Interim Report

# A Review and Update of the Virginia Department of Transportation's Cash Flow Forecasting Model

J.S. Gillespie
Research Scientist

Cherie A. Kyte
Research Associate

December 1995
VTRC
96-IR1

## Executive Summary

The Virginia Department of Transportation uses a cash flow forecasting model to predict operations expenditures by month. Components of this general forecasting model estimate line items in the VDOT budget. The cash flow model was developed in the early 1980's and has not been updated since 1988. This project involved reviewing and updating several components of the cash flow model, namely, the monthly factors submodel used to predict monthly payments to construction contracts, and the maintenance expenditures submodel, used to predict maintenance payments.

The project produced an update of the monthly factors submodel by, for the most part, following the same basic methodology used in the development of the general cash flow model. A chart of the new monthly factors parameters is in the appendix. The authors employed a new approach to formulate a forecasting model to estimate maintenance expenditures. The resulting forecasting model is on page 11.

The authors will work in conjunction with Financial Planning and Debt Management staff in the coming months to evaluate the forecasting methods suggested by the update.

## Interim Report

### A Review and Update of the Virginia Department of Transportation's Cash Flow Forecasting Model

Cherie A. Kyte and Jim S. Gillespie

## PURPOSE

The Virginia Department of Transportation (VDOT) has employed a cash flow model to forecast revenues and expenditures since 1986. This forecasting model has not been updated since 1988. The purpose of this project is to update the methodologies used currently to predict VDOT's monthly payouts to construction and maintenance activities, as requested by the Financial Planning and Debt Management Division (FPDM). The two payout streams were addressed separately in the project.

Several theoretical reasons suggest the strong probability that the predictive quality of the cash flow model has gradually declined since its last revision in 1988. The 1986-88 data on which the cash flow model's last revision was based, reflected the status of the VDOT planning organization and standard operating procedure that prevailed during that period. Given the structural and managerial changes in VDOT since 1988, the underlying relationships that affect cash flow have probably changed. For instance, it is highly probable that the increased use of contract location and design work and the introduction of new environmental permitting requirements have affected the duration and cost of preconstruction activities for the typical highway project. In addition, the loss of experienced personnel as a result of early retirement programs in 1991 and 1995 has reportedly lengthened the construction process.

### Scope

The submodels estimate the following revenue and expenditure streams for each month in the particular forecast period:

- construction and construction contract expenditures
- federal reimbursements
- state aid
- maintenance expenditures
- consultant contracts, miscellaneous contracts, and other expenditures on materials and supplies; payments to localities, other agencies, and transit properties, and expenditures on salaries and wages, equipment, and Right-of-Way.

In this project, the Virginia Transportation Research Council (VTRC) considered and updated the monthly factors submodel for predicting construction contract payouts per month and the maintenance expenditures submodel used to forecast monthly maintenance expenditures.

## PROGRESS TO DATE

### Data Requirements

The first step in updating VDOT's cash flow model involved breaking it down into its components or submodels for initial review. Interviews with VDOT staff in the Financial Planning and Debt Management Division (FPDM), Fiscal Division (FD), and the Construction Division (CD) who are familiar with the cash flow model and an examination of previous documentation produced background information.[1] The data requirements for this project were extensive. The monthly factors submodel required data on monthly construction contract payouts by contract number from 1985 to 1995 and data on large samples of individual contracts. The data was provided by the Fiscal Division (FD), FPDM, Construction Division (CD), and Information Services Division (ISD).[2] ISD provided five self-extracting files from the Bid Analysis Monitoring System (BAMS) and the Current Project Data from Fiscal Division. The files contained information on construction contract payouts; more specifically:

1) Data on estimated contract amount, number of projects in the contract, number of bids, vendor, estimated completion date, let date, award date, work begin date, actual completion date, federal award amount, federal final amount, award amount, final contract amount, contract payment amount from BAMS, estimated days to completion, final payment, liquidated damages, and workorders.
2) Project specifics such as type (urban, rural, primary, secondary, interstate), length, etc.
3) Data by contract number on project amounts and dates of payments.
4) Vendor names and amounts paid to them.
5) Payouts to projects by month and project number.

