ISSN 1648-4142 TRANSPORT
http:/www.vtu.lt/english/editions

TRANSPORT – 2003, Vol XVIII, No 6, 255–258

FORECASTING OF RAILWAY FREIGHT VOLUME: APPROACH OF ESTONIAN RAILWAY TO ARISE EFFICIENCY

Ülo Hunt

Baltic Branch of International Academy of Noosphere, Raadiku 13-75, Tallinn, Estonia. E-mail: yloetekp@zzz.ee

Received 2003 06 06; accepted 2003 11 14

Abstract. The local railroad was substantially deregulated and prepared for privatization by Estonian government in 90's. Changes in commodity mix, length of haul, shipment size, loading weight, equipment ownership, railroad costs, competition from other modes, and demand for railroad transportation have all played a significant role. This paper introduces the importance of forecasting a railway freight volume using the methodology of international knowledge.

Keywords: railroad, forecast, ARIMA, model.

1. Introduction

The Republic of Estonia has set the development of an integrated and competitive transport system as one of its goals. Taking into account the favourable position of Estonia pertaining to economic geography, the guideline of transport policy is the promotion of international transit traffic. Transport traders contribute more than 10 % of the GNP of Estonia.

Estonia is one of the shortest and cheapest transit corridors between the CIS countries and Western Europe. Together with harbours, the railway represents the main link for transit through Estonia. Promotion of international railway trade is one of the most important priorities of Estonian economy. Of the total quantity of freight passing through Estonia, 42,1 million tons or more than 95 % were conveyed by railway in 2002.

Estonia is obliged to ensure smooth transit transport of goods through its territory on the level of internationally recognised service standards. In accordance with the growth rate economy of the neighbouring regions, the activity of the customers using the railway service is developing and becoming more active, as well.

2. Efficiency and Competition

More than ever Estonian Railways Ltd. has been able to acknowledge its role upon efficient servicing of customers, being primarily oriented to international transport market. Price formation and long-term customer contracts serve as the basis for the stable customer policy.

For the Baltic railways the competition between transport modes in freight market is not the issue of the first importance. The hard competition exists between the railway companies in the freight transit market.

As far as the Estonian Railway is concerned, we are the most competitive owing to the shortest leg and the lowest prices. However, new market conditions require using of new approach.

3. Methodology

Time series models were used for obtaining short-run forecasts of railway fertilizer and timber traffic. The multiplicative seasonal autoregression integrated moving average (***ARIMA***) model is among these models.

The time series considered in this article are the sequences of observations observed at equally spaced intervals. The time series can be considered as having three basic components:

1. *Trend*
2. *Cyclic or periodic*
3. *Random or noise*

The trend represents a long term pattern, e.g., increasing sales, while the cyclic or periodic components follow the patterns such as 'summer sales are higher than winter sales' or 'there are more enquires on a Monday morning than on a Friday afternoon'. The random component is the fluctuation around the trend and cycles. Unlike regression models, the noise in time series models is not assumed to be independent but is (auto) correlated. Furthermore, the noise is usually assumed to have a zero mean and to be stationary. Being stationary means that the statistical relationship between the observations at time t and time $t+l$ is the same as the statistical relationship between the observations at time $t+t$ and $t+l+\tau$. The relationship can be described in the terms of either their joint distribution or in the terms of correlation. The theoretical correlation between the observation at time t and the observation at time $t+l$, that is at lag l is denoted by ρl.

Statistical models for time series relate the correlated noise (denoted by wt) to an independent random component or white noise (denoted by et), and hence explain the autocorrelation rl. The most common models for stationary time series are autoregressive models and moving average models. An auto-regressive model for a series wt of order p is:

$$w_t = \Phi_1 w_{t-1} + \Phi_2 w_{t-2} + \ldots + \Phi_p w_{t-p} + \varepsilon t,$$

which is denoted by AR (ρ). A moving average model of order q is:

$$w_2 = \varepsilon t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q},$$

which is denoted by MA(q). The difference between the two forms of the model is that the autocorrelation for an autoregressive model gradually dies away while the autocorrelation for a moving average model will cut off at lag q.

