An Integrated Data-Driven Method Using Deep Learning for a Newsvendor Problem with Unobservable Features

Davood Pirayesh Neghab
Department of Mechanical and Industrial Engineering, Ryerson University, Toronto, Canada
Email: dneghab@ryerson.ca

Siamak Khayyati
College of Engineering, Koç University, Rumeli Feneri Yolu, Istanbul, Turkey, 34450
Email: skhayyati13@ku.edu.tr

Fikri Karaesmen*
College of Engineering, Koç University, Rumeli Feneri Yolu, Istanbul, Turkey, 34450
Email: fkaraesmen@ku.edu.tr

We consider a single-period inventory problem with random demand with both directly observable and unobservable features that impact the demand distribution. With the recent advances in data collection and analysis technologies, data-driven approaches to classical inventory management problems have gained traction. Specially, machine learning methods are increasingly being integrated into optimization problems. Although data-driven approaches have been developed for the newsvendor problem, they often consider learning from the available data and optimizing the system separate tasks to be performed in sequence. One of the setbacks of this approach is that in the learning phase, costly and cheap mistakes receive equal attention and, in the optimization phase, the optimizer is blind to the confidence of the learner in its estimates for different regions of the problem. To remedy this, we consider an integrated learning and optimization problem for optimizing a newsvendor's strategy facing a complex correlated demand with additional information about the unobservable state of the system. We give an algorithm based on integrating optimization, neural networks and hidden Markov models and use numerical experiments to show the efficiency of our method. In an empirical experiment, the method outperforms the best competitor benchmark by more than 27%, on average, in terms of the system cost. We give further analyses of the performance of the method using a set of numerical experiments.

Key words: Inventory; Hidden Markov model; Deep neural network; Partially observed data; Integrated Estimation and Optimization

---

* Corresponding author.

This paper is dedicated to the memory of Prof. Gabor Rudolf, who is unfortunately no longer with us. We thank him for his valuable advice and comments enhancing the quality of this work.

1. Introduction

The single period random demand inventory problem is one of the central problems in inventory control and capacity management. In the standard version of the problem, it is assumed that the inventory manager chooses an order quantity before observing the random demand. The mismatch between the order quantity and the realized demand may lead to unsatisfied demand or unsold items and the implied costs of a unit lost demand and an unsold item are not symmetrical. One commonly used simplifying assumption on random demand is that its probability distribution is known with certainty and this distribution is independent and identically distributed in each period. However, this assumption might not hold given the available data in many cases.

Recent interest in data-driven approaches to inventory management has stimulated renewed interest in general versions of this problem where the random demand depends on multiple factors that are observed prior to ordering. In addition, there is past data on the observed factors and the corresponding demand that was realized that can guide a data-dependent ordering decision. This more general setup can be used for production/inventory management, and capacity planning problems in systems that are affected by disruptions in the supply chains and the consequent increase or decrease in demand for various items due to changes in the state of the environment such as the disruptions and changes caused by the recent COVID-19 pandemic. In addition to simple seasonal factors (such as day of the week or week of the month), or planning related factors (promotions, competitor's actions, etc.), many data sources that might have a potential impact on demand are being monitored daily (weather forecasts, stock market indices, currency exchange rates) and can be used in controlling the inventory systems.

A recent stream of research (Ban and Rudin 2019, Oroojlooyjadid et al. 2020) investigates the above inventory problem under such *observable features*. On the other hand, there might be other factors that may affect demand that are not directly observable at the time of the decision. These may include supplier conditions affecting competitors, business cycles, or consumer preference shifts that are observable only after a long time lag. In economics and finance, market models with such hidden or *unobservable features* that randomly evolve are used to model and analyze business cycles (Hamilton 1990) or financial market conditions (Bhar and Hamori 2004). The operations literature also includes analytical models that investigate the effects of hidden features on inventory decisions but this literature does not study how to base these decisions on limited historical data. We propose a framework that includes such *unobservable features* governed by a random process which affect demand in addition to *observable features* that are known at the time of the decision. In this framework, considering the structure of the process that governs the unobservable features evolving over time is as important as the short-term modeling of the effects of the observable features.

In this paper, we contribute to the recent stream of literature that incorporates *observable features* in data-driven inventory or capacity planning (Ban and Rudin 2019, Oroojlooyjadid et al. 2020) by investigating a single-period random demand inventory problem where random demand in each period is dependent on an unobservable factor in addition to some observable factors. We assume that past data is available on the observable factors and the corresponding demand that was observed. On the other hand, information about the unobservable factor has to be inferred from past data. The traditional approach to such a problem would be to first estimate the underlying model parameters for the effect of the observable factors on demand and infer the hidden state information and its effects. Once estimation and inference take place, the optimization problem would be solved in the second step. However, some of the recent research on this problem (see Ban and Rudin (2019) for example) has demonstrated the advantages of integrating the estimation step of the parameters for the observable factors with the optimization step using tools from machine learning. We pursue this integrated estimation and optimization approach in a model which has the additional complexity of an unobservable factor. This brings the additional challenge of considering a multi-period dynamic optimization problem because the unobservable state is estimated dynamically and depends on the entire demand sequence that was observed prior to the decision. We therefore combine estimation, inference and optimization using a multi-layered neural network. To assess the performance of this integrated approach, we compare the results from our approach against data-based methods that ignore the hidden factor information or that employ separate inference and optimization steps. Numerical examples on both a synthetic data set and on representative real data that is taken as a proxy to retail data which might have an unobservable state demonstrate that our approach compares favorably against the other benchmarks.

The remainder of this paper is organized as follows. Section 2 presents a review of the related literature. Section 3 presents the setup and the solution method we have proposed. Section 4 introduces several benchmark methods and reports the results of the numerical experiments to assess the performance of the suggested method against the benchmarks. Section 5 provides the evaluation of the methods in an example with real data. Finally Section 6 concludes the paper.

2. Literature Review

In the following we provide a review of the pertinent literature. This is presented in two parts. First, we review the literature on inventory problems with an evolving demand environment. Then, we review the related work on the data-driven newsvendor problem.

2.1. Inventory problems with an evolving demand environment

A long line of research investigates the impact of several factors such as macroeconomic shocks and cycles on inventory systems (Blinder and Maccini 1991, Shang 2012, Kesavan and Kushwaha 2014), especially through the effects of such factors on random demand. There are several papers that consider dynamic environmental factors that cause the demand distribution to be non-stationary which creates additional challenges. A common assumption in these papers is that the random environment evolves according to a Markov chain.

Earlier studies assume that the state of the Markov chain at each point in time is fully observed and the true demand distribution associated with each state is known (Sethi and Cheng 1997, Beyer and Sethi 1997, Huh et al. 2011, Gallego and Hu 2004). For instance, Feldman (1978) proposes and analyzes a model where demand depends on the state of the environment modeled by a continuous-time Markov process. Lovejoy (1992) investigates the optimality of a myopic policy with non-stationary demand which is dependent on a Markovian process over time. Song and Zipkin (1993) present an inventory model where the demand depends on the state of the world modeled by a Markov chain and derive the optimal ordering policy.

In many practical situations, the environmental states are not perfectly observable. Instead, one can observe information about the environment and can only infer the states in a probabilistic manner. Treharne and Sox (2002) categorize the literature in terms of stationarity of the demand and observability of the information into four classes. The class of decision systems with Markov modulated demand and partially observed information is known as partially observed Markov decision process (POMDP) (Monahan 1982). Treharne and Sox (2002) study several inventory policies where only the historical demand is observable and the probability distribution of the demand is determined by the non-observable state of the Markov chain. Bensoussan et al. (2005, 2007) consider the newsvendor problem with censored demand and inventory which depend on the Markov chain states. Arifoğlu and Özekici (2010) analyze a single-item periodic-review inventory system in a random environment. They extend the model of Gallego and Hu (2004) to the more general setting where the environment is only partially observable. In particular, they show that a state-dependent base-stock policy is optimal using sufficient statistics on the environment process. In a later work, Arifoğlu and Özekici (2011) investigate the optimality of a state-dependent inventory policy in a random environment where the capacity of production is random. Avci et al. (2020) model and analyze the inventory problem where the demand belongs to a probability distribution conditional on the Markovian states of the world.

In the above papers that use a POMDP model with imperfectly observed environment processes, demand state is partially revealed via past demand data and the estimation of the state of the environment is an important subproblem. This subproblem is solved using Bayesian updating to incorporate the partial observations into the inventory models and a general solution is given by the Baum-Welch algorithm. The outcome of this algorithm is the estimation of demand distribution for each state of the observed sequence. This estimation is based on maximizing the likelihood of the observed sequence. This maximization at the subproblem level does not take into account the objective function of the inventory problem which leads to a separation of estimation and optimization. However, it is seen in recent examples in the literature that integrating the estimation and optimization problems may lead to better solutions. We refer to the separated estimation and optimization procedure used in the above papers as Objective-blind Baum-Welch (ObBW) method. This work contributes to this stream of research by developing a method that integrates the estimation of the hidden states and the optimization of the system, allowing for the parameters of the objective function to guide the estimation. Table 1 displays the characteristics of this method along with other approaches that are explained in the next subsection.

2.2. Data-driven approaches to inventory problems

Many papers address the concern that the demand distribution in an inventory problem may not be completely known. Many of the works considering this topic approach the problem from the perspective of robust optimization (Scarf 1958, Gallego and Moon 1993, Perakis and Roels 2008) and Bayesian updating (Scarf 1959).

Some studies contribute to relaxations of the assumption that demand distribution is completely known by developing data-based methods. In this framework, the decision maker uses the empirical distribution obtained from past observations (Levi et al. 2007, Liyanage and Shanthikumar 2005, Huh et al. 2011, Besbes and Muharremoglu 2013). We refer to the approach in these papers which is mostly based on the sample observations of demand as the Empirical Demand Distribution (EDD) method. For instance, Bertsimas and Thiele (2005) solve the problem without estimating the distribution but assuming that all of the demand observations in the sample are assigned an equal probability $1/T$, where T denotes the number of demand observations. The optimal stock level or the order quantity is then approximated by the estimated empirical distribution. The advantage of this method is that, unlike the ObBW method, it does not assume any particular shape for the demand distribution. This is useful with real data where demand may not follow a common distribution. On the other hand, one drawback of this method is assuming that all the future observations will also belong to the same empirical distribution which may be questionable.

In recent decades, data-driven optimization under uncertainty has gained increasing attention. For instance, He et al. (2012) model the problem of setting nurse staffing levels in hospital operating rooms with the uncertainty of daily workload as a newsvendor problem. They present various models including a linear decision model that uses two features. Sachs (2015) considers ordering with different types of exogenous data such as price and temperature that might explain demand. She formulates the optimal inventory level as a linear function of those variables. In a case study with real data, she shows that the non-parametric approaches outperform the parametric ones.

This is advantageous when the true demand distribution is not completely known and several exogenous variables are available. We refer to this approach as Parameter Fitting Linear Regression (PfLR). Here, the exogenous features explain part of demand variability through a (typically) linear relationship. The approach therefore estimates the mean of the demand by a linear regression on the features. This results in a time-varying mean that depends on the features and a fixed standard deviation. In addition, the usual Gaussian assumptions are usually taken. Similar to the empirical method, this method is not able to capture the dependency between demand observations.

The integrated estimation and optimization approach is referred to as a prescriptor method (Van Parys et al. 2020). In a recent paper, Van der Laan et al. (2019) propose a new data-driven approach based on distributionally robust optimization to achieve on-target service levels. They show that the suggested approach, which bases the inventory decision directly on feature data, is more reliable than several classical approaches even with a limited number of historical observations. Several studies combine the estimation and optimization steps using tools from machine learning. Ban and Rudin (2019) consider the newsvendor problem with n observations and p features in two cases of a low and a high number of features. They propose two Machine Learning based approaches: regularization and Kernel Optimization (KO), and demonstrate some theoretical properties. They also show, in a numerical study, that using such features may lower the expected cost significantly. A recent paper by Khayyati and Tan (2020) shows that integrating the two steps of parameter estimation and optimization can improve the performance of a system in make-to-stock queues.

The general use of ML-based methods in joint estimation and optimization, however, goes back to several earlier studies such as (Efendigil et al. 2009, Goel et al. 2010, Gruhl et al. 2004, 2005). Bertsimas and Kallus (2020) combine the methods of ML with the conditional stochastic optimization problem. They include direct-effect data as well as other auxiliary information and assume that the joint probability distributions are unknown and the observations are imperfect. They develop the framework with several ML methods and show that these techniques are computationally tractable and asymptotically optimal under some conditions. This tractability is shown in the presence of dependencies in the data and censored observations.

The main contributions of these important recent papers are using informative data, benefiting from non-parametric models, decreasing the estimation errors, and making the decisions more dynamic. We categorize the approach proposed by Ban and Rudin (2019) of relating the optimal order quantity directly to the features using a functional form for the relationship as Linear Machine Learning (LML). This method combines the estimation and optimization steps by solving a nonlinear optimization problem, where the objective is minimizing directly the cost function of the newsvendor problem instead of minimizing the regression error. In this method, similar to the parameter fitting approach, there is an assumption of the linear relations between features and the optimal order quantity.