The FD provided a copy of its Contractor Payment Processing System (CPPS), which is a computer program that contains actual payment streams to projects and contractors and can be used to generate yearly, monthly, daily, supervisor daily, active projects, project summary, final payment, contractor, and escrow retainages reports.

#### *Monthly Factors Submodel*

---

[1] *A System for Forecasting and Monitoring Cash Flow: Phase I: Forecasting Payments on Construction Contracts*, Adele Shapanka and Gary R. Allen, Virginia Transportation Research Council, 1983, and *A System for Forecasting and Monitoring Cash Flow: Phase II: Forecasting Federal and State Revenues, Contracts, Other Expenditures, and Cash Balances*, Adele Shapanka and Gary R. Allen, Virginia Transportation Research Council, 1985.
[2] Jim Hewitt, Miriam Daughtry, Carolyn Payne, David Reid, and Greg Whirley of the Finance Division, Ernie Miller, and Diane Mosby of the Financial Planning and Debt Management Division, Buddy Edwards, David Nestor, and William McDowell of Construction Division, and Tom Hutton of Information Services Division assisted in the retrieval of this data.

The monthly factors submodel depends on seven "Z-Vectors", which are columns of numbers used in conjunction with detailed contract data to predict payouts to construction contracts.[3] These seven vectors are the ZA Vector, the ZM Vector, the ZF, ZS, ZL Vectors, the ZC Vector, and the ZD Vector. VTRC re-examined and updated these vectors.

The appearance, function and updated calculation of each of the vectors will be explained in turn. For the most part, the update of the vectors followed the methodology used by Gary Allen and Adele Shapanka in the development of the monthly factors submodel.[4] However, the update used greater amounts of data and encompassed a longer period of time than the original study.[5]

*ZA Vector*

- This vector is used in conjunction with initial contract estimates to predict the final cost of a project. The vector has nine entries that correspond to the nine contract size groups defined in the BASIC program.[6]
- The update of this vector was based on the original method used by Allen and Shapanka in 1983 and a later update by Allen in 1988.
- The calculation of the new ZA vector required data on construction contracts, such as contract number and estimated completion date, all sorted by contract size.[7] This data was available on disk for the calendar years 1985 through 1994.[8].
- The above databases were adjusted to remove projects that were incomplete. This left a list of contracts in each contract size group. The fraction of cost overrun for each contract was calculated by dividing the actual final contract amount by the estimated final amount.
- To alleviate the effect of outliers, which may unduly influence the outcome, all contracts whose overrun rations were greater than 2.0 or less than 0.5 were removed from the database. The number of sample contracts within each of the nine groups varied from a low of 34 to a high of 1,580 contracts.
- The column of individual percent cost overruns was averaged to determine the mean percent cost overrun for each contract size group.

---

[3]This lengthy procedure is automated in a BASIC program. Explanatory notes were added to the program in the fall of 1994.

[4] Ibid.

[5]The original study did not adjust for inflation and neither did the update since an inflation adjustment would not affect the results.

[6] The BASIC program defines nine contract size groups: (1)✂$300,000, (2)✂$500,000, (3)✂$750,000, (4)✂$1,000,000, (5)✂$2,500,000, (6)✂$5,000,000, (7)✂$8,000,000, (8)✂$10,000,000, and (9)[illegible]$10,000,000. The program categorizes individual contracts into these groups for further analysis, e.g., to determine the amount paid out to a contract by its completion date.

[7] Other data included estimated final amount (an Fiscal Division forecast), date completed, estimated days until completion, and the actual final contract amount.

[8]Dbase 4 was employed to sort the data file by contract size and display the estimated contract size and actual payment for each contract.

