A Production-based Model for Predicting Heating Oil Prices

James R. Porter and Robin O. Roundy*

School of OR&IE, Cornell University, Ithaca, NY 14853

Abstract

We attempt to determine the impact of production decisions on the price of heating oil. A model for the production of distillate fuel oil is proposed. Its solution, obtained using stochastic dynamic programming, closely matches history. By perturbing the problem in a deliberate manner, we can determine the value of additional units of inventory by examining differences in costs. Using these differences, we define a production futures price for extra units of distillate fuel oil. We then use linear regression to determine if these production futures prices have an impact on actual heating oil futures prices.

Keywords: dynamic programming, heating oil, futures prices, distillate fuel oil, convenience yield

*Corresponding author. Tel: 607-255-9137; fax: 607-255-9129; email: robin@orie.cornell.edu

1 Introduction

Distillate fuel oil is one of several petroleum products refined from crude oil, including motor gasoline, kerosene, jet fuel, and residual fuel oil. The most common distillates are diesel engine fuel and heating oil. Unfortunately, separate inventory, production, and demand data for each particular distillate is not readily available. Thus, the data used in our work will be total distillate fuel oil. While the largest use of distillate is as a fuel for diesel engines, the amount used for heating oil is substantial. Unlike diesel engine fuel, the demand for heating oil is greatly affected by seasonal factors. This seasonality can be detected when observing data for total distillate fuel oil.

Demand for heating oil is high during the winter and almost nonexistent during the summer. As a result, producers seek to hold large inventories during the winter. Figure 1 displays first-of-the-month inventory levels for total distillate fuel oil in the U.S. during 1988-1995. To maintain these levels, producers begin to accumulate stocks of heating oil

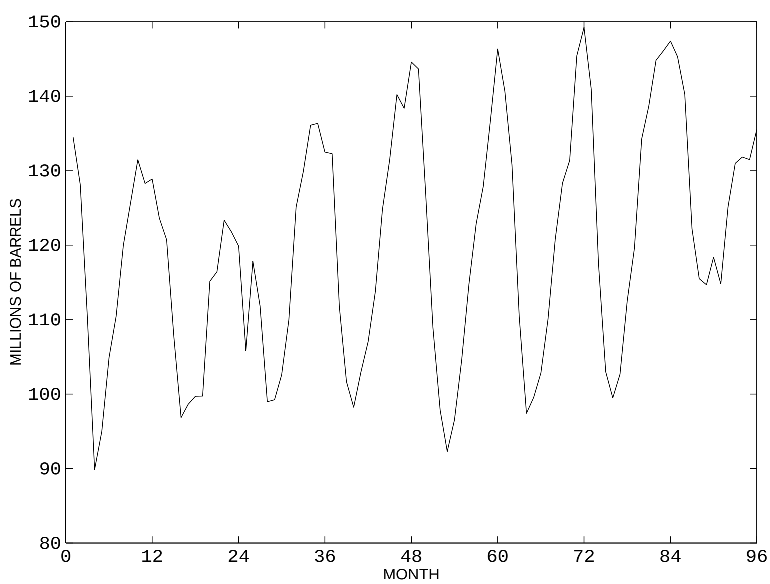

Figure 1: Total Distillate Fuel Oil Inventory, 1988-1995

during the summer and fall. Production exceeds demand, and inventory levels rise. These stocks will be stored for later use. When winter begins, demand exceeds production, and stocks built up earlier can be used along with winter production to meet demand. As winter demand is met, inventory levels fall.

We will be concerned with investigating how this seasonal pattern in distillate fuel oil data influences the price of heating oil. Since November 1978, there has been active trade in heating oil futures contracts on the New York Mercantile Exchange (NYMEX). Each contract calls for the holder of a long position to receive one thousand 42-gallon barrels of No. 2 heating oil during a pre-specified delivery month. There is no fixed delivery date. Rather, there is a delivery period of several weeks during the delivery month. The last trading day of a particular futures contract is the last business day of the month preceding the delivery month. It should also be noted that there are several active spot markets for heating oil, including one located in New York harbor. The primary goal of this research will be to determine how the production and storage of distillate fuel oil influences these prices for heating oil.

In the next section, the theory of commodity storage is discussed along with some of the relevant works in the literature. Section 3 proposes a discrete-state stochastic dynamic programming problem for distillate fuel oil production. The features of the model are discussed in detail and its solution is presented. Finally, Section 4 discusses how to forecast prices using production costs. The concept of a production futures price is discussed, and this quantity is used to help predict heating oil futures prices and spreads.

2 The Theory of Commodity Storage

Suppose we have a frictionless economy with a series of discrete trading dates $t \in [0, 1, 2, ...\tau]$. Consider a commodity which can be purchased at any date t for immediate delivery at spot price S_t. Further, assume investors can trade in a term structure of futures contracts on the commodity at any date. Let $F_{t,T}$ denote the price at time t of a futures contract maturing at date $T \in [t, \tau]$, where $F_{t,t} = S_t$. Finally, we will assume that interest rates are deterministic, and let $r_{t,T}$ represent the riskless interest rate over the period $[t, T]$. Then following Fama and French [11], the *cost of carry* relationship states that

$$F_{t,T} = S_t(1 + r_{t,T}) + w_{t,T} - c_{t,T}, \tag{1}$$

where $w_{t,T}$ represents the marginal storage cost and $c_{t,T}$ represents the marginal convenience yield from an additional unit of inventory over $[t, T]$. To better understand its meaning, let us write this formula in a slightly different way:

$$F_{t,T} - S_t = r_{t,T}S_t + w_{t,T} - c_{t,T}. \tag{2}$$

The above states that the return from buying the commodity spot at date t and selling it for future delivery at date T is equal to the sum of interest and storage costs less the convenience yield. The right-hand side of equation (2) is referred to as the cost of carry.

The *convenience yield* reflects the benefits of ownership of the commodity that are not realized by the holder of a futures contract. Perhaps the most important benefit is the ability to meet unexpected demand. Unlike interest rates or holding costs, the convenience yield is not directly observable. Given each of the other four quantities in equation (1), $c_{t,T}$ is chosen to preserve the equality.

We will let $y_{t,T}$ represent the cost of carry, so that

$$y_{t,T} = r_{t,T} S_t + w_{t,T} - c_{t,T}. \tag{3}$$

Substituting into equation (2) gives

$$F_{t,T} - S_t = y_{t,T}. \tag{4}$$

In the case where $T = t+1$, we refer to $y_{t,t+1}$ as the *instantaneous cost of carry.* It represents the cost incurred in the next instant of time. Note that we can also write

$$F_{t,T+1} - S_t = y_{t,T+1}. \tag{5}$$

Taking the opposite of equation (4) and adding it to equation (5) yields

$$F_{t,T+1} - F_{t,T} = y_{t,T+1} - y_{t,T}. \tag{6}$$

The left-hand side of the above equation is a *futures price spread.* It is the difference in price between two futures contracts on the same asset which mature during adjacent time periods. The right-hand side is known as the $(T-t)$-periods ahead *forward cost of carry.* It represents the cost, perceived at time t, of carrying the commodity during the interval $[T, T+1]$.

In the literature, the theory of storage begins with Kaldor[13]. He believed that stocks of a commodity possess a yield which must be deducted from carrying costs. One justification given for this yield was that holders of stocks could make use of them whenever they wished, a benefit Kaldor referred to as convenience. He also proposed that this yield should vary inversely with stock level. Working [18], [19] studied the Chicago wheat market. His investigations led him to believe in the existence of convenience yields which could lead to a negative net price of storage. This negative price of storage would prevail when the amount of wheat supplied was low. Telser [17], studying the storage of wheat and cotton, found support for the existence of convenience yields. He discovered that the futures price spread for these goods was heavily influenced by the seasonal pattern of their stocks. Brennan [6] studied the convenience yield for several agricultural commodities, including eggs, cheese, butter, wheat, and oats. Using linear regression, he saw significant relationships between convenience yields and stock levels. More recently, econometric techniques have been employed to study the theory of storage. Many of these analyses consider distillate fuel oil. Fama and French [11] study futures and spot prices for twenty-one commodities. They note that seasonal factors often play an important role in explaining the convenience yield, and they find some evidence suggesting that futures prices can forecast future spot prices. Bopp and Sitzer [5] focus on distillate fuel oil and argue that heating oil futures prices influence the current spot price. Lowry [14] builds a production model that contains a convenience yield arising from precautionary storage which is a quadratic function of inventory level. He applies his model to distillate fuel oil and finds support for his convenience yield specification. Finally, the theory of storage has also played an important role in many asset valuation models. The works of Gibson and Schwartz [12], Cortazar and Schwartz [8], and Miltersen and Schwartz [15] rely on a stochastic representation for the convenience yield in order to price financial instruments like options.