Finally, some recent studies in data-based optimization contribute to the literature by taking nonlinear relationships between feature data and the order quantity into consideration. Oroojlooyjadid et al. (2020) apply a deep learning approach to the newsvendor problem when such nonlinearities exist. They also consider multi-feature and multi-products extensions of the problem. It is shown that deep learning outperforms other benchmarks such as local regression, classification and regression trees, random forests, and kernel optimization, especially when demand is highly volatile. Seubert et al. (2020) develop a data-driven system of ordering for a bakery chain based on artificial neural networks. They use two different methods of sequential and joint estimation and optimization and show that both methods considerably save costs compared to human planners. Qi et al. (2020) extend the approach of Oroojlooyjadid et al. (2020) to multi-period inventory system with uncertain demand and vendor lead time. Zhang and Gao (2017) examine a supervised deep learning algorithm with two objectives. They demonstrate that the original newsvendor loss function as the training function outperforms the quadratic loss function. The algorithm has been evaluated on synthetic and real data. In this class of methods, non-linear relations between features and demand have been considered. Oroojlooyjadid et al. (2020), as the first study of this class, suggest complex functions that relate the features and the optimal order quantity using deep learning in the classical newsvendor problem. This method is able to use the features and optimize the cost function while considering a wide range of relationships in addition to linear relations. However, their method does not identify the dependencies between consecutive states of the world in an evolving environment. This work aims at addressing this issue by modeling long term dependencies using hidden Markov models.

The contribution of this paper is as follows: first, we suggest a novel approach in data-driven inventory system, which considers both observable and unobservable sources of features that affect the randomness in demand. This new model extends the existing data-driven methods to other applications, where there are limitations in identifying the factors and their volatility that influence the state of the system. Second, we use the hidden Markov model as the most used modeling of the evolving environment. Utilizing HMM in a data-driven framework enables us to model long term dependencies and take advantage of other information sources available in the form of the feature data. Third, by combining neural network modeling tools with stochastic inference, and integrating them into the optimization method, we propose an integrated solution method for the suggested model that captures nonlinear dependencies between features and order quantity while alleviating the errors that occur in the estimation step. This is different from most of the literature where the tasks of learning about the demand distribution based on the features and setting the

Table 1 Position of the suggested model in the existing literature.

	Method	Abbrev.	State-dependent	Data-driven	NV-cost function integration	Non-linear model
1	Suggested model	HMMNV	✓	✓	✓	✓
2	Empirical Demand Distribution	EDD			✓	
3	Objective-blind by Baum-Welch Alg.	ObBW	✓			
4	Parameter fitting (Linear Regression)	PfLR		✓		
5	Linear Machine Learning	LML		✓	✓	
6	Deep Learning	DNN		✓	✓	✓

order quantity are handled separately. This may lead to a misalignment between the two different objectives of minimizing the estimation error and minimizing the system cost. Finally, we show the robustness of our proposed solution method using an extensive numerical experiments with both synthetic and real data. We compare the results of the suggested model with data-driven methods that ignore the evolving environment as a hidden factor or that employ inference and optimization steps separately. Our numerical results reveal that the proposed approach performs better than other methods when there might be unobservable features and leads to the recommendation that taking into account the unobservable features might have significant benefits. In addition, we present evidence that integrating unobservable feature estimation and inventory optimization is feasible and may bring significant improvements.

3. Model

In this section, we describe the model setting and our integrated estimation and optimization approach. We consider a single item newsvendor problem where the goal is to choose the order quantity at the beginning of every period to minimize the expected costs in that period. We assume that inventory is not carried over from one period to the next and backordering of unsatisfied demand is not allowed (as in service capacity planning problems). In this setup, the decision maker solves the problem in a period independently of previous periods' inventory and his order quantity does not affect future decisions. We suppose that the demand distribution is not known but that the demand depends on the available observable feature data in addition to having some long term

Table 2 Description of the variables of the model.

Variable	Description
D_t	demand at time t
Q_t	order quantity at time t
B_i	base demand in state i
f_{nt}	n^{th} feature observed at time t
$\mathbf{F}_t$	vector of independent and identically distributed of N features at time t
T	number of periods or sample observations
β_n	linear coefficient of n^{th} feature with demand
S_t	state of the Markov chain at time t
$\mathbf{q}_t$	vector of states probability
ψ^E	emission network function which maps the features $\mathbf{F}_t$ to D_t partially
$\mathbf{W}^E$	set of parameters of emission network function ψ^E
ψ^{NV}	newsvendor network function which maps $\mathbf{F}_t$ to Q_t partially
$\mathbf{W}^{NV}$	set of parameters of the newsvendor network function ψ^{NV}
ϵ_t	error term as the nonsystematic part of the demand
$\mathcal{N}(\mu,\sigma)$	Gaussian distribution with mean μ and standard deviation σ
Λ	likelihood function of the hidden Markov model
a_{ij}	probability of transition from state i to state j
A	transition probability matrix of the Markov model
E	the vector of the probability density functions of the hidden states
π	initial probability of the hidden Markov chain
α_t	forward parameter of the Baum-Welch algorithm
γ_1	learning rate of updating $\mathbf{W}^E$
γ_2	learning rate of updating $\mathbf{W}^{NV}$
γ_3	learning rate of updating A
η	the coefficient of the trade-off between newsvendor cost and the Likelihood

dependencies on an unobservable feature modeled as a Markov chain. The optimal order quantity, therefore, must also be dependent on previous information about demand and feature data. We first present our notation and present the assumptions and then formulate the demand evolution model. A brief introduction to deep neural networks followed by our suggested approach completes the section.

3.1. Demand data, optimization and notation

The historical data of the problem can be represented using tuples of feature vectors and demand realizations as

$$\mathcal{D} = \{(\boldsymbol{F}_1, D_1), (\boldsymbol{F}_2, D_2), \ldots, (\boldsymbol{F}_T, D_T)\}, \tag{1}$$

where T denotes the number of periods. In Equation (1), the vector of $\boldsymbol{F}_t$ for each period of $t = 1, 2, \ldots, T$ consists of N different features $f_{1t}, f_{2t}, \ldots, f_{Nt}$.

Given the data $\mathcal{D}$, our focus is on the following newsvendor optimization problem where the objective is to choose an order quantity Q to minimize expected (weighted) mismatch costs:

$$\min_{\mathbf{Q}} NVC(Q) = E_D\left[h(Q-D)^+ + b(D-Q)^+ |\mathcal{D}\right], \tag{2}$$

where D denotes the random demand, Q is the order quantity with h as the unit overage cost and b as the unit underage cost. In a retail setting, the overage and underage costs may refer to the cost of unsold items and lost demand respectively. In a capacity setting (staffing) they represent the cost of unused capacity and unfulfilled demand respectively.

When the demand distribution is known, the optimal order quantity in (2) can be found by the well-known critical fractile rule. However, the data-driven environment presents an additional challenge in that the solution of the outer optimization problem depends on the inner estimation problem where the expected cost has to be estimated using past observations. The estimation problem in itself is also solved as an optimization problem. Motivated by the recent success of methods that integrate estimation and optimization (Ban and Rudin 2019, Oroojlooyjadid et al. 2020), we propose an approach that uses an integrated solution.

Finally, we should note here the optimal quantity in (2) is found separately for each period since inventory or backorders are not carried over. On the other hand, the quantity decision depends on the currently observed features and all past demand observations which carry information about the state of the unobservable feature. This makes the order quantity dependent on the entire demand sequence up to time t. Table 2 describes all the variables that are used in this study.

3.2. Special case: A newsvendor problem with Markov modulated demand and observable states

Many papers in the literature assume that demand depends on an external state of the world that evolves according to a Markov chain S (Treharne and Sox 2002, Arifoğlu and Özekici 2010). It is then natural to assume that demand D depends on the external state and therefore the demand in period t, $D_t = D|S_t$ (or $D|S_{t-1}$ depending on the filtration). Unlike the general formulation in (2), the entire demand sequence is not required here because S_t carries all the necessary information. We then look for the order quantity that minimizes the expected cost of the system as

$$\min_{\mathbf{Q}} NVC(Q) = E_D\left[h(Q-D)^+ + b(D-Q)^+|S\right], \tag{3}$$

$$Q_t^* = F_{D|S_t}^{-1}\left(\frac{b}{h+b}\right), \tag{4}$$

where $F_{D|S_t}()$ is the cumulative distribution function of demand given S_t and $F^{-1}()$ denotes its inverse.

Let us now generalize the model further and define a base demand B_i that depends on $S_t = i$ and an additional demand that depends on the values of certain observed features $f_{1t}, f_{2t}, \ldots, f_{Nt}$ that do not depend on S_t. Further, let us assume that B_i is independent of features. We can then have

$$D_t = B_i + \psi(f_{1t}, f_{2t}, \ldots, f_{Nt}) + \epsilon_t, \tag{5}$$

where ϵ_t is a random error term with $E[\epsilon_t] = 0$.

3.2.1. Example As an example, assume that there are two states of the world: (1) Good and (2) Bad, and B_i is normally distributed with parameters (μ_i, σ_i) where $i = (1), (2)$. Further assume that

$$\psi(f_{1t}, f_{2t}, \ldots, f_{Nt}) = \beta_0 + \beta_1 f_{1t} + \beta_2 f_{2t} + \ldots + \beta_N f_{Nt}. \tag{6}$$

We then have that D_t is normally distributed with mean $\mu_i + \psi(f_{1t}, f_{2t}, \ldots, f_{Nt})$ and variance $\sigma_i^2 + \sigma_\epsilon^2$.

From the known results, we then have

$$Q_t^* = \mu_i + \psi(f_{1t}, f_{2t}, \ldots, f_{Nt}) + z^* \sqrt{\sigma_i^2 + \sigma_\epsilon^2}, \tag{7}$$

where $z^* = \Phi^{-1}(b/(h+b))$ (and $\Phi()$ is the cumulative distribution function of a standard normal random variable).

Next, we present the main model in this paper which includes an unobservable environment process states.

3.3. Our model: A data-driven newsvendor problem with unobservable environment process

In an inventory system, there may be several sources of uncertainty, generated by observable sources as features and non-observable sources. The features may correspond to observations such as the weather, seasonality, and local market conditions. In this model, the base demand takes place as the non-observable source. The base demand distribution, as a part of whole demand, depends on some evolving states which have two properties; first, they follow a Markov model, second, they are not observable. The state of the Markov chain affects the system partially through the base demand. Therefore, the joint probability distribution of demand and a state varies depending on the base demand distribution in each state, and the feature-dependent part completes the demand distribution independently of the states. Let us assume that S_t is not observable but can only be inferred from past demands $D_1, D_2, \ldots, D_{t-1}$. This is similar to the setup in Treharne and Sox (2002) and Arifoğlu and Özekici (2010). One can then estimate the conditional distribution

$$\hat{S}_t = S_t | D_1, D_2, \ldots, D_{t-1}. \tag{8}$$

In our model, we assume that the realizations of the features are independent from each other and the time period. Any dependency in the sequence of the observations of a feature and any dependency between the different features do not add more information to the decision at the beginning of a period as the features are observed before setting the order quantity. Hence, the presence of these dependencies does not affect the solution. Moreover, a dependency between the states of the (unobservable) Markov chain and the features may be beneficial to the order quantity decision as the state of the Markov chain can be better inferred through the feature observations

$$\hat{S}_t = S_t | D_1, D_2, \ldots, D_{t-1}, \mathbf{F}_{t-1}. \tag{9}$$

We consider the general form of the function (6) and substitute it in the Equation (5) to get

$$D_t = \sum_{i=1}^{S} I(S_t = i) B_i + \psi(f_{1t}, f_{2t}, \ldots, f_{Nt}) + \epsilon_t, \tag{10}$$

where $I(x) = 1$ if x is true and 0 otherwise. This implies a linear relation between the Markov chain-dependent and the features-dependent parts of the demand in each state and an unknown complex relation between D_t and $\mathbf{F}_t$. More specifically, we assume that the unknown set of parameters of the function of features that constitutes the demand partially is $\mathbf{W}$. We then assume the following form

$$D_t = \sum_{i=1}^{S} I(S_t = i) B_i + \psi(\mathbf{F}_t, \mathbf{W}) + \epsilon_t. \tag{11}$$

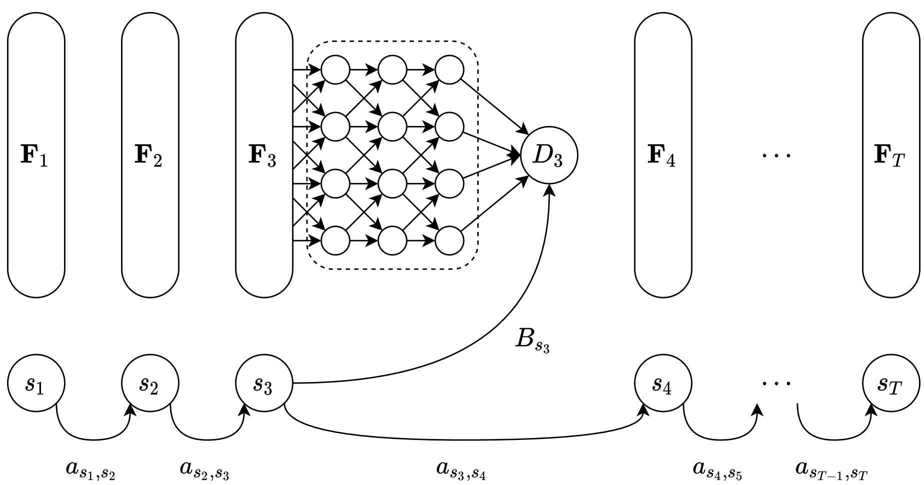

Figure 1 The effects of the features and the evolution of the state on the demand.