In the situation where the data has cyclic or periodic nature, for example days of the week, a seasonal model may be appropriate. For example, Tuesday's observation may be related to both Monday's observation and the previous Tuesday's observation. In seasonal models the terms are related to the seasonality, s, e.g., for days of the week $s = 7$. The seasonal autoregressive and moving average models are:

$$w_t = \Phi_1 w_{t-s} + \Phi_2 w_{t-2s} + \ldots + \Phi_p w_{t-ps} + \varepsilon t$$

and

$$w_2 = \varepsilon_t - \theta_1 \varepsilon_{t-s} - \theta_2 \varepsilon_{t-2s} - \ldots - \theta_Q \varepsilon_{t-Q}.$$

These models can be combined to give a seasonal ***ARMA*** model

$$w_t = \Phi_1 w_{t-s} + \Phi_2 w_{t-2s} + \ldots + \Phi_P w_{t-Ps} + e_t -$$

$$\Theta_1 e_{t-s} - \Theta_2 e_{t-2s} - \ldots - \Theta_Q e_{t-Qs},$$

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \ldots + \phi_p e_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} -$$

$$\theta_2 \varepsilon_{t-2} - \theta_q \varepsilon_{t-q},$$

where e_t is an intermediate series for the purposes of defining the model.

In the situation where the series y_t is not stationary it may be possible to make it stationary by applying differencing or seasonal differencing. First-order differencing is given by

$$\nabla y_t = y_t - y_{t-1};$$

and first-order seasonal differencing is given by

$$\nabla_s y_t = y_t - y_{t-1}.$$

Differencing can be combined with a constant term to give

$$\nabla_d \nabla_{sD} y_t = c + w_t.$$

Combining this with the ***ARMA*** model given above gives a seasonal ***ARIMA*** (AutoRegressive Integrated Moving Average) model which can be specified by seven values (p, d, q, P, D, Q, s).

Time Series Model Identification

The basic tools in the identification of a suitable time series model are data plots, the autocorrelation function (**acf**) and the partial autocorrelation function (**pacf**). Plotting the data will indicate if the series is stationary. If not, the series can be transformed using standard functions such as LOG or SQRT and differencing can be applied using function TIME_SERIES_DIFF. Given a stationary time series, the **acf** and **pacf** can be calculated using the functions ACF and PACF respectively. By looking at **acf** the order, q of a possible MA model can be identified, while the **pacf** for an **AR** (p) model will cut off at lag p.

The function PACF also gives approximate estimates of the parameters of the autoregressive model of order given by the number of partial autocorrelations requested and the predictor error variance ratio which is

$$V_l = \mathrm{var}(\varepsilon_{l,t}) / \mathrm{var}(w_t),$$

where $\mathrm{var}(w_t)$ is the variance of the stationary series and $\mathrm{var}(\varepsilon_{l,t})$ is the estimated variance of the white noise when **AR** (l) has been fitted.

If neither approach is suitable then ARMA models can be considered. The function *ARIMA_APPROX_FIT* gives an approximate fit for an ***ARIMA*** model. This may be useful in model identification or for giving a set of initial values for the full fitting process.

Time Series Model Fitting and Forecasting

The function ARIMA_FIT fits an ARIMA model using either maximum likelihood or the least squares. The model is specified by the orders p, d, q, P, D, Q and s. The function returns the parameter estimates, standard errors, t-values and significance. The results can be input into ARIMA_FORECAST to forecast values for the series.

Transfer Function Models

The ARIMA model can be used to describe a single series; however, if the (output) series can be related to several input series then a multi-input or transfer function model can be used. The form of a transfer function model is

$$y_t = z_{1,t} + z_{2,t} + \ldots + z_{m,t} + n_t,$$

where n_t follows an ARIMA model and the components of the model, $z_{1,t}$, are related to the input variables, $x_{i,t}$, by either simple linear model,

$$z_{i,t} = \omega_0 x_{i,t}$$

or *ARMA*-like model

$$z_{i,t} = \delta_1 z_{i,t-1} + \delta_2 z_{i,t-2} + \ldots + \delta_p z_{i,t-p} + \omega_0 x_{i,t-b} +$$

$$\omega_1 x_{i,t-b} - 1 + \ldots + \omega_q x_{i,t-b-q} \,,$$

where b is known as the delay. A transfer function model can be fitted using the options of ARIMA_FIT, and forecasts can be made using TRNS_FUNC_FORECAST. To forecast the output series of a transfer function model the forecasts of the input series has to be provided. Also, if the input series have been forecast using ARIMA models, these models may be supplied to TRNS_FUNC_FORECAST; this will not affect the forecast, but will adjust the forecast standard error.