3 A Model for Distillate Fuel Oil Production

3.1 Overview

The goal of this section is to propose and solve a discrete-state stochastic dynamic programming problem for monthly U.S. production of distillate fuel oil during 1988-1995. The use of dynamic programming is appropriate since decisions regarding how much to produce in the current time period will directly affect future time periods. With the objective of minimizing refinery costs, our goal will be to adjust the parameters of the model to fit historical inventory levels. The stochastic nature of the problem is derived from the uncertainty in demand. Before discussing the problem in detail, we will briefly discuss a simple method for forecasting distillate fuel oil demand.

3.2 Demand Forecasting with Winters' Method

As stated in the introduction, demand for distillate fuel oil is highly seasonal. Thus, a seasonal time series method seems to be an appropriate choice for generating forecasts. We will use an exponential smoothing model known as Winters' additive method. The details of this procedure will follow those given in Abraham and Ledolter [1].

Winters suggests a linear trend model with seasonal indicators. If we let D_t represent the variable of interest, then we can write

$$D_{t+j} = T_{t+j} + S_{t+j} + \epsilon_{t+j}. \tag{7}$$

The trend component is assumed to be linear with $T_{t+j} = \mu_{t+j} = \mu_t + \beta_t \cdot j$. Assuming that the length of the seasonal period is d, we have d seasonal factors which satisfy

$$S_i = S_{i+d} = S_{i+2d} = \ldots \text{ for } i = 1, 2, \ldots d \text{ and } \sum_{i=1}^{d} S_i = 0. \tag{8}$$

The error terms, ϵ_t, are assumed to be independent and identically distributed normal random variables with mean 0 and variance σ^2.

Given the information available, an estimate of the level of the series at time $t+1$, $\hat{\mu}_{t+1}$, can be found in two different ways. We can compute the current observation adjusted by its estimated seasonal factor, $D_{t+1} - \hat{S}_{t+1-d}$. Also, we can compute the trend estimate, $\hat{\mu}_t + \hat{\beta}_t$, where $\hat{\beta}_t$ is the slope estimate at time t. The Winters' technique employs a weighted average of these two estimates to obtain an estimate of the level at time $t+1$:

$$\hat{\mu}_{t+1} = \alpha_1(D_{t+1} - \hat{S}_{t+1-d}) + (1 - \alpha_1)(\hat{\mu}_t + \hat{\beta}_t), \tag{9}$$

where α_1 is a smoothing parameter satisfying $0 \leq \alpha_1 \leq 1$. Next, we can compute an update of the slope estimate, $\hat{\beta}_{t+1}$, using two different pieces of information. We will utilize the current estimate of the slope, $\hat{\mu}_{t+1} - \hat{\mu}_t$, as well as the previous estimate of the slope, $\hat{\beta}_t$. We again compute a weighted average using another smoothing parameter to estimate the slope at time $t+1$:

$$\hat{\beta}_{t+1} = \alpha_2(\hat{\mu}_{t+1} - \hat{\mu}_t) + (1 - \alpha_2)\hat{\beta}_t. \tag{10}$$

It is also possible to update the seasonal coefficient, $\hat{S}_{t+1}$, using the current estimate of the seasonal factor, $D_{t+1} - \hat{\mu}_{t+1}$, and the previous estimate, $\hat{S}_{t+1-d}$. Using a third smoothing parameter, we compute:

$$\hat{S}_{t+1} = \alpha_3(D_{t+1} - \hat{\mu}_{t+1}) + (1 - \alpha_3)\hat{S}_{t+1-d}. \tag{11}$$

The seasonal factors are updated once each full season. For reasons which will become clear later, however, this will not be done. In our model, *we will assume that the seasonal factors remain constant through time.* This is tantamount to choosing $\alpha_3 = 0$.

After obtaining estimates for the level, slope, and seasonal components, it is possible to compute a forecast for a future value, say D_{t+k}, from time origin t by calculating

$$\hat{D}_{t+k} = \hat{\mu}_t + \hat{\beta}_t \cdot k + \hat{S}_{t+k-d}. \tag{12}$$

3.3 The Stochastic Dynamic Programming Problem

3.3.1 Assumptions and Notation

We will begin by describing the events which take place in a typical time period. Initially, we will assume that producers create a forecast for demand using Winters' additive method. Based on this forecast and the current inventory level, a decision is made to replenish stocks. The new inventory amount, which is the sum of beginning period inventory and current period production, is referred to as the *order-up-to quantity.* After receiving new inventory, we then observe demand. We assume that demand is never backlogged, and that there are no lost sales due to consumers going elsewhere to purchase the product. Instead, in the case of unexpectedly large demand, it is assumed that producers can acquire enough of the product to meet demand at an extremely high cost. This procedure will be referred to as *emergency acquisition*, and it will ensure that beginning period inventory is always nonnegative. Finally, at the end of the time period, cost is incurred and a new inventory level is observed.

An additional issue to be considered is the existence of imports and exports. In our model, we will treat imports as part of production and exports as part of demand. This assumption seems appropriate due to their relative small size during the time period of interest.

It is now appropriate to introduce some notation. Let $X_t \geq 0$ represent the inventory level at the beginning of time period t. The order-up-to quantity for period t will be denoted by Y_t, which must satisfy $Y_t \geq X_t$. Demand will be denoted by D_t. Based on this notation, note that current period production is $Y_t - X_t$, and we have the following inventory equation:

$$(X_t + (Y_t - X_t) - D_t)^+ = (Y_t - D_t)^+ = X_{t+1}. \tag{13}$$

The quantity X_{t+1} is both the ending period inventory for time t and the beginning period inventory for time $t+1$. The positive part arises since emergency acquisition prevents beginning period inventory from falling below zero.

We will now give a general description of the dynamic programming problem. At the beginning of time period t, the state of the system, $\mathbf{s_t}$, is the following vector:

$$\mathbf{s_t} := (N\mu_t, N\beta_t, NX_t).$$

Each component of the vector above is known as a *level* of the state space. These quantities are integer-valued. Corresponding to these levels are actual values of the variables. The first two components are related to the level and slope components in Winters' method. Each possible $(N\mu_t, N\beta_t)$-pair will give rise to a probability distribution for demand. This will be explained in detail later in this chapter. The third component, NX_t, is a potential beginning period inventory level for time t. The decision variable for the problem is denoted by NY_t. It corresponds to the order-up-to level for time t and is measured on the same scale as NX_t. Also, let r_t be the discount factor for time t satisfying $0 \leq r_t \leq 1$. Following the format given in Bertsekas[4], we can denote the cost incurred in stage t by $f_t(\mathbf{s_t}, NY_t(\mathbf{s_t}), \epsilon_t)$, where

ϵ_t is a random disturbance term due to the uncertainty in demand. The total discounted cost incurred along any sample path will be

$$\sum_{t=1}^{N-1} r_t f_t(\mathbf{s_t}, NY_t(\mathbf{s_t}), \epsilon_t) + r_N f_N(\mathbf{s_N}),$$

where $f_N(\mathbf{s_N})$ is the end-of-horizon cost incurred in the final time period N. Due to the stochastic nature of the problem, our goal will be to choose decisions that minimize expected discounted cost:

$$E[\sum_{t=1}^{N-1} r_t f_t(\mathbf{s_t}, NY_t(\mathbf{s_t}), \epsilon_t) + r_N f_N(\mathbf{s_N})].$$

We will next discuss the structure of the state space in detail.

3.3.2 State Space

Recall that each component of the state space vector is an integer-valued level. Let N_μ, N_β, and N_X, respectively, represent the number of levels for each state space variable. In any time period, the components of the state space vector must satisfy $1 \leq N\mu_t \leq N_\mu$, $1 \leq N\beta_t \leq N_\beta$, and $1 \leq NX_t \leq N_X$. For each time period, this results in $N_\mu \times N_\beta \times N_X$ possible states. It should now be apparent why the seasonal component from Winters' method is not included in the state space. Depending on the length of the seasonal period, this could cause the size of the state space to explode.

For each of the state space variables, a range of values needs to be chosen. These ranges should be selected so that the state space will encompass values of the variables which are likely to be observed. There are two issues to be considered: the length of the range and its location. While the length of these ranges will remain fixed, their locations will be allowed to shift through time. First, denote the range lengths by $R\mu$, $R\beta$, and RX, respectively.

Note that given these range lengths and the number of levels, it is then possible to define state space variable *increments*:

$$I\mu = \frac{R\mu}{(N_\mu - 1)}. \tag{14}$$

$$I\beta = \frac{R\beta}{(N_\beta - 1)}. \tag{15}$$

$$IX = \frac{RX}{(N_X - 1)}. \tag{16}$$

To clarify, levels NX_t and $NX_t + 1$, for example, correspond to values of the variable X_t which are exactly IX units apart.

