Use of regression analysis to forecast cash flows

W.G. Klerck and L.E. Birkholtz
School of Business Leadership, University of South Africa, Pretoria

In this article the need to forecast the cash flows of a corporation is discussed. Although many corporations operate on a decentralized basis to a greater or lesser degree the cash resources of the corporation are often controlled from head office. The system is cash receipts are remitted to the head office treasurer on a daily, weekly, or monthly basis and he undertakes to make funds available. Owing to the size of many corporations the treasurer is far removed from activities at the branch level and it is not unknown for him to experience an unexpected inflow or outflow of funds. It is also often taken for granted by the operations' accountants that the corporate treasurer will always have sufficient funds to place at their disposal. This lack of communication can result in difficulties in providing the operations with the required funds. It is therefore necessary that the treasurer forecast his future cash position in order to have adequate cash resources at his disposal to meet the requirements of the operations over the short term. In order to assist the treasurer in making short term forecasts a model using regression analysis was developed.

S. Afr. J. Bus. Mgmt. 1984, 15: 135 – 139

In dié artikel word die noodsaaklikheid om die kontant-behoeftes van 'n maatskappy te voorspel bespreek. Alhoewel baie maatskappye gedesentraliseerd is, word die kontant-behoeftes van die maatskappy vanuit die hoofkantoor behartig. Die stelsel wat gevolg word is om behoeftes aan kontant na die hoofkantoortesourier te verwys en hy voorsien dan die nodige fondse. As gevolg van die grootte van baie onderne-mings is die tesourier nie altyd bewus van al die aktiwiteite in afdelings nie en gevolglik ondervind hy dikwels 'n onverwagte vraag na fondse. Afdelings aanvaar ook dikwels dat die tesourier altyd genoeg fondse sal hê om in hul behoeftes te voorsien. Hierdie gebrek aan kommunikasie lei daartoe dat probleme ontstaan in die voorsiening van fondse. Dit is om dié rede nodig om die toekomstige kontantbehoeftes van die maatskappy te voorspel. Ten einde hierdie voorspelling te maak is 'n model wat gebruik maak van regressie-analise ont-wikkel. Die model word gebruik om die korttermyn-kontantbe-hoeftes van die maatskappy te voorspel.

S.-Afr. Tydskr. Bedryfsl. 1984, 15: 135 – 139

W.G. Klerck and L.E. Birkholtz
School of Business Leadership, University of South Africa,
P.O. Box 392, Pretoria 0001, Republic of South Africa

Introduction

Working capital, especially cash or overdraft facilities, form the basis around which a company's affairs rotate. Stocks are transformed into sales which become debtors from which cash is collected. The cash is used to pay the creditors for the goods originally bought as well as for all the other value-added costs. This cycle is constantly repeating itself within the business and the larger the corporation or conglomerate, the larger will be the amounts involved.

There is a very close interdependence between the various factors in the cycle, and cash in particular, or the lack of it, can seriously hamper the smooth operation and function of the business.

Many corporations operate on a decentralized basis, to a greater or lesser degree, with the result that nearly all the routine transactions are negotiated at the operational level. However, one often finds that the cash resources of the corporation are controlled from the centre. The system is that cash receipts are remitted to the head office treasurer on a daily, weekly, or monthly basis and he undertakes to make funds available for cheques drawn against the operation's bank account in the normal course of events. This results in substantial amounts of cash coming under his control.

The introduction of electronic aids into the banking system, including the use of computers for processing vast amounts of documentation within a short space of time and the use of electronic data transmission between centres, facilitates the centralizing of the treasurer's function to include operations in distant towns and cities as well as those near at hand.

The banks themselves have welcomed this in principle as it allows them to monitor more easily their total exposure to a corporation, and they have aided the treasurer's function, by consolidating the various accounts a corporation uses through automatic clearing of month-end balances into a single bank account.

The treasurer's dilemma

The treasurer of the average corporation is, by virtue of the size of the corporation, far removed from the activities at the operational and branch levels. It is thus not possible for him to know with reasonable accuracy what the corporate bank balance will be, or for him alone to predict the maximum or minimum bank account balance. It is not unknown for a trea-surer to experience an unexpected sudden inflow or outflow of funds because he was not timeously informed thereof.

