Paper 239-2010

Determining Optimal Cash Allocation at ICICI Bank Branches with SAS® Enterprise Guide® and SAS/OR® Software

Sandipan Ray, ICICI Bank, Mumbai, India

ABSTRACT

Excess cash at a bank branch is a potential income loss. However, inadequate cash can cause a "cash-out" situation. Cash allocation plans are finalized at the beginning of the month, and changes are costly and might not be feasible. This paper describes an optimal cash allocation plan that minimizes expected end-of-the-day (EOD) cash holding at branches, subject to constraint on expected withdrawal rate and buffer cash level.

SAS® Enterprise Guide® is used to develop regression models to predict cash withdrawal and deposit. Expected EOD cash holding at a branch is estimated from the forecasting solution for a given cash allocation rule. Using the SAS/OR® module, an optimal cash allocation rule is developed. In the current process, deposited cash at the branch is circulated after it is processed at a central processing unit. In an alternative scenario, processing is done at the branch, and the impact on allocation and EOD cash holding is estimated to find out whether it makes economic sense for the bank to set up a processing unit at the branch level.

INTRODUCTION

ICICI Bank is India's second-largest bank with total assets of Rs. 3,562.28 billion (US$77 billion) as of December 31, 2009, and profit after tax Rs. 30.19 billion (US$648.8 million) for the nine months that ended December 31, 2009. ICICI Bank offers a wide range of banking products and financial services to corporate and retail customers through a variety of delivery channels and through its specialized subsidiaries and affiliates in the areas of investment banking, life insurance and property and casualty insurance, venture capital, and asset management. The Bank has a network of 1,646 branches and about 4,883 ATMs in India and presence in 18 countries.

At ICICI bank, the cash disbursement to branches is done from the Integrated Currency Management Center (ICMC) by a cash replenishment agency (CRA). According to the national regulator Reserve Bank of India, banks are required to hold a certain percentage of their deposit in the form of liquid cash. This ratio is called the cash reserve ratio (CRR). Cash held in the ICMC is counted towards the CRR; however any cash left at ATMs, branches, and in transit at the end of the day (EOD) is excluded. According to the current process followed by ICICI bank, cash deposited at the branch is credited to customers only after it is processed at a central processing unit (ICMC).

For all India, the average *processed cash utilization ratio* (defined as the ratio of the average processed cash disbursed in a day from branches to the processed cash disbursed and remaining outside the ICMC) is less than 45 %. In an average day, Rs. 3.5 billion (US$75 million) of cash is withdrawn from branches and Rs. 5–8 billion (US$1.08–1.7 billion) of processed cash remain at the branches at EOD. Also cash that is deposited at the branch during the day by the customers but not brought back to the ICMC by the end of the day is not counted towards the CRR. This results in a significant loss of interest income for the bank.
In the current arrangement, a cash replenishment plan for the forthcoming month is given to the CRA at beginning of the month. Any exception from the plan (called an "Add Cash Call") that requires an unplanned CRA visit is separately charged and is provided only for a limited number of instances in a month.

OBJECTIVE

The objective of this paper is to propose an optimal cash allocation strategy from the ICMC to the branches so that the expected processed and unprocessed cash holding outside ICMC are minimized subject to the following constraints on service levels:

- The ratio of expected cash withdrawal between two planned CRA visits to the peak processed cash available after an earlier CRA visit should be below a threshold level for all the days over the optimization horizon.

- The expected processed cash available at EOD should be at least a certain percentage of the expected demand for cash for the next day.

Both of these threshold levels are parameters in the optimization model, and several combinations of these are investigated to find the most appropriate level.

The following are assumed to be given because these are contracted with the CRA on a long-term basis:

- number of times that a CRA van visits the branch
- time of visit
- route plan

An alternative scenario in which processing of the deposited cash is done at the branch is introduced. Impact of this strategy on allocation and end-of-the-day cash holding is also estimated to find out whether it makes economic sense for the bank to set up a processing unit at the branch level.

EXISTING CASH ALLOCATION PROCESS

In the current process, the cash replenishment agency (CRA) receives the cash limit for the branch for the next month at the beginning of the month. When the CRA van visits the branch, it checks the amount of cash that is available at the branch. The difference between the cash limit and the amount of cash available is allocated at the branch. Unprocessed cash available at the branch at the time of CRA visit is collected and brought back to the ICMC. A few remote branches are serviced only once every two or three days; some branches in the business districts of metropolitan towns are serviced twice a day. Most of the other branches are serviced once a day by the CRA. This paper addresses the cash allocation of branches that are serviced once a day by CRA.

