Risk Hedging for Production Planning

Liao Wang (corresponding author), lwang98@hku.hk
Faculty of Business and Economics,
The University of Hong Kong, HKSAR

David D. Yao, yao@columbia.edu
Department of Industrial Engineering and Operations Research,
Columbia University, New York, NY 10027

Abstract

Traditional production planning is primarily a quantity or capacity decision, which must be made at the beginning of a planning horizon before production starts. Adding to this decision a real-time control, a risk-hedging strategy carried out throughout the horizon can better mitigate the risk involved in demand volatility. We demonstrate how this can be done in terms of jointly optimizing the capacity and the hedging decisions, addressing both the mean-variance and the shortfall objectives. Solution techniques, results and insights are highlighted. In particular, we illustrate that our approach readily accommodates data analytics and explicitly quantifies the improvement to the efficient frontier contributed by hedging.

Key words: production risk management, data analytics, hedging strategy.

History: received: January 2019; accepted: August 2019 by J. Shanthikumar after one revision.

1 Introduction

Production planning teaches how to make a production quantity (or *capacity*) decision at time zero ($t = 0$) so as to cope with demand uncertainty at a future time $t = T$, the end of the planning horizon. The decision has to be made at $t = 0$, because only then all other actions and decisions can follow — ordering raw materials and/or sub-assemblies, organizing the work force, setting up production lines, as well as carrying out the physical process that produces the goods. Demand, however, will not realize until all units are produced at $t = T$. Thus, the decision made at $t = 0$ tries to strike a balance between the risk of over-production (units that cannot be sold will incur a loss) and the risk of under-production (unmet demand means reduced profit).

The classical approach to managing this risk is embodied in the newsvendor (NV) model. It tries to find a production quantity Q so as to maximize the *expected* net profit from sales (the smaller of Q and the realized demand at T, denoted D_T) minus the net cost (cost minus salvage value). The optimal production quantity Q^{NV} is widely known to be determined by the so-called "critical ratio," $p/(p + c)$, with p being per unit profit and c per unit cost, applied to the inverse distribution function of D_T.

This simple model is lacking in several ways. First, maximizing the expected payoff (net profit) appears to be the wrong objective when it comes to managing risk. Simple calculation (in §2 below) will show that Q^{NV}, which achieves the maximal expected payoff, also maximizes the *variance* of the payoffs (for $Q \leq Q^{\mathrm{NV}}$; more on this in §2). In other words, Q^{NV} is merely one point of the mean-variance efficient frontier. Thus, instead of maximizing the expected payoff, it will be far more desirable to derive the entire efficient frontier. In this regard, other risk measures can also be considered, e.g., shortfall from a given target payoff, which only penalizes the downside risk. (The variance, in contrast, penalizes both above and below the mean.) The notion of efficient frontier also applies here — increasing the target payoff is conceivably associated with an increased risk of falling short from the target.

Second, the starting point of the NV model is the demand distribution, which is often determined by some forecasting scheme or time series analysis that typically returns a mean (or rate) plus a Gaussian noise (fluctuation from the mean). Yet, for a wide range of goods and services, demand will often depend on certain tradable assets in the financial market, or on the general economy, which can be proxied by a broad market index (such as the S&P500 index), of which there are ETF's traded in the market. Listed below are a few examples:

- Wal-Mart experienced increased demand during the last financial crisis as consumers sought lower-priced goods and its smaller-sized competitors went out of business. (*Wall Street Journal*, Nov 14, 2008, "Wal-Mart Flourishes as Economy Turns Sour.")

- The U.S. automobile industry sharply increased forecast and production when the last recession ended. (*Wall Street Journal*, Jan 14, 2014, "Auto Makers Dare to Boost Capacity: North American Factories Will Build One Million More Cars a Year.")

- A manufacturer for farming equipment experienced low demand amid falling prices of agricultural commodities. (*Forbes*, February 20, 2015, "Deere Sales Suffer From Farming Downturn.")

Thus, instead of forecasting the demand rate, a single parameter, one can (and should) do better with a rate *function* — to capture the dependence of demand on certain financial assets. This can be done by applying machine learning to data sets on both product sales and asset prices. The resulting rate function will capture the first-order functional dependence of demand on asset price (as opposed to statistical dependence like correlation between the two).

Third, in the NV model, the time-zero quantity decision is the only tool to cope with demand uncertainty at the end of horizon. And in most application context, this decision, once made, cannot be easily changed or modified without substantial penalty (as it will affect raw material ordered, equipment purchased, work force hired, etc). Yet, with the new demand model one can devise a *hedging strategy*, in the form of taking a position on the underlying tradable asset(s) and adjust it from time to time throughout the planning horizon. This real-time hedging strategy will not only add to the one-time quantity decision (at $t = 0$), it is clearly a more effective way to mitigate the risk in demand uncertainty, and to improve the risk-return tradeoff.

The main goal of this paper is to highlight *risk analytics* as a fundamental tool—applying analytical methods to turn data and problems into models and optimal decisions—much the same way as business analytics has in recent years supported, with considerable success, the end-to-end supply chains of most enterprises. The difference between the two also makes them distinctly complementary to each other. The focus of business analytics is on issues of efficiency: cost savings and revenue/profit optimization; whereas the emphasis of risk analytics is on issues of resiliency: risk-return tradeoff, catastrophe mitigation, and related resource allocation decisions and hedging strategies.

In the remainder of the paper, we start with a quick overview of related literature in §1.1, followed by a critique of the classical production planning model in §2. We then present a new demand model based on machine learning in §3, and formulate the integrated production-hedging models in §4. Solution techniques, main results and insights are highlighted in §5. These are supplemented with possible extensions in §6, along with concluding remarks in §7.

1.1 Related Literature

The mean-variance hedging model highlighted in §4 below is largely motivated by Caldentey and Haugh (2006), where the production model is more general, but the solutions obtained are not as explicit as ours, and characterizing the efficient frontier is not a concern in their study. Furthermore, to derive the optimal hedging strategy, they follow the approach of Schweizer (1992), while we follow the numeraire-based technique of Gourieroux *et al.* (1998) (also see Pham (2009)). Other related papers also include Gaur and Seshadri (2005), Wang and Wissel (2013) and Kouvelis *et al.* (2013). Gaur and Seshadri (2005) study a single-period NV problem, and model demand at T as a linear and increasing function of a tradable asset price (also at T), and thus turn the NV model into a contingent claim. They then construct a portfolio of options (and the underlying) as a hedging strategy. Wang and Wissel (2013) study mean-variance hedging of price risk born by kerosene consumers (or producers) with prespecified consumption (or production) in a continuous-time setting using crude oil futures. Kouvelis *et al.* (2013) consider a model that is similar to Wang and Wissel (2013): optimization under the mean-variance criterion, and financial hedging is in the setting of producing or storing a commodity that can be used to supply customer demand. They use a discrete-time dynamic programming approach to generate the hedging strategy, as opposed to the continuous-time stochastic control approach in Wang and Wissel (2013). Other related papers all pursue a discrete-time approach to hedging, in addition to the single-period setting in making the production or capacity decisions. Besides demands, the uncertainty linked to tradable assets that motivate hedging can include raw materials costs (Martinez-de-Albeniz and Simchi-Levi (2006)) and currency exchange rate risk (Chowdhry and Howe (1999) and Ding *et al.* (2007)).

Studies in the area of shortfall hedging are mostly motivated by asset allocation/pricing (e.g., Fishburn (1977), Harlow and Rao (1989)), which is quite different from the production planning context here. Yet, one of the models highlighted in this paper (§5.2) is to study a shortfall objective while meeting a certain "budget" constraint, we find it useful to explore the approach of convex duality in the shortfall hedging literature, in particular, the approach developed in Föllmer and Leukert (1999, 2000). Furthermore, in the operations management literature, there have been studies in recent years that use risk measures such as CVaR as objective functions in the NV model (e.g., Chen *et al.* (2015), Chen *et al.* (2009), Choi *et al.* (2011), Gao *et al.* (2011)). CVaR, being a quantile measure, relates naturally to shortfall; refer to Rockafellar and Uryasev (2000, 2002). Gao *et al.* (2011) considered demands that are negatively affected by weather, and use options written on weather indices for risk hedging. The other studies focus only on the production decision (Chen *et al.* (2009) studies pricing as well), and are not concerned with taking hedging strategies in the financial market.

