TILL CASH MANAGEMENT MODEL

by

GORDON ARTHUR SICK

B.Sc., University of Calgary, 1971
M.Sc., University of Toronto, 1972

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN BUSINESS ADMINISTRATION

in

FACULTY OF COMMERCE &
BUSINESS ADMINISTRATION

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December, 1976
© Gordon Arthur Sick, 1976
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Finance

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date December 8, 1976
This thesis develops a model for the management of till cash (currency and coin) of a branch for a Vancouver area credit union. The model is developed in two parts. First, a model is estimated to forecast cash demand and then a cash order algorithm is developed.

Two statistical models are developed to estimate cash demand. The first employs Box-Jenkins time series techniques. This model fails because the cash flow data are non-stationary, exhibiting both a growth trend and high autocorrelations at large lags. In the second model, a growth trend for real weekly cash flows is first estimated, incorporating an asymptotic capacity constraint. The real cash flow trend is converted to a nominal trend and used as the weight in a linear weighted least squares model for daily cash flows, in which the explanatory variables are dummy variables to indicate days of the week, months of the year, incidence of pay days, etc. The consistency of the resulting forecast model is also discussed.

To develop a cash order algorithm, steady state models are first considered. These models are generally based on stationary cash demand, constant delivery lag times for orders and other assumptions that are inappropriate in this till cash management setting. To relax the steady state assumptions a general dynamic programming framework is
developed for the cash management model that allows for either penalty costs for cash-outs (cash shortages) or a chance constraint involving the probability of a cash-out. Because of non-stationarity of the cash flows the dynamic program cannot be solved directly, but an approximate solution is obtained using a simulation technique. The resulting algorithm is tested on historical data and the results are discussed briefly.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1 The Nature of the Problem</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.2 Overview</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>FORECASTING THE DEMAND FOR CASH: THEORY</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2.1 Introduction</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2.2 Forecasting Cash Demand as a Time Series</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2.3 Forecasting Demand by Regression</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>2.4 Summary</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>FORECASTING THE DEMAND FOR CASH: APPLICATION</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>3.1 Introduction</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>3.2 The Data</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>3.3 Estimation by Box-Jenkins Techniques</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>3.4 Estimation by Regression Techniques</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>3.5 Summary</td>
<td>46</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Estimated Regression Coefficients for Daily Demand</td>
<td>38</td>
</tr>
<tr>
<td>2</td>
<td>Cash Delivery Order Lags</td>
<td>87</td>
</tr>
<tr>
<td>3</td>
<td>Performance of the Cash Order Model and of Management</td>
<td>87</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Growth of Real Cash Demand</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>The Cash Order Algorithm</td>
<td>82</td>
</tr>
<tr>
<td>3</td>
<td>The Minimum Cash Order Subroutine</td>
<td>83</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

The author would like to thank the officers of the credit union for the data and valuable discussions, Professor Robert White for advisory assistance and Professor William Ziemba for technical assistance. Any errors or omissions are the responsibility of the author alone.

The research was performed while the author was in receipt of a Samuel Bronfman Award and a University of British Columbia Graduate Fellowship.
Chapter 1

INTRODUCTION

1.1 The Nature of the Problem

The purpose of this study is to develop a model to order till cash (currency and coin) for a Vancouver area credit union branch, with the objective of minimizing costs, while maintaining adequate balances to meet the members' demand. The demand for cash is not deterministic, and part of the problem is to determine the size of precautionary cash balances which will reduce the risk of a cash-out to an appropriate level desired by management. A cash-out occurs if the balance of cash on hand falls below some level which management regards as the practical minimum level for smooth operation. This level may be zero, but could, more meaningfully, be about $10,000, since a smaller level would not allow all tellers to have a reasonable supply of all denominations of currency and coin.

The costs that are to be minimized are the fixed costs of placing a cash order and the opportunity costs of idle cash balances. If a policy is taken to make large orders, few orders are required and fixed order costs are low, but cash balances are high, creating large opportunity costs. Conversely, if orders are small and frequent, fixed order costs are high and opportunity costs are low. Minimization of total costs entails
finding an optimal trade-off between fixed order costs and variable opportunity costs.

Thus, the problem is essentially one of constrained optimization: minimize costs subject to some constraint involving the risk of a cash-out. In this sense, it is a standard inventory problem. Inventory theory is generally oriented towards solving analytically tractable problems which do not always fit real world situations very well. The terms of reference of this project require an operational model. Thus, in deciding whether or not a given inventory model is appropriate, the question is not whether the real problem fits the assumptions of the model (real problems rarely fit perfectly), but how well will the model perform when its specifications are violated.

Operationality of the cash order model also means that the costs of development, day to day operation and maintenance of the model should be reasonable. Since the credit union branch has on-line computer facilities, there are no major computational restrictions to the algorithm, but to minimize development costs, the algorithm should be rather straightforward. Operational costs relate mainly to the personnel time required to generate daily input. Maintenance costs are the costs of re-estimating any parameters of the model if the structure of the problem changes over time.

It is in this operational environment that the cash order model is developed.

1.2 Overview

Development of the model requires the solution of two major sub-problems: estimation of the demand for cash and construction of the actual order algorithm.
Chapter 2 discusses the theoretical aspects of forecasting the demand for cash. First, various methods of adaptive time series forecasting are considered, culminating in a discussion of Box-Jenkins time series techniques. Then, the problems of developing a linear regression model for forecasting cash demand are considered.

In Chapter 3, the results of both Box-Jenkins estimation and weighted least squares estimation are presented. The Box-Jenkins method, although preferable in the sense that it is an adaptive forecasting technique, is found to be inapplicable because the demand for cash is non-stationary. The non-stationarity arises partly because of the general growth in demand level and partly because factors of vastly different orders of seasonality are present. In order to use weighted least squares, a forecasting model for the general trend level in cash demand is fitted by least squares, and then daily cash flows, deflated by the general trend level, are regressed on dummy variables relating to months of the year, days of the week, paydays, etc. Tests of the predictive ability of the model are also discussed in this chapter.

Chapter 4 is a discussion of the general aspects of the cash management problem. Several models that have appeared in the inventory theory and cash management literature are discussed. Some of these models are based on steady state assumptions such as stationarity of the demand for cash. The appropriateness of these assumptions is discussed, and for purposes of comparison, a dynamic programming formulation of the problem is presented.

In Chapter 5, an approximate solution to the dynamic programming problem is developed. It essentially yields a variable control
limit model. Cash is ordered when the cash balance is inadequate for transactionary and precautionary requirements, which vary according to the expected demand during the order delivery lag period. The order size is selected to minimize expected costs per day during the life of the order, as computed by simulation. The chapter also discusses tests of the model on historic data.

In Chapter 6, a few concluding remarks are made about the extent to which the techniques developed here can be generalized to the cash management problem for other credit union branches.
2.1 Introduction

Forecasting is an inductive reasoning process whereby historic relationships between a variable of interest (the dependent variable) and explanatory variables (the independent variables) are analyzed, in order to predict (perhaps with error) future values of the dependent variable. A general explanatory model is of the form

\[ y_t = f(x_{1t}, \ldots, x_{nt}) + u_t \]  

(2.1.1)

where \( y_t \) is the dependent variable at time \( t \) (e.g. cash flow),
\( f \) is a function whose form is determined \( a \) priori but which may have empirically estimable parameters,
\( x_{1t}, \ldots, x_{nt} \) are the explanatory variables at time \( t \) (e.g., lagged observations of \( y_t \), or dummy variables for days of the week),
and \( u_t \) is the forecast error, the distribution of which may be important to the forecaster.

We may develop the model so that \( E(u_t) = 0 \), so that, given \( (x_1, \ldots, x_n) \),
the forecast for \( y \) is

5
\[ \hat{y} = f(x_1, \ldots, x_n) \quad (2.1.2) \]

There are generally many independent variables and many more models which may be used to explain historical values of any given dependent variable. Rarely do all of these models agree in their estimates, and a forecaster must choose which model is most appropriate. He cannot simply choose the model which best fits historical data, because the historical relationships may not continue to hold in the future. It is possible to fit a model on an early subset of historical data and test its performance on a later subset (as is done later in this paper), but there is still the outstanding question of stability of relationships. Thus, if inappropriate treatment of, or omission of, some explanatory variable causes little harm to the performance of the model on historic data, there is no assurance that the same type of error will be insignificant in the future. In the present application, for example, cash flows are increasing over time. Many forecasting methods that have increasing estimates over time can be fitted to predict well on historic data. Some of the estimates will increase without bound over time, some will increase to a finite asymptotic limit, and some will ultimately decrease after a certain time. Selection of an appropriate model cannot be made objectively with historical data only, but must involve some subjective judgment of the forecaster.

Within this framework, the theories of several potential forecasting models for cash demand at the credit union branch are analyzed. In Section 2.2, cash demand is regarded in a time series framework whereby future cash flows are estimated in terms of the series of historical flows only. Such models are frequently adaptive in the sense that certain
types of forecasting errors tend to be corrected over time. Emphasis is placed on Box-Jenkins time series models. It is pointed out that, if the cash demands have significant autocorrelations at very long lags, such as seasonal spans of one month or one year, it may not be possible to fit a parsimonious Box-Jenkins model to cash demands.

In Section 2.3, a two-part model is developed for historical cash demand. First a general trend over time is fitted to real cash demands which incorporates the \textit{a priori} notion that the branch is growing towards an asymptotic capacity constraint. Then cash demands are deflated by this trend level in a weighted least squares linear model which uses as explanatory variables dummy variables indicating the incidence of various days of the week, months of the year, paydays and holidays. This model uses a wider range of explanatory variables than the Box-Jenkins model and, as a result, may provide viable estimates even when the Box-Jenkins model fails.

2.2 \textbf{Forecasting Cash Demand as a Time Series}

In a time series model, the only independent variables in the model (2.1.1) are lagged observations of the independent variable \((y_t-1, y_t-2, \ldots)\). Many classical forecasting models such as arithmetic moving averages and exponentially weighted moving averages have this form. An excellent discussion of such models is given in Wheelwright and Makridakis [1973]. They are adaptive in the strict sense that, if there is an increase or decrease in the mean level of the variables \(y_t\), arithmetic and exponentially weighted moving average forecasts increase or decrease accordingly and converge over time to the new level. Thus absolute
forecast errors decrease in an adaptive manner, unless the general level changes again. If the time series fluctuates rapidly over time, sophisticated weighting schemes are required to reduce the response time in the adaptation process and thus reduce forecast errors. Classical theory says little about methods of selecting optimal values for the time period for moving average or the weight for an exponentially weighted average. Also, these techniques fail to provide any description of the distribution of the random error term \( u_t \) in (2.1.1).

For the problem at hand, some description of the error term \( u_t \) is valuable in analyzing the risk of a cash-out. Also, cash flows fluctuate significantly from day to day, week to week and month to month because of predictable variations in demand. A naive weighting scheme in a weighted average forecast would be highly inaccurate because of the frequent reversals in cash demand. For example, an exponential scheme would forecast high demand in January simply because there was high demand in December. The need for quantification of forecast errors and a more sophisticated weighting scheme suggests the use of Box-Jenkins techniques.

Many authors have considered two general models of a time series of observations on a single variable \( y_t \) (such as cash flow). One is the infinite autoregressive (AR) form:

\[
y_t = \delta + u_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots
\]

(2.2.1)

where \( y_t \) is the \( t^{th} \) observation in the series,

\( \delta \) is a constant,

\( \{u_t\} \) is a sequence of independent identically distributed random shocks (unobserved),
and \( \{\phi_i\} \) is a sequence of AR weights.

This is the most general form of a weighted average scheme, and expresses the current observation as a weighted sum of earlier observations plus a random shock. Note that, in general, it is not an "average" since the weights \( \phi_i \) need not sum to 1 and \( \delta \) need not be zero.

Alternatively, one can express the current observation as a weighted sum of the sequence of earlier unobserved random shocks. This is the moving average (MA) form:

\[
y_t = \delta + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \cdots
\]

(2.2.2)

where \( \{y_t\}, \delta, \{u_t\} \) are as before and \( \theta_1, \theta_2, \cdots \) are the moving average weights.

The AR and MA forms are equivalent in the sense that, given one, the other can be found, although it may not be a convergent series. For forecasting purposes, the general AR and MA forms are impossible to estimate since they have an infinite number of parameters. Box and Jenkins [1970] proposed that the forecaster should consider a finite combination of (2.2.1) and (2.2.2) which parsimoniously uses a finite number of parameters. This they call an auto-regressive moving-average (ARMA) process. Its general form is

\[
y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = \delta + u_t - \theta_1 u_{t-1} - \cdots - \theta_q u_{t-q}
\]

(2.2.3)
It can be seen that (2.2.3) represents the current observation $y_t$ as the sum of a constant $\delta$, a current random shock $u_t$, and weighted sums of past observations and past shocks (the AR and MA parts, respectively). Introducing the backshift operator $B$ where $B y_t = y_{t-1}$, $B^k y_t = y_{t-k}$, and then defining the AR polynomial

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

(2.2.4)

and the MA polynomial

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

(2.2.5)

we can rewrite (2.2.3) as

$$\phi(B) y_t = \delta + \theta(B) u_t$$

(2.2.6)

A desirable property of (2.2.6) is stationarity, which means that the form of the stochastic process does not vary over time (in particular, the scaling and location of the variation should be constant). Stationarity is equivalent to having the zeros of the AR polynomial outside the unit circle. If there is a zero strictly inside the unit circle, the series exhibits explosive behaviour. Zeros on the unit circle result in mild non-stationarity and correspond to factors in $\phi(B)$ of the form $(1 + B^s)$ or $(1 - B^s)$ for $s = 1, 2, 3, \cdots$. The form $1 - B^s$ corresponds to taking the difference between the current observation and the observation at the (seasonal) lag $s$. Powers of $(1 - B^s)$ correspond to taking differences
of differences. One can remove the difference factors from the AR part to rewrite (2.2.6), with obvious changes in notation, as

$$\phi(B) (1 - B)^d (1 - B^s)^D (1 - B^s)^s y_t = \delta + \theta(B) u_t \quad (2.2.7)$$

In practice, if the observed series $y_t$ exhibits non-stationarity (such as growth over time), the forecaster should try various forms of differencing before attempting to solve for $\phi$, $\delta$ and $\theta$.

Analogous to stationarity of the AR part is invertibility of the MA part. By recursive substitution of observations with earlier observations in (2.2.6) or (2.2.7) we can represent the current observation as an infinite AR series of past observations. The invertibility requirement stipulates that this infinite series shall converge - that is, less and less weight is placed on earlier and earlier observations to determine the current observation. This is an important requirement for forecasting purposes, since one does not know what appropriate values of early, unobserved data to use. Invertibility is equivalent to the zeros of the MA polynomial lying strictly outside the unit circle.¹

Given the invertibility and stationarity requirements, an ARMA process is uniquely determined by its autocorrelations. Thus one can look at the estimated autocorrelations of the series (along with partial autocorrelations and inverse autocorrelations which are related to them) to identify an appropriate model of the form (2.2.7) for an observed time series.

The parsimony concept requires selection of a model where $d$, $D$, $p$ and $q$ are as small as possible. This presents a potential problem
for the application of Box-Jenkins techniques to forecast cash demand, for one would expect a lot of information content in high order autocorrelations. For daily cash demands there should be high positive autocorrelations at 5 day lags because of the correspondence of days of the week, and there should also be high autocorrelations at 260 day lags because of the correspondence of months of the year (there are approximately 260 working days per year). Since the number of days in a month is not constant, the lags corresponding to semi-monthly pay days will vary in a cyclic fashion with a very long period which will create an important autocorrelation at that lag. Moreover, some weeks only have 4 days and this will upset the basic autocorrelation structure at 5 day lags. All of this suggests it may be hard to fit a parsimonious model to daily data.

The problems for daily cash flow estimation are mitigated somewhat by considering weekly cash flows. A long 260 day log is then a more reasonable (but still long) 52 week lag. The problem of holidays reducing some weeks to 4 days does not affect weekly lag structures as seriously as it does daily lag structures.

The actual estimation of (2.2.7) is accomplished by minimizing the sum of squared residuals $\sum u_t^2$, iteratively by the Marquardt algorithm. Under the hypothesis that the $u_t$'s are independent $N(0,\sigma)$, this yields maximum likelihood estimates of $\theta$, $\phi$ and $\delta$.