Data and Methodology

The following data are used in this analysis:

- daily cash withdrawal and deposit data at a branch level over the period July 2007 to December 2009
- end-of--day processed and unprocessed cash available at the branch for the period July 2007 to December 2009
- daily cash allocated to the branch from ICMC for the period July 2007 to December 2009
- cash transaction data from October 2009 to December 2009 to approximate the hourly withdrawal and deposit pattern (because data lower than the daily granularity is not available for all the branches historically)

The cash allocation problem has been divided into two parts:

1. predicting expected daily cash withdrawal and cash deposit at branches
2. minimizing the expected end-of-day cash holding, subject to the service level constraints

2

CASH WITHDRAWAL PREDICTION

Cash withdrawal is a function of seasonal factors and the recent trend in withdrawal.

Modeling Seasonal Factors

Cash withdrawal varies according to various *seasonal factors* such as the day of the week, the week of the month, holiday, salary day, and so on. The days with effect (that is, with an average withdrawal for cash that differs from that of a normal day) are to be captured and accounted for. This further varies according to the location of the branch. For holidays, two types of effects are considered: pre-holiday (up to five days before a holiday) and post-holiday (up to one day following a holiday).

The following events are considered:

- first to third days of the month
- fourth to seventh days of the month
- eighth to eleventh days of the month
- twenty-seventh day to the end of the month
- weekend (Saturday)
- pre-holiday and post-holiday impact

These events are identified based on data analysis and discussion with business team. All other days are considered to be normal days.

The hypothesis is that the preceding events have effect on withdrawal over and above the normal withdrawal level.

For each of the branches, the mean withdrawal is tested at those events to determine whether it is significantly different from the average withdrawal of normal days (based on equality of mean test). If the mean withdrawal is significantly different, then that effect is incorporated into the prediction model.

Assume that μ is the average normal day withdrawal in last six months and F_t is the seasonal factor that influences withdrawal on day t.

Let $I_k(t)$ denote the percentage increase or decrease from the average level in withdrawal due to factor k on date t. If the kth factor is not manifested on day t, then $I_k(t)$ =1.

This analysis assumes that $F_t = \prod_{i=1}^{k} I_i(t) \times \mu$

Modeling the Recent Trend

Withdrawal also varies depending on the recent trend. To capture a more recent trend of cash demand, a time series factor T_{S1} is constructed based on the moving average of the past S_1 days from the observation point to capture a more recent trend of demand. Another time series factor T_{S2} is also constructed based on the moving average of the previous S_2 days. While computing these averages, only days with normal withdrawal are considered. The reason for incorporation of the time series factor is to capture the changing transaction patterns at the branches that are due to addition of new customers and other factors that influence the withdrawal pattern but have not been considered.

Let S_t be the recent trend of withdrawal on day t:

$$S_t = aT_{S1} + (1-a)T_{S2}$$

Let D_t be the actual withdrawal and $\hat{D}_t$ be the predicted withdrawal on day t. The withdrawal is postulated as a weighted average of the seasonal factor and the short term trend of withdrawal:

$$D_t = bS_t + (1-b)F_t + \varepsilon_t \quad \varepsilon_t \sim N(0,\sigma^2)$$

a and b are parameters in the model, which is estimated by a SAS Enterprise Guide least squares technique. Large withdrawals on certain days that are not caused by any specific events are removed, because most of these transactions are due to a large transaction done at the branch by few customers and they are mandated to provide prior understanding to the branch.

After parameters are estimated, error in prediction, $\hat{e}_t = W_t - \hat{W}_t$, is computed and assumption of normality was tested using Jarque-Bera statistics.

The preceding approach is applied for cash deposit at branches too.

The following statements show the use of PROC REG to predict the debit at the branches:

```
proc reg data=WORK.QUERY2_FOR_WORK_SINGLE_BRANCH_AL noprint;
    model dnew_amt = Seasonal_factor ts1_dnew ts2_dnew / SELECTION=NONE;
    by Channel_Id;
    output out=Reg_Predictions predicted=predicted_dnew_amt;
RUN;
```