Let us now discuss the problem of locating the endpoints for the range of values associated with the state space variables. These endpoints will be allowed to vary through time. Let us let $L\mu_t$, $L\beta_t$, and LX_t, respectively, denote the left endpoints for each of the variables. We will assume that each of these quantities is specified as a multiple of its corresponding state space variable increment. Given these quantities and the increments, we can now associate each level of a state space variable with its actual value:

$$\mu_t = L\mu_t + (N\mu_t - 1) \cdot I\mu. \tag{17}$$

$$\beta_t = L\beta_t + (N\beta_t - 1) \cdot I\beta. \tag{18}$$

$$X_t = LX_t + (NX_t - 1) \cdot IX. \tag{19}$$

Consider the variable X_t, for example. At its lowest level, $NX_t = 1$, the value of the variable is simply its left endpoint, $X_t = LX_t$. At its highest level, $NX_t = N_X$, the variable is equal to the right endpoint of the range, $X_t = LX_t + (N_X - 1) \cdot IX$. Note that each value is a multiple of IX.

The evolution of the state space over time is governed by transition probabilities computed under the assumption that demand is normally distributed. After creating a forecast for demand, $\hat{D}_t$, using equation (12), we allow the error to range from -2σ to 2σ. Thus, we believe that actual demand is likely to fall in the interval $[\hat{D}_t - 2\sigma, \hat{D}_t + 2\sigma]$. To obtain possible values for demand, we round the values in these intervals and choose boundaries known as cut points which separate adjacent values. This procedure is also done in order to get updated values for μ_t and β_t. Probabilities are calculated by considering the distance between adjacent cut points. In this way, a distribution is constructed which will enable us to compute the expected discounted cost.

3.3.3 Refinery Costs

The first cost facing refiners considered in our model is known as a *production efficiency cost*. We will assume that there is some target level of production at which cost is eliminated. Deviations from this level will result in a quadratic penalty.

From Subsection 3.3.1, we know that period t production is represented by $Y_t - X_t$, the order-up-to quantity less beginning period inventory. We need to give an equation relating Y_t with its level NY_t. Our definition must be consistent with the inventory equation (13). Thus, we define

$$Y_t = LX_t + (NY_t - 1) \cdot IX. \tag{20}$$

Now we can formally define the production efficiency cost:

$$Pr(Y_t - X_t) = c \cdot ((Y_t - X_t) - L)^2. \tag{21}$$

The parameter $c > 0$ is a cost coefficient, while the parameter $L > 0$ represents the target level for production. In our model, discounted per unit production costs are accounted for by adjusting holding cost parameters. This approach is justifiable due to the inventory equation (13) assumed to hold true.

The second type of cost facing refiners will be an inventory *holding cost*. This cost will be charged on the amount of beginning period inventory. We will assume that holding costs are directly proportional to the amount of inventory held:

$$H(X_t) = h \cdot X_t. \tag{22}$$

The parameter $h > 0$ represents the per unit holding cost.

The final type of cost in our model will be a *pipeline efficiency cost*. Here, the term 'pipeline' refers to the primary product distribution system through which finished products leave the refinery. As discussed in [16], this system consists of actual pipelines, storage tanks, barges, tankers, and tank cars and trucks. As Lowry [14] notes, a large amount of product must be stored in order to maintain smooth operation of the distribution system while avoiding stockouts. For instance, a certain amount of product is necessary simply to keep specific parts of the system functioning. Pipelines, for example, must be filled in order to operate normally. The bottoms of tanks are designed so that they are never empty, preventing residue from settling at the bottom. The possibility that a refinery might have to be shut down for a period of time is another reason to accumulate stocks. The shut down might be due to a planned maintenance period, or it could be the result of an unexpected system interruption. Finally, stocks will be held in order to accommodate seasonal demand. As noted in Section 1, stocks of distillate fuel oil are built up and held during the summer in order to meet the demand for heating oil in the winter.

Our pipeline efficiency cost will have two components. The first will be the emergency acquisition cost incurred when demand can not be satisfied. We will assume that producers make unplanned purchases of the product at an extremely high cost to meet demand. The second component will be a low inventory cost. Below a certain threshold, the amount of stocks held will be insufficient to maintain smooth operation of the distribution system. In our model, producers will incur cost at a quadratic penalty when stocks fall below this level. We will model the pipeline efficiency cost as a function of the inventory position after demand is observed, denoted by Z_t, where

$$Z_t = Y_t - D_t. \tag{23}$$

Our pipeline efficiency cost function is as follows:

$$Pl(Z_t) = [M \cdot Z_t^-] + ba^2 \vee (b \cdot [(a - Z_t)^+]^2). \tag{24}$$

The first term in brackets is the emergency acquisition cost, while the remainder represents the low inventory cost. The quantities $M, a, b > 0$ are cost function parameters. Before discussing the form of the function, it will be helpful to view its graph. Consider Figure 2. When $Z_t \geq a$, no cost is incurred. As inventory falls below the threshold of a, a quadratic penalty for low inventory is imposed on the shortfall with scale parameter b. At $Z_t = 0$, the cost is ba^2. In the case where $Z_t < 0$, an emergency acquisition cost of M per unit is incurred, where it is assumed that $M > 2ab$.

An additional point needs to be considered. Note that the cost is a function of both the order-up-to quantity and demand. For computational reasons, it will be convenient to express the cost as a function of just the order-up-to quantity. Thus, we will compute the expected value of the pipeline efficiency cost:

$$EPl(Y_t) = E_D[Pl(Y_t - D_t)]. \tag{25}$$

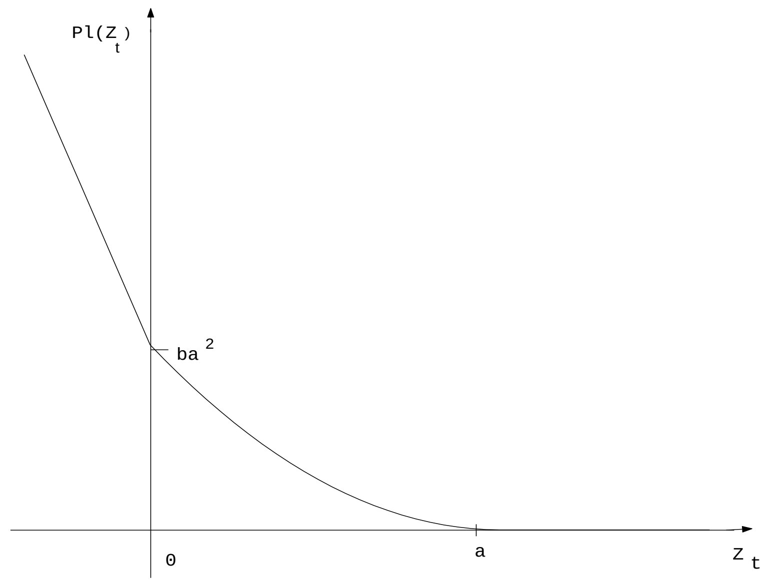

Figure 2: Pipeline Efficiency Cost

3.3.4 The Minimization Problem for a Single Stage

As discussed at the end of Subsection 3.3.1, our goal is to choose an order-up-to level that minimizes expected discounted cost. Using dynamic programming, we begin at the end of the time horizon and move backward, solving a separate optimization problem at each time period for every state. For convenience, let us write the cost incurred at any stage, f_t, simply as a function of the state: $f_t = f_t(N\mu_t, N\beta_t, NX_t)$. Taking into account the refinery costs discussed in the previous subsection, our minimization problem for $t \neq N$ is:

$$
\begin{aligned}
& f_t(N\mu_t, N\beta_t, NX_t) = H(X_t) + \\
& \min_{NY_t \geq NX_t} [Pr(Y_t - X_t) + EPl(Y_t) + \\
& r_t E_D(f_{t+1}(N\mu_{t+1}, N\beta_{t+1}, (NY_t - ND_t)^+))]. \qquad (26)
\end{aligned}
$$

The quantity r_t is the discount factor for time t, and the expectation is performed with respect to the distribution for demand. When $t = N$, the end-of-horizon cost must be

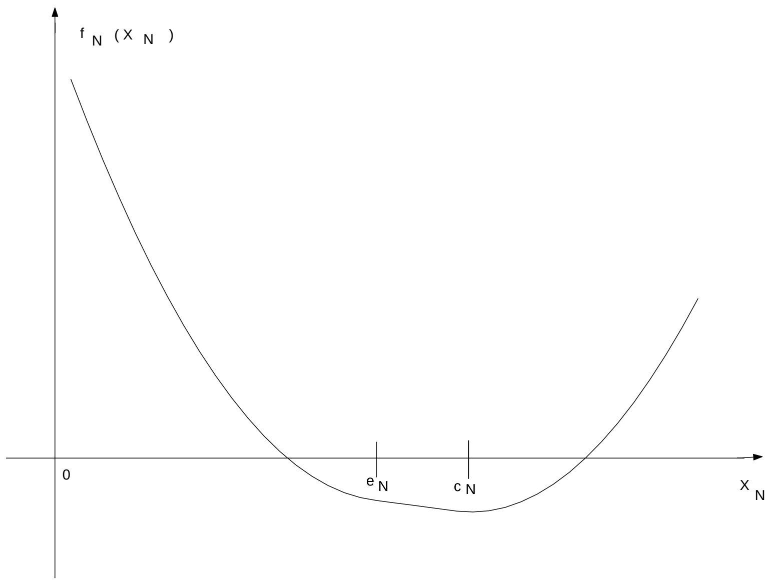

Figure 3: End-of-Horizon Cost

computed in order to begin the dynamic programming recursion. We will assume that this cost is a function of inventory only so that $f_N(N\mu_N, N\beta_N, NX_N) = f_N(NX_N)$. Written as a function of X_N, we will model these costs using the following expression:

$$f_N(X_N) = -a_N X_N + b_N \cdot [(X_N - c_N)^+]^2 + d_N \cdot [(e_N - X_N)^+]^2. \tag{27}$$