Unbiased k-day ahead forecasts that minimize mean square prediction error can be developed from the random shock form (2.2.2) or the ARMA form (2.2.3) where historical one-period forecast errors are substituted for the random shocks (and the k unrealized random shocks are set to 0). The forecast errors are normally distributed if the random
shocks are, and the k-day forecast error variance is \((1 + \theta_1^2 + \cdots + \theta_{k-1}^2) \sigma_u^2\) where the \(\theta_i\)'s are given by the random shock form (2.2.2).

If a Box-Jenkins model can be fitted, the creation of forecasts and estimation of forecast error variances is quite straightforward. The main problem to be faced when applying Box-Jenkins is the question of whether an invertible, stationary model can be fitted.

2.3 Forecasting Demand by Regression

In Section 2.2 it was pointed out that forecasts (of demand) by time series techniques only use as explanatory variables the earlier observations of the same series. It was argued that, in order to capture most of the explainable variation in cash demand, autocorrelations at long lags would be important, and the model would not likely be parsimonious. Another approach is to use more explanatory variables in the forecasting model (2.1.1).

One way to incorporate additional variables is to use a regression model:

\[ y = x\beta + u \]  

(2.3.1)

where

\[
\begin{align*}
  y &= \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \\
  x &= \begin{bmatrix} x_1' \\ \vdots \\ x_N' \end{bmatrix} \\
  u &= \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}
\end{align*}
\]

\[ E(u' u') = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_N^2 \end{bmatrix} \]
\( x_t \) is a \( k \times 1 \) vector of explanatory variables at time \( t \),
\( u_t \) is a random error, with \( \text{E} u = 0 \),
and \( \beta \) is a \( k \times 1 \) vector of regression coefficients.

Appropriate explanatory variables \( x_{it} \) include dummy variables which take the value 1 for the various days of the week, months of the year, pay days and holidays, and are zero otherwise. That is, if \( x_t' = (x_{1t}, \ldots, x_{kt}) \), we may have, for example,

\[
x_{1t} = \begin{cases} 
1 & \text{if day } t \text{ is a Tuesday} \\
0 & \text{otherwise} 
\end{cases}
\]

Another component that must be considered for \( x_t \) is the level of interest rates, since classical economic models regard it as a determinant of demand for money. There are several problems with introducing interest rates into the model. First, over the short three year period of available data, interest rates are highly collinear with time, so it would be difficult to distinguish growth effects from interest rate effects. Second, it is difficult to decide which interest rate series to use. Certainly the interest rate most likely to influence a credit union member's demand for cash is the interest rate offered on demand deposits. However, these rates are changed infrequently and any statistically estimable effects (over a three year period) would be spurious. It is difficult to argue that other interest rate series, such as commercial paper rates, have any observable effect on demand for currency. Thirdly, if some macro-economic model were available to express demand for money in terms of interest rates, such a model would be oriented towards money in the form of demand deposits.
and other major forms of cash, rather than currency. It is not necessarily true that demand for currency is as strongly affected by general economic conditions as the demand for other forms of money. Thus, for operational simplicity, interest rates were excluded from the model.

The model also must explain the historic growth in cash demand. One possible way to model this is to have some explanatory variable that increases over time (such as t). This would only model an additive growth effect and fail to model any increase in the effects of the dummy explanatory variables over time. That is, if general cash demand increases over time, one would also expect that the extra cash demand on a payday for example also should increase over time. To model this, one could allow polynomial time trending of the coefficients. That is, consider the model

\[ y_t = x_t' \beta_0 + x_t' \beta_1 t + \cdots + x_t' \beta_p t^p + u_t \quad (t = 1, \ldots, N), \]

where the \( x_t \)'s are vectors of, say, 0-1 dummy variables and the \( \beta \)'s are \( k \times 1 \) vectors of regression coefficients which are multiplied by powers of \( t \).

The degree of the polynomial \( p \) is typically less than or equal to 2 (quadratic time trending). In effect, each coefficient in \( \beta \) is allowed to follow a trend over time, where \( \beta = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p \).

This model has the linear form

\[ y_t = w_t' \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + u_t \]
where
\[
W_t = \begin{pmatrix}
(x_{1t}, tx_{1t}, \ldots, t^p x_{1t})', \\
\vdots \\
(x_{kt}, tx_{kt}, \ldots, t^p x_{kt})'
\end{pmatrix}
\]

Quadratic time trending, for example, triples the number of coefficients to be estimated, which would yield low significance levels for \( \hat{\beta} \) if there are a large number of explanatory variables. Moreover, polynomials are poor for extrapolative forecasts. That is, polynomials may have maxima or minima for \( t > T \) which simply result from the need to fit high order derivatives in the range \( 1 < t < T \), and do not in any way reflect any predictive power of the polynomial model. For example, fitting the simple trend model (for scalar \( \beta_0, \beta_1, \beta_2 \))
\[
y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + u_t
\]
to weekly cash flows \( y_t \) resulted in a polynomial that predicted a peak in demand in August 1976 (just slightly beyond the data range)! Clearly such growth models are inappropriate.

As the final choice to model the growth, let us simultaneously consider a related problem, that of heteroscedasticity. Initial regressions indicated that \( \sigma_t^2 \), the variance of \( u_t \) increases over time and, indeed, is proportional to \( d_t^2 \), the square of the general level of cash demand, at time \( t \). That is, \( \sigma_t^2 \propto d_t^2 \). Without loss of generality, we may take \( d_t \) to be the general level of weekly demand. This suggests using a multiplicative growth model to simultaneously model growth and eliminate heteroscedasticity. In the linear model (2.3.1), weighted least squares is required to eliminate
heteroscedasticity. In this case the appropriate weights are \(1/d_t\). The model is then

\[
\frac{y_t}{d_t} = \frac{1}{d_t} x_t' \beta + \tilde{u}_t \quad (t=1,\ldots,N) \tag{2.3.2}
\]

or,

\[
W = Z\beta + \tilde{u} \tag{2.3.3}
\]

where

\[
W = \begin{pmatrix}
y_1/d_1 \\
\vdots \\
y_N/d_N
\end{pmatrix}
\]

\[
Z = \begin{pmatrix}
\frac{1}{d_1} x_1 \\
\vdots \\
\frac{1}{d_N} x_N
\end{pmatrix}
\]

\[
\tilde{u} = \begin{pmatrix}
\tilde{u}_1 \\
\vdots \\
\tilde{u}_N
\end{pmatrix}
\]

and

\[
E(\tilde{U}) = 0
\]

\[
\text{cov}(\tilde{U}) = \sigma^2 I_T
\]

This can be rewritten as

\[
y_t = d_t (z_t' \beta + \tilde{u}_t) \tag{2.3.4}
\]

which is a multiplicative form. Note that \(w_t = y_t/d_t\) is the cash flow at time \(t\) expressed as a proportion of the general trend level of weekly cash demand, so we can think of \(w_t\) as being normalized cash demand at time \(t\). The \(z_t\) should be selected so as to exhibit no growth over time (that is,
as 0-1 dummy variables). The $d_t$ provide for growth of $y_t$ and eliminate heteroscedasticity at the same time.

Other possible sets of weights $\{d_t\}$ which in some way indicate the size of the branch may be considered, such the number of members or the total amount of deposits. Here it is again important to consider whether historical relationships between the "size" measures and demand will continue into the future. If care is not taken to exclude inactive members from a count of members, changes in the proportion of active members on the rolls will distort the relationship between cash demand and number of members. The branch opened a new building in May 1973, and initially experienced rapid growth for some time afterwards, but the rate of growth has decreased since then. In the early part of the period, most members were new and active, but after some time, many members became inactive, making member counts a poor measure of demand level.

The magnitude of the liabilities of the branch is also indicative of size, but this is not likely to be a stable predictor of cash flow, since liabilities can vary depending on the various retirement savings programs, home ownership savings programs, etc. that may be in vogue from time to time. That is, at various times members may change their total deposits without changing their demand for cash.

Thus, the weight $d_t$ is chosen to be proportional to the general trend level of cash flows at time $t$. In order to model the cash flow trend, let us examine the two main historic reasons for the increasing cash flow trend. These are the increase in nominal demand due to inflation and the increase in real demand due to growth of the branch. It would be folly to try to forecast inflation, but this is not necessary since Consumer Price
Index figures are available monthly and linear extrapolation beyond the latest two months of unavailable data provides an excellent forecast of the CPI level (although, of course, extrapolation is a poor predictor of changes in CPI levels). It is necessary, however, to forecast levels of nominal cash demand. Thus, let

\[ \hat{d}_t = \text{trend level of real weekly cash demand at time } t, \]
\[ p_t = \frac{\text{CP index at time } t}{\text{CP index for January 1976}}^2, \]

and

\[ d_t = \text{trend level of nominal weekly cash demand at time } t, \text{ expressed in January 1976 dollars.} \]

Then

\[ d_t = p_t \hat{d}_t. \tag{2.3.4} \]

Estimation of the trend now reduces to estimation of the trend in real demand \( \hat{d}_t \). The real weekly cash flows generally increased over the period for which data were available, but the growth rate tended to decrease over the period. Indeed, regressions to fit piecewise linear trends in time showed a monotonic increasing but concave cash trend. This is similar to the general activity levels of other branch operations because rapid growth followed the opening of a new building in May 1973, but by 1976, further growth was dampened by the general physical constraints of the building (e.g., teller wicket space is now limited and long line-ups at the tellers' wickets act to discourage members from using the branch cash services). Thus, growth in real cash demand appears to be approaching a capacity constraint, as in Figure 1, and this should be incorporated in the forecast model.
A simple way of modelling such an asymptotic constraint is by a suitably scaled and located rectangular hyperbola, namely

\[ \tilde{d}_t = c - \frac{k}{t + t_0} \]  

where \( c \) is the capacity constraint,
\( k \) is a positive scaling constant,
and \( t_0 \) is a relocation parameter for time (\( t + t_0 > 0 \) for all \( t \) in the model).

Johnston [1972, p. 52] suggests such reciprocal transformations with the artificial restriction \( t_0 = 0 \). This is a linear function of \( \frac{1}{t + t_0} \) and can be fitted by least squares. That is, let

\[ d'_t = \frac{\text{nominal weekly cash flow at time } t}{p_t} \]

For a fixed value of \( t_0 \)
\[ d_t' = c - \frac{k}{t + t_0} + e_t \]  \hspace{1cm} (2.3.6)

where, for

\[
e = \begin{pmatrix}
e_1 \\
\vdots \\
e_N 
\end{pmatrix},
\]

\[
E(e) = 0 \text{ and } E(e'e') = \Sigma_e = a\begin{pmatrix}
d_1^2 & 0 \\
0 & \ddots \\
\ddots & \ddots & \ddots \\
0 & \ddots & d_N^2
\end{pmatrix}
\]

for some \( a > 0 \). As in the regression (2.3.3) the variance of \( e_t \) is proportional to the square of the trend level in cash demand. Since \( E(e) = 0 \), the least squares estimators \( \hat{c} \) and \( \hat{k} \) are unbiased, but inefficient under the heteroscedasticity, which merely places more weight on later observations. This is the most desirable departure from homoscedasticity, in this case. To select the time location parameter \( t_0 \), it suffices to estimate (2.3.6) for various values of \( t_0 \) and choose \( t_0 \) to minimize the sum of squared residuals.

The estimated weekly real cash trend is

\[
\hat{d}_t = \hat{c} - \frac{\hat{k}}{t + t_0}.
\]

and the weekly nominal cash trend is

\[
\hat{d}_t = p_t \left( \hat{c} - \frac{\hat{k}}{t + t_0} \right). \]
These values of $\hat{d}_t$ can be used in place of $d_t$ in the weighted least squares model (2.3.3) to obtain the approximate generalized least squares estimate,

$$\hat{\beta} = (X' \hat{\Sigma}_u^{-1} X)^{-1} X' \hat{\Sigma}_u Y = (Z'Z)^{-1} Z' \hat{W}$$

where

$$\hat{\Sigma}_u = \sigma^2 \begin{pmatrix} d_1^2 & 0 \\ \vdots & \ddots \\ 0 & d_N^2 \end{pmatrix}, \quad \hat{W} = \begin{pmatrix} y_1/d_1 \\ \vdots \\ y_N/d_N \end{pmatrix}$$

For a realization of the dummy explanatory variables $z_t$; we have an estimate of nominal cash demand at time $t$ of

$$\hat{y}_t = \hat{d}_t \hat{w}_t$$

$$= p_t \left[ c - \frac{\hat{k}}{t + t_0} \right] z_t \hat{\beta}$$

(2.3.8)

where

$$\hat{y}_t = \text{cash flow forecast for time } t,$$

$$p_t = \frac{\text{CPI at time } t}{\text{CPI for January 1976} },$$

$$\hat{d}_t = c - \frac{\hat{k}}{t + t_0} = \text{forecast trend level of real weekly cash flow at time } t,$$

$$z_t = \text{vector of explanatory dummy variables for time } t,$$

$$\hat{\beta} = \text{estimated regression coefficient}.$$

Analogously, let $\hat{y}_t$ be the (unobtainable) estimate of $y_t$, given perfect knowledge of the trend. That is, let
The following proposition gives consistency results for the forecasts and variance of forecast errors.

**Proposition:** The predictions $\hat{y}_t$ and $\hat{y}_t$ are both consistent predictions of $y^*$ and a consistent estimator of the variance of $\hat{y}_t$ is

$$\hat{d}_t^2 = s^2 [1 + z_t'(Z'Z)^{-1} z_t]$$

where $s^2$ is the residual variance obtained by estimating (2.3.3) with the estimated weights $\hat{d}_t$ instead of $d_t$.

**Proof:** We have seen that $\hat{c}$ and $\hat{k}$ are unbiased. To see that they are consistent, let

$$V_{Nx2} = \begin{bmatrix} 1 (1 + t_0)^{-1} \\ \vdots \\ \vdots \\ 1 (N + t_0)^{-1} \end{bmatrix}$$

Then $\hat{c} = (V'V)^{-1} V' \hat{d}$ and the covariance matrix of $\hat{c}$ is

$$(V'V)^{-1} V' \Sigma_e V(V'V)^{-1}$$

\[(2.3.11)\]
By interpreting enlargement of the sample size \( N \) to mean the observation of a sequence of independent but stochastically identical credit union branches over the same time period, the regressor matrix \( V \) is "constant in repeated samples" (Theil [1971, pp. 364-365]), so that
\[
N^{-1}(V'V) \text{ and } N^{-1} V' \Sigma_e V \text{ both converge to positive definite matrices.}
\]
We can rewrite (2.3.11) as
\[
N^{-1}[N(V'V)^{-1} (N^{-1} V' \Sigma_e V) N(V'V)^{-1}] \text{ and this tends to 0 as } N \to \infty.
\]
Thus, the variances of \( \hat{c} \) and \( \hat{k} \) tend to 0 for large \( N \), establishing the consistency of these estimators. 3

Thus, the elements of
\[
\hat{\Sigma}_u = \sigma^2 \begin{bmatrix} \hat{\sigma}_1^2 & & 0 \\ & \ddots & \vdots \\ 0 & & \hat{\sigma}_n^2 \end{bmatrix}
\]
are consistent estimators of those of \( \Sigma_u = E(U'U) \). Also, the elements of these matrices are bounded (in probability) since the trend levels \( d_t \) are bounded by the capacity constraint \( c \). Define the positive definite matrix
\[
Q_\Sigma = \lim_{N \to \infty} N^{-1} \sigma^2 Z'Z = \text{plim}_{N \to \infty} N^{-1} X' \hat{\Sigma}_u^{-1} X.
\]
Then by the consistency and boundedness, we also have
\[
\lim_{N \to \infty} N^{-1} (X' \hat{\Sigma}_u^{-1} X) = Q_\Sigma.
\]
Then \( \text{plim}_{N \to \infty} N^{-1} X' \hat{\Sigma}_u^{-1} X = Q_\Sigma - Q_\Sigma = 0. \) The same boundedness and consistency arguments establish that
\[
\text{plim}_{N \to \infty} N^{-1} X' (\hat{\Sigma}_u^{-1} - \Sigma_u^{-1}) U = 0
\]
and
\[
\text{plim}_{N \to \infty} N^{-1} U' (\hat{\Sigma}_u^{-1} - \Sigma_u^{-1}) U = 0.
\]
Hence by Theil's Theorem 8.4 [1971, p. 339], \( \text{plim}_{N \to \infty} \sqrt{N} (\hat{\beta} - \beta) = 0, \) \( s^2 \) is a consistent estimator of \( \sigma^2 \), and the matrix \( s^2(Z'Z) \) converges in probability to the covariance matrix of \( \hat{\beta} \). In particular, note that \( \hat{\beta} \) is consistent. Thus \( \hat{w}_t = z_t' \hat{\beta} \) is a consistent estimator of \( w_t \), with variance given asymptotically by
Since \( \hat{y}_t = d_t \hat{w}_t \) and \( \hat{d}_t \) is a consistent estimator of \( d_t \), formula (2.3.10) must be the asymptotic variance of \( \hat{y}_t \). The consistency of \( \hat{y}_t \) and \( \hat{y}_t \) follows from that of \( \hat{\beta} \), \( \hat{c} \) and \( \hat{k} \). This establishes the proposition.

