Time series forecasting of quarterly railroad grain carloadings

Michael W. Babcock a,*, Xiaohua Lu b, Jerry Norton c

aEconomics Department, Kansas State University, Manhattan, KS 66506, USA
bEconomics Department, Kansas State University, Manhattan, KS 66506, USA
cUSDA/AMS/TMD/MTA, 1400 Independence Ave., SW, Room 1207, South Building, Washington, DC 20250-0266, USA

Received 11 May 1998; received in revised form 14 September 1998; accepted 23 October 1998

Abstract

The participants in the grain logistics system need forecasts of railroad grain carloads. Although forecasting studies have been conducted for virtually every mode, no forecasting studies of quarterly railroad grain transportation have been published. The objectives of the paper are (1) specify a US quarterly railroad grain transportation forecasting model, and (2) empirically estimate the model. The selection of explanatory variables requires that they have a theoretical relationship to railroad grain transportation supply and/or demand, and that the data for the explanatory variables are published in quarterly frequency. However, there are relatively few potential explanatory variables that are published quarterly and those that are available appear to have weak correlation with quarterly railroad grain carloadings. The economic process generating quarterly railroad grain carloadings is quite complex and very difficult to model with regression techniques. Given this problem and the focus on short run forecasting, a time series model was employed to forecast quarterly railroad grain carloadings. An AR(4) model was estimated using the Maximum Likelihood estimation procedure for the 1987:4–1997:4 period. The actual railroad grain carloadings for this period were compared to the forecast carloadings generated by the time series model. For 92% of the 37 quarters the percentage difference between the actual and forecast values was 10% or less. Of the 9 annual observations, the per cent difference between the actual and forecast value was less than 2.6% for 8 of the 9 years. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Time series forecasting; Railroad grain carloads

1. Introduction

Forecasting studies have been conducted by transportation analysts for virtually every mode. However there have been relatively few railroad forecasting studies. The literature includes Babcock and German (1981), Buxton (1982), Babcock and German (1982), Poff and Kuch (1982), Babcock and German (1983), McGeehan (1984), Stern and Cobb (1985) and German and Babcock
(1985). Two of these studies (Poff and Kuch, 1982; Stern and Cobb, 1985) involved forecasting models for individual railroads and there is virtually no discussion of the forecasting methodology. The Babcock and German studies (1981 and 1983) developed long term forecasts of railroad manufactured products tonnage. The only studies in the area of railroad grain transportation are Babcock and German (1982), Buxton (1982), and German and Babcock (1985). The Buxton (1982) study forecast railroad grain tonnage for a 20 year forecast period using a very simplistic methodology. Farm sales of grain are forecast and this figure is multiplied by the railroad historical share of the grain transportation market to obtain the forecast of rail grain tonnage. The Babcock and German (1982) and German and Babcock (1985) studies develop 3 to 5 year forecasts of railroad grain tonnage using single equation regression techniques. These two studies have two fundamental problems. One of these is the use of contemporaneous explanatory variables which must be forecast before the forecast of railroad grain tonnage can be obtained. This introduces a substantial source of error into the forecast. The second problem of these studies is that they develop no theoretical framework to guide the specification of the forecasting models.

No forecasting studies of quarterly railroad grain transportation have been published. The intent of this paper is to remedy that omission and develop a forecasting model that avoids the use of contemporaneous explanatory variables and is based on a sound theoretical framework.

One possible reason for the lack of forecasting studies of the type addressed in this paper is the dynamic instability of railroad grain carloadings. An examination of Fig. 1 reveals that there is a great deal of variation in quarterly railroad grain carloadings during the 1986:1 to 1997:4 period, ranging from a high of 424,658 in 1988:1 to a low of 259,033 in 1986:2.

The participants in the grain logistics system need forecasts of railroad grain carloads. Grain shippers need these forecasts to evaluate transportation equipment needs, establish marketing plans, and formulate strategies for negotiating prices and service with railroads. Federal transportation policymakers require these forecasts to measure the effects of past and prospective transportation policies. Policymakers can combine the railroad forecasts with other data to evaluate trends in modal market shares in the grain transportation market. Port authorities require forecasts of rail grain transportation to monitor port utilization and as an input to port expansion plans. Railroad car leasing companies need forecasts of railroad grain transportation demand to anticipate the demand for covered hopper cars.

