**European Journal on Decision Processes manuscript No.**
(will be inserted by the editor)

# Data-driven multiobjective decision-making in cash management

**Francisco Salas-Molina** ·
**Juan A. Rodríguez-Aguilar**

Received: date / Accepted: date

**Abstract** The volume and availability of business and finance data may continue to increase in the near future. However, the utility of such data is by no means straightforward due to a lack of integration between data-driven techniques and usual decision-making processes. This paper aims to integrate data with multiobjective decision-making in cash management by means of machine learning. To this end, we first consider cash flow forecasting as a data-driven procedure to be used as a key input to multiobjective cash management problem in which both cost and risk are goals to minimize. Next, we compute the forecasting premium, namely, how much value can be achieved in exchange of predictive accuracy. Finally, we provide cash managers with a general methodology to improve decision-making in cash management through the use of data and machine learning techniques. This methodology is based on a novel closed-loop procedure in which the estimated forecasting premium (if any) is used as a critical feedback information to find better forecasting models and, ultimately, better cost-risk results in cash management.

**Keywords** Machine learning · multiobjective decision-making · integration · cash management

**MSC Codes:** 90B05; 90C05; 90C11.

Francisco Salas-Molina
Hilaturas Ferre, S.A., Les Molines, 2, 03450 Banyeres de Mariola, Alicante, Spain
E-mail: francisco.salas.molina@gmail.com

Juan A. Rodríguez-Aguilar
IIIA-CSIC, Campus UAB, 08913 Cerdanyola, Catalonia, Spain
E-mail: jar@iiia.csic.es

## 1 Introduction

There is no doubt in the fact that the main available resource for decision-making is data. Data contains information and useful information is out there for those who need it, for those who want to use it to make better decisions. Not using this information would be equivalent to leaving the tap on wasting water. A number of examples of the utility of data in the context of business and finance have been described in Zopounidis (1999); Doumpos et al (2012); Zopounidis and Doumpos (2013); Doumpos and Grigoroudis (2013). However, there is a lack of supporting research on data-driven methods to improve decision-making in cash management.

Cash management focuses on finding the balance between cash holdings and short-term investments such as marketable securities. Cash managers make daily decisions about the firm optimal cash level for operational and precautionary purposes (Ross et al, 2002). In cash management, opportunities for improved decision making are significant, particularly in contexts that both include complex cash management systems with multiple bank accounts and involve uncertainty of the future. The economic impact of decision making in cash management is high enough to consider decision support systems. As an example, consider the potential losses derived from a bad decision leading to an overdraft when high penalty costs are charged on negative cash balances.

Since the seminal inventory approach by Baumol (1952), different cash management models based on a set of control bounds have been proposed in the literature (Gregory, 1976; Srinivasan and Kim, 1986; da Costa Moraes et al, 2015). Cash flows are usually assumed to follow a Gaussian distribution either in the form of a Wiener process (Constantinides and Richard, 1978; Premachandra, 2004; Baccarin, 2009). However, empirical data sets are hardly used with the exception of Gormley and Meade (2007); Salas-Molina et al (2016, 2017). Beyond the discussion about the most appropriate distribution, cash management models proposed in the literature present two common features: (i) cash management models assume a theoretical distribution function as a way to incorporate uncertainty about the future; and (ii) cash management models are based on a set of control bounds.

Hence, in the likely situation in which real cash flow data sets are available, one can easily identify at least two research questions worth tackling with a common key feature, namely, the use of data:

– **Question 1.** Can we follow a data-driven procedure to reduce uncertainty about the near future?
– **Question 2.** Can we benefit from a data-driven cash management model based on past observations?

In an attempt to answer the previous questions, we here propose an integrated methodology to incorporate data as a key input to decision-making in cash management. First, we assume that cash managers are interested not only in the cost but also in the risk associated to cash policies. Cost and risk are usually desired but conflicting objectives in cash management. Holding cost reductions are achieved by reducing cash balances but, at the same time, the risk of an overdraft increases. Although the main performance measure in cash management is cost, Salas-Molina et al (2016) extended the classical cost analysis to consider an additional goal, namely, risk. We use this multiobjective framework to estimate the benefits from improving forecasting accuracy.

Cash managers can leverage cash flow forecasts to reduce the uncertainty within a short-term planning horizon (Stone, 1972; Gormley and Meade, 2007; Salas-Molina et al, 2017). However, since predictability varies due to the particular characteristics of the underlying data and the forecasting technique used, we offer a comprehensive methodology for cash managers to evaluate their particular forecasting premium, i.e., the amount of cost (risk) reductions that can be achieved by forecasting cash flows with a given accuracy. By comparing the forecasting premium (if any) to the cost of designing and deploying any forecasting technique, cash managers are able to decide if the use of forecasts is worthwhile.

