must meet the stationarity–invertibility requirement. If either criterion is not met, a new model should be identified and its parameters estimated and tested. After successful estimation and testing, the model should be diagnosed. To pass diagnosis, the autocorrelation of the re-

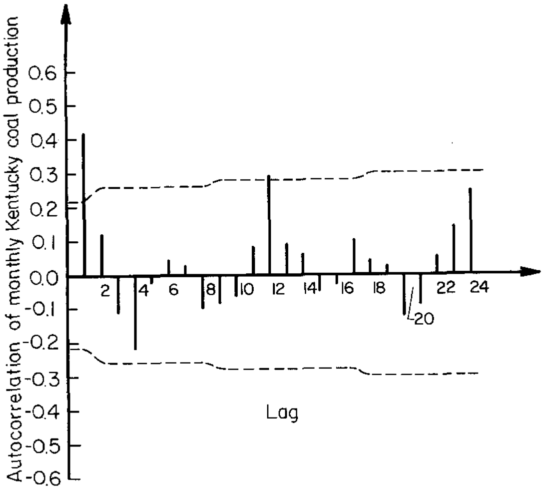

Fig. 3. Autocorrelation function of monthly Kentucky coal production.

siduals (RACFs) from the estimated model should be sufficiently small and should resemble white noise. If the residuals remain significantly correlated among themselves, a new model should be identified.

Figure 3 shows the ACFs of Kentucky monthly coal production for the period 1976–1982, and Fig. 4 shows corresponding PACFs. Dashed lines represent the 95% confidence interval. As Fig. 3 shows, the ACFs rapidly die out, and are significant at lag 1 and lag 12, suggesting the candidate MA components. Figure 4 shows that PACFs are also significant at lag 1 and lag 12, suggesting the candidate AR components. Significance of ACF and PACF at lag 12 confirm the observed seasonality of Fig. 1. It might seem sensible to incorporate regular and seasonal structures of MA and AR components additively into the ARMA model. They are usually incorporated multiplicatively, though, if their cross-product term is significant. Obviously one does not want to develop an ARMA model which includes all candidate MA and AR components suggested by ACFs and PACFs. The model should be parsimonious in this respect and at the same time pass all estimation and diagnostic checks.

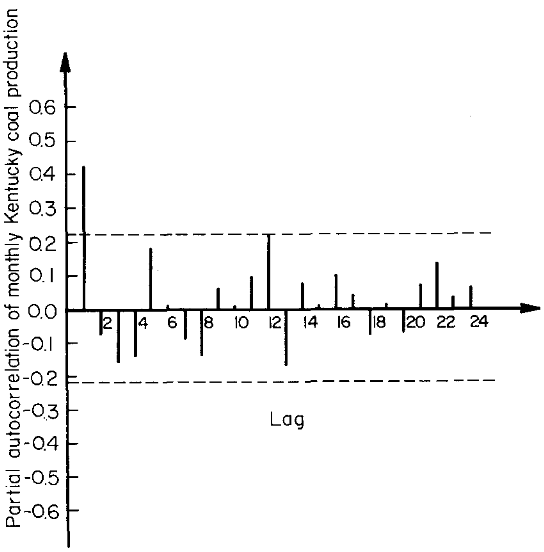

Fig. 4. Partial autocorrelation function of monthly Kentucky coal production.

The Time-Series Program, BMDP [14], was used to estimate and evaluate several candidate ARMA models. The selected model, with the smallest root-mean-square error (RMSE), has the following form:

$$(1-\phi_1 B^1)(1-\phi_{12}B^{12})y_t=(1-\theta_1 B^1)a_t \quad (9)$$

Equation (9) can be presented in the general form of relation (6) as

$$y_t=\frac{(1-\theta_1 B^1)}{(1-\phi_1 B^1-\phi_{12}B^{12}+\phi_1\phi_{12}B^{13})}a_t, \quad (10)$$

with stationarity bounds of

$$-1<\phi_1,\ \phi_{12}<+1, \quad (11)$$

and invertibility bounds of

$$-1<\theta_1<+1. \quad (12)$$

The estimated ARMA model for the 1976–1982 Kentucky monthly coal production has the following form:

$$\underset{(4.18)}{(1-0.441\,B^1)}\ \underset{(8.33)}{(1-0.781\,B^{12})}\,(Y_t-12060.00)$$