Figure 1 shows the actual versus predicted debits at Andheri branch (Sol ID: 0011), and Figure 2 shows similar data at Vallabh Vidyanagar (Sol ID: 0085). Both models use $S_1 = S_2 = 14$ days.
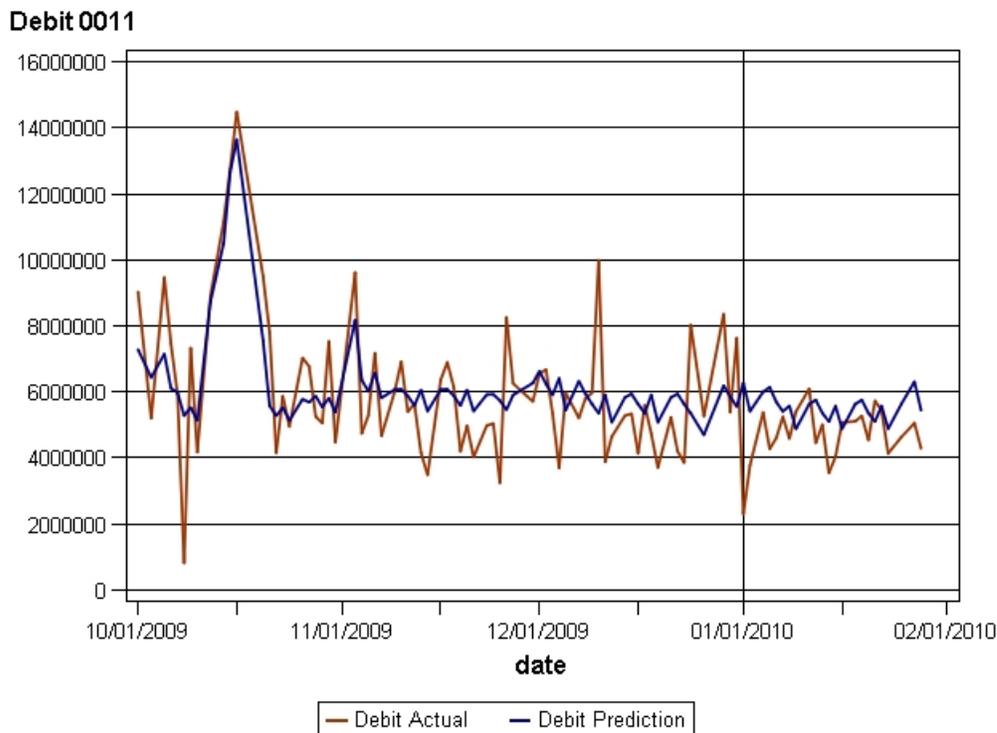
Figure 1. Actual versus Predicted Debits at the Andheri Branch

4
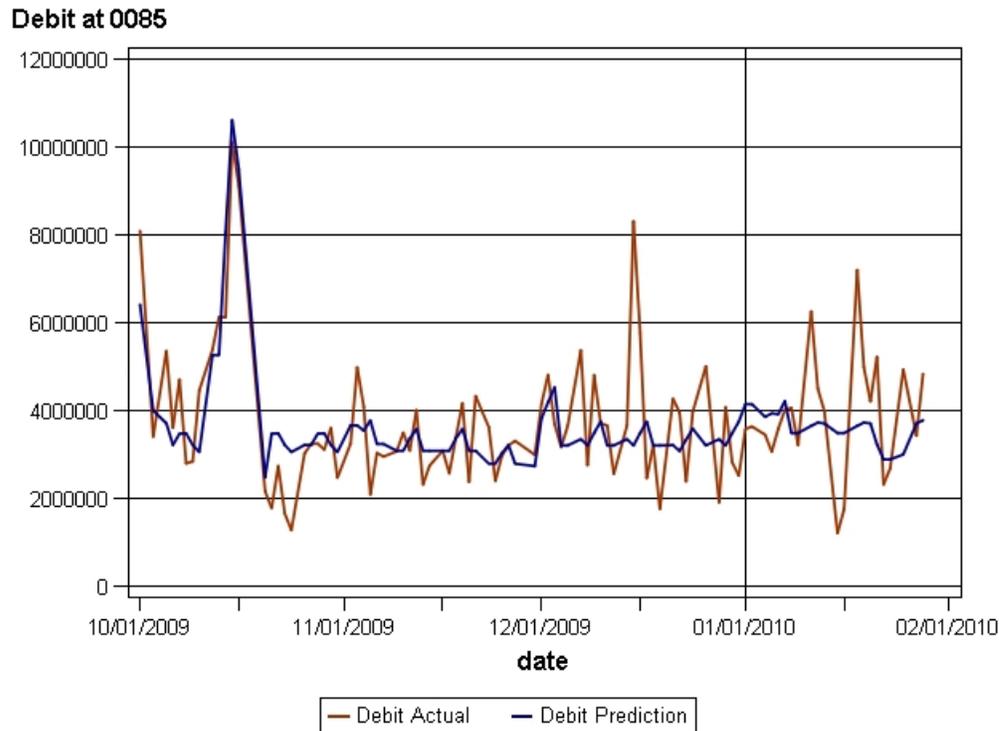
Figure 2. Actual versus Predicted Debits at the Andheri Branch

OPTIMAL ALLOCATION OF CASH

An optimal cash allocation plan is derived for branches that are serviced once a day. Two models for processing of deposited cash are assumed:

- Processing is done at only at the ICMC.
- Processing is done at both the ICMC and at branches.