We have a cost function with five parameters. To better understand this formulation, it is perhaps beneficial to view a graph of the function. Consider Figure 3. Note that c_N and e_N represent potential end-of-horizon inventory positions, with $c_N > e_N > 0$. Along the interval $[e_N, c_N]$, we will assume that each unit of inventory has a marginal value of $a_N > 0$, resulting in a negative component to the cost function. When inventory falls below e_N, we will assume that stocks are insufficient to allow for smooth operations at the refinery. A quadratic penalty will be imposed on the shortfall amount, scaled by $d_N > 0$. When inventory is greater than c_N, producers have an excess of product. A quadratic penalty is imposed on the amount of excess, scaled by $b_N > 0$. Note that a positive constant can be added to this function without changing its shape in order to make the cost nonnegative for all values of X_N. Since it will be easier to deal with positive costs in the dynamic programming problem, we will make use of this fact in our implementation.

Before discussing the solution, it is necessary to consider the issue of extrapolation. Recall from the definition of the state space in Subsection 3.3.2 that the components of the state space vector satisfy $1 \leq N\mu_t \leq N_\mu$, $1 \leq N\beta_t \leq N_\beta$, and $1 \leq NX_t \leq N_X$. During the computation of the solution to the problem, there is a possibility that a cost will be required for a vector that does not satisfy these inequalities. In these instances, we will perform a *linear extrapolation* to compute the cost. We will also monitor the amount of extrapolation performed as we solve the problem.

3.4 Solution of the Stochastic Dynamic Programming Problem

3.4.1 Introduction

In this section we will discuss the numerical solution of the dynamic programming problem in detail. We will use monthly data for beginning period inventory, demand, and production for total U.S. distillate fuel oil during 1988-1995 as reported in the *Basic Petroleum Data Book*[2]. Thus, we will implement our model under the assumption that decisions were made by a single refiner in the presence of this aggregate data. Values for quantities that define the state space will be selected to match those likely to have been chosen by the refiner. In each time period, a historical state will be identified. This particular state is the one that would have been observed by a refiner following our model. For these historical states, our goal will be to select values for cost function parameters that will cause the refiner to choose decisions that closely match historical decisions.

3.4.2 Parameter Estimation

Total demand for distillate fuel oil was forecasted using Winters' method that was described in Section 3.2. Recall that we needed to select smoothing parameters, denoted by α_1 and α_2, to help us update our estimates for the level and slope components. To obtain these smoothing parameters, we used the five-year period immediately preceding our period of interest as a base period. The values of the smoothing parameters that minimize the mean-squared error in forecasting will be considered optimal.

The base period consisted of the sixty monthly observations during 1983-1987 as reported in the *Monthly Energy Review*[10] and [2]. A seasonal period of $d = 12$ was assumed. The optimal values for the smoothing parameters were found to be $\alpha_1 = 0.20$ and $\alpha_2 = 0.04$. Initial estimates for the level and slope components were computed by performing a linear regression on an intercept and time. The procedure for estimating the seasonal factors and the variability in demand is located in the appendix.

For the state space, Table 1 records the values chosen for the time-independent quantities as defined in Subsection 3.3.2. The number of levels chosen results in 875 states for each time period. Range lengths were selected with the goal of being sufficiently wide that likely values for the variables would be contained within each one. The increment lengths are defined once the number of levels and range lengths are specified. The selection of the left endpoints of the ranges is discussed in the appendix.

Finally, since our goal is to produce a solution for the stochastic dynamic programming problem that closely matches history, it is necessary to identify the historical state and

Table 1: Time-independent State Space Parameter Values

Variable	# of Levels	Range Length	Increment Length
μ_t	$N_\mu = 7$	$R\mu = 24,000,000$	$I\mu = 4,000,000$
β_t	$N_\beta = 5$	$R\beta = 500,000$	$I\beta = 125,000$
X_t	$N_X = 25$	$RX = 60,000,000$	$IX = 2,5000,000$

decision for each time period. The details of this procedure are also located in the appendix.

3.4.3 Cost Parameter Estimation to Match History

We are now ready to discuss the solution of the dynamic programming problem. First, computer code was written in MATLAB to solve the problem. Then, some initial values were chosen for the parameters of the cost functions. In each time period, the minimization problem in Subsection 3.3.4 was solved using a variation of the Fibonacci method as described in Avriel[3]. Run time was approximately $9\frac{1}{2}$ hours on a PC with a Pentium II processor. After each iteration, the optimal decision for the historical state was compared to the historical decision. Upon analysis, the values of the cost parameters were modified in an attempt to bring these two decisions into close agreement for each stage. We will now discuss the parameter values chosen.

One quantity that needs to be defined initially is AMD. In this paper, AMD represents average monthly demand during the period 1988-1995. Although not necessary, we will find it convenient to express many of the parameters of our cost functions in terms of AMD. Its value is $AMD = 98,950,688$.

The parameters for the refinery costs were chosen by experimentation. The values selected are contained in Table 2 (Values for the end-of-horizon cost parameters are located in the appendix.). To help determine the appropriateness of our parameter selections, Figure 4

Table 2: Cost Parameter Values

Cost	Parameter Values
Production Efficiency	$c = \frac{0.6}{AMD^2}$, $L = AMD$
Holding	$h = \frac{0.015}{AMD}$
Pipeline Efficiency	$M = \frac{0.5}{AMD}$, $a = 1.27 \cdot AMD$, $b = \frac{0.09}{AMD^2}$

plots historical order-up-to quantities (unrounded) along with those predicted by our model. Before commenting on the graph, let us first present a table of errors measuring the difference

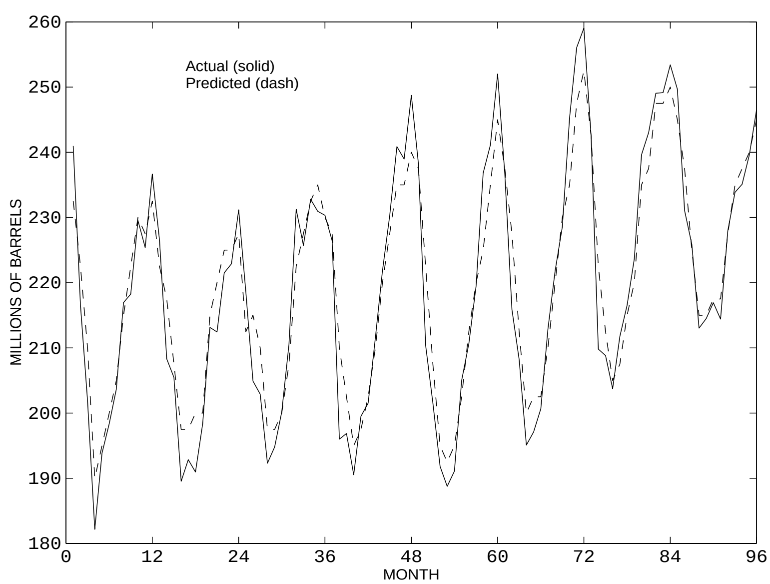

Figure 4: Actual and Predicted Order-up-to Quantities

between the historical order-up-to level and that predicted by our model. We will define this error by

$$Error = NY_t^H - NY_t^*, \tag{28}$$

where NY_t^H is the historical decision at time t and NY_t^* is the optimal decision for the historical state at time t. Table 3 records these errors for each month. The mean-squared

Table 3: Errors in Order-up-to Levels

Month	1988	1989	1990	1991	1992	1993	1994	1995
January	3	2	2	0	0	0	0	2
February	-2	-4	-4	-6	-5	-5	-5	-3
March	-3	-1	-3	-2	-2	-2	-1	0
April	-3	-3	-2	-2	-1	-2	-1	-1
May	0	-2	-1	1	-1	-2	2	0
June	-1	-4	0	0	-2	-1	1	0
July	-1	-1	1	1	1	1	1	-1
August	1	-1	3	1	-1	1	2	0
September	-2	-3	-1	1	0	-1	2	0
October	0	-1	0	2	5	4	1	-1
November	-1	-1	-2	2	2	3	1	0
December	2	1	0	3	3	3	1	1

deviation (MSD) of these errors is $MSD = 4.50$.