*ZM Vector*

- The ZM Vector is used to predict the percentage of annual payout that (typically) will occur each month. The twelve vector entries correspond to the twelve months of the year. The entries are decimals that sum to one.
- Allen and Shapanka estimated the ZM Vector by using unweighted average annual shares.
- .The initial approach used in the update used data on monthly payouts to projects, which was supplied by ISD. The data series spanned FY 86 through FY 94.
- Each month's share of annual payout was determined, and then averaged across months. A weighted average was obtained by applying the total expenditure for each fiscal year to the monthly amount in the corresponding year.
- However, this method of updating the ZM Vector was rejected when it was found to produce unsatisfactory forecasting results.
- As an alternative method, data from CPPS was used in the re-calculation of an updated vector. This data set encompassed the fiscal years 1986 through 1994.
- The monthly fractions were determined for each month and then averaged over the eight fiscal years. A weighted average based on the number of payments occurring each month was applied to the monthly fractions.[9] FY 86 was dropped from the sample payments because there were very few payments made in this fiscal year. The researcher compared the ratios of samples to populations in each of the contract size group to determine if the samples used were representative of the population. The sample sizes were judged to be representative.[10]
- The next step was to normalize these results. Normalization was accomplished by summing the column of fractions and then dividing each entry by the total, producing weights that sum to one.
- The new vector was forwarded to FPDM in October 1995. Preliminary feedback from the FPDM indicates that the new ZM Vector produces good forecasts for all the months. The forecasts for July and August are a little lower than the actuals and the ZM Vector will be re-examined in the coming months.

*ZF, ZS, ZL Vectors*

- These three vectors represent the percentages of the total payments to contractors that comprise the first, second, and final payments. The vectors are used to predict the portion of contracts paid out in the first, second, and last month of a project. Each of the vectors has nine entries to correspond to the nine contract size groups.
- The update of these vectors was based on the original method used by Allen and Shapanka.
- The estimation of these vectors required data on individual payouts and total payouts to contractors by project. Reports containing the data were available from FD's CPPS, which

[9]The number of payments ranged from a low of 215 in February to a high of 349 in January and had a mean of 277.

[10]The ratios of sample to population according to contract size were: (1)1.7%, (2) 4.2%, (3) 8.2%, (4) 16.9%, (5) 9.3%, (6) 16.9%, (7) 26.5%, (8) 17.7%, (9) 27.0%.

generates contractor reports from the names of contractors.[11] A researcher drew a random sample of approximately 30 projects per size group.

- The contractor information was entered into an Excel spreadsheet. The first, second, and final payments were divided by the total payment to the contractor for each project to obtain the percent of the contract cost that is paid out by these points in time. The resulting fractions were averaged over the sample to determine the vector entries.

*ZD Vector*

- This vector contains the expected duration (in months) for contracts in each of the nine contract size groups. Generally, the larger the contract, the longer the duration of the project.
- Allen and Shapanka estimated a regression model that included original contract amount, month in which the contract was signed, road system, and project type as explanatory variables.[12] Due to the difficulty of obtaining a like data set, a different method was used in the update.
- The data came in hard copy from from the Construction Division and included the size of the contract, final dollar value, and the completion date by project number and district from January 1, 1988 through December 9, 1994.
- The researcher searched the CD printouts for projects to fill the nine contract size groups. The researcher tried to get four projects per district with completion years close to ’88, ’90, ’92, and ’94 in order to have a representative sample.[13]
- The vector entries are the average duration of projects within a particular contract size group, rounded up to the next highest integer. Therefore this calculation included only the size of the contract as an explanatory variable.

*ZC Vector*

- The nine entries in this vector are employed to predict the amount paid out to projects by their completion date.
- The update of the ZC Vector was based on the original method used by Allen and Shapanka.
- The estimation of the vector required data on payments to completed projects and completion dates, both of which were available in the CPPS printouts from FD and the list of completed projects from CD. The CD printouts showed the completion dates by project number, and the CPPS printouts provided the payment schedules and total costs by project number. These two sets of data were examined to find matching project numbers and the data was entered into an

---

[11] Since FD did not have an historical list of contractors, the researcher opened a composite file of CPPS and printed this list. The contractor names were entered into the software and the corresponding reports were printed. The researcher tried to use only one project per contractor and get thirty projects per contract size group in order to get an unbiased sample. This process was repeated until there were thirty samples in the majority of the contract size groups.