Forecasts of railroad grain carloads are needed by the transportation firms in the grain logistics system. Class I railroads need these forecasts to formulate business plans, set railroad operations staffing levels, and as an input for equipment planning decisions. Short line railroads need these forecasts to establish business plans and assess equipment availability. Water carriers need railroad grain transportation forecasts to monitor the grain transportation market and assess their performance relative to their primary competitors—the railroads.

Given the need for quarterly railroad grain transportation forecasts by all participants in the U.S. grain logistics system, the objectives of the paper are as follows:

1. Specify a US quarterly railroad grain transportation forecasting model.
2. Empirically estimate the model developed in Objective 1.

---

1 In this study, grain includes the following commodities: barley, corn, oats, rice, rye, sorghum, wheat, and soybeans. Soybeans are included because of their agricultural significance even though soybeans are classified as an oil seed.
2. The model

2.1. Types of forecasting models

According to Pyndyck and Rubinfeld (1991) there are three general classes of forecasting models. Each class involves a different amount of complexity and each assumes a different amount of understanding concerning the structure one is trying to model.

One class of forecasting model is time series models. In these models we presume to know little or nothing about the causal relationships that affect the dependent variable that we are trying to forecast. Instead we examine the past behavior of the dependent variable in order to infer something about its future behavior. The method employed to produce the forecast could be as simple as linear extrapolation or could involve the use of a complex stochastic model. The decision to employ a time series model usually occurs if little is known about the determinants of the dependent variable, when ample historical data is available, and when the model is to be used for short term forecasting.

Another class of forecasting model is single equation regression models. In this class of models the dependent variable is explained by a single equation involving a number of independent
variables. The forecasting equation is often time dependent so that one can forecast the response over time of the dependent variable to the forecast values of the independent variables.

The third class of forecasting model suggested by Pyndyck and Rubinfeld (1991) is multi-equation simulation models. In this class of model the dependent variable may be a function of many explanatory variables which are related to each other as well as to the dependent variable through a group of equations. The construction of a simulation model starts with the specification of a group of relationships, each of which is fitted to available data. Simulation involves solving these equations simultaneously.

2.2. Specification of railroad grain transportation forecasting model

The objective of the paper is to forecast quarterly railroad grain carloadings. This variable can be thought of as a market equilibrium quantity of rail grain transportation which is determined by the interaction of railroad grain transportation demand and supply as depicted in Fig. 2.

According to Boyer (1997) demand for railroad grain transportation is derived from grain supply in origin areas and grain demand at market destinations. Assuming that grain producers will supply more grain at a higher price we can write an equation for grain supply in origin areas as:

\[
Q = aP_0
\]

where

- \( Q \) — quantity supplied by grain producers in origin areas,
- \( a \) — a constant parameter;
- \( P_0 \) — price of grain in origin areas.

Assuming grain buyers will purchase less grain at a higher price we can write an equation for grain demand at market destinations as:

\[
\text{Fig. 2. Determinants of railroad grain carloadings: equilibrium quantity of railroad grain transportation.}
\]
where
\[ Q = b - cP_d \] (2)

\[ T = P_d - P_0 \] (3)

Eq. (4) reveals that the demand for railroad grain transportation is derived from the supply of grain and the demand for grain. Thus the demand for railroad grain transportation will increase if there is an increase in grain supply in origin areas. This could be caused by a reduction in farm production costs or a change in agricultural technology. Similarly the demand for railroad grain transportation will increase if there is an increase in grain demand at destination markets. Since there is both a domestic and international demand for US grain, a wide variety of factors could affect the demand for US grain and thus the demand for railroad grain transportation. These variables would include changes in food preferences, income growth, world grain production, exchange rates, and world population growth. Thus many variables affect the US supply of and demand for grain and therefore the derived demand for railroad grain transportation.