Besides optimization of mean-variance tradeoff or CVaR, another stream of literature considers utility maximization. For example, Chen *et al.* (2007) consider a multi-period inventory and consumption planning model under which the risk-averse utility takes the form of a sum of exponential functions. While the main part of the paper does not concern financial hedging, in an extension section, a dynamic programming model is outlined to allow hedging using tradable securities.

In addition to financial hedging, there have been studies in the last decade or so advocating *operational* hedging in using production and other logistical means to increase operational flexibility to hedge against risk; refer to Boyabatli and Toktay (2004), which presents a critique on existing definitions of operational hedging in the literature, focusing on activities creating operational flexibility, thus decreasing negative impacts, without using financial instruments. To study the relationship between operational flexibility and financial hedging, Chod *et al.* (2010) consider correlated demands that are also correlated with tradable contracts, along with one-shot hedging using options. Van Mieghem (2010) gives an overview of risk mitigation in operations, and emphasizes the importance of using financial hedging as a tool to assist operational hedging. To illustrate, the author cites the example of a high correlation between a retail sales growth index and the S&P500 index, and hence the possibility to mitigate demand risk using index options. This is also the basic premise that motivates our study.

2 A Critique on Traditional Production Planning

The classical approach to production planning is the newsvendor (NV) model and its many variations. The basic model tries to find a production quantity Q at time $t = 0$ to supply a random demand D_T, which is realized at time $T > 0$, the end of the planning horizon (or production lead time) when the finished goods will supply demand; and the objective is to maximize the expected payoff, $\max_{Q \geq 0} \mathsf{E}[H_T(Q)]$, with

$$H_T(Q) := p(Q \wedge D_T) - c(Q - D_T)^+ = pQ - (p+c)(Q - D_T)^+, \tag{1}$$

where p is the unit profit (selling price minus cost), c is the net cost (cost minus salvage value) per unit, $\wedge$ denotes the min operator, and $(x)^+ := \max\{x, 0\}$. As motivated earlier, in the context of production planning, Q can be viewed as a proxy for *capacity*.

Taking derivative w.r.t. Q, and recognizing

$$\frac{d}{dQ}\mathsf{E}[(Q - D_T)^+] = \mathsf{P}(D_T \leq Q) := F(Q),$$

where $F(\cdot)$ denotes the distribution function of D_T, yields the optimal solution:

$$Q^{\mathrm{NV}} = F^{-1}\Big(\frac{p}{p+c}\Big). \tag{2}$$

This is the celebrated "critical ratio" solution.

Taking variance on $H_T(Q)$ in (1), we have

$$\mathsf{Var}[H_T(Q)] = (p+c)^2 \mathsf{Var}[(Q - D_T)^+].$$

Direct derivation yields

$$\frac{d}{dQ}\mathsf{Var}[H_T(Q)] = 2(p+c)^2 \mathsf{E}[(Q - D_T)^+][1 - \mathsf{P}(D_T \leq Q)] \geq 0.$$

That is, $\mathsf{Var}[H_T(Q)]$ is increasing in Q, for all Q. Furthermore, from (1), we know $\mathsf{E}[H_T(Q)]$ is concave in Q; and specifically, it is increasing in $Q \in [0, Q^{\mathrm{NV}}]$ and decreasing in $Q > Q^{\mathrm{NV}}$. Therefore, any $Q > Q^{\mathrm{NV}}$ does not induce efficient return-risk tradeoff and should not be considered.

In summary, we have:

Efficient Frontier of the Mean-Variance Model. Given m, let $Q_m \le Q^{\mathrm{NV}}$ be the solution to $\mathsf{E}[H_T(Q)] = m$; and let $m^{\mathrm{NV}} := \mathsf{E}[H_T(Q^{\mathrm{NV}})]$, where Q^{NV} is the NV solution in (2). Then, it is readily verified (by taking derivatives) that $\mathsf{Var}[H_T(Q_m)]$ is increasing and convex in m, for $m \in [0, m^{\mathrm{NV}}]$. That is, a larger m corresponds to a larger variance, thus forming an efficient frontier.

That the variance is increasing and *convex* in m is a sobering fact: increase in the mean will lead to acceleratingly steeper increase in variance, and the increment is at its steepest when the mean approaches its maximum m^{NV}.

Denote $v(m) := \mathsf{Var}[H_T(Q_m)]$. Then, the efficient frontier $(m, v(m))$ is an increasing and *convex* curve, meaning at a higher level of return (mean) any further increase is associated with a even steeper incremental increase in risk (variance). Thus, as we increase the production quantity towards the profit-maximizing Q^{NV}, the price we pay is the *maximal* incremental increase in risk.

Another weakness of the NV model is its lacking of an application context: who is exactly its intended user? Production planning is part of a firm's sales and operations planning process (SOP), by which the executive in charge works out production decisions together with sales and operations managers. In that context, the profit/revenue target has already been set by the firm's board. The executive's mandate is to meet or beat the target, not to set the target let alone to maximize the profit. In other words, maximizing profit is not even a relevant objective, not to add the enormous risk associated with such a pursuit as analyzed above.

With this application context in mind, it is useful to modify the classical NV model, replacing the profit-maximizing objective by minimizing a risk measure, the *shortfall*: $[m - H_T(Q)]^+$, where the constant $m \ge 0$ is a pre-specified profit target. From the H_T expression in (1), we can write

$$[m - H_T(Q)]^+ = \big[m + cQ - (p+c)D_T\big]^+, \qquad m \le pQ, \tag{3}$$

and

$$[m - H_T(Q)]^+ = m - pQ + (p+c)(Q - D_T)^+, \qquad m \geq pQ. \tag{4}$$

Clearly, the shortfall is increasing in the first case above, and it is equal to $m - \mathsf{E}[H_T(Q)]$ in the second case. Thus, direct verification leads to the following result.

Efficient Frontier of the Shortfall Model. Given m, the solution to $\min_{Q\geq 0} \mathsf{E}[m - H_T(Q)]^+ := s^{\text{NV}}(m)$ is $Q^{\text{NV}}(m) := \frac{m}{p} \wedge Q^{\text{NV}}$, with Q^{NV} following (2). Furthermore, $s^{\text{NV}}(m)$ is increasing (and convex) in m, which constitutes an efficient frontier.

Note from (4), when $m > pQ^{\text{NV}}$ the shortfall will grow linearly in m. In this case, a production-only decision becomes a handicap; and this motivates the risk-hedging strategy, to be detailed below.

3 Machine Learning for Demand Modeling

A typical demand forecasting scheme will assume a normal distribution for D_T. The rest is then carried out via some statistical analysis applied to past data (time series), so as to discern any trend (mean or rate) and to characterize the fluctuation around the mean (variance or "noise"). As motivated in §1, a better demand model is to allow the demand rate to be a *function* of certain financial assets:

$$dD_t = \tilde{\mu}(\mathbf{X}_t)dt + \tilde{\sigma} d\tilde{B}_t, \tag{5}$$

where $\mathbf{X}_t = (X_{kt})_{k=1}^K$ is a vector, with X_{kt} representing the price of k-th financial asset at time t. $\tilde{B}_t$ is a Brownian motion that represents the noise independent from $\mathbf{X}_t$ and $\sigma > 0$ is the associated parameter.

Note that the second term on the right hand side of (5), the noise part, can make dD_t negative, i.e., the cumulative demand is *not* necessarily increasing over time. While this may seem counter-intuitive, keep in mind that D_t for $t < T$ is understood as a *forecast* for cumulative demand up to t. For example, automobile industry is known to have complex supply chains, and the capacity decision for manufacturing needs to be made well in advance of the sales season. Over this planning horizon, demand forecast is updated over time and does fluctuate (e.g. certain orders may be canceled or reduced). In particular, on $t = T$, the demand forecast equals the realized demand.

We can apply machine learning (ML) to obtain the function $\tilde{\mu}(\cdot)$ from data set specific to each application. This involves: (i) identifying, from historical data of a set of financial assets in the market, the ones that impact the demand rate; and (ii) learning how the demand rate moves with the selected variables. Technically, this means finding the arguments (i.e. assets to be included in $\mathbf{X}_t$) and the functional form of $\tilde{\mu}(\cdot)$.

3.1 Learning the Demand Rate Function

We start with describing the approach with data sets of *realized* demands, and then extend the approach to incorporate demand forecasts data in §3.2. Suppose the data sets of realized demands take the following format.

- Demand: $\{D_i,\, i = 1, \cdots, N\}$, where D_i is the demand realized for the i-th period which starts on time t_i and ends on t_{i+1}. (i.e., t_i (resp. t_{i+1}) is the time "0" (resp. time "T") for the i-th period; writing $T = T_i$, then $D_i = D_{T_i}$.)