Cash allocations at branches are optimized under these combinations subject to constraints on the service level. It is assumed that the CRA visit plan is given and the optimal cash allocation plan seeks to minimize EOD cash holding at a branch over a specified time period.

Cash Processing at ICMC Only

The van comes once a day at a fixed time; it collects unprocessed cash and loads processed cash.

The objective is to minimize the EOD cash balance at a branch over a specified time period with respect to processed cash limit subject to the following constraints:

- The ratio of cash withdrawn between the previous day's CRA visit and the next day's CRA visit to the peak processed cash available after the previous day's CRA visit should be below a threshold level for all days over the optimization period.
- The total processed cash available at EOD should be a certain specific number of times the demand for cash next day.

Define the following:

- $\hat{D}(t)$ is the predicted total cash debit in a day.

- $\hat{C}(t)$ is the predicted total cash credit in a day.

- $D(t)$ is the actual cash withdrawal in a day.

- $C(t)$ is the actual cash credit in a day.

- $p(t)$ Is the estimated processed cash that remains in branch at EOD (t>0).

- $u(t)$ Is the estimated unprocessed cash that remains in branch at EOD (t>0).

- The actual processed and unprocessed cash beginning of the optimization period, ($p(0)$ and $u(0)$, respectively) is known.

- $L(t)$ is the processed cash limit for the branch.

- $O(t)$ is the unprocessed cash collected from branch by CRA.

- m is the average percentage cash withdrawal until the CRA visit.

- n is the average percentage of cash credited until the CRA visit.

Both m and n are empirically derived for a branch from the last three month's transaction data.
The van refills cash at a fixed time of the day and is aware of the exact amount of processed cash held at the branch the previous night.

By definition, $p(t) = L(t) - (1-m)\hat{D}(t)$ and $u(t) = (1-n)\hat{C}(t)$.

Let $r(t) = \dfrac{(1-m)\hat{D}(t-1) + m\hat{D}(t)}{L(t-1)}$ be defined as the *expected withdrawal rate* (the ratio of the

processed cash withdrawal until the CRA visit as a percentage of the previous day's peak processed cash holding.

To optimize the cash, perform the following calculation over the targeted days (in this case, the next month):

Minimize $\displaystyle\sum_{t=1}^{T} p(t) + u(t)$ w.r.t $L(t)$

subject to the following constraints:

- The expected withdrawal is $r(t) \leq r_0$. The ratio of expected cash withdrawal until the CRA visit the next day as a percentage of the previous day's peak processed cash holding should be below a threshold level r_0 for all days over the optimization period.

- The expected processed cash on a day $p(t)$ should be $p(t) \geq k\hat{D}(t+1)$. Processed cash available at EOD should be k times the demand for the next day for all days over the optimization period, where k is the buffer parameter.

- Various levels of r_0 and k have been explored.

- The van can come up to one hour late from the scheduled time, and for each of the branches, the expected percentage withdrawal until the van's arrival has been computed from last three months data and taken as value for k.

- For some of the branches where recent cash withdrawal pattern has a higher fluctuation (high σ^2) , the r_0 value is kept at a relatively low level.

6

- The impact on allocation of possible additional cash requests is not explicitly modeled. This is an appropriate assumption, because for branches cash buffers are kept at a reasonably high level.

