REGRESSION MODEL OF SALES VOLUME

FROM WHOLESALE WAREHOUSE

Diana Santalova
Faculty of Transport and Mechanical Engineering
Riga Technical University
1 Kalku Str., LV-1658, Riga, Latvia
E-mail: Diana.Santalova@rtu.lv

KEYWORDS
Sales forecast, stock inventory, regression model

ABSTRACT

The multiple regression model is offered for short-term forecasting of change in sales volumes from a wholesale warehouse. Testing of the model correctness was performed by means of Statistica 6.0 on the base of the company UNIFEX Ltd real sales data for two ranges of goods.
The estimation of change in sales volumes obtained by means of the offered model may be used by logistic and distribution specialists for further calculations, for example, for determination of volumes and frequency of orders for delivery of separate goods or a group of goods.

INTRODUCTION

As it is known, the logistic chain is rather complicated – starting with the research of demand pattern and ending by supply of the stock to the consumer with minimum total expenses (Kogan and Spiegel 2006, Mackay and Probert 2002).
The problem that very often comes up at logistic enterprises is precise determination of optimal volume of the ordered goods in one or some positions as well as orders frequency. Let us pick out the following task – optimal batch size determination for replenishment of the current stock so that the stock runs out or reaches some minimum (critical) level at the given moment of time. To solve this problem one needs to know the sales demand in the future, thus, the problem of demand forecast for a period of time arises. As is known, there is a lot of ways of forecasting of sales volumes and/or demand (Bayhan G. and Bayhan M 1998, Luxhoj, Riis and Stensballe 1996, Rao 1985, Profillidis 2000). There is one way of demand forecasting being considered in the given article – short-term forecasting using multiple factor regression model.
Let us consider a typical wholesale enterprise – UNIFEX Ltd that deals with purchasing of big batches of food- and non-provision goods from enterprises - producers and delivery of these goods to the shops (enterprises - customers). At UNIFEX Ltd sales volumes of each goods are recorded every week during the whole year. Actually it is the quantity of the goods delivered to all shops weekly.
For forecasting with the help of the regression model one must determine the set of independent quantitative factors (variables) supposedly effecting on sales quantity. It should be noted that a producer of a brand as usual deals with the long-term demand forecasting whereas retail vendors (shops) take care of the short-term (operative) forecast. In these two variants of forecasting the factors effecting on the demand are different; let us agree not to consider them.
Wholesale enterprises, such as UNIFEX Ltd, are the link between a producer and a consumer. In the database of UNIFEX there is the following information, supposedly effecting on sales quantity:

1. Purchasing price of the goods and its alterations;
2. Purchase volumes every week;
3. Total volume of goods dispatched to all shops every week;
3. To what enterprises-consumers and at what prices the goods were delivered;
4. Whether discounts and at what rate were granted for the given goods within the past period of time under consideration;
5. Whether there was deficiency of the given goods in the warehouse and in what volume.

There is a variety of factors affecting sales quantity, but as a rule, it is very difficult to present them in quantitative form. For example, quality of the goods is important only for perishables (the so called sort). In this case one may reduce the price as the shelf life of the given goods expires. As to food goods with long shelf life, such as chocolate and pet fodder, a supplier as a rule suggests shops to organize actions when the shelf life of the given products comes up to the end. As usual advertising allowance and discount for ending up with excessive inventories in the warehouse assumes about 15% reduction of the usual price, whereas at the expiring shelf life discount may reach 30%.

As usual it is necessary for wholesale enterprises to determine the volume of ordered goods not more than for nearest couple of weeks. Therefore, the expected demand for the given goods one should know for the same period of time. We shall forecast the change of the demand quantity for the next week, exactly as for how many times more or less has got the demand of the goods of two groups compared to the last week: A – hygiene goods, 21 items, B – pet fodder, 22 items. As sale analysis for 2004 demonstrates, the demand for the goods of these both groups is about the same during one year.

So, we have information about the sales of goods of the two groups on a weekly basis during 2004. As the demand for the given groups of goods is about the same during a year, we will forecast the change of the demand quantity for each goods from the groups one week in advance, using the data for the last 3 months.