- Asset: The *daily* (closing) price data of asset k is in form of $\{X_{kt_{ij}}, i = 1, \cdots, N, j = 1, \cdots, N_i\}$. Here t_{ij} is the time of j-th trading day within the time frame $[t_i, t_{i+1}]$ (i.e. period i), and thus N_i is the total number of trading days within this period. Then, $\mathbf{X}_{t_{ij}} = (X_{kt_{ij}})_{k=1}^{K}$ is the vector of asset prices at each time t_{ij}.

Compared to the daily asset price data, demand data is typically in longer time scale (monthly, quarterly and yearly). Therefore, for each demand data point D_i, which is associated with a time frame $[t_i, t_{i+1}]$, there are N_i asset price vectors within this time frame. Accordingly, (5) is discretized as the following:

$$D_i = \sum_{j=1}^{N_i} \delta_{ij} \tilde{\mu}(\mathbf{X}_{t_{ij}}) + \sigma \sqrt{\delta_i} \xi_i, \quad i = 1, \cdots, N. \tag{6}$$

On the right hand side, the first term (the sum) is the quadrature approximation for the time-integral in (5), and $\{\delta_{ij}\}$ are the associated weights which only depend on the time points $\{t_{ij}\}$. The second term uses the distribution of Brownian increments, and ξ_i are i.i.d. standard normal random variables (which are also independent from $\mathbf{X}_{t_{ij}}$); $\delta_i = t_{i+1} - t_i$ are also constants. The discretization in (6) effectively separates the noise (i.e. ξ_i) from the financial assets, and this defines the sum of residual squares throughout the following text.

Variable Selection Although the initial selection of asset types can be guided by risk factors recognized by the manufacturer (e.g., the risk factors disclosed in the firm's annual report), the number of potentially relevant assets could be large. The goal here is to select, from a possibly large set of candidate assets (indexed by $k = 1, \cdots, K$), a subset of assets exhibiting most relevance to the demand. In this regard, there are well-established methods; for instance, refer to Guyon and Elisseeff (2003). Among others, linear predictor with a variable selection feature as a filter is immediately applicable. Specifically, this means imposing $\tilde{\mu}(\mathbf{x}) = \beta_0 + \sum_{k=1}^{K} \beta_k x_k$ (for the purpose of variable selection) and using algorithms to reduce some of the β_k to 0; then, variables associated with non-zero β_k are the selected ones. One commonly used approach is ℓ_1-regularization, and here it amounts to solving:

$$\min_{\beta_0,\beta_1,\cdots,\beta_K} \sum_{i=1}^{N} \frac{1}{\delta_i} \Big[D_i - \beta_0 \Big(\sum_{j=1}^{N_i} \delta_{ij} \Big) - \sum_{k=1}^{K} \beta_k \Big(\sum_{j=1}^{N_i} \delta_{ij} X_{kt_{ij}} \Big) \Big]^2 + \lambda \sum_{k=1}^{K} |\beta_k|;$$

$\lambda > 0$ is a chosen tuning parameter that can be set to control the number of selected assets. The formulation above is essentially a Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani (1996)) with $\sum_{j=1}^{N_i} \delta_{ij} X_{kt_{ij}}$, the daily average asset price over the i-th period, as the k-th predictor.

Function Learning Based on (6), finding the functional form of $\tilde{\mu}(\cdot)$ in the prices of the selected assets can be cast into the following generic least squares problem:

$$\min_{\tilde{\mu} \in \boldsymbol{\mathcal{U}}} \quad \sum_{i=1}^{N} \frac{1}{\delta_i} \Big[D_i - \sum_{j=1}^{N_i} \delta_{ij} \tilde{\mu}(\mathbf{X}_{t_{ij}}) \Big]^2; \tag{7}$$

where $\boldsymbol{\mathcal{U}}$ is the function space of candidates of $\tilde{\mu}(\cdot)$.

The problem in (7) can be further reduced to structured regression by restricting $\boldsymbol{\mathcal{U}}$. Specifically, $\tilde{\mu}(\cdot)$ is represented by a linear combination of basis functions (Hastie *et al.* (2009)):

$$\tilde{\mu}(\mathbf{x}) = \sum_{m=1}^{M} \beta_m b_m(\mathbf{x}; \theta). \tag{8}$$

Each $b_m(\mathbf{x}; \theta)$ is a selected basis function with a specified functional form parameterized by θ.

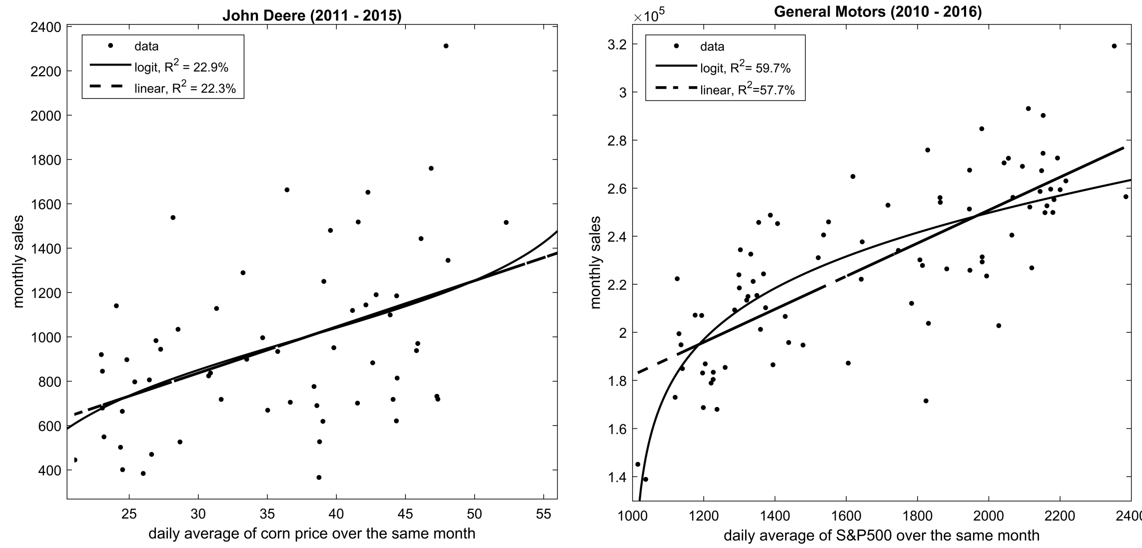

Figure 1: Monthly sales versus asset prices. y-axis represents monthly sales (in units) and x-axis represents daily average of asset price within the same month. "linear" stands for $\tilde{\mu}(x) = \beta_0 + \beta_1 x$ (where x is the daily average asset price). "logit" stands for $\tilde{\mu}(\hat{x}) = \beta_0 + \beta_1 \log(\hat{x}/(1-\hat{x}))$, where $0 \le \hat{x} \le 1$ is the transformed asset price: $\hat{x} = (x-L)/(U-L)$; $(L, U) = (15, 60)$ for corn price and $(1000, 5000)$ for S&P500.

Rearranging the terms, (7) reduces to:

$$\min_{\beta_m, \theta} \quad \sum_{i=1}^{N} \frac{1}{\delta_i} \Big[D_i - \sum_{m=1}^{M} \beta_m \Big(\sum_{j=1}^{N_i} \delta_{ij} b_m(\mathbf{X}_{t_{ij}}; \theta) \Big) \Big]^2. \tag{9}$$

Common choices of the basis functions include, e.g., polynomials, log, and logit functions.

In Figure 1, we illustrate a preliminary study that applies the above approach to data from two firms, John Deere and General Motors, respectively. For John Deere, the underlying asset is corn futures; for General Motors, it is the S&P500 index. In both cases, we study the relationship between the firm's monthly sales (y-axis) and the daily (average) asset price within the same month (x-axis). A linear fit appears to capture the relationship quite well, and the learning is further improved by using a logit function. (This study is slightly different from the demand model in (5) since the predictor is $\tilde{\mu}(\int_0^T X_t dt)$ as opposed to $\int_0^T \tilde{\mu}(X_t) dt$.)