[12] Following this methodology for the current study would have been very difficult since hard copies of individual contracts are in storage and the authors were unable to discover their exact location. Alternatively, the authors could have relied on databases from ISD, but this method would have been very labor intensive since it would have involved combing through two data files of 3,823 and 6,463 records apiece. Furthermore, the data set still would have been incomplete since the month in which the contract was signed is not included in these ISD files.

[13] See appendix.

Excel file in order to have completion dates and payment schedules by project number in the same file.[14]

- The amount paid out by completion date was divided by the actual total payouts for each project. The ZC entries were computed by averaging the sample ratios in each of the contract size groups.

Using the new Z Vectors for forecasting construction payout is relatively simple. The seven updated Z Vectors can be entered into the BASIC program in place of the older Z Vectors. Then the program may be run as before to provide forecasts of contract amounts paid out by VDOT each month.

### *Maintenance Expenditures Submodel*

The maintenance expenditures submodel is used to forecast VDOT's payouts on maintenance each month. FPDM currently relies on the historical averages developed by Allen and Shapanka to predict these expenditures. The researchers have taken a new approach to develop the maintenance disbursements forecasts.

## Progress to Date

### *Data Requirements and Updated Methodology*

The new methodology required the construction of a regression model that could be used to forecast maintenance expenditures accurately. The actual values of monthly maintenance expenditures, i.e., the dependent variable, came from FPDM and encompassed the period from FY 89 through FY 95 and were in current dollars.[15] The monthly maintenance disbursements follow a seasonal pattern of peaks in the summer and autumn months and troughs in the spring and winter months. Eight time series of likely explanatory variables and thirty different model formulations were evaluated in the model building process.[16] The independent variables tested were:

- *Annual budget/12* - the annual budget figures were compiled from VDOT's annual budget publications. Each year's total for maintenance in the Highway Maintenance Operations Fund (HMOF) was divided by 12 to give a monthly average. This variable is denoted by "AB".[17]
- *Retail sales* - monthly retail sales were obtained from the Department of Taxation. Three, six, and twelve month lags were tested to determine if the effect of retail sales on maintenance disbursements was delayed by some time period.

---

[14] See appendix.
[15] All money variables are in current dollars.
[16] A chart of these formulations is in the appendix.
[17] Each time series spanned the fiscal years 1989 through 1995.

- *New car titles* - taxable new car titles by month were obtained from the Forecasting Division of the Department of Motor Vehicles. Three, six, and twelve month lags were tested for the same reason as above.
- *Net taxable gallons of fuel* - the amount of gas and diesel sold that is destined for highway use were obtained from DMV. Three, six, and twelve month lags were tested.
- *Temperature* - average monthly temperature for the Commonwealth of Virginia was compiled from monthly publications of the National Oceanic and Atmospheric Administration[18]
- *Precipitation* - total monthly precipitation for Virginia also came from the NOAA publications
- *Trend* - a trend variable was tested to account for the upward trend in the dependent variable over time
- *Lags on maintenance disbursements* - to determine if expenditures in previous months affected expenditures in the current months, one, two, three, and four-month lags on the dependent variable were tested.

The best result obtained from using the above time series in a regression was:

[19] $MD = \alpha + \beta_1 AB + \beta_2 Temperature + \beta_3 Titles_{t-3} + \varepsilon$

The $R^2$ of a regression is a measure of its "goodness of fit". The $R^2$ of the above regression was 0.5475; in other words, the equation explains about 55% of the variation in maintenance disbursements. Since the equation still leaves much of the behavior of maintenance disbursements unexplained, additional variables were tested:

- A dummy variable for a recession was tested. This dummy variable was used to account for changes in maintenance disbursements due to a recessionary period - effects that would not otherwise be present in normal economic times. The dummy variable will indicate in which cell or month an unusual event occurs. According to quarterly estimates of gross domestic product from the United States Statistical Abstracts, a recession occurred in the fourth quarter of calendar year 1990, and the first and second quarters of 1991.[20] For the sake of analysis then, the months of October, November, and December of 1990, and January, February, March, April, May, and June of 1991 were assigned a value of one.
- Another dummy was used to account for the effect of heavy snow on maintenance disbursements. The snow dummy was assigned a value of one in months in which snowfall usually occurs and a value of zero in the remaining months. The "snow months" were varied to ensure the influencing months were captured by the dummy.