We may conjecture that the asymptotic variance (2.3.10) of \( \hat{y}_t \) is also asymptotically the variance of \( \hat{y}_t \). At any rate, that variance will be a meaningful practical approximation to the variance of \( \hat{y}_t \).

Thus, we have generalized least squares estimates for cash demand, along with asymptotic results about the estimators and their variances. This completes our analysis of the regression model.

2.4 Summary

In this chapter we have studied two statistical models which may be used to forecast cash demand. The first was a Box-Jenkins time series model in which forecasts of future cash flows are weighted sums of previous cash flows. The second model was the product of a non-linear trend model, and a weighted least squares model which uses dummy regressors to indicate days of the week, months of the year, pay days and holidays. The statistical estimation of both models will be discussed in the next chapter.
FOOTNOTES TO CHAPTER 2

1To illustrate, consider an MA process \((q=1)\) of first successive differences:

\[ y_t - y_{t-1} = \delta + u_t - \theta u_{t-1} = \delta + (1 - \theta B) u_t \]

\[ \therefore u_t = y_t - y_{t-1} - \delta + \theta u_{t-1} \]

Substituting the second equation recursively into the first to successively eliminate \(u_{t-1}, u_{t-2}, \ldots\) yields:

\[ y_t = y_{t-1} + u_t + \delta - \theta(y_{t-1} - y_{t-2} - \delta + \theta u_{t-2}) \]

\[ = u_t + (1 + \theta) \delta + (1 - \theta) y_{t-1} + \theta y_{t-2} - \theta^2 u_{t-2} \]

\[ = \ldots \]

\[ = u_t + (1 + \theta + \theta^2 + \ldots) \delta + (1 - \theta)(y_{t-1} + \theta y_{t-2} + \theta^2 y_{t-3} + \ldots) \]

Since \(E u_t = 0\) a linear unbiased forecast of \(y_t\), given \(y_{t-1}, y_{t-2}, \ldots\) is

\[ \hat{y}_t = (1 + \theta + \theta^2 + \ldots) \delta + (1 - \theta)(y_{t-1} + \theta y_{t-2} + \theta^2 y_{t-3} + \ldots) \]

Under mild restrictions on the \(y\)'s, this converges iff \(|\theta| < 1\), i.e., iff \(1 - \theta B\) has its zero outside the unit circle. Also, note that for \(\delta = 0\), \(\hat{y}_t\) is simply an exponentially weighted moving average of earlier observations, one of the classical forecasting models.

2The Consumer Price Index to be used is the Statistics Canada seasonally unadjusted monthly CPI for Vancouver. One may argue that a seasonally adjusted index would be better, since it would not distort some of the seasonal variations in cash flow that should be estimated in the model, but such a figure is not available. More importantly, one might argue that the CPI bundle of goods includes items that are paid for by cheque as well as cash. None of the indices available reflect cash payments better than the general CPI, however.
The interpretation of $V$ as being constant in repeated samples only indicates that trend estimates are consistent for time points in or shortly after the 3 year data range, $I, \cdots, N$. To test the accuracy of extrapolative forecasts the model will be fitted in Chapter 3 on the first two years of data and tested on the third year. This is a useful adjunct to the consistency results presented here which are only of an interpolative nature.

Alternatively, one can interpret sample enlargement as the addition of more observations for the same credit union as time proceeds. To consider this type of consistency, first note that

$$V'V = \begin{pmatrix} N & \sum_{t=1}^{N} (t_0 + t)^{-1} \\ \sum_{t=1}^{N} (t_0 + t)^{-1} & N/\sum_{t=1}^{N} (t_0 + t)^{-2} \end{pmatrix}$$

Let

$$(V'V)^{-1} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

The first term in the denominator converges to a finite limit as $N \to \infty$.

The square root of the second term is

$$\frac{1}{\sqrt{\sum_{t=1}^{N} (t_0 + t)^{-2} - N^{-1}(\sum_{t=1}^{N} (t_0 + t)^{-1})^2}}$$

which behaves asymptotically like

$$\frac{1}{\sqrt{N}} \int_{t_1}^{N} \frac{dt}{(t^0 + t)} = \frac{1}{\sqrt{N}} \ln \left[ \frac{t_0 + N}{t_0 + 1} \right] \to 0 \text{ as } N \to \infty.$$ Therefore, as $N \to \infty$, $m_{22} \neq 0$ and hence $(V'V)^{-1} \neq 0$.

However, $m_{11} = \frac{\sum (t_0 + t)^{-2}}{\Sigma (t_0 + t)^{-2} - (\sum (t_0 + t)^{-1})^{-2}} \to 0$ as $N \to \infty$ since the numerator and first term in the denominator have finite limits and the second term in the denominator has an infinite limit.

Similarly $m_{12} = m_{21} = \frac{1}{\Sigma (t_0 + t)^{-2} / \Sigma (t_0 + t)^{-1} - \Sigma (t_0 + t)^{-1}} \to \infty$ as $N \to \infty$ since the first term in the denominator tends to zero and the second term tends to $\infty$. Thus $(V'V)^{-1} \to \begin{pmatrix} 0 & 0 \\ 0 & \alpha \end{pmatrix}$ as $N \to \infty$ for some $\alpha > 0$. 


The elements of $V'eV$ each behave asymptotically like those of $V'V$ since the $d_t^2$ diagonal elements of $\Sigma_e$ are bounded. Hence $V'\Sigma_e V(V'V)^{-1}$ tends to a finite-valued matrix as $N \to \infty$ whose diagonal entries are non-zero. Hence as $N \to \infty$, the matrix $(V'V)^{-1} V'\Sigma_e V(V'V)^{-1}$ tends to a matrix of the form

$$
\begin{pmatrix}
0 & 0 \\
0 & \alpha'
\end{pmatrix}
$$

for $\alpha' > 0$. Thus, the estimator of the capacity constraint $\hat{c}$ is consistent, but the estimator $\hat{k}$, which describes how quickly the capacity is approached, is not consistent. This means that for projections of growth into the distant future, the forecast of ultimate capacity is "accurate" (consistent). The projected rate of growth to that capacity may be over or under estimated, but forecast errors induced by this are bounded since the projection of capacity is "accurate."
Chapter 3

FORECASTING THE DEMAND FOR CASH: APPLICATION

3.1 Introduction

In this chapter, the two cash demand models developed in Chapter 2 are estimated statistically. The next section is devoted to a discussion of the data and some of the special problems associated with them. Section 3.3 presents the results of Box-Jenkins analysis. It will be found there that attempts to estimate a stationary, invertible time series model fail. Section 3.4 presents the estimation of a cash flow trend, which is used in a weighted least squares linear model with dummy variables to forecast cash demand. This results in an estimable model, the forecasting ability of which is studied by fitting the model on the first two years of data and testing the resulting third year forecasts for bias and other sources of error.

3.2 The Data

Data on cash balances and cash orders for the credit union branch were analyzed for the period May 1973 to April 1976. This period commenced with the opening of a new building, and earlier data are likely to be structurally different at least in terms of the general level of cash flow, if not also in the pattern of variations in the flows.
Closing till cash balances were obtained for the 768 days from General Ledger accounts and information on cash deliveries was obtained from delivery invoices prepared by the British Columbia Central Credit Union, through which cash orders were placed. This allowed the computation of daily cash flows. In addition, during December 1974 several deposits were solicited from a large local merchant to cover cash shortages. These were treated in the same way as regular cash deliveries for the calculation of daily cash flows. Of the resulting 767 daily cash flows, all but 36 were negative (outflows), because the branch dealt primarily with consumer accounts rather than commercial accounts.

Collection of the data was actually quite complicated because maintenance of the General Ledger accounts was oriented towards ensuring correct figures at month end and many discrepancies crept into the accounts during the month. Three general types of discrepancies were observed, only some of which could be corrected.

The first type of discrepancy arose when two tellers exchanged cash between themselves and offsetting debits and credits to the General Ledger were not made on the same day. Alternatively, if the head teller was absent for several days and did not post the changes in the treasury cash balance, while other tellers withdrew money from the treasury (and posted their credits immediately) the same type of discrepancy would occur. These discrepancies resulted in a positive (negative) flow one day followed by a negative (positive) flow of equal magnitude a few days later. The methods of making General Ledger entries varied from time to time and for the middle half of the data period, each teller's closing balance was posted directly to the General Ledger (rather than as an aggregate with all
other tellers). Frequently this provided enough detail, in conjunction with the balance in the teller's interchanges account (which normally is zero), to infer the correct adjustment for these discrepancies. Such adjustments could not be made, of course, for the time periods when all tellers were posted to the General Ledger together.

The second type of discrepancy occurred when journal entries were inadvertently posted to the wrong account and corrected a few days later. This also would result in an positive (negative) flow one day, followed a few days later by a negative (positive) flow of equal magnitude. Little could be done to eliminate this type of discrepancy.

A third type of discrepancy could have resulted from cash delivery invoices being filed under the wrong branch. Indirect evidence of the possibility of such discrepancies was provided by the observation that invoices for other branches were found incorrectly filed with those of the branch involved in this study. This type of discrepancy would result in a positive flow one day, which, unlike the first two types of discrepancies, would not be followed by a compensating negative flow a few days later.

If the first two types of discrepancies occur randomly and independently of the dummy regressor variables (indicating days of the week, months of the year, pay days and holidays), then no bias will be introduced into the estimates \( \hat{\beta} \) and \( \hat{\gamma} \) of Section 2.3, although the estimated residual variance \( s^2 \) will be biased upwards.

The second type of discrepancy will bias some or all of the regression coefficients (for demand) downwards, although the bias is not likely to be serious since only 3 orders were found mis-filed with the branch studied.
3.3 Estimation by Box-Jenkins Techniques

When the theory of the Box-Jenkins technique was presented in Section 2.2 it was pointed out that, with significant autocorrelation information at short and long lags, a time series model for daily cash flows might be hard to fit and that aggregation to weekly cash flows would present fewer problems for estimation. If a weekly model cannot be fitted, neither can a daily model. Thus, the initial results presented here are for weekly cash flow $y_t$.

Since the cash flows generally increased over time, first differences were required to achieve stationarity. Consecutive and 4th order seasonal differencing were considered on both raw and logged data. Note that 4th order seasonal differencing corresponds approximately to a seasonal lag of one month. The autocorrelations and partial autocorrelations of these differenced series suggested a 4th or 5th order moving average process. These models are of the form

$$(1-B^s)\hat{y}_t = \delta + (1-\theta_1 B-\theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4 - \theta_5 B^5) u_t$$

where $s = 1$ or 4,

$y_t$ = weekly cash flow or logged weekly cash flow,

and $\theta_4 \neq 0$ but perhaps $\theta_5 = 0$.

For raw weekly cash flows with $s = 4$ and $\theta_5 = 0$, the model was fitted by the Marquardt least squares algorithm with initial estimates obtained by solving a system of linear Yule-Walker type equations involving the inverse autocorrelations and the MA parameters $\{\theta_i\}$. This yielded the form
\[(1-B^4)y_t = (1 - .117 B + .250 B^2 - .266 B^3 - .698 B^4) u_t + 2479 \]
\[R^2 = .382 \quad (1.84) \quad (4.25) \quad (-4.65) \quad (5.22) \quad (5.05) \quad (3.3.1)\]

The t-statistics (147 degrees of freedom) for each parameter are given in parentheses, and all except that for \( \theta_4 \) are to test the hypothesis that the parameter equals 0. The t-statistic for \( \theta_4 \) is to test the hypothesis \( \theta_4 = 1 \).

At first sight the model (3.3.1) appears to be quite reasonable in that all coefficients but \( \theta_1 \) are significantly different from zero. However, to check the invertibility of the process (which is essential to forecasting), one must examine the factored form of \( \theta(B) \):

\[\theta(B) = -.698(B + 1.34)(B - 1.05)(B + .05 - 1.01i)(B + .05 + 1.01i) \quad (3.3.2)\]

Note that the last three roots of \( \theta(B) \) are virtually on the unit circle, so that forecasts based on (3.3.1) will be unstable. Moreover, we can approximate \( \theta(B) \) as follows:

\[\tilde{\theta}(B) = - (B + 1)(B - i)(B - 1) = 1 - B^4 \quad (3.3.3)\]

The only major differences between (3.3.2) and (3.3.3) are the constant coefficient and the first factor. But (3.3.3) is simply the differencing polynomial in (3.3.1). Substituting the approximation (3.3.3) into (3.3.1) and cancelling the factor \( 1 - B^4 \) to eliminate the over-differencing yields

\[y_t = u_t + \delta'.\]
This implies that \( y_t \) is essentially white noise. Thus, the autocorrelations created by differencing constituted almost all of the autocorrelations available for estimating (3.3.1) and induced an MA polynomial that is virtually identical to the AR differencing polynomial. Of course, the t-statistics of (3.3.1) indicate that the MA polynomial is not exactly \( (1 - B^n) \), but the only difference between the polynomials is in the factor \( (B + 1.34) \). All the other factors of \( 1 - B^n \) cancel exactly.

When all of the other plausible weekly models that were mentioned earlier were fitted exactly the same phenomena happened: The factored MA polynomials had all of their roots essentially on the unit circle and the factors included those of the AR polynomial \( (1 - B) \) or \( (1 - B^n) \), which indicates over-differencing, even though differencing is required to achieve stationarity.

Thus, from a Box-Jenkins viewpoint, there is not enough information in the autocorrelations of weekly cash flows to overcome the white noise, and the time series model is not estimable. One might conjecture that if the trend is eliminated from cash demand, as with the trend model discussed in Section 2.3, a major source of the random error is eliminated and the remaining series may be estimable. Since the real requirement is to have a model for daily rather than weekly cash flows, this last hypothesis was not tested for weekly cash flows but for daily cash flows deflated by the trend. That is, in the notation of (2.3.3) let \( \hat{w}_t = \frac{y_t}{\hat{d}_t} \) be the daily cash flow \( y_t \) deflated by the trend \( \hat{d}_t \). For all combinations of first order and seasonal \( (s = 5 \text{ days}) \) differencing, the series of autocorrelations failed to die out at large lags, indicating that there was still too much information at long lags to allow the fitting of a parsimonious Box-Jenkins model.
In summary, the Box-Jenkins estimation failed because the series of cash flows is non-stationary. The non-stationarity occurs in two forms. One source of non-stationarity is the trend. The other source of non-stationarity is a seasonal drift related to the autocorrelations at long lags.

3.4 Estimation by Regression Techniques

Since the attempt at Box-Jenkins estimation of the demand for cash was a failure we consider here a different collection of explanatory variables and use the model developed in Section 2.3, which employs a non-linear trend estimate multiplied by a weighted least squares linear model.

The non-linear trend (2.3.6) of 156 real weekly cash flows, $d_t$, was fitted by least squares yielding

$$\hat{d}_t = 170,000 - 28.9 \times 10^6 (t + 240)^{-1}$$  \hspace{1cm} (3.4.1)

where $t$ is the number of working days from May 1, 1973 and $t$-statistics are given in parentheses. As expected, the residuals displayed increasing variance over time. As a result, the estimated $t$-statistics and variance of the error term are all biased. The estimated coefficients, however, are unbiased and consistent.

The asymptotic capacity level of real weekly cash demand that is forecast by (3.4.1) is $170,000$ (expressed in January 1976 dollars). Substituting $t = 767$ yields an estimated trend level for May 1976 of $141,000$
or 83% of the asymptotic level of cash demand. This figure is thought
to be reasonable in light of the actual remaining potential for growth
at the branch.