As indicated by Fig. 2 the demand for railroad grain transportation is also affected by the price and service performance of competing transportation modes, primarily motor carriers and water carriers. The substitutability of railroads, trucks, and water carriers and thus the impact of competing mode price and service on railroad grain transportation demand, depends on such factors as the type of grain, the distance of the shipment, and the distance of the shipper from a water port. In general, motor carriers have a comparative cost advantage for short hauls while railroads and water carriers have a comparative cost advantage for long hauls. In the domestic market, corn is shipped a relatively short distance to corn processing plants and livestock feeding operations. These movements are dominated by motor carriers. However, corn is also shipped a long distance to US ports for export and railroads and water carriers dominate these shipments (Fruin, Halbach and Hill, 1990, p. 10). In contrast, wheat is generally shipped a relatively long distance to both domestic flour mills and to US export ports. Thus railroads dominate these
shipments (Reed and Hill, 1990, p. 14). The ability of water carriers to compete with railroads depends on the distance of the grain shipper from a water port. Trucks and water carriers are complimentary modes with water carriers depending on trucks to deliver grain to water ports. As the distance of the shipper from a water port increases, the motor carrier price increases and the cost of the combined motor carrier–water carrier shipment rises relative to the railroad cost.

The above discussion indicates that motor carriers and railroads are not close substitutes in most grain transportation markets. However, this is not the case for railroads and water carriers (with trucks often playing a supplementary role). Thus changes in water carrier price and service (i.e., speed, dependability, safety) will affect the demand for railroad grain transportation.

According to the forecasting framework in Fig. 2 the equilibrium quantity of railroad grain transportation is also affected by the supply of railroad grain transportation service. Railroad input price changes, technological improvements (such as the development of unit trains), and the supply of covered hopper cars are three important factors that affect the supply of railroad grain transportation service and thus the equilibrium quantity of railroad grain transportation.

Summarizing, the equilibrium quantity of railroad grain transportation is determined by the demand for and supply of railroad grain transportation service. The demand for railroad grain transportation service is affected by the factors that affect the supply of grain in origin areas, the variables that influence the demand for grain at market destinations, and the price and service performance of competing transportation modes (relative to railroads). The supply of railroad grain transportation service depends on such factors as railroad input prices, technological changes, and the supply of covered hopper cars.

2.3. Forecasting model specification and data availability

Since the objective is to forecast quarterly railroad carloadings of grain, the selection of explanatory variables requires that they have a theoretical relationship to the dependent variable and that the data for the explanatory variables is published in quarterly frequency. This dual requirement has the effect of limiting the number of potential explanatory variables for the forecasting model.

The stock of covered hopper cars affects railroad grain transportation supply and quarterly values of this variable can be derived. Unfortunately the correlation between the quarterly stock of covered hopper cars and quarterly railroad carloadings of grain is not very high for any of several lagged and unlagged specifications of the stock of covered hopper cars. Also the sign of the correlation coefficients is negative which is inconsistent with theoretical expectations since an increase in grain transportation supply should produce an increase in the equilibrium quantity of railroad grain carloadings.

Railroad cost is another variable that affects the supply of railroad grain transportation and is available in quarterly frequency. The Association of American Railroads (AAR) publishes the Rail Cost Adjustment Factor (RCAF), a quarterly forecast of railroad inflation based on

\footnote{Correlation coefficients for the quarterly stock of covered hopper cars and quarterly railroad carloadings of grain were computed for the contemporaneous relationship and four quarterly lagged specifications for the 1982:1 to 1996:4 period.}
the AAR’s All Inclusive Price Index which measures changes in the price levels of inputs to railroad operations. A correlation analysis of RCAF and quarterly railroad grain carloadings for the period 1988:1 to 1997:3 yielded relatively low correlation coefficients for the contemporaneous relationship as well as several lagged specifications. Also, the positive sign of the correlation coefficients is not consistent with theoretical expectations. An increase in railroad costs would decrease railroad supply and the equilibrium quantity of quarterly railroad grain carloads.

The US Department of Agriculture publishes several quarterly data series that measure the demand for US grain and therefore the quarterly derived demand for railroad grain transportation. The international demand for US grain is measured by US grain exports and railroad grain cars released from ports. Domestic disappearance measures the demand for US grain within the United States. Other variables such as total disappearance and volume moved reflect both the international and domestic demand for US grain.3

Table 1 contains correlation coefficients between quarterly railroad grain carloads and the variables described in the previous paragraph for the contemporaneous relationship as well as various lagged relationships. With the exception of the contemporaneous correlation for US grain exports (0.54) and for railroad grain cars released from ports (0.81), the correlation coefficients of the US grain demand variables with quarterly railroad grain carloadings are quite low. The high contemporaneous correlations for the two measures of international demand for US grain are of limited usefulness for forecasting purposes. This is because the contemporaneous value of the explanatory variable must be forecast before a forecast of quarterly railroad grain carloadings can be obtained which introduces another source of error into the forecast.