We also establish a link between machine learning best practices, such as time-series forecasting, and multiobjective decision-making that is ready to be used for profit by cash managers. More precisely, we present a general method to approach cash management from a data-driven multiobjective perspective. We incorporate cash flow forecasts as a data-driven procedure to reduce uncertainty about the near future. In addition, we propose a novel closed-loop procedure in which the estimated forecasting premium (if any) is used as a critical feedback information to find better forecasting models.

Summarizing, a body of work on the use of forecasting in cash management was initiated by Stone (1972) and followed by Gormley and Meade (2007); Salas-Molina et al (2017). Our paper aims, to extend this body of work by:

1. Introducing the concept of forecasting premium.
2. Relying on a data-driven procedure to set minimum balances.
3. Linking the concepts of machine learning and multiple criteria decision-making in cash management by means of a general methodology.

In what follows, we first provide useful background on a recent formulation of the cash management problem as a framework for analysis purposes in Section 2. In Section 3, we first propose a general data-driven cash management model using forecasts as a key input and we later present a methodology to evaluate if a data-driven procedure such as forecasting is worthwhile. In Section 4, we describe a general methodology to integrate forecasts in cash management within a multiobjective framework. Section 5 concludes and provide natural extensions of this work.

## 2 Background: the cash management problem

In this section, we provide useful background on the cash management problem (CMP). More precisely, we briefly introduce the CMP formulation proposed by Salas-Molina et al (2016). Intuitively, the CMP can be viewed as a control problem. Think of a water tank like the one depicted in Figure 1. The level of the tank needs to be monitored to keep the water between two bounds, for instance, a low bound and a high bound. To this end, some control actions can be taken to increase or decrease the level of water. Replace water with money and you will be dealing with a cash management problem.

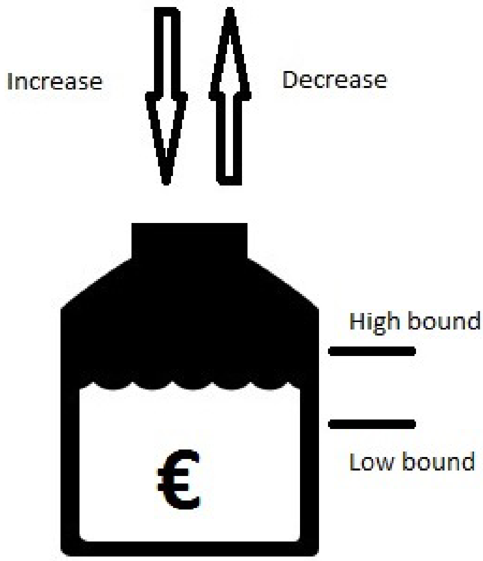


Fig. 1: A graphical definition of the cash management problem.

Formally, the CMP is defined as an optimization problem that, given a set of past cash flow observations determining an initial cash balance, aims to find the best sequence of transactions $X = \{x_t : t = 1, 2, \ldots, \tau\}$ with $x_t \in \mathbb{R}$, what is called a policy, that minimizes some cost function over a time horizon of $\tau$ time steps. The CMP is characterized by its particular cost structure. Any ordering transaction into a cash account may have a cost, which may include a fixed part ($\gamma_0^+$) and a variable part ($\gamma_1^+$). On the other hand, a return transaction from a cash account may also have a cost with a fixed part ($\gamma_0^-$) and a variable part ($\gamma_1^-$). Furthermore, at the end of each time step $t$, a holding cost $v$ per money unit is charged if a positive cash balance occurs, or a penalty cost $u$ per money unit is charged if a negative cash balance occurs. According to this cost structure, a general daily cost $c(x_t)$ is expressed as:

$$c(x_t) = \Gamma(x_t) + H(b_t) \tag{1}$$

where $x_t$ is the transfer made at time step $t$, $\Gamma(x_t)$ is a transfer cost function, $H(b_t)$ is a holding/shortage cost function, and $b_t$ is the cash balance at the end of time step $t$, determined by the following cash balance state equation:

$$b_t = b_{t-1} + x_t + f_t \tag{2}$$

being $f_t$ the net cash flow occurred at time step $t$. Then, the transfer cost function $\Gamma(x_t)$ is defined as:

$$\Gamma(x_t) = \begin{cases} \gamma_0^- - \gamma_1^- \cdot x_t & \text{if } x_t < 0, \\ 0 & \text{if } x_t = 0, \\ \gamma_0^+ + \gamma_1^+ \cdot x_t & \text{if } x_t > 0. \end{cases} \tag{3}$$

Additionally, the holding/shortage cost function is expressed as:

$$H(b_t) = \begin{cases} -u \cdot b_t & \text{if } b_t < 0; u > 0, \\ v \cdot b_t & \text{if } b_t > 0; v > 0. \end{cases} \tag{4}$$