To test the goodness of fit for (3.4.1), a quadratic polynomial
in time was added to the regression and it did not significantly increase
$R^2$. Fitting the model for the first and last halves of the data separately
did not result in significantly different estimates for the regression
coefficients $\hat{c}$ and $\hat{k}$. (Again, we must emphasize that standard significance
tests are not meaningful in such an heteroscedastic situation, and these
statements of significance are heuristic only.)

Note that since precisely three years of data were used to
estimate the trend, no bias will be introduced as a result of seasonal
fluctuations in demand.

The trend estimates multiplied by the CPI levels $p_t$ were then
used as weights $d_t$ in the generalized least squares model (2.3.3) to
estimate the fluctuations in cash demand about the trend line. The inde­
pendent variables $z_{it}$ were dummy variables indicating the incidence or
proximity of days of the week, months of the year, semi-monthly and
monthly pay days and holidays. An initial regression was run using all
of these variables as regressors and the residuals were examined for
evidence of the first two types of data discrepancies discussed in Section
3.2. Such discrepancies manifest themselves as reversed pairs of residuals
of large magnitude, separated by only a few days, and correspond to a
bookkeeping error and its correction. In order to reduce any tendency of
these errors to bias the estimated regression coefficients, adjustments
to the data were made whenever there was a pair of residuals with opposite
signs, each having a magnitude in excess of 3 standard errors of the regression. The adjustment was made so that the sum of the cash flows was unchanged and the resulting residuals were equal. Such adjustments were made in 8 cases, out of the 759 data observations that were used for the demand estimation.

The final form of the demand equation (2.3.3), with \( k = 28 \) explanatory variables, is presented in Table 1.

To avoid multi-collinearity the regression was run with no dummy variable for January. Since the other winter months, November and February, had coefficients that were insignificantly different from zero, they were dropped from the model and hence equated to January. The regression coefficients for Tuesday and Wednesday mid-month and month-end pay dates were essentially identical, so they were included as one dummy variable.

A few comments about the regression coefficients are in order. Recall that the equation is for daily cash demand expressed as a proportion of the general weekly trend level of cash flow, and the explanatory variables are either 0 or 1. Thus the \( i \)th regression coefficient indicates the estimated increase in cash demand (expressed as a proportion of the weekly trend level) that is caused by incidence of the event for which the \( i \)th variable is an indicator function. For example, an ordinary Tuesday in January that is not on or prior to a pay day will have a estimated cash demand of 5.32% of the general weekly trend level.

From the first five coefficients we can see that the demand for cash rises through the week, peaking on Saturday (even though Saturday has shorter hours of operation than the other days).
Table 1
Estimated Regression Coefficients for Daily Demand

<table>
<thead>
<tr>
<th>Coefficient Subscript i</th>
<th>Estimated Coefficient $\hat{\beta}_i$</th>
<th>Standard Error of $\hat{\beta}_i$</th>
<th>Dummy Variable $z_{it}$ is a 0-1 Indicator Function for:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.0532</td>
<td>.0120**</td>
<td>Tuesday</td>
</tr>
<tr>
<td>2</td>
<td>.0844</td>
<td>.0114**</td>
<td>Wednesday</td>
</tr>
<tr>
<td>3</td>
<td>.1139</td>
<td>.0123**</td>
<td>Thursday</td>
</tr>
<tr>
<td>4</td>
<td>.2358</td>
<td>.0123**</td>
<td>Friday</td>
</tr>
<tr>
<td>5</td>
<td>.2848</td>
<td>.0121**</td>
<td>Saturday</td>
</tr>
<tr>
<td>6</td>
<td>.0178</td>
<td>.0138</td>
<td>March</td>
</tr>
<tr>
<td>7</td>
<td>.0375</td>
<td>.0144**</td>
<td>April</td>
</tr>
<tr>
<td>8</td>
<td>.0233</td>
<td>.0139*</td>
<td>May</td>
</tr>
<tr>
<td>9</td>
<td>.0349</td>
<td>.0140**</td>
<td>June</td>
</tr>
<tr>
<td>10</td>
<td>.0530</td>
<td>.0139**</td>
<td>July</td>
</tr>
<tr>
<td>11</td>
<td>.0193</td>
<td>.0136</td>
<td>August</td>
</tr>
<tr>
<td>12</td>
<td>.0210</td>
<td>.0140*</td>
<td>September</td>
</tr>
<tr>
<td>13</td>
<td>.0291</td>
<td>.0137*</td>
<td>October</td>
</tr>
<tr>
<td>14</td>
<td>.1249</td>
<td>.0148**</td>
<td>December</td>
</tr>
<tr>
<td>15</td>
<td>.0931</td>
<td>.0198**</td>
<td>Tuesday or Wednesday and Pay Day</td>
</tr>
<tr>
<td>16</td>
<td>.1285</td>
<td>.0440**</td>
<td>Thursday and Mid-Month Pay Day</td>
</tr>
<tr>
<td>17</td>
<td>.1817</td>
<td>.0439**</td>
<td>Thursday and Month-End Pay Day</td>
</tr>
<tr>
<td>18</td>
<td>.0707</td>
<td>.0261**</td>
<td>Friday and Mid-Month Pay Day</td>
</tr>
<tr>
<td>19</td>
<td>.1345</td>
<td>.0263**</td>
<td>Friday and Month-End Pay Day</td>
</tr>
<tr>
<td>20</td>
<td>.0936</td>
<td>.0260**</td>
<td>Saturday and Mid-Month Pay Day</td>
</tr>
<tr>
<td>21</td>
<td>.0820</td>
<td>.0254**</td>
<td>Saturday and Month-End Pay Day</td>
</tr>
</tbody>
</table>

CONTINUED
<table>
<thead>
<tr>
<th>Coefficient Subscript $i$</th>
<th>Estimated Coefficient $\hat{\beta}_i$</th>
<th>Standard Error of $\hat{\beta}_i$</th>
<th>Dummy Variable $z_{it}$ is a 0-1 Indicator Function for:</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>-.0207</td>
<td>.0212</td>
<td>Thursday and Pay Day Occurs Earlier in Week</td>
</tr>
<tr>
<td>23</td>
<td>-.0505</td>
<td>.0194**</td>
<td>Friday and Pay Day Occurs Earlier in Week</td>
</tr>
<tr>
<td>24</td>
<td>-.0377</td>
<td>.0190*</td>
<td>Saturday and Pay Day Occurs Earlier in Week</td>
</tr>
<tr>
<td>25</td>
<td>.0481</td>
<td>.0165**</td>
<td>Tuesday and Pay Day Occurs Later in Week</td>
</tr>
<tr>
<td>26</td>
<td>.0630</td>
<td>.0171**</td>
<td>Wednesday and Pay Day Occurs Later in Week</td>
</tr>
<tr>
<td>27</td>
<td>.0546</td>
<td>.0197**</td>
<td>Thursday and Pay Day Occurs Later in Week</td>
</tr>
<tr>
<td>28</td>
<td>.1336</td>
<td>.0260**</td>
<td>Holiday Occurs Next Day (May, June, July, August, September and December only)</td>
</tr>
</tbody>
</table>

$R^2 = .5468$  
759 Observations

Durbin Watson Statistic = 2.08

Standard Error of the Regression $s = .0950$

* Significant at 10% level

** Significant at 1% level

**NOTE**: Mid-month and month-end pay days occur on the 15th or month-end, respectively, unless that day is a holiday or on a weekend, in which case the pay day is shifted ahead. Monday pay days are recorded on Tuesday. A Saturday pay day is a day after a Friday pay day.
Because of Christmas, December has by far the largest cash demand, but the other winter months have light demand. Demand in April is high, likely because the branch is located near a university where cash demand would increase as students and faculty prepare for vacations.

As to be expected, there is a sharp increase in the demand for cash on monthly and semi-monthly pay dates. For Thursday and Friday pay dates, the cash demand was much heavier at month-end than at mid-month, probably because many people are paid only at the end of the month. *A priori* it was anticipated that on Fridays and Saturdays following a pay day there would be an increased demand for cash since people would wait until these convenient days to cash their cheques. However, the results indicate that people generally cash their cheques right on a pay day or perhaps even cash post-dated cheques before pay day, since there is a significant increase in demand on the days preceding a pay day. As a result, there is actually a significant decrease in demand on Fridays and Saturdays following a pay day.

As to be expected, holidays precipitated a large increase in the demand for cash on the day prior to the holiday. This effect was not observed for Easter, Thanksgiving or Remembrance day holidays, however, so these holidays were excluded from consideration by only considering holidays in May, June, July, August, September and December.

The Durbin-Watson statistic shows no significant autocorrelation of the errors. Furthermore, a Box-Jenkins identification of the residuals was performed and, with only one exception, all the autocorrelations and partial autocorrelations out to 60 day lags were within a 95% confidence interval about 0. In particular, for the 5 day (1 week) lag, the autocorrelation was .01 and the partial autocorrelation...
was 0. At the 5 day (2 week) lag the autocorrelation and partial autocorrelation were both -.02. This indicates that there is no observable effect on cash demand as a result of bi-weekly pay periods (as opposed to the semi-monthly pay periods used in the regression). Furthermore, the lack of any significant autocorrelation structure indicates that the first two types of data discrepancies either are insignificant or occur at random lags.

There was no evidence of heteroscedasticity, except for an increased variance of the residuals around Christmas. A plot of squared residuals against squared estimates, however, indicated there was no general reason to believe that residual variance increases with the general level of cash demand.

The data period was divided into two equal sub-periods and a Chow test was performed on the null hypothesis that the estimated coefficients were unchanged from one period to the next. The null hypothesis was barely rejected at the 5% level, but not at the 1% level. On the whole, it appears best to use the estimates based on all of the data and avoid any attempts to forecast a trend for each coefficient. Polynomial time trending is likely the best way of modelling a trend in the coefficients, but as discussed earlier, polynomial projections can be very unreliable.

Plotting the residuals of the regression against a cumulative normal distribution revealed a fat-tailed distribution, somewhat skewed to the right. After pooling the residuals into 20 classes in the ±3.5 standard error range and performing a chi-square goodness of fit test, the normality hypothesis had to be rejected at the 1% level. Similarly a Kolmogorov-Smirnov goodness of fit test implied rejection of the normality
hypothesis at the 1% level. This is largely a result of the data errors that arose from the bookkeeping discrepancies.

The same chi-square and Kolmogorov-Smirnov tests implied rejection (at the 1% level) of the hypotheses that the residuals were distributed as Poisson or binomial.

In order to test the forecasting ability of the model, the first two years of data were used to re-estimate the coefficients of the model, using exactly the same methods as discussed for the full set of data. This model was used to forecast nominal cash flows in the third year. It was assumed that CPI data is available with a 50 working day lag, so in forecasting the cash demand in the third year, instead of using the actual CPI, the CPI was estimated by a linear projection of the CPI trend over the 250 working days prior to the latest available CPI figures. Thus, if $p_t$ is the published CPI level for the month which includes day $t$, the CPI estimate is:

$$\hat{p}_t = p_{t-50} \left( 1 + \frac{1}{5} \left( \frac{p_{t-50} - p_{t-300}}{p_{t-300}} \right) \right).$$

This CPI estimate is used in the usual manner to forecast cash demand in the final year.

Theil [1966, pp. 26-36] has suggested several techniques for analyzing the prediction errors of a forecast method. Let

$$\hat{y}_t = \text{prediction for day } t \ (t=1,\cdots,n)$$

$$y_t = \text{realized value for day } t \ (t=1,\cdots,n).$$
Then the mean square error of prediction is

\[
MSEP = \frac{1}{n} \Sigma (\hat{y}_t - y_t)^2
\]

Letting \( \hat{y} = \frac{1}{n} \Sigma y_t \), \( \bar{y} = \frac{1}{n} \Sigma y_t \)

and

\[
s_p^2 = \frac{1}{n} \Sigma (\hat{y}_t - \hat{y})^2, \quad s_A^2 = \frac{1}{n} \Sigma (y_t - \bar{y})^2
\]

and

\[
r = \frac{1}{n} \frac{\Sigma (y_t - \bar{y})(y_t - \hat{y})}{s_p s_A}
\]

the following decomposition holds:

\[
MSEP = (\hat{y} - \bar{y})^2 + (s_p - s_A)^2 + 2(1 - r) s_p s_A
\]

Let

\[
u^m = \frac{(\hat{y} - \bar{y})^2}{MSEP}
\]

\[
u^s = \frac{(s_p - s_A)^2}{MSEP}, \quad \text{and}
\]

\[
u^c = \frac{2(1 - r) s_p s_A}{MSEP}
\]

These are termed inequality proportions, and \(u^m + u^s + u^c = 1\). \(u^m\) is the proportion of \(MSEP\) that results from bias in the general forecast level, and for a good forecast, it should be small. \(u^s\) is the proportion due to a poor forecast of the variance of the series, and should also be minimized. \(u^c\) is the variation due to unequal co-variation of forecasts and realizations, and is something over which the forecaster has little control (except,
perhaps, to use more explanatory variables in his forecasts). For the forecasts of cash flows in the last year,

\[
\text{MSEP} = 1.729 \times 10^8 \\
\mu^m = 0.0045 \\
\mu^s = 0.0628 \\
\mu^c = 0.9327
\]

These results indicate that virtually no error arose because of bias in the general trend level estimates, and little error results from incorrectly estimating the level of variation. On the other hand, 93% of the error is due to different covariation (essentially, noise). These results suggest that the model performance could not have been improved without the use of more explanatory variables.

Another breakdown of the errors is:

\[
\text{MSEP} = (\hat{y} - \bar{y})^2 + (s_p - rs_A)^2 + (1 - r) s_A^2
\]

Let

\[
\mu^R = \frac{(s_p - rs_A)^2}{\text{MSEP}} \\
\mu^D = \frac{(1 - r^2)s_A^2}{\text{MSEP}}
\]

Then \(\mu^m + \mu^R + \mu^D = 1\). If one regresses

\[
y_t = a + \hat{y}_t + e_t \quad (t=1,\ldots,n)
\]
then perfect forecasts correspond to $a = e_t = 0$ and $b = 1$. Clearly, $e_t \neq 0$ in general, so this is an irreducible component of forecast error. The proportion of MSEP due to this component is $u^D$. On the other hand, if $a$ differs greatly from zero, there is substantial bias in the forecasts and $u^m$ is large. If the regression slope $b$ differs greatly from 1, $u^R$ is large. For a good forecast $u^m$ and $u^R$ are small. We have already seen that $u^m$ is small, and $u^R = 0.0367$ which is also small. The mean absolute error of prediction was $9084$ while the mean error of prediction is only $883$. This means that on average the model over-estimated the general trend level in the last year by less than $1000$/day.

As a benchmark we may consider a naive forecast model that simply extrapolates nominal cash flows linearly from the first two years to the last year. For this model, the MSEP is $3.000 \times 10^8$. As an analogue to Theil's second U-statistic (Theil [1966, p. 28]), we may consider the ratio of the MSEP of the full model to the MSEP of the benchmark. (The smaller the ratio, the better the forecast.) This ratio is .58, indicating that the full model does indeed outperform naive extrapolation (which, of course, has a ratio of 1). One might have expected the ratio to be smaller, but the high level of white noise (i.e. low $R^2$ in the regressions), which affects both the model and the benchmark, prevents the ratio from approaching zero.

In summary the model performs quite well in tests of predictive ability and in particular, the estimation of the trend curve appears to be good.
3.5 Summary

In this chapter, we have studied the results of the estimation of the two main models of cash demand. The Box-Jenkins time series technique was found to be inapplicable due to non-stationarity of the cash flow data due to a growth trend and the presence of significant autocorrelation information at very long lags.

A satisfactory multiplicative model for cash demand was obtained in two parts. First, a trend for real cash demand that incorporated a capacity constraint was estimated. Then daily cash flows deflated by this trend were regressed on dummy variables indicating days of the week, months of the year, paydays and holidays. The residuals of this regression were too fat-tailed to be normally distributed. This resulted in part from bookkeeping discrepancies which created pairs of reversed residuals of large magnitude.

The second (multiplicative) model was tested by fitting it on the first two years of data and studying the prediction errors that resulted in the third year. Most of the mean squared prediction error was a result of random noise, and very little was due to bias or incorrectly estimated scaling of the variation in cash flows. In short, the model is good for prediction purposes.

The two-part multiplicative model will be used in the cash order algorithm to predict cash demand and provide a probability distribution for the prediction errors, which will be used in the determination of precautionary cash buffer sizes.
The treasury is a balance of vault cash maintained by the head teller from which the tellers withdraw cash as they need it. Cash parcels would go first to the treasury and then to individual tellers as needed.