It is clear from both the graph and the table that our model does a good job of matching historical inventory positions. We are able to match the seasonal pattern in distillate fuel oil that is observed each year. There are two important points to note when considering the types of errors that are made. From Table 3, we can see that the model consistently underestimates the order-up-to level in December and January while overestimating it in February, March, and April. A potential remedy for this problem may be to allow time-dependent cost parameters in the model. For example, to encourage higher levels in December and January, the threshold parameter a in the pipeline efficiency cost could be shifted to the right in those months. Similarly, to decrease levels in February, March, and April, the holding cost parameter h could be increased. Since we wanted our cost parameters to remain stationary through time, this idea was not seriously considered.

Finally, it should be noted that the amount of extrapolation required to solve the problem was well within tolerable limits. This concludes our discussion of the stochastic dynamic programming problem.

4 Price Forecasting with Production Costs

4.1 Overview

In this section we will attempt to use our cost data from the optimal solution of the stochastic dynamic programming problem as an aid in predicting heating oil prices. A production futures price will be constructed by performing a slight alteration to the dynamic programming problem. We will then investigate to determine if these production futures prices have any value in predicting either actual heating oil futures prices or their price spreads. Linear regression models will be built to test for relationships between the variables. To begin with, we will define what is meant by a production futures price.

4.2 The Production Futures Price

Recall from Section 1 that a trader holding a long position in a futures contract on an asset will receive a pre-specified quantity of that asset during a pre-specified delivery date. In the context of our dynamic programming problem, we could give the refiner additional units of inventory in a future time period. Once this additional inventory is received, we would be interested in measuring its impact on costs not only in the current time period but in earlier time periods as well. We will define a new quantity that represents the value of receiving this additional inventory in the future.

First, suppose we wish to give our refiner additional inventory in time period t. Given our model, the most natural way to do this is to give the refiner exactly IX barrels at the same time that stocks are replenished. In essence, this has the effect of increasing the order-up-to level from NY_t to $NY_t + 1$. The minimization problem for stage t is now the following:

$$\begin{aligned} &f_t(N\mu_t, N\beta_t, NX_t) = H(X_t) + \\ &\min_{NY_t \geq NX_t} [Pr(Y_t - X_t) + EPl(Y_t + IX) + \\ &r_t E_D(f_{t+1}(N\mu_{t+1}, N\beta_{t+1}, (NY_t + 1 - ND_t)^+))]. \end{aligned} \tag{29}$$

Since this additional inventory is not present at the beginning of the period, the holding cost is unaffected. Also, the production efficiency cost is unaltered since this inventory was not actually produced. The expected pipeline efficiency cost and the discounted expected future cost will both be impacted by this additional inventory. After making this modification, we can solve the problem using the method and parameters given in Section 3. Clearly, we would expect this perturbation to have an impact on the optimal decision and cost in the current stage as well as in earlier stages.

Let us introduce some notation. Denote by T the time period in which IX barrels of extra inventory are given. Then, we can write the new cost for the historical state at time $t \leq T$ as $f_t^T(N\mu_t^H, N\beta_t^H, NX_t^H)$. Since we wish to measure the value of receiving this additional inventory, we can consider the following expression:

$$P_{t,T} = f_t(N\mu_t^H, N\beta_t^H, NX_t^H) - f_t^T(N\mu_t^H, N\beta_t^H, NX_t^H). \tag{30}$$

The right-hand side of equation (30) represents the difference between the cost observed for the usual model and that for the model with the extra inventory. This difference should represent the value at time t of additional inventory to be received at time T. For this reason, we refer to $P_{t,T}$ as the *production futures price* at time t for additional inventory in time period T. When $T = t$, we have a production spot price.

A point should be made regarding the name of $P_{t,T}$. It could be argued that this price should be referred to as a production *forward* price. It should be noted, however, that forward prices and futures prices are equivalent when interest rates are assumed to be deterministic. Since we make that assumption in this work, we will use the term 'futures' to describe this price.

We will be interested in comparing these production futures prices to actual heating oil futures prices. We note that the scale of the quantity $P_{t,T}$ is unknown. While an actual heating oil futures contract has a size of one thousand barrels, the number of barrels associated with the production futures price is 2.5 million. This must be considered when trying to relate the two prices. Also, unlike an actual futures price, it could be the case that $P_{t,T} < 0$.

This is not unrealistic since there may be certain times during the year when more inventory is undesirable. For instance, a refiner would not want to receive additional units of heating oil at the end of the heating season.

4.3 Predicting Heating Oil Futures Prices

Now that we have defined production futures prices, it is natural to be curious about their relationship to actual heating oil prices. We will build regression models in which the heating oil futures price will be the response. A set of independent variables, which includes the production futures price, will be used for forecasting.

Data for heating oil futures prices consists of first-of-the-month settlement prices for contracts traded on the New York Mercantile Exchange (NYMEX) during the period 1988-1995. We will consider contracts with time to maturity varying from one to five months. These contracts are by far the most liquid with respect to trading volume. A separate regression model will be built for each specific time to maturity. Prices are quoted in cents per gallon. Denote by $F^{HO}_{t,t+\Delta}$ the price at the beginning of period $t = 1, 2, \ldots 96$ for a futures contract on heating oil maturing at date $t + \Delta$, where $\Delta = 1, 2, 3, 4, 5$ represents time to maturity. We will restrict ourselves to contracts that mature during the eight-year period. Thus, as time to maturity increases by one month, we will lose one observation.

The most interesting independent variable in our model will be the production futures price. Defined in the previous section, we will denote these by $P_{t,t+\Delta}$ where $t = 1, 2, \ldots 96$ is the current time period and $t + \Delta, \Delta = 1, 2, 3, 4, 5$ is the period in which extra inventory is received.

In addition to the production futures prices, the contemporaneous crude oil futures price will be used as an explanatory variable. It seems proper to include the crude oil price since heating oil is refined from crude oil. This should compensate for the fact that our dynamic programming problem does not model the impact of this input to production. Crude oil futures contracts, like their heating oil counterparts, are also traded on NYMEX. Contracts with time to maturity varying from one to five months are heavily traded, and their settlement prices are readily available. We will let $F^{C}_{t,t+\Delta}$ represent the price in cents per gallon at the beginning of period $t = 1, 2, \ldots 96$ for a futures contract on crude oil maturing at date $t + \Delta$, where $\Delta = 1, 2, 3, 4, 5$.

Next, we will include in our regression models monthly indicator variables to control for seasonality. We will let I_{tj} represent these seasonal variables, which satisfy

$$I_{tj} = \begin{cases} 1 & \text{if observation } t \text{ is in period } j \\ 0 & \text{otherwise,} \end{cases}$$

for $t = 1, 2, \ldots 96$ and $j = 1, 2, \ldots 11$ (There are only eleven indicators since the model contains a constant term.).

One area of concern when conducting an analysis with linear regression is serial correlation in the error terms. The assumption that the errors are uncorrelated allows us to make inferences about the significance of various model terms. If serial correlation is present in the residuals, then we can not do this inference. As a means of eliminating correlation in the errors, we will include lagged variables in our model. We will denote the one-period lagged values for heating oil futures price, production futures price, and crude oil futures price by $F^{HO}_{t-1,(t-1)+\Delta}$, $P_{t-1,(t-1)+\Delta}$, and $F^{C}_{t-1,(t-1)+\Delta}$, respectively. The inclusion of these lagged values, particularly that of the response, will help reduce the amount of serial correlation in the errors. For each of our models, we will perform the Ljung-Box test as described in [7] to see if correlation is present. Let us briefly describe this test.

Suppose we wish to test the null hypothesis that the first h autocorrelations of the residuals are equal to zero against the alternative that at least one is nonzero. In a model with n observations, we compute the following statistic:

$$Q_{LB} = n \cdot (n+2) \sum_{j=1}^{h} \frac{\hat{\rho}^2(j)}{n-j}, \tag{31}$$

where $\hat{\rho}(j)$ is the sample autocorrelation function of the residuals at lag j. If the null hypothesis is true, then Q_{LB} has an approximate chi-squared distribution with h degrees of freedom. The null hypothesis is rejected for large values of the test statistic. We will perform this test for each of our models, setting $h = 3$.

Before discussing the regression models, let us present a time series plot of heating oil futures prices during 1988-1995. Figure 5 plots the time series of prices for the 1-month and 5-month time-to-maturity contracts. There are two interesting time periods to consider. First, for the 1-month contract, observation 25 is extremely high. As reported in the *1990 CRB Commodity Year Book* [9], a record cold winter caused the price of heating oil to jump. While it had a dramatic effect on the 1-month price, this event did not significantly affect contracts with other maturity times. Thus, when we fit our regression models, we will remove this observation in the $1-$month model. The second time period of note is the set of observations from time periods 33 to 37. This is the Persian Gulf War period, ranging from September, 1990 to January, 1991. The heating oil futures price for contracts on all maturities is greatly affected during this period. Rather than eliminate each of these points, we will find it acceptable to leave them in each model.