3.2 Integration with Forecasting

Demand forecasts data can be incorporated with the machine learning component. For simplicity, denote the i-th production cycle over $[t_i, t_{i+1}]$ as $[0, T_i]$. Suppose within the firm there are demand forecasts (say, generated by existing blackbox routine) of D_i ($:= D_{T_i}$) taking place on dates $0 < s_{i1} < s_{i2} < \cdots < s_{in_i} < T_i$, denoted by $\{\hat{D}_{ij}, j = 1, \cdots, n_i\}$. Following usual convention to let $s_{i(n_i+1)} = T_i$, hence $\hat{D}_{i(n_i+1)}$ is recognized as D_i (note this means forecast made at the time of demand realization is just the realized demand). Each $\hat{D}_{ij}$ relates to the demand model in (5) by:

$$\begin{aligned}\hat{D}_{ij} = \mathsf{E}[D_i \,|\, \mathcal{F}_{s_{ij}}] &= \mathsf{E}\Big[D_{s_{ij}} + \int_{s_{ij}}^{T_i} \tilde{\mu}(\mathbf{X}_u)du + \tilde{\sigma}(\tilde{B}_{T_i} - \tilde{B}_{s_{ij}}) \Big| \mathcal{F}_{s_{ij}}\Big] \\ &= \int_0^{s_{ij}} \tilde{\mu}(\mathbf{X}_s)ds + \tilde{\sigma}\tilde{B}_{s_{ij}} + \mathsf{E}\Big[\int_{s_{ij}}^{T_i} \tilde{\mu}(\mathbf{X}_u)du \Big| \mathcal{F}_{s_{ij}}\Big]. \qquad (10)\end{aligned}$$

$\mathcal{F}_{s_{ij}}$ stands for all the information accumulated over $[0, s_{ij}]$, to which $D_{s_{ij}}$ is known and $B_{T_i} - B_{s_{ij}}$ is independent.

Apply discretization to (10) and define:

$$\begin{aligned}\Delta\hat{D}_{i0} &:= \hat{D}_{i1} = \sum_{t_{i\ell}=0}^{s_{i1}} \delta_{i\ell}\tilde{\mu}(\mathbf{X}_{t_{i\ell}}) + \sigma\sqrt{\hat{\delta}_{i0}}\xi_{i0} + g_{i1}; \\ \Delta\hat{D}_{ij} &:= \hat{D}_{i(j+1)} - \hat{D}_{ij} \\ &= \sum_{t_{i\ell}=s_{ij}}^{s_{i(j+1)}} \delta_{i\ell}\tilde{\mu}(\mathbf{X}_{t_{i\ell}}) + \sigma\sqrt{\hat{\delta}_{ij}}\xi_{ij} + g_{i(j+1)} - g_{ij},\ j = 1, \cdots, n_i. \qquad (11)\end{aligned}$$

$t_{i\ell}$ stands for the ℓ-th trading day of period i and the terms involving sums are the quadrature approximations of $\int_{s_{ij}}^{s_{i(j+1)}} \tilde{\mu}(\mathbf{X}_s)ds$ (with $\delta_{i\ell}$ being the weights). $\hat{\delta}_{ij} = s_{i(j+1)-}s_{ij}$ are constants ($s_{i0} := 0$) and ξ_{ij} are i.i.d. standard normal variables. The quantity g_{ij} is defined as follows:

$$g_{ij} := \mathsf{E}\Big[\int_{s_{ij}}^{T_i} \tilde{\mu}(\mathbf{X}_u)du \Big| \mathcal{F}_{s_{ij}}\Big] = \int_{s_{ij}}^{T_i} \mathsf{E}\Big[\tilde{\mu}(\mathbf{X}_u) \Big| \mathcal{F}_{s_{ij}}\Big]du. \qquad (12)$$

(The switch of order of expectation and integral requires some mild technical condition on $\tilde{\mu}(\cdot)$.) The computation of g_{ij}, in particular the conditional expectation involved, can be done via Monte Carlo by simulating sample paths of $\mathbf{X}_u$. This, in turn, requires first fitting an asset price model for $\mathbf{X}_t$ (such as the geometric Brownian motion). For the structured regression discussed above, the computation of g_{ij} over realized samples of $\mathbf{X}_u$ is straightforward for each chosen set of basis functions.

The i.i.d. noises ξ_{ij} in (11) are separated out to formulate the least squares problem:

$$\min_{\tilde{\mu}} \sum_{i=1}^{N} \Big\{ \sum_{j=1}^{n_i} \frac{1}{\hat{\delta}_{ij}} \Big[\Delta \hat{D}_{ij} \quad - \sum_{t_{i\ell}=s_{ij}}^{s_{i(j+1)}} \delta_{i\ell} \tilde{\mu}(\mathbf{X}_{t_{i\ell}}) - g_{i(j+1)} + g_{ij} \Big]^2 + \frac{1}{\hat{\delta}_{i0}} \Big[\Delta \hat{D}_{i0} \quad - \sum_{t_{i\ell}=0}^{s_{i1}} \delta_{i\ell} \tilde{\mu}(\mathbf{X}_{t_{i\ell}}) - g_{i1} \Big]^2 \Big\}. \tag{13}$$

Restricting $\tilde{\mu}(\cdot)$ to (8), the problem in (13) is cast as a structured regression. Solving this problem is more challenging than solving (9), as evaluation of g_{ij} in (12) is computationally expensive. Efficient global optimization algorithms are appropriate; refer to Jones *et al.* (1998) and Gutmann (2001).

4 Production-Hedging Models

To facilitate exposition, assume the demand model in (5) depends on a single asset price, X_t; and assume X_t follows the geometric Brownian motion (GBM), a standard asset price process:

$$dX_t = X_t(\mu dt + \sigma dB_t); \tag{14}$$

where μ and σ are parameters, and B_t is a Brownian motion independent from the demand noise, $\tilde{B}_t$, in (5).

With hedging, there are two components in the payoff function — the terminal wealth (at $t = T$) from hedging (χ_T) as well as from production (H_T):

$$H_T(Q) := pQ - (p + c)(Q - D_T^+)^+, \quad \chi_T(\vartheta) := \int_0^T \theta_t dX_t$$

where $H_T(Q)$ follows the same expression in the NV model in §2, except here we use D_T^+ to enforce the non-negativity of the realized demand at T (which is needed when D_T is assumed to have a normal noise component); θ_t is the position (number of shares) taken on the asset with per share price X_t, and hence χ_T the terminal wealth due to hedging. The decisions are Q and $\vartheta := \{\theta_t,\, t \in [0, T]\}$, with the latter required to be adapted to the information dynamics of both demand forecast and asset price, i.e., $\theta_t \in \mathcal{F}_t$ for all $t \in [0, T]$, where $\{\mathcal{F}_t\}$ is the filtration associated with $\{(D_t, X_t)\}$.

For a mean-variance formulation, the problem we want to solve is:

$$B(m) := \inf_{Q \geq 0, \vartheta} \{\mathsf{Var}[H_T(Q) + \chi_T(\vartheta)] \,|\, \mathsf{E}[H_T(Q) + \chi_T(\vartheta)] \geq m\}, \tag{15}$$

i.e., we want to minimize the variance of the total payoff, subject to its mean at or above a given target level m. This formulation essentially follows the same spirit as Markowitz mean-variance portfolio optimization model, refer to Markowitz (1987).

Write the terminal wealth at $t = T$ as $W_T := H_T + \chi_T$. It can be shown that at optimality the constraint in (15) must be binding, i.e., $\mathsf{E}(W_T) = m$. (More on this at the end of §5.1 below.) Hence, the objective reduces to minimizing the second moment of W_T, $\mathsf{E}(W_T^2)$, given the constraint $\mathsf{E}(W_T) = m$.

In general, minimizing the variance, $\mathsf{Var}(W_T)$, will suffer from the so-called "time inconsistency" (see, e.g., Basak and Chabakauri (2010)), mainly because it does not satisfy the law of iterated conditioning: $\mathsf{Var}(W_T) \neq \mathsf{E}[\mathsf{Var}(W_T|\mathcal{F}_t)]$. Yet, this is not an issue in the formulation above: we do have $\mathsf{E}(W_T^2) = \mathsf{E}[\mathsf{E}(W_T^2|\mathcal{F}_t)]$, for all $t \in [0, T]$.

As the variance objective penalizes both up- and down-side deviation from the target, alternatively, we can minimize a shortfall objective as follows.

$$\inf_{Q \geq 0, \vartheta} \mathsf{E}\Big\{\big[m - H_T(Q) - \chi_T(\vartheta)\big]^+\Big\} \tag{16}$$

$$\text{s.t.} \quad \chi_t := \int_0^t \theta_s dX_s \geq -C, \quad \theta_t \in \mathcal{F}_t^X, \qquad t \in [0, T],$$

where $C > 0$ is a given limit to bound the possible loss from trading, and $\{\mathcal{F}_t^X\}$ is the filtration generated by $\{X_t\}$ (or, equivalently, by $\{B_t\}$). These are two features not present in the mean-variance model. So, let's motivate.

