A Data-Driven Approach for a Class of Stochastic Dynamic Optimization Problems

Thuener Silva\textsuperscript{1}, Davi Valladão\textsuperscript{1}, and Tito Homem-de-Mello\textsuperscript{2}

\textsuperscript{1}Industrial Engineering Department, PUC-Rio, Rio de Janeiro, Brazil
\textsuperscript{2}School of Business, Universidad Adolfo Ibáñez, Santiago, Chile

October 16, 2019

Abstract

Dynamic stochastic optimization models provide a powerful tool to represent sequential decision-making processes. Typically, these models use statistical predictive methods to capture the structure of the underlying stochastic process without taking into consideration estimation errors and model misspecification. In this context, we propose a data-driven prescriptive analytics framework aiming to integrate the machine learning and dynamic optimization machinery in a consistent and efficient way to build a bridge from data to decisions. The proposed framework tackle a relevant class of dynamic decision problems comprising many important practical applications. The basic building blocks of our proposed framework are: (i) a Hidden Markov Model as a predictive (machine learning) method to represent uncertainty; and (ii) a distributionally robust dynamic optimization model as a prescriptive method that takes into account estimation errors associated with the predictive model and allows for control of the risk associated with decisions. Moreover, we present an evaluation framework to assess out-of-sample performance in rolling horizon schemes. A complete case study on dynamic asset allocation illustrates the proposed framework showing superior out-of-sample performance against selected benchmarks. The numerical results shows the practical importance and applicability of the proposed framework since it extracts valuable information from data to obtain robustified decisions with an empirical certificate of out-of-sample performance evaluation.

1 Introduction

Dynamic stochastic optimization models have a long history in operations research, with applications in many different areas. Such models represent a sequential decision-making process whereby information is revealed in stages and decisions are made based on the information available up to that point. There are of course multiple ways to solve dynamic stochastic optimization models — for instance, (approximate) dynamic programming, Markov decision processes, and multistage
stochastic programs. The choice for a particular approach depends, of course, on the structure of the problem under study.

In this paper we consider dynamic convex stochastic optimization models whereby the uncertainty has a certain structure. Our starting point are well-known models in the literature that assume that the uncertainty observed in each time period does not depend on the uncertainty observed in other time periods — i.e. the stochastic process representing the uncertainty is stagewise independent. This class of models was considered in the seminal paper of Pereira and Pinto (1991), who proposed in that work the now well-established Stochastic Dual Dynamic Programming algorithm (SDDP) to solve such problems. In a nutshell, SDDP works by alternating forward and backward steps: in the forward step a sample path of the process is generated, a corresponding solution is obtained, and in the backward step the value function in each period is successively approximated by linear cuts. While the stagewise independence assumption is indeed strong, it avoids the “curse of dimensionality” observed in traditional multistage stochastic programs, which allows for the solution of large problems with long time horizons. Moreover, oftentimes it is possible to work around that assumption by making changes in the model. For instance, the case where the underlying stochastic process follows an auto-regressive model can be written in the stagewise independent form by augmenting the state space (see, e.g., Infanger and Morton (1996)).

Models with stagewise independence, combined with SDDP as a solution technique, have indeed become very popular in the literature, especially for long-term energy planning (Pereira and Pinto 1991; Shapiro et al. 2013) but also in finance (Valladao et al. 2018), transportation (Fhoula et al. 2013) and even mine planning (Reus et al. 2019). SDDP has two particularly attractive features: first, it can solve large-scale stochastic dynamic problems, and second, it provides a policy rather than just a numerical solution. Such features are accomplished through the development of a sequence of the aforementioned state-independent piecewise-linear approximations of the value function in each stage. Once the approximations are built, one can evaluate the optimal policy for an arbitrary realization of the stochastic process by solving a sequence of linear programs.

Despite the success of models assuming stagewise independence, it is unquestionable that in many practical situations such an assumption is not satisfied, even with “tricks” such as state-space augmentation. Some authors have replaced the stagewise independence assumption with that of Markovian dependence, though in that case the solution methods are not as efficient as in the stagewise independent case (Mo et al. 2001; Philpott and de Matos 2012; Löhndorf and Shapiro 2019). Moreover, such models require the user to input the transition probability matrices.

In this paper we take a different approach: we use a Hidden Markov Model (HMM) to learn the dependence directly from the data, see Rabiner (1989) for a comprehensive tutorial. Although HMMs has been studied for decades, it appears that their use in the context of optimization models is limited. Roughly speaking, HMMs create a Markov chain from the data to model the evolution of the process, but the actual states of this Markov chain are not observable. Since each (unobservable) state corresponds to situations where the underlying stochastic process behaves similarly, we replace
the stagewise independence assumption with the following:

**Assumption 1.** The distribution of each element $\xi_t$ of the underlying vector-valued stochastic process depends only on $t$ and on the current (unobservable) state of the Markov chain $\{K_t\}$.

Assumption 1 allows for modeling different “states” of the system. For instance, in the financial model discussed in Section 5 the states correspond to the market being in a “bull”, “bear”, or “regular” state. It is important to stress that such states are unobservable to the user — rather, they are learned directly from the data. Assumption 1 also implies that, conditionally on each given (unobservable) state of the Markov chain $\{K_t\}$, the underlying stochastic process $\{\xi_t\}$ is stagewise independent. That is,

$$P (\xi_t \in A, \xi_{t+1} \in B \mid K_t = j, K_{t+1} = k) = P (\xi_t \in A \mid K_t = j) \cdot P (\xi_{t+1} \in B \mid K_{t+1} = k).$$

One drawback of HMMs, however, is that since the probabilities of transition between pairs of states are estimated from the data, the solutions of the optimization model that uses such transition probabilities will be very much dependent on the observed data. Such dependence may then lead to poor *out-of-sample performance* of the resulting policies. We prevent such phenomenon from happening by employing a distributionally robust optimization (DRO) approach that allows for variations in the estimated transition probability matrix of the HMM. Our DRO model leads to tractable formulations that do not increase the complexity of the model; moreover, they can be solved by a variation of the SDDP. Unlike standard SDDP, however, the calculation of lower and/or upper bounds for the optimal value of the problem cannot be estimated simply from Monte Carlo simulations of the objective function at the current solution. For the DRO model, we provide lower and upper bounds and show that arguments for convergence of standard SDDP can be applied to our case.

The concern about of out-of-sample performance is, unfortunately, often overlooked in the stochastic optimization literature, particularly in the case of dynamic models — oftentimes the user simply implements the first-stage decision given by the model, and then re-solves the model in every time period in order to obtain new decisions. As we discuss in the paper, such procedure can be expensive and wasteful, as it discards the value function approximations obtained in previous steps of a rolling horizon scheme. To counter that effect, we propose an extra step in the out-of-sample evaluation that allows for improvement of the current value function approximations for updated values of the previous decisions (initial conditions of the current problem). Such feature makes the policies generated by the algorithm easier to use and quicker to evaluate.

We illustrate our ideas with a dynamic asset allocation problem. Such problem consists of decision processes under uncertainty with complex characteristics embedding the investors risk tolerance, transaction cost and price dynamics. By building upon previous work, we propose a Data-Driven DRO approach that builds an HMM for the return process and allows for ambiguity in the transition probability matrix with the thrust to enhance out-of-sample performance. Our
numerical tests show that the resulting portfolio can yield excellent results, with enhanced out-of-sample performance over selected benchmarks including the equal-weight-allocation strategy which has been shown to be optimal under certain assumptions [Esfahani and Kuhn 2017, Pflug et al. 2012] with an competitive out-of-sample (empirical) performance [DeMiguel et al. 2007].

2 A data-driven prescriptive analytics framework

We start by presenting a data-driven prescriptive analytics framework that integrate all the machine learning and optimization machinery in a consistent and efficient way to build a bridge from data to decisions. The basic building blocks of our proposed framework are: (i) a predictive (machine learning) method to represent uncertainty; and (ii) a prescriptive (optimization) model that takes into account estimation errors associated with the predictive model and allows for control of the risk associated with decisions. In our context, the predictive model is a Hidden Markov Model, and the prescriptive model is a distributionally robust dynamic optimization model with risk-based constraints that induces a (parameterized) level of robustness over the HMM transition probabilities given that they might be polluted with estimation errors.

In what follows we describe these building blocks in more detail. Before that, however, we establish the notation for the dynamic stochastic optimization problem of interest. We consider a filtered probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \), where \( \mathcal{F} = \mathcal{F}_T \) and \( \mathcal{F}_0 \subseteq \ldots \subseteq \mathcal{F}_T \). The input process is represented by a stochastic process \( \{\xi_t\} \) with values in \( \mathbb{R}^m \), and a Markov chain \( \{K_t\} \) with unobservable states such that Assumption 1 is satisfied. The sigma-algebra generated by \( \{\xi_s, K_s\}_{s=0}^t \) is contained in \( \mathcal{F}_t \). The feasibility set in each time period consists of \( \mathcal{F}_t \)-adapted solutions \( x_t \in \mathbb{R}^n \) such that \( x_t \in \mathcal{X}_t(x_{t-1}, \xi_t) \), where the set \( \mathcal{X}_t(x_{t-1}, \xi_t) \) is given by linear inequalities which may involve \( x_{t-1} \) and \( \xi_t \). We will detail the formulation of the model shortly.

2.1 Predictive method: The Hidden Markov Model framework

Following standard practice in the literature, given historical data of the underlying stochastic process \( \{\xi_t\} \) the database is split into two parts: training and test datasets. We define the training dataset from the oldest entry until a pre-specified date, while the test dataset comprises the remainder. The training data are used to estimate the parameters of the HMM; for instance, the standard EM (expectation-maximization) algorithm [Moon 1996] can be used to estimate for instance means, variances and covariances from data, as well as the nominal transition probability matrix.
In general, the HMM parameters estimated from data are the transition probabilities $\hat{p}_j(k) = P(K_{t+1} = k \mid K_t = j), \forall t = 0, \ldots, T - 1,$ and the coefficients $\Theta$ associated with conditional density function $p(\xi_t \mid K_t = k; \Theta)$. Once the parameters are estimated, the HMM can then be used to classify the current state $j$ of the Markov chain and therefore the (conditional) distribution of stochastic process $\{\xi_t\}$, which due to Assumption 1 depends only on $j$. The current state classification is obtained by conducting statistical inference of the current state given all information so far. More specifically, we obtain from the HMM parameters the quantity $P(K_t = k \mid \xi_t, \xi_{t-1}, \ldots, \xi_1), \forall k \in K$, hereinafter referred to as the posterior probability of state $k$ at time $t$, which is defined as

$$P(K_t = k \mid \xi_t, \xi_{t-1}, \ldots, \xi_1) := \frac{p(K_t = k, \xi_t, \xi_{t-1}, \ldots, \xi_1)}{\sum_{j \in K} p(K_t = j, \xi_t, \xi_{t-1}, \ldots, \xi_1)}, \quad (1)$$

where $p(\xi_t, \xi_{t-1}, \ldots, \xi_1, K_t = k)$ is the joint probability density function evaluated at the observed sample path and the current Markov state being $k$. This joint probability density function is obtained by an iterative procedure called the HMM forward pass, see Bishop (2006). Now, we can classify the current state

$$k^*_t \in \arg\max_{k \in K} P(K_t = k \mid \xi_t, \xi_{t-1}, \ldots, \xi_1) \quad (2)$$

as the most probable Markov state given all available information at time $t$. The process is illustrated in Figure 2.