For the suggested model no specially organized structures of data are necessary – all data may be obtained from the enterprise database. The demand estimated with the help of the model further may be used for calculation of the optimal ordered goods batch. It is also important to note that a mere purchase manager as usual does not possess any knowledge in such spheres like inventory management theory and mathematical statistics as well as does not possess specific mathematical software. This is why for him/her it would be preferable to use the already given formulae for ordered batch optimal volume calculation, where he/she realizes them, for example, in Excel. Testing of the regression model soundness has been performed in the Statistica 6.0 package. The suggested approach will help enterprise workers who deal with distribution and logistic to solve the above indicated problem and will allow enterprise to reduce the expenses for purchasing and storing goods.

MULTIPLE REGRESSION MODEL FOR SALE VOLUME FORECAST

Let us consider the standard regression model (Draper and Smith 1998). The dependent variable $Y(t)$ is the proportion of the given goods demand quantity during the current week t to the demand quantity during the past week t-1. Let us pick out the factors effecting on the demand quantity change, basing on the information that we already have:

1. Ordinal week number.
2. Goods realization price.
3. Discount.
4. Stock condition.

The regression model will be written as follows:

$$Y(t) = \beta_0 t^{\beta_1} P^{\beta_2} A^{\beta_3} S^{\beta_4} e^{\varepsilon}, \qquad (1)$$

where t is ordinal week number, t = 1, 2, ... , whereas t =1 means the first week of the year;

$Y(t)$ is the dependent variable, demand quantity change in the week t concerning the previous week:

$$Y(t) = \frac{W(t)}{W(t-1)}, \quad t = 2, 3, \ldots ,$$

where $W(t)$ and $W(t\text{-}1)$ – demand in the t-th and (t-1)-th week, accordingly;

P – this variable may be presented in various ways:
1) planned average goods price in the t-th week;
2) difference of prices in the t-th and (t-1)-th weeks; variable equals 0 if in the end of the (t-1)-th week there has not been any change of the price.

In the given model the first way of determination of the variable P has been chosen.

A – variable is equal to the discount portion if during the t-th week any enterprise-consumer had any discount for the given goods; 1 – otherwise;

S – index of goods deficiency, demonstrates the stock of the given goods in the warehouse in the t-th week. This index is calculated on the base of data about arrival, depletion and stock of goods in the warehouse. The index demonstrates the stock of the given goods in the warehouse during a certain period of time after the necessary quantity of the goods has been loaded. It takes values from 0 to 1.0 – situation “out of stock”, stock of goods is equal to zero (or lower than the determined critical level), 1 – maximum possible determined stock of goods. The rest values from 0 to 1 are determined as the proportion of the current stock to the maximum possible one;

ε – odd component, having log-normal distribution.

So, in the regression model we have 4 independent variables: t, P, A, S, and 5 unknown coefficients: β_0, β_1, β_2, β_3, β_4. Our goal is to evaluate these coefficients and check their values at all data that we already have for two groups of the goods under consideration. It should be noted that in the mathematical model (1) goods belonging to one or another group does not get estimated. Nevertheless we keep the classification of the goods to the two groups, following the registration technology accepted at UNIFEX.

As it is obvious, the suggested model is a non-linear one (multiplicative). After having performed logarithms to the left and to the right, we obtain its following parameter-linear form:

$$\ln Y(t) = \ln\beta_0 + \beta_1 \ln t + \beta_2 \ln P + \beta_3 \ln A + \beta_4 \ln S + \varepsilon.$$

To this modified model one may apply standard estimation formulae (the least squares method).

ANALYSIS OF REGRESSION MODEL CORRECTNESS

There has been made a series of experiments to check the correctness of the suggested regression model. With this aim a mode *Multiple Regression* of Statistica 6.0 package has been used. The given procedure allows estimating parameters of parameter-linear regression

model, using the least squares method. The check has been made using the data of the sales of each out of 43 types of goods of the two groups during 2004.
To analize the correctness of the suggested model the following criteria was used (Srivastava 2002):

1. Fisher Criterion of the hypothesis check of regression insignificance, level of significance α=5%;
2. Student Criterion of the hypothesis check of the *k*-th independent variable insignificance, level of significance α=5%;
3. Multiple determination coefficient R^2.