First, the constraint $\chi_T \geq -C$ is necessary to make the shortfall formulation above a meaningful model; in particular, C can be viewed as a "budget" to carry out the hedging strategy. With an infinite budget, it is conceivable that one can always achieve any given target m, and hence, zero shortfall; and this can be done with hedging alone, with no need for production ($Q = 0$). On the practical side, a limited C value not only controls the possible loss due to trading (i.e., hedging), it may also reflect the firm's desire not to deviate from its main line of business, which is production.

Second, by restricting the hedging strategy to the filtration generated by $\{X_t\}$ alone, we are pursuing a formulation with "partial information". This is often a good representation of reality: in practice the hedging/trading decision is often a real-time decision taking input from the financial market, as embodied in the filtration $\{\mathcal{F}_t^X\}$. It would be unrealistic to assume that one could simultaneously also keep track of demand projection, which typically involves piecing together disperse information garnered from polling the sales force, and is hence updated much less frequently, at much longer time scales. Thus, in the research literature partial information is an important model in its own right; see, e.g., Caldentey and Haugh (2006).

5 Approaches, Solutions and Insights

5.1 Mean-Variance Hedging

To solve the problem in (15), first assume Q is given (and hence omit it from the arguments of B and H_T), so that we can focus on deriving the hedging strategy ϑ. Consider the *conjugate dual* of $B(m)$:

$$A(\lambda) := \inf_{\vartheta} \mathsf{E}\Big\{\Big[\lambda - H_T - \chi_T(\vartheta)\Big]^2\Big\}, \tag{17}$$

where λ is a given parameter, in parallel to m. The pair, $A(\lambda)$ and $B(m)$ are connected as follows. Write $Y := H_T + \chi_T$, and note

$$\begin{aligned} B(m) &= \mathsf{Var}(Y) = \mathsf{Var}(Y - \lambda) = \mathsf{E}[(Y-\lambda)^2] - [\mathsf{E}(Y-\lambda)]^2 \\ &= \mathsf{E}[(\lambda - Y)^2] - (m-\lambda)^2 = A(\lambda) - (m-\lambda)^2. \end{aligned} \tag{18}$$

Thus, finding the optimal hedging strategy ϑ via solving the $B(m)$ problem in (15) is the same as via solving the $A(\lambda)$ problem in (17), since the difference between the two, $(m-\lambda)^2$, is independent of ϑ, involving given parameters only. Furthermore, the relation in (18) above leads to (see Proposition

6.6.5 in Pham (2009)):

$$A(\lambda) = \min_m \Big[B(m) + (m-\lambda)^2\Big], \quad B(m) = \max_\lambda \Big[A(\lambda) - (m-\lambda)^2\Big]. \tag{19}$$

The problem in (17) belongs to a class of control problems known as MSE (mean-square error), from which the optimal hedging strategy (given Q) can be derived using a numeraire-based approach (Gourieroux *et al.* (1998)), and expressed explicitly as follows (refer to Wang and Yao (2017) for details):

$$\theta_t^* = -\xi_t(Q) + \frac{\eta}{\sigma X_t}\big[\lambda - V_t(Q) - \chi_t^*\big], \tag{20}$$

where $\eta := \mu/\sigma$, and $V_t(Q) := \mathsf{E}^M[H_T(Q)|\mathcal{F}_t]$. Here E^M denotes expectation with respect to the risk-neutral measure P^M, as opposed to the original, physical measure P, and the change of measures is induced by η. Specifically, the Radon-Nikodym derivative is

$$\frac{d\mathsf{P}^M}{d\mathsf{P}} = e^{\eta B_T - \eta^2 T/2} := Z_T. \tag{21}$$

Note that V_t is a P^M-martingale, and admits the following representation:

$$V_t(Q) = V_0(Q) + \int_0^t \xi_s(Q) dX_s + \int_0^t \delta_s(Q) d\tilde{B}_s, \tag{22}$$

and the two processes ξ_t and δ_t can be derived explicitly via Itô's formula.

Having solved the $A(\lambda)$ problem in (17), for a given λ, we can obtain $B(m)$ from (19) by further optimizing λ. The resulting solution can be expressed as follows (with Q now added back to emphasize the dependence on Q, as well as the fact that Q has yet to be optimized):

$$B(m, Q) = \frac{[m - V_0(Q)]^2}{e^{\eta^2 T} - 1} + \int_0^T e^{-\eta^2 (T-t)} \mathsf{E}[\delta_t^2(Q)] dt. \tag{23}$$

Furthermore, λ and m relate to each other as follows:

$$\lambda = \frac{m e^{\eta^2 T} - V_0(Q)}{e^{\eta^2 T} - 1}. \tag{24}$$

From (23), $B(m, Q)$ is (strictly) increasing in m when $m \geq V_0$ (for any given Q). This also explains why the constraint in (15) must be binding. Furthermore, $B(m, Q)$ is (strictly) decreasing in m for $m < V_0$. Hence, setting the target m at any level less than V_0 is not efficient.

5.2 Shortfall Hedging

Again, we first assume Q is given and focus on deriving the optimal hedging strategy. Applying Jensen's inequality, along with conditional expectation on X_T and $A_T := \int_0^T \tilde{\mu}(X_t)dt$, we can turn the real-time hedging problem in (16) into a *static* optimization problem below:

$$\min_{V_T} \mathsf{E}\big[(m - H_T - V_T)^+\big] \quad \text{s.t.} \quad V_T \geq -C, \quad \mathsf{E}^M(V_T) \leq 0\ . \tag{25}$$

It is important to point out that, due to Jensen's inequality, the objective function in the above optimization problem is a *lower bound* of the original objective in (16). In this regard, the above cannot be equivalent to the original problem. Yet, the constraint $\mathsf{E}^M(V_T) \leq 0$ is not present in the original problem; and with this additional constraint, the two problems can be shown to be equivalent. This constraint follows from χ_t being a P^M-supermartingale. Here, as in the mean-variance model above, the change of measure is induced by $\eta = \mu/\sigma$ via the Radon-Nikodym derivative in (21).

This dual (lower-bound) problem can be solved by a Lagrangian multiplier approach (for details, refer to Wang and Yao (2019)), and the solution is:

$$V_T^* = (p + c)(Q - \hat{D}_T^+)^+ + (m - pQ + C)\mathbf{1}\{\lambda^* Z_T \leq 1\} - C, \tag{26}$$

where Z_T follows (21), λ^* is the positive Lagrangian multiplier, and $\hat{D}_T := A_T + \tilde{\sigma}\sqrt{T}\Phi^{-1}(\lambda^* Z_T)$ is the "proxy" for D_T (as the latter is not accessible due to partial information) with Φ^{-1} being the inverse distribution function of the standard normal random variable. Once $V_T^*(= \chi_T^*)$ is derived, the optimal hedging strategy θ_t^* follows from Itô's Lemma along with martingale representation theorem, similar to what's outlined in the mean-variance model.

Substituting V_T^* in (26) back into the shortfall objective, and denoting the latter as $s(m, Q)$, we can derive the following expression:

$$s(m,Q) = (p+c)\mathsf{E}\big[(Q \wedge \hat{D}_T^+ - D_T^+)^+\big] + (m - pQ + C)\mathsf{P}(\lambda^* Z_T \geq 1). \tag{27}$$

5.3 Optimal Production Quantity and Efficient Frontiers

The shortfall under the optimal hedging strategy, $s(m,Q)$ in (27) can be shown (with considerable effort, as both λ^* and $\hat{D}_T$ also depend on Q) to be convex in Q. Thus, finding the optimal solution $Q^*(m)$, jointly with the optimal hedging strategy, is a readily solvable convex minimization problem. To make it even better, a universal upper bound (i.e., independent of m) on the optimal Q can be explicitly identified, and this further facilitates the (line) search for $Q^*(m)$. Furthermore, it can be shown that $s(m, Q^*(m))$ is increasing in m, hence constitutes an efficient frontier — setting a higher target will lead to a higher shortfall.

Similarly, the optimal production quantity in the mean-variance model can also be obtained from a line search on the $B(m,Q)$ expression in (23), which, however, need not be convex in Q (its first term is, but not the second one). Also, $B(m, Q^*(m))$ is increasing in m, forming an efficient frontier — a higher mean return corresponds to a higher variance.