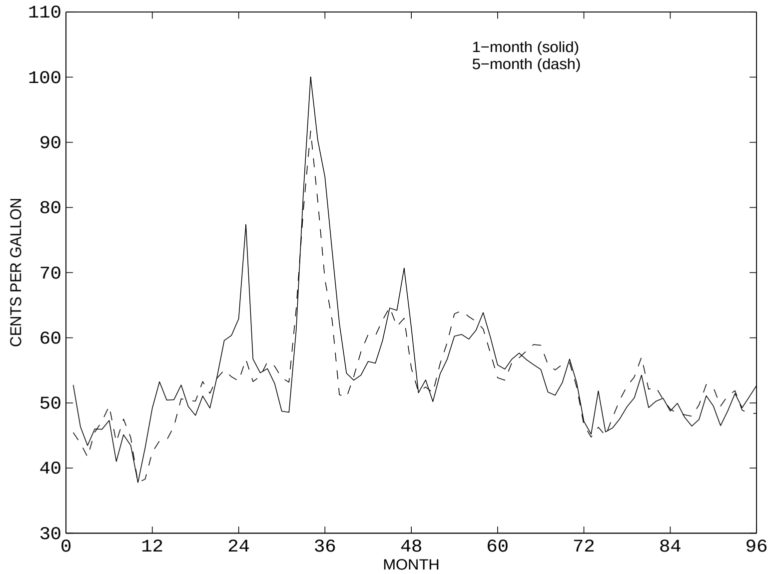

Figure 5: Heating Oil Futures Prices, 1988-1995

We are now ready to fit our regression models. The regression equation to be estimated is

$$\begin{aligned} F^{HO}_{t,t+\Delta} = \quad & B_0 + B_C \cdot F^C_{t,t+\Delta} + B_P \cdot P_{t,t+\Delta} + \textstyle\sum_{j=1}^{11} B_j \cdot I_{tj} + B_{HO_L} \cdot F^{HO}_{t-1,(t-1)+\Delta} + \\ & B_{C_L} \cdot F^C_{t-1,(t-1)+\Delta} + B_{P_L} \cdot P_{t-1,(t-1)+\Delta} + \epsilon_t, \end{aligned} \tag{32}$$

where $\Delta = 1, 2, 3, 4, 5$ and $1 \le t \le 96 - \Delta$. The results of these linear regressions are located in Table 4. Each column represents a different time to maturity. For each regression coefficient, the least squares estimate is given along with its standard error in parentheses. Coefficients statistically different from zero at a significance level of 0.05 are in boldface. The next-to-last row gives the adjusted R^2 value for each model, while the final row gives the value of the Ljung-Box test statistic for $h = 3$ lags and its p-value in parentheses. In all five models, the crude oil futures price was overwhelmingly the most important predictor variable. The estimated coefficient was near unity, and the value of the associated t-statistic

Table 4: Regression Model Results for Predicting Futures Prices

Quantity	$\Delta = 1$	$\Delta = 2$	$\Delta = 3$	$\Delta = 4$	$\Delta = 5$
$\hat{B}_0$	**5.67**(1.97)	**4.63**(1.72)	2.89(1.46)	2.38(1.40)	1.46(1.27)
$\hat{B}_C$	**1.02**(0.06)	**1.03**(0.05)	**1.05**(0.05)	**1.03**(0.05)	**1.05**(0.05)
$\hat{B}_P$	**3538**(705)	**2989**(775)	**3175**(893)	1482(1013)	942(1091)
$\hat{B}_1$	-0.48(1.31)	-1.97(1.02)	-0.14(0.98)	-0.99(0.82)	-0.53(0.79)
$\hat{B}_2$	-2.25(1.21)	-1.40(1.10)	-0.64(0.95)	-0.67(0.93)	0.12(0.91)
$\hat{B}_3$	**-4.53**(1.19)	**-4.29**(1.13)	**-2.70**(1.05)	-1.24(1.26)	0.01(1.26)
$\hat{B}_4$	-1.68(1.31)	-2.32(1.19)	-1.09(1.08)	0.00(1.32)	1.50(1.19)
$\hat{B}_5$	**-4.00**(1.29)	**-3.57**(1.18)	**-2.37**(1.11)	-0.44(1.35)	0.71(1.24)
$\hat{B}_6$	**-2.75**(1.32)	**-2.97**(1.19)	-1.27(1.09)	0.13(1.34)	1.23(1.20)
$\hat{B}_7$	**-3.66**(1.24)	**-3.24**(1.15)	-1.64(1.06)	0.33(1.31)	1.49(1.23)
$\hat{B}_8$	-2.16(1.26)	-1.64(1.13)	-0.22(1.01)	1.26(1.25)	**2.23**(0.96)
$\hat{B}_9$	-0.92(1.20)	-0.09(1.07)	1.54(0.97)	**2.91**(1.08)	**2.98**(0.80)
$\hat{B}_{10}$	-2.35(1.20)	-1.98(1.14)	0.21(1.00)	1.30(0.94)	0.55(0.84)
$\hat{B}_{11}$	-1.01(1.14)	-0.76(1.01)	1.34(0.84)	0.27(0.85)	0.54(0.77)
$\hat{B}_{HO_L}$	**0.48**(0.10)	**0.63**(0.11)	**0.64**(0.11)	**0.62**(0.11)	**0.60**(0.11)
$\hat{B}_{C_L}$	**-0.49**(0.10)	**-0.65**(0.11)	**-0.68**(0.11)	**-0.64**(0.11)	**-0.63**(0.11)
$\hat{B}_{P_L}$	**-3620**(713)	**-3048**(788)	**-3194**(901)	-1439(1018)	-835(1095)
$R^2(adj)$	95.3%	96.6%	97.4%	97.5%	97.5%
$Q_{LB}(h=3)$	2.74(0.43)	4.96(0.17)	3.83(0.28)	2.35(0.50)	2.02(0.57)

was statistically significant at the 0.05 level. For time to maturity ranging from one to three months, the production futures price coefficient was statistically significant and positively correlated with heating oil price. For the two longest maturity contracts, however, it was not significantly different from zero. The impact of the seasonal indicators varied greatly, with some coefficients being statistically significant only for certain maturity levels and others not at all. Also, note that the lagged terms were important in forecasting the heating oil futures price. They also eliminated serial correlation in the model, as we would fail to reject the null hypothesis of uncorrelated errors based on our Ljung-Box statistics. The overall explanatory power of each model was excellent, with adjusted R^2 values in excess of 95%.

Finally, using our regression models, we can determine the importance of the production futures price in reducing the total variability in heating oil futures prices. Consider the following. For a fixed maturity level, the total variability in the heating oil futures price is calculated by computing the sum of the squared deviations about its sample mean. In a linear regression model, this quantity, known as the total sum of squares, is partitioned into the sum of squares due to regression and the sum of squares due to error. As explanatory variables are added to our regression model, the regression sum of squares grows while the error sum of squares is reduced. For each particular term added to the model, we can compute its sequential sum of squares. This is defined as the further reduction in error sum of squares beyond that achieved by the previous predictors. In our models, the two most important variables to consider are crude oil futures price and production futures price. Let SS_{HO} represent the total sum of squares for the heating oil futures price. Denote the sequential sums of squares for the crude oil futures price and the production futures price by SS_C and SS_P, respectively, as each term is successively added to the model. Table 5 records

Table 5: Selected Sums of Squares when Forecasting Prices

Quantity	$\Delta = 1$	$\Delta = 2$	$\Delta = 3$	$\Delta = 4$	$\Delta = 5$
SS_{HO}	9314.19	9069.28	8453.67	7337.2	6333.1
SS_C	8213.60	8262.76	7735.93	6740.7	5846.9
SS_P	121.92	144.03	188.72	186.1	169.6

these quantities for each of our forecasting models. Note that total variability in the heating oil futures price decreases as the time to maturity increases. The crude oil futures price is responsible for a large reduction in error sum of squares, while the production futures price contributes to a lesser degree. Still, the production futures price does explain part of the variability in the heating oil futures price. Thus, it adds value to our model.

4.4 Predicting Heating Oil Futures Price Spreads

Recall the theory of commodity storage discussed in Section 2. The difference in price between contemporaneous spot and futures prices on the same asset was known as the cost of carry. It was conjectured that this quantity depended on interest costs, holding costs, and a convenience yield earned from storing the asset. Along the same line of reasoning, the difference in price between two futures prices on the same asset with adjacent maturity dates was known as the forward cost of carry. This quantity could be interpreted as the current belief about the cost of carrying the asset during the time period defined by the aforementioned maturity dates. Since a spot price can be thought of as a futures price with time to maturity zero, each of these price differences can be thought of as a futures price spread. Let us introduce some notation consistent with the previous subsection. Consider the following expression:

$$FPS_{t,\Delta}^{HO} = F_{t,t+\Delta+1}^{HO} - F_{t,t+\Delta}^{HO}. \tag{33}$$

The left-hand side represents the futures price spread for heating oil at time t on contracts with time to maturity Δ and $\Delta + 1$. The right-hand side is the Δ-periods ahead forward cost of carry. It is the cost, perceived at time t, of carrying the asset during the period $[(t+\Delta),(t+\Delta+1)]$. For heating oil, we would expect to observe a seasonal pattern in these futures price spreads.

