Optimal Use and Replenishment of Two Substitutable Raw Materials in a Stochastic Capacitated Make-to-Order Production System

Qi (George) Chen
London Business School, Regent's Park, London, U.K.

Izak Duenyas, Stefanus Jasin
Stephen M. Ross School of Business, University of Michigan, Ann Arbor, Michigan, U.S.

Problem definition: We study a multi-period, nonstationary, make-to-order, joint production and inventory model where two kinds of input raw materials with availability uncertainties and different output conversion rates can be blended and then processed in a production line with stochastic capacity to produce the output product. **Academic/practical relevance:** The problem is motivated by the practice in coal-fired power plants, an important part of the energy sector, where two types of coal with different energy content per unit mass are blended for electrical power generation. Our model is the first to capture the key operational features in this context. **Methodology:** We model the problem as a Markov decision process and develop a novel approximate optimization approach to analyze and characterize the structure of the optimal policy. **Results:** We show that a Use-down-to/Balancing Production Policy and modified Order-up-to Ordering Policy is optimal. We also propose a heuristic policy based on piece-wise linear value function approximation. While computing the value function approximation via brute-force is time-consuming due to the curse of dimensionality, we leverage the structure of the optimal policy to develop an algorithm which greatly improves the computational time of the value function approximation. Our numerical studies on both a synthetic data set and real world data show that the proposed heuristic provides significant profit improvement over three simpler straw policies some of which are used in practice. **Managerial implications:** Our paper suggests the significant profit improvement opportunity of using our proposed policy and demonstrates how one can develop computationally more efficient heuristic policies by leveraging the structure of the optimal policy.

Key words: production/inventory, capacity, nonstationarity, stochastic dynamic program, optimal policy

1. Introduction

Coal is a major source of energy for power generation worldwide. In the United States for example, around 27.4% of total electricity generation comes from coal-fired power plants in 2018 (see Table 7.2a in U.S. Energy Information Administration (2019)). For these plants, managing the inventory and usage of coal is critical to their business success because the cost of purchasing and shipping coal is massive: The total expenditure on purchasing and shipping coal across all coal-fired power

[1] The numbers are calculated from U.S. Department of Energy (2019). Note that among all 294 plants, the data of average total cost per million BTUs is only available for 198 plants that are operated by regulated electric utilities. Thus we calculated the total and average expenditure for those 198 plants.

plants in the U.S. in 2018 is more than 17.7 billion dollars with an average of 89.4 million dollars per plant (U.S. Department of Energy 2019).[1] While there has been a recent global trend of substituting unabated coal with natural gas and renewable sources as inputs in the production of electricity due to coal's high carbon emission and impact on global warming, such transition may take decades to complete due to various economic and political factors (Evans and Pearce 2020). This suggests that at least in the next couple decades, coal would still remain an important power source and thus the high expenditures on it need to be properly managed.

However, based on our experience with a coal-fired power plant we recently worked with, managing the replenishment and the usage of coal for power generation effectively remains challenging due to the nature and the practice of coal-fired power generation process. In such a process, the energy content in coal is released and *proportionally* transformed into electrical power. What is interesting is that there are two types of coal in use for power generation and they have different levels of energy density, i.e., the amount of energy per unit mass measured in British Thermal Unit (BTU) per pound: *bituminous* coal and *sub-bituminous* coal (Environmental Protection Agency 2010). The bituminous coal has a higher energy density and can thus generate more electricity for the same amount of mass used than sub-bituminous coal; but its total landed cost per energy content is more expensive (i.e., for one short ton of bituminous coal, its cost of purchase and shipping is more expensive than that of an amount of sub-bituminous coal with equivalent total energy content). This means that if a coal-fired power plant were to consistently have ample production capacity to process enough coal to meet the electricity demand, it would only use the cheaper sub-bituminous coal. However, this is not the case in practice. Specifically, the coal-fired power plants face the following challenges.

- **Coal availability.** The coal-fired power plants typically source coal from coal suppliers under fixed schedules (e.g., weekly) and face uncertain coal availability when deciding replenishment quantities. Specifically, coal is loaded onto cart trains and then shipped by rail from coal mines, so the actual delivery quantity of coal is limited by some uncertain factors such as coal dumper outage at the coal mines or unexpected railroad and weather conditions. The impact is quite pronounced: Based on the operational data of the power plant we worked with, the full quantity of the order is delivered only 23.8% of the time; when full quantity is not delivered, on average only about 75.9% of the ordered quantity is delivered. The undelivered amount of the current order will not be automatically added to the next order, but the power plant decides the total amount of coal in the next order in the next coal replenishment period.
- **Production capacity.** The maximum total mass of coal that can be processed at the power plant between subsequent replenishment periods is both stochastic and non-stationary due to

both unexpected downtime and scheduled maintenance of machines (e.g., the coal pulverizer which is used to turn coal pieces into fine particles before they are sent to the boiler).

- **Demand**. The power generators are connected to the power grid and the plant earns revenue for the amount of power it supplies to the grid to meet the demand load. The demand load during subsequent replenishment periods is both stochastic and non-stationary due to, for example, residential power usage patterns.

The uncertainty and nonstationarity of the operational environment makes the selection and use of fully substitutable inputs a non-trivial problem. Most existing coal-fired plants in the U.S. keep inventory of both types of coal and leverage this flexibility in inputs to meet demand effectively. Specifically, the two types of coal are pulverized into fine particles and blended into a mixture before they go through the thermo-physical process to generate power. The idea is that if using purely the cheaper coal to the maximum production capacity is not sufficient to cover the demand, the plant can blend in the other coal to increase the average energy density of the mixture to increase power generation. This poses several interesting operational questions: When should the firm use both types of coal? In doing so, how much of each type of coal should the firm use? How does this affect the replenishment decisions? Note that the key operational features we identify above for the coal-fired power plants are also present in other industries such as primary metal manufacturing.[2]

While there is a large body of production and inventory management literature which we review later, we are not aware of any model that captures the features we outlined above. Thus, in this paper, we model the coal replenishment and usage problem as a multi-period joint inventory and production problem where a firm can use either of two kinds of raw materials (or their mixture) to produce an end product. The two types of materials have different conversion rates (i.e., the amount of end product that can be produced using one unit mass of a particular material) and vary in their purchasing costs. The production system operates in a make-to-order fashion: When demand for the end product arrives, the firm decides how much of each type of raw material to use in the production line. Although the probability distribution of the production capacity is known, how much production capacity the firm has to convert the inputs to output in every period is a

[2] For example, we have recently worked with a leading industrial aluminum supplier who supplies aluminum sheets to major auto OEMs as well as to can manufacturers (for soft drinks). Since producing aluminum using recycled aluminum scrap requires 95% less energy than creating virgin aluminum, most industrial aluminum manufacturers use aluminum scrap as input to produce the desired aluminum sheets via a recovery process. (Economist 2007) The aluminum supplier we worked with uses different sources of aluminum scrap such as recycled beverage cans and the automobile stamping scrap. Obtaining these scraps has different costs and the amount of scrap available for any source material in any period is uncertain. The different sources of recycled material in the same recovery process result in different rates of recovery for the same amount of raw material. The recovery process is capacitated by the available machine hours.

random variable due to the maintenance that needs to be performed on the production line and other interruptions, and the actual realization of the production capacity becomes known only at the beginning of the period when production decisions need to be made. The produced end product is then sold at a deterministic price in each period, and any unsatisfied demand becomes lost sales. In our motivating example, while electricity price fluctuates every several minutes, coal usage decisions are planned out at a much larger time-scale (e.g., on a daily or weekly basis); for planning purposes, we use the forecasted *average* electricity price during the planning period as an approximation of the per unit revenue the firm earns from selling electricity and model it as a deterministic (but nonstationary) price in each period. In addition to the raw material usage decision, the firm also has to decide in each period the amount of raw materials to replenish from upstream suppliers in the next period. The availability of raw materials that can be delivered is uncertain when the firm places the order. As a result, the actual amount of materials to be delivered equals the minimum of what the firm orders and the amount of the materials available in that period. Since the unit prices of raw materials are specified by long-term contracts with the suppliers and thus do not experience a lot of variation, we model those as deterministic (but nonstationary) prices in each period. To operate such a system efficiently, the firm needs to coordinate the use and replenishment of both raw materials over time under uncertain production capacity and material availability. The main contributions of this paper are summarized as follows.

1. Managerially, we characterize the structure of the optimal replenishment and production policy of the problem and provide insights on how to coordinate both inputs. To the best of our knowledge, this is the first characterization of the optimal policy for a joint production and replenishment problem where two types of substitutable materials with uncertain supply can be blended and converted into an end product in a non-stationary setting.
2. One of the main challenges for analyzing the optimal policy is that the multi-variate optimal value functions are not necessarily differentiatiable, which makes it difficult to use the standard derivative-based first-order optimality condition approach. To address the technical challenge posed by this problem, we develop a novel approximate optimization approach where we transform the original problem into a sequence of approximate problems and use the structure of their optimal solutions to analyze the structure of the optimal policy of the original problem (see Section 4.3.1).
3. Computationally, we develop a heuristic policy based on piece-wise linear approximation of the optimal value function. Since coal usage decision depends on the realizations of demand and capacity, the problem has a high-dimensional state space, so the brute-force approach to compute the approximate value functions (Algorithm 1) can be time-consuming. However, the structure of the optimal policy we characterize allows us to reduce the dimensionality of

the computation of the approximate value functions and develop an algorithm (Algorithm 2) which greatly improves the computational time of the proposed heuristic.

4. We assess the practical benefit of our proposed heuristic by comparing its performance to alternative straw policies on both a synthetic data set and a real data set. Our results suggest that our proposed heuristic provides significant profit improvement over the straw policies, and Algorithm 2 significantly improves the computational time over Algorithm 1. This highlights the practical benefits of our proposed heuristic and the characterization of the optimal policy.

The remainder of the paper is organized as follows. The relevant literature is reviewed in Section 2. We formulate the problem as a Markov decision process in Section 3. We then analyze the model using stochastic dynamic programming and establish its structural properties (Lemma 1) in Section 4.2, characterize a jointly optimal production and ordering policy (Theorems 1 and 2) in Sections 4.3 and 4.4, and develop a heuristic policy in Section 4.5. We then provide evidence of the practical benefit of the proposed heuristic using numerical studies in Section 5. Extensions of the main model are considered in Section 6. Finally, Section 7 concludes the paper.

2. Related Literature

Motivated by the coal-fired power plant example, the model we study in this paper involves using *two* types of substitutable inputs to make one type of outputs via a *shared* production line in a *nonstationary* business environment. This model is related to the production and inventory management literature which focuses on how a firm should use capacitated resources (e.g., equipment and labor) to transform input items purchased from upstream suppliers into output items sold to downstream buyers. One stream of such work focuses on a context where there is a simple one-to-one mapping from input items to output items (e.g., Evans (1967), DeCroix and Arreola-Risa (1998), Nahmias and Schmidt (1984), DeCroix and Arreola-Risa (1998) and Glasserman (1996)); while they also consider the impact of finite production capacity on operations, they do not allow different input items to be blended to produce the output items. Another stream of work allows for more general contexts where either an output item requires multiple input items or an input item is used to make multiple output items, the so-called assemble-to-order production systems, but does not consider the impact of finite production capacity on inventory and production decisions. The general complex mapping between the input items and output items makes it difficult to characterize the optimal policy and its structure remains poorly understood; thus most existing work on assemble-to-order systems focuses on performance evaluation of heuristic allocation controls and inventory policies (typically base-stock policies). (See Song and Zipkin (2003) for a literature review on assemble-to-order systems.) Existing studies that do investigate the structure of the optimal policy typically focus on the stationary setting and derive conditions under which

a myopic policy is optimal. For example, van Mieghem and Rudi (2002) consider in a multi-period *stationary* setting a class of the so-called *newsvendor networks* which involve multiple input and output items, processing and storage units that interact in a very general way. They derive insights on when the problem can be decomposed into repeated and independent single-period problems, and study in those cases the structural property of the optimal policy and the solution approach. While a stationary setting may be a reasonable approximation in certain industries, it does *not* approximate other industries well, for example, the electricity industry where demand and electricity prices fluctuate over time. In contrast to their work, we study a two inputs and one output system in a *nonstationary* setting and fully characterize the structure of the optimal policy which cannot be decomposed into independent single-period problems and cannot be solved via their approach.

The uncertain availability of materials in our model is also related to the literature on stochastic capacity. The stochastically capacitated inventory model was first investigated by Ciarallo et al. (1994) in a single-product setting where they show that the base-stock policy is optimal. This modeling technique has recently been applied to more complex two-dimensional production settings. For example, Hu et al. (2008) study the problem of producing *identical* products in two locations to satisfy local demand where supply is stochastically capacitated and the firm can make lateral transhipment of products to satisfy the demand in the other location. Demirel et al. (2015) study a context where a firm produces two types of products on dedicated production lines both of which face capacity uncertainties, and are then further calibrated on a shared resource. While the two papers have different components in their model (i.e., transhipment versus calibration), they show that the optimal production policies have the same structure. In establishing the optimal production policy in their problems, they use a critical structural property of their dynamic programs elegantly explained in a follow up paper Chen et al. (2015), i.e., $L^\natural$-convexity (i.e., $\{\mathbb{A}_1^k\}_k$ on page 884 in Hu et al. (2008), the "second-order properties" defined on page 828 in Demirel et al. (2015), and Definition 1 in Chen et al. (2015)). Unfortunately, in our problem, the blending feature in the production stage destroys the $L^\natural$-convexity; we can generate counter examples where this property does *not* hold in our setting. Therefore, while we show that the optimal ordering policy has the same structure as the optimal production policies in Hu et al. (2008) and Demirel et al. (2015), we need to use a different argument based on contraction mapping to establish the result.

Our problem is also related to the literature on dual-sourcing (multisourcing) under uncertain supply with random yields, e.g, Anupindi and Akella (1993), Swaminathan and Shanthikumar (1999), Federgruen and Yang (2008, 2011, 2014). All these papers focus on the structure of the firm's optimal ordering decisions, i.e., supplier selection and the allocation of the order across selected suppliers. For example, in Federgruen and Yang (2011), although per unit purchasing costs

vary across suppliers, once the orders are delivered, the usable items from all suppliers are identical and thus the inventory is single dimensional, which leads to the optimality of using a scoring rule to determine which suppliers to source from. However, the scoring rule is not optimal in our setting because different types of materials have different conversion rates in the production stage so the inventory is not exchangeable and the problem cannot be reduced to a single dimensional problem; this makes the optimal policy in our setting much more complicated.

