Causal Demand Forecasting

- Demand model: $y_t = a + b \cdot x_t + \varepsilon_t$
- Function of one (or more) factors (e.g., price, interest rate, weather)
- Regression → Least squares estimation

$$\hat{b} = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \qquad \hat{a} = \bar{y} - \hat{b}\bar{x} \qquad \bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \qquad \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$$

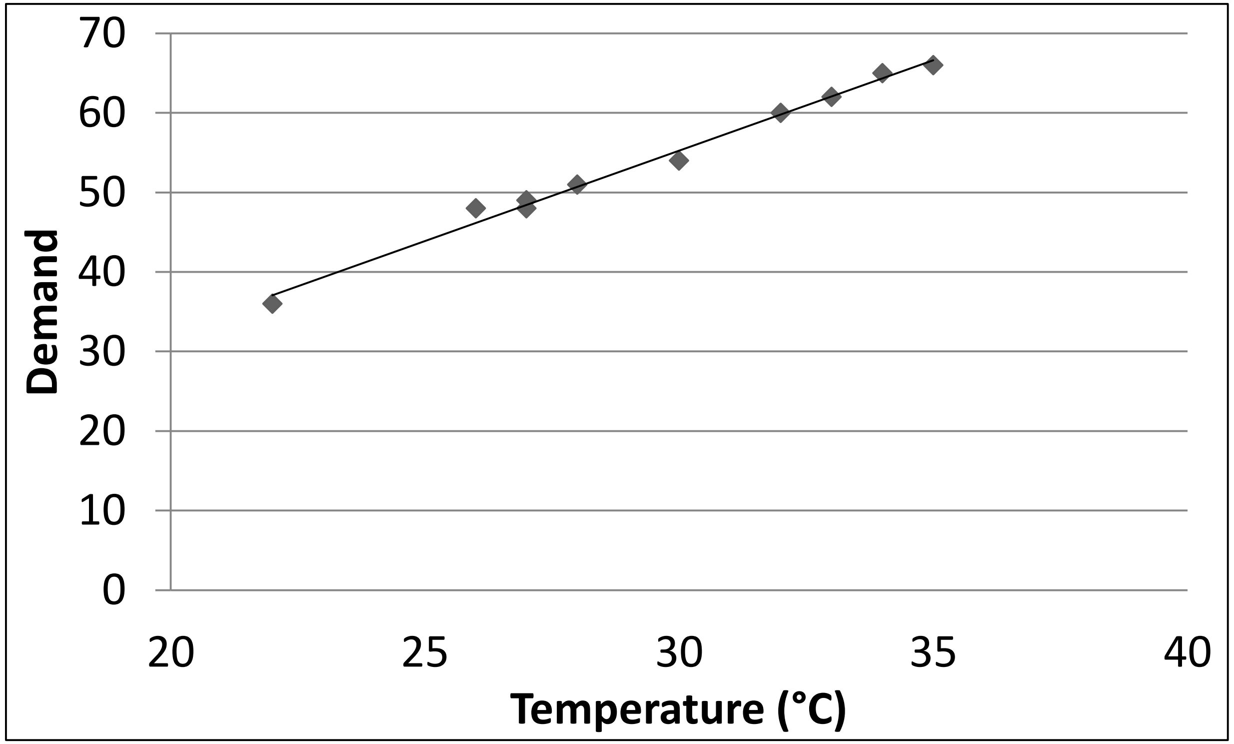
Forecast for period i:

$$p_{t+i} = \hat{a} + \hat{b} \cdot x_{t+i}$$

- Goodness of fit: $R^2 = \frac{\sum_{t=1}^{n}(p_t - \bar{y})^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$ with $0 \le R^2 \le 1$