Figure 1 shows how demand is generated by features and states: D_3 is a function of the observed features F_3 and the unobservable state s_3.

The inventory optimization problem can be written as:

$$\min_{\mathbf{Q}} NVC(Q) = E_D \left[h(Q-D)^+ + b(D-Q)^+ | D_1, D_2, \ldots, D_{t-1}, \mathbf{F}_{t-1} \right]. \tag{12}$$

We propose to fit a joint estimation and optimization model to sample data and test the performance of the model out of sample with the objective of minimizing the cost in the out of sample data. Minimizing the cost function requires deriving an accurate belief about the state of the Markov chain at each point in time and solving the inventory optimization problem at that time. These two problems can be stated as:

(1) Hidden Markov model problem: First, we consider the problem of maximizing the likelihood of the observed sequence of demands that undergoes a hidden Markov model.

In our proposed problem setting, we assume that demand observations are produced by a continuous stochastic process. The problem of interest is characterizing the properties of demand observations. According to our data structure and several important applications, we know that the source of demand observations is nonstationary and its property varies over time. Here we use the mathematical structure of HMM to explain the theoretical basis to characterize the statistical properties of the demand observations (Rabiner 1989, Picone 1990).

Given a set of demand observations $D = D_{1:T}$, HMM with a finite set of S distinct hidden states changes the state of the system according to a set of probabilities associated with each state. In order to present a full probabilistic description of this system, the other elements of an HMM rather than the observations and the number of states are defined as follows referred to as the triple model parameters $\lambda = (\pi, A, E)$. Here, $\pi = \{\pi_i = P(S_1 = i)\}$ is the prior probabilities of s_i being the initial state of the demand observations. $A = \{a_{ij}\}$ is the state transition probabilities matrix. $E = \{p_1, p_2, \dots, p_S\}$ is the probability density functions of the observations in hidden states where $p_j(D_t) = P(D_t|S_t = j)$. Given this form of HMM, three basic problems must be solved for the model[1]. We use the mathematical programming formulation of HMM to facilitate representing these fundamental problems implicitly (Qin et al. 2000). The optimal solution of the following formulation is the model parameters set λ that is most likely to generate the observed demand sequence

$$\begin{aligned} \max_{\pi, A, E,} \quad & \Lambda = \log[P(D|\lambda)] \\ \text{s.t.:} \quad & A \cdot 1 = 1 \\ & \pi \cdot 1 = 1, \end{aligned} \tag{P-HMM}$$

where the first constraint characterizes the transition probability from hidden state S_{t+1} into S_t , and $\sum_j \{a_{ij}\} = 1$, and the second one satisfies the relation between observation D_t and hidden state S_t at time t, and $\sum_j \{p_j(D_t)\} = 1$.

[1] Rabiner (1989) clarifies that HMM design involves three problems; evaluating the probability or likelihood of a sequence of observations using particular parameters of HMM; identifying the best sequence of states; and adjusting the model parameters so that they explain the occurrence of the observations as much as possible.

(2) Newsvendor problem: The second problem is finding the network that sets the best order quantity given the state sequence $S = S_{1:T}$ estimated from the first problem that has most probably generated the demand sequence

$$\begin{aligned} \min_{Q,\mathbf{W}^{NV},B_i} \quad & NVC(Q|S) \\ \text{s.t.:} \quad & Q_t = B_i + \psi^{NV}(\mathbf{F}_t, \mathbf{W}^{NV}) \quad ;\forall t = 1,\dots,T,\ i \in S \\ & \mathbf{W}^{NV} \in \mathcal{R}. \end{aligned} \tag{P-NV}$$

However, these two objectives do not always align necessarily. We use a linear scalarization method to formulate these two problems as a single-objective optimization

$$\begin{aligned} \min_{Q,\mathbf{W}^{NV},B_i,\lambda,\mathbf{q}} \quad & \text{Loss} = -\eta\Lambda + (1-\eta)NVC \\ \text{s.t.:} \quad & A \cdot 1 = 1 \\ & \pi \cdot 1 = 1 \\ & Q_t = B_i + \psi^{NV}(\mathbf{F}_t, \mathbf{W}^{NV}) \quad ;\forall t = 1,\dots,T,\ i \in S \end{aligned} \tag{P-HMMNV}$$

where $\mathbf{q}$ is the probability of the states effective at the time of making decision on order quantity Q. In Appendix B Section B.1, we show that how one can adjust the order quantity by changing the trade-off coefficient η in (P-HMMNV) to counter the effect of the probability of the states[2].

To solve the above problem, we consider deep learning as it is one of the machine learning methodologies that can model both highly non-linear functions and the Markov chain using historical data and its training can be performed efficiently using gradient methods[3]. In the following, we describe the deep neural network briefly.

3.4. Deep neural networks

Deep neural networks are a sub-category of neural networks. Neural networks are machine learning models originally inspired by biological processes. Neural networks are extremely capable of approximating highly non-linear functions. Neural networks are widely studied and widely used in machine learning and have various applications, especially in image and speech recognition (Gurney 2018).

A neural network consists of several nodes that are connected, forming a directed graph. Each node/neuron function receives signals from its upstream neurons and passes the aggregate signal

[2] We thank an anonymous referee for this suggestion.

[3] It is important to note that, in most of the data-driven approaches of the newsvendor problem, a regularization term is added to the objective function (Oroojlooyjadid et al. 2020). However, this issue is more critical in the cases labeled as "fat data" where there are many input feature variables in the model (Ban and Rudin 2019). In the present study, the number of features is small and we do not incorporate the regularization term explicitly in the objective function, rather we handle overfitting by performing training and evaluation in the training step.

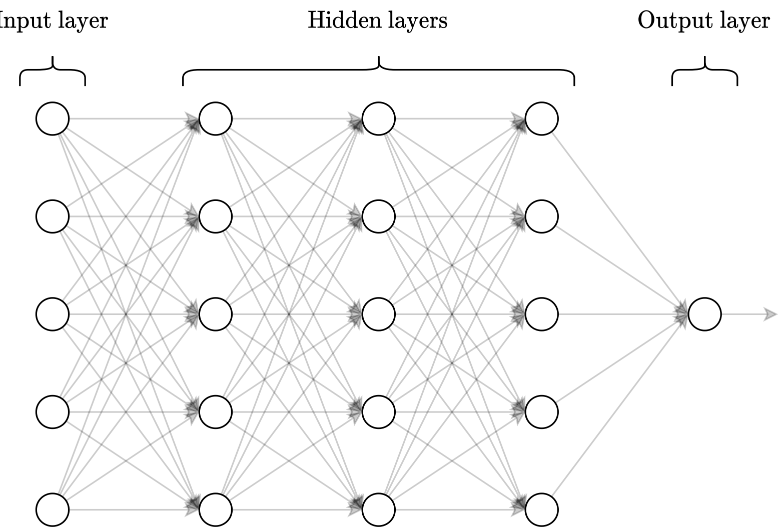

Figure 2 **A deep neural network.**

to an activation function. The output of the activation function then in turn is passed to the downstream neurons. Activation functions are typically monotone increasing functions that map the set of real numbers to a finite interval e.g. [0 1]. Some commonly used activation functions include the sigmoid and tanh functions. A deep neural network is a neural network with various hidden layers between the inputs and the outputs. Figure 2 depicts a symbolic neural net.

The data that we are interested in modeling in this work has a time dimension that is important in understanding the demand. The time dimension can be incorporated into a neural network by folding a neural network that takes the inputs from different times via different neurons into a network that takes in similar features from different time periods, from the same neuron. The folded network is referred to as a recurrent network.

A neural network is fitted to a given set of data points by changing the weights of the arcs that connect the neurons. Namely, the goal in training a neural network is decreasing the total error of the model in predicting the output variable by changing the network weights. This problem, in general, can be very complicated, however, it has become computationally much less burdensome thanks to the backpropagation algorithm. The backpropagation algorithm is a gradient-based method that iterates between forward and backward passes through the network modifying the weights of the arcs and calculating the gradients in the network.

3.5. Deep neural network for solving Markov modulated and data-driven newsvendor

In this study, we propose a new algorithm referred to as HMMNV that utilizes the machine learning method of Deep Neural Networks (DNN) for solving the proposed model. We unify the objective functions of the two problems (P-HMM) and (P-NV) by integration of Markov chain and the newsvendor problem in a network. To this end, we propose a two-head neural network in which these objective functions are combined in a single function as in problem (P-HMMNV) and optimized simultaneously. The suggested network comprises two networks of HMM and the newsvendor. The most likely sequence of states is obtained from the available data by the HMM network. The state information, $\mathbf{q}$, that influences the order quantity partially along with the available features as the inputs of the newsvendor network completes the order quantity at each time period. Figure 3 shows this integration and represents the folded version of the expanded recurrent network over time.

In addition to these two networks, we estimate the base demand B and the function ψ in Equation (11) by a separate neural network in each state. We refer to this network as the emission network whose outputs are used as likelihoods of an observation given some model parameters. The emission network is embedded into the HMM network.

In order to estimate the hidden Markov model and train its network we can unfold the recurrent network and treat it as a feed forward network. Feeding and backpropagation steps of a neural network estimation is equivalent to the forward and backward steps of the well-known Baum-Welch algorithm which is used to estimate HMM. The likelihood of each demand observation is estimated by the probability density functions based on a normal distribution as

$$p_i(D_t) \sim \mathcal{N}\left(\mu_i + \psi^E(\mathbf{F}_t, \mathbf{W}^E), \sqrt{\sigma_i^2 + \sigma_\epsilon^2}\right), \tag{13}$$

where, μ_i is the mean of the base demand in state i. In the forward step, a forward term is defined at each time for each state of the model as

$$\alpha_t(s) = P(D_{1:t}, S_t = s) = P(D_{1:t}, S_t = s, S_{t-1} = i),\ \ D_{1:t} = \{D_1, D_2, \cdots, D_t\}, \tag{14}$$

where $D_{1:t} = \{D_1, D_2, \cdots, D_t\}$. Using the chain rule and rewriting for $P(D_{1:t}, S_t = s, S_{t-1} = i)$, we then have

$$\begin{aligned}\alpha_t(s) = \sum_{i=1}^{S} & P(D_{1:t}|S_t = s, S_{t-1} = i, D_{1:t-1}) \\ & P(S_t = s|S_{t-1} = i, D_{1:t-1}) P(S_{t-1} = i, D_{1:t-1}).\end{aligned} \tag{15}$$

Since the last observation D_t is conditionally independent of everything but S_t, and in the Markov model, we know that S_t only depends on S_{t-1}, Equation (15) could be written as

$$\alpha_t(s) = P(D_{1:t}|S_t = s) \sum_{i=1}^{S} P(S_t = s|S_{t-1} = i)\alpha_{t-1}(i). \tag{16}$$