---

[18]These publications contain data from the National Climatic Data Center.
[19]Where $\alpha$ is an intercept term and $\varepsilon$ is the error term (discussed later in the report)..
[20]Statistical Abstract of the United States. 114th edition. 1994. The National Data Book,. United States Department of Commerce, Economics and Statistics Administration, Bureau of the Census.

A few of the data points in the maintenance disbursement series appeared to be unusually low when compared to the same months in other years. For example, the amount for June of 1989 was only $17,000 while for every other June, the amounts were ranged from $39,687,000 to $76,169,000. A similar phenomenon occurred in December of 1988. Conversations with Fiscal Division staff revealed that the reason for these anomalous entries was that the Commonwealth Accounting Report Systems (CARRS) report for FY89 was flawed. To alleviate the skewing that these odd points caused, "breakpoint dummies" were used. These dummy variables have a value of one to correspond to the particular point in question, and a value of zero at every other point. In total, nine breakpoint dummies were tested.

The researchers tested two alternative devices to account for the seasonal pattern of maintenance disbursements. The first involved employing eleven month dummies, each of which would indicate the number of months remaining in the fiscal year. The second alternative was to use a trigonometric function. Both alternatives are discussed in the following sections.

*Month Dummy Models*

These models incorporated the eleven month dummies and all the other variables in significance testing in order to build the appropriate regression model. This process involved running many preliminary regressions to determine which variables contributed significantly to the explanation of the behavior of maintenance disbursements. The dependent variable is used in a regression with each of the variables in turn. The variable with the highest t-statistic is chosen and then the dependent variable is run on this variable and all others, again in turn. This process continues until all the significant variables are chosen. Combinations of variables are also tested to determine if adding two or three variables at once can add to their significance. This process was repeated for every model formulation. The different models can then be compared and judged on the criterion of $R^2$. The model with the highest $R^2$ was as follows:

[21]

$$MD = \alpha + \beta_1 AB + \beta_2 Titles_{t-3} + \beta_3 DJUNE89 + \beta_4 DDEC88 + \beta_5 JULY + \beta_6 AUGUST + \beta_7 FEB + \beta_8 MAY + \varepsilon$$

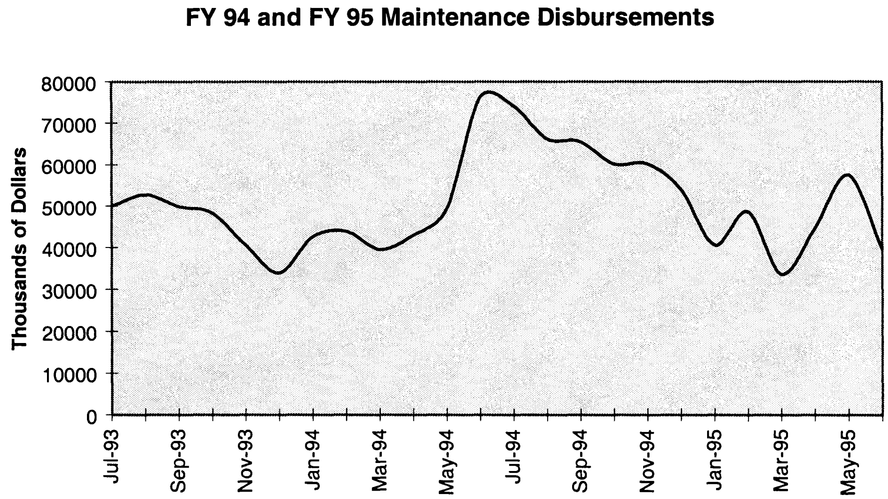
where $\varepsilon$ is a random disturbance. The disturbance arises for several reasons. The main reason is that we cannot capture every influence on disbursements, such as all the inherent randomness of human behavior. The net effect of such intangible factors is captured in the disturbance term. The $R^2$ of the above model is 0.6555; in other words, this particular model explains about two thirds of the variation in maintenance disbursements.