Finally, due to the curse of dimensionality, our model cannot be solved to optimum within a reasonable time frame. Thus, we propose an approximation approach which leverages the policy structure we characterize and provides a heuristic policy. This approach is related to a line of work in the approximate dynamic programming (ADP) literature which leverages the structural properties of the dynamic program to come up with computationally tractable heuristic policies. For example, Secomandi (2008) studies a class of heuristic control policies whose control parameters are determined by re-solving a math program which approximates the underlying Markov decision process. Jiang and Powell (2015) utilize monotonicity properties of the value function of a class of dynamic program to propose an iterative ADP algorithm which has a fast rate of convergence. There is also a line of work which leverage convex properties to develop ADP algorithms. For example, the popular stochastic dual dynamic programming method leverages the convex property of the value function to construct cutting planes to form outer approximation of the value function, and this approach has been recently extended to handle random processes which are not necessarily stage-wise independent (Löhndorf and Shapiro 2019). Nadarajah and Secomandi (2018) utilize the convexity of a class of stochastic dynamic program and generalize the least squares Monte Carlo method, which approximates the true value function with a weighted average of a collection of basis functions whose weights are determined by a least square regression based on estimates of value of certain states obtained by Monte Carlo simulation, to multi-dimensional state space cases. Finally, Halman and Nannicini (2019) leverage affine and convex properties in the objective function of the class of stochastic dynamic program they consider to develop a Finite Polynomial Time Approximation Scheme for stochastic dynamic programs with multidimensional actions and scalar states. While we show that our stochastic dynamic program has convex value functions, in contrast to the approach above which utilizes convexity to directly approximate the value function, our approach is novel in that we leverage the explicitly characterized structure of the optimal policy and discretize it to obtain an approximately optimal policy, and then use that to compute an approximation of the value function. Our approach suggests that by leveraging the explicit characterization of the optimal policy structure of the problem, one could develop efficient and effective ADP algorithms.

3. Model

Before introducing the model and formulating the problem into a Markov decision process (MDP), we introduce some notation. Let $\mathbb{R}_+, \mathbb{Q}_+, \mathbb{Z}_+$ (resp. $\mathbb{R}_{++}, \mathbb{Q}_{++}, \mathbb{Z}_{++}$) denote the set of nonnegative (resp. strictly positive) reals, rationals and integers. Let $\sqcup$ denote the union of sets that do not intersect. Denote by $a \wedge b$ (resp. $a \vee b$) the minimum (resp. maximum) of a and b. All vectors are denoted in boldface.

Consider a make-to-order system as illustrated in Figure 1 where the firm can use either of the two raw materials (or their mixture) to produce a product over a planning horizon of T periods indexed forward by t. In each period t, the demand is D_t, and the firm earns p_t for satisfying each unit of demand. We assume that D_t are independent non-negative random variables with finite support $\mathcal{D}$ (in Section 6, we show our results hold if the demands are correlated and follow a Markov-modulated process). Unmet demands are lost.

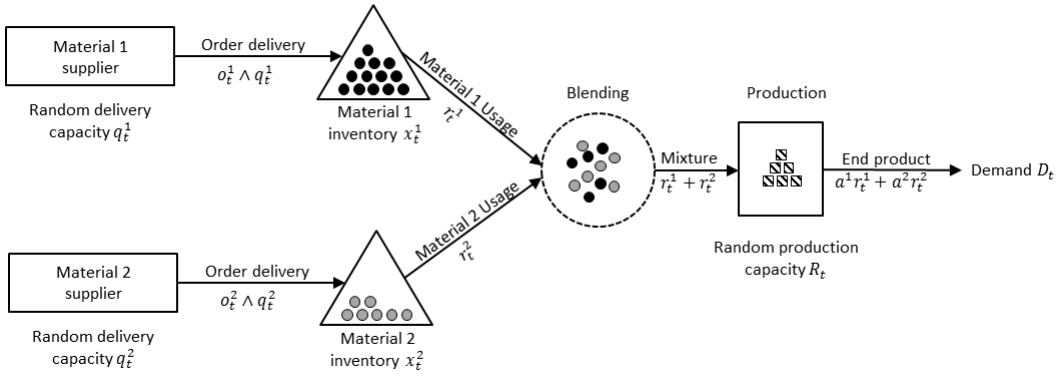


Figure 1 **Problem setting.**

On the production side, we use a non-negative random variable R_t with finite support $\mathcal{R}$ to denote the production capacity of the production line in period t. The production line can process two types of raw materials indexed by j. Due to the nature of the problem motivating this research (e.g., both coal and primary metal are measured by weight and ordered in large quantities), we use continuous variables to model the quantity of both types of raw materials. Denote by a^j the conversion rate of type-j material, i.e., the production line can convert each unit of type-j material into a^j units of the end product. We assume without loss of generality that $a^1 < a^2$, i.e., type-1 material has a lower conversion rate. Let c_t^j denote the per unit purchasing cost of type-j material in period t and let h_t^j denote its per unit holding cost in period t. Due to delivery uncertainty mentioned earlier, the amount of raw materials available to be delivered in period t is stochastic and denoted by a discrete random variable q_t^j. We assume that q_t^j are independent across t and j and has a finite and nonempty support $\mathcal{Q}_t^j$ (in Section 6, we show our results hold if the material

availability distributions are correlated and follow a Markov-modulated process). In particular, we label the elements in $\mathscr{Q}_t^j$ in an increasing order, i.e., $\mathscr{Q}_t^j = \{q_{(k)t}^j : 1 \le k \le |\mathscr{Q}_t^j|\}$ and, for any $1 \le k < k' \le |\mathscr{Q}_t^j|$, we have $q_{(k)t}^j < q_{(k')t}^j$. For generality, we also allow $\mathscr{Q}_t^j$ to include ∞. Note that, when $|\mathscr{Q}_t^j| = 1$, this model reduces to the deterministic capacity case: $q_{(1)t}^j = \infty$ corresponds to unlimited material supply whereas $q_{(1)t}^j < \infty$ corresponds to limited supply.

The firm decides in period t for each type-j material, how much to use for production, r_t^j, and how much to order to replenish the inventory in the next period (period $t+1$), o_{t+1}^j. Let y_t^j denote the inventory level of type-j material before production takes place in period t. In this make-to-order system, the production decision is modeled as a recourse decision, i.e., the firm decides r_t^1 and r_t^2 after observing the realization of D_t and R_t.[3] Specifically, the timeline of the events in the model is as follows. Before the planning horizon starts (i.e., $t = 0$), the firm starts with an initial inventory of x_1^j for type-j material and decides the order quantity o_1^j for type j material to be delivered in period 1. Then for each period $t = 1, \dots T$, the following events occur in sequence:

E1 The available type-j material that can be delivered in period t, i.e., q_t^j, is realized and $o_t^j \wedge q_t^j$ units of type-j material arrive at the beginning of period t. In the meantime, the firm incurs the purchasing cost $\sum_{j=1}^2 c_t^j(o_t^j \wedge q_t^j)$. (Recall that the undelivered portion of the order is not automatically delivered in the next period; instead, the firm decides in the next period again how much material it wants.)

E2 The demand D_t and capacity R_t are realized and observed by the firm.

E3 The firm decides r_t^1 and r_t^2. In total, these convert into $a^1 r_t^1 + a^2 r_t^2$ end products and reduce the type j material's inventory level to $y_t^j - r_t^j$. The firm earns a total revenue of $p_t(a^1 r_t^1 + a^2 r_t^2)$, and incurs a total holding cost of $\sum_{j=1}^2 h_t^j(y_t^j - r_t^j)$.

E4 The firm places an order of o_{t+1}^j type-j material to be delivered in period $t+1$. (Note that this event does not occur when $t = T$ because it is the end of the planning horizon.)

For notational convenience, let $\mathbf{r}_t = (r_t^1, r_t^2), \mathbf{o}_t = (o_t^1, o_t^2)$, $\mathbf{y}_t = (y_t^1, y_t^2)$ and $\mathbf{q}_t = (q_t^1, q_t^2)$. We define a *non-anticipative policy* $\pi = (\pi_0, \pi_1, \pi_2, \dots, \pi_T)$ as a set of $T+1$ functions where for each $t = 0, \dots, T$, π_t maps the information the firm has observed at the beginning of period t into decisions $\mathbf{r}_t \in \mathbb{R}_+^2$ and $\mathbf{o}_{t+1} \in \mathbb{R}_+^2$ (note that when $t = 0$, the decisions are $\mathbf{o}_1 \in \mathbb{R}_+^2$; when $t = T$, the decisions are $\mathbf{r}_T \in \mathbb{R}_+^2$). (We suppress the decisions' dependency on π for notational simplicity). Due to the production capacity constraint and the fact that excess production of electrical power cannot be stored, a non-anticipative policy is *feasible* if $r_t^j \le y_t^j$, $r_t^1 + r_t^2 \le R_t$ and $a^1 r_t^1 + a^2 r_t^2 \le D_t$. The firm's objective is to find a feasible non-anticipative policy which maximizes its expected profit. Note

[3] In our motivating example, while the demand load is stochastic in between decision periods, the plant uses its demand load forecast and the predicted production capacity based on machine maintenance schedules and the revealed labor shortages.

that at the beginning of each decision period, the only pay-off relevant variables are $\mathbf{y}_t$, D_t and R_t. Hence, we formulate the problem into a Markov decision process (MDP) as follows:

- *State space* in period $t \in \{0, 1, \ldots, T\}$: $\mathcal{S}_0 := \mathbb{R}^2_+$ and $\mathcal{S}_t := \mathbb{R}^2_+ \times \mathcal{D} \times \mathcal{R}$ for $t \neq 0$.
- *Action space* in period $t \in \{0, 1, \ldots, T\}$: For any state $\mathbf{s}_0 = (x_1^1, x_1^2) \in \mathcal{S}_0$, $\mathcal{A}_0(\mathbf{s}_0) := \{(o_1^1, o_1^2) \in \mathbb{R}^2_+\}$; for $t \neq 0$, for any state $\mathbf{s}_t = (y_t^1, y_t^2, D_t, R_t) \in \mathcal{S}_t$,

$$\mathcal{A}_t(\mathbf{s}_t) := \begin{cases} \{(r_T^1, r_T^2) \in \mathbb{R}^2_+ : \sum_{j=1}^2 r_T^j \leq R_T, \sum_{j=1}^2 a^j r_T^j \leq D_T, r_T^j \leq y_T^j\}, & \text{if } t = T \\ \{(o_t^1, o_t^2, r_t^1, r_t^2) \in \mathbb{R}^4_+ : \sum_{j=1}^2 r_t^j \leq R_t, \sum_{j=1}^2 a^j r_t^j \leq D_t, r_t^j \leq y_t^j\}, & \text{otherwise} \end{cases}.$$

- *Transition probability* to state $\mathbf{s}_{t+1} \in \mathcal{S}_{t+1}$ in period $t+1$ from state $\mathbf{s}_t \in \mathcal{S}_t$ in period t given the action $\mathbf{a}_t \in \mathcal{A}_t(\mathbf{s}_t)$, for any $t \in \{0, 1, \ldots, T-1\}$: When $t = 0$, for any $\mathbf{s}_1 = (y_1^1, y_1^2, D_1, R_1)$, $\mathbf{s}_0 = (x_1^1, x_1^2)$, and $\mathbf{a}_0 = (o_1^1, o_1^2)$, the transition probability is defined as $\sigma_0(\mathbf{s}_0, \mathbf{s}_1, \mathbf{a}_0) := \mathbb{P}(\mathbf{s}_1 | \mathbf{s}_0, \mathbf{a}_0) = \mathbb{P}(y_1^j = x_1^j + o_1^j \wedge q_1^j, \forall j = 1, 2, D_1, R_1)$; otherwise, for any $\mathbf{s}_{t+1} = (y_{t+1}^1, y_{t+1}^2, D_{t+1}, R_{t+1})$, $\mathbf{s}_t = (y_t^1, y_t^2, D_t, R_t)$, and $\mathbf{a}_t = (r_t^1, r_t^2, o_{t+1}^1, o_{t+1}^2)$, the transition probability is defined as $\sigma_t(\mathbf{s}_t, \mathbf{s}_{t+1}, \mathbf{a}_t) := \mathbb{P}(\mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{a}_t) = \mathbb{P}(y_{t+1}^j = y_t^j - r_t^j + o_{t+1}^j \wedge q_{t+1}^j, \forall j = 1, 2, D_{t+1}, R_{t+1})$.
- *Period t Reward function* if transitioning to state $\mathbf{s}_{t+1} \in \mathcal{S}_{t+1}$ in period $t+1$ from state $\mathbf{s}_t \in \mathcal{S}_t$ in period t with the action $\mathbf{a}_t \in \mathcal{A}_t(\mathbf{s}_t)$, for any $t \in \{0, 1, \ldots, T-1\}$: When $t = 0$, for any $\mathbf{s}_1 = (y_1^1, y_1^2, D_1, R_1)$, $\mathbf{s}_0 = (x_1^1, x_1^2)$, and $\mathbf{a}_0 = (o_1^1, o_1^2)$, the reward function is defined as $f_0(\mathbf{s}_0, \mathbf{s}_1, \mathbf{a}_0) := -\sum_{j=1}^2 c_1^j (y_1^j - x_1^j)$; otherwise, for any $\mathbf{s}_{t+1} = (y_{t+1}^1, y_{t+1}^2, D_{t+1}, R_{t+1})$, $\mathbf{s}_t = (y_t^1, y_t^2, D_t, R_t)$, and $\mathbf{a}_t = (r_t^1, r_t^2, o_{t+1}^1, o_{t+1}^2)$, the reward function is defined as $f_t(\mathbf{s}_t, \mathbf{s}_{t+1}, \mathbf{a}_t) := p_t \sum_{j=1}^2 a^j r_t^j - \sum_{j=1}^2 h_t^j (y_t^j - r_t^j) - \sum_{j=1}^2 c_{t+1}^j (y_{t+1}^j - y_t^j)$. *Period T Reward function* if in state $\mathbf{s}_T = (y_T^1, y_T^2, D_T, R_T) \in \mathcal{S}_T$ with the action $\mathbf{a}_T = (r_T^1, r_T^2) \in \mathcal{A}_T(\mathbf{s}_T)$ is defined as $f_T(\mathbf{s}_T, \mathbf{a}_T) := p_T \sum_{j=1}^2 a^j r_T^j - \sum_{j=1}^2 h_T^j (y_T^j - r_T^j)$.