Finally, the probability of being in state j for a new time $t+1$ is

$$P(S_{t+1} = j) = \sum_{i=1}^{S} \alpha_t(i) a_{ij}, \tag{17}$$

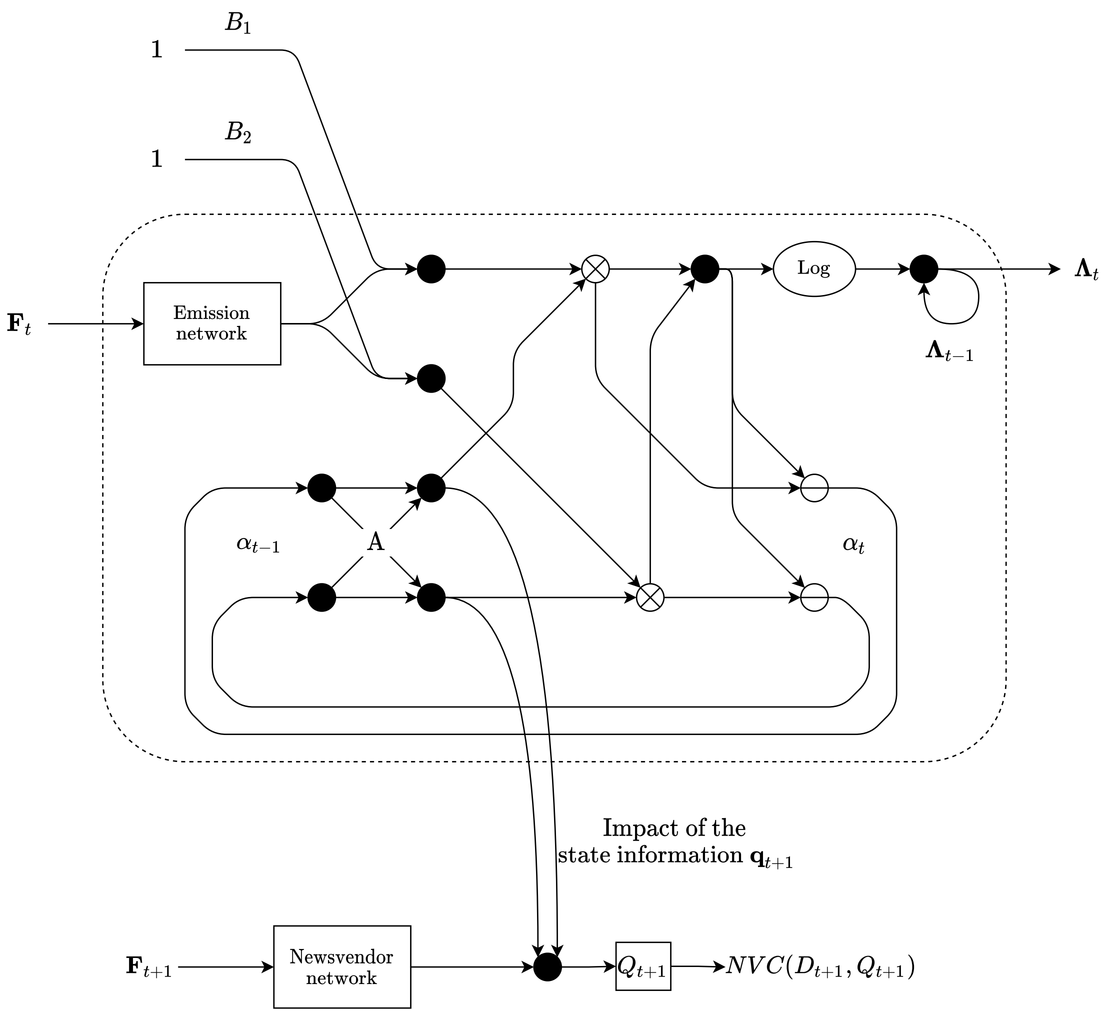

Figure 3 **HMMNV network. This figure includes three neural networks: 1- Hidden Markov Model (HMM) network (dashed rectangle), 2- Newsvendor network, and 3- Emission network. The emission network is embedded into HMM network. The network weights denoted by B and A are the base demand associated with each state and the transition probabilities between states, respectively. All other connections have weight=1. Filled units (circles) denote summation, crossed units multiply their inputs, and units divided by a line, divide one input by the other. In fact, this network is repeated in the number of observations so that constructs a network over time (recurrent network).**

where a_{ij} is the most recent estimation of the probability of transition from state i to state j. Let $q_{t+1}(j)$ denote the probability obtained by Equation (17). The vector $\mathbf{q}_{t+1} = [q_{t+1}(j = 1), \ldots, q_{t+1}(j = S)]$ contains the probabilities of all states at time $t+1$ which sum to one. Vector $\mathbf{q}_{t+1}$, as the state information, is then multiplied by some network weights and builds the base demand which is represented as the connection between HMM and newsvendor networks in Figure 3. The sum of the values obtained from state-dependent and feature-dependent is the estimation for the order quantity.

In the estimation process, scaling is required in implementation of forward-backward algorithm. Let us consider the forward term of Equation (14) in the re-estimations procedure and write as

$$\begin{aligned}\alpha_t(s) &= P_\lambda(D_0, \cdots, D_t, S_t = s)\\ &= \sum_{s_1 s_2 \ldots s_{t-1}} P_\lambda(D_0, \cdots, D_t, S_t = s \,|s_1 s_2 \ldots s_{t-1}) P_\lambda(s_1 s_2 \ldots s_{t-1})\\ &= \sum_{s_1 s_2 \ldots s_{t-1}} \left[\prod_{i=1}^{t} p_{s_i}(D_i) \prod_{i=1}^{t-1} a_{s_i s_{i+1}}\right]\end{aligned} \tag{18}$$

3.5.1. Scaling the terms: All the involving terms in Equation (18) are on a probability scale meaning that they are less than one. Therefore, the summation rapidly drops to zero with an exponential rate. The result is too small and may exceed the machine precision and relative errors round to zero in floating point. To solve this problem, a scaling is proposed and used by Levinson et al. (1983) to keep all $\alpha_t(s)$'s bounded at each induction step. This scaling factor only depends on time t and not the current state s. Corresponding computations include two parts:

- Initialization

$$\begin{aligned}\ddot{\alpha}_1(i) &= \alpha_1(i),\\ c_1 &= \frac{1}{\sum_{i=1}^{S} \ddot{\alpha}_1(i)},\\ \hat{\alpha}_1(i) &= c_1 \ddot{\alpha}_1(i)\end{aligned} \tag{19}$$

- Induction

$$\begin{aligned}\ddot{\alpha}_1(i) &= \sum_{j=1}^{S} \hat{\alpha}_{t-1} a_{ji} p_i(D_t),\\ c_t &= \frac{1}{\sum_{i=1}^{S} \ddot{\alpha}_t(i)},\\ \hat{\alpha}_t(i) &= c_t \ddot{\alpha}_t(i)\end{aligned} \tag{20}$$

The coefficient c_t at each step depends on t. $\hat{\alpha}_t(i)$ is the modified forward variable which sums to one, $\sum_{i=1}^{S} \hat{\alpha}_t(i) = 1$. It is easy to see that

$$\hat{\alpha}_t(i) = \left(\prod_{\tau=1}^{t} c_\tau\right) \alpha_t(i). \tag{21}$$

By using this new forward algorithm, obtained in the last step, we have

$$\begin{aligned}1 &= \sum_{i=1}^{S} \hat{\alpha}_T(i) = \sum_{i=1}^{S} \left(\prod_{t=1}^{T} c_t\right) \alpha_T(i)\\ &= \left(\prod_{t=1}^{T} c_t\right) \sum_{i=1}^{S} \alpha_T(i) = \left(\prod_{t=1}^{T} c_t\right) P(D|\lambda).\end{aligned} \tag{22}$$

Let $\mathbf{C} = \prod_{\tau=1}^{t} c_\tau$, then $P(D|\lambda) = 1/\mathbf{C}_T$. The logarithmic form of the likelihood function is then

$$\Lambda = \log[P(D|\lambda)] = -\sum_{t=1}^{T} \log c_t. \tag{23}$$

In the next step, we obtain the partial derivatives of the function Λ with respect to all network parameters.

3.5.2. Backpropagation step: The backpropagation step of the HMM network includes the partial derivative of the likelihood function (23) with respect to transition probabilities a_{ij} and emissions $p_i(D_t)$, which are calculated as

$$\frac{\partial \Lambda}{\partial a_{ij}} = \sum_{t=1}^{T} \frac{\partial \Lambda}{\partial c_t} \frac{\partial c_t}{\partial \ddot{\alpha}_t(i)} \frac{\partial \ddot{\alpha}_t(i)}{\partial a_{ij}}, \tag{24}$$

$$\frac{\partial \Lambda}{\partial a_{ij}} = \sum_{t=1}^{T} c_t \hat{\alpha}_{t-1}(i) p_j(D_t), \tag{25}$$

$$\frac{\partial \Lambda}{\partial p_i(D_t)} = \sum_{t=1}^{T} \frac{\partial \Lambda}{\partial c_t} \sum_{j=1}^{S} \frac{\partial c_t}{\partial \ddot{\alpha}_t(j)} \frac{\partial \ddot{\alpha}_t(j)}{\partial p_i(D_t)}, \tag{26}$$

$$\frac{\partial \Lambda}{\partial p_i(D_t)} = \sum_{t=1}^{T} \sum_{j=1}^{S} c_t \hat{\alpha}_{t-1} a_{ji}. \tag{27}$$

Derivatives obtained from Equation (27) are used for updating two parameter sets of the base demand vector $\mathbf{B} = [B_1, B_2, \ldots, B_S]$ and the weights of the emission network $\mathbf{W}^E$

$$\mathbf{B}_{n+1} = \mathbf{B}_n + \gamma_1 \left(\frac{\partial \Lambda}{\partial p_i(D_t)} \frac{\partial p_i(D_t)}{\partial \mathbf{B}_n} \right). \tag{28}$$

The input associated with base demand is the unit vector as in Figure 3. Therefore, the second partial derivative of $\frac{\partial p_i(D_t)}{\partial B_n} = 1$.

Regarding the update of the emission network weights we have

$$\mathbf{W}^E_{n+1} = \mathbf{W}^E_n + \gamma_1 \left(\frac{\partial \Lambda}{\partial p_i(D_t)} \frac{\partial p_i(D_t)}{\partial \mathbf{W}^E_n} \right), \tag{29}$$

where, the first partial derivative is obtained in (27) and the second one, partial derivatives of emissions $p_i(D_t)$ with respect to the network weights, is related to the backpropagation step in emission network $\mathbf{W}^E_n$, which is explained in Appendix A.

The newsvendor network is a feed forward network which is trained by backpropagation algorithm and the weights are updated using the derivatives of the cost function with respect to $\mathbf{W}^{NV}$

$$\mathbf{W}^{NV}_{n+1} = \mathbf{W}^{NV}_n - \gamma_2 \left(\frac{\partial NVC}{\partial \mathbf{W}^{NV}_n} \right). \tag{30}$$

Further details of the partial derivative term are provided in Appendix A.

The transition matrix A is the joint part of the HMM and the newsvendor networks. Consequently, both partial derivatives of which in Equation (25) plus the partial derivative of the newsvendor cost function with respect to A are used to update the transition matrix

$$A_{n+1} = A_n - \gamma_3 \left(-\eta \frac{\partial \Lambda}{\partial A_n} + (1-\eta) \frac{\partial NVC}{\partial A_n} \right), \tag{31}$$

where the term in parenthesis is the partial derivative of the single objective function (P-HMMNV) with respect to A

$$A_{n+1} = A_n - \gamma_3 \left(\frac{\partial \text{Loss}}{\partial A_n} \right). \tag{32}$$

3.5.3. Network specification: We consider two hidden layers in the newsvendor network following the rule proposed by Huang (2003) that determines the number of hidden nodes in each layer. Further, we consider a single hidden layer for the emission network in which the number of hidden nodes are specified based on the formula suggested by Ke and Liu (2008). In training procedure of the HMMNV model, we choose the candidates for the trade-off coefficient η from the set of $\{0.001, 0.01, 0.1, 0.9, 0.99, 0.999\}$. In order to find the best learning rates for each network, a grid search over the set $\{0.001, 0.01, 0.1, 2, 10\}$ is used for all learning rates. We choose the best candidate among these parameters based on a cross validation step on the training data set. We then train and test the model on all data set using the best parameter chosen by cross validation.

Figure 4 indicates how the sample observations are divided into different sets so that one can implement the algorithms. First, we pick a smaller set of the data and divide it into training and test samples for inner cross-validation to find the best parameters. We then train the algorithm with the best selected parameter on the entire smaller set that we chose initially and test the model on the other test set to evaluate the performance of the model.

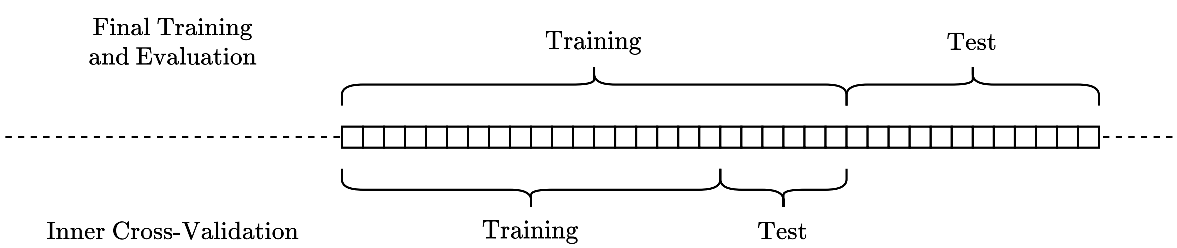

Figure 4 **Different sets of the sample observations used for cross-validation, training, and test.**

4. Analyses

In this section, we test the performance of HMMNV model using simulated data. In Section 4.1, we introduce several alternative methods as benchmarks. In Section 4.2, we describe our numerical setup to evaluate the performance of HMMNV model with different benchmarks, and present and highlight the results in Section 4.3.

4.1. Benchmarks and true model

In the following, we list the benchmark methods and explain more specifically how they solve the intended newsvendor problem. We also provide results for the newsvendor cost that is obtained by the true model as a near perfect benchmark. The true model is used as a baseline to compare all the methods.

EDD: The empirical demand distribution approach uses only the demand observation and finds the optimal solution based on the empirical distribution which assumes equal weights for each observation. The optimal quantity is obtained by the known formula.

ObBW: The Objective-blind Baum-Welch algorithm considers the Markov chain which subordinates the demand (Avci et al. 2020, Treharne and Sox 2002). Consequently, it finds the most probable sequence of states by the predetermined distribution for demand. Therefore, this is a parametric approach that fits a distribution on demand and estimates the appropriate parameters for each state of the Markov model. Baum-Welch algorithm is a well-known forward-backward method in the estimation of hidden Markov states using observable data (i.e. demand). Details of this algorithm are described by Rabiner (1989).