It is frequently taken for granted by the operations' accoun-tants that the corporate treasurer will always have sufficient

funds to place at their disposal to meet whatever cheques they may issue in the conduct of their normal affairs. This lack of communication can result in difficulties in providing the operations with the required funds.

It becomes clear then that the treasurer has a need to forecast his future cash position in order to have adequate cash resources or overdraft facilities at his disposal to meet the requirements of the operations on the short-term daily basis.

The treasurer's role

The treasurer's role in an organization includes amongst others the duty to maximize the use of the corporate cash resources or alternatively to minimize the cost of borrowings. He does this by arranging short-term credits or deposit facilities with the bankers at the best possible interest rates. For instance, if he knows that he has a parcel of cash that will not be called upon for a number of weeks, he is then able to negotiate a slightly better interest rate with a banker, on condition he lays this money out for a fixed term.

Conversely, should he require a certain amount and is aware at the time that interest rates are 'hardening', he may consider borrowing short-term money at a fixed rate for a period, expecting that after this he will be able to redeem the loan from future expected cash inflows. It causes him a certain amount of embarrassment should he forecast a surplus of cash which he then lays out on fixed deposit, for say 15% p.a., only to find that he is short of funds and has to temporarily borrow the required amounts at a higher interest rate, of say 19% p.a. Effectively he makes a loss (quite unnecessarily) of 4% p.a. on these transactions.

The problem is exacerbated during times when interest rates are running at record levels, as currently being experienced in the business world.

It should be clear then that the treasurer who is able to forecast his cash resources or requirements accurately can minimize the cost of short-term borrowings. Whereas it may be possible only in exceptional circumstances to forecast future cash flows exactly, it is desirable to find a technique that will show greater accuracy than that demonstrated in Table 1.

The question of the corporation's standing with its bank

Table 1 Month-end cash balances (R000s)

	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
Forecast	1197	75	716	1649	142	2236
Actual	429	251	196	982	916	1810

or bankers also comes into play in that it is usual to negotiate loans for facilities up to a certain maximum level, and should this be exceeded, it causes embarrassment for the corporation, especially if such an overdraft is unexpected and the treasurer is not in a position to take prior steps to avoid the over-limit positions.

Thus it becomes clear that the treasurer needs to be able to forecast his future cash position as accurately as possible, in order to:

(a) Provide in advance for the cash requirements of his operations;

(b) maximize the resource, i.e. cash for which he is responsible or, conversely, to minimize the cost of borrowings; and

(c) avoid possible embarrassment and loss of standing with

his bankers.

Description of the company

The company upon which the forecasting model is based was formed at the beginning of 1980, as an Operating Division within a large South African company quoted on the Johannesburg Stock Exchange. It is one of twelve divisions within the corporation. Annual turnover of the corporation was in excess of R600 million for the 1981/82 financial year.

Each of the Operating Divisions is constituted with a Chairman, Managing Director, and Board of Directors who are responsible for the strategy as well as for the running of their particular divisions. The divisions in turn are responsible to the Corporate Board to whom they submit for approval their budgets, capital votes, and strategies on an annual basis. They report progress against these on a monthly basis. All the divisions within the corporation use the services of the Centralized Banking Service which is administered by the Corporate Treasurer.

The particular Division (or company) with which we are concerned is involved in the conversion of paper and plastic resin into sacks and bags of various constructions and sizes, for sale to industrial customers for their packaging requirements.

The division buys its raw materials from six main suppliers. It then converts this into packaging to suit individual customer's requirements. Very few stock lines are carried as the customers all specify different requirements for size, construction, printing, closure, etc. and the goods are only manufactured once an order has been received.

The customers this division deals with are all industrial concerns whose involvement is mainly in the foodstuffs sector (i.e. sugar, flour, maize meal, etc.) and in the dry chemical products sector (i.e. fertilizer, lime, cement, etc.).

Sales and marketing are conducted mainly on a national level, through a national marketing manager and staff, although sales to local customers are frequently negotiated by the local operation. There are very few cash sales and credit terms are 'payment within thirty days of statement'.

As there is a time lag between order and delivery by suppliers, raw materials stocks approximating two-months' supply are held to ensure continued production.

By arrangement with the suppliers, payment for stock is made thirty days after the month-end. This means that the company must fund the raw materials for an additional thirty days before they are committed to manufacture.

Manufacturing takes place in response to customers' orders and because of production-scheduling requirements, the finished goods are held in stock for ten days before delivery, at which point they are sold. At this transformation point, the finished goods have changed into debtors via the sales transaction.