### 2.2 Basic prescriptive method: Risk-constrained dynamic stochastic programming

In addition to the aforementioned constraints $x_t \in X_t(x_{t-1}, \xi_t)$, we also consider risk-based constraints in our optimization model. The introduction of such constraints allows us to control the risk associated with decisions. The incorporation of risk control in dynamic stochastic optimization models has been subject of much research in the literature, since issues such as time consistency must be taken into account; see, e.g., Shapiro et al. (2014), Rudloff et al. (2014) for discussions. As we shall see shortly, our risk-constrained model satisfies time consistency. This dynamic stochastic
programming model will then be extended to a distributionally robust dynamic model in Section 2.3.

Before describing the risk-constrained dynamic model, we briefly recall the notion of Conditional Value-at-Risk (CV@R) risk measure defined in Rockafellar and Uryasev (2000). Given a random variable $Z$ representing some quantity such that larger values are less favorable (for instance, losses), we write

$$CV@R_\beta[Z] = \min_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{1-\beta} \mathbb{E} \left[ (Z - \eta)_+ \right] \right\}.$$  

This risk measure is concentrated on the right tail of the distribution of $Z$. When the random variable of interest is such that larger values are more favorable, then it is more appropriate to refer to acceptability functionals rather than to risk measures (see, e.g. Roorda et al. (2005)). We call this variable of interest “wealth” and denote it by $W$. Also, we denote by $\phi_\alpha[W]$ the acceptability functional corresponding to CV@R, i.e., $\phi_\alpha[W] := -CV@R_{1-\alpha}[-W]$, which can be written as (we omit the subscript $\alpha$ as it is fixed throughout the paper)

$$\phi[W] = \max_{z \in \mathbb{R}} \left\{ z - \frac{1}{\alpha} \mathbb{E} \left[ (z - W)_+ \right] \right\}. \quad (3)$$

Following Artzner et al. (1999), Roorda et al. (2005), a coherent acceptability functional $\phi$ satisfies the following properties for any random wealth variables $W, W^1, W^2$:

(i) $\phi[W + c] = \phi[W] + c$ for any constant $c$.

(ii) $\phi[\lambda W^1 + (1-\lambda)W^2] \geq \lambda \phi[W^1] + (1-\lambda)\phi[W^2]$ for any $\lambda \in [0,1]$.

(iii) $\phi[W^1] \leq \phi[W^2]$ if $W^1 \leq W^2$ a.s.

---

1Because of the direct one-to-one relationship between $\phi$ and $CV@R_{1-\alpha}$, in the paper we will often refer to $\phi$ as “CV@R”, with the meaning understood from the context.
We return now to our model. In each period \( t \), given a feasible solution \( x_t \in \mathbb{R}^n \) we define a function \( g_t(x_t, \xi_{t+1}) \) to represent the “wealth” resulting from the decision \( x_t \). Note that \( g_t \) depends on the random variable \( \xi_{t+1} \) which has not been realized yet at time \( t \), but it does not depend on future values \( \xi_{t+2}, \ldots, \xi_T \). To simplify the notation, let \( W_{t+1} := g_t(x_t, \xi_{t+1}) \). Then, we apply the acceptability functional \( \phi \) defined in \( [3] \) to \( W_{t+1} \) conditionally on the current state \( j \) of the HMM, resulting in the quantity

\[
\phi_{p_j}[W_{t+1}] := \max_{z \in \mathbb{R}} \left\{ z - \frac{1}{\alpha} \sum_{k \in K} \mathbb{E}\left[(z - W_{t+1})_+ | K_{t+1} = k\right] \hat{p}_j(k) \right\}. \tag{4}
\]

With this notation, the risk-based constraint can then be expressed as

\[
\phi_{p_j}[g_t(x_t, \xi_{t+1})] \geq 0.
\]

Finally, in each period \( t \), given a feasible solution \( x_t \in \mathbb{R}^n \) and a realization of \( \xi_t \), a reward of \( f_t(x_t, \xi_t) \) is accrued. For each state \( j \) of the HMM, the optimization model is then written as

\[
Q^j_0 := \max_{x_0 \in X_0} f_0(x_0) + \sum_{k \in K} \mathbb{E}\left[Q^1_k(x_0, \xi_1) | K_1 = k\right] \hat{p}_j(k) \tag{5}
\]

\[
\text{s.t. } \phi_{p_j}[g_0(x_0, \xi_1)] \geq 0. \tag{6}
\]

where \( X_0 \) represents linear constraints on \( x_0 \) and, for each \( t = 1, \ldots, T - 1 \) and for each state \( j \) of the HMM,

\[
Q^j_t(x_{t-1}, \xi_t) := \max_{x_t \in X_t(x_{t-1}, \xi_t)} f_t(x_t, \xi_t) + \sum_{k \in K} \mathbb{E}\left[Q^1_k(x_t, \xi_{t+1}) | K_{t+1} = k\right] \hat{p}_j(k) \tag{7}
\]

\[
\text{s.t. } \phi_{p_j}[g_t(x_t, \xi_{t+1})] \geq 0. \tag{8}
\]

The final-stage function \( Q_T \) is defined as

\[
Q^j_T(x_{T-1}, \xi_T) := \max_{x_T \in X_T(x_{T-1}, \xi_T)} f_T(x_T, \xi_T). \tag{9}
\]

Equations \( (5)-(9) \) define the risk-averse dynamic stochastic optimization we would like to solve. It is important to notice that, since the risk function is applied only locally in each period through the constraints \( \phi_{p_j}[g_t(x_t, \xi_{t+1})] \geq 0 \), time consistency is ensured (see Valladão et al. (2018)) — after all, the problem still has a nested form, which is a basic condition for time-consistency (see, e.g., Shapiro et al. (2014), Homem-de Mello and Pagnoncelli (2016)). Such constraints could be used, for example, to provide a convex approximation of probabilistic constraints of the form \( \mathbb{P}(g_t(x_t, \xi_{t+1}) \leq 0) \leq \alpha \).

We will make the following assumptions for the remainder of the paper (for convenience we set
Let $g_T \equiv 0$:

**Assumption 2.** For any $t = 0, \ldots, T$ and any realization of $\xi_1, \ldots, \xi_{t+1}$, the functions $f_t(\cdot, \xi_t)$ and $g_t(\cdot, \xi_{t+1})$ are affine.

Assumption 2 can be relaxed; for example, to the case where $f_t$ and $g_t$ are defined as the minimum of affine functions. Nevertheless, we keep the linear assumption for simplicity.

**Assumption 3.** For any $t = 1, \ldots, T$, the set $X_t(x_{t-1}, \xi_t)$ is non-empty and consists of vectors $x_t \in \mathbb{R}^n$ satisfying linear inequalities of the form

$$A_t(\xi_t)x_t = b_t(\xi_t) - B_t(\xi_t)x_{t-1}$$

$$x_t \geq 0.$$

The set $X_0$ has the form $\{A_0x_0 = b_0, \ x_0 \geq 0\}$.

Assumption 3 imposes a polyhedral structure on the set $X_t(x_{t-1}, \xi_t)$, which will be useful in the developments that follow.

### 2.3 Extended prescriptive method: A data-driven distributionally robust dynamic model

The HMM approach described in Section 2.1 has the advantage of learning directly from the data. However, the estimation of probabilities of the HMM is very sensitive to changes in the input parameters. Such sensitivity may cause considerable instability in the optimization model, with similar observed data leading to different performances of the corresponding optimal solutions. Moreover, poorly estimated probabilities will likely lead to poor out-of-sample performance of the solutions proposed by the model. As stated in Section 1, our main goal is to provide a data-driven approach for dynamic stochastic optimization problems which performs well out of sample. Therefore, it is critical to address this estimation shortcoming.

Our approach to circumvent the estimation issue is to use a *distributionally robust optimization* (DRO) model for the problem. The idea of DRO is to construct an ambiguity set for the distributions of the random variables of the problem and then to optimize the worst-case within the ambiguity set. DRO problems have long been studied in the literature (albeit with a different terminology), starting from a seminal paper by Scarf (1958), then followed by Záčková (1966) and later by Shapiro and Kleywegt (2002), Shapiro and Ahmed (2004) and Goh and Sim (2010). Much of the recent literature on this topic focuses on ways of constructing the ambiguity set (call it $\mathcal{P}$) that ensure tractability of the resulting problem. For example, Delage and Ye (2010) define $\mathcal{P}$ as the set of distributions that have a given mean and covariance matrix. Another popular approach is to define $\mathcal{P}$ as the set of distributions that are not “too far” from a some reference distribution. Of course, such a notion requires defining an appropriate way to measure the “distance” between
Several such distances exist, for instance the Kantorovich and Wasserstein distances, the Kullback-Leibler divergence, and Chi-squared distance (and more generally phi-divergences), among others. This is a growing field with a lot of current activity; we refer to Pflug and Wozabal (2007), Ben-Tal et al. (2013), Bayraksan and Love (2015), Esfahani and Kuhn (2017), Rahimian et al. (2018) for some of the work in this area. The benefit of using DRO — under certain settings — to improve out-of-sample performance can be formally demonstrated; see Van Parys et al. (2017).