PfLR: The parameter fitting linear regression is the first and the simplest method which takes the feature data into account. It consists of two steps: estimating the parameters of the demand distribution and then use the estimations in optimization problem (Ban and Rudin 2019). However, this approach ignores the dependency between demand observations through the hidden Markov states. That said, it is a dynamic approach that makes a linear relationship between features and demand by the following linear regression

$$D_t = \beta_0 + \beta_1 f_{1t} + + \beta_2 f_{2t} + \ldots + \beta_N f_{Nt} + \epsilon_t, \tag{33}$$

where $t = 1, \ldots, T$, and N is the number of features. The time-varying mean for demand and the standard deviation of the estimated normal distribution is calculated by the coefficients and residuals of the regression in (33). The standard deviation of the demand distribution is equivalent to the standard deviation of the residuals. The optimal order quantity would be

$$Q_t^* = \beta_0 + \beta_1 f_{1t} + \beta_2 f_{2t} + \ldots + \beta_N f_{Nt} + z^* \sigma_\epsilon, \tag{34}$$

where $z^* = \Phi^{-1}(b/(h+b))$.

LML: Ban and Rudin (2019) suggest a linear decision rule for the order quantity as

$$Q = \left\{ q : \mathcal{X} \to \mathbf{R} : q_t(\beta) = \beta_0 + \sum_{n=1}^{N} \beta_n f_{nt} \right\}. \tag{35}$$

They substitute this linear formula into the newsvendor problem and solve a nonlinear program for the optimal order quantity.

DNN: Oroojlooyjadid et al. (2020) apply a deep neural network to the newsvendor problem which is a nonlinear extension to the LML model. In this approach, a neural network maps the feature data to the optimal decision. The goal of the network is to provide the minimum average cost value over all periods

$$\min_{\mathbf{W}} \frac{1}{T}\sum_{t=1}^{T} NVC(\theta(\mathbf{F}_t, \mathbf{W}), D_t), \tag{36}$$

where $\mathbf{W}$ is the matrix of the network weights, $\mathbf{F}_t$ is the vector of input features at time t, and the network is indicated by the mapping function θ. In order to have a network structure in DNN method similar to HMMNV method, we use the same hyper parameters selecting rule (network specification including the number of hidden layers and the number of hidden nodes in each layer) explained for HMMNV method.

True model: The true model is used as a benchmark where the prediction of the state and the effects of features are done perfectly. Although this scenario is impossible in practice, it can be used as a benchmark to capture the difference between the performance of the methods discussed here with an *ideal* method. In the true model, we assume that one knows the exact parameters of the model. Two sets of parameters exist, first, the states of the system which are hidden and unobservable to all the other methods, second, the distribution of the demand at each time which is the sum of two normal distributions related to states referred to as base demand and the feature part of the demand. More specifically, current state i, associated mean and variance of the base demand denoted by μ_i and σ_i^2, respectively, the exact function $\psi(\mathbf{F}_t, \mathbf{W})$ and the variance σ_ϵ^2 are known. Therefore, the factors that cause some costs in the true model are the two variance terms of σ_ϵ^2 and σ_i^2 related to the features part and base demand distributions. The true optimal order quantity is then obtained by the known formula as $Q_t^* = \mu_i + \psi(\mathbf{F}_t, \mathbf{W}) + z^*\sqrt{\sigma_i^2 + \sigma_\epsilon^2}$, where $z^* = \Phi^{-1}(b/(b+h))$ and $\Phi()$ is the cumulative distribution function of a standard normal random variable.

In the next section, for each method, the results are reported as the percentage deviation of the cost obtained by a method with respect to the cost for the true model,

$$\text{Percentage deviation of a method} = \frac{\text{Cost of a method} - \text{Cost of the true model}}{\text{Cost of the true model}} \times 100.$$

4.2. Numerical Experiments

In this section, we first present the experimental setup that we have used to evaluate the performance of our method compared to the benchmarks. In our experimental setup, the demand in each period is a normally distributed random variable. The mean of this distribution in each period is determined based on the state of a Markov chain and a number of randomly generated

Table 3 **The set of parameters used in the experimental setup.**

Parameter	Description	Domain
e	transition probability	0.01 0.1 0.2
μ_1	mean of the base demand in state 1	1 2 3
μ_2	mean of the base demand in state 2	1
σ_1	standard deviation of the base demand in state 1	0.5 1
σ_2	standard deviation of the base demand in state 2	0.5
σ_ϵ	standard deviation of the demand by features	0.5
N	number of features	1 3 5
W	network weights	1 2 3
L	number of hidden layers	1 2 3
b/h	newsvendor cost parameters ratio	2 5 10
T	number of observations	200 400 800

features for that period. Table 3 shows the domain for each parameter of the system. According to Figure 1, we evolve the base demand by a two-state Markov chain model, and the additional part of the demand is generated by using a neural network and feature data. The final demand is the summation of state-dependent and feature-dependant demand values. Parameter e relates to the transition probability between two states of the Markov chain. The transition matrix is then determined as

$$\begin{array}{c c} & \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \end{array} & \begin{pmatrix} 1-e & e \\ e & 1-e \end{pmatrix} \end{array},$$

where s_1 and s_2 indicate two states of the Markov chain. Regarding the base demand in each state, we assume that they have normal distribution with parameters (μ_1, σ_1) and (μ_2, σ_2), respectively. We change the parameters of the first state within the range specified in Table 3 and fix those of the second state on $(\mu_2 = 1, \sigma_2 = 0.5)$. N is the number of features observed before the realization of the demand. All the features are randomly drawn from the standard normal distribution. We consider three different network weights denoted by W. Each network weight is a random variable with the standard normal distribution. The structure of the network, including the number of hidden layers makes the relation between the features and demand more complex as the number of hidden layers increases. In our experiments, we consider up to three layers denoted by L. We consider 10 hidden nodes in each hidden layer where a sigmoid function serves as the activation function. We also consider three different values for the ratio between the cost rates in the problem (b/h). Finally, the parameter T is the number of observations. We evaluate the methods on all 5832 combinations that these parameter sets provide.

4.3. Results

In the following, we give a summary of the results of our experiments. In pairwise comparisons, HMMNV outperforms DNN in 64 % of the cases. HMMNV outperforms EDD in 60 % of the cases.

HMMNV outperforms all the methods in 37 % of the cases. EDD outperforms all methods in 15 % of the cases.

Figures 5 presents the results of these experiments. In this figure, each point on the plot corresponds to a data with a certain parameter set where the x-axis is labeled by the algorithms and the value of y-axis represents the percentage deviation from the cost of the true model associated with each parameter set. A box plot is also depicted on the scattered result points of each algorithm. HMMNV obtained a lower mean and deviation from the mean compared to other methods. ObBW and EDD do not seem to perform well. The other three methods PfLR, LML, and DNN perform similarly. This grouping in performance implies that both sources that affect demand including observable features and non-observable Markov states have to be taken into account.

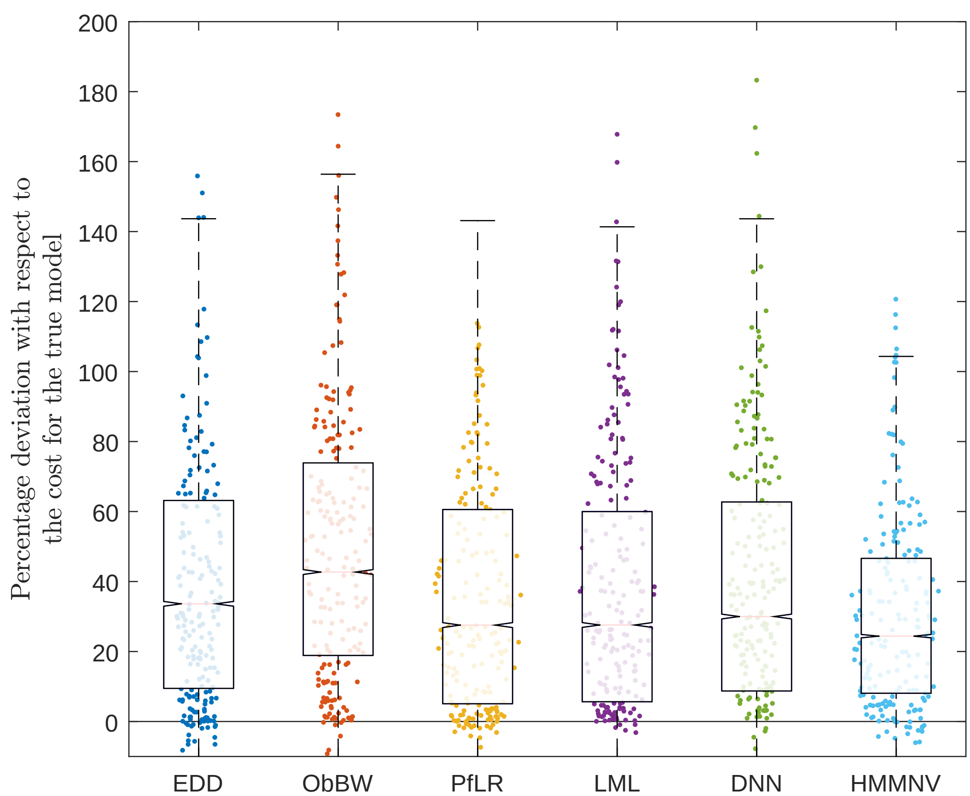

Figure 5 The result of all algorithms for different parameter sets.

To show the outperformance of each method over the others, we also implement a t-test on cost values two by two. This test proves if there is a meaningful and statistically significant difference between the results of any two algorithms. Table B.2 shows the p-values of the test. If the algorithm in row i has a lower cost than algorithm in column j, the cell ij of the table is filled by their p-value

Table 4 p-values of the t-test between the cost of algorithms.

	EDD	ObBW	PfLR	LML	DNN	HMMNV
EDD		3.7894e-100				
ObBW						
PfLR	7.7860e-13	9.8089e-168		0.8860	1.0743e-06	
LML	3.5083e-12	1.1297e-163			2.6823e-06	
DNN	0.0312	2.7515e-115				
HMMNV	1.6786e-80	4.6684e-299	2.4323e-35	1.5093e-35	9.6652e-63	

obtained from the test otherwise it is left empty. A p-value lower that 1 % indicates that the mean cost of an algorithm is less than the other one at 1 % significance level. The HMMNV has a lower cost compared to others and the p-values close to zero confirm this. EDD as the simplest method outperforms only the ObBW. ObBW has shown higher cost compared to all other methods. PfLR and LML are very close and they outperform the DNN. DNN is better than EDD and ObBW at 5 % and 1 % levels, respectively.

In order to investigate the effect of each parameter on the results and show how the algorithms differ over the range of the parameters, we plot the deviations of the methods for each parameter separately. Figure 6 represents the results for parameter values in table 3. The parameter e which refers to the transition probability of the hidden Markov is a representative for the information about the non-observable states. A lower value of e indicates more stability for Markov chain and consequently greater importance and effect caused from states on demand. As a result we can see

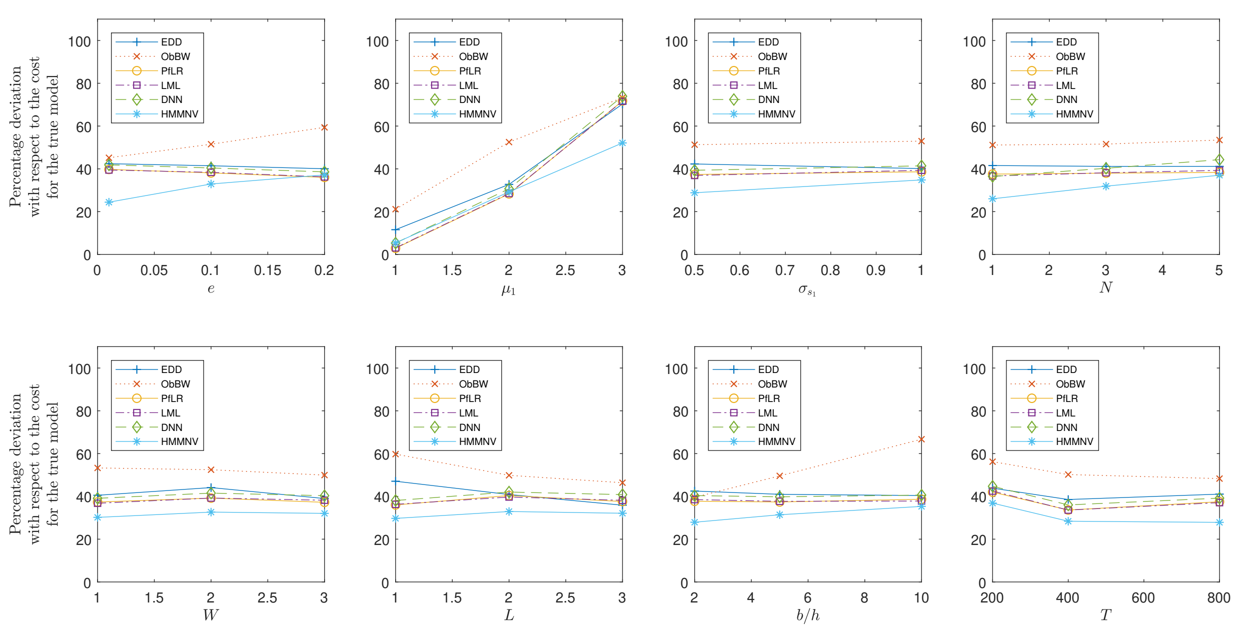

Figure 6 Deviations of algorithms' cost from the true model against the range of each parameter.

that the HMMNV method have lower cost for the smallest value of e. As e increases the methods except ObBW converges which means that the systematic effect of Markov states on randomness of the demand decreases. The parameter associated with the mean of the base demand in two states affects the results more than other parameters. As μ_1 increases, the long term dependencies influence the demand more than observable features. As a result, HMMNV performs better than the benchmarks in cases with higher imbalance in base demands. Other parameter of the base demand which is the standard deviation in not effective as much as the mean and all methods except HMMNV are insensitive to that. Similar to the mean of the base demand, standard deviation imposes more costs on our method as its value in one state differ more from the other state. The fourth plot shows the effect of the number of observable features indicated by N. It is reasonable that methods which don't use features like EDD and ObBW are not sensitive to N. Other methods converge as N increases. This results from the noises that features add to the demand. If features compose the underlying pattern, one can gets better results by incorporating them to the model. A specific value of the parameter W which indicates the relations between features and demand is not meaningful compared to its other values but lower deviations over all W show the power of the HMMNV algorithm in achieving optimal policy with different random relations compared to others. The complexity of these relations stems from the structure of the networks and the associated number of hidden layers. It is observed that HMMNV results in lower cost regardless of the nonlinear complexity. The outperformance of the HMMNV method is obvious over different cost imbalance imposed by the b/h ratio. Deviations increase as this ratio increases in all methods. Regarding the number of observations, the results represent that our method works better as T increases. This is actually derived by the Markov sequence which is well captured using more observations. Briefly, all these analyses prove the robustness of the suggested algorithm that results in a lower average cost with respect to other methods.