Thus, for any feasible non-anticipative policy π, the expected profit equals

$$V^\pi(x_1^1, x_1^2) = \mathbf{E}^\pi \left[\sum_{t=0}^{T-1} f_t(\mathbf{s}_t, \mathbf{s}_{t+1}, \mathbf{a}_t) + f_T(\mathbf{s}_T, \mathbf{a}_T) \right],$$

where $\mathbf{a}_t = \pi_t(\mathbf{s}_t)$, and the expectation is taken over $\mathbf{s}_{t+1} \sim \sigma_t(\mathbf{s}_t, \mathbf{s}_{t+1}, \pi_t(\mathbf{s}_t))$. Thus, the firm should solve $\sup_\pi V^\pi(x_1^1, x_1^2)$. Note that by the theory of Markov decision process, it is without loss of optimality to focus on the class of *Markov policies* which satisfies the following: π_t is a function that maps any $\mathbf{s}_t \in \mathcal{S}_t$ into an action $\mathbf{a}_t \in \mathcal{A}_t(\mathbf{s}_t)$. This allows us to use backward induction to characterize the optimal policy in the next section.

4. Analysis

In this section, we first formulate the optimality conditions of the problem via a collection of Bellman equations which decomposes each decision period into two stages, the production stage and the ordering stage, in Section 4.1. This gives rise to a two-stage stochastic dynamic program. We then use backward induction to characterize the properties of the value functions in Section 4.2 which we leverage to provide a full characterization of the jointly optimal production and replenishment policy in Sections 4.3 and 4.4. Finally, we propose a heuristic policy in Section 4.5 and explain how the structure of the optimal policy we characterize help increase computational efficiency.

4.1. Preliminary Material

It is convenient to denote by $\mathbf{w}_t = (w_t^1, w_t^2) := \mathbf{y}_t - \mathbf{r}_t$ the inventory levels after production and by $\mathbf{x}_t = (x_t^1, x_t^2) := \mathbf{w}_{t-1}$ the inventory level before the period t raw material delivery arrives. By definition, we have: $\mathbf{y}_t = \mathbf{x}_t + \mathbf{o}_t \wedge \mathbf{q}_t$. Note that, once the production decision $\mathbf{r}_t$ is made, $\mathbf{o}_{t+1}$ only depends on the inventory levels after production $\mathbf{w}_t$. Hence, we can formulate this problem as a two-stage dynamic program characterized by the following Bellman equations for each period $t \in \{1, \ldots, T\}$:

$$\hat{V}_t(\mathbf{x}_t) = \max_{\mathbf{o}_t \in \mathbb{R}^2_+} \; -\mathbb{E}_{\mathbf{q}_t}\left[\sum_{j=1}^{2} c_t^j (o_t^j \wedge q_t^j)\right] + \mathbb{E}_{D_t, R_t, \mathbf{q}_t}\left[\hat{U}_t((\mathbf{o}_t \wedge \mathbf{q}_t) + \mathbf{x}_t, D_t, R_t)\right],$$

where $\hat{U}(.,.,.)$ is defined as

$$\begin{aligned}\hat{U}_t(\mathbf{y}_t, D_t, R_t) = \max_{\mathbf{r}_t} \; & p_t(a^1 r_t^1 + a^2 r_t^2) - \sum_{j=1}^{2} h_t^j (y_t^j - r_t^j) + \hat{V}_{t+1}(\mathbf{y}_t - \mathbf{r}_t) \\ \text{s.t.} \quad & a^1 r_t^1 + a^2 r_t^2 \le D_t, \;\; r_t^1 + r_t^2 \le R_t, \;\; 0 \le r_t^j \le y_t^j, \;\; \forall j = 1, 2\end{aligned}$$

In addition, we also have the boundary condition $\hat{V}_{T+1}(\mathbf{x}_{T+1}) = 0$ for all $\mathbf{x}_{T+1} \in \mathbb{R}^2_+$. One can view $\hat{V}_t(\mathbf{x}_t)$ as the value function for the replenishment stage and $\hat{U}_t(\mathbf{y}_t, D_t, R_t)$ as the value function for the production stage. The above profit maximization problem can be cast into an equivalent cost minimization problem as follows. Define $V_{T+1}(\mathbf{x}_{T+1}) := -\hat{V}_{T+1}(\mathbf{x}_{T+1}) + \sum_{j=1}^{2} c_{T+1}^j x_{T+1}^j$ where $c_{T+1}^1 = c_{T+1}^2 = 0$ and, for all $t = 1, \ldots, T$

$$\begin{aligned}U_t(\mathbf{y}_t, D_t, R_t) &:= -\hat{U}_t(\mathbf{y}_t, D_t, R_t) + a^1 p_t y_t^1 + a^2 p_t y_t^2, \\ V_t(\mathbf{x}_t) &:= -\hat{V}_t(\mathbf{x}_t) + c_t^1 x_t^1 + c_t^2 x_t^2, \\ J_t(w_t^1, w_t^2) &:= (a^1 p_t + h_t^1 - c_{t+1}^1) w_t^1 + (a^2 p_t + h_t^2 - c_{t+1}^2) w_t^2 + V_{t+1}(w_t^1, w_t^2), \\ G_t(\mathbf{y}_t, \mathbf{x}_t) &:= \mathbb{E}_{\mathbf{q}_t}[\bar{G}_t(\mathbf{y}_t \wedge (\mathbf{x}_t + \mathbf{q}_t))], \forall t = 1, \ldots, T, \text{where} \\ \bar{G}_t(y_t^1, y_t^2) &:= \textstyle\sum_{j=1}^{2} (c_t^j - a^j p_t) y_t^j + \mathbb{E}_{D_t, R_t}\left[U_t(\mathbf{y}_t, D_t, R_t)\right].\end{aligned}$$

Then, we can write the above Bellman equations for each $t \leq T$ as follows:

$$\begin{aligned}
\mathbf{OPT1}_t(\mathbf{x}_t): \qquad & V_t(\mathbf{x}_t) = && \min_{\mathbf{y}_t \geq \mathbf{x}_t} \; G_t(\mathbf{y}_t, \mathbf{x}_t) \\
\mathbf{OPT2}_t(\mathbf{y}_t, D_t, R_t): \qquad & U_t(\mathbf{y}_t, D_t, R_t) = && \min_{\mathbf{w}_t} \quad J_t(w_t^1, w_t^2) \\
& && \text{s.t.} \quad 0 \leq w_t^j \leq y_t^j, \;\; \forall j = 1,2 \\
& && \quad a^1 w_t^1 + a^2 w_t^2 \geq a^1 y_t^1 + a^2 y_t^2 - D_t, \qquad (1) \\
& && \quad w_t^1 + w_t^2 \geq y_t^1 + y_t^2 - R_t. \qquad (2)
\end{aligned}$$

In addition, we have the boundary condition $V_{T+1}(\mathbf{x}_{T+1}) = 0$ for all $\mathbf{x}_{T+1} \in \mathbb{R}^2_+$. Define $\mathscr{E}_2(\mathbf{y}_t, D_t, R_t)$ as the set of feasible solutions to $\mathbf{OPT2}_t$ and define $\mathscr{E}_1(\mathbf{x}_t, M) := \{\mathbf{y}_t \in \mathbb{R}^2_+ : x_t^j \leq y_t^j \leq \max\{x_t^j, M\}, j = 1,2\}$ for any $M > 0$; note that $\mathscr{E}_1(\mathbf{x}_t, \infty)$ is the feasible region of $\mathbf{OPT1}_t$. Finally, for notational brevity, whenever there is no confusion, we will often drop the dependency of **OPT1** (resp. **OPT2**) on $\mathbf{x}_t$ (resp. $\mathbf{y}_t, D_t, R_t$) and t.

4.2. Structural properties of the Stochastic Dynamic Program

The two-stage dynamic program formulation of our joint production and replenishment problem has the following structural properties. For all $t = 0, \ldots, T$, the following holds:

$$\begin{aligned}
&\mathbb{H}_t : V_{t+1}(.) \text{ is convex, continuous and supermodular on } \mathbb{R}^2_+, \\
&\text{and } \liminf_{x_{t+1}^j \to \infty} \tfrac{V_{t+1}(\mathbf{x}_{t+1})}{x_{t+1}^j} \geq c_{t+1}^j \text{ for both } j = 1,2.
\end{aligned}$$

Note that $\{\mathbb{H}_t\}_{t=1}^T$ holds by backward induction: By definition of $V_{T+1}(\mathbf{x}_{T+1})$, $\mathbb{H}_T$ holds trivially and serves as our inductional basis; the induction step, i.e., $\mathbb{H}_t$ implies $\mathbb{H}_{t-1}$, along with other useful properties are established in Lemma 1 below. (Unless otherwise noted, all proofs can be found in the Online Appendix.)

LEMMA 1. *Suppose $\mathbb{H}_t$ holds. Then: (a) $J_t(.)$ is convex, continuous, supermodular, and $\lim_{w_t^j \to \infty} J_t(\mathbf{w}_t) = \infty$ for all j; (b) for any D_t, R_t, $U_t(., D_t, R_t)$ is convex, continuous, supermodular and $\liminf_{y_t^j \to \infty} \frac{U_t(\mathbf{y}_t, D_t, R_t)}{y_t^j} \geq a^j p_t + h_t^j$ for all j; (c) $\bar{G}_t(.)$ is convex, supermodular and $\lim_{y_t^j \to \infty} \bar{G}_t(\mathbf{y}_t) = \infty$ for all j; (d) $\mathbb{H}_{t-1}$ holds, i.e., $V_t(.)$ is convex, continuous, supermodular and $\liminf_{x_t^j \to \infty} \frac{V_t(\mathbf{x}_t)}{x_t^j} \geq c_t^j$ for all j.*

The structural properties of the value functions V_t and U_t help to decouple the analysis of the optimal production and ordering decisions and allow us to separately characterize the structure of the production policy and the ordering policy next.

4.3. Optimal Production Policy

We will show in this subsection that an optimal production policy follows a *state-dependent Use-down-to/Balancing* (**UB**) policy. Before formally defining a **UB** policy, we first introduce some notation and provide a graphical illustration of the constraint set of **OPT2**. Let $K^j(D_t,R_t) := \min\{R_t, \frac{D_t}{a^j}\}$ for $j=1,2$ and define two continuous functions:

$$L^1(D_t,R_t)=\begin{cases} K^1(D_t,R_t), & D_t \le a^1R_t \\ \frac{a^2R_t-D_t}{a^2-a^1}, & a^1R_t < D_t < a^2R_t \\ 0, & D_t \ge a^2R_t \end{cases}, \quad L^2(D_t,R_t)=\begin{cases} 0, & D_t \le a^1R_t \\ \frac{D_t-a^1R_t}{a^2-a^1}, & a^1R_t < D_t < a^2R_t \\ K^2(D_t,R_t), & D_t \ge a^2R_t \end{cases}.$$

Note that $K^j(D_t,R_t)$ is the maximum possible quantity of type-j material the firm can use in period t, and the usage combination $(L^1(D_t,R_t),L^2(D_t,R_t))$ produces most end products and uses most production capacity among all usage combinations that satisfy the demand and capacity constraints, i.e., (1) and (2). For brevity, we will suppress the dependency of L^j, K^j on D_t, R_t. We denote by $\mathscr{F}(D_t,R_t)$ the convex hull of the points $(0,0),(K^1,0),(0,K^2),(L^1,L^2)$. The set $\mathscr{F}(D_t,R_t)$ can be interpreted as the set of feasible *usage quantities* when there is unlimited inventory of both types of materials. Figure 2 illustrates the shape of the feasible region of **OPT2**, $\mathscr{E}_2(\mathbf{y}_t,D_t,R_t)$ under different $\mathbf{y}_t, D_t, R_t$. Note that $\mathscr{E}_2(\mathbf{y}_t,D_t,R_t)$ (shaded areas in Figure 2) can be written as the intersection of $\mathbf{y}_t - \mathscr{F}(D_t,R_t)$ (areas within the bold lines in Figure 2) and $\mathbb{R}^2_+$. Note also that the shape of $\mathscr{F}(D_t,R_t)$ depends on D_t and R_t and can be categorized into three cases depending on the ratio of demand and production capacity, i.e. D_t/R_t (when $R_t=0, D_t/R_t := \infty$). In the case where the *demand/capacity ratio is low* (i.e., $D_t/R_t \le a^1$), even though production is capacitated, using only the cheaper material with lower conversion rate already satisfies the demand; so, as illustrated in the left graph in Figure 2, ignoring the capacity constraint (2) (which corresponds to the red dotted line) of **OPT2** does not affect the feasible region. In the case where the *demand/capacity ratio is high* (i.e., $D_t/R_t \ge a^2$), even using the material with higher conversion rate and with full production capacity cannot satisfy all the demand; so, as illustrated in the right graph in Figure 2, ignoring the demand constraint (1) (which corresponds to the blue dashed line) does not affect the feasible region. In the case where the *demand/capacity ratio is medium* (i.e., $a^1 < D_t/R_t < a^2$), as illustrated in the middle graph in Figure 2, neither the demand constraint (1) nor the capacity constraint (2) can be ignored without affecting the feasible region.

We are now ready to formally define our **UB** policy below.