Under this cost structure, the ultimate goal of the CMP within a biobjective framework is to find the policy $X$ that minimizes the expected cost and risk over the time horizon $\tau$. To this end, expected cost $C(X)$ is measured by the average daily cost:

$$C(X) = \frac{1}{\tau}\sum_{t=1}^{\tau} c(x_t) = \frac{1}{\tau}\sum_{t=1}^{\tau}[\Gamma(x_t) + H(b_t)] \tag{5}$$

and expected risk $R(X)$ is measured by the standard deviation of the daily cost:

$$R(X) = \left(\frac{1}{\tau}\sum_{t=1}^{\tau}(C(X) - c(x_t))^2\right)^{1/2}. \tag{6}$$

Then, under the framework of compromise programming (Yu, 1985; Zeleny, 1982; Ballestero and Romero, 1998), the goal is to find policy $X$ that minimizes the sum weighted distances to an ideal point where expected cost and risk are zero:

$$\min \left[\frac{w_1}{C_{max}} \cdot C(X) + \frac{w_2}{R_{max}} \cdot R(X)\right] \tag{7}$$

subject to:

$$X \in S \tag{8}$$

where $C_{max}$ and $R_{max}$ are normalization factors to avoid meaningless summation of goals and $S$ is the set of all possible policies established by a cash management model. In what follows, we propose a cash management model with no restriction of the form of $S$.

## 3 Extracting value from data: the forecasting premium

The size of business data bases may continue to increase on a daily basis. In addition, recent results (Gormley and Meade, 2007; Salas-Molina et al, 2017) must encourage cash managers to produce better cash flow forecasts since improvements in accuracy can be converted into important benefits. Thus, a data-driven approach in which decision-making is based on the analysis of data, rather than purely on intuition should result in better decisions in cash management. In this section, we first propose a multiobjective cash management model using forecasts as a key input. Next, we describe a new data-driven procedure to extract useful knowledge from data to obtain what we call the forecasting premium, i.e., how much value can be achieved in exchange of accuracy.

### 3.1 A data-driven multiobjective cash management model

In most previous models presented in the literature, cash managers have to determine the bounds which minimize the sum of transaction and holding costs. However, the ultimate goal of cash managers is not to find the best set of bounds, but the best sequence of control actions. To this end, cash managers can leverage forecasts to derive policies (Salas-Molina et al, 2017). Given a cash flow data set $\boldsymbol{f} = \{f_j : j = 1, 2, \ldots, N\}$, cash managers can rely on state-of-the-art forecasting techniques to obtain in-sample forecasts $\hat{\boldsymbol{f}} = \{\hat{f}_j : j = 1, 2, \ldots, N\}$ and out-of-sample forecasts $\hat{\boldsymbol{f}}_\tau = \{\hat{f}_t : t = 1, 2, \ldots, \tau\}$ for a given planning horizon $\tau$, e.g., the next five working-days. As a result, given an initial cash balance $b_0$, the solution to the CMP, namely, the policy $X$ that minimizes the sum of transaction and holding costs up to time step $\tau$, can be obtained by solving the following program:

$$\min \quad \left[\frac{w_1}{C_{max}} \cdot C(X) + \frac{w_2}{R_{max}} \cdot R(X)\right] \tag{9}$$

subject to:

$$\hat{b}_t = \hat{b}_{t-1} + x_t + \hat{f}_t \tag{10}$$

$$\hat{b}_t \geq 0 \tag{11}$$

$$t = 1, 2, \ldots, \tau. \tag{12}$$

where actual balances $b_t$ and cash flows $f_t$ in equation (2) are replaced with predictions $\hat{b}_t$ and $\hat{f}_t$. Since cash managers usually discard policies including overdrafts, we restrict the feasibility space to non-negative cash balances which is equivalent to set $u = \infty$ in equation (4). Unlike in constraint (8), we do not impose any form of policy $X$ allowing a more general and flexible policy than bound-based models. As a result, if both $C(X)$ and $R(X)$ are linear we are dealing with a linear program. However, if $R(X)$ is defined as the standard deviation of daily as in Section 2, we are dealing with a quadratic program.

Both formulations can be solved using commercial mathematical programming solvers such as CPLEX or Gurobi.

Within this framework, considering a purely random cash flow process is equivalent to replace forecasts with random shocks sampled from a particular distribution function such as a Gaussian. However, we here assume that forecasts with a certain degree of accuracy are available to be used as a key input towards a data-driven cash management. Next, we propose a data-driven procedure to deal with the uncertainty introduced by cash flow forecasts in the cash management problem.

## 3.2 Dealing with cash flow uncertainty

The classical approach to deal with uncertainty is stochastic programming (Birge and Louveaux, 2011; Prékopa, 2013). However, this approach may even lead to a soft constrained problem as pointed out by Ben-Tal and Nemirovski (1999). In this paper, we use the concept of robust counterpart of an optimization problem (Ben-Tal and Nemirovski, 1999; Ben-Tal et al, 2009) to encode the cash management problem as an optimization problem with hard constraints, namely, those which must be satisfied whatever the realization of the uncertainty introduced by cash flow forecasting errors.