In Figure 2, we compare the efficient frontier of the NV model (with a shortfall objective) with the efficient frontier of the shortfall hedging model. In the left panel, we plot the two efficient frontiers. Clearly, the hedging strategy leads to a significant improvement: for any target value (x-axis), the corresponding shortfall (y-axis) is substantially lower. Of particular interest is the range of high target values. Recall from (4), for $m \geq pQ^{\mathrm{NV}}$ ($= 4100$ in this problem instance), the shortfall of the NV model will increase linearly in m, at a slope of 1; and this is confirmed here in the figure. (In particular, note the right panel, which plots the rate of increase of the frontier curves.) In contrast, with the addition of the hedging strategy, the increase of the shortfall is at a substantially lower slope (peaked at slightly below 0.7), and this is achieved via a modest hedging budget of $C = 0.1m$, one-tenth of the target.

Similarly, for mean-variance hedging, Figure 3 compares the efficient frontiers of the NV model and the hedging model, for two problem instances, corresponding to the two panels in the figure. The improvement contributed by the hedging strategy is also quite substantial, in particular at the higher end of the mean, which is where the risk increases more steeply. (Note here all the frontier

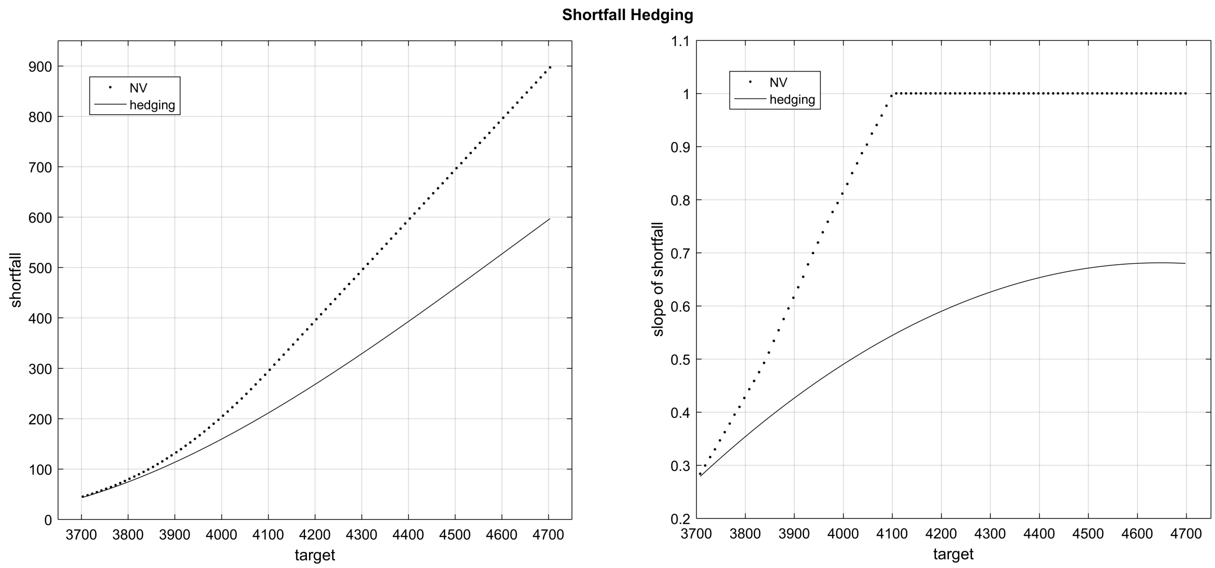

Figure 2: Shortfall Efficient Frontiers. The right panel shows the derivative (w.r.t. the target) of the two curves in the left panel.

curves are increasing and *convex*.)

5.4 Insights

First, let's take a closer look at the hedging strategies. The optimal mean-variance hedging in (20), θ_t^*, maintains two positions in the asset (X_t) at any time t:

- a position to "cancel" out the ξ_t in V_t;
- a position equal to the gap between the current wealth $(V_t + \chi_t^*)$ and the target (λ), weighted by the ratio η/σ, so as to "catch up."

For the shortfall hedging, the V_T^* expression in (26) has two terms, both try to close the gap from m left by the production payoff, $m - H_T(Q) = (p+c)(Q - D_T^+)^+ + (m - pQ)$.

- The first term, $(p+c)(Q - \hat{D}_T^+)^+$, can be viewed as a "put option." It tries to close the first part of the gap, but needs to use $\hat{D}_T$ as a surrogate for D_T due to partial information.
- The second term, $(m - pQ + C)\mathbf{1}\{\lambda^* Z_T \leq 1\}$, is a "digital option." It aims to close the other part of the gap (after subtracting C).

In summary, the hedging strategy in the mean-variance model is to dynamically maintain a portfolio of two positions on the underlying financial asset. The first position cancels out the tradable

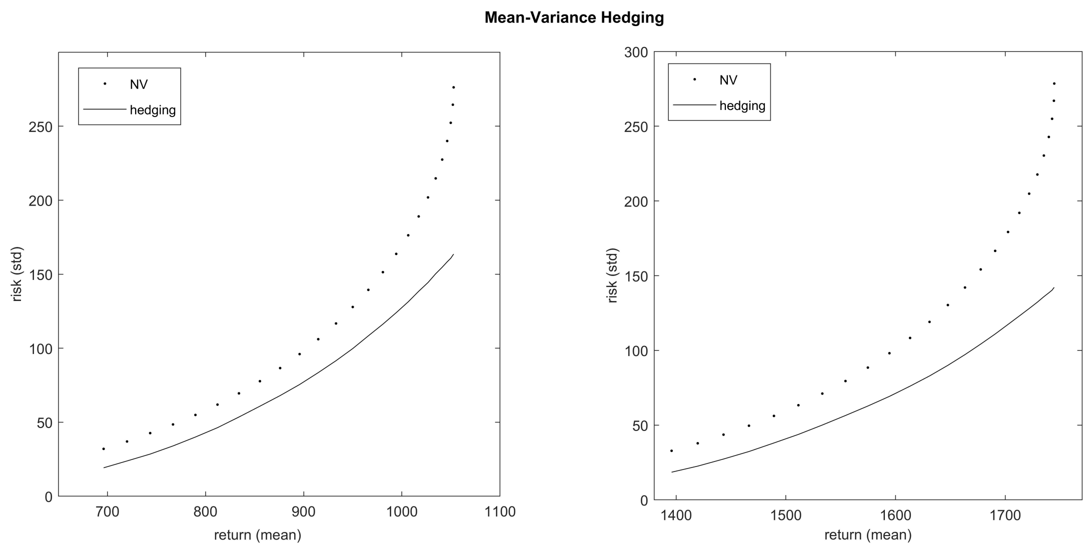

Figure 3: Mean-Variance Efficient Frontiers. Each panel corresponds to one problem instance.

component of the projected wealth from production, i.e., revenue from supplying demand; thus, its function is one of pure risk mitigation. The second position uses trading gains to catch up with the target mean return. In contrast, the shortfall hedging strategy takes the form of two options, a put option and a digital option, and the underlying for both is the "surrogate demand" $\hat{D}_T$, necessitated by the lack of full information on the real demand. (Note, the Lagrangian multiplier λ^* also depends on $\hat{D}_T$, since it is the solution to $\mathsf{E}^M(V_T^*) = 0$, the binding constraint in (25).) Since the risk measure is shortfall, both options are designed to contribute more terminal wealth so as to supplement production and help closing the gap from the target. In other words, there is no "cancelling" component in the hedging strategy; and this is only natural as the shortfall measure does not penalize any upside risk.

Next, let's specify the contribution of the hedging strategies, i.e., what they can add to the production-only decision. Specifically, for the mean-variance hedging, we examine the variance reduction it can achieve, from the variance of the NV model. Suppose Q is given, and $m = \mathsf{E}[H_T(Q)] + \mathsf{E}(\chi_T^*)$, where χ_T^* is the terminal wealth attained by the optimal hedging strategy that minimizes $B(m, Q)$ in (23): $B(m, Q) = \mathsf{Var}[H_T(Q) + \chi_T^*]$. Denote

$$m_{\min} := \mathsf{E}[H_T(Q)] \wedge \mathsf{E}^M[H_T(Q)], \qquad m_{\max} := \mathsf{E}[H_T(Q)] \vee \mathsf{E}^M[H_T(Q)].$$

Then, we can show

$$B(m, Q) \leq \mathsf{Var}[H_T(Q)] - \mathsf{Var}(\chi_T^*), \qquad \forall\, m \in [m_{\min}, m_{\max}]. \tag{28}$$

In particular, when $m = m_{\min}$ or $m = m_{\max}$, we have $B(m, Q) = \mathsf{Var}[H_T(Q)] - \mathsf{Var}(\chi_T^*)$. That is, the hedging strategy reduces the variance by an amount that is at least $\mathsf{Var}(\chi_T^*)$.