*Trigonometric Function Models*

[21] MD is Maintenance Disbursements, AB is the annual budget for the year divided by 12, $Titles_{t-3}$ is the three month lag car titles, DJUNE89 is the breakpoint dummy for June of 1989, DDEC88 is the breakpoint dummy for December of 1988, JULY, AUGUST, FEB, and MAY are the month dummies.

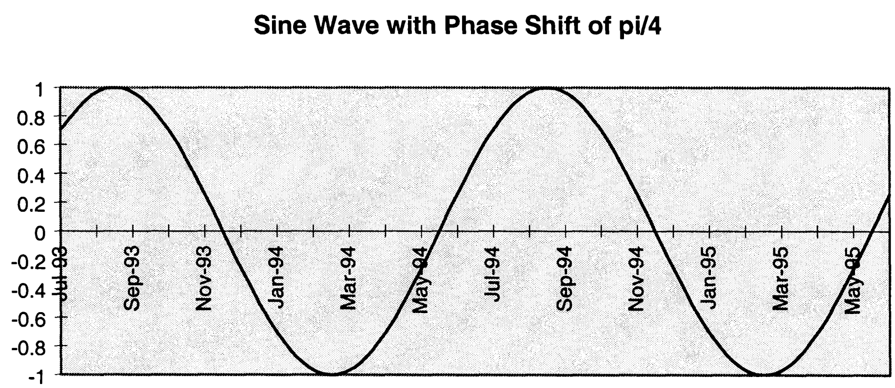
The trigonometric function's purpose is to model the seasonal pattern of maintenance expenditures. Figure 1 shows the pattern of maintenance disbursements for FY 94 and FY 95.

Figure 1




It was conjectured that a sine wave with a period of one year and a phase shift of about pi/4 would imitate much of the seasonal variation. Such a sine wave relative to a July 1993 zero is shown for FY 94 and 95 in Figure 2.

Figure 2




The addition of a second term with a half year period and the same phasing adds more flexibility at the cost of estimating one additional parameter. The continuous trigonometric function is rendered into twelve discrete month long pieces by integration. This means the area under the curve can be found. The equation is:

$$y_m = \int_{\frac{\pi}{6}m-1}^{\frac{\pi}{6}m} [\sin(t+\theta) + A\sin(t+\theta)]dt$$

where
m=month index=1 for July, 12 for June, etc.
$\theta$=Phase shift variable (different $\theta$ move the sine wave laterally through time)
A=weight on the second term.

The value of the equation varied by month. For instance, the equation for July is:

$$y_{july} = \int_0^{\frac{\pi}{6}} [\sin(t+\theta) + A\sin(t+\theta)]dt$$

Once the above equation was integrated, the form for our sample month was:

$$y_{july} = -\cos(\frac{\pi}{6}+\theta) - A/2\cos 2(\frac{\pi}{6}+\theta) + \cos(0+\theta) + A/2\cos 2(0+\theta)$$

To avoid the arduous task of iterating over both A and $\theta$ to determine which values of each result in the highest $R^2$ for the trigonometric function, the integrated equation was split into two parts:[22]

$$y_{1july} = -\cos(\frac{\pi}{6}+\theta) + \cos(0+\theta)$$

$$y_{2july} = -\cos 2(\frac{\pi}{6}+\theta) + \cos 2(0+\theta)$$

The two trigonometric equations were modified according to the which functional form of the general maintenance disbursement regression. For example, y1 and y2 were multiplied by the annual budget term and in another case, were multiplied by the square root of the annual budget term. For example:

$$MD = \alpha + \beta_1 AB + \beta_2 y_1 AB + \beta_3 y_2 + \varepsilon$$

Because $\theta$ does not enter the equation linearly, the equation cannot be estimated by Ordinary Least Squares (OLS), so the researchers iterated over different values of $\theta$ and then estimated the rest of the parameters by OLS.[23] Optimal values of $\theta$ were found for each form of the trigonometric equations. The trigonometric functions became columns of values once values for t were substituted into the equations.