On average it takes fifteen days from the date of sale to the date the statement is issued at the month-end. Although the credit terms to the customers are thirty days, experience shows that the customers pay their accounts 45 days from statement date — i.e. at day 130 of the cycle depicted above. This then completes the cycle and shows there is a 100-days cash cycle from the time cash is paid out for raw materials until it is received back from customers (day 130 minus day 30).

Forecasting cash flows

Forecasting of future cash flows is done at the operational level by the accountant or his assistant. The aim is to attempt to forecast as accurately as possible the net cash flow for the ensuing month, so that the Corporate Treasurer is in a position

to assess his immediate situation and to enable him to plan the best strategy for either utilizing surplus moneys or funding short-term deficits.

Sales

Sales form the link between the operation and the customer in the trading cycle. Once a sale has been concluded and the goods despatched, the operation acquires an additional debt for future cash inflow. It takes on average 60 days to receive the money after a sale has taken place. For this reason sales will be lagged by one, two, or three months.

Debtors (Cash inflow only)

In the transformation process, sales are translated into debtors. Debtors form part of the working capital of the operations and are constantly being turned over. As debtors are part of the operation's investment, cash inflow is increased with a reduction in the overall debtors' level increases. Debtors lag the cash flow by one month, in that an increase or decrease in debtors' levels over the month determines whether there has been an outflow or an inflow of cash.

Creditors (Cash outflow only)

This variable includes all the amounts due to suppliers of raw materials held mainly in stock. By negotiation with the major suppliers, these outstanding amounts must be settled by the 28th of the succeeding month, failing which the organization will usually be penalized by having to pay interest on the overdue balance. Because this variable is paid during the following month it is necessary to lag it by one month.

Trading days

Experience had indicated that the cash flow in those months with more trading days was normally higher than in those months with less trading days. For this reason the number of trading days was included as an independent variable in the model.

Time

It was believed that as time went by there might be an increase (or even decrease) in sales and therefore also a corresponding change in cash flows. For this reason time (month number) was included in the model.

Dummy variables

The nature of the business conducted by the operations shows a distinct seasonal trend. Business activity reaches a low point during the December/January period when customers usually close their plants and reduce their activities. From this point activity slowly improves till it reaches a peak in the July to November period. In anticipation of the increased business activity the operation usually increases its stock levels in advance of the peak season and then allows these levels to decline as the season goes through its peak. On the other hand debtors' levels increase during the peak and only reduce after the peak has been reached. During the slow-down period levels reduce accordingly, but stock levels are not allowed to fall below 'safe' levels. These seasonal trends were catered for by the introduction of dummy variables.

The data

Having identified the variables it was then necessary to collect the data for each of the variables. The data span a period of forty months, from March 1980 to August 1983. There are no data further back as the division for which the forecasting is done was only constituted during March 1980.

The last three observations were not used in selecting an appropriate regression model. These observations were used to test the performance of the model. Although three observations are not enough to test the model, more observations could not be used, owing to a lack of data.

Selecting an appropriate regression model

Since there is no 'theory' linking cash flows to a specific set of independent variables, the approach adopted was necessarily an empirical one, i.e. variables likely to influence cash flows were used to build various regression models and then to check the models for adequacy of fit.

The criterion of model-building in regression is parsimony, i.e. a model is developed which includes the fewest number of independent variables that permit an adequate interpretation of the cash flows. Multiple regression models that have fewer independent variables are easier to analyse and interpret.

In this study the most appropriate regression model was identified by evaluating all possible regression models. In the examination of all possible regression models two selection procedures were used:
(a) The R^2p criterion (coefficient of determination); and
(b) Mallows Cp,
where p refers to the number of independent variables considered.

The R^2p Criterion

The R^2p criterion involves the comparison of the coefficients of determination for all possible regression models. Since R^2p cannot decrease as additional independent variables are introduced, the highest R^2p value will occur when all possible independent variables are included in the model. The goal in utilizing R^2p is to compare alternative models so that it can be determined when the introduction of additional independent variables does not produce a commensurate increase in R^2p.