$$=\underset{(4.53)}{(1-0.669\,B^1)a_t}, \quad (13)$$

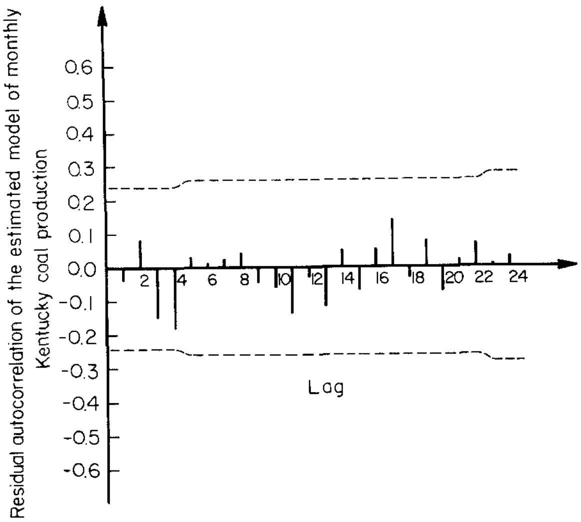

Fig. 5. Residual autocorrelation function of the estimated model of monthly Kentucky coal production.

with RMSE = 1692.40, where Y_t is Kentucky monthly coal production in month t in thousands of short tons, a_t is the white noise, RMSE is the root-mean-sqaure error, and 12,060 is the mean monthly coal production for the period of 1976–1982. The t statistics for parameter estimates are shown in parentheses below each parameter and are significant at 95% confidence. Equation (13) satisfies the stationarity and invertibility requirements (11) and (12). Figure 5 shows the autocorrelation of the residuals (RACFs) of the model. As can be seen, all the values are inside the range of 95% confidence interval and are therefore not significant. To check whether the entire residual autocorrelation is different from what could be expected of white noise, the portmanteau test was performed [9].

Table 1 summarizes the test conclusions. The Q statistic is the sum of the squares of first K residual autocorrelations multiplied by the number of observations for the period of time-series study. Q values distributed approximately chi-square with the degree of freedom equal to K minus the number of

TABLE 1

Portmanteau test

K	Q	Degree of freedom	Level of significance
12	7.76	9	0.56
24	12.79	21	0.91

estimated parameters. Table 1 shows the results are not significant at the 0.05 level.

MODEL FORECASTING AND UPDATING

Equation (13) was used to forecast Kentucky monthly coal production for the 12-month period beginning in January 1983, assuming the stochastic process in the derived model would continue throughout 1983. Figure 6 presents model forecasts with a 95% confidence interval for the period 1982–83. The model projection extends past seasonality throughout 1983.

The results suggest the ARMA model of eqn. (13) can be enhanced by combining 1976–82 and 1983 data and re-estimating the parameters. The revised model, using 8 years

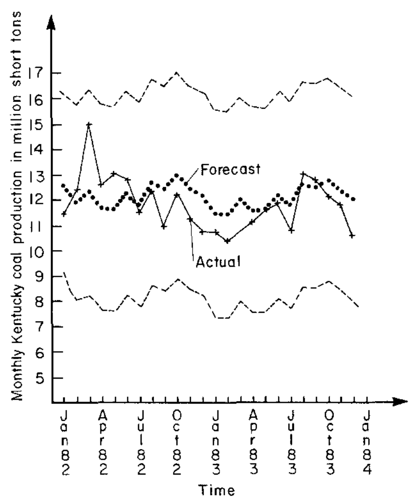

Fig. 6. Comparison of actual and forecast monthly Kentucky coal productions.

TABLE 2

Comparison of forecast and actual 1982–83 monthly Kentucky coal production in thousands of short tons

Month	Actual production	1976–82 ARMA model forecast	1976–83 ARMA model forecast
1982:			
Jan.	11496	12606	12492
Feb.	12434	11942	11832
Mar.	15036	12317	12287
Apr.	12621	11648	11614
May	13049	11600	11570
Jun.	12838	12282	12267
Jul.	11540	11801	11717
Aug.	12393	12758	12724
Sep.	11003	12584	12553
Oct.	12206	13015	12982
Nov.	11204	12516	12498
Dec.	10845	12223	12159
Total	146665	147292	146695
1983:			
Jan.	10737	11504	11345
Feb.	10478	11535	11395
Mar.	11911	12070	12018
Apr.	11100	11654	11597
May	11642	11664	11614
Jun.	11915	12217	12193
Jul.	10834	11851	11765
Aug.	13080	12601	12575
Sep.	12850	12467	12441
Oct.	12159	12805	12784
Nov.	11790	12416	12398
Dec.	10446	12187	12127
Total	138942	144971	144256
Total 1982–83	285607	292263	290951

of time-series, reflects improvements as evidenced by the lower RMSE and higher t statistics for the parameters. The re-estimated ARMA model is:

$$\underset{(4.61)}{(1 - 0.488\, B^1)}\, \underset{(10.35)}{(1 - 0.799\, B^{12})}\, (Y_t - 11999.81) = \underset{(5.65)}{(1 - 0.690\, B^1)}\, a_t, \qquad (14)$$

with RMSE = 1582.78. The revised model satisfies all the estimation and diagnosis requirements, and has a smaller RMSE than the former model, eqn. (13).