The following SAS statements show the use of PROC OPTMODEL from the SAS/OR package to solve this optimization problem:

```
%let maxcnt=71;
%let k=0.28;/* processed cash at eod to take care of demand at next day till
refill 2/9( a ) is minimum*/
%let a=0.1;

%let x=0.85;
%let y=0.9;
%let b=.35;
%let strtDr= 9829871;
%let strtCr= 6220990;

data work.single_branch_new;
set WORK.SINGLE_BRANCH_S(obs=&maxcnt);
retain date1 0;
date1+1;
dmt=debit*&a;
cmt=credit*&b;
u=cmt;
drop date;
run;

proc optmodel;
    set dates init 1..&maxcnt;
       num debit{dates};
       num credit{dates};
       num dmt{dates};
       num cmt{dates};
       var p{dates}>=0;
       var q{dates}>=0 <=&y;
       num u{dates};
       var l{dates}>=0 <=20000000;
      read data work.single_branch_new into dates=[date1] debit=Debit
credit=Credit dmt cmt u;

       num r{d in dates} = if d in {3..&maxcnt} then ((1-&a)*Debit[d-1] +
&a*Debit[d]) / (l[d-1].sol + p[d-2].sol - &a*Debit[d-1])
       else if d = 2 then ((1-&a)*debit[1] + &a*debit[2]) / (l[1].sol-
&a*debit[1] + &strtdr);

       /*constraints*/
       /* write a constraint to derieve/define p*/
       con pchk{d in dates : d>1 & d<=&maxcnt}: p[d]= l[d] - debit[d] + p[d-1];
       con pchk0{d in dates : d=1}: p[1]= l[d] - debit[1] + &strtdr;

       /* ensure that cash withdrawaltill CRA visit next day as a percentage of
yesterdays peak processed cash holding*/
       /* constraint for r of utilization*/

       con rchk{d in dates : d>2 & d<=&maxcnt}:
       ((1-&a)*Debit[d-1]+&a*Debit[d]) <= &x * (l[d-1] + p[d-2] - &a*Debit[d-
1]);

       con rstrt:
               ((1-&a)*debit[1]+&a*debit[2]) <= &x * (l[1]-&a*debit[1]+&strtdr);

       /*constraint to check if the processed cash at EOD is sufficient for next
day demand: p(t)>kd(t+1)*/
```

```
con peodchk{d in dates: d>1 & d<&maxcnt}: p[d] >= &k*debit[d+1];

/*Objective Function*/
min buffer = sum{d in dates}(u[d]+p[d]);
expand buffer;
solve obj buffer;
print u;
print r;
print l;
create data new_results_v2 from [date]=dates u_cash=u p_cash=p r=r
limit=l debit_p=debit;
quit;
```

Cash Processing at Branch and the ICMC

The van comes once a day at a fixed time; it collects soiled cash and loads processed cash. It is assumed that the branch has a processing unit that continuously processes deposited cash and puts it in reuse. The percentage of cash it can process is assumed to be constant. The remaining cash is soiled or counterfeit and is taken away to the ICMC. Let s be the percentage of deposited cash that can be processed at the branch.

By definition,

$$p(t) = L(t) - (1-m)\hat{D}(t) + (1-n)s\hat{C}(t)$$

$$u(t) = (1-n)(1-s)\hat{C}(t)$$

Let $r(t) = \dfrac{L(t-1) - (1-m)\hat{D}(t-1) - m\hat{D}(t) + s(1-n)\hat{C}(t-1) + sn\hat{C}(t)}{L(t-1)}$ be defined as the ratio of expected processed cash available until the CRA visit as a percentage of the previous day's peak processed cash holding.

To optimize the cash, perform the following calculation over the targeted days (in this case, the next month):

Minimize $\sum_{t=1}^{T} p(t) + u(t)$ w.r.t $L(t)$

subject to the following constraints:

- The expected withdrawal is $r(t) \geq r_0$. The ratio of processed cash available until the CRA visit as a percentage of yesterday's peak processed cash holding should be above a threshold level r_0 for all days over the optimization period.

- The expected processed cash on a day $p(t)$ should be $p(t) \geq k\hat{D}(t+1)$. Processed cash available at EOD should be k times the demand for next day for all days over the optimization period k is buffer parameter.