Definition 1. For each t, a **UB** policy is characterized by two non-increasing functions $\gamma_t^j(.): \mathbb{R}_+ \to \mathbb{R}_+$ $(j=1,2)$ and two vector-valued functions $\alpha_t(.): \mathbb{R}_+ \to \mathbb{R}^2_+$ and $\beta_t(.): \mathbb{R}_+ \to \mathbb{R}^2_+$, all of which intersect at one point $\bar{\mathbf{w}}_t \in \mathbb{R}^2_+$. These functions jointly define the following partition of state space (see Figure 3(a) for an illustration) and decision rules :

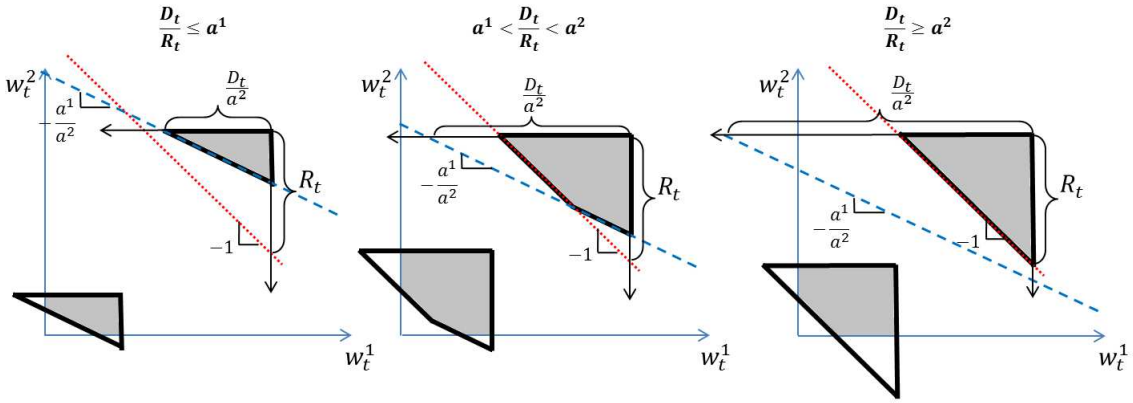


Figure 2 **Illustration of $\mathbf{y}_t - \mathscr{F}(D_t, R_t)$ and $\mathscr{E}_2(\mathbf{y}_t, D_t, R_t)$ under different $\mathbf{y}_t, D_t, R_t$. The bold line area corresponds to $\mathbf{y}_t - \mathscr{F}(D_t, R_t)$, the shaded area corresponds to $\mathscr{E}_2(\mathbf{y}_t, D_t, R_t)$, the red dotted line corresponds to the capacity constraint and the blue dashed line corresponds to the demand constraint.**

1. State Space Partition. For each D_t, R_t, we partition the whole state space into:

$$
\begin{aligned}
\mathscr{R}_0 &:= \left\{ \mathbf{y}_t \in \mathbb{R}_+^2 : \begin{array}{l} y_t^1 \le \bar{w}_t^1 + L^1 \text{ or } y_t^2 \le \bar{w}_t^2 + L^2; \\ y_t^1 \le \bar{\gamma}_t^1(y_t^2) \text{ if } y_t^2 \le \bar{w}_t^2 + L^2; y_t^2 \le \bar{\gamma}_t^2(y_t^1) \text{ if } y_t^1 \le \bar{w}_t^1 + L^1. \end{array} \right\}, \\
\mathscr{R}_1 &:= \mathscr{R}_{1a} \cup \mathscr{R}_{1b}, \\
\mathscr{R}_2 &:= \left\{ \mathbf{y}_t \in \mathbb{R}_+^2 - \mathscr{R}_0 : \begin{array}{l} \alpha_t^1(a^1 y_t^1 + a^2 y_t^2 - D_t) < y_t^1 < \alpha_t^1(a^1 y_t^1 + a^2 y_t^2 - D_t) + L^1 \text{ and} \\ a^1 y_t^1 + a^2 y_t^2 > a^1 \bar{w}_t^1 + a^2 \bar{w}_t^2 + D_t \end{array} \right\}, \\
\mathscr{R}_3 &:= \mathscr{R}_{3a} \cup \mathscr{R}_{3b} \cup \mathscr{R}_{3c}, \\
\mathscr{R}_4 &:= \left\{ \mathbf{y}_t \in \mathbb{R}_+^2 - \mathscr{R}_0 : \begin{array}{l} \beta_t^1(y_t^1 + y_t^2 - R_t) + L^1 < y_t^1 < \beta_t^1(y_t^1 + y_t^2 - R_t) + K^1 \text{ and} \\ y_t^1 + y_t^2 > \bar{w}_t^1 + \bar{w}_t^2 + R_t \end{array} \right\}, \\
\mathscr{R}_5 &:= \mathscr{R}_{5a} \cup \mathscr{R}_{5b},
\end{aligned}
$$

where $\mathscr{R}_{1a}, \mathscr{R}_{1b}, \mathscr{R}_{3a}, \mathscr{R}_{3b}, \mathscr{R}_{3c}, \mathscr{R}_{5a}, \mathscr{R}_{5b}$ are defined as

$$
\begin{aligned}
\mathscr{R}_{1a} &:= \left\{ \mathbf{y}_t \in \mathbb{R}_+^2 - \mathscr{R}_0 : y_t^1 \le \alpha_t^1(a^1 y_t^1 + a^2 y_t^2 - D_t) \text{ and } a^1 y_t^1 + a^2 y_t^2 > a^1 \bar{w}_t^1 + a^2 \bar{w}_t^2 + D_t \right\} \\
\mathscr{R}_{1b} &:= \left\{ \mathbf{y}_t \in \mathbb{R}_+^2 - \mathscr{R}_0 : \begin{array}{l} y_t^1 < \bar{w}_t^1 \text{ and } a^1 y_t^1 + a^2 y_t^2 \le a^1 \bar{w}_t^1 + a^2 \bar{w}_t^2 + D_t \\ \text{and } y_t^1 + y_t^2 \le \bar{w}_t^1 + \bar{w}_t^2 + R_t \text{ and } y_t^2 > \bar{\gamma}_t^2(y_t^1) \end{array} \right\} \\
\mathscr{R}_{3a} &:= \left\{ \mathbf{y}_t \in \mathbb{R}_+^2 - \mathscr{R}_0 : \begin{array}{l} a^1 y_t^1 + a^2 y_t^2 > a^1 \bar{w}_t^1 + a^2 \bar{w}_t^2 + D_t \text{ and } y_t^1 + y_t^2 \le \bar{w}_t^1 + \bar{w}_t^2 + R_t \\ \text{and } y_t^1 \ge \alpha_t^1(a^1 y_t^1 + a^2 y_t^2 - D_t) + L^1 \end{array} \right\} \\
\mathscr{R}_{3b} &:= \left\{ \mathbf{y}_t \in \mathbb{R}_+^2 - \mathscr{R}_0 : \begin{array}{l} a^1 y_t^1 + a^2 y_t^2 \le a^1 \bar{w}_t^1 + a^2 \bar{w}_t^2 + D_t \text{ and } y_t^1 + y_t^2 > \bar{w}_t^1 + \bar{w}_t^2 + R_t \\ \text{and } y_t^1 \le \beta_t^1(y_t^1 + y_t^2 - R_t) + L^1 \end{array} \right\} \\
\mathscr{R}_{3c} &:= \left\{ \mathbf{y}_t \in \mathbb{R}_+^2 - \mathscr{R}_0 : \begin{array}{l} a^1 y_t^1 + a^2 y_t^2 > a^1 \bar{w}_t^1 + a^2 \bar{w}_t^2 + D_t \text{ and } y_t^1 + y_t^2 > \bar{w}_t^1 + \bar{w}_t^2 + R_t \\ \text{and } \alpha_t^1(a^1 y_t^1 + a^2 y_t^2 - D_t) + L^1 \le y_t^1 \le \beta_t^1(y_t^1 + y_t^2 - R_t) + L^1 \end{array} \right\} \\
\mathscr{R}_{5a} &:= \left\{ \mathbf{y}_t \in \mathbb{R}_+^2 - \mathscr{R}_0 : y_t^1 \ge \beta_t^1(y_t^1 + y_t^2 - R_t) + K^1 \text{ and } y_t^1 + y_t^2 > \bar{w}_t^1 + \bar{w}_t^2 + R_t \right\} \\
\mathscr{R}_{5b} &:= \left\{ \mathbf{y}_t \in \mathbb{R}_+^2 - \mathscr{R}_0 : \begin{array}{l} y_t^2 < \bar{w}_t^2 \text{ and } y_t^1 + y_t^2 \le \bar{w}_t^1 + \bar{w}_t^2 + R_t \\ \text{and } a^1 y_t^1 + a^2 y_t^2 \le a^1 \bar{w}_t^1 + a^2 \bar{w}_t^2 + D_t \text{ and } y_t^1 > \bar{\gamma}_t^1(y_t^2) \end{array} \right\}
\end{aligned}
$$

and the functions $\bar{\gamma}_t^j : [0, \bar{w}_t^{3-j} + L^{3-j}] \to \mathbb{R}_+$ for $j = 1, 2$ are defined as

$$
\bar{\gamma}_t^1(y_t^2) = \begin{cases} \gamma_t^1(y_t^2) + K^1, & 0 \le y_t^2 \le \bar{w}_t^2 \\ \bar{w}_t^1 + K^1 + (y_t^2 - \bar{w}_t^2)\frac{L^1 - K^1}{L^2}, & \bar{w}_t^2 < y_t^2 \le \bar{w}_t^2 + L^2 \end{cases} \tag{3}
$$

$$\bar{\gamma}_t^2(y_t^1) = \begin{cases} \gamma_t^2(y_t^1) + K^2, & 0 \le y_t^1 \le \bar{w}_t^1 \\ \bar{w}_t^2 + K^2 + (y_t^1 - \bar{w}_t^1)\frac{L^2 - K^2}{L^1}, & \bar{w}_t^1 < y_t^1 \le \bar{w}_t^1 + L^1 \end{cases} \tag{4}$$

2. Usage Decision Rule. Let $\hat{\gamma}_t^j(y_t^{3-j}) := \gamma_t^j(y_t^{3-j}) \vee \bar{w}_t^j$. The usage decisions $\mathbf{w}_t^* = (w_t^{*1}, w_t^{*2})$ are

a. (Use-down-to) For $\mathbf{y}_t \in \mathscr{R}_0$, $w_t^{*j} = y_t^j \wedge \hat{\gamma}_t^j(y_t^{3-j})$;

b. (Balancing) For $\mathbf{y}_t \in \mathbb{R}_+^2 - \mathscr{R}_0$, we divide into five cases:

i. For $\mathbf{y}_t \in \mathscr{R}_1$, $w_t^{*1} = y_t^1, w_t^{*2} = y_t^2 - K^2$;

ii. For $\mathbf{y}_t \in \mathscr{R}_2$, $w_t^{*1} = \alpha_t^1(a^1 y_t^1 + a^2 y_t^2 - D_t), w_t^{*2} = \alpha_t^2(a^1 y_t^1 + a^2 y_t^2 - D_t)$;

iii. For $\mathbf{y}_t \in \mathscr{R}_3$, $w_t^{*1} = y_t^1 - L^1, w_t^{*2} = y_t^2 - L^2$;

iv. For $\mathbf{y}_t \in \mathscr{R}_4$, $w_t^{*1} = \beta_t^1(y_t^1 + y_t^2 - R_t), w_t^{*2} = \beta_t^2(y_t^1 + y_t^2 - R_t)$;

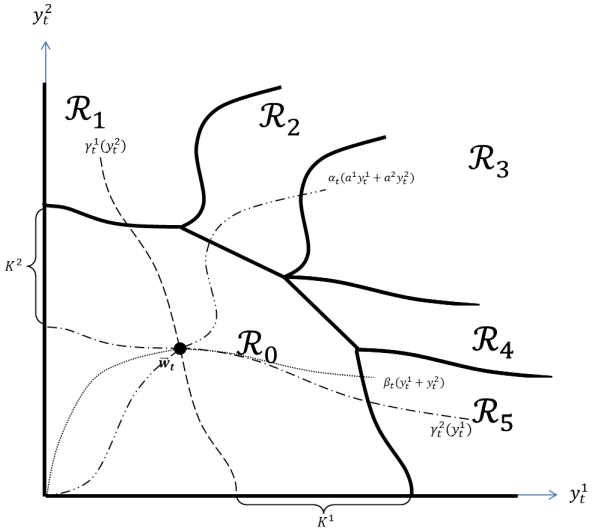
v. For $\mathbf{y}_t \in \mathscr{R}_5$, $w_t^{*1} = y_t^1 - K^1, w_t^{*2} = y_t^2$.

We state our main result in this section.

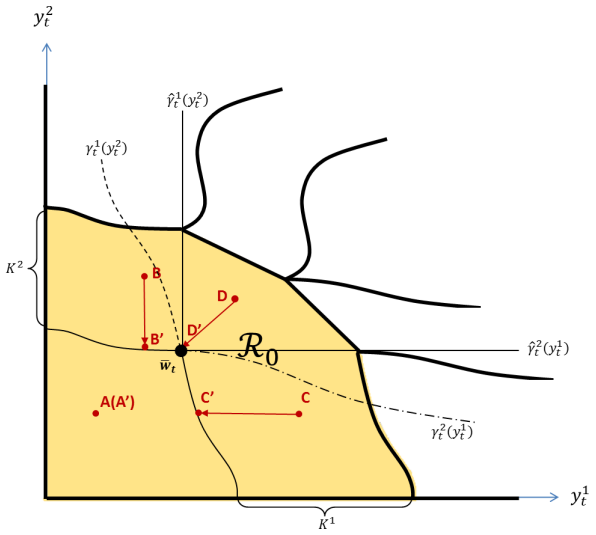
THEOREM 1. *For any t, there exist two non-increasing functions $\gamma_t^j(.) : \mathbb{R}_+ \to \mathbb{R}_+$ for $j = 1, 2$ and two vector-valued functions $\alpha_t(.) : \mathbb{R}_+ \to \mathbb{R}_+^2$ and $\beta_t(.) : \mathbb{R}_+ \to \mathbb{R}_+^2$ that all intersect at a point $\bar{\mathbf{w}}_t \in \mathbb{R}_+^2$, such that the corresponding* **UB** *policy is optimal. Moreover, $\bar{\mathbf{w}}_t \in \arg\min_{\mathbf{w}_t \in \mathbb{R}_+^2} J_t(\mathbf{w}_t)$, $\gamma_t^j(y_t^{3-j}) \in \arg\min_{w_t^j \in \mathbb{R}_+} J_t(w_t^j, y_t^{3-j})$ for any $j = 1, 2$, and $\alpha_t(x) \in \arg\min_{\mathbf{w}_t \in \mathbb{R}_+^2} \{J_t(\mathbf{w}_t) : a^1 w_t^1 + a^2 w_t^2 = x\}$ and $\beta_t(x) \in \arg\min_{\mathbf{w}_t \in \mathbb{R}_+^2} \{J_t(\mathbf{w}_t) : w_t^1 + w_t^2 = x\}$ for all $x \in \mathbb{R}_+$.*

Theorem 1 provides the structure of an optimal production policy and also establishes results that significantly simplify the computation of the optimal policy compared to solving the dynamic program $\{\mathbf{OPT1}_t, \mathbf{OPT2}_t\}_{t=1}^T$ directly. Indeed, note that the state space of the dynamic program for the production stage (see $\mathbf{OPT2}_t$) has four dimensions (i.e., y_t^1, y_t^2, D_t, R_t) and the action space has two dimensions (i.e., w_t^1 and w_t^2). Therefore, a naive brute-force implementation of this high-dimensional dynamic program can be very time consuming: for each t, one needs to solve a two-variable optimization problem for each state in a four dimensional state space. Interestingly, Theorem 1 states that the task of solving this high-dimensional dynamic program can be broken down into solving, for each t, four single-variable optimizations for each state in a one dimensional state space, and one two-variable optimization. This reduction of dimensionality greatly improves the computational time of the optimal policy, which we discuss in detail in Section 4.5.