For shortfall hedging, its contribution to reducing the shortfall has been illustrated in the efficient frontiers in Figure 2. Here's a different but related insight. One concern is that by minimizing a shortfall objective, one might do poorly in the *expected* terminal wealth, in comparison against the NV model (which maximizes the expected terminal wealth as its objective). It turns out this concern is unwarranted. It can be shown that the shortfall-minimizing objective leads to a jointly optimized production-hedging decision that will contribute *more profit* than the profit-maximizing production-only decision of the NV model. Indeed, the improvement in expected total terminal wealth (at T), above and beyond the NV model, can be quantified as follows:

$$\mathsf{E}[H_T(Q^*(m)) + V_T^*] - \mathsf{E}[H_T(Q^{\mathrm{NV}}(m))] \geq \beta(m - pQ^{\mathrm{NV}})^+ + C\psi \geq 0, \tag{29}$$

where $\beta \in [0, 1]$ and $\psi \in [0, \lambda^*]$. (Recall λ^* is the Lagrangian multiplier from solving the hedging problem.)

6 Extensions

6.1 Multiple Products

A firm typically produces multiple products; and customer demands on its various product lines are not only correlated in a statistical manner, they may actually form certain functional relations: a demand surge on one product typically leads to decreases on other products. This functional dependence may also originate from the demand on each product depending (in its own way) on multiple financial assets such as commodities, and on the general economy as well. Consider, for instance, the following cases widely reported in business news:

- During the (quite recent) period when oil price was plunging, many car buyers switched out of smaller models into SUV's and other gas guzzlers. Yet, due to relevant standards and regulations (such as Corporate Average Fuel Economy), car producers had already increased the production of more fuel-efficient models only to see them suffering from reduced demand. (*Wall Street Journal*, November 19, 2014, "Ford Presses Ahead With Developing Fuel-Efficient Vehicles."; *Wall Street Journal*, January 13, 2015, "Clash Looms Over Fuel Economy Standard.")

- Quite recently, automakers was financing the development of electric vehicles by shifting production to SUV's (which has higher profit margin) from the smaller, fuel-efficient sedans. Then, the fuel price began rising, and customers switched from the gas guzzlers to smaller cars. (*Bloomberg*, April 26, 2018, "It's a Bad Time to Stop Making Small Cars: Automakers Are Financing Their Electric-Vehicle Research by Selling SUV's, But Rising Gas Prices Will Turn Off Consumers.")

Of course, in addition to dependence on fuel price, the above demand can also be a function of the general economy, as indicated in the second bullet point in §1 ("automakers"). Therefore, we want to study the integrated risk hedging and capacity planning in a multi-product setting.

There are J products, indexed by $j = 1, \dots, J$. Each product j has a demand rate that is a function of K financial assets, indexed by $k = 1, \dots, K$. Let $\mathbf{X}_t := (X_{kt})_{k=1}^{K}$ denote the vector of asset prices at time t. Assume the asset prices follow the stochastic differential equations below:

$$dX_{kt} = X_{kt}\Big(\mu_k dt + \sum_{\ell=1}^{K} \sigma_{k\ell} dB_{\ell t}\Big), \qquad k = 1, ..., K; \tag{30}$$

where $B_{\ell t}$, $\ell = 1, \dots, K$, are components of $\mathbf{B}_t$. μ_k and $\sigma_{k\ell}$ are constant parameters. Denote $\mu := (\mu_k)_{k=1}^{K}$, the vector of rates; and $\Sigma := [\sigma_{k\ell}]_{k,\ell=1}^{K}$, the covariance matrix. Assume Σ is invertible, and denote

$$\boldsymbol{\eta} := \Sigma^{-1}\boldsymbol{\mu}. \tag{31}$$

Then, $\boldsymbol{\eta}$ is the vector of "market price of risk" that induces the risk-neutral measure. (Assume, as before and without loss of generality, that the risk-free interest rate is zero.)

Next, we specify the demand processes. Let D_{jt} denote the (forecast) cumulative demand up to t for product j:

$$dD_{jt} = \tilde{\mu}_j(\mathbf{X}_t)dt + \sum_{i=1}^{K} \tilde{\sigma}_{ji} d\tilde{B}_{it}, \qquad j = 1, ..., J; \tag{32}$$

where $\tilde{\mu}_j(\cdot) : \Re^K_+ \mapsto \Re_+$ is a nonnegative function and $\tilde{\sigma}_{ji}$ are constants. Note here, $\tilde{B}_{it}$'s are the random factors that model demand forecasts errors and how many of these factors one needs is often an issue of modeling. Here the choice of K, same as the number of random factors in the asset price model, is only for notational simplicity. (It can be changed to K', for instance.) $\tilde{B}_{it}$, $i = 1, \cdots, K$, are components of $\tilde{\mathbf{B}}_t$, and recall they are independent from the $(B_{kt})_{k=1}^K$ involved in the asset prices in (30).

We expect the approach that derives the optimal hedging strategy (with the production decision given) as outlined in §5 can be extended, with effort, to the high-dimensional demand/asset model as spelled out above. Optimize the production quantities $Q := (Q_j)_{j=1}^J$, *jointly* with the hedging strategy, will be a challenge, due to the lack of second-order properties in Q. (Recall, as explained above, even in the single-dimensional case, the objective function is not convex in Q.) Numerical computation will be needed. Another line of attack is to dig deeper into the problem structure, and to explore special cases. (For instance, assume the product demands have independent or positively correlated forecast noise — due to using a common forecasting scheme/tool.)

6.2 General Asset-Price Processes

The geometric Brownian motion (GBM) model in (14) has been a standard (as well as the most basic) model for financial assets (e.g., Hull (2014) and Luenberger (1998)). Yet, the price dynamics of many asset classes have been observed to follow a mean-reverting pattern that cannot be captured by the GBM. These include futures, commodities, interest rates and exchange rates (e.g., Anthony and MacDonald (1998), Casassus and Collin-Dufresne (2005), Larsen and Srensen (2007)), and even equities (e.g., Gropp (2004) and Poterba and Summers (1988)). In addition, mean-reversion is also a key feature in many trading strategies such as pairs trading (e.g., Leung and Li (2015)).

A popular diffusion process that is mean reverting is the Ornstein-Uhlenbeck (OU) process:

$$dX_t = \mu(\kappa - X_t)dt + \sigma dB_t \quad \Rightarrow \quad X_t = X_0 e^{-\mu t} + \kappa(1 - e^{-\mu t}) + \int_0^t \sigma e^{-\mu(t-s)} dB_s,$$

where μ and σ are the same parameters as in (14) and so is the Brownian motion B_t; in addition, κ is the mean to which the process will revert (in the long run). Clearly, $\mathsf{E}(X_t) = \mathsf{E}(X_0)e^{-\mu t} + \kappa(1-e^{-\mu t}) \to \kappa$ when $t \to \infty$. We plan to extend both the mean-variance hedging and the shortfall hedging models, allowing the asset price X_t to follow the OU process instead of the GBM in (14). Even more generally, we want to consider asset-price models following general diffusion processes.

7 Concluding Remarks

Our approaches to deriving the optimal hedging strategies outlined above are "duality" based. Consider shortfall hedging in §5.2: instead of solving the original problem, we consider its "dual", the lower-bound problem, which is easy (or at least, easier) to solve. However, to satisfy primal feasibility, and thereby closing the duality gap, is the key. And that is represented by identifying the additional constraint $\mathsf{E}^M(V_T) \leq 0$. Indeed in a more abstract setting involving general utility functions as objectives, this falls into so-called convex-duality approach; refer to Föllmer and Leukert (2000) and Pham (2002).) At a general level, this is also very much in the same spirit as the approach outlined in §5.1, where instead of directly solving $B(m, Q)$, we solve its conjugate dual $A(\lambda)$.

This simple and elegant idea has been effectively used in other applications. Consider, for instance, the real-time scheduling of server(s) among different job classes in a queueing network. Finding the optimal control policy (the primal problem) can be solved in two steps: (i) first solve a dual, optimization problem to find the best "state", an equilibrium measure, in which to operate the system; (ii) then find a control policy that drives the system to the best state. It turns out that the dual problem in step (i) has a very appealing polymatroid structure, which not only facilitates finding the (dual) optimal solution, but also suggests how to enforce primal feasibility. For a broad class of queueing systems, this primal feasibility turns out to take the form of so-called *strong conservation laws* (or its variation and generalization). Refer to Shanthikumar and Yao (1992); also, Chen and Yao (2001), Yao (2002).