Table 1: The Results of the Experiments for the Group A Goods

Goods №	Beta0	Beta1	Beta2	Beta3	Beta4	R^2
1	**4,4**	**-0,1522**	**-19,5**	**22,0**	**1,1**	**0.88**
2	**6,6**	0,0916	**-31,7**	**34,5**	0,6	**0.84**
3	**-5,9**	**-0,0818**	**-17,2**	**16,3**	**0,7**	**0.86**
4	**5,1**	-0,0933	**-20,8**	**20,5**	0,7	0.75
5	**18,0**	-0,0236	**-21,0**	**20,7**	0,2	**0.87**
6	**-3,4**	**-0,1212**	**-11,4**	4,2	**1,1**	**0.93**
7	**13,9**	0,0774	**-60,8**	**60,9**	**1,6**	0.72
8	**15,2**	0,0541	**-69,0**	**68,6**	**0,7**	**0.85**
9	**4,2**	0,2481	**-12,1**	5,3	**0,9**	0.72
10	**0,9**	-0,0876	**-14,2**	**10,5**	**1,6**	**0.93**
11	**0,8**	0,0404	**-20,2**	**19,4**	**1,7**	**0.87**
12	**-2,5**	0,1062	**-22,8**	**23,2**	**1,2**	**0.91**
13	**-3,1**	-0,0497	**-29,8**	**30,9**	**0,8**	**0.93**
14	**5,9**	0,0652	**-29,4**	**28,1**	0,6	0.77
15	**-2,4**	0,0161	**-21,1**	**20,8**	0,4	**0.88**
16	**-2,3**	0,0156	**-22,3**	**22,5**	0,8	**0.85**
17	**5,8**	0,0761	**-30,5**	**30,3**	0,7	0.76
18	**5,4**	0,0124	**-25,4**	**26,9**	**0,7**	**0.85**
19	**0,4**	-0,0995	**-9,1**	**8,3**	**0,8**	**0.85**
20	**-2,1**	-0,1654	**-19,2**	**18,6**	0,3	**0.88**
21	**8,9**	-0,0766	**-45,4**	**44,8**	0,7	**0.86**

Table 2: The Results of the Experiments for the Group B Goods

Goods №	Beta0	Beta1	Beta2	Beta3	Beta4	R^2
1	**55,6**	0,1215	**-26,7**	**26,1**	**0,4**	**0.88**
2	**90,1**	-0,0014	**-37,4**	**33,8**	0,2	**0.84**
3	**160,9**	0,1618	**-67,1**	**50,2**	0,3	**0.91**
4	**61,8**	-0,0744	**-24,6**	**42,2**	0,1	0.78
5	**65,7**	0,0000	**-29,7**	**19,4**	**0,8**	**0.91**
6	**79,7**	0,2065	**-38,4**	**40,7**	0,2	**0.84**
7	**79,9**	0,0668	**-38,4**	4,2	0,2	**0.85**
8	**70,0**	-0,0027	**-27,9**	**28,5**	-0,1	0.75
9	**78,1**	-0,0533	**-35,5**	**24,1**	-0,2	**0.88**
10	**154,6**	-0,1791	**-63,9**	15,4	0,2	0.77
11	**61,2**	-0,1656	-25,0	21,6	**1,3**	0.75
12	**23,4**	0,4181	-9,4	4,1	**2,6**	**0.85**
13	**12,9**	-0,0846	-6,4	3,9	0,7	0.73
14	**10,9**	0,0466	**-5,6**	10,5	**0,9**	**0.82**
15	**6,7**	0,0450	**-3,4**	**3,2**	**1,1**	**0.89**
16	**44,3**	-0,0018	**-20,8**	**20,7**	0,6	0.71
17	**45,1**	-0,0021	**-21,2**	**20,7**	0,3	0.74
18	**29,8**	-0,0300	-13,8	14,1	0,7	0.66
19	**22,3**	0,0505	**-10,5**	2,2	**0,7**	**0.89**
20	**35,4**	0,1507	**-16,8**	1,7	0,6	**0.87**
21	**57,4**	0,0485	**-27,2**	2,7	-0,1	0.78
22	**28,1**	-0,0072	**-15,2**	**14,7**	0,2	**0.80**

The results of the experiments (coefficient estimation values for variables and multiple determination coefficient R^2) for group A goods are shown in the Table 1, for group B goods the results are shown in the Table 2. The calculated values of Fisher Criterion for the goods of the both groups are shown in the Table 3.