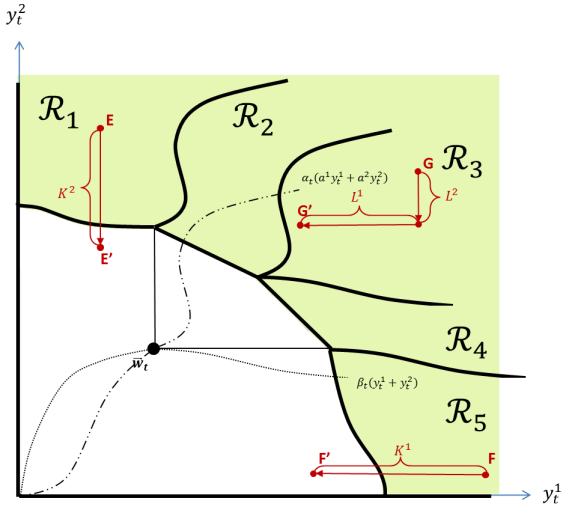
Managerially, a conceptual illustration of the **UB** policy for the case where the *demand/capacity ratio is medium* ($a^1 < D_t/R_t < a^2$) is depicted in Figure 3 where, for any initial inventory levels (e.g., B), the red arrow shows where the optimal post-production inventory levels should be (e.g., B'). Specifically, the state space of inventory levels of both types of materials is partitioned into six regions (see Figure 3(a)), where $\mathscr{R}_0$ is the use-down-to region whereas $\mathscr{R}_1$ to $\mathscr{R}_5$ are the balancing regions. We first consider the optimal production decision in the use-down-to region $\mathscr{R}_0$ (see



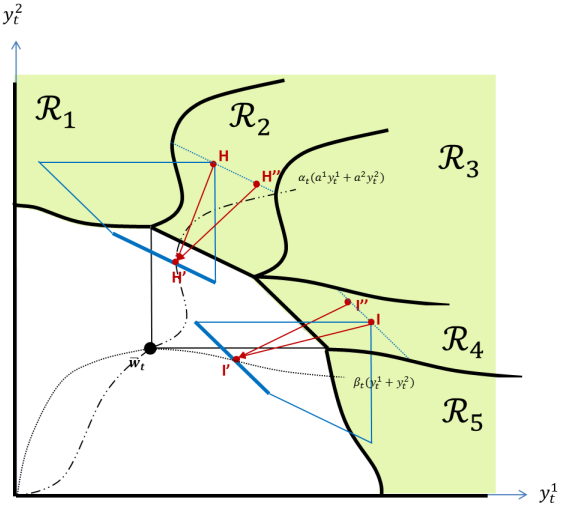
(a) Partition of the state space under a **UB** policy.



(b) **UB** policy in the use-down-to region $\mathscr{R}_0$.



(c) **UB** policy in the balancing regions $\mathscr{R}_1, \mathscr{R}_3, \mathscr{R}_5$.



(d) **UB** policy in the balancing regions $\mathscr{R}_2, \mathscr{R}_4$.

Figure 3 **Illustration of a UB policy. A UB policy is fully determined by the switching curves γ_t^1 (black dashed curve), γ_t^2 (black dash-dot curve), α_t (black dash-double-dot curve), β_t (black dotted curve). Figure 3(a) illustrates the state space partition (bold black solid curves). The arrows in Figures 3(b)-3(d) show the optimal production decision in the use-down-to region and the balancing regions. The area within the blue solid lines in Figure 3(d) is the feasible region under the corresponding state. The blue dotted line segment in Figure 3(d) indicates the set of states that have the same optimal ending inventory.**

Figure 3(b)). Starting at any starting inventory levels $\mathbf{y}_t = (y_t^1, y_t^2)$ in $\mathscr{R}_0$ (e.g., A, B, C, D) , the use-down-to levels for type-j material is $\hat{\gamma}_t^j(y_t^{3-j})$ (recall that $\hat{\gamma}_t^j$ is defined in the Usage Decision Rule of Definition 1 and is independent of D_t and R_t) and the ending inventory should be $y_t^j \wedge \hat{\gamma}_t^j(y_t^{3-j})$ (e.g., A', B', C', D'). (For example, when the starting inventory levels is C, then its first coordinate

is larger than $\hat{\gamma}_t^1(y_t^2)$ and its second coordinate is smaller than $\hat{\gamma}_t^2(y_t^1)$, so the firm should only use type-1 material and bring the ending inventory to C'.) The intuition behind the optimality of use-down-to policy in this region is as follows. In the use-down-to region $\mathscr{R}_0$, both types of materials have low inventory levels. If the current period price for the end product is relatively low, it is more cost effective to hold the inventory for future periods' use than to fully satisfy the demand in current period. Therefore, neither the capacity constraint (2) nor the demand constraint (1) is binding, and the optimal production policy becomes a state-dependent use-down-to policy with use-down-to levels $\hat{\gamma}_t^j(y_t^{3-j})$ for $j = 1, 2$. (At the other extreme, if the current period price of the end product is fairly high, then $\bar{\mathbf{w}}_t$ will be located at origin. This means that the use-down-to region completely vanishes and it is not optimal to hold back inventory.)

We now consider the optimal production decision in the balancing regions $\mathscr{R}_1 - \mathscr{R}_5$ (see Figures 3(c) and 3(d)). Note that, in the balancing regions, at least one type of material has a lot of inventory. Thus, inventory holding cost becomes a major cost driver and it is more cost effective to use the material with a lot of inventory to fully satisfy the demand and to reduce the holding cost. However, since the total volume that can be produced is limited by production capacity, the firm may not be able to fully satisfy the current period demand. Therefore, either the capacity constraint or the demand constraint is binding in the balancing regions. Consider regions $\mathscr{R}_1$ and $\mathscr{R}_2$. Given starting inventory levels $\mathbf{y}_t = (y_t^1, y_t^2)$ in $\mathscr{R}_1$ (e.g., E in Figure 3(c)), it is optimal to use as much type-2 material as possible and bring the ending inventory to $(y_t^1, y_t^2 - K^2)$ (e.g., E'). Given starting inventory levels $\mathbf{y}_t = (y_t^1, y_t^2)$ in $\mathscr{R}_2$ (e.g., H in Figure 3(d)), it is optimal to bring the ending inventory to $(\alpha_t^1(a^1 y_t^1 + a^2 y_t^2 - D_t), \alpha_t^2(a^1 y_t^1 + a^2 y_t^2 - D_t))$ (e.g., H') which coincides with the intersection of the curve α_t (the black dash-double-dot curve) and the portion of the boundary of $\mathscr{E}_2(H, D_t, R_t)$ where the demand constraint is binding (the bold blue line segment portion on the boundary of the feasible region of H in Figure 3(d)). (Note that H' is independent of R_t, and it depends on $D_t, \mathbf{y}_t$ only via their linear combination $a^1 y_t^1 + a^2 y_t^2 - D_t$. This means that for any point H'' in $\mathscr{R}_2$ that is on the line segment with a slope of $-\frac{a^1}{a^2}$ that goes through H (the blue dotted line), the ending inventory levels are the same as that of H, i.e., H'.) Why is this usage decision rule optimal? In regions $\mathscr{R}_1$ and $\mathscr{R}_2$, type-2 material is "in excess of" type-1 material, so the firm should use more of type-2 material to reduce its inventory cost. A higher usage of this high conversion rate material allows the firm to fully satisfy demand before hitting the capacity constraint. Hence, in both regions $\mathscr{R}_1$ and $\mathscr{R}_2$, the demand constraint (1) is always binding at optimum. This reduces the optimal production problem into a single dimensional optimization problem. Specifically, in region $\mathscr{R}_2$ where type-2 material is only "*moderately* in excess of" type-1 material, an optimal

usage quantity equals $\alpha_t(a^1y_t^1+a^2y_t^2-D_t) \in \arg\min_{\mathbf{w}_t\in\mathbb{R}^2_+}\{J_t(\mathbf{w}_t): a^1w_t^1+a^2w_t^2=a^1y_t^1+a^2y_t^2-D_t\}$ (note that, based on Definition 1, the band-shaped region of $\mathscr{R}_2$ is defined by appropriately shifting $\alpha_t(.)$ which ensures that under any state in $\mathscr{R}_2$, $\alpha_t(a^1y_t^1+a^2y_t^2-D_t)$ is feasible to **OPT2** as Figure 3(d) shows); whereas in region $\mathscr{R}_1$ where type-2 material is "*critically* in excess of" type-1 material, the overstocking cost of type-2 material is a major cost driver and the firm should only use type-2 material and as much as possible. The intuition of $\mathscr{R}_4$ and $\mathscr{R}_5$ are similar to $\mathscr{R}_1$ and $\mathscr{R}_2$.

Finally, given starting inventory levels $\mathbf{y}_t=(y_t^1,y_t^2)$ in $\mathscr{R}_3$ (e.g., G in Figure 3(c)), it is optimal to use a fixed quantity L^j (defined earlier) of each type-j material and bring the ending inventory to $(y_t^1-L^1, y_t^2-L^2)$ (e.g., G'). In region $\mathscr{R}_3$, the inventories of both types of materials are balanced. Thus, the usage quantities of both types of materials are also well-balanced. This results in demand being fully satisfied at production capacity limit and the resulting inventory levels are precisely the intersection of the demand constraint and the capacity constraint.

4.3.1. Technical Challenge of the Analysis and Our Approach Although the multivariate optimization problem **OPT2** is a convex optimization with linear constraints, its objective J_t is not necessarily differentiable, so it cannot be easily analyzed using the traditional derivative-based approach. To address this technical challenge, we develop a novel *approximate optimization approach* (to be explained later) to analyze **OPT2** and show that a **UB** policy is optimal. We provide a road-map of our approach below.

Step 1: Approximate Optimization and Its Solution. We first introduce an approximation of **OPT2** which is parameterized by $n\in\mathbb{Z}_{++}$ and characterize the structure of its optimal solution:

$$\mathbf{OPT2}^{(n)}(\mathbf{y}_t,D_t,R_t) \quad U_t^{(n)}(\mathbf{y}_t,D_t,R_t)=\min_{\mathbf{w}_t\in\mathscr{E}_2(\mathbf{y}_t,D_t,R_t)} J_t^{(n)}(w_t^1,w_t^2)$$

where $J_t^{(n)}(w_t^1,w_t^2):=J_t(w_t^1,w_t^2)+\frac{1}{n}[(w_t^1)^2+(w_t^2)^2]$. Note that $\mathbf{OPT2}^{(n)}$ is a strictly convex program and has a unique optimal solution (it is always feasible since $\mathbf{y}_t$ is a feasible point). While $\mathbf{OPT2}^{(n)}$ is still not necessarily differentiable and hard to analyze using the conventional derivative-based approach, the uniqueness of the optimal solution to $\mathbf{OPT2}^{(n)}$ for all states (due to strict convexity) allows us to *progressively* partition the state space into several regions, and reduce $\mathbf{OPT2}^{(n)}$ to a single-variable optimization problem in each region by identifying at least one binding constraint. This idea allows us to prove the structure of the optimal solution to $\mathbf{OPT2}^{(n)}$ in Proposition 1.

PROPOSITION 1. *An optimal solution of* $\mathbf{OPT2}^{(n)}$ *is defined by a* **UB** *policy which is characterized by* $\gamma_t^{(n)1}(.),\gamma_t^{(n)2}(.)$, $\alpha_t^{(n)}(.),\beta_t^{(n)}(.)$ *that intersect at* $\bar{\mathbf{w}}_t^{(n)}$, *where*

$$\bar{\mathbf{w}}_t^{(n)}:=\arg\min_{\mathbf{w}_t\in\mathbb{R}^2_+} J_t^{(n)}(\mathbf{w}_t);$$

$$\gamma_t^{(n)j}(y_t^{3-j}) := \arg\min_{w_t^j \in \mathbb{R}_+} J_t^{(n)}(w_t^j, y_t^{3-j}), \forall j = 1,2, \forall y_t^{3-j} \in \mathbb{R}_+;$$
$$\alpha_t^{(n)}(s) := \arg\min_{\mathbf{w}_t \in \mathbb{R}_+^2} \{J_t^{(n)}(\mathbf{w}_t) : a^1 w_t^1 + a^2 w_t^2 = s\}, \forall s \in \mathbb{R}_+;$$
$$\beta_t^{(n)}(s) := \arg\min_{\mathbf{w}_t \in \mathbb{R}_+^2} \{J_t^{(n)}(\mathbf{w}_t) : w_t^1 + w_t^2 = s\}, \forall s \in \mathbb{R}_+.$$

Proposition 1 shows that a **UB** policy with properly defined switching curves is optimal to **OPT2**$^{(n)}$. Since the approximate optimizations **OPT2**$^{(n)}$ "converge" to **OPT2** as n tends to infinity, one would expect the optimal production policy also "converges", and a **UB** policy is optimal to **OPT2**. We do this in the next step.

Step 2: Convergence of the Optimal Production Policy. The idea in this step is to construct a **UB** policy whose switching curves are the "limit" of the switching curves that define the optimal policies of **OPT2**$^{(n)}$; we then show this constructed **UB** policy is optimal to **OPT2**. Specifically, using the four functions $\alpha_t^{(n)}, \beta_t^{(n)}, \gamma_t^{(n)1}, \gamma_t^{(n)2}$ defined earlier, we now construct four *limiting functions*, $\alpha_t : \mathbb{R}_+ \to \mathbb{R}_+^2, \beta_t : \mathbb{R}_+ \to \mathbb{R}_+^2, \gamma_t^1 : \mathbb{R}_+ \to \mathbb{R}_+, \gamma_t^2 : \mathbb{R}_+ \to \mathbb{R}_+$, which we will use to characterize an optimal solution of **OPT2**. Since the construction involves showing the convergence of functions, we construct the limiting functions in two steps: we first define these functions on dense subsets of their respective domains, and then expand the definition to their whole domains.

Formally, we first show (by Cantor's diagonal argument) that there exists a subsequence $\{n_k \in \mathbb{Z}_{++}\}_{k \in \mathbb{Z}_{++}}$ of $\{1,2,3,\dots\}$ such that the following are well-defined: $\bar{\mathbf{w}}_t = (\bar{w}_t^1, \bar{w}_t^2) := \lim_{k\to\infty} \bar{\mathbf{w}}_t^{(n_k)}$, and for all $j = 1,2$ and $s \in \mathbb{Q}_+$

$$\gamma_t^j(s) := \begin{cases} \lim_k \gamma_t^{(n_k)j}(s), & s \neq \bar{w}_t^{3-j} \\ \bar{w}_t^j, & s = \bar{w}_t^{3-j} \end{cases}, \alpha_t(s) := \begin{cases} \lim_k \alpha_t^{(n_k)}(s), & s \neq a^1\bar{w}_t^1 + a^2\bar{w}_t^2 \\ \bar{\mathbf{w}}_t, & s = a^1\bar{w}_t^1 + a^2\bar{w}_t^2 \end{cases}$$
$$\beta_t(s) := \begin{cases} \lim_k \beta_t^{(n_k)}(s), & s \neq \bar{w}_t^1 + \bar{w}_t^2 \\ \bar{\mathbf{w}}_t, & s = \bar{w}_t^1 + \bar{w}_t^2 \end{cases}.$$

Then we show, for any $x \in \mathbb{R}_+$, there exist sequences $\{s_m^\alpha(x) \in \mathbb{Q}_+\}_{m\in\mathbb{Z}_{++}}$, $\{s_m^\beta(x) \in \mathbb{Q}_+\}_{m\in\mathbb{Z}_{++}}$, $\{s_m^{\gamma,j}(x) \in \mathbb{Q}_+\}_{m\in\mathbb{Z}_{++}}$ that all converge to x, such that the following are well-defined: for all $j = 1,2$

$$\gamma_t^j(x) := \begin{cases} \lim_m \gamma_t^j(s_m^{\gamma,j}(x)), & s \neq \bar{w}_t^{3-j} \\ \bar{w}_t^j, & x = \bar{w}_t^{3-j} \end{cases}, \alpha_t(x) := \begin{cases} \lim_m \alpha_t(s_m^\alpha(x)), & x \neq a^1\bar{w}_t^1 + a^2\bar{w}_t^2 \\ \bar{\mathbf{w}}_t, & x = a^1\bar{w}_t^1 + a^2\bar{w}_t^2 \end{cases}$$
$$\beta_t(x) := \begin{cases} \lim_m \beta_t(s_m^\beta(x)), & x \neq \bar{w}_t^1 + \bar{w}_t^2 \\ \bar{\mathbf{w}}_t, & x = \bar{w}_t^1 + \bar{w}_t^2 \end{cases}.$$

Using $\bar{\mathbf{w}}_t, \gamma_t^1, \gamma_t^2, \alpha_t, \beta_t$, we can construct a **UB** policy by Definition 1. The following holds.