In the DRO model, equations (7)-(8) are replaced with the following:

\[
Q^t(x_{t-1}, \xi_t) = \max_{x_t \in X_t(x_{t-1}, \xi_t)} f_t(x_t, \xi_t) + \left\{ \min_{p_j \in P_j} \sum_{k \in K} \mathbb{E} \left[ Q_{t+1}^k(x_t, \xi_{t+1}) \right] | K_{t+1} = k \right\} p_j(k) \quad (10)
\]

subject to

\[
\left\{ \min_{p_j \in P_j} \phi_{p_j} \left[ g_t(x_t, \xi_{t+1}) \right] \right\} \geq 0 \quad (11)
\]

and equations (5) and (6) are also replaced accordingly. In the above model, \( P_j \) is the ambiguity set for the distribution of the next state of the HMM, conditionally on the current state \( j \), and is defined as

\[
P_j = \left\{ p_j \in \mathbb{R}^{\vert K \vert} \mid \begin{array}{c}
1^T p_j = 1 \\
d(p_j, \hat{p}_j) \leq \Delta \\
p_j \geq 0
\end{array} \right\}, \quad (12)
\]

where \( d(p_j, \hat{p}_j) \) measures the total variation distance between \( p_j \) and \( \hat{p}_j \) (recall that \( \hat{p}_j \) is the vector of state-\( j \) probabilities estimated for the HMM), i.e.,

\[
d(p_j, \hat{p}_j) := (1/2) 1^T | p_j - \hat{p}_j |. \quad (13)
\]

In the above formulation, the parameter \( \Delta \) controls the level of ambiguity allowed in the model — a value of \( \Delta = 0 \) indicates that the estimated probabilities can be fully trusted, whereas a value of \( \Delta = 1 \) ignores the estimated probabilities and simply optimizes with respect to the worst-case state of the HMM.

The use of a distributionally robust model for the transition probabilities of the HMM affects both the objective function and the CV@R constraint — note that the expression \( \min_{p_j \in P_j} \{ \cdot \} \) appears in both places. Using the same worst-case probability distribution for both the objective function and the CV@R constraint makes the separation between the inner and outer problem impossible (Pflug and Wozabal 2007). Nevertheless, if we allow for two separated worst-case probability distributions (one for objective function and other for the constraint), the problem becomes more

\footnote{Note that here we use “distance” as an abuse of terminology, since in some of these cases the function is not symmetric, i.e. \( d(P, Q) \neq d(Q, P) \).}
tractable. This “separation” can be conceptually motivated by the idea that the objective function is a “worst case” expectation while the risk constraint must be feasible for any transition probability in the ambiguity set, i.e., \( \phi_{p_j} \left[ g_t(x_t, \xi_{t+1}) \right] \geq 0, \forall p_j \in P_j \). For convenience we will use the same ambiguity set for both the objective function and the constraint, however one could use different sets (for instance, defined with different \( \Delta \)s) to allow for different levels of robustness in each expression.

It is also worthwhile mentioning that, while other probability distance functions could be used instead of the total variation distance in (13), the choice for the total variation is natural in this setting where there are only a finite number of Markov states. Moreover, with the total variation distance the model can be efficiently solved because the robust counterpart is a linear optimization problem (Ben-Tal et al. 2013).

3 Solution methodology

In this section we propose three steps to efficiently solve the data-driven distributionally robust dynamic model: (i) first, we develop a computationally tractable dual reformulation; (ii) second, we generate a Sample Average Approximation (SAA) of the original continuously distributed problem; (iii) finally, we adapt the Stochastic Dual Dynamic Programming (SDDP) algorithm to suit the proposed model. Major modifications are needed in SDDP, in particular, we highlight the development of a deterministic lower bound (for a maximization problem) which, while related to results recently proposed in the literature, has some novel aspects and a practical appeal. In the following subsections we describe all three steps in detail.

3.1 A tractable dual reformulation

In this section we present a tractable formulation of (10)-(11) based on the dual of the inner minimization problems in those equations. Consider initially the inner problem in (10). For a fixed \( x_t \in X_t(x_{t-1}, \xi_t) \), the inner problem is written as

\[
\min_{p_j, e} \sum_{k \in K} E \left[ Q_{t+1}^k (x_t, \xi_{t+1}) \mid K_{t+1} = k \right] p_j(k) \tag{14}
\]

\[\text{s.t.} \quad p_j(k) - e_k \leq \hat{p}_j(k), \quad \forall k \in K : \theta_k^- \]
\[p_j(k) + e_k \geq \hat{p}_j(k), \quad \forall k \in K : \theta_k^+ \]
\[\sum_{k \in K} e_k \leq 2\Delta : \lambda \]
\[\sum_{k \in K} p_j(k) = 1 : \eta \]
\[p_j \geq 0.\]
In order to write the inner optimization problem as a maximization problem so it can be merged with the outer one, we construct the dual formulation. Denote by $\theta_k^-, \theta_k^+, \lambda$ and $\eta$ the dual variables corresponding to the constraints in (14). Then, the dual formulation of the inner problem in (10) becomes

$$\max_{\theta^-, \theta^+, \lambda, \eta} \sum_{k \in K} \hat{p}_j(k)(\theta_k^+ - \theta_k^-) - \eta - 2\Delta \lambda$$

s.t. 
- $-\theta_k^- + \theta_k^+ - \eta \leq E \left[ Q_{t+1}^k(x_t, \xi_{t+1}) \mid K_{t+1} = k \right]$, $\forall k \in K$
- $\theta_k^- + \theta_k^+ - \lambda = 0$, $\forall k \in K$
- $\theta^-, \theta^+, \lambda \geq 0$.

By using a similar approach it is possible to construct a dual formulation to write (11) in a more tractable way. Note that from (4) we can write

$$\phi_{p_j} [g_t(x_t, \xi_{t+1})] = \max_{z \in \mathbb{R}} \left\{ h(z, p_j) := z - \frac{1}{\alpha} \sum_{k \in K} E \left[ (z - g_t(x_t, \xi_{t+1}))_+ \mid K_{t+1} = k \right] p_j(k) \right\}.$$ 

The function $h$ is concave in $z$ and linear in $p_j$. Thus, given that $P$ is compact, we can deduce from Sion’s minimax theorem (Sion et al. 1958) that $\min_{p_j \in P_j} \max_{z \in \mathbb{R}} h(z, p_j) = \max_{z \in \mathbb{R}} \min_{p_j \in P_j} h(z, p_j)$ and hence it follows that for any given $z$, $\min_{p_j \in P_j} h(z, p_j)$ can be written as

$$\min_{p_j \in P_j} \int \left\{ z - \frac{1}{\alpha} E \left[ (z - g_t(x_t, \xi_{t+1}))_+ \mid K_{t+1} = k \right] p_j(k) \right\}$$

s.t. 
- $p_j(k) - e_k \leq \hat{p}_j(k)$, $\forall k \in K$ : $\tilde{\theta}_k^-$
- $p_j(k) + e_k \geq \hat{p}_j(k)$, $\forall k \in K$ : $\tilde{\theta}_k^+$
- $\sum_{k \in K} e_k \leq 2\Delta$ : $\tilde{\lambda}$
- $\sum_{k \in K} p_j(k) = 1$ : $\tilde{\eta}$
- $p_j \geq 0$. 

11
By writing the dual of (16) analogously to (15), it follows that the left-hand side of (11) can be written as the optimization problem

\[
\max_{z, \theta^-, \theta^+, \lambda, \tilde{\eta}} \quad z + \sum_{k \in K} \hat{p}_j(k)(\tilde{\theta}_k^+ - \tilde{\theta}_k^-) - \tilde{\eta} - 2\Delta \tilde{\lambda}
\]

s.t. \[ -\tilde{\theta}_k^- + \tilde{\theta}_k^+ - \tilde{\eta} \leq -\frac{1}{\alpha} \mathbb{E} \left[ (z - g_t(x_t, \xi_{t+1}))_+ \right| K_{t+1} = k], \quad \forall k \in K \]

\[ \tilde{\theta}_k^- + \tilde{\theta}_k^+ - \tilde{\lambda} = 0, \quad \forall k \in K \]

\[ -\tilde{\theta}_k^- + \tilde{\theta}_k^+ - \tilde{\lambda} \geq 0, \]

Finally, to write the complete reformulation of (10)-(11) we must merge the two constructed dual models (15) and (17). Notice first that, since the left-hand side of (11) is given by the maximization problem (17), the max operator present in (17) can be removed as it is equivalent to imposing that there exist at least one value of \( z, \tilde{\theta}^-, \tilde{\theta}^+, \lambda \) and \( \tilde{\eta} \) satisfying

\[ z + \sum_{k \in K} \hat{p}_j(k)(\tilde{\theta}_k^+ - \tilde{\theta}_k^-) - \tilde{\eta} - 2\Delta \tilde{\lambda} \geq 0 \]

and

\[ -\tilde{\theta}_k^- + \tilde{\theta}_k^+ - \tilde{\eta} \leq -\frac{1}{\alpha} \mathbb{E} \left[ (z - g_t(x_t, \xi_{t+1}))_+ \right| K_{t+1} = k], \quad \forall k \in K \]

\[ \tilde{\theta}_k^- + \tilde{\theta}_k^+ - \tilde{\lambda} = 0, \quad \forall k \in K \]

\[ -\tilde{\theta}_k^- + \tilde{\theta}_k^+ - \tilde{\lambda} \geq 0, \]

Thus, by adding the above inequalities and variables to (15) we obtain the final reformulation of (10)-(11):

\[
Q_t^k(x_{t-1}, \xi_t) = \max_{x_t, \theta^-, \theta^+, \lambda, \eta, \tilde{\eta}} \quad f_t(x_t, \xi_t) + \sum_{k \in K} \hat{p}_j(k)(\theta_k^+ - \theta_k^-) - \eta - 2\Delta \lambda
\]

s.t. \[ z + \sum_{k \in K} \hat{p}_j(k)(\theta_k^+ - \theta_k^-) - \eta - 2\Delta \lambda \geq 0 \]

\[ -\theta_k^- + \theta_k^+ - \eta + \frac{1}{\alpha} \mathbb{E} \left[ (z - g_t(x_t, \xi_{t+1}))_+ \right| K_{t+1} = k] \leq 0, \quad \forall k \in K \]

\[ -\theta_k^- + \theta_k^+ - \eta - \mathbb{E} \left[ Q_{t+1}^k(x_t, \xi_{t+1}) \right| K_{t+1} = k] \leq 0, \quad \forall k \in K \]

\[ \theta_k^- + \theta_k^+ - \lambda = 0, \quad \forall k \in K \]

\[ \theta_k^- + \theta_k^+ - \lambda = 0, \quad \forall k \in K \]

\[ \theta^-, \theta^+, \tilde{\theta}^-, \tilde{\theta}^+ \geq 0 \]

\[ x_t \in \mathcal{X}_t(x_{t-1}, \xi_t). \]
Problem (18)-(25) is multistage stochastic program, which is convex by virtue of Assumptions 2 and 3. It provides a conceptually tractable reformulation of (10)-(11). The word “conceptually” refers to the fact that such model cannot be directly implemented, for two reasons: first, inequalities (20) and (21) involve expectations and second, inequality (21) involves the unknown value function $Q_{t+1}^k(x_t, \xi_{t+1})$. The first difficulty can be dealt with by means of sample average approximations, as will see in Section 3.2. The second difficulty appears more daunting due to the curse of dimensionality. For instance, when no assumptions are made about the input process \{\xi_t\} there is a very large number of possible outcomes at each stage and the number of scenarios grow exponentially with the number of stages. As we shall see in Section 3.3 however, under Assumption 1 we can adapt the SDDP method to our setting, which allows us to approximate the value function $Q_{t+1}^k(x_t, \xi_{t+1})$ by piecewise-linear functions and so standard optimization methods can be used to solve the problem.