In Section 3.1, we assumed the availability of two data sets of length $N$: one with past observations $\boldsymbol{f}$, and another one with predictions $\hat{\boldsymbol{f}}$. From those vectors, we are in a position to estimate the uncertainty introduced by cash flow forecasts by means of a third vector with (not-necessarily Gaussian) forecasting errors as follows:

$$\boldsymbol{e} = \boldsymbol{f} - \hat{\boldsymbol{f}}. \tag{13}$$

Thus, following the robust optimization approach as suggested by Ben-Tal and Nemirovski (1999), we replace equation (11) with a minimum balance constraint:

$$\hat{b}_t \geq b_{min} \tag{14}$$

where $b_{min}$ is set to:

$$b_{min} = \xi \cdot \text{std}(\overline{e}(\tau)). \tag{15}$$

The operator std computes the standard deviation of the elements of a given vector and parameter $\xi \in \mathbb{R}_+$ is a subjective value chosen by cash managers to reflect their attitude towards risk. The larger the value of $\xi$, the more averse to risk they are. For instance, if we assume normally distributed forecasting errors, setting $\xi = 3$ would be approximately equivalent to ensure a positive cash balance with probability 0.99. Finally, $\overline{e}(\tau)$ is the cumulative forecasting error for a planning horizon of length $\tau$ that can be easily obtained by time-series embedding (Serra et al, 2012).

### 3.3 The forecasting premium

In Section 3.1, we proposed to take a first data-driven step by using forecasts as a key input to derive optimal policies. In Section 3.2, we set a minimum cash balance $b_{min}$ based on past forecasting errors. In what follows, we propose a simple methodology to estimate the reward that can be obtained by improving predictive accuracy. However, since either designing or enhancing any forecasting model has a cost in terms of both time and money, it is important to know if this cost is going to offset the benefits derived from using forecasts. We can estimate the savings associated to predictive accuracy by obtaining a number of synthetic predictions and by evaluating the corresponding policy costs as suggested in Salas-Molina et al (2017). We next generalize this analysis: (i) to a multiobjective framework in which both cost and risk are desired goals; and (ii) to consider any forecasting error distribution.

To this end, let us assume that both a vector of forecasts $\hat{\boldsymbol{f}}_0$ and a vector of actual observations $\boldsymbol{f}_0$ are available. Consequently, a vector of forecasting errors can be computed as $\boldsymbol{e}_0 = \boldsymbol{f}_0 - \hat{\boldsymbol{f}}_0$. Then, by varying a non-negative parameter $p_i$ with $i : 1, 2, \ldots M$, we can generate a number of synthetic predictions from perfect prediction to purely random shocks:

$$\boldsymbol{e}_i = p_i \cdot \boldsymbol{e}_0. \tag{16}$$

Given a vector $\boldsymbol{p}$ with a number of accuracy variations $p_i$, e.g., ranging in $[0.1, 1]$, we can generate accuracy-controlled predictions by deducting $\boldsymbol{e}_i$ from actual cash flow $\boldsymbol{f}_0$:

$$\hat{\boldsymbol{f}}_i = \boldsymbol{f}_0 - \boldsymbol{e}_i \tag{17}$$

Finally, we can evaluate the impact of such accuracy variations for a cash management model using forecasts in terms of both cost and risk with respect to a baseline forecast $\hat{\boldsymbol{f}}_0$ as detailed in Algorithm 1. This method that can be applied to any cash management model that accepts forecasts as a key input. Since cash managers usually obtain forecasts for short prediction horizons, e.g., the next five working days, we sample only $\tau$ forecasts from the available data set (line 9).

### 3.4 An illustrative example

In order to illustrate our forecasting premium method, we experiment on a real data set of 2717 past cash flow observations $\hat{\boldsymbol{f}}_0$ from an industrial company in Spain. We obtain forecasts using random forests and a set of date-related explanatory variables as described in Salas-Molina et al (2017). From the set of forecasts and actual observations, we compute forecasting error vector $\boldsymbol{e}_0$. Major cash flows (Stone and Miller, 1987), such as loan payments or taxes, are usually known with certainty for short-term planning horizons. In addition, cash managers recompute policies after a short period of time when new information about actual cash balances is available. As a result, we here limit