PROPOSITION 2. *$\{\mathscr{R}_j\}_{j=0}^5$ forms a partition of $\mathbb{R}_+^2$ and there exists an optimal solution to* **OPT2***, which we denote by $\mathbf{w}_t^*$, such that:*

a. If $\mathbf{y}_t \in \mathscr{R}_0$, $\mathbf{w}_t^* = (y_t^1 \wedge \hat{\gamma}_t^1(y_t^2), y_t^2 \wedge \hat{\gamma}_t^2(y_t^1))$;

b. Otherwise, i.e., when $\mathbf{y}_t \in \mathbb{R}^2_+ - \mathscr{R}_0$*, the following hold: (b1) When* $D_t \le a^1 R_t$*, (1) is binding at* $\mathbf{w}^*_t$*; (b2) When* $a^1 R_t < D_t < a^2 R_t$*, (1) is binding if* $\mathbf{y}_t \in \mathscr{R}_1 \sqcup \mathscr{R}_2$*, (2) is binding if* $\mathbf{y}_t \in \mathscr{R}_4 \sqcup \mathscr{R}_5$*, both (1) and (2) are binding if* $\mathbf{y}_t \in \mathscr{R}_3$*; (b3) When* $D_t \ge a^2 R_t$*, (2) is binding at* $\mathbf{w}^*_t$*.*

From Proposition 2, it is easy to prove that the constructed **UB** policy is an optimal production policy: Part (a) establishes the optimality for states in the use-down-to region; Part (b) identifies at least one binding constraint in **OPT2** for states in balancing regions, which reduces **OPT2** to a single-variate convex optimization problem whose optimal solution structure is well-understood and corresponds to the decision rules in the balancing regions of **UB** in Definition 1.

4.4. Optimal Ordering Policy

We now characterize an optimal ordering policy which, together with the optimal production policy characterized in Section 4.3, completes the characterization of a jointly optimal production and ordering policy of our problem. The following theorem characterizes an optimal ordering policy given that the firm makes optimal production decisions.

THEOREM 2. *For any* t*, there exist two nonincreasing functions* $\zeta^j_t(.) : \mathbb{R}_+ \to \mathbb{R}_+$ *for* $j = 1, 2$ *that intersect at a point* $\bar{\mathbf{y}}_t \in \mathbb{R}^2_+$*, where* $\bar{\mathbf{y}}_t \in \arg\min_{\mathbf{y}_t \in \mathbb{R}^2_+} \bar{G}_t(\mathbf{y}_t)$ *and*

$$\zeta^j_t(x^{3-j}_t) = \begin{cases} \bar{\zeta}^j_t(x^{3-j}_t),\ \forall x^{3-j}_t > \bar{y}^{(n)(3-j)}_t \\ \hat{\zeta}^j_t(x^{3-j}_t),\ \forall x^{3-j}_t \le \bar{y}^{(n)(3-j)}_t \end{cases}$$

with $\bar{\zeta}^j_t(x^{3-j}_t) \in \arg\min_{x^j_t \in \mathbb{R}_+} \bar{G}_t(\mathbf{x}_t)$ *and* $\hat{\zeta}^j_t(x^{3-j}_t) \in \arg\min_{y^j_t \in \mathbb{R}_+} \{\min_{y^{3-j}_t \in \mathbb{R}_+} \mathbb{E}_{q^{3-j}_t}[\bar{G}_t(y^j_t, y^{3-j}_t \wedge (x^{3-j}_t + q^{3-j}_t))]\}$*, such that an optimal ordering target* $\mathbf{y}^*_t(\mathbf{x}_t)$ *satisfies the following:*

a. If $x^j_t \ge \zeta^j_t(x^{3-j}_t)$ *for some* $j = 1, 2$*, then* $\mathbf{y}^*_t(\mathbf{x}_t) = (x^1_t \vee \zeta^1_t(x^2_t), x^2_t \vee \zeta^2_t(x^1_t))$*;*

b. If $x^j_t < \zeta^j_t(x^{3-j}_t)$ *for both* $j = 1, 2$*, then* $\mathbf{y}^*_t(\mathbf{x}_t) \in \arg\min_{\mathbf{y}_t \in \mathbb{R}^2_+} G_t(\mathbf{y}_t, \mathbf{x}_t)$*.*

Putting Theorems 1 and 2 together, we have fully characterized the structure of an optimal policy. The proof of Theorem 2 follows a similar road-map as in Theorem 1, with a properly defined sequence of approximate optimizations of **OPT1**; but since **OPT1** is not a convex optimization, the analysis of the approximate optimizations is very different from Step 1 in Section 4.3.1. For expositional clarity, we defer the full proof to the Online Appendix. Next, we explain the optimal ordering policy and its intuition in more detail.

Consider first the case when the initial inventory levels satisfy $x^j_t \ge \zeta^j_t(x^{3-j}_t)$ for some $j = 1, 2$. (An optimal ordering policy in this case is illustrated in the shaded region of Figure 4.) By Theorem 2 part (a), ζ^j_t can be viewed as type-j material's order-up-to levels and the optimal ending inventory should be $x^j_t \vee \zeta^j_t(x^{3-j}_t)$. (For example, when the starting inventory levels is B, then its first coordinate is larger than $\zeta^1_t(x^2_t)$ and its second coordinate is smaller than

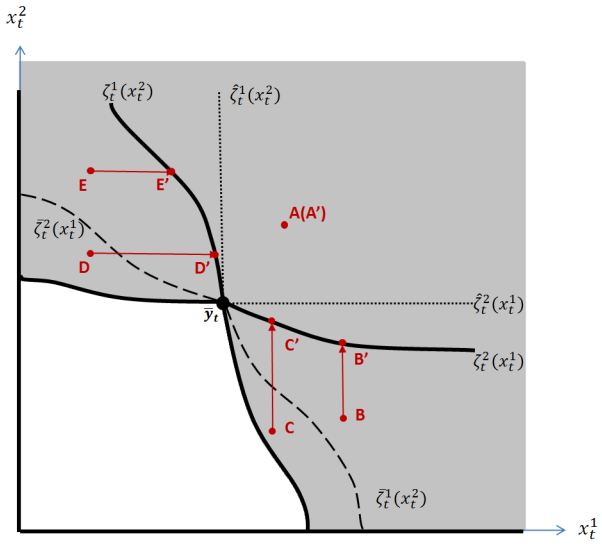


Figure 4 **Illustration of an optimal ordering policy for the states such that either $x_t^1 \geq \zeta_t^1(x_t^2)$ or $x_t^2 \geq \zeta_t^2(x_t^1)$ (shaded region). The arrows show the optimal ordering decisions. The solid curves correspond to the functions ζ_t^j ($j=1,2$) which are piecewise defined by the dashed curves $\bar{\zeta}_t^j$ and dotted curves $\hat{\zeta}_t^j$.**

$\zeta_t^2(x_t^1)$, so the firm should only order type-2 material and bring the ending inventory to B'.) Note that the order-up-to levels of type-j material *only* depend on the initial inventory of the other material but is independent of its own inventory. The intuition is as follows. When the firm has enough type-j inventory x_t^j (i.e., $x_t^j \geq \zeta_t^j(x_t^{3-j})$ for some $j=1,2$), it should not order additional type-j material, so the optimal ending inventory for type-j material is x_t^j. (This is as if the firm orders up to $\zeta_t^j(x_t^{3-j})$ since $x_t^j \geq \zeta_t^j(x_t^{3-j})$.) Given that, the optimization reduces to $\min_{y_t^{3-j} \geq x_t^{3-j}} G_t(x_t^j, y_t^{3-j}, \mathbf{x}_t) = \min_{y_t^{3-j} \geq x_t^{3-j}} \mathbb{E}_{q_t^{3-j}}[\bar{G}_t(x_t^j, y_t^{3-j} \wedge (x_t^{3-j} + q_t^{3-j})]$. The objective is unimodal in y_t^{3-j} and the firm should aim at an order-up-to level of $\zeta_t^{3-j}(x_t^j)$ (independent of x_t^{3-j}) as the ending inventory of type-$(3-j)$ material whenever possible.

We now consider the second case when the initial inventory levels satisfy $x_t^j < \zeta_t^j(x_t^{3-j})$ for both $j=1,2$. In this case, the structure of the optimal decision does not have the simple order-up-to structure as the case above. However, Theorem 2 part (b) states that in this case, instead of optimizing G_t over a constraint set $\mathscr{E}_1(\mathbf{x}_t, \infty)$, one can simply solve for an unconstrained optimizer of G_t which automatically satisfies the constraints. The intuition is that in this case, since the initial inventory for both types of materials is already very low (i.e., $x_t^j < \zeta_t^j(x_t^{3-j})$ for both $j=1,2$), the firm is not strictly better-off by having less inventory of either type of material. Hence, the constraint that the ending inventory cannot be lower than the initial inventory does not affect the optimal cost the firm can attain, and ignoring the constraints can still lead to an optimal solution of the constrained problem.

4.5. Proposed heuristic policy

Note that the MDP we have analyzed above has a continuum of states (i.e., while there are finite number of possible states for demand and production capacity, the inventory levels of both types of raw materials are continuous) which results in value functions with continuous variables; thus, the optimal policy is not directly amenable to computational methods. To address this computational challenge, we propose a heuristic policy which uses value function approximations to make decisions. Before explaining in more detail how to calculate approximations of the value functions, we first lay out the heuristic below.

Proposed heuristic. Input parameter: $\mathbf{x}_1$

Step 1: Offline value function approximation. Call an algorithm which returns approximations of the value functions of the production stage, $\tilde{U}_t$, and the value functions of ordering stage, $\tilde{V}_t$.

Step 2: Online policy implementation. For $t = 1, \ldots, T$, **do**:

a. Replace U_t by $\tilde{U}_t$ in $\mathbf{OPT1}_t(\mathbf{x}_t)$ and solve for the solution $\mathbf{y}_t^*$. Order $\mathbf{o}_t = \mathbf{y}_t^* - \mathbf{x}_t$.

b. Observe D_t, R_t, and $\mathbf{q}_t$. Set $\mathbf{y}_t = \mathbf{x}_t + \mathbf{o}_t \wedge \mathbf{q}_t$.

c. Replace V_{t+1} by $\tilde{V}_{t+1}$ in $\mathbf{OPT2}_t(\mathbf{y}_t, D_t, R_t)$ and solve for the solution $\mathbf{w}_t^*$. Use $\mathbf{r}_t = \mathbf{y}_t - \mathbf{w}_t^*$ for production.

d. Set $\mathbf{x}_{t+1} = \mathbf{w}_t^*$.

Next, we explain how we approximate the value function in Step 1 and the algorithms we develop to operationalize this value function approximation. Recall that the value functions U_t and V_t are both jointly convex in their continuous variables. This motivates us to use *state-space discretization* and *local linear interpolation* to generate piece-wise linear convex approximations of the value function via back-ward induction. In a nutshell, in each step of the backward induction, when calculating the value function, only a *selected* finite number of states are calculated via optimization while the value of the remaining states are linearly interpolated based on the value of the selected states. Mathematically,

- *State space discretization.* Let $\delta \in \mathbb{R}_+$ denote the discretization step size for inventory levels and let $\bar{\kappa}\delta$ denote the maximum possible inventory level[4]. Let $\mathcal{I}(\bar{\kappa}) := \{0, \delta, 2\delta, \ldots, \bar{\kappa}\delta\}$. Then $\mathcal{I}(\bar{\kappa})^2 \subset \mathbb{R}_+^2$ corresponds to a grid of discrete states in the inventory space; the total number of states is $\bar{\kappa}^2 \kappa_D \kappa_R$ where $\kappa_D := |\mathcal{D}|$ and $\kappa_R := |\mathcal{R}|$.
- *Local linear interpolation.* Note that we need to linearly interpolate the approximation of the value function of the production stage (resp. ordering stage) $\tilde{U}_t(., D_t, R_t) : \mathcal{I}(\bar{\kappa})^2 \to \mathbb{R}_+^2$ (resp. $\tilde{V}_t(.) : \mathcal{I}(\bar{\kappa})^2 \to \mathbb{R}_+^2$) to a new function $\tilde{U}_t : [0, \bar{\kappa}\delta]^2 \to \mathbb{R}_+^2$ (resp. $\tilde{V}_t(.) : [0, \bar{\kappa}\delta]^2 \to \mathbb{R}_+^2$). We propose the following local linear interpolation procedure for $\tilde{U}(., D_t, R_t)$. (We use the same approach

[4] Typically, κ is chosen to be a large number so that it will not affect the optimal policy. In the motivating coal-fired power plant example, purchased coal are dumped into large open space called the coal yard, and the power plant we have worked with does not have a binding storage capacity for coal.

for $\tilde{V}_t$.) Fix any D_t, R_t. Then for any $\mathbf{y} = (y^1, y^2) \in [0, \bar{\kappa}\delta]^2$, there exist k^1, k^2 such that $\mathbf{y}$ lies in the square $[k^1\delta, (k^1+1)\delta] \otimes [k^2\delta, (k^2+1)\delta]$. If $\mathbf{y}_t$ is in the lower-left corner of the square (i.e., $y^1 + y^2 \leq (k^1 + k^2 + 1)\delta$), then there exists $\lambda, \mu \in [0,1]$ such that $\lambda + \mu \leq 1$ and $\mathbf{y} = \lambda(k^1\delta, k^2\delta) + \mu((k^1+1)\delta, k^2\delta) + (1-\lambda-\mu)(k^1\delta, (k^2+1)\delta)$; so we let $\tilde{U}_t(\mathbf{y}, D_t, R_t) := \lambda\tilde{U}_t(k^1\delta, k^2\delta, D_t, R_t) + \mu\tilde{U}_t((k^1+1)\delta, k^2\delta, D_t, R_t) + (1-\lambda-\mu)\tilde{U}_t(k^1\delta, (k^2+1)\delta, D_t, R_t)$. Interpolate similarly if $\mathbf{y}_t$ is in the upper-right corner of the square.