In a separate experiment, we examine the performance of the suggested algorithm in some extreme situations when historical demand observations contain some outliers that have low-frequency records. We generate new sample sets using different model parameters and show the robustness of our model in these extreme scenarios. The results of these experiments are given in Appendix B Section B.2[4].

5. Real Data Experiment: Crude Oil Demand as a Proxy

We evaluate the performance of our algorithm and the benchmark methods using real data of the U.S. weekly crude oil demand. This data can be taken as a proxy for a product or service whose demand is closely correlated with U.S. crude oil. Food crops and agricultural commodities including

---

[4] We thank an anonymous referee for this suggestion.

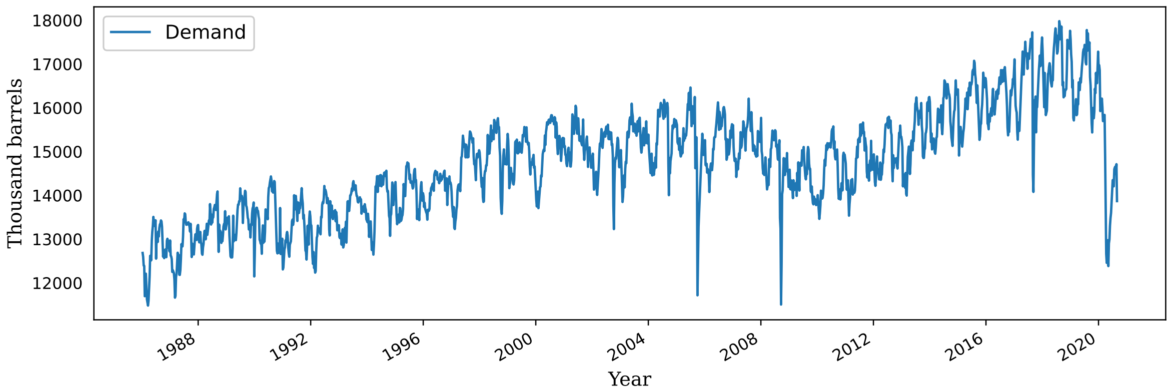

Figure 7 **Time series of the weekly demand for crude oil from January 1986 to August 2020. The average demand is 14731 thousand barrels per week.**

corn, wheat, rice, and sugar are some newsvendor products that are shown to have a correlation with crude oil (Du et al. 2011, Mokni and Youssef 2020).

The crude oil demand data consists of weekly demand and covers the period from January 1986 to August 2020[5]. The series of the demand is shown in Figure 7. The set has 1800 weekly data sample observations. The set of features for the model includes 16 dummy variables representing the quarter of the year and the month of the year. These features capture the observable variation in demand which derives the seasonality. However, the variations caused by environmental randomness are not observable, we assume that they exist and follow a Markov chain model with two states. Our algorithm use the feature data to explain the additional demand and model the residuals as base demand which are not captured by features. Base demand is modeled by Markov chain and evolves the demand over time along with the additional demand.

We pick a time window with the length of 500 weeks and train the algorithms on the first 400 weeks and test the trained models on the remaining 100 weeks. We roll this window by 100-week forward steps and obtain 14 test periods in total. Table 5 summarizes the result of the algorithms. This table shows the average cost of each method over 14 test periods (1400 weeks) for each of newsvendor cost parameters ratio (b/h).

It is observed that our proposed method outperforms other approaches for all values of cost parameters ratio. The second best approach is ObBW that uses demand and fits a Markov model. Its outperformance suggests that the states of the environment have long dependency and influence the demand in a state-dependent manner. The benchmark methods which use features have similar results with minor differences. Briefly, ObBW and EDD are still the best benchmarks among others.

[5] Data is collected from the U.S. Energy Information Administration at: https://www.eia.gov

Table 5 **Average cost of the crude oil ordering policy by each algorithm (values are divided by 10^3).**

b/h	EDD	ObBW	PfLR	LML	DNN	HMMNV
2	1.161	0.840	1.150	1.123	1.117	0.477
5	1.660	1.262	1.751	1.701	1.703	0.650
10	2.158	1.625	2.282	2.274	2.244	1.557
20	2.871	2.134	2.847	3.104	3.178	1.832

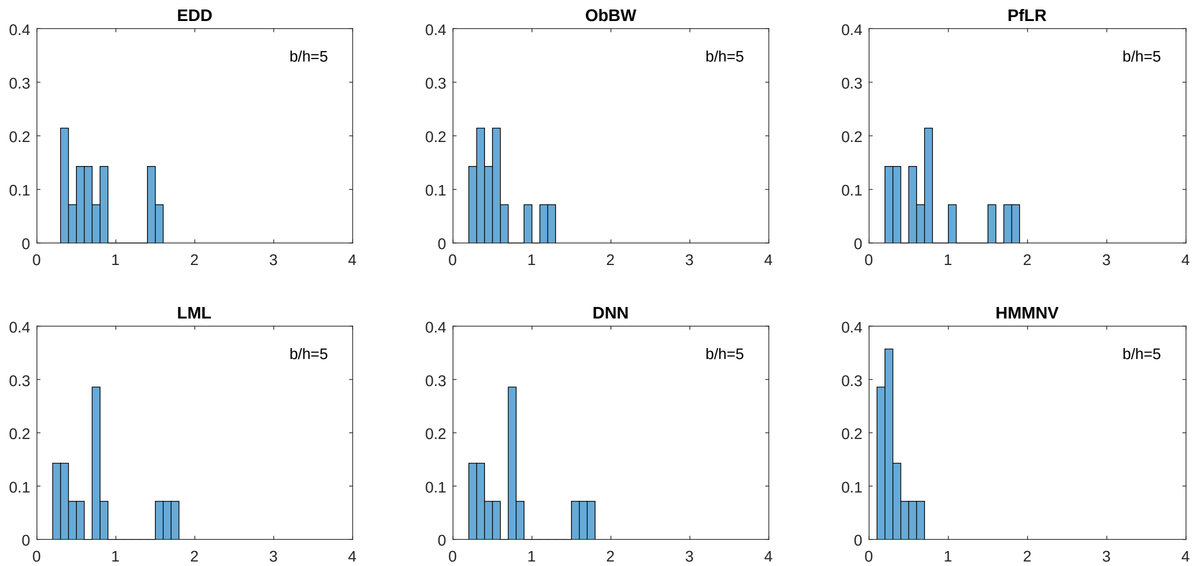

Figure 8 **The frequency of the cost values obtained by each algorithm for the real data over all test periods.**

We conclude that if the randomness of the features and the environment are combined in a model, the cost decreases drastically. [6]

Figure 8 shows the frequency of costs over 14 test periods for all methods. The suggested method offers an ordering policy that results in lower costs than other methods. In order to show the states obtained by HMMNV method and the optimal policy, we plot the demand and the order quantity for 200 recent out of sample weeks in Figure 9. It is observed that the order quantity tracks the demand closely, imitating the present variations in the demand. The bottom panel of this figure shows the estimated sequence of the states during this period. When comparing the order and

[6] In addition to these five benchmarks, we also consider the well-known method of Holt-Winters that explicitly accounts for seasonality and trend in time-series forecasting literature. To use this method in the newsvendor problem setting, we utilize it to forecast the demand and then use the forecast as the order quantity. This approach is referred to as Estimate-As-Solution (EAS) in the literature (Oroojlooyjadid et al. 2020). For the trend and seasonality components in the Holt-Winters method, we consider the trend component as an additive and the seasonality components as multiplicative, since the Holt-Winters model performs well using these components compared to other types of components. The mean values of newsvendor cost for the Holt-Winters method are as follows. For b/h =2, 5 ,10, and 20, the mean of costs are 0.907, 1.826, 3.370, and 6.419, multiplied by 10^3, respectively. Comparing these results with those in Table 5 indicates that the means of all costs are smaller when we use HMMNV to find the optimal order quantity compared to the Holt-Winters approach. We thank an anonymous referee for this valuable suggestion.

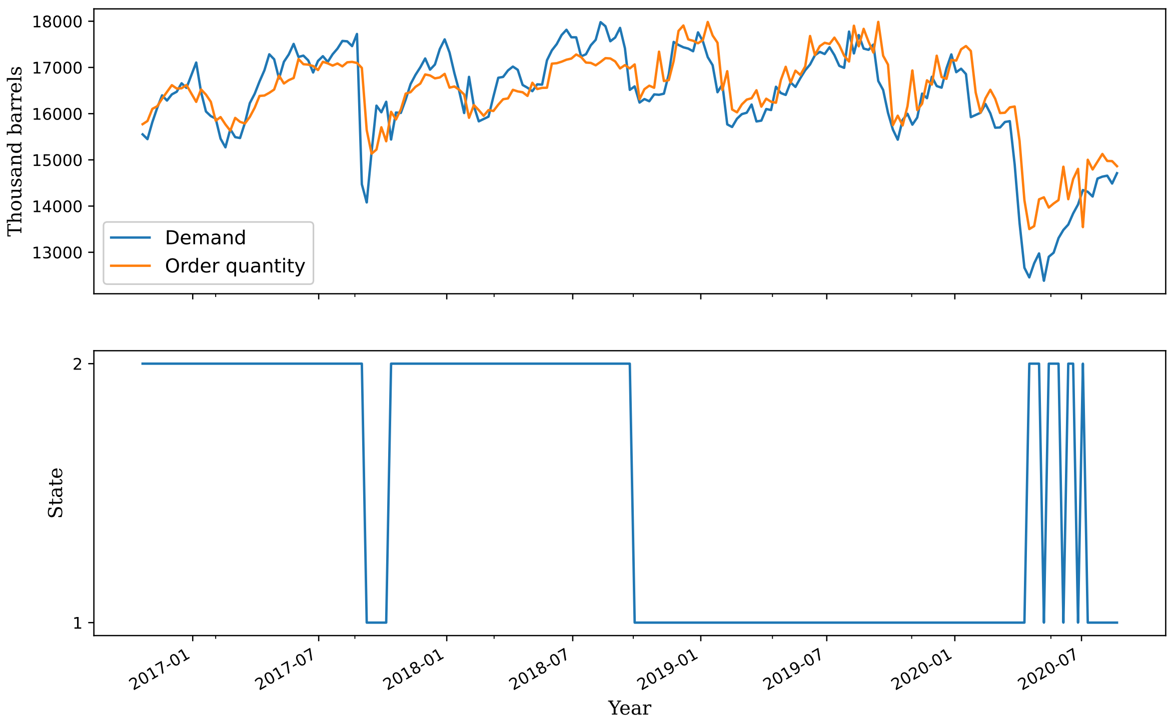

Figure 9 **Optimal order quantity obtained by HMMNV for $b/h = 5$ during 200 weeks of test periods from October 2016 to August 2020 (top panel). Bottom panel shows the corresponding state sequence of the Markov model estimated by this algorithm.**

demand series to the states, we observe that the HMMNV algorithm assigns demands with higher variations to state 1 while the data labeled by state 2 has lower variations. Consequently, the order quantity by HMMNV has the variations similar to that presented in the demand.

We further show the robustness of the suggested method in the real experiment by performing the algorithms using monthly crude oil demand observations. Additionally, we take into account the well-studied features in the literature of crude oil forecasting. These features such as unemployment, S&P 500 stock index, personal disposable income, and etc. capture microeconomic effects on the dynamics of the oil market. Details of the features and the results are reported in Appendix B Section B.3[7].

6. Conclusions

We presented an integrated learning and optimization method based on deep learning for the data-driven newsvendor problem with observable and unobservable features. Feature data about the demand for a product and underlying dynamics of the market that have often dependencies across multiple decision periods are both common factors in the ordering decision and this work provides an approach for effectively learning from both these categories of information.