**Algorithm 1:** Forecasting premium algorithm

**Input**: A vector of actual cash flows $\boldsymbol{f}_0$, a vector of errors $\boldsymbol{e}_0$, a cash management model $m$, policy cost $C(X)$, policy risk $R(X)$, accuracy vector $\boldsymbol{p}$, cost structure $\beta$, risk parameter $\xi$, number of replicates $r$, planning horizon $\tau$;
**Output**: Cost and risk forecasting premium;
**for** *each element* $p_i$ **do**
  Compute vector $\boldsymbol{e}_i = p_i \cdot \boldsymbol{e}_0$;
  Compute vector $\hat{\boldsymbol{f}}_i = \boldsymbol{f}_0 - \boldsymbol{e}_i$;
  Compute $\tau$-cumulative vector $\overline{e}_i(\tau)$ and set $b_{min} = \xi \cdot \text{std}(\overline{e}_i(\tau))$;
  **for** $j = 1, 2, \ldots, r$ **do**
    Set an initial balance $b_0$;
    Sample $\tau$ ordered forecasts from $\hat{\boldsymbol{f}}_i$ and store in $\hat{\boldsymbol{f}}_\tau$;
    Obtain policy $X$ using model $m$, $\hat{\boldsymbol{f}}_\tau$, and structure $\beta$;
    Compute cost $C(X)$ and $R(X)$ for actual cash flows $\boldsymbol{f}_0$;
  **end**
  Compute average cost and risk for each element $p_i$;
**end**

our analysis to robust optimal policies for a planning horizon of $\tau = 5$ days, equivalent to a working-week. Current bank practices in Spain are adjusted to the ranges proposed in da Costa Moraes and Nagano (2014) for numerical evaluation. Then, we select from those ranges the following representative cost structure:

$$\beta = \{\gamma_0^+ = 20€, \gamma_0^- = 20€, \gamma_1^+ = 0.01\%, \gamma_1^- = 0.01\%, h = 0.02\%\} \qquad (18)$$

Following the time delay embedding procedure described in Serra et al (2012), we obtain a 5-cumulative empirical error vector to set the minimum balance constraint with parameter $\xi = 3$, equivalent to three standard deviations of the empirical error distribution. Finally, we follow the multiobjective approach described in this paper to minimize cost and risk. Hence, we define cost $C(X)$ as the average daily cost as in equation (5) and risk $R(X)$ as the standard deviation of daily cost as in equation (6).

Given a cost structure $\beta$, an initial cash balance $b_0$ set to an arbitrary value 20% above $b_{min}$, and a number of controlled-accuracy forecasts $\hat{\boldsymbol{f}}_i$, the ultimate goal is to solve the CMP encoded from equation (9) to (12). We replicate this problem 100 times for different error proportions $p_i$ ranging from 0.1 to 1 taking steps of size 0.1. Since we sample forecasts of length five, we are evaluating an equivalent planning horizon longer than two years. The median results from this experiment grouped by error proportion ($p$) are shown in Figure 2.

Losses are computed by means of equation (9). We report median values instead of averages as a way to reduce the impact of extreme values on averages. Since we set $C_{max}$ and $R_{max}$ to the expected cost and risk of a trivial policy consisting in taking no control action, the values below one indicate a better performance than a trivial strategy. As a result, the lower the loss the higher the opportunity to achieves savings from predictive accuracy.

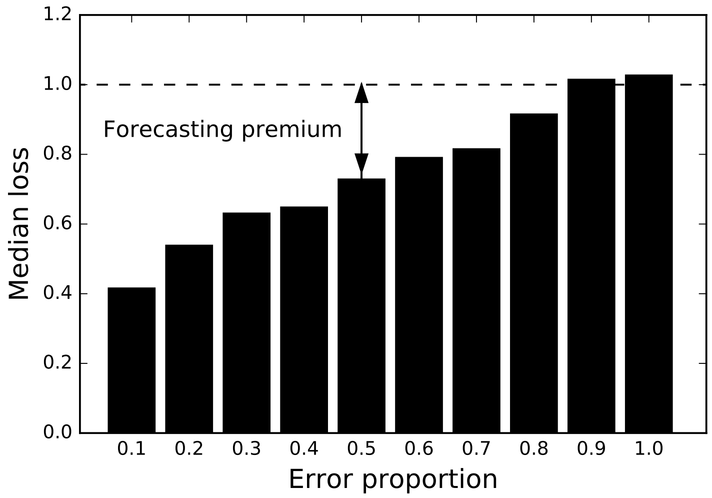


Fig. 2: The forecasting premium.

These results show that remarkable loss savings can be achieved for low error proportions. For instance, if a hypothetical cash manager is able to reduce forecasting errors up to a 50% from the benchmark, it is expected that 30% of the total loss will be saved. In other words, if the current loss derived from taking no control action is one million euros, three hundred thousand euros can be saved by deploying a data-driven cash management model with forecasts as a key input. This saved amount is the forecasting premium. As a rule of thumb, the forecasting premium for a given error proportion is one minus the expected loss (e.g. 30% for error 0.5). As a result, if the cost of improving forecasting accuracy is less than the forecasting premium, the effort to achieve such an accuracy is worthwhile.

## 4 An integrated data-driven approach to cash management

The forecasting premium computed in the previous section may support cash management decision-making in a closed-loop manner. In this section, we integrate this forecasting premium in a data-driven methodology in which machine learning best practices are used in combination to multiple criteria decision-making techniques to solve the CMP. The state-of-the-art approaches mainly focus on finding the cash management model that best fits a particular cash flow distribution to derive cash policies as depicted in Figure 3.