A natural algorithm to calculate value function approximation explained above is as follows.

Algorithm 1. Input parameters: $\delta, \bar{\kappa}$

Set $\tilde{V}_{T+1}(\mathbf{x}_{T+1}) = c^1_{T+1}x^1_{T+1} + c^2_{T+1}x^2_{T+1}$ for all $\mathbf{x}_{T+1} \in \mathcal{I}(\bar{\kappa})^2$. Set $t = T$.

While $t \geq 1$, **do**:

a. **Production stage.** For any $D_t \in \mathcal{D}, R_t \in \mathcal{R}$:

a1. For all $\mathbf{y}_t \in \mathcal{I}(\bar{\kappa})^2$, replace V_{t+1} by $\tilde{V}_{t+1}$ in **OPT2**$(\mathbf{y}_t, D_t, R_t)$ and solve for the optimal value, $\tilde{U}_t$.

a2. Use local linear interpolation to interpolate $\tilde{U}_t(\mathbf{y}_t, D_t, R_t)$ for $\mathbf{y}_t \in [0, \bar{\kappa}\delta]^2 - \mathcal{I}(\bar{\kappa})$.[5]

b. **Ordering stage.**

b1. For any $\mathbf{x}_t \in \mathcal{I}(\bar{\kappa})^2$, replace U_t by $\tilde{U}_t$ in **OPT1**$(\mathbf{x}_t)$ and solve for the optimal value $\tilde{V}_t$.

b2. Use local linear interpolation to interpolate $\tilde{V}_t(\mathbf{x}_t)$ for $\mathbf{x}_t \in [0, \bar{\kappa}\delta]^2 - \mathcal{I}(\bar{\kappa})$.

c. Reduce t by 1.

End

While it can be shown that the approximate value function calculated via Algorithm 1 converges to the true value function as $\delta \to 0$, it is worth mentioning that Algorithm 1 requires solving a total of $T(\bar{\kappa}^2\kappa_D\kappa_R + \bar{\kappa}^2)$ *constrained* optimizations with *two decision variables*, which is computationally time-consuming when $\bar{\kappa}, \kappa_D, \kappa_R$ are large. This is also known as the curse of dimensionality when solving stochastic dynamic program with multiple dimensional state spaces. Fortunately, thanks to the characterization of the structure of the optimal policy in Sections 4.3 and 4.4, it is possible to reduce the computational complexity by the following Algorithm 2 which only requires $\Theta(T\bar{\kappa})$ *single-variate unconstrained convex optimizations* and fewer than $T\bar{\kappa}^2$ constrained optimizations with two decision variables. Note that the total number of optimizations required in Algorithm 2 is independent of κ_D and κ_R because we leverage the structural property of the optimal policy; thus, Algorithm 2 can significantly improve the computational time of Algorithm 1 when the number of potential demand scenarios (κ_D) and production capacity scenarios (κ_R) are large. In Section 5.3, we use a synthetic data set to demonstrate the empirical computational benefit of Algorithm 2 over Algorithm 1.

[5] Note that all steps related to interpolation are only implicitly conducted when the interpolated function is used in the subsequent optimizations.

Algorithm 2. Input parameters: $\delta, \bar{\kappa}$

Set $\tilde{V}_{T+1}(\mathbf{x}_{T+1}) = c^1_{T+1}x^1_{T+1} + c^2_{T+1}x^2_{T+1}$ for all $\mathbf{x}_{T+1} \in \mathcal{I}(\bar{\kappa})^2$. Set $t = T$.

While $t \geq 1$, **do**:

a. **Production stage.**

a1. Replace V_{t+1} in J_t by $\tilde{V}_{t+1}$, and denote the new function as $\tilde{J}_t$.

a2. Compute $\bar{\mathbf{w}}_t \in \arg\min_{\mathbf{w}_t \in \mathbb{R}^2_+} \tilde{J}_t(\mathbf{w}_t)$ and the following:

$$\begin{aligned}
\gamma^j_t(y^{3-j}_t) \in \arg\min_{w^j_t \in \mathbb{R}_+} \tilde{J}_t(w^j_t, y^{3-j}_t), \forall y^{3-j}_t \in \mathcal{I}(\bar{\kappa}) \text{ and } y^{3-j}_t \geq \bar{w}^{3-j}_t, j = 1, 2,\\
\alpha_t(x) \in \arg\min_{\mathbf{w}_t \in \mathbb{R}^2_+} \{\tilde{J}_t(\mathbf{w}_t) : a^1 w^1_t + a^2 w^2_t = x\}, \forall x \in \mathcal{I}(\lceil (a^1 + a^2)\bar{\kappa} \rceil) \text{ and } x \geq a^1 \bar{w}^1_t + a^2 \bar{w}^2_t,\\
\beta_t(x) \in \arg\min_{\mathbf{w}_t \in \mathbb{R}^2_+} \{\tilde{J}_t(\mathbf{w}_t) : w^1_t + w^2_t = x\}, \forall x \in \mathcal{I}(\lceil 2\bar{\kappa} \rceil) \text{ and } x \geq \bar{w}^1_t + \bar{w}^2_t.
\end{aligned}$$

a3. Linearly interpolate the switching curves $\gamma^1_t, \gamma^2_t, \alpha_t, \beta_t$ in their respective domains.

a4. For all D_t, R_t, and $\mathbf{y}_t \in \mathcal{I}(\bar{\kappa})^2$, compute $\tilde{U}_t(\mathbf{y}_t, D_t, R_t)$ according to the production decision in Definition 1.

a5. For all D_t, R_t, use local linear interpolation to interpolate $U_t(., D_t, R_t)$ for the remaining states.

b. **Ordering stage.**

b1. Replace U_{t+1} in $\bar{G}_t$ by $\tilde{V}_{t+1}$, and denote the new function as $\tilde{\bar{G}}_t$.

b2. Compute the following:

$$\bar{\zeta}^j_t(x^{3-j}_t) \in \arg\min_{y^j_t \in \mathbb{R}_+} \tilde{\bar{G}}_t(y^j_t, x^{3-j}_t), \forall x^{3-j}_t \in \mathcal{I}(\bar{\kappa}), j = 1, 2$$

b3. Linearly interpolate the curves $\bar{\zeta}^1_t$ and $\bar{\zeta}^2_t$ in their respective domains.

b4. For all $\mathbf{x}_t \in \mathcal{I}(\bar{\kappa})^2$, compute $\tilde{V}_t(\mathbf{x}_t)$ according to the following optimal ordering decisions: If $x^{3-j}_t \geq \bar{\zeta}^{3-j}_t(x^j_t)$ for some $j = 1, 2$, the ordering decision is $(x^1_t \vee \bar{\zeta}^1_t(x^2_t), x^2_t \vee \bar{\zeta}^2_t(x^1_t))$; otherwise, replace U_t by $\tilde{U}_t$ in $\mathbf{OPT1}(\mathbf{x}_t)$ and solve for the optimal ordering decisions.

b5. Use local linear interpolation to interpolate $U_t(., D_t, R_t)$ for the remaining states.

c. Reduce t by 1.

End

5. Numerical studies

The characterization of the structure of the optimal policy allows us to develop a heuristic policy in the previous section. To assess the practical benefit of the proposed heuristic policy (unless otherwise noted, we use Algorithm 2 for Step 1 of our proposed heuristic), we compare the performance of our proposed heuristic to three straw policies introduced in Section 5.1 on a real-world data set in Section 5.2 and a synthetic data set in Section 5.3.

5.1. Considered Policies

In our numerical study, we consider the following three straw policies:

- *Straw-P heuristic.* This is a policy which *prioritises* the use of the cheaper material. Specifically, for production, the firm always uses the cheaper material (material 1) first, and switches to the higher conversion rate material (material 2) only when material 1 is stocked out; in the ordering stage, the firm ignores the material availability uncertainty and follows the optimal order-up-to policy computed given its production policy.[6]
- *Straw-L heuristic.* This is a policy which uses a myopic *linear program* to determine the production decisions. Specifically, for production, the firm uses the optimal solution of the myopic linear program whose objective only involves the current period production stage costs but ignores the cost-to-go. Given this production policy, in the ordering stage, the firm ignores the material availability uncertainty and follows the optimal order-up-to policy. (Note that Straw-L can be viewed as a slightly modified version of the policy van Mieghem and Rudi (2002) developed, which is optimal in their setting.)
- *Straw-M-h heuristic.* This is a *model predictive control* policy with a rolling planning horizon of h periods. Specifically, at the beginning of each decision stage, the firm formulates a deterministic optimization problem for the next h decision periods where the random variables are replaced by their averages. After solving this optimization, the firm only uses the decision for the current decision stage. This process is repeated in every decision stage.

5.2. Study 1: Real world coal-fired power plant data

In this subsection, we calibrate our model based on data of the coal-fired power plant we worked with. This power plant is connected to the power grid operated by Midcontinent Independent System Operator (MISO). We calibrate our model with three data sources.

D1 Publicly available dataset from U.S. Energy Information Administration (Form 2019M-EIA923). This dataset includes monthly data on fuel receipts and costs, power generation, fuel consumption and stocks at power plant levels.

D2 Publicly available dataset from EnergyOnline® which tracks and publishes the hourly data on wholesale energy price and demand load from all of the major Independent System Operators in the U.S. which includes MISO.

D3 The power plant's proprietary operational data. This dataset includes fuel orders, delivery and consumption at a monthly level, and the plant's coal processing capacity measured in short tons per hour.

Table 1 summarizes the data sources we use to calibrate the model in our paper. Specifically, we consider the weekly planning decisions of coal replenishment and usage for power generation for 13

[6] The power plant we worked with used a simplified version of this heuristic where the order-up-to levels are determined by experience.

weeks from week 27 to week 39 in 2019 (i.e., this corresponds to the 3^{rd} quarter of 2019 from July 2^{nd} to September 30^{th}). The energy price information is obtained from **D2** by averaging the hourly Local Marginal Price (LMP) at the Michigan hub of MISO (where the power plant connects to the power grid) during each week. For energy demand load, we would like to figure out the demand load distribution faced by the plant. **D2** provides the hourly demand load aggregated for the whole MISO network. Due to variability of demand load at different hours of the day, we first use hourly demand load information to fit a normal distribution for the demand load at different hours of the day for each week during the planning horizon (for all different weeks and different hours of the day, the data pass the Kolmogorov-Smirnov normality test). Then, for each week, we aggregate the fitted normal hourly demand distributions into a normal demand distribution for the whole week for the MISO network. Next, we incorporate the actual monthly energy generation of the plant during the planning horizon from **D1** to estimate the fraction of total MISO demand load which the plant is accounted for, and use this fraction to generate a normal weekly demand load distribution faced by the plant. Finally, we use this normal weekly demand distribution to generate a discrete demand load distribution with a support of 20 discrete values which are uniformly spaced between 3 standard deviations above and below the mean of the normal distribution. The heat content of fuel and the fuel cost can be directly obtained from **D1**. Since the majority of the holding cost of coal is the financial cost, we use 6% weighted average cost of capital of the company which owns the power plant for the annual percentage holding cost. The coal availability distribution is calibrated from **D3** which contains information on the amount of coal ordered and the actual amount of coal delivered at a monthly level. When these two quantities are the same, it means that the actual amount of coal that is available at the suppliers is higher than the actual delivery; otherwise, when the order is larger than delivery, it means that the coal availability equals the actual delivery. We assume that the underlying distribution of coal availability is a normal distribution truncated from below at 0, and use maximum likelihood to estimate the parameter of the truncated normal.[7] Finally, the distribution of plant capacity is not easy to estimate since there are many activities and stakeholders involved in the power generation process which we do not have direct access to. Nonetheless, we have both the maximum capacity of the power plant (w.r.t. the amount of coal it can process) and the actual monthly consumption of coal from **D3**.

[7] Note that we observe the coal availability only when the amount of coal delivered is less than ordered. Due to this limitation of data, we have chosen to use a *parametric approach* to estimate the underlying distribution of coal availability. We use normal distribution based on our position that the limited coal availability is driven by independent random disruptions of the coal production processes at the coal mines, e.g., uncertain yield of the coal extracted, random coal dumper outages, etc. As a robustness check, we have also used Beta distribution to fit the coal availability distribution. The performance comparison across different policies under this calibrated model remains quantitatively very similar to the main model, which suggests that our numerical results are robust with respect to the choice of distribution for modeling coal availability.

The variability in monthly consumption of coal we observe come from many sources, one of which is plant capacity. Thus, we compute the fraction of each month's fuel consumption over the maximum monthly consumption across all months in 2019, and use these 12 fractions to generate a discrete distribution of the *fraction* of the actual plant capacity and the maximum plant capacity. We then model the distribution of the plant's capacity by multiplying this discrete random variable *fraction* with the maximum plant capacity.

Table 1 Calibration of the Model

Model Parameter (Notation, unit of measure)	**Data Calibration**
Planning horizon (T, week)	13 weeks
Energy price (p_t, \$/megawatt-hours)	D2
Demand distribution (D_t, megawatt-hours)	D1, D2
Heat content of fuel (a^j, megawatt-hours/short-ton)	D1
Fuel cost (c_t^j, \$/short-ton)	D1
Weekly fuel holding cost as a fraction of fuel cost (h_t^j/c_t^j, n/a)	$6\% \times \frac{1}{52}$
Fuel availability distribution (q_t^j, short-tons)	D3
Plant capacity distribution (R_t, short-tons/week)	D3

To evaluate the profit under different heuristics, we use Monte Carlo simulation with 1000 sample runs. Table 2 reports the simulation results. (For computing the approximate value function in our proposed heuristic, we set $\delta = 15000$ and $\bar{\kappa} = 100$.) It shows that in our calibrated model, our proposed heuristic provides more than 4% profit improvement over all three straw policies. Specifically, compared to Straw-P, the type of policy used in the power plant we worked with, our proposed heuristic provides a quarterly profit improvement of \$1.1 million. Our numerical result strongly suggest that our proposed heuristic policy can provide significant improvement of the current practice.

Table 2 Comparison of quarterly profit of the proposed heuristic versus straw policies

	Proposed	**Straw-P**	**Straw-L**	**Straw-M-1**
Mean	\$19,452,391.60	\$18,327,572.94	\$18,639,264.51	\$17,925,565.35
Standard error	\$30,838.49	\$42,771.94	\$41,676.78	\$31,070.36
Absolute improvement	-	\$1,124,818.65	\$813,127.08	\$1,526,826,24
Relative improvement	-	6.14%	4.36%	8.52%

Note: Straw-M-h heuristic achieves the highest profit when $h = 1$; so only Straw-M-1 is reported.