Let us comment the obtained results. Firstly, let us note that the signs of the estimated coefficients, as a rule, fit the physical sense of the influence of the considered factors.

So, coefficients *Beta2* and *Beta3* (for the variables *P* and *A* accordingly) for all the sets of data have one and the same sign. It means that the chosen model behaves approximately in the same way with all the researched sets of data. The negative coefficient by the variable *P* that shows the price of the given goods shows that if the prices increase the sales volume decreases.

The positive coefficient by the variable *A* that shows the value of the discount for the given goods, shows that if the discount increases the sales volume increase too. Coefficient *Beta1* by the variable *t*, showing the ordinal week number, and coefficient *Beta4* by the variable *S*, showing the stock state, take positive as well as negative signs. The negative *Beta1* shows that gradually with time the given goods sales volume decreases, and the positive one demonstrates its increasing. The positive *Beta4* shows that the bigger the stock of the given goods in the warehouse, the bigger the given goods sales volume, that means, the demand for given goods is stable; the negative sign of it shows that there has not been any increase in sales volume not withstanding sufficient given goods stock in the warehouse – that means, there has been the so called „over stock", frozen residue – the present stock is much bigger than the present demand. The *Beta0* is the intercept of the presented regression model and is not discussed.

In bold face there are the coefficients of the most important independent variables (by Student criterion). As it has already been supposed, the most important factors affecting the demand changes have occurred to be the price of the given goods along with the given discount. Let us set the variables in the order of decreasing importance: price *P*, discount *A*, stock state *S*, week ordinal number *t*. Let us consider now the calculated values of Fisher criterion.

Table 3: The Calculated Values of Fisher Criterion

Goods №	The group A goods	The group B goods	Goods №	The group A goods	The group B goods
1	40,92	15,41	12	57,88	23,55
2	28,74	26,95	13	76,78	10,29
3	34,86	36,85	14	15,40	18,74
4	15,09	14,36	15	38,66	34,21
5	37,89	45,86	16	30,69	9,24
6	74,16	21,64	17	15,77	11,42
7	12,36	24,76	18	31,51	6,99
8	29,83	12,13	19	30,98	33,60
9	12,43	31,57	20	38,78	27,59
10	73,41	13,76	21	28,27	14,48
11	34,54	11,70	22	-	16,48

For the experiment of the group A goods data the theoretical value of Fisher criterion (by the freedom degree 4 and 47) equals 2.56, but for the experiment of the group B goods data the theoretical value of Fisher criterion (by the freedom degree 4 and 37) equals 2.62. It is obvious that for all experiments the value of the calculated Fisher criterion is higher than the theoretical one. This shows that the hypothesis of the regression insignificance is not accepted by the significance level 5%. Therefore this shows that our suggestions of the chosen factors' influence on the demand's change cannot be rejected.

Let us consider now the values of the multiple determination coefficient R^2. This coefficient characterizes changeability of the regression and shows what portion of the changeability of the dependent variable is determined by the changeability of the regression (Srivastava 2002). For adequate models its value should be not less than 0.8. For the group A goods in 16 experiments out of 21 this coefficient exceeds the level of 0.8. The situation with group B goods is worse – the coefficient exceeds the level of 0.8 only in 13 experiments out of 22.

The obtained models may be used for forecast of the change of demand value for each goods separately. Further on we shall consider models that are common for all goods of one group. Let us call these models group models.

GROUP MODELS' ANALYSIS

In this case the analyzed massive are the present data of sales of goods of each group. As a result we have the regression model equations for forecast of the change of demand for the both groups of goods.