However, recent research (Gormley and Meade, 2007; Salas-Molina et al, 2017) on the use of forecasts to reduce costs must encourage cash managers to incorporate machine learning in cash management. Synthetically, machine learning is concerned with the question of how to construct computer programs

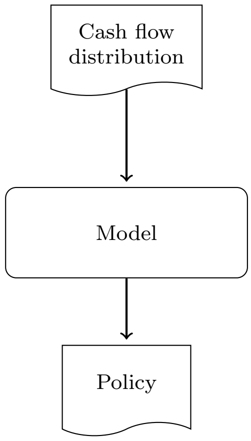


Fig. 3: The state-of-the-art approach to cash management.

that automatically improve with experience (Mitchell, 1997). More formally, machine learning is the field of artificial intelligence that deals with making computers modify or adapt their actions so that these actions get more accurate, where accuracy is measured by how well the chosen actions reflect the correct ones (Marsland, 2009).

Within the context of machine learning, we pay special attention to time-series forecasting techniques as a way to reduce uncertainty about the near future as we showed in Section 3.3. The main concern of time-series forecasting is the construction a model to predict the value of a variable of interest with sufficient accuracy. This model usually takes the form of a mathematical expression relating the variable of interest (cash flows $f_t$, in our particular case) with a number of independent variables (e.g., summarized in vector $\boldsymbol{z}_t$):

$$f_t = g(\boldsymbol{z}_t) + \varepsilon_t \tag{19}$$

where $\varepsilon_t$ is the estimation error. Fitting this model is equivalent to find a particular form of function $g(\boldsymbol{z}_t)$ that minimizes errors when using past observations. Later, the model is used to obtain predictions $\hat{f}_t$ of the variable of interest for an instance of the vector of dependent variables. Alternative techniques for time-series forecasting can be found in general forecasting works such as Box et al (2008) and Makridakis et al (2008). For a specific application to cash flow forecasting, we refer the interested reader to the works by Miller and Stone (1985); Gormley and Meade (2007) and Salas-Molina et al (2017).

As a result, we here propose a data-driven multiobjective procedure comprising the main findings in this paper. Figure 4 graphically depicts the workflow of our methodology. Since we rely on cash flow forecasts to reduce the uncertainty about the near future in daily cash management, cash managers must first select and validate a cash flow forecaster (see e.g. Salas-Molina et al (2017)). In the usual context of high penalty costs for negative cash balances

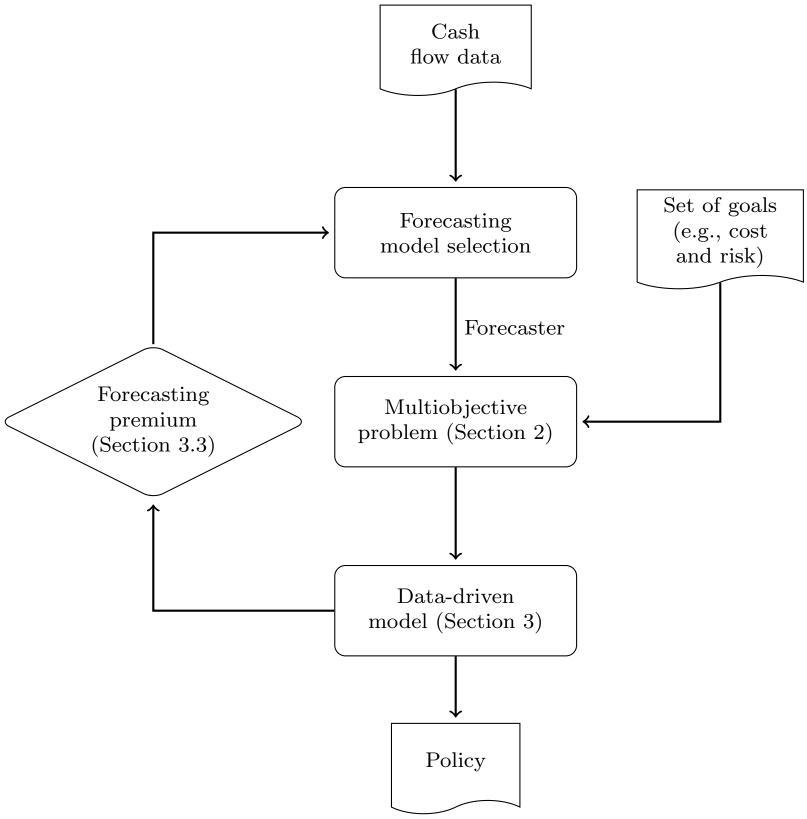


Fig. 4: Data-driven multiobjective cash management.

in comparison to holding costs for positive ones, cash managers may be interested in considering risk as an additional goal to be minimized. Then, once a forecaster has been deployed, we argue that the utility of forecasts can be exploited not only in a single objective framework but in a multiobjective one in which not only the cost but also the risk of alternative policies are considered as described in Section 2. Within this framework, a particularly useful model for cash management is the one presented in Section 3.1, which departs from state-of-the-art bound-based models by placing no restriction on the form of the policy. This boundless model uses forecasts as a key input as a first data-driven procedure. However, the use of forecasts necessarily affects the performance of the model due to predictive accuracy.