5.3. Study 2: Synthetic data set

Study 1 provides the practical benefit of our proposed heuristic in the context of the specific coal-fired power plant we worked with. To provide some insights into the robustness of the benefit of our proposed heuristic, we compare the performance of different heuristics on a synthetic data set that consists of 288 problem scenarios with different problem parameters which we explain in more

detail below. Moreover, we use this synthetic data set to provide empirical evidence that when computing approximate value functions in our proposed heuristic, Algorithm 2 provides significant computational benefit compared to Algorithm 1.

For all scenarios in the synthetic data set, we set a planning horizon of $T = 10$ periods . For the conversion rates of the two materials, we fix $a^1 = 1$ and vary a^2 among $1.5, 2, 2.5$. We let the electricity price p_t be cyclic in every three periods and follow the "$p^h - p^l - p^l$" pattern where $p^h = 10$ and $p^l = 4$. We let the purchasing and holding costs be stationary.Specifically, the holding cost is set at 10% of the purchasing cost in each period for both products. We vary the purchasing costs of both materials j to attain different *nominal service levels* (NSL^j) defined as $\mathsf{NSL}^j = \frac{a^j p^h}{a^j p^h + h^j}$. (Note that given NSL^j, a^j, p^h and $h^j = 10\% c^j$, c^j is uniquely determined.) NSL^j can be interpreted as the optimal service level the firm should aim at in high price periods if it only uses type-j material. We vary NSL^j among $93\%, 95\%, 97\%$ to reflect that in high penalty cost periods the firm should aim at a relatively high service level. We use discrete triangular distribution with a support of 10 values to model the demand, production capacity and material availability. In each problem scenario, the demand (resp, production capacity, material availability) is independent and identical across periods. Across different scenarios, we set the mean of demand to be 18 and vary its coefficient of variation ($\mathsf{CVDemand}$) between 0.3 and 0.6; we also vary the mean of production capacity to attain *production utilizations* ($\mathsf{UtilProd}$) of 0.6 and 0.9 ($\mathsf{UtilProd} := \frac{\mathbb{E}(D)}{0.5[a^1\mathbb{E}(R)+a^2\mathbb{E}(R)]}$) and vary its coefficient of variation (CVProd) between 0.3 and 0.6. Finally, we let the material availability distributions to be symmetric across the two materials and vary the proportion of demand to the total product that could be produced by the *average* materials available to be delivered, denoted by ρ, to be between 0.6 and 0.9 ($\rho := \frac{\mathbb{E}(D)}{(a^1\mathbb{E}(q_1)+a^2\mathbb{E}(q_2))}$) and also vary its coefficient of variation (σ) between $0.3, 0.6$. This leads to a combination of 288 different scenarios. Table 3 summarizes our experiment design.

Table 3 Experiment Design. $a^1 = 1$, $(p_1, \ldots, p_{10}) = (10, 4, 4, 10, 4, 4, 10, 4, 4, 10)$, $h_t^j = 10\% c_t^j$ **and** $\mathbb{E}(D) = 18$.

Model Parameter	Values
a^2	1.5, 2, 2.5
$(\mathsf{NSL}^1, \mathsf{NSL}^2)$	(.97, .95), (.97, .93), (.95, .93)
CVDemand	0.3, 0.6
UtilProd	0.6, 0.9
CVProd	0.3, 0.6
ρ	0.6, 0.9
σ	0.3, 0.6

For each of the 288 scenarios, we generate 1000 sample paths and compute the average profit under our proposed heuristic (for computing the approximate value function in our proposed heuristic, we set $\delta = 1$ and $\bar{\kappa} = 40$) and the three straw policies (for Straw-M, we consider rolling horizons

$h = 2, 4, 6, 8$, for each scenario and pick the best performing one), and then use these average profits to compute the relative profit improvement of our proposed heuristic over the straw policies. Table 4 provides the summary statistics of these metrics across all 288 scenarios. Our numerical results show that our proposed heuristic greatly improves the profit over all three straw policies which provides further evidence that its benefit is robust within the range of model parameters in this synthetic data set.

Table 4 Summary statistics of the relative profit improvement of proposed heuristic over straw policies

	Mean	Standard deviation	25-percentile	Median	75-percentile	No. of observations
Straw-P	31.85%	13.57%	21.39%	30.32%	39.19%	288
Straw-L	15.07%	9.33%	7.64%	13.48%	22.14%	288
Straw-M	26.28%	18.52%	12.26%	23.98%	35.49%	288

Note that the approximate value functions in our proposed heuristic can also be calculated using Algorithm 1. However, as Table 5 illustrates, Algorithm 1 can be time-consuming, taking over six hours on average to compute a 10-period problem with $\bar{\kappa} = 40, \kappa_D = \kappa_R = 10$. Thus, for a real application with many more scenarios of demand and production capacity scenarios, Algorithm 1 may not be a feasible solution due to long computational time. In contrast, Algorithm 2 is on average 26.6 times faster, taking on average below 15 minutes to solve the same problems. This suggests that Algorithm 2 can help greatly reduce the computational burden and makes our proposed heuristic more amenable to practical implementation.

Table 5 Comparison of the computational time (in minutes) of the approximate value functions for the first 10 problem scenarios. Prod. (resp. Ordering, Total) stands for the computational time of the production (resp. ordering, both) stages. The programs are coded in Matlab and run in a PC with Intel(R) Core 7-3770 processor and 8 GB RAM.

	Algorithm 1			Algorithm 2		
Scenario	Prod.	Ordering	Total	Prod.	Ordering	Total
1	326.23	30.57	356.79	3.84	9.80	13.64
2	358.86	43.96	402.82	3.74	10.60	14.33
3	303.56	25.47	329.02	3.78	9.19	12.98
4	356.81	39.10	395.90	4.53	12.74	17.27
5	354.97	32.62	387.59	3.77	9.86	13.64
6	357.69	39.69	397.38	3.84	11.35	15.19
7	340.14	25.46	365.59	3.85	8.15	12.00
8	338.24	33.89	372.13	3.76	9.99	13.74
9	326.38	31.62	358.00	4.00	9.94	13.94
10	323.91	36.26	360.16	4.02	10.26	14.28

6. Extensions

In this section, we consider a couple model extensions and show that the structure of the optimal policy generalizes to those contexts as well.

Temporal statistically dependent distributions. So far we have assumed that the data process of the uncertainties in our main model are independent across time; however, they may exhibit serial correlation in practice. In our motivating application, electricity demand may have serial correlation due to, for example, weather conditions. There are different ways to model temporal dependence of the underlying data process, and two approaches have been suggested in the literature (see Löhndorf and Shapiro (2019) and the references therein for mode details): One approach is to model the data process as an autoregressive process and treat the realizations of the data process as state variables; an alternative approach is based on Markov Chain discretization of the Markovian data process using optimal quantization. While the two approaches use different ways to model the temporal dependence of the data process, they both essentially result in a Markov-Modulated data process, a model which has been used in the classic Markov-Modulated Demand (MMD) inventory literature (Song and Zipkin 1993). Following this literature, in this extension, we introduce an exogenous stochastic process $\{S_t \in \mathcal{S}\}_{t=1}^{T}$ where $\mathcal{S}$ is a finite set and S_t represents the state of the business environment in period t which is informative for predicting the demand, capacity, and supply availability in the next period. Specifically, we assume that, conditioning on $S_t = s \in \mathcal{S}$, the joint distribution of $D_{t+1}, R_{t+1}, \mathbf{q}_{t+1}$ is stationary over time. The stochastic process $\{S_t\}$ evolves according to an exogenous discrete-time Markov chain with stationary transition probabilities. We assume that the firm observes S_t along with the realization of demand, capacity, and supply availability for that period, i.e., the firm observes S_t in **E2** of the timeline in the base model. Our next result shows that the optimal policy structure generalizes in this model extension. However, it would be more computationally challenging to solve for the optimal policy because the firm needs to compute switching curves for each state in $\mathcal{S}$.

THEOREM 3. *For the finite horizon problem with Markov-Modulated Demand, an optimal production and inventory policy is one which has the same structure as characterised in Theorems 1 and 2 in the paper for each period t and state $S_t \in \mathcal{S}$.*

Infinite horizon with discounted cost. Our main model features a finite horizon with no discounting which is appropriate for medium-term planning purposes. However, for some applications with much longer planning horizons, the financial cost of capital need to be considered. To capture a much longer planning horizon with financial discounting, consider a generalized model with Markov-Modulated Demand but with a discount factor $\xi < 1$ and an infinite planning horizon

$T = \infty$. Then, the structure of the optimal policy in the original finite horizon problem generalizes to this model as well. Specifically, the following result holds.

THEOREM 4. *For the discounted infinite horizon problem with Markov-Modulated Demand, the optimal production and inventory policy follows the same structure as characterised in Theorems 1 and 2.*

7. Closing Remarks

In this paper, motivated by the coal procurement and production management in coal-fired power plants, we have studied a joint production and replenishment problem in a capacitated make-to-order system where the end product can be made from either of two types of raw materials (or their mixture) which not only vary in their costs and conversion rates but also have uncertain availability. We show that an optimal policy can be jointly characterized by a Use-down-to/Balancing Production Policy and a modified Order-up-to Ordering Policy. Since the state variables include continuous variables, the exact optimal policy is computationally difficult to calculate, we propose a heuristic policy based on piece-wise linear approximations of the optimal value functions. We leverage the structure of the optimal policy to develop an algorithm which improves the computational time of the approximate value functions in our proposed heuristic compared to a brute-force algorithm. To demonstrate the practical benefit of our proposed heuristic, we consider three straw policies which are motivated by practice and academic literature, and compare the performance via numerical studies. We show that in a model calibrated with real-world data of a coal-fired power plant, our proposed heuristic provides significant profit improvement over the straw policies. We further provide evidence of the robustness of the profit improvement of our proposed heuristic in a synthetic data set with a range of different model parameters.

Note that we consider two types of input materials whose quantity are modeled as continuous variables in this paper due to the operational nature of the motivating example. These features may not hold in other applications, and two future research directions naturally arise. One direction is to generalize the model to allow more than two types of input raw materials. In some special cases, one can show that using only two types of input raw materials is optimal; thus, the problem can be reduced to the one studied in this paper; but in general, increasing the number of input raw materials introduces combinatorial complexity into the selection of the type of inputs to order and use, and significantly complicates the structure of the optimal policy. The other direction is to consider a model with discrete inventory quantities. The integrality constraint of this extension breaks down the argument we employ in this paper to characterize the structure of the optimal policy, so we believe a different approach is needed to analyze the discrete setting.

References

Anupindi, R., R. Akella. 1993. Diversification under supply uncertainty. *Management Sci.* **39** 944–963.

Chen, X., X. Gao, Z. Hu. 2015. A new approach to two-location joint inventory and transshipment control via l-natural-convexity. *Operations Research Letters* **43** 65–68.

Ciarallo, F. W., R. Akella, T. E. Morton. 1994. A periodic review, production planning model with uncertain capacity and uncertain demand-optimality of extended myopic policies. *Management Sci.* **40** 320–332.

DeCroix, G. A., A. Arreola-Risa. 1998. Optimal production and ivnentory policy for multiple products under resource constraints. *Management Sci.* **44** 950–961.

Demirel, S., I. Duenyas, R. Kapuscinski. 2015. Production and inventory control for a make-to-stock/calibrate-to-order system with dedicated and shared resources. *Oper. Res.* **63** 823–839.

Economist. 2007. The price of virtue. june 7th, 2007 .

Environmental Protection Agency, (EPA). 2010. *Available and Emerging Technologies for Reducing Greenhouse Gas Emissions From Coal-Fired Electric Generating Units*. Pennyhill Press.

Evans, R. V. 1967. Inventory control of a multiproduct system with a limited production resource. *Naval Research Logistics Quarterly* **14** 173–184.

Evans, Simon, Rosamund Pearce. 2020. (2020, march 26). mapped: The world's coal power plants. *Carbon Brief. Retrieved from https://www.carbonbrief.org/* .

Federgruen, A., N. Yang. 2008. Selecting a portfolio of suppliers under demand and supply risks. *Oper. Res.* **56** 916–936.

Federgruen, A., N. Yang. 2011. Procurement strategies with unreliable suppliers. *Oper. Res.* **59** 1033–1039.

Federgruen, A., N. Yang. 2014. Infinite horizon strategies for replenishment systems with a general pool of suppliers. *Oper. Res.* **62** 141–159.

Glasserman, P. 1996. Allocation production capacity among multiple products. *Oper. Res.* **44** 724–734.

Halman, Nir, Giacomo Nannicini. 2019. Toward breaking the curse of dimensionality: an fptas for stochastic dynamic programs with multidimensional actions and scalar states. *SIAM Journal on Optimization* **29**(2) 1131–1163.

Hu, X., I. Duenyas, R. Kapuscinski. 2008. Optimal joint inventory and transshipment control under uncertain capacity. *Oper. Res.* **56** 881–897.

Jiang, Daniel R, Warren B Powell. 2015. An approximate dynamic programming algorithm for monotone value functions. *Operations Research* **63**(6) 1489–1511.

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

Nadarajah, Selvaprabu, Nicola Secomandi. 2018. Merchant energy trading in a network. *Operations Research* **66**(5) 1304–1320.

Nahmias, S., C. P. Schmidt. 1984. An efficient heuristic for the multi-item newsboy problem with a single constraint. *Naval Research Logistics Quarterly* **31** 463–474.

Secomandi, Nicola. 2008. An analysis of the control-algorithm re-solving issue in inventory and revenue management. *Manufacturing & Service Operations Management* **10**(3) 468–483.

Song, J., P. Zipkin. 2003. Supply chain operations: Assemble-to-order systems. A. G. de Kok, S. C. Graves, eds., *Handbooks in Operations Research and Management Science*, chap. 11. Elsevier, 561–596.

Song, Jing-Sheng, Paul Zipkin. 1993. Inventory control in a fluctuating demand environment. *Operations Research* **41**(2) 351–370.

Swaminathan, J., J. Shanthikumar. 1999. Supplier diversification: effect of discrete demand. *Operations Research Letters* **24** 213–221.

U.S. Department of Energy, the Energy Information Administration (EIA). 2019. Eia-923 monthly generation and fuel consumption time series file. 2019 feburary .

U.S. Energy Information Administration, (EIA). 2019. Monthly energy review march 2019 .

van Mieghem, J. A., N. Rudi. 2002. Newsvendor networks: Inventory management and capacity investment with discretionary activities. *Manufacturing Service Oper. Management* **4** 313–335.