The Cp Criterion

The Cp criterion developed by Mallows (see Daniel & Wood, 1971) measures the total squared error of a regression model with p parameters. The total squared error consists of a bias component and a random error component. The bias component represents the difference in the predicted Y values obtained from the fitted regression model and the 'true' regression model. The random error component represents the variability around the fitted line of regression. When a regression equation with p independent variables does not contain any bias component, the average value of Cp is p, the number of parameters. The goal is therefore to identify those regression models for which Cp is close to p.

The model for cash inflow

The independent variables used in developing the model for cash inflow were:

$X1$	– sales lagged 1 month
$X2$	– sales lagged 2 months
$X3$	– number of tradding days
$X4$	– the time variable (month number)
$X5$	– debtors
$X6$ to $X16$	– dummy variables for each month of the year.

The first run on the computer revealed that observations

number 6, 9 and 24 were outliers i.e. the standardized residual for these observations exceeded 2. As outliers tend to have a significant effect on the regression results, these points were ignored in the development of a model.

In Table 2 the values for Mallows Cp and the coefficient of determination R^2p are shown for selected combinations of the independent variables.

Only those combinations for which Mallows Cp are the

Table 2 Cp and R^2p values for selected regression models

Independent variables included in regression model	Number of parameters p	R^2p	Cp
$X2$	2	0,578	94,18
$X2, X13$	3	0,683	65,37
$X1, X2, X13$	4	0,764	43,54
$X1, X2, X3, X11$	5	0,853	19,38
$X1, X2, X3, X11, X13$	6	0,898	8,12
$X1, X2, X3, X11, X12, X13,$	7	0,912	5,75
$X1, X2, X3, X11, X12, X13, X16$	8	0,920	5,44
$X1, X2, X3, X4, X11, X12, X13, X16$	9	0,927	5,59
$X1, X2, X3, X4, X9, X11, X12, X13, X16$	10	0,930	6,59
$X1, X2, X3, X4, X9, X11, X13, X14, X16$	11	0,932	8,14

smallest are shown. Because Cp is a measure of total squared error, an attempt was made to find the set of independent variables that leads to the smallest Cp value while at the same time minimizing the bias component (see Daniel & Wood). Since Cp is lowest for the $X1, X2, X3, X11, X12, X13, X16$ model ($C7 = 5,44$) and the $X1, X2, X3, X4, X11, X12, X13, X16$ model ($C8 = 5,59$) it would appear that one of these two models should be selected.

The $X1, X2, X3, X11, X12, X13, X16$ model was selected since its Cp value is slightly lower and it appears to contain less bias (in fact its Cp value is even less than p, the average total squared error when no bias exists). The coefficients of determination for the two best models do not differ much and therefore the above choice seems reasonable.

The complete output obtained from the above model is given in Table 3.

The F-value for the regression is 42,96 and this is significant at the 95% confidence level. Thus the independent variables as a whole provide a statistically significant causal relationship. The R^2s term shows a reasonably close fit with 92% of the variation in cash inflows being explained by the independent variables.

The t-values indicate that all the regression coefficients with the exception of $X16$ differ significantly from zero.

The assumptions of regression analysis

The four basic assumptions normally made in regression analysis must be met before a regression model can be used. These assumptions are:
(a) Linearity
(b) Independence of residuals
(c) Homoscedasticity
(d) Normality of residuals

Linearity

Linear regression assumes that the independent variables are linearly related with the dependent variable. The reasonably

Table 3 Regression equation for cash inflow

Independent variables	Regression coefficient	Estimated standard error	t-value
Intercept	−4,62089	0,85388	−5,41
Sales Lagged 1 Month ($X1$)	0,000373	0,00009	4,33
Sales Lagged 2 Months ($X2$)	0,000732	0,00009	8,08
Days ($X3$)	0,20469	0,03544	5,78
Dummy Variable No. 6 ($X11$)	−1,52675	0,35013	−4,36
Dummy Variable No. 7 ($X12$)	0,67773	0,27981	2,42
Dummy Variable No. 8 ($X13$)	−1,01780	0,29714	−3,43
Dummy Variable No. 11 ($X16$)	0,43415	0,27130	1,60

$R^2s = 0,920$

$F = 42,96$

Standard Error of Estimate = 0,4139
Durbin-Watson 1,6655

high value of R^2p and the significant F-value indicated that there seemed to be a linear relationship between the dependent and independent variables.