This discussion has indicated that there are relatively few potential explanatory variables that have published data in quarterly frequency. Those variables for which data are published quarterly appear to have weak correlation to railroad quarterly grain carloadings. This result indicates that the process generating quarterly railroad grain carloadings is quite complex and would be very difficult to model with single equation regression or multi-equation simulation. Given this problem and the focus on short run forecasting, we turn our attention to developing a time series model to forecast quarterly railroad grain carloadings.

2.4. The time series forecasting model

The forecasting model for railroad grain carloads is generated from the Autoregressive Integrated Moving Average (ARIMA) model popularized by Box and Jenkins (1976). Like other time series models, it is assumed that the environment in which a time series was generated will not change significantly in the near future so that the model which is built on the history of the time series can be used to forecast the future values of the time series.

Let \( y_1, y_2, \ldots, y_T \) be a time series. A general ARIMA model for \( \{y_t\} \) can be expressed as follows:

\[
Z_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_p Z_{t-p} + \epsilon_t - \alpha_1 \epsilon_{t-1} - \ldots - \alpha_q \epsilon_{t-q}
\]

---

3 Volume moved is defined as total disappearance minus US grain used on farms.
where, \( \{Z_t\} \) is a stationary time series from \( \{y_t\} \) obtained by differencing \( d \) times, where \( \{\varepsilon_t\} \) is a series of white noise, \( p \) is called the order of autoregression, and \( q \) is called the order of moving average.\(^4\)

The Box–Jenkins methodology is an approach to finding, for a given set of data, an ARIMA model that adequately represents the data generating process. The methodology involves three steps known as identification, estimation and diagnostic checking, and forecasting.

---

\(^4\) A stationary process is defined as one whose joint probability distribution and conditional probability distribution are both invariant with respect to time. This implies that the mean, variance and covariance of the distributions are also invariant with respect to time.
In the identification step, autocorrelation functions and partial autocorrelation functions are employed to find the appropriate \( d, p \) and \( q \) to reduce the number of models that could be fit. The autocorrelation functions of the series are defined as:

\[
\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (k = 0, 1, 2, \ldots)
\]

where, \( \gamma_k = \text{cov}(y_t, y_{t+k}) = E[(y_t - E[y_t])(y_{t+k} - E[y_{t+k}])] \) is the autocovariance function of \( y_t \), which gives us an idea of how the members of a time series depend on one another, and \( \gamma_0 \) is the variance of \( y_t \), \( k \) is the number of lags.

The autocorrelation functions \( \rho_k(k = 0, 1, 2, \ldots) \) have some important properties which enable us to use them as a tool to identify the order of difference (\( d \)) to obtain a stationary time series from \( y_t \), and to identify, together with the partial autocorrelation functions, the order of moving average (\( q \)) and the order of autoregression (\( p \)) in a potential ARIMA process.

The first property of the autocorrelation functions is that \( \rho_k(k = 0, 1, 2, \ldots) \) decline quickly toward zero with increasing lag \( k \) if the process is stationary. This property of autocorrelation functions can be easily illustrated by an AR(1) process \( y_t = \rho y_{t-1} + \epsilon_t \), which is stationary when \( \rho < 1 \). It can be calculated that \( \rho_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \rho^k (k = 0, 1, 2, \ldots) \) in this case. The autocorrelation function approaches zero quickly with the increasing lag \( k \). Therefore, the failure of the autocorrelation functions to approach zero rapidly suggests that the underlying process is non-stationary and differencing of the time series is necessary to obtain a stationary time series so that an ARMA model could be applied. Differencing continues until the resulting series is stationary. The theoretical autocorrelation functions are unknown. Thus, the estimated autocorrelation functions (which tend to follow the behavior of the theoretical autocorrelation functions) are calculated from the sample and used as an indicator of stationarity.

\[
\left( \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} (k = 0, 1, 2, \ldots, N) \text{ where } \hat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y}), \bar{y} = \frac{1}{T}\sum_{t=1}^{T}y_t, \text{ and } N \text{ is small relative to } T \right).
\]

The second property of autocorrelation functions is that for a moving average process of order \( q, \text{MA}(q) \), the autocorrelation functions will have a cutoff at \( q \), i.e., \( \rho_q \neq 0 \) but \( \rho_k = 0 \) for all \( k > q \). Consequently, if the estimated autocorrelation functions are significantly not equal to zero for \( k = q \) and equal to zero for all \( k > q \), then it is reasonable to believe that the process has a moving average component of order \( q \).