3.2 A Sample Average Approximation of the model

In this section we construct a SAA of problem (18)-(25), which allows us to replace the expectations in (20) and (21) with averages of random realizations sampled from the “true” distributions. First, for each state $k \in K$ of the HMM, we draw i.i.d. samples from the conditional distribution of $\xi_{t+1}$ given $K_{t+1} = k$. We denote those samples by $\{\xi_{t+1}^k(s)\}_{s \in S_k}$. Next, define the probability $q_k(s)$ of scenario $s$ conditional on state $k$ of the HMM as

$$q_k(s) := P(\xi_{t+1} = \xi_{t+1}^k(s) \mid K_{t+1} = k).$$

Moreover, we introduce variables $y_{ks}$ for each $k \in K$ and each $s \in S_k$ in order to linearize the “plus” function in (20).
We can then write the SAA version of (18)-(25) as follows.

\[
Q_j^t(x_{t-1}, \xi_t) = \max_{x_t, z, y, \theta^-, \theta^+, \lambda, \eta, \tilde{\theta}^-, \tilde{\theta}^+, \tilde{\lambda}, \tilde{\eta}} f_t(x_t, \xi_t) + \sum_{k \in K} \hat{p}_j(k)(\theta_k^+ - \theta_k^-) - \eta - 2\Delta \lambda
\]  

(26)

s.t. \[
\begin{align*}
& z + \sum_{k \in K} \hat{p}_j(k)(\tilde{\theta}_k^+ - \tilde{\theta}_k^-) - \tilde{\eta} - 2\Delta \tilde{\lambda} \geq 0 \\
& -\tilde{\theta}_k^- + \theta_k^+ - \eta - \sum_{s \in S_k} y_{ks} \frac{q_k(s)}{\alpha} \leq 0, \forall k \in K \\
& -\theta_k^- + \theta_k^+ - \lambda = 0, \forall k \in K \\
& \tilde{\theta}_k^- + \tilde{\theta}_k^+ - \tilde{\lambda} = 0, \forall k \in K \\
& z - g_t(x_t, \xi_{t+1}(s)) - y_{ks} \leq 0, \forall k \in K, \forall s \in S_k \\
& x_t \in X_t(x_{t-1}, \xi_t) \\
& \theta^-, \theta^+, \tilde{\theta}^-, \tilde{\lambda}, \lambda, \tilde{\lambda} \geq 0.
\end{align*}
\]  

(27)  

(28)  

(29)  

(30)  

(31)  

(32)  

(33)  

(34)

where \[
Q_{t+1}^k(x_t) := \sum_{s \in S_k} Q_{t+1}^k(x_t, \xi_{t+1}(s)) q_k(s).
\]  

(35)

### 3.3 Modified Stochastic Dual Dynamic Programming algorithm

With model (26)-(34) at hand, the only remaining issue is dealing with the value function in constraint (29). As discussed earlier, we adapt the SDDP method for this purpose. The SDDP algorithm is mainly characterized by two steps: a forward-in-time simulation and a backward-in-time recursion. The forward step generates trial solutions that are later used in the backward step to construct cutting-plane approximations of the future value function. When there is Markovian dependency, the forward step must generate (i) a path of states of the HMM and (ii) sample paths of the process \( \{\xi_t\} \) conditionally on each sampled state of the HMM. Then, trial solutions are created by solving the problem with the current value function approximations at each stage using the sampled processes. The backward step uses trial solutions and goes in the opposite time direction, from \( t = T \) to \( t = 0 \), adding cuts to improve the outer approximation of the value function. In the context of a maximization problem, we can obtain a deterministic upper bound using the outer approximation generated by the SDDP backward procedure.

We remark that model (26)-(34) is not in the standard form of problems solved by SDDP, since the value function appears in the constraint (29) rather in the objective function as customary in the literature. A similar situation arises in the model studied by Philpott and de Matos (2012), albeit in a somewhat different context since that paper deals with nested risk measures. Thus,
for the sake of completeness we detail the steps and show how to construct an upper (i.e., outer) approximation for the value function. In Section 3.4 we will discuss how to construct a lower (inner) approximation.

Suppose we are in iteration \( \nu \) of the algorithm. Consider problem (26)-(34) with constraint (29) replaced with an upper approximation \( \overline{Q}_{t+1}^{j,\nu}(x_t) \) given by linear inequalities, and denote the optimal value of the approximated problem by \( \overline{Q}_t^{j,\nu}(x_{t-1}, \xi_t) \). We will detail shortly how to construct \( \overline{Q}_{t+1}^{j,\nu}(x_t) \). We have, for \( t = T - 1 \) to \( t = 0 \),

\[
\overline{Q}_t^{j,\nu}(x_{t-1}, \xi_t) := \max_{x_t, z, y, \theta^-, \theta^+, \lambda, \eta, \theta^{-}, \theta^{+}, \tilde{\lambda}, \tilde{\eta}, u} f_t(x_t, \xi_t) + \sum_{k \in K} \hat{p}_j(k)(\theta^+_k - \theta^-_k) - \eta - 2\Delta \lambda - 2\Delta \tilde{\lambda} \geq 0
\]

\[
\eta - \theta^+_k - \theta^-_k - \tilde{\eta} - 2\Delta \tilde{\lambda} \leq 0, \quad \forall k \in K
\]

\[
-\theta^-_k + \theta^+_k - \eta - \overline{Q}_{t+1}^{j,\nu}(x_t) \leq 0, \quad \forall k \in K
\]

\[
\theta^-_k + \theta^+_k - \lambda = 0, \quad \forall k \in K
\]

\[
\theta^-_k + \theta^+_k - \tilde{\lambda} = 0, \quad \forall k \in K
\]

\[
z - g_t(x_t, \xi^k_t(s)) - y_{ks} \leq 0, \quad \forall k \in K, \forall s \in S_k
\]

\[
u = x_{t-1} \quad : \pi_t^j(\xi_t)
\]

\[
x_t \in \mathcal{X}_t(u, \xi_t)
\]

\[
\theta^-, \theta^+, \theta^{-}, \theta^{+}, y, \lambda, \tilde{\lambda} \geq 0.
\]

For \( t = T \) we have, at all iterations \( \nu \), the simpler problem

\[
\overline{Q}_T^{j,\nu}(x_{T-1}, \xi_T) := \max_{x_T, u} f_T(x_T, \xi_T)
\]

\[
s.t. \quad u = x_{T-1} \quad : \pi_T^j(\xi_T)
\]

\[
x_T \in \mathcal{X}_T(u, \xi_T).
\]

Note that \( \overline{Q}_T^{j,\nu}(x_{T-1}, \xi_T) = Q_T^{j,\nu}(x_{T-1}, \xi_T), \) i.e., there is no approximation in the last stage.

It is important to clarify the role of the auxiliary variable \( u \) introduced in the problem, which appears in constraints (43), (44), (47) and (48). This variable is just a generic artifact to obtain a subgradient\(^3\) of the function \( \overline{Q}_t^{j,\nu}(x_{t-1}, \xi_t) \) with respect to \( x_{t-1} \). From (43) and (47) we see that this subgradient is given by the dual variable \( \pi_t^j(\xi_t) \). Let \( \tilde{x}_t^j \) be the optimal solution of (36)-(45).

\(^3\)In reality, it is a supergradient since the function is concave, but we will call it a subgradient as this terminology is more common in the literature.
(and (46)-(48) in the case \(t = T\)) generated by the forward step in iteration \(\nu\) for stage \(t\).

In the backward step, we solve (36)-(45) (and (46)-(48) in the case \(t = T\)) for each \(j \in \mathcal{K}\) and each scenario \(\xi^j_\ell = \xi^j_\ell(s)\), \(s \in S_j\), with \(x_{t-1} = \hat{x}^\nu_{t-1}\). Let \(\pi^\nu_{t,s} := \pi^\nu_1(\xi^j_\ell(s))\) denote the corresponding dual variable obtained from (43) (and from (47) in the case \(t = T\)). Then, we construct the Benders cut

\[
\ell^\nu_t(x_{t-1}) := \tilde{Q}^\nu_t(x_{t-1}) + (\pi^\nu_t)^\top (x_{t-1} - \hat{x}^\nu_{t-1})
\]

for the function \(\tilde{Q}^\nu_t(\cdot) := \sum_{s \in S_j} \tilde{Q}^\nu_t(\cdot, \xi^j_\ell(s)) q_j(s)\), using the average dual decision vector \(\pi^\nu_t = \sum_{s \in S_j} \pi^\nu_{t,s} q_j(s)\). As discussed earlier, \(\pi^\nu_{t,s}\) is a subgradient of \(\tilde{Q}^\nu_t(\cdot, \xi^j_\ell(s))\) at \(\hat{x}^\nu_{t-1}\) and thus it follows that \(\ell^\nu_t(x_{t-1}) \geq \tilde{Q}^\nu_t(x_{t-1})\) for all \(x_{t-1}\). Moreover, since the function \(Q^k_{t+1}(x_t)\) in (29) is replaced with the upper approximation \(\tilde{Q}^k_{t+1}(x_t)\) in (39), it follows that \(\tilde{Q}^\nu_t(x_{t-1}, \xi^j_\ell(s)) \geq Q^j_t(x_{t-1}, \xi^j_\ell(s))\) for all \(s \in S_j\) and thus \(\tilde{Q}^\nu_t(x_{t-1}) \geq Q^j_t(x_{t-1})\).