For both mean-variance and shortfall objectives, we have measured the terminal wealth using the *physical* (or "real world") measure P (as opposed to the martingale, or risk-neutral, measure P^M). Recall, the terminal wealth has two components, one from production (H_T), and the other from hedging/trading (χ_T). It goes without saying that the former should be measured by P; after all, production contributes real wealth. In the same spirit, hedging/trading also leads to gain/loss that adds to (or subtracts from) the production wealth. Thus, measuring χ_T by P is only natural, and consistent as well. This said, it is also worth noting that the martingale measure P^M does come out at the right places in our models. In the mean-variance hedging, V_t is a P^M-martingale, and a key component of the hedging strategy, ξ_t, is derived from the martingale representation of V_t; refer to (20) and (22). The same applies to the the shortfall model, where the role of P^M is even more striking: as noted before, it is the constraint $\mathsf{E}^M(V_T) \leq 0$ in (25) that closes the duality gap of the lower-bound problem by enforcing no-arbitrage.

References

Anthony, M. and MacDonald, R. 1998. On the mean-reverting properties of target zone exchange rates: Some evidence from the ERM. *European Economic Review,* **42(8)**, 1493-1523.

Basak, S. and Chabakauri, G. 2010. Dynamic Mean-Variance Asset Allocation. *The Review of Financial Studies,* **23(8)**, 2970-3016.

Boyabatli, O. and Toktay, B. 2004. Operational hedging: A review with discussion. *Research Collection Lee Kong Chian School Of Business.*

Caldentey, R. and Haugh, M. 2006. Optimal control and hedging of operations in the presence of financial markets. *Mathematics of Operations Research,* **31(2)**, 285-304.

Casassus, J. and Collin-Dufresne, P. 2005. Stochastic convenience yield implied from commodity futures and interest rates. *Journal of Finance,* **60(5)**, 2283-2331.

Chen, H. and Yao, D.D. 2001. *Fundamentals of Queueing Networks: Performance, Asymptotics and Optimization.* Springer-Verlag, Applications of Mathematics Series, vol. 46, , New York.

Chen, L.G., Long, D.Z. and Perakis, G. 2015. The impact of a target on Newsvendor decisions. *Manufacturing & Service Operations Management,* **17(1)**, 78-86.

Chen, X., Sim,M., Simchi-Levi, D. and Sun, P. 2007. Risk aversion in inventory management. *Operations Research,* **55(5)**, 828-842.

Chen, Y., Xu, M. and Zhang, Z.G. 2009. A risk-averse Newsvendor model under the CVaR criterion. *Operations Research,* **57(4)**, 1040-1044.

Chod, J., Rudi, N. and Mieghem, J. 2010. Operational flexibility and financial hedging: complements or substitutes. *Management Science,* **56(6)**, 1030-1045.

Choi, S., A. Ruszczynski, and Y. Zhao. 2011. A multiproduct risk-averse Newsvendor with law-invariant coherent measures of risk. *Operations Research,* **59(2)**, 346-364.

Chowdhry, B. and Howe, J. 1999. Corporate risk management for multinational corporations: financial and operational hedging policies. *Review of Finance,* **2(2)**, 229 - 246.

Ding, Q., Dong, L. and Kouvelis, P. 2007. On the integration of production and financial hedging decisions in global markets. *Operations Research,* **55(3)**, 470-489.

Fishburn, P. 1977. Mean-risk analysis with risk associated with below-target returns. *American Economic Review,* **67(2)**, 116 - 126.

Föllmer, H. and Leukert, P. 1999. Quantile hedging. *Finance and Stochastics,* **3(3)**, 251-273.

Föllmer, H. and Leukert, P. 2000. Efficient hedging: cost versus shortfall risk. *Finance and Stochastics,* **4(2)**, 117-146.

Gao,F., Chen, F. and Chao, X. 2011. Joint optimal ordering and weather hedging decisions: mean-CVaR model. *Flexible Services and Manufacturing Journal,* **23(1)**, 1-25.

Gaur, V. and Seshadri, S. 2005. Hedging inventory risk through market instruments. *Manufacturing & Service Operations Management,* **7(2)**, 103-120.

Gropp, J. 2004. Mean reversion of industry stock returns in the US, 1926–1998. *Journal of Empirical Finance,* **11(4)**, 537-551.

Gourieroux, C., Laurent, J. and Pham, H. 1998. Mean-variance hedging and numeraire. *Mathematical Finance,* **8(3)**, 179-200.

GUTMANN, H.M. 2001. A radial basis function method for global optimization. *Journal of Global Optimization,* **19(3)**, 201-227.

GUYON, I. AND ELISSEEFF, A. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* **3**, 1157-1182.

HARLOW, W. AND RAO, R. 1989. Asset pricing in a generalized mean-lower partial moment framework: theory and evidence. *Journal of Financial and Quantitative Analysis,* **24(3)**, 285-311.

HASTIE, T., TIBSHIRANI, R. AND FRIEDMAN. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag New York.

HULL, J.C. 2014. *Options, Futures, and Other Derivatives*, 9th ed., Prentice Hall, Upper Saddle River, New Jersey.

JONES, D.R., SCHONLAU, M. AND WELCH, W.J. 1998. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization,* **13(4)**, 455-492.

KOUVELIS, P., PANG, Z. AND DING, Q. 2013. Dynamic financial hedging strategies for a storable commodity with demand uncertainty. Working Paper, Olin School of Business, Washington University in St. Louis, St. Louis, MO.

LARSEN, K.S. AND SRENSEN, M. 2007. Diffusion models for exchange rates in a target zone. *Mathematical Finance,* **17(2)**, 285 - 306.

LEUNG, T. AND LI, X. 2015. Optimal mean reversion trading with transaction costs and stop-loss exit. *International Journal of Theoretical & Applied Finance,* **18(3)**.

LUENBERGER, D. 1998. *Investment Science,* Oxford University Press, New York.

MARTINEZ-DE-ALBENIZ, V. AND SIMCHI-LEVI, D. 2006. Mean-variance trade-offs in supply contracts. *Naval Research Logistics,* **53(7)**, 603-616.

MARKOWITZ, H.M. 1987. *Mean-variance analysis in portfolio choice and capital markets.* Basil Blackwell, New York.

Pham, H. 2002. Minimizing shortfall risk and applications to finance and insurance problems. *Annals of Applied Probability,* **12(1)**, 143-172.

Pham, H. 2009. *Continuous-time Stochastic Control and Optimization with Financial Applications*, Springer-Verlag, Berlin/Heidelberg.

Poterba, J.M. and Summers, L.H. 1988. Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics,* **22(1)**, 27-59.

Rockafellar, R. T. and S. Uryasev. 2000. Optimization of conditional Value-at-Risk. *Journal of Risk,* **2(3)**, 21 - 42.

Rockafellar, T.R. and Uryasev, S. 2002. Conditional Value-at-Risk for general loss distributions. *Journal of Banking & Finance* **26**, 1443-1471.

Schweizer, M. 1992. Mean-variance hedging for general claims. *Annals of Applied Probability,* **2(1)**, 171-179.

Shanthikumar, J.G. and Yao, D.D. 1992. Multiclass queueing systems: polymatroid structure and optimal scheduling control. *Operations Research,* **40(3)**, 293-299.

Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B,* **58(1)**, 267-288.

Van Mieghem, J. A. 2010. Risk Management and Operational Hedging: An Overview. In: *Handbook of Integrated Risk Management in Global Supply Chains* (eds P. Kouvelis, L. Dong, O. Boyabatli and R. Li), Chapter 2, 13 - 48, John Wiley & Sons, Inc.

Wang, L. and Yao, D.D. 2017. Production with risk hedging: optimal policy and efficient frontier. *Operations Research,* **65(4)**, 1095-1113.

Wang, L. and Yao, D.D., 2019. Production planning with shortfall hedging under partial information and budget constraint. Working Paper, Faculty of Business and Economics, The University of Hong Kong, Hong Kong.

Wang, L. and Wissel, J. 2013. Mean-variance hedging with oil futures. *Finance and Stochastics,* **17(4)**, 641-683.

Yao, D.D. 2002. Dynamic scheduling via polymatroid optimization. In: *Performance Evaluation of Complex Systems: Techniques and Tools,* M.C. Calzarossa and S. Tucci (eds.), Springer-Verlag, , 89 - 113.