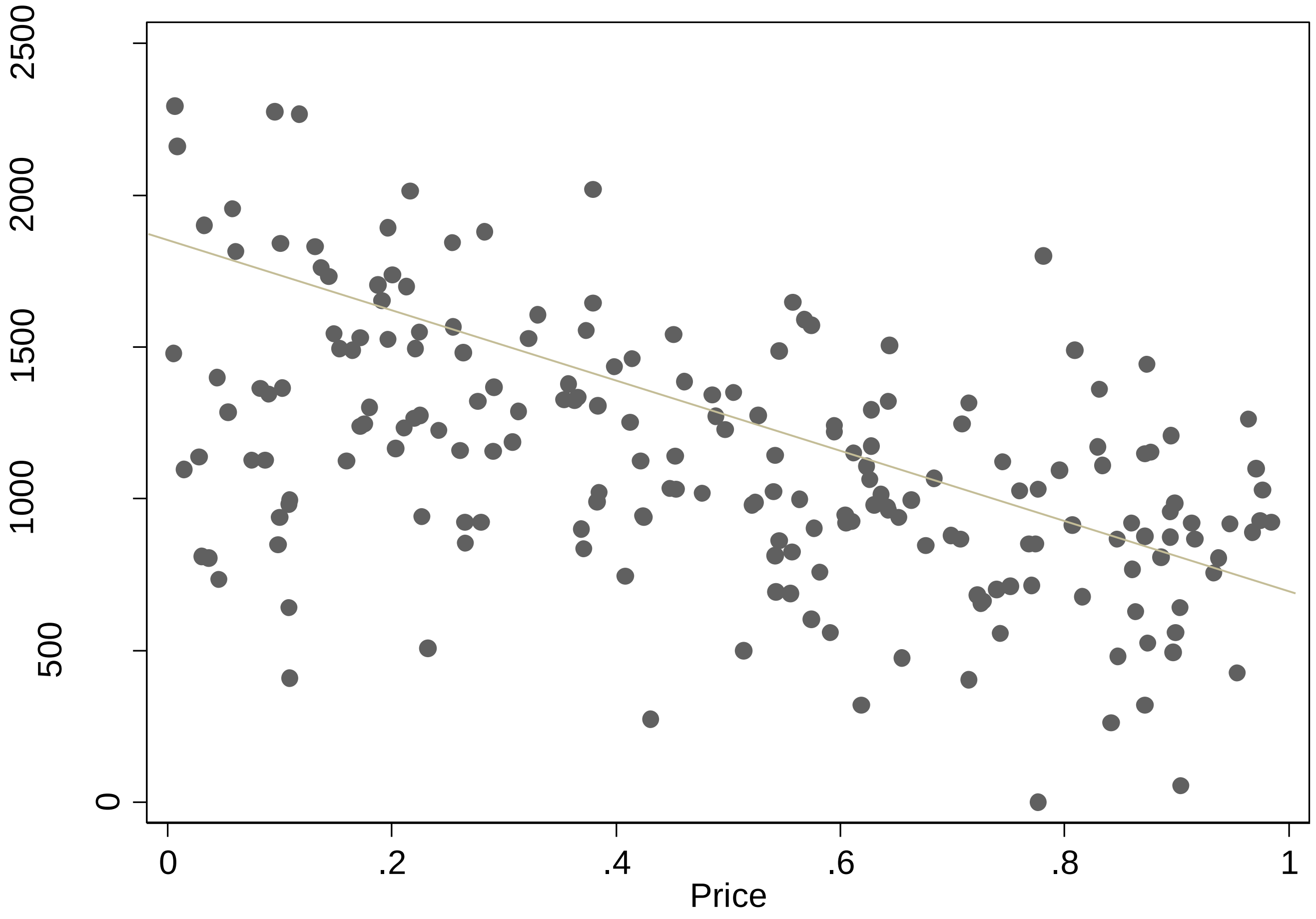
Example

t	y	x (Temp.)	p
1	48	27	48.44
2	36	22	37.08
3	49	27	48.44
4	65	34	64.36
5	54	30	55.26
6	60	32	59.81
7	48	26	46.17
8	51	28	50.72
9	62	33	62.08
10	66	35	66.63



$a=-12.84$, $b=2.27$, $R^2=0.99051$

Demand Observations & Inventory Function



Causal Demand Forecasting

Problem

- Inaccurate forecasts
- Time series information often not sufficient

Question

- Which additional information can improve forecast accuracy?

Approach

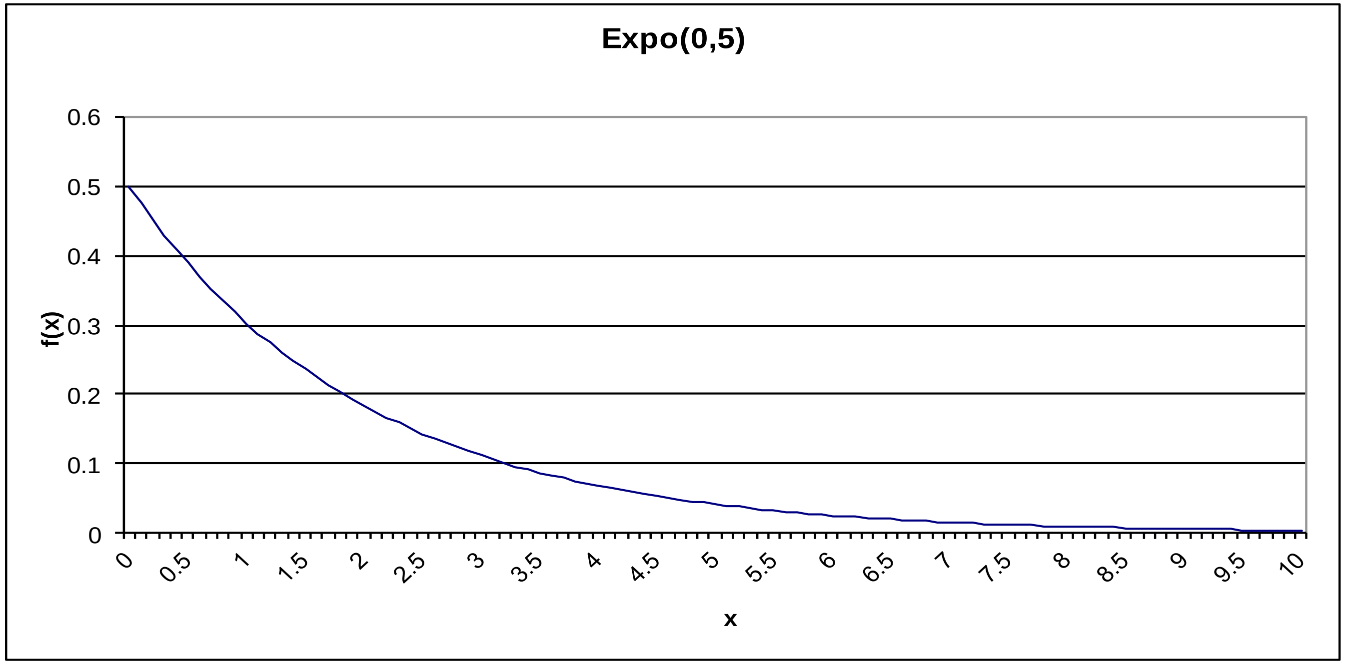
- Comparison
  - Time-series methods
  - Causal
    - Installed base
    - Time-to-failure
    - Usage profiles