Table 2 compares historical values and model forecast values for coal production. For most of the months, forecasts are accurate to within 5% of actual production. The ARMA model developed on the basis of 1976–82 time-series data, eqn. (13), overestimates 1983 coal production by about 4%. Nonetheless, the forecasts are very reasonable and accurately replicate the 1983 cyclical patterns as depicted in Fig. 6. The updated model for 1983, eqn. (14), provides better prediction with a smaller average percentage overestimation.

CONCLUSIONS AND RECOMMENDATIONS

The superiority of ARMA modeling for short-term coal production forecasting lies mostly in the simplicity of structure and limited data requirements. Basically univariate ARMA models can capture seasonality and stochastic behavior of a stationary process without addressing causal relationships among various economic and institutional variables. The applied ARMA modelling for Kentucky monthly coal production consisted of iterative stages of identification, estimation and diagnosis. The selected model for 1976–83 time-series data showed that Kentucky monthly coal production has a seasonality of 12 months, and also depends on the past month of production and white noise. Although the selected ARMA model is dependent on the 1976–83 data, the same methodology can be applied to any coal industry activity that has a time-series stationary behavior.

Short-term forecasting, which closely tracks coal production and is updated periodically, can be incorporated with longer-term econometric models to enhance the effectiveness of state planning programs. Furthermore, short-range predictions can fortify decisions affecting coal-industry activities as well as national energy policies.

Other applications of ARMA modelling within the industry may be suggested by time-series stationary fluctuations in:

(1) Equipment productivity as affected by cyclic maintenance and replacement schedules;
(2) Storage area requirements as affected by transport capacity response to herein noted cyclic production variations;
(3) Transport facility rolling/floating stock and/or conveyer requirements as affected by cyclic shifting of coal sources and destinations;
(4) Manpower requirements as affected by cyclic production variations.

All such cyclic effects, requirements and activities within the coal industry should be evaluated for potential management decisions. They may be obtained through short-term forecasting techniques and can be easily updated as autoregressive moving averages. Together with any forthcoming direct multivariate structural relationships between such time-series stationary fluctuations, powerful management tools can be developed.

REFERENCES

1 E.R. Slatich, Coal data, a reference. Office of Coal, Nuclear Electric and Alternate Fuels, Report No. DOE/EIA-0064 (84), Energy Information Admin., U.S. Dept. of Energy, Washington, D.C., January 1985, 77 pp.

2 J. Duffield, Forecasting coal demand in the Western United States. In: G. Fenhann (Ed.), Proc. Riso Int. Conf. on Use of Simulation Models in Energy Planning, Riso National Laboratory, Roskilde, Denmark, May 1983, pp. 101–113.

3 W.C. Labys, F. Shahrokh, Y.J. Wang and R.L. Sanford, A dynamic forecasting model of the United States coal market. In: Y.J. Wang and R.L. Sanford (Eds.), Proc. 1st Conf. on Use of Computers in Coal Industry, Dept. of Mineral Resource Economics, College of Mineral and Energy Resources, West Virginia Univ., Morgantown, WV, 1983, pp. 619–627.

4 R.S. Pindyck, The Structures of World Energy Demand. MIT Press, Cambridge, MA, U.S.A., 1979, 300 pp.

5 U.S. Dept. of Energy, Short-Term Energy Outlook, Volume 2: Methodology. Report No. DOE/EIA-0202 (82/2Q)-2, Energy Inf. Admin., Office of Energy Markets and End Use, Washington, D.C., May 1982.

6 J.P. Cleary and H. Levenbach, The Professional Forecaster, The Forecasting Process Through Data Analysis, Lifetime Learning, Belmont, CA, U.S.A., 1982, 402 pp.

7 R.S. Pindyck and D.L. Rubinfeld, Econometric Models and Economic Forecasts. 2nd Edn, McGraw Hill, NY, 1982, 630 pp.

8 H. Theil, Principles of Econometrics. Wiley, NY, U.S.A., 1971, 736 pp.

9 G.E. Box and G.M. Jenkins, Time Series Analysis: Forecasting and Control. Holden-Day, London, 1976, 575 pp.

10 R. McCleary and R.A. Hay, Applied Time Series Analysis for the Social Sciences. Sage, London, 1980, 330 pp.

11 S.M. Pandit and S.M. Wu, Time Series and System Analysis with Applications. Wiley, NY, U.S.A., 1983, 650 pp.

12 Kentucky Coal Journal, Frankfort, KY, U.S.A., (monthly, 1975–present).

13 Kentucky Dep. Mines and Minerals, Annual report. Lexington, KY, U.S.A., December 1983.

14 W.J. Dixon (Ed.), BMDP Statistical Software, 1983 printing with additions. Univ. California, Los Angeles, CA, U.S.A., 1983, 735 pp.