Independence of residuals

To establish whether the residuals are independent a commonly used plot, namely that of the standardized residuals against the fitted values, was utilized (see for example Chatterjee & Price, 1977). The plot revealed that the standardized residuals all tended to fall between 2 and −2 and were randomly distributed about zero. It therefore seemed that the residuals were independently distributed. Furthermore, the Durbin-Watson statistic lay within the range 1,10 to 2,90, indicating that there seemed to be very little or no autocorrelation.

Homoscedasticity

A plot of the standardized residuals against time revealed that the error term had a constant variance and therefore the requirement of homoscedasticity was met.

Normality of residuals

A normal probability plot of the residuals revealed that the residuals were approximately normally distributed.

It would thus appear that the multiple linear regression model given above is an appropriate model for the cash in-

Table 4 Regression equation for cash outflow

Independent variables	Regression coefficient	Estimated standard error	t-value
Incercept	−5,95423	1,40697	−4,23
Sales lagged 2 months ($X2$)	0,000749	0,00012	6,09
No. of trading days ($X3$)	0,299313	0,06072	4,93
Dummy Variable No. 5 ($X11$)	−0,96651	0,53150	−1,82

$R^2_4 = 0,68$

$F = 21,68$

Standard Error of Estimate = 0,879
Durbin-Watson = 1,86

Table 5 Test results of the regression models

Cash In (000ᶜ)			Cash Out (000ᶜ)			Net Cash Flow		
Actual	Predicted	Error	Actual	Predicted	Error	Actual	Predicted	Error
5,374	5,341	0,033	4,392	3,930	0,462	0,982	1,411	−0,429
6,752	6,318	0,434	5,836	5,487	0,349	0,916	0,831	0,085
5,643	6,184	−0,541	3,833	3,954	−0,121	1,810	2,230	−0,420

flow problem. One advantage of the above model is that the independent variables are all known with certainty and need not be forecast.

The model for cash outflow

A similar approach was adopted in developing a model for the cash outflow. The independent variables used were:

X1	−	sales lagged 1 month
X2	−	sales lagged 2 months
X3	−	sales lagged 3 months
X4	−	number of trading days
X5	−	the time variable
X6	−	creditors
X7 − X17	−	dummy variables

Preliminary investigations revealed that observations number 6, 7, and 16 were outliers and were consequently omitted from the analysis. The results of the analysis are contained in Table 4. The most appropriate regression model (in terms of Cp and R^2p) contained the variables $X2$, $X4$, $X5$, and $X11$.

The F-value for this model is significant at the 95% confidence level. The coefficient of determination is 68%, which is less than that obtained from the previous model. The t-values again indicate that all the regression coefficients except dummy variable number 5 are significant.

The above model does not seem to violate any one of the four assumptions mentioned previously and can therefore be accepted.

Testing the model

The predictive ability of regression models is usually determined by selecting a set of data not used in developing the model and using these data to evaluate the performance of the model. Owing to a lack of data (only 40 observations were available) it was not possible to evaluate the performance of the model adequately.

In order to obtain some idea of the predictive ability of the model the last three observations were not used in developing the model and were consequently used to test the model.

The results of the test are given in Table 5. The percentage errors for the cash inflow model are 0,6%, 64%, and 9,6%. For the cash outflow model they are 10,5%, 6,0%, and 3%. The percentage errors in the forecast of the net cash flow are 43%, 9,3%, and 23,3%. The present forecasting system gives the following percentage errors for the same three observations: 67,9%, 84,5%, and 23,5%.

These percentage errors are generally higher than those produced by the model. The limited test carried out cannot however prove the superiority of the regression model. The regression model should be implemented and the results obtained compared with the present system. Only then can the actual predictive ability of the model be established.

Conclusion

The treasurer of any corporation needs to forecast as accurately as possible the future cash position of his company so as to:
(a) Provide in advance for the cash requirements of his operations;
(b) minimize the cost of borrowings; and
(c) avoid loss of standing with his bankers.

To aid the treasurer in forecasting his short term cash flows, two regression models were developed, one to forecast cash inflows and the other to forecast cash outflows. The results obtained from these models indicate that the models may yield superior forecasts to those obtained from the present system. Owing to lack of data the test cannot be considered adequate but nevertheless the model should be implemented to establish its actual predictive performance.

References

Chatterjee, S. & Price, B. 1977. *Regression Analysis By Example*. New York: John Wiley & Sons, Inc.

Daniel, C.S. & Wood, F.S. 1971. *Fitting Equations To Data*. New York: John Wiley & Sons, Inc.