For group A:

$$Y(t)=0.92t^{-0.05}\,P^{-0.04}\,A^{2.05}\,S^{1.64}. \quad (2)$$

For group B:

$$Y(t)=0.79t^{-0.03}\,P^{-0.02}A^{1.97}\,S^{1.57}. \quad (3)$$

It is apparent that for the both groups the variable coefficients have the same signs and are the values of one and the same magnitude. The variables *A* and *S* were declared by Student criterion to be the most important for both experiments. The values of Fisher criterion are 123.01 and 128.79 for the equations (2) and (3) accordingly. The theoretical value of Fisher criterion is 2.38 (as by freedom degree 4 and 715 for the first experiment, so by freedom degree 4 and 914 for the second experiment). It is obvious that also here the calculation values of Fisher criterion exceed the theoretical ones, therefore, the hypothesis of the regression insignificance also is not accepted at the level of significance 5%. The value of the multiple determination coefficient R^2 is equal to 0.64 and 0.60 for equations (2) and (3) accordingly, what is, as it was supposed, much worse than in the case of usage of an individual model.

Thus, with the use of the group model received forecast will be less precise than with the use of the individual model.

As a result the model was acknowledged to be adequate for all experiments and the most significant variables were selected.

CONCLUSIONS

In the future it would be expedient to include a variable into the model that shows the arrival of goods to the warehouse of the enterprise UNIFEX.

It should be noted that the opportunity to build forecast of the demand and to plan purchase volumes, using mathematical methods and software, in fact does not get used in any commercial or logistic enterprise of Latvia. At such enterprises the building of forecast is based on so-called manual approach of average intensity calculation notwithstanding that there are special subprograms within the known management systems (for example, MS Navision). So let us mark the originality of the suggested approach, because building demand forecast using the above considered regression model allows one to mark the tendency of the demand behavior more precisely. The obtained demand values estimations one may use for the further calculations – the volume of the ordered goods batch using, for example, *N*-staged dynamic model of stock inventory (Taha 2002). It is important that the calculations based on the suggested models may be done in Excel, which is extremely comfortable for distribution and logistic specialists.

ACKNOWLEDGEMENTS

I'm really grateful to Professor Alexander Andronov for his help in writing this paper.

REFERENCES

Bayhan G.M., Bayhan M. 1998. "*Forecasting using autocorrelated errors and multicollinear predictor variables*". Computers and Industrial Engineering. Vol.34, N 2, pp. 413 – 421.

Draper N.R., Smith H. 1998. "Applied Regression Analysis." Third edition. John Wiley & Sons, Inc., New York.

Kogan K., Spiegel U. 2006. "*Optimal policies for inventory usage, production and pricing of fashion goods over a selling season*". Journal of the Operation Research Society. Vol.57, N 3, pp. 304 – 315.

Luxhoj J., Riis J., Stensballe B. 1996. "*A hybrid econometric-neutral network modeling approach for sales forecasting*". International Journal of Production Economics. Vol.43. N 2, pp. 175 – 192.

Mackay R., Probert S. 2002. "*Forecasting the United Kingdom's supplies and demands for fluid fossil-fuels*". Fuel and Energy Abstracts. Vol.43, N 4, p. 243.

Profillidis V. 2000. "*Econometric and fuzzy models for the forecast of demand in the airport of Rhodes*". Journal of Air Transport Management. Vol.6, N 2, pp. 95 – 100.

Rao S-k. 1985. "*An Empirical Comparison of Sales Forecasting Models*". Journal of Product Innovation Management. Vol.2, N 4, pp. 232 – 242.

Srivastava M. S. 2002. "Methods of Multivariate Statistics." John Wiley & Sons, Inc., New York.

Taha H.A. 2002. "Operation Research. An introduction." Seventh edition. Prentice Hall, London.

AUTHOR's BIOGRAPHY

DIANA SANTALOVA was born in Riga, Latvia. She graduated from Riga Aviation University in 1998, having received a degree of the master of Computer Technologies. Since 1998 teaches such disciplines as Optimisation Methods, Computer Systems in Business, Information Management Systems, etc. at Riga Aviation University, and at Riga Technical University. Now the subject of her research work is: "Semi-parametric regression models for forecasting of passenger streams in the Baltic countries". Her e-mail address is Diana.Santalova@rtu.lv.