Partial autocorrelation functions, which are defined below, is another tool which is frequently used with autocorrelation functions to identify the orders of autoregression and moving average.

An autoregressive process of order \( k \), \( \text{AR}(k) \) can be expressed as

\[
y_t = \theta_{k1}y_{t-1} + \theta_{k2}y_{t-2} + \ldots + \theta_{kk}y_{t-k} + \epsilon_t.
\]

An important property of \( \text{AR}(k) \) is that the \( k \)th coefficient \( \theta_{kk} \) which measures the correlation between \( y_t \) and \( y_{t-k} \) not accounted for by an \( \text{AR}(k-1) \) is non-zero. We call the sequence
\( \theta_{kk}(k = 1, 2, \ldots) \) of partial autocorrelations the partial autocorrelation functions. \( \theta_{kk} \) can be estimated from the equation \( y_k = Y_k \theta + \varepsilon_k \), where,

\[
Y_k = \begin{bmatrix} y_k & y_{k-1} & \cdots & y_1 \\ y_{k+1} & y_k & \cdots & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ y_{T-1} & y_{T-2} & \cdots & y_{T-k} \end{bmatrix}, \quad y_k = \begin{bmatrix} y_{k+1} \\ y_{k+2} \\ \vdots \\ y_T \end{bmatrix}, \quad \theta_k = \begin{bmatrix} \theta_{k1} \\ \theta_{k2} \\ \vdots \\ \theta_{kk} \end{bmatrix}, \quad \text{and} \quad \varepsilon_k = \begin{bmatrix} \varepsilon_{k+1} \\ \varepsilon_{k+2} \\ \vdots \\ \varepsilon_T \end{bmatrix},
\]

and \( \hat{\theta}_{kk} \) is the last coordinate of \((Y_k'Y_k)^{-1}Y_k'y_k\). If the estimated \( \hat{\theta}_{pp} \neq 0 \) but \( \hat{\theta}_{kk} = 0 \) for all \( k > p \), it is reasonable to believe that the underlying process has an AR\((p)\) component.

In summary, the rules which are used to identify an ARIMA\((p, d, q)\) model are as follows:

1. If the autocorrelation functions taper off slowly or do not approach zero quickly as \( k \) increases, nonstationarity is indicated and differencing is suggested until stationarity is obtained.
2. If the autocorrelation functions have a cut-off at point \( q \), i.e., \( \hat{\rho}_q \neq 0 \) but \( \hat{\rho}_k = 0 \) for all \( k > q \), while the partial autocorrelation functions taper off, then the underlying process is likely an MA\((q)\).
3. If the partial autocorrelation functions \( \hat{\theta}_{pp} \neq 0 \) but \( \hat{\theta}_{kk} = 0 \) for all \( k > p \), while the autocorrelation functions taper off, then the underlying process is likely an AR\((p)\).
4. If neither the autocorrelation nor the partial autocorrelation functions have a cutoff point, an ARMA model may be adequate. The orders of AR and the MA have to be inferred from the particular pattern of the autocorrelation and partial autocorrelation functions.

The second step of the ARIMA methodology is estimation. Three different procedures are available to estimate an identified autoregressive moving average model: (1) the Conditional Least Squares (CLS) estimation procedure which minimizes the residual sum of squares conditional on the assumption that the past unobserved errors are equal to zero; (2) the Unconditional Least Squares (ULS) estimation procedure which minimizes the residual sum of squares without condition; and, (3) the Maximum Likelihood (ML) estimation procedure which maximizes the log likelihood function. CLS is an approximation of ULS. ML is more accurate in some circumstances than ULS, but with a good model specification and a sufficiently long time series, the estimates from all three procedures will be about the same.