The implications derived from cash flow forecasting errors are twofold. On the one hand, some precautionary cash balance is required to avoid undesirable overdraft situations. In order to deal with the uncertainty introduced by forecasting errors, we suggest in Section 3.2 a second data-driven procedure to set this minimum cash balance in terms of the empirical error distribution. On the other hand, cash managers may be interested in estimating the reward that can be achieved by improving the predictive accuracy, i.e., the forecasting premium. Finally, if this forecasting premium is high enough to dedicate efforts in selecting a new forecaster, the process may eventually start again allowing the improvement of forecasting techniques and, ultimately, the performance of deployed cash management policies.

Summarizing, we show in Table 1 the most relevant works related to the methodology proposed in this paper in terms of the use of cash flow forecasts in multiobjective cash management. To this end, we analyze four important dimensions of the cash management problem: (i) the cash management model used; (ii) the forecasting technique used; (iii) the goals considered for optimization purposes; and (iv) the treatment of the inherent uncertainty associated to forecasting errors. From the analysis of Table 1, it is important to highlight that the use of forecasts in the cash management problem dates back to the model proposed in Stone (1972). Although no forecaster was proposed in this paper, the author published several works on cash flow forecasting based on seasonal interaction models (Miller and Stone, 1985; Stone and Miller, 1987). Furthermore, the author proposed a method based on error tolerances to control uncertainty associated to forecasting errors. Interestingly, Gormley and Meade (2007) did not compare their model based on forecasts to the model proposed by Stone (1972). To the best of our knowledge, Salas-Molina et al (2016) were the first to propose a multiobjective framework to deal with the cash management problem considering not only the cost but also the risk of alternative policies and using the Miller and Orr (1966) model. On the other hand, Salas-Molina et al (2017) compared alternative forecasting models to show that predictive accuracy is correlated with cost savings in cash management using the Gormley and Meade (2007).

In this paper, we follow the multiobjective cash management model described in Salas-Molina et al (2016) linking the concepts of machine learning and multiple criteria decision-making. Since we here propose a general data-driven methodology (see Figure 4), we do not recommend the deployment of a any forecaster and we leave this decision to practitioners. A particular feature of our proposal is the use of data-driven techniques no only to predict future cash flows but also to deal with uncertainty by means of setting minimum cash balances based on cumulative forecasting errors as described in Section 3.2.

Table 1: Summary of related works. SI: Seasonal Interaction; AR: Autoregression; R: Linear Regression, RBF: Radial Basis Functions, RF: Random Forest.

| Work | Cash model | Forecaster | Goals | Uncertainty |
|---|---|---|---|---|
| Stone (1972) | Stone (1972) | SI | None | Tolerances |
| Gormley and Meade (2007) | Gormley and Meade (2007) | AR | Cost | None |
| Salas-Molina et al (2016) | Miller and Orr (1966) | None | Cost, risk | None |
| Salas-Molina et al (2017) | Gormley and Meade (2007) | AR, R, RBF, SI, RF | Cost | None |
| This paper | Salas-Molina et al (2016) | None | Cost, risk | Minimum balances |

## 5 Conclusions

In this paper, we provide useful tools to extract knowledge from cash flow data in an attempt to improve decision-making in cash management by closing the gap between machine learning theory and daily financial practice. Unlike existing literature, instead of assuming a particular cash flow distribution, we rely on data-driven multiobjective optimization techniques to derive cash policies. We base our approach on the use of forecasts as a key input in cash management as a data-driven procedure to reduce uncertainty about the near future. We use both past forecasts and actual observations to compute the forecasting premium as an estimation of savings that can be obtained in exchange of accuracy. Since cash managers may be interested in risk analysis to avoid an overdraft, we consider both cost and risk (but possibly others) as suitable goals to estimate potential savings. We also rely on a data-driven procedure to set minimum balances based on past error forecasts to deal with the inherent uncertainty associated to cash management.

As a result, we provide cash managers with a general methodology to improve decision-making through the use of available data rather than assuming any theoretical probability distribution. On the one hand, we suggest to design cash management models based on forecasts rather than on a set of bounds. A further advantage of our approach is that we can formulate the optimization problem as a linear-quadratic program that can be solved using state-of-the-art solvers such as CPLEX or Gurobi. On the other hand, by comparing the forecasting premium (if any) to the cost of deploying any forecasting technique, cash managers are able to decide if the use of forecasts is worthwhile.