[22]To test the significance when the trigonometric function is kept as a composite, one form of the composite y was used. The y term was multiplied by the square root of AB. However, the results were not as precise as the y1 and y2 results since the composite y parameters were determined by iteration.
[23]By running repeated regressions with different values of $\theta$ the best value for $\theta$ can be reached. The corresponding values of A were found by dividing the coefficient for y2 by the coefficient for y1.

These trigonometric equations were then put in the general regression and all variables were tested for significance. The following equation had the highest $R^2$ of all forms tested:

$$MD = \alpha + \beta_1 AB + \beta_2 y_1 AB + \beta_3 y_2 AB + \beta_4 Titles_{t-3} + \beta_5 DJUNE89 + \beta_6 DDEC88 + \varepsilon$$

The $R^2$ for this equation was 0.6840, therefore the above equation explains about 70% of the variation in maintenance disbursements. Since this model had the highest $R^2$ of any tested, it was chosen as the model on which forecasts should be based. This final model was tested for heteroskedasticity (non-equal error variances) and autocorrelation (interdependent errors). The explanatory variables were tested for multicollinearity.[24] Tests for heteroskedasticity, autocorrelation, and multicollinearity revealed that the model is free of these problems.[25]

As a final test, the final model was used to forecast maintenance disbursements from FY 90 to FY 95.[26] The current forecasting model was also used to generate forecasts for these fiscal years. Comparing these two sets of results to the actual maintenance disbursements for these years shows that the new method has an average absolute percentage error of 16% while the current method has an average absolute percentage error of 20%.

## FINDINGS AND RECOMMENDATIONS

The forecasting model is as follows:

$$MD = -22708.56 + 0.00108 * AB + 0.00043 * y_1 AB + 0.00007 * y_2 AB + 0.68210 * Titles_{t-3} + \varepsilon$$

It should be noted that the breakpoint dummies are not included in the forecasting model since they were simply used to neutralize the impact of a rare event whose recurrence is not anticipated.

To forecast monthly maintenance disbursements, each month's corresponding annual budget amount, y1 and y2 figures, and the titles figure from three months previous are entered into the forecasting equation. This can be done in an Excel spreadsheet. The formula may simply be copied to all months in the forecast period. For example, suppose we want to forecast maintenance disbursements for the month of June FY 95 and we have the following information:

- the annual budget divided by twelve is $51,050,379.6
- the value of the function y1 at this month is 13,545,556

---

[24]The presence of any of these phenomena can lead to imprecise models and, therefore, imprecise and inaccurate forecasts.

[25]A detailed discussion of the problems of heteroskedasticity, autocorrelation, and multicollinearity, as well as the particulars of their tests, appears in the appendix.

[26]FY 89 was not included in the comparison because of the anomalous entries in December and June.

- the value of the y2 function is 44,937,398
- the number of car titles three months previous is 30,788

By placing each of these values into the forecasting equation, we can predict the amount of maintenance disbursements for June of FY 95. In this case, the equation gives the number $62588.26 in thousands of dollars.

The values of the trigonometric functions must be calculated for each year of the forecast period. If the new car titles figures are not available for the required months, these values may be generated using a simple growth equation.

The updated forecasting construction and maintenance forecasting methodologies will be forwarded to FPDM for evaluation in late December. The authors will work in conjunction with FPDM in the coming months to fine tune the models. The final models should be in place by May, 1996.