[7] We thank an anonymous referee for this suggestion.

Through extensive numerical experiments based on synthetic and real data, we assess the performance of a variety of methods for the data-driven control of a newsvendor system and show that our method outperforms the others in a variety of settings.

Although we consider only the ordering problem for a single-item newsvendor in this study, our approach can be extended to multiple items whose demands could be correlated. Another interesting extension is that of a joint price and inventory optimization problem where the demand is dependent on the selling price. In this case, one can investigate how to switch between pricing strategies during various hidden states and use environmental and local features as price drivers.

Acknowledgements: Research leading to these results has received funding from the EU ECSEL Joint Undertaking under grant agreement no. 737459 (project Productive4.0) and from TUBITAK (217M145).

References

Arifoğlu, K. and S. Özekici (2010). Optimal policies for inventory systems with finite capacity and partially observed Markov-modulated demand and supply processes. *European Journal of Operational Research 204*(3), 421–438.

Arifoğlu, K. and S. Özekici (2011). Inventory management with random supply and imperfect information: A hidden Markov model. *International Journal of Production Economics 134*(1), 123–137.

Avci, H., K. Gokbayrak, and E. Nadar (2020). Structural results for average-cost inventory models with Markov-modulated demand and partial information. *Production and Operations Management 29*(1), 156–173.

Ban, G.-Y. and C. Rudin (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research 67*(1), 90–108.

Bensoussan, A., M. Cakanyıldırım, and S. P. Sethi (2005). On the optimal control of partially observed inventory systems. *Comptes Rendus Mathematique 341*(7), 419–426.

Bensoussan, A., M. Çakanyıldırım, and S. P. Sethi (2007). A multiperiod newsvendor problem with partially observed demand. *Mathematics of Operations Research 32*(2), 322–344.

Bertsimas, D. and N. Kallus (2020). From predictive to prescriptive analytics. *Management Science 66*(3), 1025–1044.

Bertsimas, D. and A. Thiele (2005). A data-driven approach to newsvendor problems. *Technical report, Massachusetts Institute of Technology, Cambridge, MA.*

Besbes, O. and A. Muharremoglu (2013). On implications of demand censoring in the newsvendor problem. *Management Science 59*(6), 1407–1424.

Beyer, D. and S. P. Sethi (1997). Average cost optimality in inventory models with Markovian demands. *Journal of Optimization Theory and Applications 92*(3), 497–526.

Bhar, R. and S. Hamori (2004). *Hidden Markov models: applications to financial economics*, Volume 40. Springer Science & Business Media.

Blinder, A. S. and L. J. Maccini (1991). The resurgence of inventory research: what have we learned? *Journal of Economic Surveys 5*(4), 291–328.

Du, X., L. Y. Cindy, and D. J. Hayes (2011). Speculation and volatility spillover in the crude oil and agricultural commodity markets: A Bayesian analysis. *Energy Economics 33*(3), 497–503.

Efendigil, T., S. Önüt, and C. Kahraman (2009). A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis. *Expert Systems with Applications 36*(3), 6697–6707.

Feldman, R. M. (1978). A continuous review (s, s) inventory system in a random environment. *Journal of Applied Probability 15*(3), 654–659.

Gallego, G. and H. Hu (2004). Optimal policies for production/inventory systems with finite capacity and Markov-modulated demand and supply processes. *Annals of Operations Research 126*(1-4), 21–41.

Gallego, G. and I. Moon (1993). The distribution free newsboy problem: review and extensions. *Journal of the Operational Research Society 44*(8), 825–834.

Goel, S., J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts (2010). Predicting consumer behavior with web search. *Proceedings of the National Academy of Sciences*.

Gruhl, D., L. Chavet, D. Gibson, J. Meyer, P. Pattanayak, A. Tomkins, and J. Zien (2004). How to build a webfountain: An architecture for very large-scale text analytics. *IBM Systems Journal 43*(1), 64–77.

Gruhl, D., R. Guha, R. Kumar, J. Novak, and A. Tomkins (2005). The predictive power of online chatter. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pp. 78–87. ACM.

Gurney, K. (2018). *An introduction to neural networks*. CRC press.

Hamilton, J. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics 45*, 39–70.

He, B., F. Dexter, A. Macario, and S. Zenios (2012). The timing of staffing decisions in hospital operating rooms: incorporating workload heterogeneity into the newsvendor problem. *Manufacturing & Service Operations Management 14*(1), 99–114.

Huang, G.-B. (2003). Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Transactions on Neural Networks 14*(2), 274–281.

Huh, W. T., R. Levi, P. Rusmevichientong, and J. B. Orlin (2011). Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Operations Research 59*(4), 929–941.

Ke, J. and X. Liu (2008). Empirical analysis of optimal hidden neurons in neural network modeling for stock prediction. In *Computational Intelligence and Industrial Application, 2008. PACIIA'08. Pacific-Asia Workshop on*, Volume 2, pp. 828–832. IEEE.

Kesavan, S. and T. Kushwaha (2014). Differences in retail inventory investment behavior during macroeconomic shocks: Role of service level. *Production and Operations Management 23*(12), 2118–2136.

Khayyati, S. and B. Tan (2020). Data-driven control of a production system by using marking-dependent threshold policy. *International Journal of Production Economics 226*, 107607.

Levi, R., R. O. Roundy, and D. B. Shmoys (2007). Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research 32*(4), 821–839.

Levinson, S. E., L. R. Rabiner, and M. M. Sondhi (1983). An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell System Technical Journal 62*(4), 1035–1074.

Liyanage, L. H. and J. G. Shanthikumar (2005). A practical inventory control policy using operational statistics. *Operations Research Letters 33*(4), 341–348.

Lovejoy, W. S. (1992). Stopped myopic policies in some inventory models with generalized demand processes. *Management Science 38*(5), 688–707.

Mokni, K. and M. Youssef (2020). Empirical analysis of the cross-interdependence between crude oil and agricultural commodity markets. *Review of Financial Economics 38*(4), 635–654.

Monahan, G. E. (1982). State of the art—a survey of partially observable Markov decision processes: theory, models, and algorithms. *Management Science 28*(1), 1–16.

Oroojlooyjadid, A., L. V. Snyder, and M. Takáč (2020). Applying deep learning to the newsvendor problem. *IISE Transactions 52*(4), 444–463.

Perakis, G. and G. Roels (2008). Regret in the newsvendor model with partial information. *Operations Research 56*(1), 188–203.

Picone, J. (1990). Continuous speech recognition using hidden Markov models. *IEEE ASSP Magazine 7*(3), 26–41.

Qi, M., Y. Shi, Y. Qi, C. Ma, R. Yuan, D. Wu, and Z.-J. M. Shen (2020). A practical end-to-end inventory management model with deep learning. *Available at SSRN 3737780*.

Qin, F., A. Auerbach, and F. Sachs (2000). A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophysical Journal 79*(4), 1915–1927.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*(2), 257–286.

Sachs, A.-L. (2015). The data-driven newsvendor with censored demand observations. In *Retail Analytics*, pp. 35–56. Springer.

Scarf, H. (1958). A min-max solution of an inventory problem. *Studies in the Mathematical Theory of Inventory and Production*.

Scarf, H. (1959). Bayes solutions of the statistical inventory problem. *The Annals of Mathematical Statistics 30*(2), 490–508.

Sethi, S. P. and F. Cheng (1997). Optimality of (s, s) policies in inventory models with Markovian demand. *Operations Research 45*(6), 931–939.

Seubert, F., N. Stein, F. Taigel, and A. Winkelmann (2020). Making the newsvendor smart–order quantity optimization with anns for a bakery chain. *Working Paper*.

Shang, K. H. (2012). Single-stage approximations for optimal policies in serial inventory systems with nonstationary demand. *Manufacturing & Service Operations Management 14*(3), 414–422.

Song, J.-S. and P. Zipkin (1993). Inventory control in a fluctuating demand environment. *Operations Research 41*(2), 351–370.

Treharne, J. T. and C. R. Sox (2002). Adaptive inventory control for nonstationary demand and partial information. *Management Science 48*(5), 607–624.

Van der Laan, N., R. H. Teunter, W. Romeijnders, and O. Kilic (2019). The data-driven newsvendor problem: Achieving on-target service levels. Technical report, Working paper, University of Groningen, SOM research school.

Van Parys, B. P., P. M. Esfahani, and D. Kuhn (2020). From data to decisions: Distributionally robust optimization is optimal. *Management Science*.

Zhang, Y. and J. Gao (2017). Assessing the performance of deep learning algorithms for newsvendor problem. In *International Conference on Neural Information Processing*, pp. 912–921. Springer.

Appendices

Appendix A Backprobagation Algorithm

In order to calculate the derivatives based on backpropagation algorithm for the newsvendor network, we consider the network with L layers including input and output layers, so it has $L-2$ hidden layers in each of which there are $M^{(l)}$ hidden nodes for $l=2,\dots,L-1$. $w_{ij}^{(l,\,l+1)}$ denotes the weight between node i in layer l and node j in layer $l+1$. Let $net_i^{(l)}$ and $o_i^{(l)}$ define the input and output of each node i in layer l. It is seen that $o_i^{(1)} = s\left(net_i^{(1)}\right)$, and in output unit the activation function is a simple addition function so simply the output of this node is identical to the transmitted value to it and they also represent the order quantity $Q = o^{(L)} = net^{(L)}$.

If we consider a sample being fed into the network, we could drive the calculations for it, and then extend them for all samples in a matrix form. At the end of feed-forward step, once we get the output from the network in output node, the corresponding newsvendor cost will be computed by newsvendor cost function.

Backpropagation step begins right after feed-forward step. In this step we look for partial derivatives $\frac{\partial NVC}{\partial w_{ij}^{(l,l+1)}}$ for all i, j, and $l=1,\dots,L-1$. So, we need to compute the derivatives of activation functions in each node and find the path to a certain edge. This takes us to the intended partial derivative. The derivative of the newsvendor cost function NVC with respect to the output Q or equivalently $o^{(L)}$ is

$$\frac{\partial NVC}{\partial o^{(L)}} = \begin{cases} h & if\ D < Q \\ b & if\ Q < D \end{cases}. \tag{A.37}$$

As the activation function in the output unit is a simple additional function, the derivative of it with respect to its input value, $\frac{\partial o^{(L)}}{\partial net^{(L)}}$ is just equal to 1. In hidden layer units, sigmoid function has a property that its derivative results in an expression that consists only the function itself. So if the output of hidden node j in layer l is $o_j^{(l)}$, the derivative would be $o_j^{(l)}(1-o_j^{(l)})$. Now for simplicity of chain rule in taking the derivatives we define a term referred as to backpropagation error for node $j=1,\dots,M^{(l)}$ in layer $l=2,\dots,L$. We are moving from the right of the network to the left, so firstly we can compute partial derivatives of $\frac{\partial NVC}{\partial w_{j1}^{(L-1,L)}}$. To this end we define backpropagation error of output unit as

$$\delta^{(L)} = \frac{\partial NVC}{\partial o^{(L)}} \frac{\partial o^{(L)}}{\partial net^{(L)}}. \tag{A.38}$$

In this problem $\delta^{(L)}$ would be

$$\delta^{(L)} = \begin{cases} \delta^{(L)}(h) = h.1 & if\ D < Q \\ \delta^{(L)}(b) = b.1 & if\ Q < D \end{cases}. \tag{A.39}$$

Now, one more multiplication is left to get to the partial derivative $\frac{\partial NVC}{\partial w_{j1}^{(L-1,L)}}$ which is $\frac{\partial net^{(L)}}{\partial w_{j1}^{(L-1,L)}} = o_j^{(L-1)}$ Finally we have

$$\frac{\partial NVC}{\partial w_{j1}^{(L-1,L)}} = \delta^{(L)} o_j^{(L-1)}, \tag{A.40}$$

which is

$$\frac{\partial NVC}{\partial w_{j1}^{(L-1,L)}} = \begin{cases} \delta^{(L)}(h) o_j^{(L-1)} & if\ D < Q \\ \delta^{(L)}(b) o_j^{(L-1)} & if\ Q < D \end{cases}. \tag{A.41}$$

The remaining set of partial derivatives $\frac{\partial E}{\partial w_{ij}^{(l,l+1)}}$ for $l = 1, \ldots, L-2$ are obtained similarly by computing backprobagtion errors. In order to compute the backpropagation error $\delta_j^{(l)}$ of node j in the hidden layer l, all possible backward paths which reach to hidden node j should be considered so that an integration weighted of backpropagation errors of layer $l+1$, $\delta^{(l+1)}$, will transit to hidden node j

$$\delta_j^{(l)} = \sum_{k=1}^{M^{(l+1)}} \delta_k^{l+1} w_{jk}^{(l,l+1)}. \tag{A.42}$$

Similar to the previous set of derivatives, the term of $\frac{\partial o_j^{(l)}}{\partial net_j^{(l)}} = o_j^{(l)}(1 - o_j^{(l)})$ should be multiplied by $\delta_j^{(l)}$. Then backpropagation error for node j in the hidden layer l is calculated as

$$\delta_j^{(l)} = o_j^{(l)} \left(1 - o_j^{(l)}\right) \delta_j^{(l)}. \tag{A.43}$$