We then update the upper approximation of the value function (so it can be used in period \(t - 1\)) as

\[
\tilde{Q}^k_t(x_{t-1}) := \min_{i=1, \ldots, \nu} \ell_t^k(x_{t-1}), \quad \forall k \in \mathcal{K}.
\]

It follows that when we solve (36)-(45) in period \(t - 1\), by using (50) in constraint (39) we have

\[
-\theta^-_k + \theta^+_k - \eta \leq \left\{ \min_{i=1, \ldots, \nu} \ell^k_{t+1}(x_t) \right\}, \quad \forall k \in \mathcal{K},
\]

which is equivalent to the set of linear constraints

\[
-\theta^-_k + \theta^+_k - \eta \leq \ell^k_{t+1}(x_t) \leq 0, \quad \forall k \in \mathcal{K}, \forall i = 1, \ldots, \nu.
\]

Therefore, the outer approximation of the SAA problem can be represented as model (36)-(45), with constraint (39) replaced with inequalities (51). Note also that, because of Assumption 3, the constraints given by (44) are linear in \((x_t, u_t)\). Thus, since \(f_t(\cdot, \xi^j_\ell)\) and \(g_t(\cdot, \xi^j_\ell)\) are linear by Assumption 2, model (36)-(45) is just a linear program. Moreover, we can rely on the result of Philpott and Guan (2008) that shows convergence of the optimal solutions of the outer approximations to an optimal solution of the original problem in finitely many iterations, assuming that every scenario in the problem is eventually sampled in the forward pass. In our context, if the transition probability matrix obtained from HMM is irreducible, then it is possible to generate any scenario with nonzero probability and hence the proof of Philpott and Guan (2008) can be applied.

### 3.4 Deterministic lower bound

For standard SDDP applications, one can obtain a statistical lower bound by evaluating the current policy via Monte Carlo simulation and compute an estimator of the objective function, see details in Shapiro (2011). However, it is not practical to obtain a statistical objective function assessment within the distributionally robust framework (10)-(11). The issue here is that, in order
to evaluate the objective function in (10), we would need to know the optimal worst-case transition probability matrix in the corresponding inner problem, but this is not possible since we only have an approximation of value function \( Q^k_{t+1} \). Thus, if we simulate scenarios using any (suboptimal) transition probability matrix, the statistical evaluation of the objective function will not be a valid lower bound.

Our approach is to explore an extended inner approximation of \( Q^k_{t+1}(\cdot) \) to construct a valid lower bound to problem (26)-(34). The standard inner approximation method uses a convex combination of evaluated trials points instead of the Benders cuts (outer approximation). A similar approach was first proposed by Philpott et al. (2013) who ensured feasibility of the convex combination by pre-evaluating all vertices of the uncertainty support, e.g., a multidimensional hypercube. The contribution by Philpott et al. (2013) notwithstanding, the approach proposed in that work is not efficient in practice since the number of vertices grows exponentially with the uncertainty dimension. More recently, Baucke (2018) enhanced the method assuming that Lipschitz constants for the value function are known, but in many practical applications it may be difficult to estimate such constants. In our approach, we ensure feasibility by assuming there exists a simple (suboptimal) policy whose objective value is relatively easy to compute. The thrust of the method is to decompose, in an optimal manner, the function in two terms: one given by the convex combination of trial points, while the other is described by the simple (suboptimal) policy evaluation. The latter term is crucial as it allows for evaluation of the upper bound function outside the convex hull of the trial points.

Consider the expected value function \( Q^j_{t+1}(x_t) \) defined in (35), and suppose that in iteration \( \nu \) of the algorithm we have a concave lower (inner) approximation \( \hat{Q}^{j,\nu}_{t+1}(\cdot) \) for \( Q^j_{t+1}(\cdot) \) given by linear inequalities. Let \( \{\hat{x}_t^i\}_{i=1,...,\nu} \) be the solutions obtained from the previous forward steps of the algorithm. We detail next how to construct the function \( \hat{Q}^{j,\nu}_t \).

As in the case of the outer approximation discussed in Section 3.3, the algorithm goes backwards in time, from \( t = T \) until \( t = 0 \), and \( \hat{Q}^{j,\nu}_t \) is constructed from \( \hat{Q}^{j,\nu}_{T-1} \). For \( t = T \) we set \( \hat{Q}^{j,\nu}_T(x_{T-1}) := \sum_{s \in S_j} Q^j_T(x_{T-1},\xi^j_T(s)) q_j(s) \) in all iterations \( \nu \), where \( Q^j_T(\cdot) \) is defined in (9). Let \( R \) denote the set of points satisfying constraint (11), and define, for \( t < T \),

\[
\hat{Q}^{j,\nu}_t(x_{t-1},\xi_t) := \max_{x_t \in \hat{R}(x_{t-1},\xi_{t-1})} f_t(x_t,\xi_t) + \min_{p_j \in \mathcal{P}_j} \sum_{k \in \mathcal{K}} \hat{Q}^{k,\nu}_{t+1}(x_t) p_j(k)
\]

and \( \hat{Q}^{j,\nu}_t(x_{t-1}) := \sum_{s \in S_j} \hat{Q}^{j,\nu}_t(x_{t-1},\xi_t(s)) q_j(s) \). Note that, since \( \hat{Q}^{j,\nu}_{T-1}(\cdot) \) is piecewise-linear concave, it follows from Assumptions 2 and 3 that \( \hat{Q}^{j,\nu}_t(\cdot,\xi_t) \) and \( \hat{Q}^{j,\nu}_t(\cdot) \) are also piecewise-linear concave. Moreover, since \( \hat{Q}^{j,\nu}_{T-1}(\cdot) \) is a lower bound for \( Q^j_{T-1}(\cdot) \), by comparing (10)-(11) and (52) we see that \( \hat{Q}^{j,\nu}_t(x_{t-1},\xi_t) \leq Q^j_t(x_{t-1},\xi_t) \) and thus it follows that

\[
\hat{Q}^{j,\nu}_t(x_{t-1}) \leq Q^j_t(x_{t-1}).
\]
necessarily tight) lower bound \( Q_j^i(x) \leq Q_j^i(x) \) in a relatively inexpensive manner. For instance, in the dynamic asset allocation case, \( Q_j^i(x) \) is the current wealth minus the cost of moving all risky allocation to the risk free asset, see Section 5. Another example are the energy planning models where one can compute the overall cost of deficit of the system. Then, let \( \epsilon \) be an arbitrary number in \((0, 1)\) and define the function \( Q_{j,\nu}^i(x_{t-1}) \) as

\[
Q_{j,\nu}^i(x_{t-1}) := \max_{x', x'', \mu} \sum_{i=1}^{\nu} \mu_i (1 - \epsilon) \hat{Q}_{j,\nu}^i(\hat{x}_{t-1}) + \epsilon Q_j^i(x'')
\]

s.t. 
\[
(1 - \epsilon) x' + \epsilon x'' = x_{t-1}
\]
\[
\sum_{i=1}^{\nu} \mu_i \hat{x}_{t-1} - x' = 0
\]
\[
\sum_{i=1}^{\nu} \mu_i = 1
\]
\[
\mu, x', x'' \geq 0.
\]

Proposition 4. \( Q_{j,\nu}^i(\cdot) \) is a piecewise-linear concave lower bound for \( Q_j^i(\cdot) \).

Proof. For \( t = T \) we have \( Q_{j,\nu}^i(x_{T-1}) = Q_j^i(x_{T-1}) \) by definition and so the statement is true. Suppose \( t < T \). Given feasible \( x_{t-1} \), let \((\mu^*, x', x'')\) be an optimal solution to problem (54). Then, we have that

\[
Q_{j,\nu}^i(x_{t-1}) = \sum_{i=1}^{\nu} \mu_i^* (1 - \epsilon) \hat{Q}_{j,\nu}^i(\hat{x}_{t-1}) + \epsilon Q_j^i(x'')
\]

\[
\leq (1 - \epsilon) \hat{Q}_{j,\nu}^i(x') + \epsilon Q_j^i(x'')
\]

\[
\leq (1 - \epsilon) Q_j^i(x') + \epsilon Q_j^i(x'')
\]

\[
\leq Q_j^i(x_{t-1}).
\]

The inequality in (55) follows from concavity of \( \hat{Q}_{j,\nu}^i(\cdot) \), whereas the inequality in (56) follows from (53) and the choice of \( \Omega_j^i(\cdot) \). Finally, the inequality in (57) follows from concavity of \( Q_j^i(\cdot) \). As discussed earlier the function \( \hat{Q}_{j,\nu}^i(\cdot) \) is piecewise-linear concave, and \( \Omega_j^i(\cdot) \) is assumed to be piecewise-linear concave as well. It follows that the function \( Q_{j,\nu}^i(x_{t-1}) \) defined in (54) is also piecewise-linear concave.

We close this section by noting that \( \hat{Q}_{j,\nu}^i(x_{t-1}, \xi_i) \) in (52) can be computed as the solution of a linear program, similarly to (36)-(45) but with \( \Omega_{j,\nu}^i(x_{t-1}) \) in (39) replaced with \( Q_{j,\nu}^i(x_{t-1}) \). For more details about the deterministic lower and upper bounds algorithms see Appendix ??.
4 Assessing out-of-sample performance in a rolling horizon scheme

Most SDDP applications use a rolling horizon scheme to mitigate the end-effect of the terminal time stage. One way to interpret this usage is that the actual problem has an infinite horizon and it is approximated by a finite horizon model with many time stages such that the “end of the world” has a small influence on the first stage decision. This is the case for long term energy planning, portfolio selection and asset-liability management problems, to name a few. In this section, we establish a generic out-of-sample evaluation framework and develop an acceleration scheme for the particular time-homogeneous case where the parameters of the problem (i.e. the functions \( f_t(x_t, \xi_t) = f(x_t, \xi_t) \) and \( g_t(x_t, \xi_t) = g(x_t, \xi_t) \), and the coefficients in the set \( \mathcal{X}_t \)) do not depend on the time period.

The framework for the rolling horizon scheme is defined as follows. Consider a testing horizon of length \( H \) and let \( t_1, \ldots, t_H \) denote the times at which the model is solved and the corresponding first-stage optimal solution is implemented. A suitable way to emulate the actual decision making process is to concatenate five steps for a given time \( t \in \{t_1, \ldots, t_H\} \): (i) the HMM parameters are estimated via EM algorithm using as input the sequence of observed uncertainty realization \( \{\xi_1, \ldots, \xi_t\} \); (ii) an SAA version of the problem is generated as in Section 3.2; (iii) a Markov state classification is performed. An easy classification method is to use as initial state the one with highest posterior probability of occurrence given the historical path of the process \( \{\xi_t\} \), see (2). In step (iv), the SDDP algorithm for the problem with \( T \) stages is run until convergence (according to Algorithm ??) for problem (26)-(34) assuming a given previous implemented decision \( x_{t-1} \) and the current uncertainty realization \( \xi_t \); (v) the first-stage decision \( x_t \) is implemented, the time \( t \) is updated and we go to step (i). Note that this procedure is computationally intensive since a SDDP is run until convergence for each time step of the simulation. Finally, for the implementation of optimal policy \( x_t \) in step (v) it is necessary to use a method—step (iii)— to infer the initial state of the HMM (recall that such states are not observable).