Inverse autocorrelations were developed by Cleveland [1972] and are dual to the autocorrelations in the sense that reversing the roles of the AR and MA parameters also reverses the roles of the autocorrelations and inverse autocorrelations.
Chapter 4

THE CASH MANAGEMENT PROBLEM IN GENERAL

4.1 Introduction

In this chapter we shall discuss the general considerations involving the management of till cash.

Section 4.2 discusses some of the aspects that make till cash management at the branch involved in this study different from other cash management problems and inventory problems in general. Important points raised there include the non-stationarity of cash demand and the variable lag times for delivery of cash orders.

Section 4.3 presents some cash management models that have been built on a steady-state framework, in which the basic structure of the problem does not change over time. Steady state models include those by Baumol, Miller-Orr and Eppen-Fama. Models that give exact solutions for their optimal parameters are generally steady state.

Section 4.4 casts the cash management problem as a general dynamic programming model which does not require any steady state assumption. Optimal solutions are hard to calculate from this model, but the model is examined in Section 4.5 to find which approximations can be made to simplify computation with a minimal loss of optimality. One of these approximations is used in Chapter 5 to develop a cash order algorithm.
4.2 The Problem and Its Environment

In this section, the cash management problem is defined and the general context of the problem is presented.

The objective of cash management is to select a policy that will minimize the present value of the expected costs of maintaining an inventory of cash. These costs include the costs of ordering cash, as well as the opportunity costs of assets held in the form of non-interest earning cash. This problem may be cast in the form of a recourse problem or a chance constrained problem. In the recourse problem, the objective function also includes a penalty for cash-outs (or cash shortages). In the chance constrained problem, the objective function only includes order costs and opportunity costs, but the objective is minimized subject to the constraint of holding the probability of a cash-out below some predetermined level $\alpha$. A more specific discussion of the recourse and chance constrained formulations will be presented in Section 4.4.

The most important features of the problem are as follows:

1. The demand for cash is a non-stationary stochastic process for which forecasts and the distribution of residual error are available (by the analysis of Chapter 3).

2. The distribution of forecast errors is skewed to the right and too fat-tailed to be normal, nor is the distribution Poisson or binomial.

3. Cash flows are generally net outflows (only 30 net inflows in 767 daily observations) so that virtually no loss of optimality is involved by considering only non-negative transfers or orders.
4. The lag time between placement and receipt of orders is two to four working days (depending on the day of the week when the order is placed), and the length of this lag is of the same order of magnitude as the expected optimal time between re-order points.

5. Cash can be ordered daily, but the interaction of variable delivery lags makes ordering cash on some days irrational. That is, it is irrational to order if the order will arrive no earlier than will an order placed a day or two later when more information on cash balances is available.

6. There are two assets: non-interest earning cash and short-term, liquid interest earning assets. The extension to more assets of varying terms to maturity is beyond the scope of this study, since it would entail a comprehensive asset and liability management problem. The credit union offers interest on the minimum daily balance for some accounts, so it is reasonable to suppose that the interest earning asset is just a deposit in such an account, thereby avoiding the need to consider more than two assets. The opportunity cost of cash is the daily yield \( i \) on the interest earning asset.

7. Cash orders are only available for use the day after they are received, since cash parcels are counted after hours. Opportunity costs of cash accrue as of the actual delivery time.

8. Opportunity costs of cash are incurred over weekends even though other parts of the model, such as cash demand and the ability to take deliveries, are not operational then. That is, there is a discontinuity in the demand process every weekend.
9. The cost of placing a cash order $Q$ is the piecewise linear function

$$F(Q) = \begin{cases} a + bQ & \text{if } Q > 0 \\ 0 & \text{if } Q = 0 \end{cases}$$

10. A cash-out occurs when the cash balance falls below a pre-specified balance $\bar{x} > 0$. It may be that $\bar{x} > 0$, reflecting the fact that the branch is effectively out of cash if not all tellers have an adequate supply of all denominations of currency and coin.

11. Emergency cash can be obtained on short notice by soliciting deposits from certain local merchants, however there are certain intangible costs associated with the risk of transporting the cash without an armoured car.

The cash order models presented in this chapter and the next should be evaluated on the basis of how well they will perform in this environment.

4.3 Steady State Models

In this section, we will review some of the cash management models which assume a steady unchanging inventory problem over time. For these models, the optimal order policy can be cast in terms of parameters that are constant over time. The steady state assumptions are generally violated by the following features of the problem: non-stationarity of the demand for cash, variable delivery lags for cash orders, the irrationality
of ordering cash on certain days of the week and the discontinuity of the cash demand process on weekends. Asserting that the steady state assumptions are generally violated by various features of the problem requires two qualifications.

First, it is often possible to enlarge the state space of the problem so that steady state assumptions do hold on the enlarged state space. For example, in Section 4.4, a model, which was developed by Iglehart and Karlin for non-stationary demand, is discussed, where stationarity is induced by enlarging the state space from the set of cash demands to the Cartesian product of the set of cash demands and demand states. (The demand states are defined so that cash demand is stationary within them.) Such refinements generally leave solutions that are too complicated to be computed.

Second, we should observe that a steady state model may be fitted even when its assumptions are violated, and then the most appropriate way to measure the extent to which the assumptions are violated is by the amount of increased cost caused by the resulting sub-optimality. We shall evaluate the steady state models in this framework.

One of the best-known cash management models is by Baumol [1952]. He assumes a non-stochastic, constant demand for cash over time, and his model can readily be modified to allow for a delivery lag period for orders, since the cash requirement during the delivery period is assumed to be known with certainty. Let \( y \) be the daily demand for cash, \( b \) the brokerage fee or fixed order cost, \( i \) the daily opportunity interest cost of idle cash balances, and \( Q^* \) the optimal order size. By elementary calculus, one can verify that \( Q^* = \sqrt{2b y/i} \). This illustrates the classical trade-off
between order costs and holding (interest) costs in inventory theory: if (fixed) order costs are high relative to the interest cost, large orders should be made, but relatively infrequently; if order costs are relatively low, small orders should be made, but more often. This is a control limit policy of simple $(s,S)$ form, where the trip level $s$ is the demand for cash during the delivery lag period and $S = Q^*$ is the return point. That is, when the cash balance falls below $s$, order an amount $S$. This gives a pure transactions demand for cash.

If there is a need for precautionary balances, one simply has to add the size of the precautionary buffer to $s$ and $S$. Archer [1966] discusses some of the important considerations in the determination of precautionary cash balances. In effect, he points out that if $\tilde{y}_d$ is the stochastic demand for cash during the delivery lag period and management desires to risk a cash-out only with probability $\alpha$, then $s$ should be the $1 - \alpha$ point of the cumulative probability distribution for $\tilde{y}_d$. Then the re-order point $s$ includes provision for both precautionary and transactions demand for cash during the delivery lag period. For example, if the daily demands for cash are independent and identically distributed $N(\tilde{y}, \sigma^2)$ and the order lag period is $n$ days, and $N_{1-\alpha}$ is the one-tailed $1 - \alpha$ point of the standard normal distribution, then $\tilde{y}_d$ is distributed $N(n\tilde{y}, n\sigma^2)$ and $s = n\tilde{y} + N_{1-\alpha} \sqrt{n} \sigma$ so that $n\tilde{y}$ is the transactions demand and $N_{1-\alpha} \sqrt{n} \sigma$ is the precautionary demand for cash.

Since the Baumol model was developed in the context of a constant demand for cash, it is inconsistent to add precautionary requirements to the control limit, while disregarding the fact that the stochastic demand also makes the objective function stochastic.
Miller and Orr [1966] have developed a steady state model which assumes purely stochastic cash flows. Their model is based on the following assumptions:

1. For some small fraction \( \frac{1}{t} \) of the working day, the cash balance will increase by \( +m \) dollars with probability \( p \) and decrease by \( -m \) dollars with probability \( 1-p \). These Bernoulli trials are the only exogenous changes that can occur in the cash balance.

2. There is a two asset environment: cash and an instrument yielding daily interest at the rate \( i \).

3. Transfers between the asset accounts may be made instantaneously in either direction, for a fixed cost \( b \).

4. The cash balance must not fall below zero.

They appeal to arguments by Karlin [1958] that an appropriate objective function will be of simple form, which in their case is an \((h,z)\) control limit model where the cash balance is returned to the point \( z \) whenever it exceeds \( h \) or falls below 0.

If \( p = \frac{1}{2} \), the stochastic process of cumulative cash flows has no drift because inflows and outflows are equally probable. In this case the optimal values of \( z \) and \( h \) are given by

\[
z = \left( \frac{3bm^2t}{4i} \right)^{1/3}
\]

\[
h = 3z
\]

They also show that \( m^2t \) is the daily variance of cash flows \((\sigma^2)\), so that, in this no-drift case, the optimal control parameters can be evaluated in terms of the known quantities \( b \), \( i \) and \( \sigma^2 \), and no specific assumptions are needed about the parameters \( m \) and \( t \).
Since there is a pronounced downward drift in cash balances in our case, we have $0 < p < \frac{1}{2}$. In this case, Miller and Orr present complicated formulae for the optimal values of $h$ and $z$ which specifically depend on $m$ and $t$, and cannot be summarized in terms of $\sigma^2$. Implementation of such a model then requires evaluation of the parameters $p$, $m$ and $t$, which together describe a binomial distribution of cash flows. For small values of $p$, such as in our case, this can be approximated by a Poisson distribution, so that, in principle, such a model could be made operational, by fitting a Poisson distribution to the observed cash flows.

A more serious difficulty in implementing the Miller-Orr model in our problem results from the fact that the demand for till cash is non-stationary, so that the parameters $m$ and $p$ vary over time. They point out that if the seasonality of demand is of a long term nature only, $m$ and $p$ can be varied seasonally to obtain a good approximation to optimal behaviour. But if the period of the seasonal process is approximately the same length as a typical period between order and reorder, it is not clear that the modified model will approach optimal behaviour. Indeed, Orr [1970] makes this specific caveat:

... if adjustment transfers are made at three day intervals on average, while the period of pronounced draft is a month or more, there is no large problem ... if ... cash flows are characterized by extreme drift over short periods, then the steady state mode of analysis may be inappropriate: transient effects may dominate the longer-term movements that are well-handled in the steady-state approach.

In Section 4.4, it will become apparent that there are really two problems that arise from the short-term non-stationarity of cash flows. First, there is the problem that Orr mentions of sub-optimality arising
from the fact that a myopic order policy improperly accounts for the costs that occur after \( m \) and \( p \) are changed and leaves transient effects when \( m \) and \( p \) are changed. The second problem arises because selection of the parameters \( m \) and \( p \) requires some description of the demand process that will be in effect until the next order arises. But the time of the second order depends on the size of the first order, which in turn depends on the values of \( m \) and \( p \) used for the first order. This is a circular problem, for which there is no obvious solution.

Another major violation of the Miller-Orr axioms is the fact that, not only is there a major order delivery lag time (almost as long as the life of a typical order), but the length of this lag changes according to the day of the week. There is no obvious way to adjust the Miller-Orr algorithm to take account of this or the fact that it is irrational to order cash on some days of the week, either.

These problems make it highly unlikely that an operational version of the Miller-Orr model can be applied to the till cash management problem. There are two other minor violations of the Miller-Orr assumptions, which do not undermine the basic feasibility of the model. The first arises from the discontinuities of the problem on weekends: cash demand stops, while opportunity costs continue. The second arises from the question of whether the cash demand process is really Bernoullian. For sufficiently small divisions of the day, the Bernoulli process approximates various continuous processes for various values of \( m \) and \( p \), however, so this is not likely to be a serious problem. In a later paper, Miller and Orr [1968] provide a rather substantial defence for the approximation of cash demand by a Bernoulli process.
Many other authors have considered the cash management problem in a steady state form. Eppen and Fama [1968] assume that holding and penalty costs are proportional to the size of cash balances, transfer costs are linear functions of the amount transferred, cash deliveries are instantaneous and that the sequence of cash balances forms a discrete Markovian stochastic process. They use a simple for \((u, U; D, d)\) inventory policy whereby the cash balance is moved up to \(U\) if it falls below \(u\), down to \(D\) if it rises above \(d\), and no action is taken if the balance is between \(u\) and \(d\). They set up a linear program to minimize the present value of expected costs over an infinite horizon.

Constantinides [1976] models the cash management problem in a continuous time framework with a Wiener process for demand. (A Wiener process has stationary independent increments, and at each point in time the cash balance is normally distributed.) He assumes proportional penalty and holding costs, and fixed transfer costs. He solves for the optimal parameters of a \((u, U; D, d)\) inventory policy by minimizing

\[
\lim_{N \to \infty} N^{-1} \left( EC_N \right)
\]

where \(C_N\) is the total stochastic inventory cost incurred out to a horizon \(N\).

All of these papers yield specific solutions for inventory control parameters, but only by relying on the stationarity of the underlying cash demand process. The non-stationarity of the demand process prevents these models from being applied in precisely the same way it prevents the Miller-Orr model from being applied in our problem. Similarly the varying delivery lag times and discontinuities undermine the steady state assumptions of all models discussed in this chapter.
The price paid to obtain a simple steady state solution appears to be too high in terms of the resulting sub-optimality of decisions. In the next section, the problem will be cast in a more general dynamic programming framework, which has a complementary set of attributes: the solutions are optimal, but too complicated to compute.

4.4 Dynamic Programming Formulation

Cash management problems are really just special inventory theory problems. The two main features that distinguish most cash management problems from ordinary inventory problems are the assumption of zero lead time for transactions and the possibility of increases in cash balances as well as decreases. We have seen that our problem differs from the standard cash management problem in these two respects, so perhaps it is best to consider the till cash management problem as a more general inventory theory problem. This generality will also allow us to relax the assumption in the previous section that the problem has a steady state nature.

The general solution to inventory theory problems is often cast in a dynamic programming framework. A general discussion of dynamic programming is given, for example, by Ziemba [1975] and specific formulations for cash management are given by Eppen and Fama [1969], Daellenbach and Archer [1969] and many other authors. The model can be formulated as either a recourse model (with a penalty function for cash-outs) or as a chance constrained model (constraining the risk of a cash-out), and the following discussion derives both.

To be consistent with the statistical portion of the paper we shall enumerate working days, not calendar days. Thus if t is a Saturday,
t+1 is a Tuesday. Also, to remain consistent with the rest of the paper, we shall enumerate time in a forward rather than backward fashion.

Let $x_t$ be the opening cash balance on working day $t$, $\gamma$ be the one-day discount rate ($0 < \gamma \leq 1$), $F(Q)$ be the cost of ordering $\$Q \geq 0$, $H_t(x_t)$ the opportunity cost of the opening cash balance $x_t \geq 0$ on day $t$, $p(x_t)$ be the penalty cost for a cash-out of size $x_t < 0$, and $y_t$ be the stochastic cash flow on day $t$ with cumulative distribution function $\phi_t(y_t)$. Define the cost of holding the opening cash balance $x_t$ on day $t$ as

$$C_t(x_t) = \begin{cases} H_t(x_t) & \text{if } x_t \geq 0 \\ p(x_t) & \text{if } x_t < 0 \end{cases}.$$

Note that the opportunity cost of the cash balance $H_t$ depends on $t$ to the extent that it is higher on the day after a weekend, in order to fully reflect the opportunity cost incurred over the weekend.

Let $g_t(Q_t, x_t)$ be the expected discounted cost as of time $t$ of current and future order policy if the current (opening) cash balance is $x_t$ and an order of size $Q_t$ is placed at time $t$ (temporarily assuming a zero delivery lag) and an optimal policy is followed for all subsequent orders. Then we have a dynamic program in the following recursive form:

$$g_t(Q_t, x_t) = F(Q_t) + \gamma \int_{-\infty}^{\infty} \left\{ \min_{Q_{t+1} \geq 0} g_{t+1}(Q_{t+1}, x_t - y + Q_t) \right\} d\phi_t(y)$$

$$+ \gamma \int_{-\infty}^{\infty} C_{t+1}(x_t - y + Q_t) d\phi_t(y). \quad (4.4.1)$$
That is, the expected cost at time \( t \) is the sum of the order cost, the discounted expected optimal cost at time \( t+1 \), and the discounted expected holding cost of the opening balance at time \( t+1 \).