The major concern here is that the residuals are systematically distributed across the series (e.g., they could be negative in the first part of the series and approach a zero in the second part) or that they contain some serial dependency which may suggest that the ARIMA model is inadequate. The analysis of ARIMA residuals constitutes an important test of the model. The estimation procedure assumes that the residuals are not (auto-) correlated and that they are normally distributed.

The ARIMA method is appropriate only for a time series that is stationary (i.e., its mean, variance, and autocorrelation should be approximately constant through time) and it is recommended when there are at least 50 observations in the input data. It is also assumed that the values of the estimated parameters are constant throughout the series.

The showed methodology united with other approaches of a statistical analysis will give the establishment for the calculation of several estimation data for "describing better future".

4. Model, Data and Results

We try to use next statistical data for short-term forecasting of fertilizers and timber traffic. Data til 2002-12 is actual and the rest is forecast (Table 1).

We use for forecasting the next ARIMA model values (Table 2).

The working model gives us the following estimation parameters (Table 3) for a forecasting model.

The result of a forecasting model is shown on the following charts. It is clear, that ARIMA seasonal model gives better results, than a traditional statistical trend model.

Forecasts standard errors are shown in the Table 4 below.

The model developed in this study had good forecasting accuracy on a month basis. Comparing this estimation with the real data for February 2003 we have got the difference for timber –12,3 % and for fertilizer +10,4 %.

Table 1. Data of forecasting and timber

Transportation seasonal freight on Estonlan Rallways

Time	Fertilizers	Timber	Time	Fertilizers	Timber
1996-1	110.773	56.829	2000-1	206.976	81.025
1996-2	79.516	55.190	2000-2	257.892	113.821
1996-3	116.655	63.306	2000-3	152.249	144.178
1996-4	70.355	48.034	2000-4	134.179	123.971
1996-5	21.597	45.347	2000-5	82.641	97.051
1996-6	33.646	46.725	2000-6	126.621	70.609
1996-7	13.340	46.469	2000-7	179.958	87.571
1996-8	18.533	41.720	2000-8	121.002	62.170
1996-9	37.497	35.758	2000-9	116.674	55.094
1996-10	34.030	41.526	2000-10	66.623	63.199
1996-11	90.049	37.817	2000-11	100.540	70.724
1996-12	81.079	39.675	2000-12	197.236	76.090
1997-1	73.388	45.147	2001-1	200.298	100.034
1997-2	120.317	62.896	2001-2	227.782	108.499
1997-3	101.578	84.336	2001-3	204.607	139.737
1997-4	68.467	87.533	2001-4	175.703	125.736
1997-5	69.592	77.406	2001-5	156.753	93.221
1997-6	32.024	76.079	2001-6	119.882	67.455
1997-7	14.432	73.331	2001-7	155.738	49.414
1997-8	31.969	72.150	2001-8	146.721	56.230
1997-9	31.500	69.454	2001-9	205.981	62.212
1997-10	55.069	76.583	2001-10	141.635	66.800
1997-11	104.699	62.778	2001-11	229.258	60.029
1997-12	97.708	78.441	2001-12	300.383	79.266
1998-1	126.820	72.826	2002-1	412.332	87.584
1998-2	124.391	84.850	2002-2	286.723	88.534
1998-3	116.367	108.960	2002-3	260.961	112.219
1998-4	99.643	119.443	2002-4	226.732	123.901
1998-5	43.655	101.105	2002-5	198.674	91.871
1998-6	26.148	82.643	2002-6	220.391	69.994
1998-7	18.489	64.182	2002-7	132.656	58.213
1998-8	43.196	70.950	2002-8	174.170	51.235
1998-9	54.566	59.959	2002-9	179.126	70.875
1998-10	53.338	64.156	2002-10	210.634	73.678
1998-11	100.021	65.854	2002-11	182.526	86.659
1998-12	179.587	90.311	2002-12	248.614	90.119
1999-1	151.344	82.329	2003-1	323.870	101.733
1999-2	137.147	100.355	2003-2	292.061	103.294
1999-3	194.185	149.535	2003-3	271.401	128.171
1999-4	131.628	148.444	2003-4	241.480	132.850
1999-5	88.233	112.694	2003-5	219.042	100.293
1999-6	27.214	81.057	2003-6	208.484	77.139
1999-7	83.284	70.909	2003-7	189.353	63.570
1999-8	80.422	64.345	2003-8	202.885	59.879
1999-9	112.214	72.931	2003-9	237.959	75.983
1999-10	143.779	73.444	2003-10	216.334	79.145