## APPENDIX

### ZA Vector Sample Sizes

The initial sample size and the sample size used after incomplete contracts, blank cells, and outliers were discarded are, respectively, by contract group: (1) 1810, 1580, (2) 713, 638, (3) 366, 314, (4) 177, 160, (5) 321, 285, (6) 177, 155, (7) 113, 101, (8) 34, 30, (9) 111, 95.

**ZD Sample Size**

The sample sizes for the nine groups are: (1) 36, (2) 36, (3) 34, (4) 35, (5) 34, (6) 28, (7) 19, (8) 5, (9) 10.

**ZC Sample Size**

The sample sizes for the nine contract groups are: (1) 10, (2) 12, (3) 11, (4) 10, (5) 10, (6) 13, (7) 7, (8) 5, (9) 10.

**Heteroskedasticity, Autocorrelation, and Multicollinearity**

Heteroskedasticity occurs when the error terms have different variances. This implies that the variance of maintenance disbursements changes over time. This means that the results of the regression will be biased, therefore the coefficients will be biased. The estimates of the coefficients will be either too low or too high and therefore the predictive quality of the model will suffer. For example, suppose a model with heteroskedasticity produced a coefficient value of 3 for the variable car titles. This coefficient would be inaccurate and therefore the forecast would be inaccurate. Autocorrelation occurs when the disturbance term relating to any observation is influenced by the disturbance term relating to any other observation. For example, if there is autocorrelation, a disruption in one of the variables may be carried over into the next month. The presence of autocorrelation will also result in biased estimators. Multicollinearity occurs when there is a linear relationship between some or all of the explanatory variables. If perfect multicollinearity is present, the regression coefficients of the explanatory variables are indeterminate and their standard errors are infinite. If multicollinearity is less than perfect, the coefficients possess large standard errors, which means the coefficients cannot be estimated with great accuracy.

*Tests for Heteroskedasticity*

The problem of heteroskedasticity is more likely to appear in models with cross-sectional data than in models with time series data. Even though all the formations of the maintenance model are based on time series data, the final model was tested for the presence of heteroskedasticity in two ways. First, the Glejser test was performed. The hypothesis of the test is that is that the heteroskedasticity in the model comes from one of the independent variables, which in this case, is the annual budget term. The absolute values of the residuals from the final model are regressed against different functional forms of the annual budget term and an intercept. A significant coefficient on the annual budget variable indicates the presence of heteroskedasticity. The Glejser test revealed that none of the AB coefficients were significant - in fact, they were highly insignificant, which means the hypothesis is disproved and there is no heteroskedasticity in the model.

A second test that is more general in nature than the Glejser test, known as the White Test, was also performed on the model. The White test is an indicator of *any* heteroskedasticity between any variable or combination of variables in the model. The squared residuals from the final model were regressed on as many variables and combinations of variables as Excel could accommodate (16). The $R^2$ of this regression was multiplied by the number of observations (84). This critical value was then compared to the test value of the chi-square at a 95% level of confidence with 15 degrees of freedom. The value from the test was 9.65 and the critical chi-square value was 25. Since 9.65<25, the hypothesis is rejected and therefore no heteroskedasticity was found in this model.

*Test for Autocorrelation*

The Durbin-Watson Test was performed on the model. The test is based on the calculation of the *d statistic*, which is the ratio of the sum of squared differences in successive residuals (errors) to the sum if the squared residuals. Since the d statistic is computed from the residuals which are dependent on the explanatory variables, the distribution of the *d* is difficult to find. Unlike other tests such as the t-test or the F-test, there is no unique critical value that will lead to the acceptance or rejection of the hypothesis that there is no autocorrelation. In the case of the model chosen, the *d* should not differ from the value of 2 - otherwise there is autocorrelation. The performance of the test on this model resulted in a value of 2.01, which is not significantly different from 2, therefore, there is no heteroskedasticity in this model.

*Test for Multicollinearity*

Multicollinearity exists when the correlation coefficient between variables approaches a value of one. In practice, any value above 0.7 can cause problems in a regression model. In the case of the final model, the highest correlation coefficient for any combination was only 0.4. Therefore, there is an absence of multicollinearity in this particular model.