The following SAS statements show the use of PROC OPTMODEL from the SAS/OR package to solve this optimization problem:

```
%let maxcnt=26;
%let k=0.3;/* processed cash at eod to take care of demand at next day till
refill 2/9( a ) is minimum*/
```

```
%let a=.15;

%let x=0.15;
%let y=0.3;
%let b=.2;
%let s=0.8;

/* Average values of aug/sep09  data*/
%let strtDr= 5589871 ;
%let strtCr= 6120990;

/*85% forecast code*/
data work.single_branch_new;
set WORK.SINGLE_BRANCH_S(obs=&maxcnt);
retain date1 0;
date1+1;
dmt=debit*&a;
cmt=credit*&b;
u=cmt;
drop date;
run;


proc optmodel;
    set dates init 1..&maxcnt;
       /*set<string> ATMS init {};*/
       num debit{dates};
       num credit{dates};
       num dmt{dates};
       num cmt{dates};
       var p{dates}>=0;
       num u{dates};
       var l{dates}>=0 <=10000000;
    read data work.single_branch_new into dates=[date1] debit=Debit
credit=Credit dmt cmt u;

/*constraints*/
       /* write a constraint to derieve/define p*/
       con pchk{d in dates : d>1 & d<=&maxcnt}: p[d]= p[d-1]+l[d]-
debit[d]+credit[d]*&s;

       con pchk0{d in dates : d=1}: p[1]= &strtdr+l[d]-debit[1]+credit[1]*&s;

       /* ratio of processed cash remained at branch till CRA visit
        as a percentage of yesterday's peak processed cash holding.*/
       num r{d in dates} = if d in {3..&maxcnt} then (p[d-1].sol -&a*Debit[d]
+&s*&b*credit[d]) / (l[d-1].sol + p[d-2].sol - &a*Debit[d-1] + credit[d-
1]*&s*&b)
              else if d=2  then (p[1].sol -&a*Debit[2] +&s*&b*credit[2]) /
(l[1].sol + &strtdr - &a*Debit[1] + credit[1]*&s*&b);
       /* constraint for r of utilization*/

       con r_ens{ d in dates: d>2 & d<=&maxcnt}:
       (p[d-1] -&a*Debit[d] +&s*&b*credit[d]) >= &x * (l[d-1] + p[d-2] -
&a*Debit[d-1] + credit[d-1]*&s*&b);
       con r_ens0{ d in dates: d=2}:
       (p[1] -&a*Debit[2] +&s*&b*credit[2]) >= &x* (l[1] + &strtdr - &a*Debit[1]
+ credit[1]*&s*&b);

       /* ratio of processed cash remained at branch at EOD as a percentage of
peak processed cash holding  */
       num q{d in dates} = if d in {2..&maxcnt} then (p[d].sol)/(l[d].sol+p[d-
1].sol-&a*Debit[d]+credit[d]*&s*&b)
```

9

```
        else if d=1 then (p[1].sol)/(l[1].sol + &strtdr-&a*Debit[1] +
credit[1]*&s*&b);

        con q_ens{ d in dates: d>1 & d<=&maxcnt}:
        (p[d]) >= &y*(l[d]+p[d-1]-&a*Debit[d]+credit[d]*&s*&b);
        con q_ens0{ d in dates: d=1}:
        (p[1]) >= &y*(l[1]+&strtdr-&a*Debit[1]+credit[1]*&s*&b);

        /*constraint to check if the processed cash at EOD is sufficient for next
day demand: p(t)>kd(t+1)*/
        con peodchk{d in dates: d>1 & d<&maxcnt}: p[d]>= &k*debit[d+1];

        /*Objective Function*/
    min buffer = sum{d in dates}(u[d]+p[d]);
        expand buffer;
        solve obj buffer;
        print u;
        print r;
        print q;
        print l;
        create data new_results from [date]=dates u_cash=u p_cash=p r=r q=q
limit=l debit_p=debit credit_p=credit;
quit;
```

RESULTS

The solution has been rolled out to 20 large branches acrosss India which typically have high EOD cash. Figure 3 and Figure 4 show the result of the 'pre-pilot' optimization exercise for select branches. There is some fluctuation in cash deposit on specific days which the cash forecasting model is not able to predict satisfactorily. Two solutions are explored to tackle this problem. One is obviously to improve the demand forecasting model. The second is to vary the buffer level r_0 and k optimally so that the service level and the utilization rate are increased simultaneously.

The first part of the exercise involves predicting the withdrawals at the branches. Figure 1 and Figure 2 show the predictions of withdrawals at two branches with Andheri (Sol ID 0011) and Vallabh Vidyanagar (Sol ID 0085). Also shown in the same graphs are the out-of-time predictions (from January 1, 2010, to January 28, 2010).