We will use first-of-the-month heating oil settlement prices from NYMEX during 1988-1995 to help us construct the price spreads. In addition, we will also use first-of-the-month New York harbor spot prices, denoted by S_t^{HO}, where

$$S_t^{HO} = F_{t,t}^{HO}. \tag{34}$$

Note that possible values for Δ are $0, 1, 2, 3, 4$. We will again restrict ourselves to contracts that have a maturity date within the eight-year period.

We will again use linear regression modeling to help us forecast these futures price spreads. One potential explanatory variable is a price spread constructed using our production futures prices. Consider the following expression:

$$FPS_{t,\Delta}^{P} = P_{t,t+\Delta+1} - P_{t,t+\Delta}. \tag{35}$$

The quantity $FPS_{t,\Delta}^{P}$ is the production futures price spread with definition similar to that of the heating oil futures price spread. We would expect it to be useful in predicting that price spread.

Although the futures price of crude oil was extremely useful in predicting the heating oil futures price, we will not include these prices in our model. Unlike heating oil spreads, the crude oil futures price spreads do not exhibit seasonality. Thus, we conjecture that they will not be useful in forecasting the heating oil futures price spreads.

Identical to the previous models, we will include eleven seasonal indicator variables in our regressions. Also, we will make use of lagged values to help us eliminate serial correlation in the residuals. We will use the following one-period lagged values:

$$FPS^{HO}_{t-1,\Delta} = F^{HO}_{t-1,t+\Delta} - F^{HO}_{t-1,(t-1)+\Delta}.$$

$$FPS^{P}_{t-1,\Delta} = P_{t-1,t+\Delta} - P_{t-1,(t-1)+\Delta}. \tag{36}$$

Before fitting the models, let us look at a graph of heating oil futures price spreads during 1988-1995. Figure 6 plots the time series for $FPS^{HO}_{t,\Delta}$ for $\Delta = 0$ and $\Delta = 4$. The

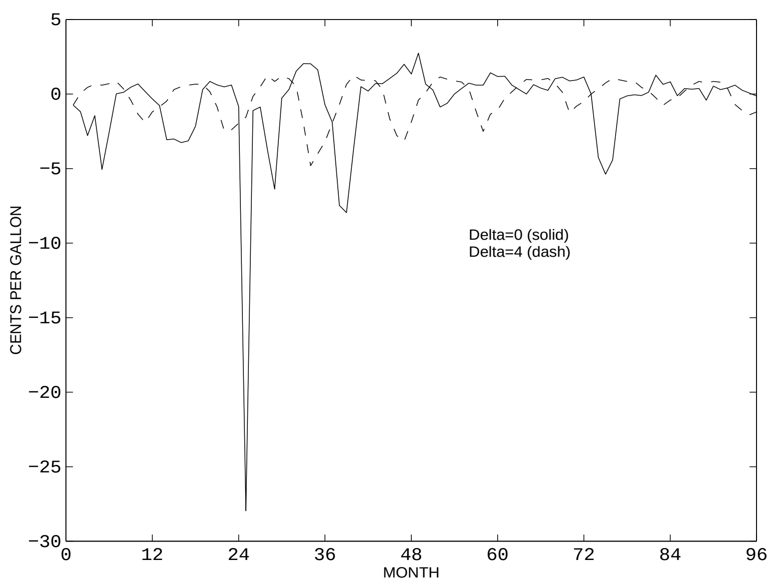

Figure 6: Heating Oil Futures Price Spreads, 1988-1995

seasonal pattern predicted by the theory of commodity storage is apparent, particularly for the $\Delta = 4$ series. Clearly, observation 25, discussed in the previous subsection, is having an enormous impact on the spread for $\Delta = 0$. This is to be expected since it is a function of the 1-month futures price. Note that for this same reason, we would also expect the spread corresponding to $\Delta = 1$ to be affected. The other price spreads, like the one corresponding to $\Delta = 4$, will be unaffected. Regarding the Persian Gulf War period, we will again leave the observations in our models. Thus, for the models corresponding to $\Delta = 0$ and $\Delta = 1$, we will remove observation 25. We can now estimate the parameters of our regression models. The regression equation to be estimated is

$$\begin{aligned} FPS_{t,\Delta}^{HO} = B_0 + B_P \cdot FPS_{t,\Delta}^{P} + \sum_{j=1}^{11} B_j \cdot I_{tj} + \\ B_{HO_L} \cdot FPS_{t-1,\Delta}^{HO} + B_{P_L} \cdot FPS_{t-1,\Delta}^{P} + \epsilon_t, \end{aligned} \tag{37}$$

where $\Delta = 0, 1, 2, 3, 4$ and $1 \leq t \leq 95 - \Delta$. The results of these regressions are reported in Table 6. Each column represents a different value of Δ. The format of the table is identical to Table 4. From the table, we can see that the production futures price spread is clearly statistically significant from zero for each model with the exception of $\Delta = 4$, when it is still marginally significant. It is positively correlated with the heating oil futures price spread. The significance of the seasonal indicator variables again varies across models. In particular, for the models corresponding to $\Delta = 0$ and $\Delta = 1$, only one of the eleven coefficients is statistically significant from zero. Yet for the model corresponding to $\Delta = 2$, eight of the seasonal coefficients were significant. For the lagged variables, the significance of the coefficients varied. In particular, note that the lag of the heating oil price spread became more significant as Δ increased. Of course, the primary reason for including these lagged variables was to eliminate serial correlation in the errors. Based on the Ljung-Box statistics,

Table 6: Regression Model Results for Predicting Futures Price Spreads

Quantity	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 3$	$\Delta = 4$
$\hat{B}_0$	-0.81(1.04)	0.13(0.37)	**-1.06**(0.24)	0.59(0.37)	0.44(0.27)
$\hat{B}_P$	**8004**(1636)	**3376**(619)	**3418**(616)	**2687**(663)	1107(575)
$\hat{B}_1$	-1.68(1.24)	**-1.12**(0.51)	**1.52**(0.38)	-0.64(0.48)	-0.02(0.24)
$\hat{B}_2$	-1.33(1.25)	-0.66(0.43)	0.64(0.35)	-0.42(0.47)	-0.08(0.28)
$\hat{B}_3$	**4.31**(1.64)	0.26(0.47)	**1.63**(0.32)	-0.09(0.42)	-0.13(0.34)
$\hat{B}_4$	2.94(1.89)	-0.08(0.58)	**1.08**(0.34)	-0.76(0.55)	-0.23(0.38)
$\hat{B}_5$	2.59(1.77)	0.25(0.55)	**0.75**(0.37)	-0.53(0.51)	-0.68(0.39)
$\hat{B}_6$	1.80(1.75)	-0.32(0.60)	**1.20**(0.36)	-0.92(0.52)	-0.52(0.40)
$\hat{B}_7$	-0.09(1.62)	-0.21(0.50)	**0.80**(0.37)	-0.68(0.50)	-0.62(0.40)
$\hat{B}_8$	-0.74(1.33)	-0.39(0.50)	**0.80**(0.38)	-0.87(0.51)	-0.80(0.43)
$\hat{B}_9$	-1.25(1.27)	-0.36(0.46)	0.73(0.39)	**-1.06**(0.45)	**-1.28**(0.40)
$\hat{B}_{10}$	0.30(1.24)	-0.07(0.49)	0.57(0.39)	**-0.89**(0.35)	**-1.79**(0.26)
$\hat{B}_{11}$	-1.29(1.24)	-0.33(0.45)	**1.31**(0.39)	**-1.63**(0.41)	-0.26(0.36)
$\hat{B}_{HO_L}$	0.11(0.11)	0.07(0.08)	**0.52**(0.10)	**0.70**(0.08)	**0.73**(0.07)
$\hat{B}_{P_L}$	-1016(2064)	-786.4(810)	**-1724**(721)	-1413(721)	-71(599)
$R^2(adj)$	59.1%	79.7%	78.3%	84.2%	90.4%
$Q_{LB}(h=3)$	1.22(0.75)	11.21(0.01)	2.20(0.53)	4.97(0.17)	5.21(0.16)

however, there is evidence to suggest that the errors in the model corresponding to $\Delta = 1$ are correlated. Finally, overall explanatory power is good for each model, with adjusted R^2 values mostly rising as Δ increases.