Again in this problem we have two types of backpropagation error

$$\delta_j^{(l)} = \begin{cases} \delta_j^{(1)}(h) & if\ D < Q \\ \delta_j^{(1)}(b) & if\ Q < D \end{cases}. \tag{A.44}$$

According to the chain rule, to obtain the partial derivative $\frac{\partial NVC}{\partial w_{ij}^{(1,l+1)}}$ the last term which is derivative of $\frac{\partial net_j^{(l+1)}}{\partial w_{ij}^{(l,l+1)}} = {o_j}^{(l)}$ is multiplied by $\delta_j^{(l+1)}$, hence;

$$\frac{\partial NVC}{\partial w_{ij}^{(l,l+1)}} = \begin{cases} \delta_j^{(l)}(h) {o_j}^{(l)} & if\ D < Q \\ \delta_j^{(l)}(b) {o_j}^{(l)} & if\ Q < D \end{cases}. \tag{A.45}$$

When all partial derivatives are computed the network weights would be updated in the negative gradient direction. Introducing a constant γ as a learning rate, the corrections for the weights will be

$$\Delta w_{ij}^{(1,l+1)} = \begin{cases} -\gamma \delta_j^{(l)}(h) {o_j}^{(l)} & if\ D < Q \\ -\gamma \delta_j^{(l)}(b) {o_j}^{(l)} & if\ Q < D \end{cases}. \tag{A.46}$$

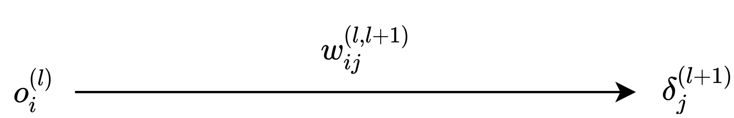

Figure A.10 Input and backpropagated error on an edge

There are more than one sample which are fed into the network. So if we have T pairs of samples like $\{(\boldsymbol{F}_1, D_1), (\boldsymbol{F}_2, D_2), \dots, (\boldsymbol{F}_T, D_T)\}$, the weight corrections are computed for each sample and we get, for example, the corrections as $\Delta_1 w_{ij}^{(l,l+1)}, \Delta_2 w_{ij}^{(l,l+1)}, \dots, \Delta_T w_{ij}^{(l,l+1)}$ for weight $w_{ij}^{(l,l+1)}$. Therefore the total amount of correction in the gradient direction is

$$\Delta w_{ij}^{(l,l+1)} = \sum_{t=1}^{T} \Delta_t w_{ij}^{(l,l+1)}. \tag{A.47}$$

In the emission network, the corresponding partial derivatives and the chain rule is similar to the newsvendor network. The emissions whose partial derivatives with respect to network weights are calculated is $\frac{\partial p(D)}{\partial \mathbf{w}^E}$, where $p(D)$ is a probability density function of normal distribution at each time.

$$p(D) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(D-\mu)^2}{2\sigma^2}} \tag{A.48}$$

First we need to compute the derivative of the above function with respect to D as the output of the emission network, $o_i^{(L)}$, where $i \in S$, then

$$\frac{\partial p(D)}{\partial o_i^{(L)}} = -\frac{o_i^{(L)}}{\sigma^2} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(o_i^{(L)}-\mu)^2}{2\sigma^2}} \right). \tag{A.49}$$

This value is multiplied by previous derivatives of the HMM network reached to the emission network and completes the initial derivatives that are used for updating the emission network weights. First, we have

$$\frac{\partial \Lambda}{\partial p(D)} \frac{\partial p(D)}{\partial o_i^{(L)}}. \tag{A.50}$$

We sum up the above derivatives over all states to obtain a single value for updating the network weights as they are identical in each state, we show the result as

$$\frac{\partial \Lambda}{\partial o^{(L)}}. \tag{A.51}$$

The backpropagation error term for the last layer is

$$\delta^{(L)} = \frac{\partial \Lambda}{\partial o^{(L)}} \frac{\partial o^{(L)}}{\partial net^{(L)}}. \tag{A.52}$$

Now, one more multiplication is left to get to the partial derivative $\frac{\partial \Lambda}{\partial w_{j1}^{(L-1,L)}}$ which is $\frac{\partial net^{(L)}}{\partial w_{j1}^{(L-1,L)}} = o_j^{(L-1)}$. Finally we have

$$\frac{\partial \Lambda}{\partial w_{j1}^{(L-1,L)}} = \delta^{(L)} o_j^{(L-1)}. \tag{A.53}$$

Derivatives of the function Λ to other network weights $w_{ij}^{(l,l+1)}$ between node i in layer l and j in layer $l+1$ for $l = 1, \dots, L-1$ are obtained similar to newsvendor network by chain rule and computing backpropagation errors for each node and layer.

Appendix B Further Analyses

B.1 Trade-off coefficient η analysis

Figure B.1 depicts the effect of the trade-off coefficient η on the final cost of the model for test sample. The curve on the plot is obtained by optimizing the function in (P-HMMNV) for different values of η. Accordingly, the network is trained and the cost for the out-of-sample is shown on Y-axis for the corresponding value of η. The optimal value for η is 0.64 (indicated by the red vertical line) by a grid search over the range $[0.001, 0.999]$. However, this value is not available prior to setting the order quantity. For this reason, the HMMNV method finds the best value for η equal to 0.70 by implementing the cross-validation process (indicated by the black vertical line). It is observed that the proposed algorithm and the cross-validation detect near-optimal value for η. One can also observe that the plot is unimodal, confirming that ignoring the best trade-off between hidden states and observable features as two sources of randomness will lead to an increase in the final objective value.

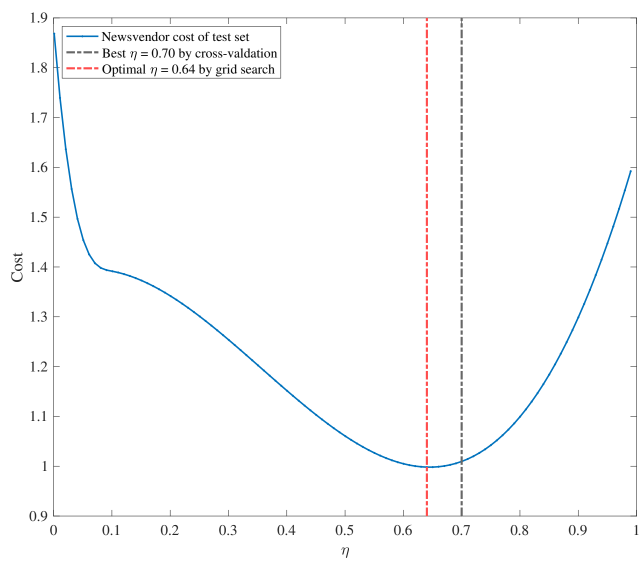

Figure B.1 **Effect of the trade-off coefficient η on the out-of-sample newsvendor cost.**

B.2 Extreme scenarios

We generate some samples to simulate the suggested extreme scenarios. In this data set, we assume that some demand outliers with low-frequency historical records belong to the state that is rarely visited by the Markov chain and the demand value in that scarce state is significantly different from the demand observations in the other state(s). We consider a two-state Markov chain with the parameter e that is the probability of transition from state 1 to state 2. The transition matrix can be set as

$$\begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{pmatrix} 1-e & e \\ 1-e & e \end{pmatrix} \end{array},$$

where the value of e is chosen so that $1-e$ is big enough to ensure that the system is in state 1 most of the time. We set the other parameters of the model as given in Table B.1. The small e values ensure that most of the demand observations come from state 1. The big difference between μ_1 and μ_2 ($\mu_2 - \mu_1 = 4$) causes the demand observations from state 2 to have extreme values and act as outliers with low frequency in the demand sequence.

Table B.1 The set of parameters used in the experimental setup for extreme scenarios with demand outliers.

Parameter	Description	Domain
e	transition probability	0.01 0.05
μ_1	mean of the base demand in state 1	1
μ_2	mean of the base demand in state 2	5
σ_1	standard deviation of the base demand in state 1	0.5
σ_2	standard deviation of the base demand in state 2	0.5
σ_ϵ	standard deviation of the demand by features	0.5
N	number of features	1 3
W	network weights	1
L	number of hidden layers	1
b/h	newsvendor cost parameters ratio	2 5 10
T	number of observations	100 200

Figure B.2 and Table B.2 give the results for these experiments. The results indicate that both the HMMNV and the DNN method are able to perform well given the presence of states that are rarely visited.

Table B.2 p-values of the t-test between the cost of algorithms in extreme scenarios.

	EDD	ObBW	PfLR	LML	DNN	HMMNV
EDD		0.2685				
ObBW						
PfLR	0.1940	0.0150				
LML	0.0265	6.4659e-04	0.3785			
DNN	0.0192	4.2083e-04	0.3115	0.8878		
HMMNV	0.0091	9.9872e-05	0.2308	0.7778	0.8952	

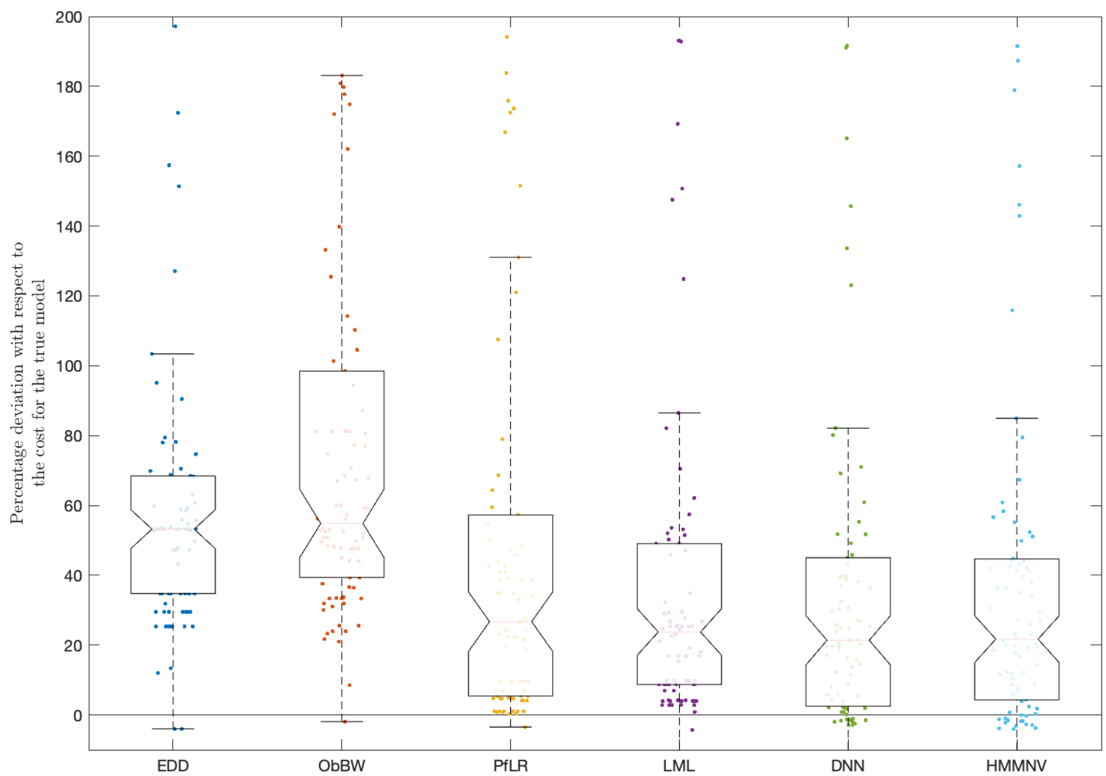

Figure B.2 **The result of all algorithms for extreme scenarios.**

B.3 Real experiment with macroeconomic features

We have used monthly data related to suggested extra features including lagged price of crude oil, lagged unemployment, lagged SP 500 stock index, and lagged personal disposable income for assessing the effect of using extra features. This data covers the period from September 1986 to August 2020 and it consists of 408 monthly sample observations. Accordingly, we resample the demand for crude oil on a monthly basis. We pick a time window with the length of 120 months and feed the first 96 months (80 % of the total periods in a window) to the algorithms and then test the trained models on the following 24 months. We roll this window by 24-month forward steps and obtain 13 test periods in total. Table B.3 gives the average cost of the monthly ordering policy for crude oil. We observe that the proposed HMMNV method performs well compared to other benchmark methods, however, as expected, this advantage is not present in all cases (e.g. when $b/h = 10$, DNN method has an average cost lower than HMMNV method). This is due to the informative input features that capture almost all the uncertainties in demand observations and have more out of sample forecasting power. Moreover, it should be noted that both HMMNV and DNN methods perform well compared to other methods since they consider nonlinear relations in modeling as well.

Table B.3 Average cost of the crude oil ordering for the newsvendor problem for monthly data by each algorithm using macroeconomic features (values are divided by 10^5).

b/h	EDD	ObBW	PfLR	LML	DNN	HMMNV
2	4.791	4.589	5.321	7.875	5.518	4.350
5	6.479	7.511	7.244	10.534	6.580	5.892
10	7.594	11.23	8.604	9.057	6.916	8.009
20	8.668	17.054	10.430	7.329	8.232	7.372