Using the debit predictions shown, the optimum allocation schedule that is to be sent to the ICMC for the next month is calculated. The SAS/OR OPTMODEL procedure (PROC OPTMODEL) is used to find the optimum limits.
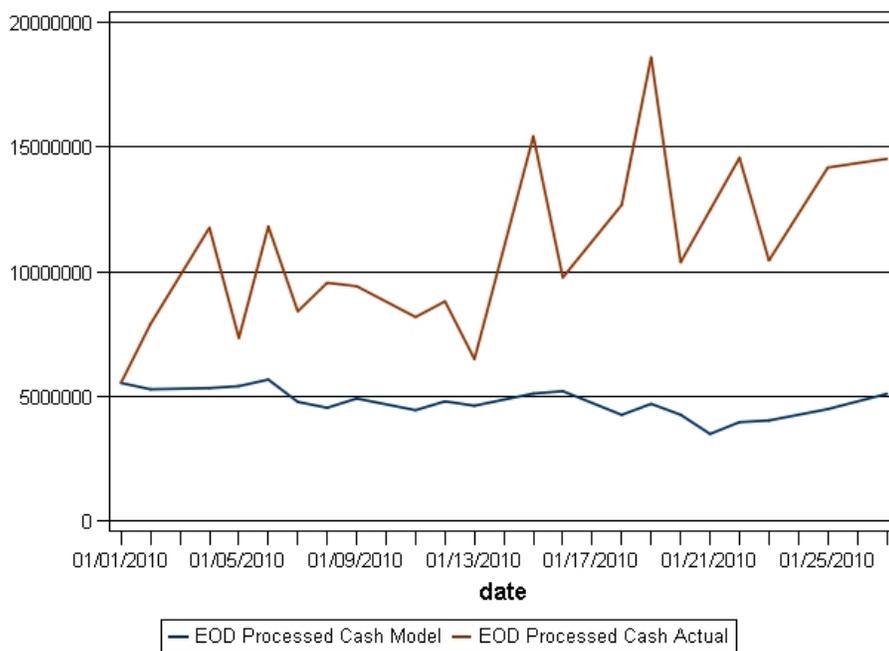
Figure 3 and Figure 4 show the impact of the model on the EOD processed cash at these branches.

EOD Processed Cash at 0011
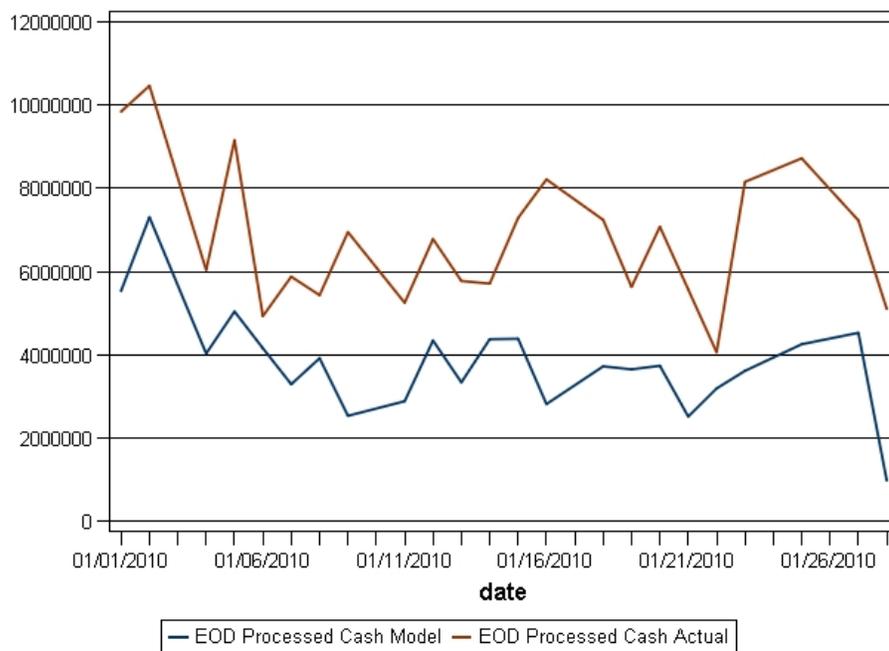
Figure 3. EOD Processed Cash Reduction for the Andheri Branch

EOD Processed at 0085

Figure 4. EOD Processed Cash Reduction for the Andheri Branch

Figure 4 shows that, for the month of January 2010, the EOD processed cash would have been reduced from an average value of Rs. 6,806,568 to Rs. 3,833,742. This gives the Bank a profit of (Rs. 6,806,568 – Rs. 3,833,742) x 4% = Rs. 118,913.05 at a single branch (0011). Similarly, at 0085 the profit would have been (Rs. 11,146,571.6 – Rs. 4,592,899.19) x 4% = Rs. 262,147.

The average net profit for the entire bank (all branches included) would be around Rs. 125,000,000 (estimated value).