Table 2. Values of ARIMA model

ARIMA Model		
Number of autoregressive terms	p	2
Order of non-seasonal differencing	d	0
Number of moving average terms	q	1
Number of seasonal autoregressive terms	P	1
Order of seasonal differencing	D	1
Number of seasonal moving average terms	Q	0
The seasonality	s	12

Table 3. Results of simulation

Fertilizers				
Parameter Estimates				
Parameter	Estimate	S.D.	t value	Sig.
phi_1	-0,09499731	0,39685723	-0,23937403	0,81154633
phi_2	0,19079644	0,16059548	1,18805611	0,23900578
theta_1	-0,36692407	0,38447241	-0,95435734	0,34333345
Phi_1	-0,55435912	0,13235746	-4,18834825	8,392E-05
Constant	28277,7675	4985,4818	5,67202301	3,2656E-07
Timber				
Parameter Estimates				
Parameter	Estimate	S.D.	t value	Sig.
phi_1	0,44779236	0,35085477	1,27628978	0,20625955
phi_2	0,21066067	0,26277443	0,80167875	0,42557299
theta_1	-0,23326823	0,33071097	-0,70535377	0,48303525
Phi_1	-0,25010224	0,1459259	-1,7138989	0,09117059
Constant	5554,06845	4451,05719	1,24780883	0,2164446

However, because the residuals are assumed to be normally distributed, ARIMA models don't handle time series with irregular, large amplitude bursts very well.