The optimal inventory policy at time \( t \) is to select an order \( Q_t^* > 0 \) such that

\[
g_t(Q_t^*, x_t) = \min_{Q_t} g_t(Q_t, x_t)
\]

However, the recursive relation (4.4.1) does not specify any values of \( g_t \) unless we give it a starting value for some time \( T \). We may either minimize expected costs for arbitrarily large horizons (with \( \gamma < 1 \) to ensure convergence), or we may minimize total expected costs out to some fixed horizon \( T \). In the latter case one may set \( \gamma = 1 \) (no discounting) and initialize at the horizon \( T \) by setting \( g_T(Q_T, x_T) = 0 \). Then \( T^{-1} g_t(Q_t^*, x_t) \) can be computed recursively by means of (4.4.1) and is, approximately, the long run average cost of an optimal cash management policy.

In order to generalize the problem to allow for varying delivery lag times, we may extend a formulation provided by Bellman et al. [1955, p. 87] by enlarging the state space to include historical but unreceived orders. Define the order history \( Q_t = (Q_{t-1}, Q_{t-2}, Q_{t-3}, \ldots) \). Define the delivery lag \( L(t) > 0 \) so that an order placed on day \( t \) is received on day \( t + L(t) - 1 \), counted after hours and placed into use on day \( t + L(t) \).

The total of past orders that will become available for use on day \( t \) is

\[
R(Q_t) = \sum_{j=1}^{\infty} Q_{t-j} \delta L(t-j), j
\]  

(4.4.2)
where \[ \delta_{ij} = \begin{cases} 0 & \text{if } i+j \\ 1 & \text{if } i=k \end{cases} \]

In our application, the maximum delivery time is 4 days so \(0 < L(t) \leq 4\) and the sum (4.4.2) has only 4 terms. Note that an order placed at time \(j\) enters \(R(Q_t)\) only for \(t = j + L(j)\). We now require a convention regarding the opening cash balance \(x_t\). We shall say \(x_t\) includes the orders due to be used on day \(t\). That is, \(x_t = x_{t-1} - y_{t-1} + R(Q_t)\). This notation is convenient because holding costs apply to the closing balance of day \(t-1\) as well as the orders that are received and counted that night, for use the next day. We shall also adopt the convention that \(Q_t = 0\) whenever \(t + L(t) = t + 1 + L(t + 1)\). That is, we shall never consider ordering cash on an "irrational" day for which an order placed next day with more information will arrive just as soon.

Letting \(h_t(Q_t; Q_t, x_t)\) be the present value of the expected order policy costs given the order history \(Q_t\) and cash balance \(x_t\), when the current order is \(Q_t\) and all subsequent orders are selected optimally, we have the following analogue of (4.4.1).

\[
h_t(Q_t; Q_t, x_t) = F(Q_t) + \gamma \int_{-\infty}^{\infty} c_{t+1}(x_t - y + R(Q_{t+1})) \, d\phi_t(y) \\
+ \gamma \int_{-\infty}^{\infty} \min_{Q_{t+1} \geq 0} \left\{ h_{t+1}(Q_{t+1}; Q_{t+1}, x_t - y + R(Q_{t+1})) \right\} \, d\phi_t(y)
\]

(4.4.3)
Note that $Q_{t+1}$ on the right hand side of (4.4.3) is well-defined given $Q_t$ and $Q_t$ on the left side. As before, set $h_T(Q_T; Q_T, x_T) = 0$ for some large $T$, in order to recursively compute and minimize (4.4.3) for any given time $t$.

To formulate the model as a chance constrained model set the penalty cost $P(x_t) \equiv 0$ and define a cash-out to be the occurrence of a cash balance below $\bar{x} > 0$. Suppose management will tolerate cash-outs with a probability $\alpha$ ($0 < \alpha < 1$). To determine the day for which the cash-out constraint is to be applicable note that the cash order is not available before $t + L(t)$, so it is senseless to set a constraint regarding the cash-out risk on any earlier day. Also, for an order on day $t$, one cannot be content with constraining the risk of a cash-out on day $t + L(t)$ since that will not provide protection against a cash-out on the next day if $t + L(t) + 1 < t + L(t + 1)$, which is the earliest time an order placed on day $t + 1$ can be used. This can occur when the delivery lag times vary so that $L(t) + 1 = L(t + 1)$. Hence, a more realistic constraint is to constrain the risk of a cash-out on day $t + L(t + 1)$, which is the last day before an order placed at $t + 1$ will be available. To do this, we should perform the minimization in (4.4.3) so that for $t=1,2,3,\ldots$

$Q_t > 0$ and

$$\text{Prob} \left\{ x_t - y_t + \sum_{j=t+1}^{t+L(t+1)} (-y_j + R(Q_j)) \geq \bar{x} \right\} \geq 1 - \alpha. \quad (4.4.4)$$

If $\phi_j(y_j) \sim N(E(y_j, \sigma_j^2)$, and the $y_j$ are independent, this has the deterministic form
\[
\left[ E(y_t) + \sum_{j=t+1}^{t+L(t+1)} \left( E(y_j) - R(Q_j) \right) \right] + \sqrt{\sigma_t^2 \cdots \sigma_{t+L(t+1)}^2} N_{1-\alpha} \leq x_t - \bar{x} \quad (4.4.5)
\]

where \( N_{1-\alpha} \) is the \( 1-\alpha \) point of the standard normal distribution. Note that the term in square brackets is the expected cumulative cash flow (after adjusting for the receipt of past and current orders) between \( t \) and the day before an order placed at \( t+1 \) will be received, so it is the net transactions demand for cash. The other term on the left is the buffer of cash required to hold the probability of a cash-out below \( \alpha \), and hence is the precautionary demand for cash. By increasing the size of the current order \( Q_t \), the term in square brackets can be decreased to satisfy the inequality, unless \( t + L(t) > t + L(t+1) \) in which case an order placed on day \( t+1 \) will arrive as early as an order placed on day \( t \). In this last case, it is irrational to order money and we set \( Q_t = 0 \).

This model accommodates all the features of the problem mentioned in Section 4.2. In particular, it incorporates non-stationarity of the cash flows because \( Q_t \) depends on \( t \). The variable delivery time for orders is incorporated in \( L(t) \) and \( R(Q_t) \). The discontinuity of the model on weekends (demand stops, opportunity costs continue) is incorporated in the fact that the costs \( H_t \) and \( C_t \) depend on \( t \).

The model has, however, one major flaw. It is far too complicated to compute all of the recursive formulae (4.4.3). The order cost \( F(Q_t) \) is easy to evaluate (in our case it is a linear function for \( Q_t > 0 \) and is zero for \( Q_t = 0 \)). The second term in (4.4.3) is the expected total holding and penalty cost \( E_y (C_{t+1}(x_t - y_t + R(Q_{t+1}))) \) of the stochastic opening cash balance on day \( t+1 \), and is readily computable by numerical integration, or by simulation. The third term is the discounted expected
value of optimal future order policy given the order $Q_t$ on day $t$ (and order history $Q_{t-1}$). The integrand has no tractable representation because it is itself a solution to an optimization problem, and the character of this recursive optimization problem changes significantly with $t$, due to the non-stationarity of $y_t$, variable delivery lags and discontinuities on weekends. Since there is no steady state characterization of the recursive optimization problem, it must be re-solved at each iteration. If numerical methods are used, the number of calculations required grows exponentially with the number of iterations of (4.4.3), so that the problem is virtually insoluble for large values of $T$.

To appreciate the level of complexity introduced just by non-stationarity, we may consider a model by Iglehart and Karlin [1962] for a non-stationary inventory process. They assumed $k$ demand states with Markov transition probabilities between states. Each state has demand density $\phi_i(y)$. In our application, we could let the states $i$ be the days of the year ($k = 365$), in which case the probability transition matrix would have a 1 corresponding to the transition from day $t-1$ to $t$ and zeros elsewhere. Under some restrictive conditions (including $y_j \geq 0$ and proportional order costs only) they show that the optimal policy is characterized by $k$ critical numbers $\bar{x}_i$ such that, if the inventory falls below $\bar{x}_i$ in state $i$, one should order up to the level $\bar{x}_i$. The solution requires solving up to $k!$ transcendental equations and $(k-1)$ renewal equations. Even for $k = 20$ states, this is virtually impossible.

For operational purposes, the dynamic programming problem (4.4.3) cannot be solved to yield optimal inventory policy. However, it can be used as a benchmark for evaluation of operational approximations. This will be discussed in the next section.
4.5 Operational Approximations to Optimal Policies

It was pointed out in the last section that the main stumbling block to solving the dynamic programming problem (4.4.3) was the evaluation of the last term which is the present value of the expected cost of future optimal order policy given that $Q_t$ is ordered on day $t$. An operation solution of the problem requires some sort of approximation for this term. There are two main types of approximation that can be done.

The first type of approximation is to avoid performing the minimization in (4.4.3), and instead parameterize an heuristic order policy and simulate the operation of the model with various parameter assignments, selecting those parameter values which minimize the long run costs. The problem is to determine what sort of heuristic policy to use, bearing in mind that a more realistic policy has more parameters, while the number of computer runs required to find an optimal parameterization grows exponentially with the number of parameters. For example one could simulate the use of a simple $(s, S)$ order policy to find which values of the parameters $s$ and $S$ order policy to find which values of the parameters $s$ and $S$ yield the least long run cost. However, the problem is not a steady problem and the optimal values of the parameters for an order arriving before a Friday pay day in December will certainly not be the same as the optimal values of the parameters for an order arriving on a Tuesday in January with the next pay day two weeks away. Thus it would be better to use a family $\{(s_i, S_i): i \in I\}$ of order policies where the state index $i$ depends on the sequence of expected demands over the next few days. The form of this dependence is heuristic. Furthermore, for a comprehensive range of states $I$, testing the model for $N$ different...
values of each of the \((s_1, S_1)\) would require \(N^{|I|}\) computer simulation runs which would be prohibitively expensive.

The second type of approximation involves the assumption that the costs of future optimal order policy beyond the next re-order point will be approximately the same as the costs of the best order policy before the next order point. To formalize this notion, we shall cast the dynamic program (4.4.3) in terms of the average expected cost per day of the best policy out to horizon \(T\) starting at the present time \(t=1\). We may set the daily discount rate \(\gamma\) to 1.

Condition all of the probability distributions on the event that the next optimal order occurs at time \(\tau + 1 > 1\). That is, condition on the event that \(Q_2 = \cdots = Q_\tau = 0\) and \(Q_{\tau+1} \neq 0\). (In the presence of fixed order costs it often occurs that the optimal order is to place no order at all.) Then recursive substitution of (4.4.3), noting that 
\[ F(Q_2) = \cdots = F(Q_\tau) = 0, \]
yields the following formula for long run costs per day:

\[
h_1(Q_1; Q_1, x_1) = A(\tau, Q_1) + E(\min h_{\tau+1})
\]

(4.5.1)

where

\[
A(\tau, Q_1) = F(Q_1) + \\
\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{j=1}^{T} C_{j+1}(x_1 + \sum_{i=1}^{j} (-y_i + R(Q_i + 1))) \, d\phi_1(y_1) \cdots d\phi_T(y_\tau)
\]
\[
E(\min h_{\tau+1}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \min_{Q_{\tau+1} \geq 0} h_{\tau+1}(Q_{\tau+1}; Q_{\tau+1}, x_1) + \\
\sum_{i=1}^{\tau} (-y_i + R(Q_{i+1})) d\phi_i(y_i) \cdots d\phi_{\tau}(y_{\tau})
\]

This gives expected costs out to the horizon \( T \) for the order \( Q_t \) as the sum of expected costs out to \( \tau \) (in \( A(\tau, Q_1) \)) and the expected costs from \( \tau+1 \) to \( T \) (in \( E(\min h_{\tau+1}) \)). Note that \( \tau \) is a function of the cash flows \( y_t \), so that conditioning on \( \tau \) means that the distribution functions \( \phi_t(y_t) \) are conditioned on \( \tau \). For simplicity this dependence is not indicated explicitly in the notation of (4.5.1), since no confusion will arise by this. The first term, \( A(\tau, Q_1) \), can be evaluated by simulating the cash flows. The second term, \( E(\min h_{\tau+1}) \), is much more difficult to evaluate. A reasonable approximation for the second term is \( \frac{T-\tau}{T} h_1(Q^*_i; Q_1, x_1) \), where \( Q_1^* \) is the optimal order at time \( t = 1 \). This is the expected cost for the whole period (given \( \tau \)) multiplied by the factor \( (T-\tau)/T \) to adjust for the shorter time to the horizon. That is, we approximate the expected daily cost of optimal policy over the period \( \tau+1 \) to \( T \) with the expected daily cost of optimal policy over the whole period \( 1 \) to \( T \). Now, suppose \( Q_1 \) is close to \( Q_1^* \), so that we can use \( h_1(Q_1; Q_1, x_1) \) in place of \( h_1(Q_i; Q_1, x_1) \). Then we can approximate (4.5.1) as

\[
h_1(Q_1; Q_1, x_1) = A(\tau, Q_1) + \frac{T-\tau}{T} h_1(Q^*_i; Q_1, x_1)
\]

or

\[
\tau h_1(Q_1; Q_1, x_1) = T A(\tau, Q_1)
\]

(4.5.2)
Thus, the average cost per day to the horizon $T$ is approximately $\tau^{-1} A(\tau, Q_i)$, conditional on the next re-order occurring $\tau$ days from now, where $Q_i$ is a near-optimal order. Now compute the expectation with respect to the random variable $\tau$ to obtain the following approximation to the expected average cost per day objective function:

$$E_{\tau}(\tau^{-1} A(\tau, Q_i)) \quad \text{(4.5.3)}$$

A simulation technique for evaluating the objective function, for a given value of $Q_i$, is presented in Section 5.6. A grid search can be used to select $Q^*$. Note that as $Q$ approaches $Q^*$ in the grid search, the quality of the approximation (4.5.2) improves because $h_1(Q_i; Q_i, x_1)$ approaches $h_1(Q^*_i; Q_i, x_1)$. If the model is cast in a chance constraint framework, the grid search should be performed only over values of $Q_i$ that satisfy the chance constraint (4.4.4) or (4.4.5).

Note that the approximation of $E(\min h_{\tau+1})$ is made by assuming that the structure of demand after the placing of the next order is approximately the same as the structure before the next order is placed. This is essentially a steady state assumption, so it is instructive to consider how a steady state model like that of Miller-Orr would fit into this framework. The Miller-Orr model gives optimal expected costs per day (assuming steady state parameters) which could be used in place of $E_{\tau}(\tau^{-1} A(\tau, Q_i^*))$. However, by examining (4.5.1) and the definition of $A(\tau, Q_1)$ we see that it depends on the distribution of cash demands out to the reorder point $\tau+1$. In order to use Miller-Orr, we would require a summary of these distributions in order to evaluate the steady state
parameters. But such a summary requires knowledge of \( \tau \). That is, if a re-order is required before a holiday or some other day of heavy demand, the summary should not forecast such heavy cash requirements as if the re-order is to be received after the holiday. In order to evaluate \( \tau \) for the Miller-Orr model, we would have to select \( Q_i \) and then perform some sort of simulation of the cash demand to determine the next order point. But the value for \( Q_i \) depends on the summary parameters derived from \( \tau \). Thus, some sort of iteration back and forth between values of \( \tau \) and values of \( Q_i \) would be required to evaluate \( E_{\tau}(\tau^{-1} A(\tau, Q^*)) \). It seems simpler to evaluate this expectation directly by simulation. In this way, we would be making a steady state approximation like that required to invoke a Miller-Orr (or any other) steady state solution, but we are choosing the steady state parameters in a more meaningful and simpler way (by simply conditioning on \( \tau \)) than one could if selecting Miller-Orr parameters by some sort of iteration process.

4.6 Summary

In Section 4.2 we examined several of the distinguishing features of the till cash management problem, and foremost amongst these were the non-stationarity of the data, the variable delivery lag times for orders and the discontinuity of the cash flow/opportunity cost relationship on weekends.

In Section 4.3 several steady state models for cash management were considered, including those of Baumol, Miller and Orr, and Eppen and Fama. It was pointed out that the steady state assumptions of these models were too seriously violated to yield good solutions in this content.
In Section 4.4, a dynamic programming model for cash management was presented. Although the dynamic programming solution is virtually impossible to compute, it was pointed out that certain approximations allow the model to be presented as one of minimizing expected average costs per day over the life of the current order. This is a steady state approximation to the problem.