Although the results in this paper extend the body of work on the use of data-driven techniques to solve the cash management problem, we restrict ourselves to the use of forecasts (and forecasting errors) to deal with uncertainty. Other machine learning techniques are in place to improve cash management in a similar way. In addition, multiple criteria decision-making in cash management is currently an active area of research and alternative techniques to those described in this paper can be considered to enhance the understanding of the problem.

On the other hand, our results imply that cash management can be enriched by combining machine learning best practices and multiobjective decision-making. Furthermore, we pay special attention to application issues in an attempt to connect theory and practice in cash management. Indeed, both the algorithms and the mathematical programs presented in this paper are ready to be implemented in cash management decision support systems. Natural extensions of this work include the analysis of cash management systems with multiple bank accounts and the study of the pros and cons of alternative measures of risk.

## References

Baccarin S (2009) Optimal impulse control for a multidimensional cash management system with generalized cost functions. Eur J Op Res 196(1):198–206

Ballestero E, Romero C (1998) Multiple criteria decision making and its applications to economic problems. Springer Science & Business Media

Baumol WJ (1952) The transactions demand for cash: An inventory theoretic approach. The Q J Econ 66(4):545–556

Ben-Tal A, Nemirovski A (1999) Robust solutions of uncertain linear programs. Oper Res Lett 25(1):1–13

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) Robust optimization. Princeton University Press

Birge JR, Louveaux F (2011) Introduction to stochastic programming. Springer Science & Business Media

Box GE, Jenkins GM, Reinsel GC (2008) Time series analysis: forecasting and control. John Wiley & Sons, New Jersey

Constantinides GM, Richard SF (1978) Existence of optimal simple policies for discounted-cost inventory and cash management in continuous time. Oper Res 26(4):620–636

da Costa Moraes MB, Nagano MS (2014) Evolutionary models in cash management policies with multiple assets. Econ Model 39:1–7

da Costa Moraes MB, Nagano MS, Sobreiro VA (2015) Stochastic cash flow management models: A literature review since the 1980s. In: Decision Models in Engineering and Management, Springer, pp 11–28

Doumpos M, Grigoroudis E (2013) Multicriteria decision aid and artificial intelligence: links, theory and applications. John Wiley & Sons

Doumpos M, Zopounidis C, Pardalos PM (2012) Financial decision making using computational intelligence, vol 70. Springer Science & Business Media

Gormley FM, Meade N (2007) The utility of cash flow forecasts in the management of corporate cash balances. Eur J Op Res 182(2):923–935

Gregory G (1976) Cash flow models: a review. Omega 4(6):643–656

Makridakis S, Wheelwright SC, Hyndman RJ (2008) Forecasting methods and applications. John wiley & sons, New Jersey

Marsland S (2009) Machine learning: an algorithmic perspective. CRC press

Miller MH, Orr D (1966) A model of the demand for money by firms. The Q J Econ pp 413–435

Miller TW, Stone BK (1985) Daily cash forecasting and seasonal resolution: Alternative models and techniques for using the distribution approach. J Financ Quant Anal 20(03):335–351

Mitchell TM (1997) Machine learning. McGraw-Hill

Prékopa A (2013) Stochastic programming, vol 324. Springer Science & Business Media

Premachandra I (2004) A diffusion approximation model for managing cash in firms: An alternative approach to the miller–orr model. Eur J Op Res 157(1):218–226

Ross SA, Westerfield R, Jordan BD (2002) Fundamentals of corporate finance, sixth edn. McGraw-Hill

Salas-Molina F, Pla-Santamaria D, Rodriguez-Aguilar JA (2016) A multi-objective approach to the cash management problem. Annals Oper Res DOI 10.1007/s10479-016-2359-1

Salas-Molina F, Martin FJ, Rodríguez-Aguilar JA, Serrà J, Arcos JL (2017) Empowering cash managers to achieve cost savings by improving predictive accuracy. Int J Forecast 33(2):403–415

Serra J, Kantz H, Serra X, Andrzejak RG (2012) Predictability of music descriptor time series and its application to cover song detection. IEEE Transactions on Audio, Speech, and Language Processing 20(2):514–525

Srinivasan V, Kim YH (1986) Deterministic cash flow management: state of the art and research directions. Omega 14(2):145–166

Stone BK (1972) The use of forecasts and smoothing in control-limit models for cash management. Financ Manag pp 72–84

Stone BK, Miller TW (1987) Daily cash forecasting with multiplicative models of cash flow patterns. Financ Manag 16(4):45–54

Yu PL (1985) Multiple criteria decision making: concepts, techniques and extensions. Plenum Press

Zeleny M (1982) Multiple criteria decision making. McGraw-Hill

Zopounidis C (1999) Multicriteria decision aid in financial management. Eur J Op Res 119(2):404–415

Zopounidis C, Doumpos M (2013) Multicriteria decision systems for financial problems. Top 21(2):241–261