The estimation step also includes diagnostic checking. Usually, the identification step can only reduce the number of possible models that could be fit since in most cases the autocorrelation functions and partial autocorrelation functions will not have clear-cut patterns. Therefore, it is necessary to fit several different models and select the best of them. The criterion which are commonly used in the model selection includes the variance of residuals, Akaike’s Information Criterion (AIC) which is defined as \(-2\ln(L) + 2k\), where \( L \) is the likelihood function and \( k \) is the number of free parameters, and Schwarz’s Bayesian Criterion (SBC) which is defined as \(-2\ln(L) + \ln(n)k\), where \( n \) is the number of residuals that can be computed for the time series. The lower the variance of residuals, AIC or SBC, the better the goodness of fit.
The third step is forecasting using the estimated equation. It has been shown that the forecast is unbiased which means that on average the forecast is equal to the true future value of the series. The forecast variance for the h-step-ahead forecast can be expressed as follows:

\[ V(h) = \sigma^2 \left\{ 1 + \varphi_1^2 + \varphi_2^2 + \ldots + \varphi_{h-1}^2 \right\} \]

where, \( \sigma^2 \) is the mean square error, \( \varphi_1 = \alpha_1 + \theta_1 \), and \( \varphi_i (i = 2, 3, \ldots, h - 1) \) are functions of the estimated coefficients of the model which can be calculated using the formula:

\[
\varphi_i = \begin{cases} 
\alpha_i + \sum_{j=1}^{\min(i,p)} \theta_j \varphi_{i-j} & (i = 1, 2, \ldots q) \\
\sum_{j=1}^{\min(i,p)} \theta_j \varphi_{i-j} & (i > q) 
\end{cases}
\]

3. Empirical results of the time series forecasting model

The identification step for developing a forecasting model for quarterly railroad grain carloadings begins with an examination of the seasonal pattern of the time series (see Fig. 1). This examination reveals an observable seasonal pattern for which railroad grain carloadings are relatively high in the first and fourth quarters and relatively low in the second and third quarters. Given this annual seasonal pattern it is reasonable to assume that the number of carloads of the same quarter last year may be a good estimation for the number of carloads of the corresponding quarter for this year. To eliminate this seasonal effect we difference the series by a one year lag \( y_t = x_t - x_{t-4} \). The autocorrelation functions of \( \{y_t\} \) decrease rapidly as \( k \) increases. This result indicates that the underlying process of railroad grain carloadings is stationary. Also, the partial autocorrelation functions are significantly below zero at \( k = 4 \), and insignificant when \( k > 4 \). This suggests that \( \{y_t\} \) is likely to be an AR(4).

In the estimation step we use the Maximum Likelihood (ML) estimation procedure. The initial estimation of AR(4) for the deseasonalized series \( y_t \) indicates that the term \( y_{t-2} \) is not statistically significant, therefore, we exclude it from the model. Examination of the quarterly railroad grain carloading data indicated that the time series patterns prior to and after 1987:4 are significantly different (see Fig. 1) indicating that a structural change in carloadings may have occurred in 1987:4. Thus the model is estimated for the 1987:4-1997:4 period. The estimation result with the exclusion of \( y_{t-2} \) is:

\[ y_t = -9018.4 + 0.48275y_{t-1} + 0.19774y_{t-3} - 0.57888y_{t-4} + e_t. \]  

(9)

To check the goodness-of-fit of the model, the hypothesis that the residuals \( e_t (t = 1, 2, \ldots, T) \) are white noise \( (E(e_t) = 0, V(e_t) = \sigma^2, \text{and } E(e_t e_{t+k}) = 0, \text{where } j \neq k) \) is tested by chi-square statistics which are computed using the Ljung–Box formula (Lyung and Box, 1978).
Table 2
Comparison of actual to forecast quarterly grain carloadings 1988:4–1997:4