For the time-homogeneous case, it is appropriate to use only the first stage problem to implement every decision in the rolling horizon scheme. This is motivated by the fact that the problem structure does not depend on the time period. In this context, we propose a relatively fast evaluation framework that is divided it into two parts: (i) Estimation and sampling; and (ii) out-of-sample evaluation. For that, we split our data in two: training and testing datasets. Figure 3 illustrates the process.

In the estimation and sampling part, the training dataset is used as input for the EM (expectation-maximization) algorithm to estimate the HMM parameters, i.e., nominal transition probabilities and conditional probability distributions of the uncertain vector. Those conditional distributions are sampled using Latin Hypercube Sampling (LHS), which performs better than Monte Carlo sampling method as shown in Homem-de Mello et al. (2011), to construct the SAA scenario tree.

For the out-of-sample evaluation, a rolling horizon scheme is used over the testing dataset to
simulate historical (out-of-sample) performance. In essence, we follow three steps for a given time $t$: (i) the Markov state classification is performed using (2); (ii) a SDDP is run until convergence (again, according to Algorithm ??) for problem (26)-(34) assuming a given previous stage decision $x_{t-1}$ and the current uncertainty realization $\xi_t$; (iii) the first stage decision $x_t$ is implemented, the time $t$ is updated and we go to step (i). Note that the steps are very similar to the initial decision process laid out earlier; however, in the out-of-sample evaluation we do not re-estimate the parameters of the HMM, nor do we generate a new SAA version of the model. Thus, the convergence of SDDP in step (ii) should be much faster as it can use the value function approximations constructed in the previous steps as described below.

Given that HMM parameters are fixed, the value function for each state and time period remains the same and can be reused over the rolling horizon scheme. However, the value function might not be well approximated given the updated value of the initial condition $x_{t-1}$. Therefore, we use the current approximation of the value function of the first stage to perform a convergence test using deterministic upper and lower bounds (see Appendix ??) to evaluate the gap given the updated initial condition. If the gap is not sufficiently small, we restart the SDDP algorithm to improve the value function until it achieves a satisfactory gap. Once the algorithm converges, a current optimal solution is obtained and implemented. The whole procedure is now repeated one-step ahead, given the previous optimal decision and the current observed uncertainty realization. The whole evaluation process iterates until it reaches the last period to be simulated.

5 Case study: a risk-constrained dynamic asset allocation model

In this section we illustrate an application of the framework laid out in the previous sections to an asset allocation problem. The model learns the asset returns from the data and solves a dynamic optimization problem where the goal is to maximize the expected final wealth, taking into account the transaction costs in each period. Other papers use learning approaches for this problem; for example, Lim et al. (2012) apply a regret-optimization approach find the best (single-period) portfolio choice using historical data as input. We build upon the work of Valladão et al. (2018), which allows us to use their results as a benchmark since that paper does not deal with out-of-sample performance. In subsection 2.1 we recap the stochastic model for asset returns while in subsection 5.2 we present an equivalent formulation for the risk constrained dynamic asset allocation model proposed by Valladão et al. (2018).

5.1 The HMM learning methodology for asset returns

The uncertain returns $r_t$ are represented by a Hidden Markov Model (HMM). In the context of financial market, HMM methodology is frequently used to model asset returns (Hassan and Nath 2005, Elliott and Siu 2014, Elliott and Van der Hoek 1997, Mamon and Elliott 2007, 2014).
Such paradigm postulates that the probability distribution of asset returns depends only on the current state of the market — for instance, “bull market”, “bear market”, “regular market” — that evolve according to a discrete-time finite-state Markov Chain. Such states, however, cannot be observed, hence the need for a Hidden Markov Model. Conditionally to each state, the log-returns are independent and identically distributed, with distribution given by multivariate Gaussian estimated from data. This modeling choice is suitable for financial time series since it empirically reproduces most of the stylized facts for asset return series (Granger and Ding 1994, Rydén et al. 1998).

As before, we denote by $K_t$ the (random) Markov state at time $t$, by $\mathcal{K}$ the set of states of the Markov chain and by $\hat{P}$ the corresponding estimated transition matrix with dimension $|\mathcal{K}| \times |\mathcal{K}|$, with $\hat{p}_j(k)$ denoting the probability to transition from state $j$ to state $k$. The standard EM (expectation-maximization) algorithm (Moon 1996) is used to estimate from data means, variances and covariances as well as the nominal transition probability matrix.
5.2 A CV@R-constrained dynamic asset allocation model

The model proposed in Valladão et al. (2018) is a multistage stochastic program which maximizes, in each stage, the future value function that represents the conditional expectation of the terminal wealth, subject to a CV@R constraint. Using the notation defined in (5)-(9), that model can be written as follows. Given an initial wealth $W_0$ and the stochastic return process $r_t$, we denote $\xi_t = (1 + r_t)$ and solve, for each possible initial state $j$, the problem

$$Q_j^0 := \max_{x_0 \in \mathbb{R}^{N+1}_+} \sum_{k \in K} \mathbb{E} \left[ Q_k^t(x_0, \xi_1) \middle| K_{t+1} = k \right] \hat{p}_j(k)$$

subject to

$$\begin{align*}
&\phi_{\hat{p}_j} \left[ \xi_1^\top x_0 \right] \geq (1 - \gamma)W_0 \\
&(1 + \hat{c})^\top x_0 = W_0
\end{align*}$$

where $\hat{c}$ is the vector $\begin{bmatrix} 0 \\ c \end{bmatrix}$, and $Q_j^i (t = 1, \ldots, T - 1)$ is defined recursively as

$$Q_j^i(x_{t-1}, \xi_t) = \max_{x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)} \sum_{k \in K} \mathbb{E} \left[ Q_k^{i+1}(x_t, \xi_{t+1}) \middle| K_{t+1} = k \right] \hat{p}_j(k)$$

subject to

$$\begin{align*}
&\phi_{\hat{p}_j} \left[ \xi_{t+1}^\top x_t \right] \geq (1 - \gamma)(\xi_t^\top x_{t-1}).
\end{align*}$$

In the above equations, the set $\mathcal{X}_t(x_{t-1}, \xi_t)$ is defined as

$$\left\{ x_t \in \mathbb{R}^{N+1}_+ \middle| \exists b_t, d_t \in \mathbb{R}_+^N: \begin{array}{l}
x_{0,t} + (1 + c)^\top b_t - (1 - c)^\top d_t = x_{0,t-1} \\
x_{i,t} - b_{i,t} + d_{i,t} = \xi_{i,t} x_{i,t-1}, \quad \forall i \in A.
\end{array} \right\}$$

and the end-of-horizon function is

$$Q_T^j(x_{T-1}, \xi_T) = \xi_T^\top x_{T-1}, \quad \forall j \in K.$$ 

From (63), we see that the allocations $x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)$ satisfy the equation

$$1^\top x_t = \xi_t^\top x_{t-1} - c^\top (b_t + d_t),$$

that is, the amount of money available at time $t$ is the return of the investment made at time $t - 1$, minus the transaction costs of assets that were bought ($b_t$) and sold ($d_t$).

A few words about the above model are in order. First, notice that the objective functions in (58) and (61) maximize the expected future value of the allocation in each period, where the expectation is taken with respect to both the returns and the Markov states. Constraint (60) reflects the fact that the transaction costs are incurred before the returns are realized; thus, assuming that the initial wealth $W_0$ is in cash, the initial allocation $1^\top x_0$ plus the corresponding purchase costs
must be equal to that amount. This constraint is generalized to an arbitrary time period $t$ by means of the set $X_t(x_{t-1}, \xi_t)$ defined in (63), which accounts for the transaction costs resulting from both purchases and sales of assets (note that $b_{i,t}$ and $d_{i,t}$ are never simultaneously positive as nothing is gained from buying and selling the same asset in a given time period).

Constraint (62) can be written as $\text{CV@R}_{1-\alpha} [W_t - W_{t+1}] \leq \gamma W_t$ where $W_t = \xi_{t}^\top x_{t-1}$ is the wealth—prior to discounting transaction costs, cf. (65)—in period $t$. That is, the constraint limits the loss between periods $t$ and $t + 1$ to a percentage of the wealth at time $t$ (note that constraint (59) applies the same idea at $t = 0$). The parameter $\gamma$ can then be interpreted as the level of risk-aversion of the decision maker: at one extreme ($\gamma = 0$) we have $\text{CV@R}_{1-\alpha} [W_t - W_{t+1}] \leq 0$ which in particular implies that $P(W_{t+1} < w_t | W_t = w_t) \leq \alpha$, i.e., the probability of a loss between periods $t$ and $t + 1$ must be very low; at the other extreme ($\gamma = 1$) we do not impose any risk constraints and so when there are no transaction costs the optimal portfolio will invest only in the asset(s) with highest expected return at each time $t$ (“all eggs in the same basket”).

5.3 Numerical results

To analyze how our approach behaves in practice, we test the model with real data. The data sets used in the experiments come from Kenneth R. French data set. The stocks from NYSE, AMEX and NASDAQ are represented by capitalization-weighted indexes for each industry sector. We use monthly data of 5 industrial portfolios (“Cnsmr”, “Manuf”, “HiTec”, “Hlth” and “Other”). For simplicity, we use excess returns, i.e., the incremental return over the risk free asset. This way, the risk free asset presents $r_{0,t} = 0$, $\forall t = 1, \ldots, T$.

The framework was implemented in Julia language 0.6, using JuMP (Dunning et al. 2017) and CPLEX 12.7.1.0 to solve linear programming problems. All experiments were conducted on Intel Xeon E5-2680 2.7 GHz with 128GB RAM machine, while reported computational times are associated with single core usage. The hmmlearn 0.2.0 library was used to construct the return distributions assuming that, conditional to each Markov state, log (excess) returns follows a multivariate Gaussian distributions.