In the next chapter an algorithm, based on the material in Section 4.4, is presented to select orders with the objective of minimizing expected costs per day out to the first re-order point.
FOOTNOTES TO CHAPTER 4

1Daniel Orr [1970, p. 82].
Chapter 5

A STOCHASTIC PROGRAMMING MODEL TO ORDER TILL CASH

5.1 Introduction

In this chapter the stochastic programming model to determine near-optimal till cash orders is developed. Several technical considerations are discussed in Section 5.2. The main finding there is the contention that the chance constrained model is more appropriate than the recourse model. Section 5.3 develops an approximation to the deterministic equivalent of the chance constraint. A method of evaluating the value of objective function by simulation for any order size $Q$ is discussed in Section 5.4. The objective function in Section 5.4 is more complicated than the one presented in Chapter 4, and the dynamic program underlying this variant of the objective function is presented in the Appendix. The actual cash order algorithm is presented in Section 5.5 and is tested on the three years of historical cash flow data in Section 5.6. This allows evaluation of the effective insurance premium, in terms of annual operating costs, that is required to constrain the risk of a cash-out to any given level. The model performance is also compared to the historical performance by management over the same data period.
5.2 Preliminary Considerations

In order to develop a model based on the discussion in the previous chapter, two technical decisions have to be made. One must choose between recourse and chance constrained stochastic optimization and the functional form of the order costs and holding costs must be considered. These points will be discussed in this section.

As discussed in Sections 4.2 and 4.4 of the previous chapter, the recourse formulation of the dynamic programming model uses the penalty function $P(x_t)$ to encourage the holding of adequate cash balances that will reduce the risk of cash-outs. The chance constrained formulation eliminates the penalty function but adds a constraint to the selection of the $Q_t$'s which reduces the risk of a cash-out to some level set by management. The chance constraint has a deterministic equivalent such as (4.4.4) or (4.4.5). Both formulations require a rather precise knowledge of the tails of the distributions of the daily cash flows since cash-outs occur only for extreme cash flows. In Chapter 3 it was found that the residuals of the estimated cash flows formed a distribution that was too fat-tailed to be normal and it was also argued that a lot of the outliers could have simply been a result of bookkeeping discrepancies rather than actual cash flows. Thus, we do not have a very good knowledge of the tails of distributions of the daily cash demands. It is not clear what adjustment, if any, could be made to the tails of the distributions in the recourse model so that the estimated expected penalty cost of a given cash order policy will be a reasonable approximation of its true value. However, in Section 4.4 it was shown that for normal forecast errors the chance constraint can be expressed as a sum of net transactionary and net
precautionary requirements. The precautionary requirements are in the form of the buffer \( \sqrt{\sigma^2_{t} + \cdots + \sigma^2_{t+L(t+1)}} N_{1-\alpha} \) where \( y_j \sim N(E(y_j), \sigma_j^2) \) and the \( y_j \) are independent. A mis-specification of the tails of the distributions of the \( y_j \)'s will mis-specify the risk level \( \alpha \), but the concept of a precautionary buffer requirement expressed as a sum of money is nevertheless something which management can relate to previous experience. Thus, a manager may be a little uneasy about specifying a risk of a cash-out of 1% (especially if he understands that the distribution of the tails of the forecast errors is not known accurately), but he is more likely to feel at ease if he specifies a precautionary buffer of, say, $60,000 when the order delivery lag is 3 days. For these reasons, the problem will be formulated in a chance-constrained framework, rather than as a recourse model.

Now we may consider the form of order cost function. In Section 4.2 it was suggested that an adequate form would be

\[
F(Q_t) = \begin{cases} 
  b + aQ_t & \text{if } Q_t > 0 \\
  0 & \text{if } Q_t = 0 
\end{cases}
\]

Since only non-negative orders are considered and the long run total of the cash orders will equal the long run net cash outflow we must have

\[
\sum_{t=1}^{T} aQ_t = \sum_{t=1}^{T} y_t \text{ for large } T.
\]

Thus the contribution to long run costs by the variable order cost term is independent of order policy and without loss of generality we can assume \( a = 0 \), so that

\[
F(Q_t) = \begin{cases} 
  b & \text{if } Q_t > 0 \\
  0 & \text{if } Q_t = 0 
\end{cases}
\]

(5.2.1)
We shall also be more explicit at this point about the cost function of the opening cash balance $C_t(x_t)$. This is a piecewise linear function

$$C_t(x_t) = \begin{cases} H_t(x_t) = i x_t (D(t) - D(t-1)) & \text{if } x_t \geq 0 \\ P(x_t) = p x_t & \text{if } x_t < 0 \end{cases}$$

(5.2.2)

where $i$ is the daily interest rate of the interest earning asset, $D(t)$ is the number of calendar days after day one, and $p \geq 0$ is the penalty cost rate for negative balances ($p = 0$ for the chance constrained model). Note for example that $D(t) - D(t-1) = 3$ if day $t$ is Tuesday (to take account of weekend cash holdings) and $D(t) - D(t-1) = 1$ if day $t$ is any other working day of the week not following a holiday.

5.3 The Chance Constraint

As discussed in the last section it is most appropriate to formulate the problem in a chance-constrained framework. Although the forecast errors are too fat-tailed to be normally distributed, it was pointed out that we can model the chance constraint as though the errors were normally distributed as long as we remember that the probabilistic constraint only gives a relative, but not absolute, quantification of the risk of a cashout. Thus, we shall use the chance constraint (4.4.5) which is based on normal forecast errors.

In Section 2.3, an estimate of the variance of the daily cash flow estimate $\hat{y}_t$ given in formula (2.3.10) is
\[ \hat{d}_t^2 s^2 (I + z_t^T (Z'Z)^{-1} z_t) \]  

(5.3.1)

where \( \hat{d}_t \) is the forecast trend level, \( s^2 \) is the standard error of the regression, \( z_t \) is the vector of dummy variables for day \( t \) and \( s^2 (Z'Z)^{-1} \) is the estimated covariance matrix of the regression coefficients \( \hat{\beta} \).

For the purposes of calculating the precautionary buffer size in the chance constraint, it is a little difficult to continually compute a quadratic form involving the 28 x 28 covariance matrix. The dummy components of the vector \( z_t \) are described in Table 1 in Chapter 2 along with the regression coefficients and their standard errors. A careful examination of Table 1 indicates that at most 4 components of \( z_t \) are non-zero at any time: one each for the day of the week, month of the year, pay-day and holiday. Also, the largest standard error of the regression coefficients is .044, and the standard errors are typically .015 to .020.

Consider the following type of day where the coefficients of the non-zero dummy variables have the largest possible standard errors: a Thursday in June with a month-end pay day, just before a holiday. Then \( z_{3t} = z_{9t} = z_{17t} = z_{28t} = 1 \) and all the other dummy variables are zero. Then from the estimated covariance submatrix corresponding to the non-zero variables, we have

\[
\begin{bmatrix}
1.51 & -0.50 & -1.20 & 0.18 \\
-0.50 & 1.95 & 0.22 & -0.37 \\
-1.20 & 0.22 & 19.31 & -0.11 \\
0.18 & -0.37 & -0.11 & 6.74
\end{bmatrix}
\]

\[ s^2 z_t^T (Z'Z)^{-1} z_t = 10^{-4} (1,1,1,1) \]

\[ = 0.0026 \]
Now $s^2 = (0.095)^2 = 0.0090$. From (5.3.1) we see that the forecast variance is

$$0.0026 \hat{d}_t^2 + 0.0090 \hat{d}_t^2 = 0.0116 \hat{d}_t^2.$$  

This is an example where the first term is larger than usual (using 4 coefficients with large standard errors), yet it is relatively small compared to the total forecast variance. Since we are not using the forecast variances to compute absolute levels of cash-out risk, but only to compute approximate precautionary buffers, it is reasonable to neglect the first term which is complicated to compute and instead suppose that the forecasts $\hat{y}_t$'s have a relatively constant variance given approximately by $s^2 \hat{d}_t^2$. Thus, the deterministic equivalent of the chance constraint (4.4.5) involving the risk of a cash-out becomes

$$\hat{y}_t + \sum_{j=t+1}^{t+L(t+1)} (\hat{y}_j - R(Q_j)) + \sqrt{L(t+1)} \cdot s \hat{d}_t \cdot N_{1-\alpha} \leq x_t - \bar{x} \quad (5.3.2)$$

where $x_t$ is the opening balance on day $t$, $\bar{x}$ is the trip balance for a cash-out, $\hat{y}_j$ is the forecast of the cash flow for day $j$, $R(Q_j)$ is the total of orders to be received on day $j$, $L(t)$ is the number of days in the delivery lag for an order placed on day $t$ and $N_{1-\alpha}$ is the $1-\alpha$ point of the standard normal distribution. This constraint specifies that the probability of running out of cash just before an order placed next day $(t+1)$ is received shall be less than $\alpha$. As before, we shall only require the chance constraint to hold on days when it is rational to order cash (that is, when $t + L(t) < t + 1 + (L(t+1))$. 

Finally we observe that, without loss of generality, we may set $\tilde{x} = 0$. If management feels that it has essentially run out of cash when the cash balance is $\tilde{x} = \$5000$, say, then the $\$5000$ is not a manageable item. Any holding costs associated with $\tilde{x}$ cannot be increased or decreased by any cash management policy. Thus, when reporting cash balances $x_t$ we may suppose that $\tilde{x}$ has already been deducted, so that a cashout occurs when $x_t < 0$.

5.4 The Objective Function

The dynamic programming formulation of equations (4.4.3) and (4.5.1) is not the only dynamic programming formulation of the inventory problem - it was presented there only because it is the simplest to understand. Hence the approximate objective function (4.5.3), which is

$$E(\tau^{-1} A(\tau, Q)),$$

is not the only objective function that can be used in the problem. From the dynamic programming formulation (4.5.1), we see that $\tau^{-1} (A(\tau, Q_1))$ is the average of the order cost $F(Q_1)$ and the holding costs $C_{t+1}(x_{t+1})$ of opening balances over the period starting on the working day when the current order is placed ($t=1$) to the working day before the next order is placed ($t = \tau$).

However, the selection of the order size $Q_1$ affects holding costs over a different period in time. If cash is ordered on working day $t$, it is ready for use on day $t + L(t)$ but it actually arrives at
the branch on day \( t + L(t) - 1 \) in order to be counted after hours for use on the next day. Thus the selection of order size \( Q_1 \) affects the holding costs of the opening cash balances on days \( 1 + L(1) \) to \( t + L(t+1) \), namely \( C_{1+L(1)}(x_{1+L(1)}), \ldots, C_{t+L(t+1)}(x_{t+L(t+1)}) \). Furthermore, we indexed the holding cost function by \( t \) to account for the increased cost of holding cash balances over weekends, while \( \tau \) is a count of working days only, so it is more appropriate to average the costs over the number of calendar days rather than working days. Accordingly, let working day \( t \) be \( D(t) \) calendar days after day one. For example if day 1 is a Saturday, \( D(2) = 3 \). Since the first order is placed on day 1 and the next order is placed on day \( \tau + 1 \) (a random variable), a more reasonable objective function is the expected average daily cost (on a calendar day basis) of ordering and holding costs which are affected by the size of the order \( Q_1 \) namely

\[
E_{\tau} \left( \frac{A'(\tau, Q_1)}{D(\tau+L(\tau+1))} \right) \tag{5.4.2}
\]

where

\[
A'(\tau, Q_1) = \mathcal{F}(Q_1) + \\
\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\tau+L(\tau+1)} C_{t+L(t+1)}(x_1 + \sum_{i=2}^{t+1} (-y_{i-1} + R(Q_{i-1}))) \, d\phi_1 \cdots d\phi_{t+L(t+1)} \tag{5.4.3}
\]

(For simplicity of notation we have replaced the differential \( d\phi_i(y_i) \) with \( d\phi_i \). Recall also, that it is conditional on \( \tau \).) A dynamic program corresponding to this objective function is given in the Appendix at the
end of this chapter. Again the reader is reminded that we set $Q_t = 0$ whenever $t + L(t) = t + 1 + L(t+1)$. That is, we do not order cash if a later order will arrive on the same day.

Suppose a simulation model generates a new sequence of cash flows $\{y_t\}$ in each iteration. Then we can condition on $\{y_t\}$, noting that given $\{y_t\}$, there is a well defined $\tau^*$ such that cash will be ordered at $\tau^* + 1$. Then we can use (5.4.3) to evaluate the conditional expectation

$$E_{\tau^*} \{ A' (\tau, Q_i) \} = \frac{A' (\tau^*, Q_i)}{D(\tau^* + L(\tau^* + 1))}$$

$$F(Q_i) + \sum_{\tau^* + L(\tau^* + 1)}^{\tau^* + L(\tau^* + 1)} C_{t+1} (x_{t+2} + \sum_{i=2}^{t+1} (-y_{i-1} + R(Q_i)))$$

Now, taking the expectation with respect to $\{y_t\}$,

$$E_{\tau^*} \{ A' (\tau, Q_i) \} = E_{\{y_t\}} E_{\tau^*} \{ A' (\tau, Q_i) \}.$$  (5.4.5)

Thus (5.4.5) can be evaluated by calculating (5.4.4) for each iteration of a simulation of $\{y_t\}$ and taking the average of the resulting values which simply amounts to taking the expectation with respect to $\{y_t\}$. This can be done for any value of $Q_i$, so that the objective functions can be minimized by a grid search. This is the essence of the cash order algorithm to be presented in the next section.
5.5 The Cash Order Algorithm

In this section the formulae presented in the previous sections are used to develop an algorithm to order cash by selecting an order size to minimize the expected average costs per day (as computed by a simulation model) subject to a constraint involving the risk of incurring a cash-out. A flow-chart of the order algorithm is presented in Figure 2. The algorithm is to be applied each morning with the opening cash balance as input. As before, the days are re-numbered so that \( t = 1 \) is the current day, for which an order is being considered.

Since there are fixed costs of ordering cash and opportunity costs of holding excess cash balances, it is always best to postpone making an order on day \( t \) unless the chance constraint \((5.3.2)\) is violated with \( Q_t = 0 \). Thus the minimum acceptable order on day \( t \) is

\[
Q_{\text{MIN}}(t) = \sqrt{t(t+1)} \cdot s \cdot \hat{d}_t \cdot N_{1-\alpha} + \sum_{j=t}^{t+L(t+1)} \hat{y}_j + \bar{x} - x_t - \sum_{j=t+1}^{t+L(t)-1} R(Q_j) \tag{5.5.1}
\]

where \( s \) is the standard error of the regression in Table 1, \( \hat{d}_t \) is the estimated weekly trend level in cash flows, \( N_{1-\alpha} \) is the \( \alpha \) point of the standard normal distribution, \( \hat{y}_j \) is the forecast cash flow on day \( j \) and \( R(Q_j) \) is the total of previously made orders that are due to arrive on day \( j \). Note that \( R(Q_{t+L(t)}) = Q_t \) and \( R(Q_j) = 0 \) for \( t+L(t) < j \leq t+L(t+1) \). Thus we can omit the dates \( t+L(t) + 1, \ldots, t+L(t+1) \) in the last summation of \((5.5.1)\) and the date \( t+L(t) \) corresponds to \( Q_{\text{MIN}}(t) \).

If \( Q_{\text{MIN}}(t) > 0 \) an order is placed, unless \( 1 + L(t+1) = L(t) \), in which case an order placed on day \( t+1 \) would arrive just as soon as an order.
Start
Current day is $t = 1$

$Q_{MIN}(1) \leq 0$
or $1 + L(2) = L(1)$

Yes

Stop
$Q_1 = 0$

No

$Q = Q_{MIN}(1)$
$QBEST = Q$

$N = 1$
$EXPCOST(Q) = 0$
$BOUND = 0$

$Q = Q + QSTEP$

$t = 2$

$x_t = x_{t-1} + R(Q_t)$
$- \frac{\gamma}{t-1} - d_{t-1} * RES(1 + 758 * FRAN)$

$Q_{MIN}(t) \leq 0$
or $1 + L(t+1) = L(t)$

Yes

$t = t + 1$

No

$t = t - 1$

$\tau = t - 1$

$EXPCOST(Q) = EXPCOST(Q) + \frac{1}{NSIM} * \frac{A'(\tau, Q)}{D(\tau + 1 + L(\tau + 1))}$
$BOUND = BOUND + \frac{1}{NSIM} * \frac{A'(\tau, Q)}{D(\tau + 1 + L(\tau + 2))}$

$N = N + 1$

$N > NSIM$

Yes

$EXPCOST(Q_{BEST}) < EXPCOST(Q)$

No

$EXPCOST(Q_{BEST}) < BOUND$

No

Yes $QBEST = Q$

$Q_1 = QBEST$

No

Stop

Figure 2. The Cash Order Algorithm.
Figure 3. The Minimum Cash Order Subroutine.

\[ QMIN(t) = \sqrt{L(t+1)} \hat{s}_t N_{1-\alpha} + \sum_{j=t}^{t+L(t+1)} y_j ^{\hat{z}} + \tilde{x} - x_t - \sum_{j=t+1}^{t+L(t)-1} R(Q_j) \]
order placed on day \( t \), making it irrational to use less information and order on day \( t \).