As we did in the previous subsection, we can measure the contribution made by specific predictor variables in reducing the total variability in the heating oil futures price spread by examining sequential sums of squares. For our models, the variable of interest will be the production futures price spread. For simplicity, let SS_{HO} represent the total sum of squares for the heating oil futures price spread, and denote by SS_P the sequential sum of squares for the production futures price spread. Table 7 records these values. First, note

Table 7: Selected Sums of Squares when Forecasting Price Spreads

Quantity	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 3$	$\Delta = 4$
SS_{HO}	1136.42	275.44	152.64	177.15	149.97
SS_P	360.25	132.83	102.03	118.59	109.84

the extremely large value for SS_{HO} when $\Delta = 0$. This is due to the presence of observation 25. More importantly, note the relatively large values for SS_P. Contrast these to the models in the previous subsection, where the production futures price helped reduce variability in heating oil futures prices only slightly compared to the crude oil futures price. Clearly, the production futures price spread is extremely useful for predicting the heating oil futures price spread. This is not surprising. Let us write the production futures price spread in terms of actual production costs. Using equations (30) and (35), we obtain the following:

$$FPS^P_{t,\Delta} = P_{t,t+\Delta+1} - P_{t,t+\Delta}.$$

$$
\begin{aligned}
&= [f_t(N\mu_t^H, N\beta_t^H, NX_t^H) - f_t^{t+\Delta+1}(N\mu_t^H, N\beta_t^H, NX_t^H)] \\
&- [f_t(N\mu_t^H, N\beta_t^H, NX_t^H) - f_t^{t+\Delta}(N\mu_t^H, N\beta_t^H, NX_t^H)]. \\
&= f_t^{t+\Delta}(N\mu_t^H, N\beta_t^H, NX_t^H) - f_t^{t+\Delta+1}(N\mu_t^H, N\beta_t^H, NX_t^H). \qquad (38)
\end{aligned}
$$

The above is simply the difference in time t production cost when additional inventory is received in future time period $t + \Delta$ as opposed to when it is received in period $t + \Delta + 1$. This represents the value, viewed from time t, of holding this additional inventory during the period $[(t + \Delta), (t + \Delta + 1)]$. It is natural to expect this quantity to be related to the corresponding heating oil futures price spread. Seasonality in production and inventory levels generate seasonals in the production futures price spreads. This seasonality is observed in the actual heating oil futures price spreads.

4.5 Conclusions and Future Research

From the preceding analysis, it is clear that decisions about the production and storage of distillate fuel oil influence the price of heating oil. This is consistent with the theory of commodity storage. The production futures prices introduced in this work are helpful in explaining observed heating oil futures prices and especially heating oil futures price spreads.

A potential direction for future research is to perform out-of-sample forecasts for prices. For instance, since our model covers the time period from 1988-1995, we could attempt to see if our parameters would allow us to match historical states and decisions in 1996. We could then take the costs associated with these states and use them in our fitted regression models to predict prices. Alternatively, we could attempt the more difficult problem of trying to predict prices in the future. Since the future is not known, historical states and decisions have yet to be observed. We could, however, identify for each time period a set of states likely to be realized. Based on these likely states, we may be able to infer potential decisions and their associated costs. Using this data, we could then attempt to forecast prices.

Appendix

In this appendix we discuss several details related to parameter estimation and selection for the stochastic dynamic programming problem.

With respect to the implementation of Winters' method, a point needs to be made regarding the estimates of the seasonal factors. Recall that these estimates are not being updated since $\alpha_3 = 0$ in equation (11). To compensate for this fact, when we generated forecasts for demand during the period 1988-1995, we did not use the estimates from the base period. Instead, to more accurately reflect the seasonality during this period of interest, we estimated seasonal factors using the data from 1988-1995 as reported in [2]. Furthermore, one additional quantity that needs to be estimated is the standard deviation of demand, σ. Using the data for demand from 1988-1995, we performed a linear regression on an intercept, time, and seasonal indicators. Our estimate for σ is the standard error of the residuals. This estimate was found to be $\hat{\sigma} = 4,753,565$.

For the left endpoints of the ranges for the state variables, recall that these quantities depend on time. For μ_t and β_t, we will take the following approach. At the beginning of time period t, estimates for the level and slope, computed at the end of time period $t-1$, are known. Round these quantities to an integer by dividing by their respective increments. Next, multiply these rounded values by their increments. To compute the left endpoints of the ranges, subtract an integer multiple of the increment from these quantities. For the left endpoints of X_t, the following approach was taken. Seasonal values for the left endpoints were selected for each month by looking at historical beginning period inventory levels during 1988-1995. Each value is an integer multiple of IX.

Recall that our goal is to choose parameters for cost functions that will produce a solution to the dynamic programming problem that matches actions taken during the period 1988-1995. To do so, we must first determine the historical state faced by the refiner as well as his decision in each time period.

Let the historical state at time t be denoted by the following vector:

$$\mathbf{s_t^H} := (N\mu_t^H, N\beta_t^H, NX_t^H).$$

The first two components of this vector are determined easily. Recall from the previous subsection that the ranges for μ_t and β_t are defined so that they contain the historical values. Thus, $N\mu_t^H$ and $N\beta_t^H$ are identified easily. To compute NX_t^H, we will note the historical beginning period inventory, X_t^H, and make use of equation (19):

$$X_t^H = LX_t + (NX_t^H - 1) \cdot IX. \tag{A.1}$$

We solve this expression for NX_t^H, noting that $\frac{X_t^H}{IX}$ will need to be rounded:

$$NX_t^H = \left\lceil \left(\frac{X_t^H}{IX} - 0.5 \right) \right\rceil - \frac{LX_t}{IX} + 1. \tag{A.2}$$

By defining LX_t carefully, we can be certain that $1 \leq NX_t^H \leq N_X$ for all t.

The historical decision at time t will be denoted by NY_t^H. We can compute this quantity in exactly the same way as we computed NX_t^H. Given Y_t^H, equation (20) gives

$$Y_t^H = LX_t + (NY_t^H - 1) \cdot IX. \tag{A.3}$$

Thus, similar to equation (40) we obtain

$$NY_t^H = \left\lceil \left(\frac{Y_t^H}{IX} - 0.5 \right) \right\rceil - \frac{LX_t}{IX} + 1. \tag{A.4}$$

Finally, to compute the end-of-horizon cost, we noted the amount of inventory held at the end of December, 1995. This is also the beginning period inventory for January, 1996, denoted by X_N^H. Upon rounding to a multiple of IX, we selected parameter values for the cost function given in equation (27) that minimized the cost associated with this historical level. This will help eliminate errors in matching history at the end of the horizon. The values chosen for the parameters were $a_N = 1$, $b_N = 10$, $c_N = 1.26 \cdot AMD$, $d_N = 10$, and $e_N = AMD$. To ensure that the cost is always nonnegative, the constant (1.29) was added to the function in each state.

References

[1] Abraham, B. and J. Ledolter. *Statistical Methods for Forecasting.* John Wiley & Sons, Inc., 1983.

[2] American Petroleum Institute. *Basic Petroleum Data Book*, Vol. XVI, No. 1. American Petroleum Institute, Washington, DC, 1996.

[3] Avriel, M. *Nonlinear Programming: Analysis and Methods.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1976.

[4] Bertsekas, D. *Dynamic Programming: Deterministic and Stochastic Models.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1987.

[5] Bopp, A. E. and S. Sitzer. On the Efficiency of Futures Markets-Another View. *Energy Economics*, **10**: 199-205, 1988.

[6] Brennan, M. J. The Supply of Storage. *American Economic Review*, **48**: 50-72, 1958.

[7] Brockwell, P. J. and R. A. Davis. *Introduction to Time Series and Forecasting.* Springer-Verlag, New York, 1996.

[8] Cortazar, G. and E. S. Schwartz. Valuation of Commodity-Contingent Claims. *Journal of Derivatives*, **1**: 27-39, 1994.

[9] Commodity Research Bureau. *1990 CRB Commodity Year Book.* Commodity Research Bureau, New York, 1990.

[10] Energy Information Administration. *Monthly Energy Review.* Energy Information Administration, Washington, DC, December, 1985.

[11] Fama, E. F. and K. R. French. Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage. *Journal of Business*, **60**: 55-74, 1987.

[12] Gibson, R. and E. S. Schwartz. Stochastic Convenience Yield and the Pricing of Oil Contingent Claims. *The Journal of Finance*, **45**: 959-977, 1990.

[13] Kaldor, N. Speculation and Economic Stability. *Review of Economic Studies*, **7**: 1-27, 1939.

[14] Lowry, M. N. Precautionary Storage of Refinery Products-The Case of Distillate Fuel Oil. *Energy Economics*, **10**: 254-260, 1988.

[15] Miltersen, K. R. and E. S. Schwartz. Pricing of Options on Commodity Futures with Stochastic Term Structures of Convenience Yields and Interest Rates. *Journal of Financial and Quantitative Analysis*, **33**: 33-59, 1988.

[16] National Petroleum Council. *Petroleum Inventories and Storage Capacity.* National Petroleum Council, Washington, DC, 1984.

[17] Telser, L. G. Futures Trading and the Storage of Cotton and Wheat. *Journal of Political Economy*, **66**: 233-244, 1958.

[18] Working, H. Theory of the Inverse Carrying Charge in Futures Markets. *Journal of Farm Economics*, **30**: 1-28, 1948.

[19] Working, H. The Theory of the Price of Storage. *American Economic Review*, **39**: 1254-1262, 1949.