<table>
<thead>
<tr>
<th>(1) Year/Quarter</th>
<th>(2) Actual carloadings</th>
<th>(3) Forecast carloadings</th>
<th>(4) (3)/(2)−1×100, % difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988:4</td>
<td>376,627</td>
<td>394,789</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>387,652</td>
<td>406,086</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>362,687</td>
<td>397,047</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>330,292</td>
<td>336,804</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>392,918</td>
<td>357,461</td>
<td>−9.0</td>
</tr>
<tr>
<td>Total</td>
<td>1,473,549</td>
<td>1,497,398</td>
<td>1.6</td>
</tr>
<tr>
<td>1989:1</td>
<td>406,216</td>
<td>397,395</td>
<td>−2.2</td>
</tr>
<tr>
<td>:2</td>
<td>356,006</td>
<td>389,063</td>
<td>9.3</td>
</tr>
<tr>
<td>:3</td>
<td>341,556</td>
<td>345,545</td>
<td>1.2</td>
</tr>
<tr>
<td>:4</td>
<td>375,984</td>
<td>384,494</td>
<td>2.3</td>
</tr>
<tr>
<td>Total</td>
<td>1,479,762</td>
<td>1,516,497</td>
<td>2.5</td>
</tr>
<tr>
<td>1990:1</td>
<td>348,679</td>
<td>377,872</td>
<td>8.4</td>
</tr>
<tr>
<td>:2</td>
<td>301,851</td>
<td>326,223</td>
<td>8.1</td>
</tr>
<tr>
<td>:3</td>
<td>344,326</td>
<td>297,442</td>
<td>−13.6</td>
</tr>
<tr>
<td>:4</td>
<td>381,273</td>
<td>367,645</td>
<td>−3.6</td>
</tr>
<tr>
<td>Total</td>
<td>1,376,129</td>
<td>1,369,182</td>
<td>−0.5</td>
</tr>
<tr>
<td>1991:1</td>
<td>378,567</td>
<td>365,729</td>
<td>−3.4</td>
</tr>
<tr>
<td>:2</td>
<td>315,232</td>
<td>340,074</td>
<td>7.9</td>
</tr>
<tr>
<td>:3</td>
<td>357,920</td>
<td>342,126</td>
<td>−4.4</td>
</tr>
<tr>
<td>:4</td>
<td>405,122</td>
<td>382,582</td>
<td>−5.6</td>
</tr>
<tr>
<td>Total</td>
<td>1,456,841</td>
<td>1,430,511</td>
<td>−1.8</td>
</tr>
<tr>
<td>1992:1</td>
<td>386,186</td>
<td>367,323</td>
<td>−4.9</td>
</tr>
<tr>
<td>:2</td>
<td>338,835</td>
<td>305,750</td>
<td>−9.8</td>
</tr>
<tr>
<td>:3</td>
<td>339,935</td>
<td>358,059</td>
<td>5.3</td>
</tr>
<tr>
<td>:4</td>
<td>363,695</td>
<td>376,039</td>
<td>3.4</td>
</tr>
<tr>
<td>Total</td>
<td>1,428,651</td>
<td>1,407,171</td>
<td>−1.5</td>
</tr>
<tr>
<td>1993:1</td>
<td>328,204</td>
<td>358,342</td>
<td>9.2</td>
</tr>
<tr>
<td>:2</td>
<td>294,940</td>
<td>285,523</td>
<td>−3.2</td>
</tr>
<tr>
<td>:3</td>
<td>329,030</td>
<td>312,862</td>
<td>−4.9</td>
</tr>
<tr>
<td>:4</td>
<td>384,743</td>
<td>362,844</td>
<td>−5.7</td>
</tr>
<tr>
<td>Total</td>
<td>1,336,917</td>
<td>1,319,571</td>
<td>−1.3</td>
</tr>
<tr>
<td>1994:1</td>
<td>371,083</td>
<td>355,148</td>
<td>−4.3</td>
</tr>
<tr>
<td>:2</td>
<td>352,852</td>
<td>330,791</td>
<td>−6.3</td>
</tr>
<tr>
<td>:3</td>
<td>388,575</td>
<td>359,360</td>
<td>−7.5</td>
</tr>
<tr>
<td>:4</td>
<td>373,223</td>
<td>401,681</td>
<td>7.6</td>
</tr>
<tr>
<td>Total</td>
<td>1,485,733</td>
<td>1,446,980</td>
<td>−2.6</td>
</tr>
<tr>
<td>1995:1</td>
<td>378,734</td>
<td>344,049</td>
<td>−9.2</td>
</tr>
<tr>
<td>:2</td>
<td>328,319</td>
<td>326,694</td>
<td>−0.5</td>
</tr>
<tr>
<td>:3</td>
<td>267,043</td>
<td>331,882</td>
<td>24.3</td>
</tr>
<tr>
<td>:4</td>
<td>333,555</td>
<td>314,633</td>
<td>−5.7</td>
</tr>
<tr>
<td>Total</td>
<td>1,307,651</td>
<td>1,317,258</td>
<td>0.7</td>
</tr>
<tr>
<td>1996:1</td>
<td>330,205</td>
<td>342,202</td>
<td>3.6</td>
</tr>
<tr>
<td>:2</td>
<td>279,919</td>
<td>286,960</td>
<td>2.5</td>
</tr>
<tr>
<td>:3</td>
<td>280,292</td>
<td>298,084</td>
<td>6.3</td>
</tr>
<tr>
<td>:4</td>
<td>313,381</td>
<td>345,216</td>
<td>10.2</td>
</tr>
<tr>
<td>Total</td>
<td>1,203,797</td>
<td>1,272,462</td>
<td>5.7</td>
</tr>
</tbody>
</table>