5.3.1 Results for the predictive model.

The training dataset comprises 444 months (prior to January 2007) while the testing dataset for historical simulation uses 96 months (from January 2007 to June 2018) to validate the proposed framework. Following Valladão et al. (2018) we select three Markov states and, conditional to each state, 750 return realizations obtained using LHS. All simulations start with $1 in the risk-free asset, therefore if the strategy ends the simulation with $2 it implies an accumulated excess return of 100%. Figure 4 illustrates the posterior probability of each Markov state as in [1], and the

4http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data library.html
5https://github.com/hmmlearn/hmmlearn
solid line indicates the simulated wealth of the equal-weighted portfolio (as a proxy for the general behavior of the market) with cumulative return on the right axis. The dashed line corresponds to 0%.

![Figure 4: Markov states and equal-weight portfolio wealth.](image)

| $K_t$ | 1   | 2   | 3   |
|-------|-----|-----|-----|
| 1     | 69.28 | 28.58 | 2.14 |
| 2     | 58.23 | 40.66 | 1.11 |
| 3     | 0.77  | 8.79  | 90.44 |

Table 1: Markov transition matrix in percentage.

| State | Cnsmr | Manuf | HiTec | Hlth | Other |
|-------|-------|-------|-------|------|-------|
| 1     | 1.17  | 1.13  | 0.91  | 1.29 | 1.21  |
| 2     | 1.86  | 1.81  | 2.08  | 1.29 | 1.94  |
| 3     | -1.24 | -0.93 | -1.81 | -0.81| -1.60 |

Table 2: Assets expected returns (%) conditional to each Markov state.

A closer look at the Markov transition matrix Table 1 the individual asset returns in Table 2 and the corresponding standard deviations in Table 3 in conjunction with Figure 4 allows us to infer how HMM is classifying the historical data and how to interpret the states. State 1 has low positive returns and low probability to transition to state 3. State 2 has high probability to transition to state 1, it also has higher returns than state 1 and it has more volatility. State 3 has
negative returns, is almost absorbent with 90% chance to transition to itself and has almost no probability of transition to state 1. Therefore, states 2 and 3 can be seen as bull and bear states, respectively. It is more difficult to infer the role of state 1, it seems to be a more volatile regular state since it is the most probable state during the whole simulation (Figure 4).

### 5.3.2 Results for the prescriptive model.

As discussed in Section 4, we implement the algorithm in a rolling-horizon fashion. The horizon (number of periods) in each problem is \( T = 16 \) months with monthly decisions. For the out-of-sample evaluation described in 4 we start with an estimated HMM, a SAA of the original problem and a converged SDDP—comprising a first stage problem and set of \( T \) future value functions. We shall denote the time periods of the testing dataset \( t \in \{ t_1, \ldots, t_H \} \) and, define \( R_t = \frac{W_t - W_{t-1}}{W_{t-1}} \) the portfolio percentage profit given by the proposed strategy at time \( t \) (recall that \( W_t := \xi_t^x_t \) is the corresponding wealth, prior to discounting transaction costs, cf. (65)). It is important to reiterate that the implemented decisions \( x_t^* \) are obtained as the first stage solution of a \( T \)-stage problem given the current (inferred) Markov state—the most probable state given all information available up to \( t \). In order to compare out-of-sample performance of the different models, we use the “ex-post” average return of the portfolio strategy, \( \hat{R}_{EP} := \frac{1}{H} \sum_{t=1}^{H} R_t \), and the “ex-post” CV@R of the returns, defined as \( -\phi_{EP} \), where \( \hat{\phi}_{EP} = \max_{z \in \mathbb{R}} \left\{ z - \frac{1}{H} \sum_{t \in \{ t_1, \ldots, t_H \}} (z - R_t) \right\} \) following the expression in (4). Note that a comparison of the ex-post CV@R with the parameter \( \gamma \) can be interpreted as an out-of-sample evaluation of constraint (62), since as remarked earlier that constraint can be written as \( CV@R_{1-\alpha} [W_t - W_{t+1}] \leq \gamma W_t \), i.e.,

\[
-\phi_{p_j} [R_{t+1}] \leq \gamma.
\]

Despite the differences with the ex-ante counterparts, ex-post metrics are widely used, specially within the context of financial markets.

For a better assessment of out-of-sample performance, several experiments were made with different combinations of \( \gamma \) and \( \Delta \). It is important to stress the difference between these parameters. The former quantifies the decision maker’s level of risk aversion (cf. (62)), whereas the latter establishes the confidence in the estimated distribution (cf. (12)). In this context, \( \gamma \) restricts the possible decisions, however even if the risk restriction is met, the confidence (or lack thereof) in the estimated probabilities (\( \hat{p} \)) will still impact the optimal portfolio decision.

### Table 3: Standard deviations (%) for each asset and state in percentage.

| State | Cnsmr | Manuf | HiTec | Hlth | Other |
|-------|-------|-------|-------|------|-------|
| 1     | 0.14  | 0.10  | 0.12  | 0.13 | 0.14  |
| 2     | 0.20  | 0.15  | 0.18  | 0.16 | 0.19  |
| 3     | 0.51  | 0.39  | 0.46  | 0.35 | 0.50  |
We illustrate the compound effect of the ambiguity aversion ($\Delta$) and the risk aversion ($\gamma$) coefficients over the optimal allocation. The effect of these coefficients can be seen in Figure 5, where we present the optimal portfolio on a particular date as a function of $\Delta$ for a few values of $\gamma$. We see for example that, for a low value of $\gamma$ — i.e., a more risk-averse decision maker — the optimal portfolio is less sensitive to variations in $\Delta$, as the optimal portfolio puts a high percentage on the risk-free asset regardless of the value of $\Delta$. For a slightly less risk-averse decision maker ($\gamma = 0.05$) the optimal portfolio is diversified, with the components changing according to $\Delta$; note that for values of $\Delta$ larger than 0.35 the ambiguity set includes all distributions and so the min-max problem will always assume the worst possible state, so it is not surprising that the optimal portfolio from that value of $\Delta$ on puts everything on the risk-free asset. With $\gamma = 0.1$ we have essentially a risk-neutral decision maker and so for most values of $\Delta$ the optimal portfolio consist only of the asset with largest expected return, though that asset changes based on the level of confidence on the parameters of the HMM given by $\Delta$.

To further analyze this distinction between the parameters and how the Markov state impacts the final portfolio we show the portfolio allocation during the simulation for specific $\gamma$ and $\Delta$ values. This comparison is depicted in Figure 6 where we present the allocation policy for two $\Delta$ values, 0.01 and 0.05.
0.0 and 0.2, with $\gamma = 0.07$ for the whole simulation period. The choice for $\gamma = 0.07$ is due to the fact that it is the ex-post CV@R corresponding to the equal-weight portfolio. The left axis shows the allocation in percentage of each asset and the right axis shows the wealth for our DRO model and for the equal-weight portfolio for comparison purposes. Some observations can be made: first, notice that during the period between 2008-2009 (which corresponds to the subprime crisis) the optimal portfolios for both values of $\Delta$ learn from the HMM that the market is in a bear state (cf. Figure 4) and thus allocate almost everything into the risk-free asset. Second, while the equal-weight portfolio clearly dominates the portfolio for the case $\Delta = 0$, it is outperformed by the robustified portfolio with $\Delta = 0.2$ as the latter strategy yields better protection during the “bear” times and provides good diversification and good returns during the remaining time periods.

Figure 7 depicts the ex-post CV@R of policies obtained with various values of $\Delta$, compared with $\gamma$. We observe that for lower values of $\Delta$ such as 0.0, 0.025 and 0.05, we have poor out-of-sample risk control, as the ex-post CV@R is higher than the tolerated value of $\gamma$. Here, we can see that higher values of the parameter $\Delta$ do indeed enforce better out-of-sample performance. Also, for $\gamma$ values higher than 0.1 the ex-post CV@R did not increase, thus corroborating our earlier observation that the risk constraint is not active for $\gamma \geq 0.1$. The figure also shows that the ex-post CV@R of the equal-weight portfolio is equal to 0.07.

The risk-return curves for different values of $\gamma$ and $\Delta$ are shown in Figure 8. Naturally, for portfolios with the same risk the ones with more return are preferred, whereas for portfolios with the same return the ones with less risk are preferred. In the figure, each line corresponds to one value of $\Delta$, whereas each dot corresponds to one value of $\gamma$. We see that the efficient frontier consists of portfolios corresponding to $\Delta$ around 0.25-0.3, regardless of $\gamma$. This, it appears—based on these experiments—that the right choice for $\Delta$ ensures good performance regardless of the decision maker’s risk tolerance. The use of $\gamma$, however, is still important for sensitivity purposes, as we can see that lower values of $\Delta$ combined with high values of $\gamma$ can yield portfolios with inferior performance. Notice also that, as remarked earlier, values of $\Delta$ above 0.35 lead to excessive robustness as there is no trust in the HMM parameters and thus the optimal portfolio consists only of the risk-free asset. That is, it is important to collect sufficient data so one has some confidence in the HMM parameters, but it is better not to trust such parameters blindly.

As shown by other authors (DeMiguel et al. 2007, Esfahani and Kuhn 2017, Pflug et al. 2012), the equal-weight portfolio is a good benchmark strategy to compare with as it has competitive out-of-sample performance, especially when the model faces extreme uncertainty or when the transactions costs are high. In Figure 8, we see that the equal-weight portfolio is dominated by most strategies. It confirms the superior out-of-sample performance of our proposed model, except in the cases where the ex-post CV@R is high, although that can only occur when the pre-specified risk tolerance $\gamma$ is high so the decision maker is nearly risk-neutral.

Finally, in order to assess the effect of $\Delta$ on the quality of the portfolio, we need to use a metric that summarizes risk and return. The ICV@R (Fernandes et al. 2016) was inspired by the Sharpe
Figure 6: Allocation during the simulation for $\Delta$ values 0.00 and 0.20 with $\gamma = 0.07$.

Figure 6: Allocation during the simulation for $\Delta$ values 0.00 and 0.20 with $\gamma = 0.07$.

ratio in that it measures return by unit of risk. It is computed as the ratio between the average return $\hat{R}_{EP}$ and the deviation between the average return and the average tail $\hat{\phi}_{EP}$ (Kalinichenko et al. 2012), that is, $ICV@R := \frac{\hat{R}_{EP}}{\hat{R}_{EP} - \hat{\phi}_{EP}}$. Figure 9 illustrates the deviation concept.

Figure 10 depicts the values of the ICV@R index for various values of $\Delta$ and $\gamma$. We see that the ICV@R function is almost always monotonically non-decreasing for $\Delta \leq 0.325$ regardless of the
value of \( \gamma \), which suggests again that in order to have better out-of-sample performance one should use higher (but not too high) \( \Delta \) values. Moreover, we see again that values of \( \Delta \) around 0.25-0.325 yield the highest values of ICV@R (and thus the best portfolios according to this criterion).
regardless of the value of $\gamma$. The figure also shows that the equal-weight portfolio is dominated by our distributionally robust approach for all cases with $\Delta \in [0.125, 0.325]$.