At \( t=1 \), if an order is to be placed, the grid search starts with the minimal order \( Q = Q_{\text{MIN}}(1) \) and examines larger values, in increments of \( Q_{\text{STEP}} \). This ensures that the chance constraint is satisfied.

For each order size \( Q \), the cash flows are simulated \( N_{\text{SIM}} \) times. In each iteration the sequence \( \{x_1, x_2, x_3, \ldots\} \) is calculated by setting

\[
y_{t-1} = y_{t-1} + \hat{d}_{t-1} \ast \text{RESIDUAL}(1 + 758 \ast \text{FRAND})
\]

\[
x_t = x_{t-1} - y_{t-1} + R(Q_t)
\]

where \text{RESIDUAL} is the array of 759 unsorted residuals from the regression in Table 1, Section 3.4 and \text{FRAND} is a random number generator that gives independent uniform variables on \([0,1]\). The sequence of cash flows is continued until a re-order is required on day \( t = \tau + 1 \) (using the order criteria discussed earlier). This allows computation of the following expectation (conditioned on \{y_t\}) by formula (5.4.4):

\[
E_{\tau \mid \{y_t\}} \left[ \frac{A^\prime(\tau, Q_1)}{D(\tau + L(\tau + 1))} \right]
\]

This is then averaged into the objective function which is the unconditional expectation

\[
\text{EXPCOST} = E_\tau \left[ \frac{A^\prime(\tau, Q_1)}{D(\tau + L(\tau + 1))} \right].
\]
Having computed the objective function \( \text{EXPCOST}(Q) \) after \( \text{NSIM} \) iterations of the simulation, the objective is compared to that for the last best order \( (\text{QBEST}) \), and if better, \( \text{QBEST} \) is updated and the objective is computed for the next larger order, \( Q + \text{QSTEP} \).

If the order \( Q \) is not better than \( \text{QBEST} \), then \( \text{QBEST} \) is a local minimum of the objective function. For a given cash flow sequence \( \{y_t\} \) the numerator of (5.4.4) increases strictly with \( Q \) since larger orders cause larger cash balances with longer times to re-order. However, the denominator is integer-valued and is only non-decreasing in \( Q \). Thus, increasing \( Q \) slightly may increase the numerator but not the denominator, although increasing \( Q \) by a greater amount may cause an increase in the denominator, as well, which causes an overall decrease in the objective function. Thus local minima need not be global minima. However to obtain a lower bound for the objective function suppose that for a given \( \{y_t\} \), increasing the order from \( Q \) to \( Q' \) causes an increase in \( \tau^* \) of one working day. Then \( \frac{A'_{(\tau^*+1, Q')}}{D(\tau^*+1+L(\tau^*+2))} > \frac{A'_{(\tau^*, Q)}}{D(\tau^*+1+L(\tau^*+2))} \). That is, at best the denominator can be increased while the numerator is unchanged by the increase in order size. This lower bound is computed at each iteration of the simulation and averaged into the expectation \( \text{BOUND} \). Then if the last best order is locally optimal (i.e. \( \text{QBEST} < Q \)) it is also globally optimal if \( \text{EXPCOST}(\text{QBEST}) < \text{BOUND} \). This provides a stopping rule for the algorithm. Note that for very large orders, the total opportunity costs grow quadratically with order size (since both \( \tau \) and the average balance grow asymptotically with \( Q \)) while the denominator increases only linearly with \( Q \). Thus, a finite order is optimal and the stopping rule will stop the algorithm in a finite number of steps.
The algorithm essentially provides a variable $(s,S)$ control limit model. The decision to order is made by evaluation of the chance constraint and the decision of how much to order is determined by the simulation and grid search. Both decisions depend on the cash flow forecasts and delivery lag times so that the parameters $s$ and $S$ vary over time.

5.6 Application of the Model

In this section we shall study the performance of the model with various values for the cost and risk parameters, as well as compare the performance of the model to actual management performance.

The tests were performed on all three years of data. In order to be able to take averages of the costs over the three years, all of the data were rescaled to a constant $140,000/week trend level. That is, the cash flow $y_t$ was replaced by $\frac{140,000}{d_t} y_t$. For simplicity, a cash-out was regarded as the occurrence of a zero cash balance so that $x = 0$. All tests of the model were performed with $NSIM = 100$ iterations of the cash flow simulation for each order size and orders were selected in steps of $QSTEP = $5000 over the minimum order.

The delivery lag period $L(t)$ for an order placed on day $t$ varies according to the day of the week because the order is received from a bank which operates on a Monday to Friday work week in contrast to the Tuesday to Saturday work week of the credit union. Orders are placed in the vault on the day of a delivery and counted after hours, for use the next day. A list of the delivery lags is given in Table 2. Note that
Table 2
Cash Delivery Order Lags

<table>
<thead>
<tr>
<th>Order Placed Before Noon on Day t</th>
<th>Order Delivered on</th>
<th>Order Put to Use on</th>
<th>Lag L(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday</td>
<td>Thursday</td>
<td>Friday</td>
<td>3</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Friday</td>
<td>Saturday</td>
<td>3</td>
</tr>
<tr>
<td>Thursday</td>
<td>Tuesday</td>
<td>Wednesday</td>
<td>4</td>
</tr>
<tr>
<td>Friday</td>
<td>Tuesday</td>
<td>Wednesday</td>
<td>3</td>
</tr>
<tr>
<td>Saturday</td>
<td>Tuesday</td>
<td>Wednesday</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3
Performance of the Cash Order Model and of Management

<table>
<thead>
<tr>
<th>Fixed Order Cost $b</th>
<th>Opportunity Cost of Cash 365 x i%</th>
<th>Risk Level (Normal Errors) $\alpha$</th>
<th>Probability of a Cash-Out in Simulation</th>
<th>Number of Orders Placed</th>
<th>Number of Actual Cash-Outs</th>
<th>Average Cost Per Year ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>-</td>
<td>-</td>
<td>153</td>
<td>2</td>
<td>15,700</td>
</tr>
<tr>
<td>Model:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>.025</td>
<td>.0103</td>
<td>192</td>
<td>0</td>
<td>11,100</td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>.01</td>
<td>.0084</td>
<td>198</td>
<td>0</td>
<td>11,600</td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>.001</td>
<td>.0019</td>
<td>192</td>
<td>0</td>
<td>13,900</td>
</tr>
<tr>
<td>25</td>
<td>10</td>
<td>.001</td>
<td>.0011</td>
<td>203</td>
<td>0</td>
<td>18,900</td>
</tr>
<tr>
<td>40</td>
<td>7</td>
<td>.001</td>
<td>.0011</td>
<td>159</td>
<td>0</td>
<td>15,200</td>
</tr>
<tr>
<td>40</td>
<td>10</td>
<td>.001</td>
<td>.0017</td>
<td>186</td>
<td>0</td>
<td>20,200</td>
</tr>
</tbody>
</table>
it is irrational to order cash on Thursday or Friday, since a Saturday order arrives on the same day.

Table 3 outlines the performance of the model and of management, under various assignments of the cost and risk parameters. These parameter assignments are justified as follows:

Fixed order costs b include the cost of armoured car delivery and the bank charge for preparing United States and Canadian coin and currency parcels. These total $18.50, but other fixed costs, such as the cost of the tellers' time to count the cash, make $25 and $40 fixed costs more realistic. The model was tested under both cost assumptions.

The credit union offers 7% annual interest on the minimum daily balance in one type of savings account, and this may be taken as the opportunity cost of cash. Alternatively, aggressive investment of excess cash could yield rates of up to 10% per year, so the model was tested with this opportunity cost as well.

The model was tested with three risk levels \( \alpha \) of a cash-out: \( \alpha = .001 \), \( \alpha = .01 \), \( \alpha = .025 \). These risk levels are only valid under the assumption that the errors are normally distributed. Since the forecast residuals were too fat-tailed to be normal, the cash-out risk may be higher or lower than \( \alpha \). To estimate this level of risk, a count of the number of cash-outs in the simulations corresponding to the optimal orders on each order date was kept. This number, divided by the total number of iterations (where the total number of iterations = NSIM x number of orders placed) is a good indication of the risk of a cash-out and is listed beside \( \alpha \) under the heading "Probability of a Cash-Out in Simulation."

In addition a count of the total number of cash-outs that would have
occurred with the actual cash flows, when the optimal order policy is followed, is provided.

Referring to Table 3, we may first compare performance of management with the model, using fixed costs of $25 per order and an opportunity cost of cash of 7% per year. In the 2 year and 11 month period from May 1973 to March 1976, management generally placed weekly orders, incurring costs of $15,700 per year. On two occasions (or for 1.3% of the orders placed), deposits were solicited from a local merchant to cover cash shortages ($2600 and $20,800). In contrast, with \( \alpha = .025 \) the model yielded only a 1% risk of a cash-out in simulated performance and no cash-outs in practice, while total costs were only $11,100/year. Decreasing \( \alpha \) to .001 resulted in a 0.2% cash-out risk in simulated performance while the costs were still less than those of management ($13,900/year). Over the 35 month period the model ordered cash about 200 times, compared to management's 153 orders, which suggests that management's commitment to weekly orders was a large source of the sub-optimality of its historical performance.

Another way of regarding the simulated cash-out probabilities is to note that the "insurance premium" associated with reducing the probability of a cash-out from .010 to .008 is $11,600-$11,100 = $500 per year, while the "insurance premium" for reducing the cash-out risk from .010 to .002 is $2800 per year.

When the fixed order costs and opportunity costs of holding cash were varied with \( \alpha = .001 \), the simulated cash-out probability was always between .001 and .002 and the model responded in a predictable manner in terms of the number of orders placed. That is, increased
fixed costs caused fewer orders to be placed and increased opportunity costs caused more orders to be placed.

In summary, the model provides more protection against cash-outs while incurring smaller operating costs than management's historical performance. Also, the model readjusts its long run policy in the correct direction as the ratio of fixed order costs to opportunity costs is varied.

5.7 Summary

In this chapter, the algorithm for ordering till cash was developed. In Section 5.2 it was asserted that a chance constrained model would be more appropriate than a recourse model, and the exact forms of the order and holding cost functions were established. The calculation of the deterministic equivalent of the chance constraint was discussed in Section 5.3. In Section 5.4, it was pointed out that a somewhat more complicated objective function than that proposed in Section 4.5 would yield a closer approximation to truly optimal decisions and an operational method of evaluating the objective function by simulation was presented. The actual cash order algorithm for minimizing the objective function by means of a grid search was presented in Section 5.5. The performance of the cash order model was compared to the cash order performance of management in Section 5.6. It was established there that implementation would yield lower costs and/or greater protection against the risk of cash-outs.
APPENDIX TO CHAPTER 5

THE DYNAMIC PROGRAM CORRESPONDING TO THE OBJECTIVE FUNCTION (5.4.2)

Let $\tau = t_0 < t_1 < t_2 < t_3 < \cdots$ be the days on which it is rational to order cash. This is a list of those days $t_n$ such that an order placed after $t_n$ cannot arrive at the same time as an order placed later (i.e. $t + L(t) > t_n + L(t_n)$ whenever $t > t_n$). Let

$$C_t^n (x_{t_n} + L(t_n), \ldots, x_{t_n+1} + L(t_{n+1})) = \sum_{t=t_n+L(t_n)}^{t_{n+1}+L(t_{n+1})-1} C_t(x_t) \quad (5.A.1)$$

This is the sum of the holding costs of opening cash balances which would be affected by the size of the order $Q_{t_n}$, but not by the size of the next possible order $Q_{t_{n+1}}$. Then let

$$h_t^n (Q_{t_n}, Q_{t_n}; x_{t_n})$$

$$= F(Q_{t_n}) + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} C_t^n d\phi_{t_n} \cdots d\phi_{t_{n+1}+L(t_{n+1})-1}$$

$$+ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \min_{Q_{t_{n+1}} > 0} h_t^n (Q_{t_n+1}, Q_{t_{n+1}}; x_{t_{n+1}}) d\phi_{t_n} \cdots d\phi_{t_{n+1}-1} \quad (5.A.2)$$

Note that the minimizations are performed subject to the constraint (5.3.2)
and for $t_n \leq t \leq t_{n+1} + L(t_{n+1})$, $x_t = x_{t_n} + \sum_{j=t_{n+1}}^{t} (-y_j) + R(Q_j)$.

This is a dynamic program where we iterate from $n = N$ to $n = 0$ and the horizon is $T = t_N$. Now if we condition on the time $t + 1 = t_{n+1}$ of the next order, recursive substitution of (5.A.2) for $h_{t_0}', \ldots, h_{t_n}'$ yields the following analogue of (4.5.1) where $t_0 = 1$:

$$h_i'(Q_i; Q_i, x_i) = A'(\tau, Q_i) + E(\min_{t_{n+1}} h_{t_{n+1}}'),$$

(5.A.3)

where $A'$ is defined in (5.4.3). Again we approximate the expectation in (5.A.3) by $h_1'$, adjusted now for the different number of calendar days associated with the two terms. That is, $E(\min_{t_{n+1}} h_{t_{n+1}}') = \frac{D(T) - D(t_{n+1} + L(t_{n+1}) - 1)}{D(T)} h_1$.

Substituting into (5.A.3) yields the following approximation to the average daily cost:

$$\frac{h_1'}{D(T)} = \frac{A'(\tau, Q_i)}{D(t_{n+1} + L(t_{n+1}) - 1)} = \frac{A'(\tau, Q_i)}{D(\tau + L(\tau + L))}.$$

Computing the expectation over $\tau$ of the above expression yields (5.4.2).
Chapter 6

CONCLUSIONS

At this point it is appropriate to briefly review some of the general lessons that were learned by developing this project.

From a statistical viewpoint it is perhaps surprising that so much random variation is present in the recorded daily cash flows. If, as suspected, a large portion of the variation is in the record keeping, it is because the records are oriented to being accurate on an aggregate monthly basis rather than a daily basis.

Despite the fact that the cash demand at the credit union varies in a cyclic manner, Box-Jenkins time series techniques cannot be used to model the demand process. This is a result of non-stationarity induced by the trend towards increasing cash demand as well as effective non-stationarity arising from the long periods of the cycles (up to one year long) and the even greater least common multiple of these periods.

The estimation of the demand for cash finally had to be performed in a manner that was custom fitted to the specific case at hand and cannot be readily generalized to other branches. It entailed fitting a non-linear trend to real cash demand that incorporated the notion of an asymptotic capacity level. Such a trend model may not be appropriate for other
branches. The subsequent estimation of demand variation about the trend was performed by regression techniques that can readily be generalized to other branches, however.

In terms of the development of a cash order algorithm, it is apparent that the till cash management problem is not so sharply distinguished from standard inventory theory problems as is the classical cash management problem, in which the lag time for transactions is zero and the stochastic cash balance can drift up or down. However, the main problems associated with till cash management appear to be the non-stationarity of demand and other features that make it hard to develop a steady state solution. An approximate dynamic programming solution to the till cash management problem at a potential order point was obtained by conditioning on the time of the next order and replacing the expected costs of optimal behaviour after that time with the average cost prior to that time (rescaled because of the shorter time to the horizon). Greater accuracy, at much greater computational expense, might be obtained by conditioning on the time of the third or some later order, in order to incorporate more iterations in the dynamic program before invoking an approximation. Unfortunately the number of calculations increases exponentially with the number of dynamic programming iterations.

The major trade-off between theoretical correctness and operationality occurs in the cash order algorithm, rather than the statistical estimation. Since the algorithm does actually out-perform management and can readily be implemented with existing on-line computer facilities it is a success from an operational point of view.
LIST OF REFERENCES