\[ \chi^2_m = n(n + 2) \sum_{k=1}^{m} \frac{r_k^2}{(n-k)} \]

where, \[ r_k = \frac{\sum_{t=1}^{n-k} e_t e_{t+k}}{\sum_{t=1}^{n} e_t^2} \]

For \( m = 6, 12, 18, 24 \), none of the \( \chi^2_m \)'s are statistically significant indicating that there is no evidence that the residuals are not white noise.

With regard to diagnostic checking, we estimated many ‘nearby’ models including AR(3), AR(5), AR(6), and ARMA(4,1) and compared them with AR(4). The AR(4) model has the lowest AIC and SBC and is selected as the model for quarterly railroad grain loadings.

To obtain the forecasting equation for quarterly railroad grain loadings we substitute the difference equation, \( y_t = x_t - x_{t-4} \), into Eq. (9) which yields:

\[ x_t = x_{t-4} - 9018.4 + 0.48275(x_{t-1} - x_{t-5}) + 0.19774(x_{t-3} - x_{t-7}) - 0.57888(x_{t-4} - x_{t-8}) + e_t. \]

The one-period-ahead forecasting equation, therefore, can be expressed as

\[ \hat{x}_t = X_{t-4} - 9018.4 + 0.48275(X_{t-1} - X_{t-5}) + 0.19774(X_{t-3} - X_{t-7}) - 0.57888(X_{t-4} - X_{t-8}) \]

Table 2 compares actual railroad grain loadings for the 1988:4 to 1997:4 period to the forecast loadings obtained using Eq. (12). Of the 37 quarterly observations in Table 2, the percentage difference between the actual and forecast value is 5% or less for 17 (46%) of the 37 quarters. The percentage difference between the actual and forecast value is 5.1 to 10% for 17 (46%) quarters. Thus for 92% of the 37 quarters the percentage difference between the actual and forecast values is 10% or less. Of the 9 annual observations in Table 2, the percent difference between the actual and forecast value is less than 2.6% for 8 of the 9 years. The lone exception is 1997 which has a percentage difference of 5.7%. Thus while the deviation of the forecast from the actual grain loadings may be relatively large for some quarters, on an annual basis Eq. (11) does a good job of estimating railroad grain loadings.

4. Conclusion

Quarterly railroad grain loadings are dynamically unstable and thus inherently difficult to forecast. Grain carloads depend on the demand and supply of railroad grain transportation. However, the data for several variables that have a theoretical relationship to the demand and supply of railroad grain transportation service is not published quarterly, obviously hindering the
ability to forecast quarterly railroad grain carloadings. As shown in this paper, potential explanatory variables that are published in quarterly frequency have a relatively low correlation to railroad grain carloadings and in some cases the sign of the correlation coefficients is not in accordance with theoretical expectations. This latter result indicates that the economic process generating quarterly railroad grain carloadings is quite complex. Given this problem and the focus on short term forecasting we developed a time series model to forecast quarterly railroad grain carloadings.

Time series models may forecast well as long as the underlying economic process generating the forecast does not change. In contrast, structural change in the process will generate large forecast errors. As noted above, a structural change in the process generating quarterly railroad grain carloadings may have occurred in 1987:4 so we estimated the time series model for the 1987:4–1997:4 period. A structural shift may also have occurred in 1997 since the 1997 grain carloadings of 1,203,797 is 8% less than the 1996 level and 19% less than 1995 carloadings. The well publicized congestion problems of the western railroads may have contributed to the abnormally low grain carloadings in 1997. Of course these types of events cannot be forecast with a time series model. Thus while time series model forecasts can be useful, firms in the grain logistics system should use them along with other information in making their business decisions.

Acknowledgements

The research on which this paper is based was sponsored by the US Department of Agriculture, Agricultural Marketing Service, Transportation and Marketing Division. The authors are solely responsible for the accuracy of the facts and analysis presented in the paper.

References

Association of American Railroads, various annual issues. Railroad Facts, Washington, DC.