6 Conclusions

The evolution of computer power, new theoretical results and the development of specialized software tools have made stochastic dynamic optimization models widely applicable in recent years. In our opinion, however, the increase in utilization of such models has not been accompanied by a similar development in the treatment of data for this type of problems—indeed, in most applications reported in the literature some form for the underlying stochastic process $\{\xi_t\}$ is assumed
(after some study of available data), the problem is solved and the optimal solution given by the model is implemented. Our goal is in this paper is to bring this practice closer to a new reality of data-driven problems, where information can be inferred automatically from the data via some machine learning technique, and an independent validation procedure is applied in order to evaluate the decisions yielded by the model.

To accomplish such goal, we have presented a framework for data-driven stochastic dynamic decision models with certain structure which is applicable in many different contexts. Our approach combines a Hidden Markov Model (HMM) as the predictive engine with a dynamic Distributionally Robust Optimization (DRO) model as the prescriptive methodology. Notwithstanding the HMM flexibility to approximately capture the dynamics of a variety of stochastic processes, it is subject to estimation errors as well as model misspecification. Therefore, a distributionally robust dynamic optimization model is a suitable choice to embody the uncertainty dynamics represented by the HMM and at the same time to robustify decisions against the uncertainty over the HMM parameters.

We have provided a tractable reformulation of the problem and shown that we can adapt the well-known Stochastic Dual Dynamic Programming (SDDP) algorithm to solve the proposed model. Along the way we have developed a deterministic lower bound (for a maximization problem) which, although related to recent literature results, it has a practical appeal of considering user-defined simple policies evaluations to improve computational tractability and solution efficiency.

For the validation procedure, we have presented an evaluation framework to assess out-of-sample performance of the optimal policies yielded by the model in a rolling horizon scheme. We have also presented an acceleration scheme in case of computationally intensive problems which is applicable when the problem structure does not depend on the time period. We have illustrated the power and flexibility of the propose data-driven prescriptive analytics framework with a complete case study on dynamic asset allocation. The numerical results show superior out-of-sample performance against selected benchmarks. The case study reiterates the practical importance and applicability of the proposed framework since it emulates the actual decision process of a dynamic asset allocation problem, extracting valuable information from data to obtain robust decisions with an empirical certificate of suitable out-of-sample performance.

While the case study focuses on one type of problem (dynamic asset allocation), we believe that the framework presented in the paper can be useful in other contexts as well. For example, in long-term energy planning — a type of problems for which SDDP has been extensively used — the input stochastic process (e.g., water inflows, solar radiation) could be inferred directly from the data, using machine learning techniques as described in this paper. We believe that the presented work raises important questions for future development on the integration of machine learning methods and dynamic optimization under uncertainty, and hope it will stimulate further research in this area.
References

Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent Measures of Risk. *Mathematical Finance* 9(3):203–228.

Baucke R (2018) *Risk aversion in multistage stochastic optimisation problems*. Ph.D. thesis, Operations Research, URL https://researchspace.auckland.ac.nz/handle/2292/45173.

Bayraksan G, Love DK (2015) Data-driven stochastic programming using phi-divergences. *Tutorials in Operations Research*, 1–19 (INFORMS).

Ben-Tal A, Den Hertog D, De Waegenaere A, Melenberg B, Rennen G (2013) Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* 59(2):341–357.

Bishop CM (2006) *Pattern recognition and machine learning* (Springer).

DeMiguel V, Garlappi L, Uppal R (2007) Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies* 22(5):1915–1953.

Dunning I, Huchette J, Lubin M (2017) Jump: A modeling language for mathematical optimization. *SIAM Review* 59(2):295–320, URL http://dx.doi.org/10.1137/15M1020575.

Elliott RJ, Sin TK (2014) Strategic asset allocation under a fractional hidden markov model. *Methodology and Computing in Applied Probability* 16(3):609–626.

Elliott RJ, Van der Hoek J (1997) An application of hidden markov models to asset allocation problems. *Finance and Stochastics* 1(3):229–238.

Esfahani PM, Kuhn D (2017) Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 1–52.

Fernandes B, Street A, Valladão D, Fernandes C (2016) An adaptive robust portfolio optimization model with loss constraints based on data-driven polyhedral uncertainty sets. *European Journal of Operational Research* 255(3):961–970.

Fhoula B, Hajji A, Rekik M (2013) Stochastic dual dynamic programming for transportation planning under demand uncertainty. *2013 International Conference on Advanced Logistics and Transport*, 550–555 (IEEE).

Goh J, Sim M (2010) Distributionally robust optimization and its tractable approximations. *Oper Res* 58:902–917.

Granger CW, Ding Z (1994) Stylized facts on the temporal and distributional properties of daily data from speculative markets. *UCSD Department of Economics Discussion Paper* 94–19.

Hassan MR, Nath B (2005) Stock market forecasting using hidden markov model: a new approach. *Intelligent Systems Design and Applications, 2005. ISDA’05. Proceedings. 5th International Conference on*, 192–196 (IEEE).

Homem-de Mello T, De Matos VL, Finardi EC (2011) Sampling strategies and stopping criteria for stochastic dual dynamic programming: a case study in long-term hydrothermal scheduling. *Energy Systems* 2(1):1–31.

Homem-de Mello T, Pagnoncelli BK (2016) Risk aversion in multistage stochastic programming: A modeling and algorithmic perspective. *European Journal of Operational Research* 249(1):188–199.
Infanger G, Morton DP (1996) Cut sharing for multistage stochastic linear programs with interstage dependency. *Math Program* 75:241–256.

Kalinchenko K, Uryasev S, Rockafellar RT (2012) Calibrating risk preferences with the generalized capital asset pricing model based on mixed conditional value-at-risk deviation. *The Journal of Risk* 15(1):45.

Lim AE, Shanthikumar JG, Vahn GY (2012) Robust portfolio choice with learning in the framework of regret: Single-period case. *Management Science* 58(9):1732–1746.

Löhndorf N, Shapiro A (2019) Modeling time-dependent randomness in stochastic dual dynamic programming. *European Journal of Operational Research* 273(2):650 – 661.

Mamon RS, Elliott RJ, eds. (2007) *Hidden Markov Models in Finance*, volume 104 of *International Series in Operations Research & Management Science* (Boston, MA: Springer US).

Mamon RS, Elliott RJ, eds. (2014) *Hidden Markov Models in Finance*, volume 209 of *International Series in Operations Research & Management Science* (Boston, MA: Springer US).

Mo B, Gjelsvik A, Grundt A (2001) Integrated risk management of hydro power scheduling and contract management. *Power Systems, IEEE Transactions on* 16(2):216–221.

Moon TK (1996) The expectation-maximization algorithm. *IEEE Signal Processing Magazine* 13(6):47–60, ISSN 1053-5888, URL [http://dx.doi.org/10.1109/79.543975](http://dx.doi.org/10.1109/79.543975).

Pereira MVF, Pinto LMVG (1991) Multi-stage stochastic optimization applied to energy planning. *Math. Program.* 52(2):359–375, ISSN 0025-5610, URL [http://dx.doi.org/10.1007/BF01582895](http://dx.doi.org/10.1007/BF01582895).

Pflug GC, Pichler A, Wozabal D (2012) The 1/n investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance* 36(2):410–417.

Philpott A, de Matos V, Finardi E (2013) On solving multistage stochastic programs with coherent risk measures. *Operations Research* 61(4):957–970.

Philpott AB, de Matos VL (2012) Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. *European Journal of Operational Research* 218(2):470–483.

Philpott AB, Guan Z (2008) On the convergence of stochastic dual dynamic programming and related methods. *Operations Research Letters* 36(4):450–455.

Rabiner LR (1989) A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2):257–286, ISSN 0018-9219, URL [http://dx.doi.org/10.1109/5.18626](http://dx.doi.org/10.1109/5.18626).

Rahimian H, Bayraksan G, Homem-de Mello T (2018) Identifying effective scenarios in distributionally robust stochastic programs with total variation distance. *Mathematical Programming* 1-38.

Reus L, Pagnoncelli B, Armstrong M (2019) Better management of production incidents in mining using multistage stochastic optimization. *Resources Policy* 63:101404, ISSN 0301-4207, URL [http://dx.doi.org/https://doi.org/10.1016/j.resourpol.2019.101404](http://dx.doi.org/https://doi.org/10.1016/j.resourpol.2019.101404).

Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. *Journal of risk* 2:21–42.

Roorda B, Schumacher JM, Engwerda J (2005) Coherent acceptability measures in multiperiod models. *Mathematical Finance* 15(4):589–612.

Rudloff B, Street A, Valladão DM (2014) Time consistency and risk averse dynamic decision models: Definition, interpretation and practical consequences. *European Journal of Operational Research* 234(3):743–750.
Ryden T, Teräsvirta T, Asbrink S (1998) Stylized facts of daily return series and the hidden markov model. *Journal of applied econometrics* 13(3):217–244.

Scarf H (1958) A min-max solution of an inventory problem. Arrow K, Karlin S, Scarf H, eds., *Studies in the Mathematical Theory of Inventory and Production*, 201–209 (Stanford, CA,: Stanford University Press).

Shapiro A (2011) Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research* 209(1):63–72.

Shapiro A, Ahmed S (2004) On a class of minimax stochastic programs. *Siam J Optimiz* 14(4):1237–1249.

Shapiro A, Dentcheva D, Ruszczyński A (2014) *Lectures on stochastic programming : modeling and theory* (SIAM), 2nd edition.

Shapiro A, Kleywegt A (2002) Minimax analysis of stochastic problems. *Optim Meth and Software* 17:523–542.

Shapiro A, Tekaya W, da Costa JP, Soares MP (2013) Risk neutral and risk averse stochastic dual dynamic programming method. *European Journal of Operational Research* 224(2):375–391.

Sion M, et al. (1958) On general minimax theorems. *Pacific Journal of mathematics* 8(1):171–176.

Valladão D, Silva T, Poggi M (2018) Time-consistent risk-constrained dynamic portfolio optimization with transactional costs and time-dependent returns. *Annals of Operations Research* 1–27.

Van Parys BP, Esfahani PM, Kuhn D (2017) From data to decisions: Distributionally robust optimization is optimal. *arXiv preprint arXiv:1704.04118*.

Záčková J (1966) On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky* 91(4):423–430.