Also, we would like to see the impact of adding cash processing capabilities at the branches. The debits and credits are also predicted for the next month, based on which we found the optimal limits of the cash to be indented from ICMC. Figure 5 shows a comparison of the current scenario with the simulated scenario of cash processing at branch. Figure 5 also shows the impact of the optimization model without the cash processing case.
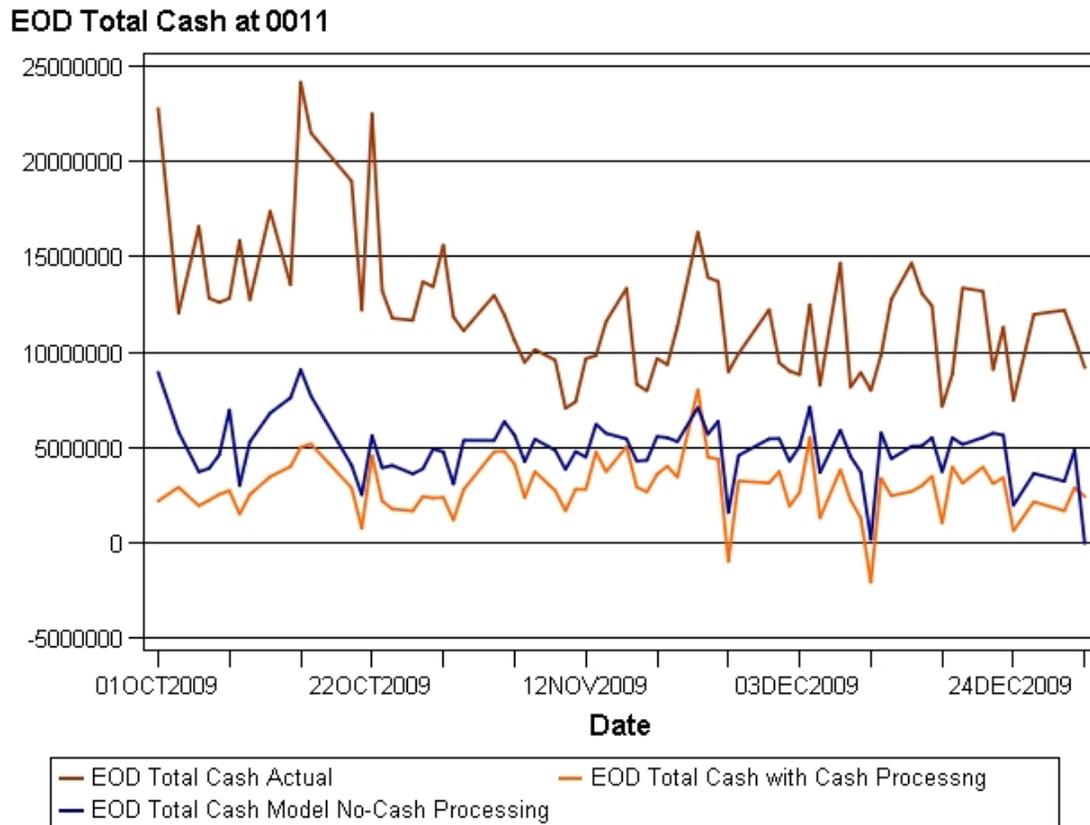
Figure 5. Impact of Cash Processing at Branches (Andheri Branch Case)

The monetary impact this would bring is:

(Average(EOD Total Cash Actual) – Average(EOD Total Cash with Cash Processing)) x 4%
= (Rs. 12,140,829.88 – Rs. 2,931,448.8) x 4% = Rs. 368,375

(If cash processing had not been done, the profit would be Rs. 289,010.)

All these figures are at a single branch (0011) only.

CONCLUSION

This paper is a work in progress, and the model will be applied to all the branches where cash servicing is done once a day. Subsequently, optimization rules will be developed for the branches where servicing is done more than once or done after couple of days. As of now, results indicate that using optimization will significantly increase (around 65%–75%) cash utilization level at the branch without causing a cash-out

situation. This program provides an indicative level of cash demand at the branch. Branch managers deduct the process cash available at the branch from the limit and add the amount for which they have prior intimation for large cash withdrawals. However, the key lesson is that cash processing at a branch might not be optimal unless the branch is very large. A similar model for allocating cash at the ATMs is under development. ATM cash allocation is more complicated because it usually works with a lower cash buffer. An allocation strategy needs to incorporate both the time of low cash and the refilling strategy used.

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Sandipan Ray
ICICI Bank
A-WING, 7th floor, AUTUMN ESTATE, CHANDIVALI FARM ROAD, OPP.MADA COLONY, CHANDIVALI
Mumbai, India - 400075
+91-22-42514047
sandipan.ray@icicibank.com
www.icicibank.com