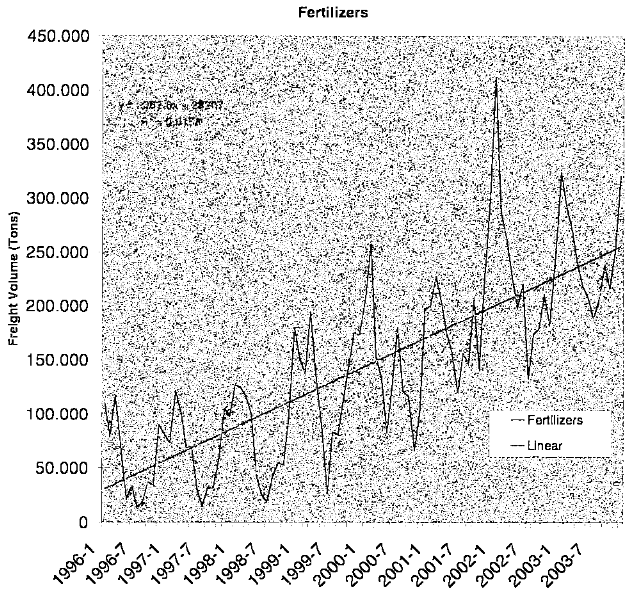

Fig 1. Forecasting fertilizers transport by rail

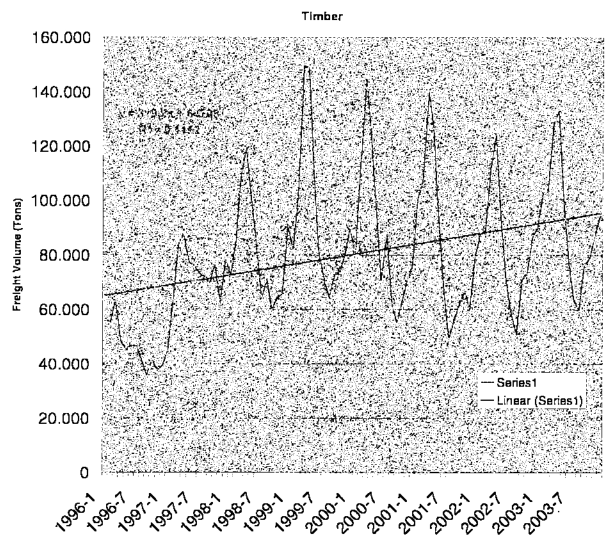

Fig 2. Forecasting timber transport by rail

Table 4. Forecasts standart errors

|Standard Errors of Forecasts(12)|		|Standard Errors of Forecasts(12)|
40.821,7194		12.754,9670
42.304,0691		15.432,1581
42.836,6972		16.775,1934
42.862,1949		17.441,5726
42.877,4701		17.793,9644
42.877,8204		17.980,5814
42.878,2957		18.080,6183
42.878,2979		18.134,3394
42.878,3141		18.163,2673
42.878,3141		18.178,8562
42.878,3147		18.187,2627
42.878,3147		18.191,7971

Another area for further research is forecasting of other types of railway traffic such as petroleum products, coal, grain and others. It would be interesting to observe the degree of temporal instability in the movement of these commodities. Finding of suitable models is necessary for railway traffic forecasts.

5. Summary

The tendencies shown at the last time in the development of the world economy, including the transport sector, witness their globalisation. The specialisation of different continents and regions in production is asking for more efforts for resolving the bottleneck issues in the world economy. New methods of analysis are welcome.

And the big business will set the requirements for the cargo to move. In the vast territories of Eurasia, where railways play the most important part in passenger and freight transport, the drawing of technically different railway networks closer to each other certain potentials for the development of connections between West and East. Estonian Railway is open to start partnership with all freight forwarders to increase its competitiveness on the market.

Literature

1. Aladjev, V. Z.; Veetõusme, R. A.; Hunt, Ü. J. General Theory of Statistics. Tallinn: TRG & Salcombe Eesti Ltd., 1995. 201 p. (in Russian with English summary).
2. Aladjev, V. Z.; Hunt, Ü. J.; Shishakov, M .L. Course of the General Theory of Statistics. Ed. Acad. A. D. Ursul. Gomel: BELGUT Press, 1995. 201 p.
3. Hunt, Ü. J.; Shishakov, M. L; et al. Probability Theory and Mathematical Statistics. Ed. Acad. V. Z. Aladjev. Gomel: TRG, 1997. 180 p.
4. Hunt, Ü. J. Some Methods of Calculation of Carrying Capacity of the Rail-ways of thc Baltic Region. In: Proc. Intern. Conf. TRANSBALTICA-99, April 1999. Vilnius, p. 392–398 (in Russian with English summary).
5. Hunt, Ü. J. Usage of new tariff model: Approach of Estonian Railway to arise efficiency. In: Proc. Intern. Conf. TRANSBALTICA-2002, April 2002, Vilnius, p. 214-220.
6. Aladjev, V. Z.; Hunt, Ü. J. A Workstation for mathematicians. In: Proc. of Conf. Improvement of Control Mechanism, April 1999, Grodno, p. 95–99 (in Russian with English summary).
7. Aladjev, V. Z.; Hunt, Ü. J. A Workstation for mathematicians. In: Conf. TRANSBALTICA-99, April 1999, VTU, Vilnius, p. 392–395 (in Russian with extended English summary).
8. Michael, W.; Babcock, Xiaohua Lu. Forecasting inland waterway grain traffic. In: Transportation Research. Part E: Logistics and Transportation Review. V. 38, 2002, p. 65–74.
9. Statistical Add-Ins for Excel, User Guide. Second Edition. The Numerical Algorithms Group Limited, 2000.
10. Lucio Pompeo, Ted Sapountzis. Freight expetations. The McKinsey Quarterly, 2002, No 2.
11. Electronic Statistics Textbook. Tulsa, OK, StatSoft, Inc., 2002.