Fill-Rate	Exp. 1. Ord.	Exp. 2. Ord.	Naïve	Trend	Exp. adaptive	Causal
99 %	17.042	16.627	17.042	13.850	13.634	9.733
95 %	11.995	11.703	11.995	9.748	9.596	6.851

- Causal methods improve forecast accuracy and reduce required inventories

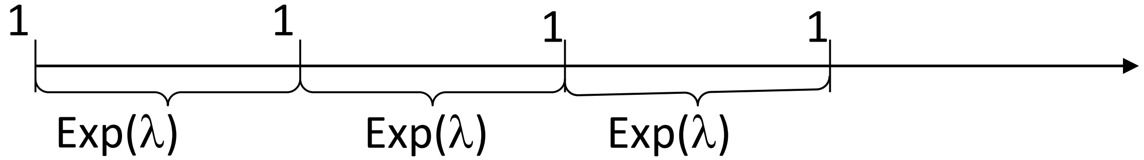
Exponential distribution

- Density $$f(x) = \lambda \cdot e^{-\lambda x} \quad x \geq 0$$
- Distribution $$F(x) = 1 - e^{-\lambda x} \quad x \geq 0$$
- E(X)=1/λ, Var(X)=(1/λ)^2



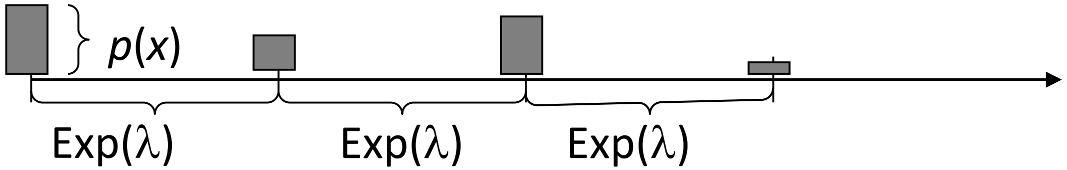
Stochastic processes

- Poisson-Process
  - Unit demands
  - Exponentially distributed interarrival times with rate λ (number of customers per period)



  - Number of demands in an interval of length *L*
    - Poisson distributed with parameter λL

- Compound-Poisson-Process
  - Exponentially distributed customer arrivals with rate λ
  - Discrete demand size distribution $p(x)$, x=1,2,...



Negative binomial distribution

- Probability density

$$P(D=d)=\binom{\alpha+d-1}{\alpha-1}p^{\alpha}(1-p)^{d}$$

- Moment fitting
  - Mean: $\alpha(1-p)/p$, variance: $\alpha(1-p)/p^2$

$$\alpha=\frac{\mu^2}{\sigma^2-\mu},\quad p=\frac{\mu}{\sigma^2}$$

Serially correlated demand

- ARMA(p,q)-process: Autoregressive moving average process

$$D_t = \underbrace{\boxed{\phi_1 D_{t-1} + \phi_2 D_{t-2} + \ldots + \phi_p D_{t-p}}}_{\text{Autoregressive terms}} + \varepsilon_t \underbrace{\boxed{- \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}}}_{\text{Moving average terms}}$$

- Special cases
  - First-order autoregressive process AR(1) $D_t = \phi_1 D_{t-1} + \epsilon_t$
  - First order moving average process MA(1) $D_t = \theta_1 \epsilon_{t-1} + \epsilon_t$
- Overview
  - Identification of demand process: correlation diagrams
  - Inventory management under correlated demand

Machine Learing for Demand Prediction

Machine Learning (ML):

•“Learn“ patterns from historical data (e.g., causal relationships, auto-regression, trends, seasonality)

• Objective: Improvement of out-of-sample generalization

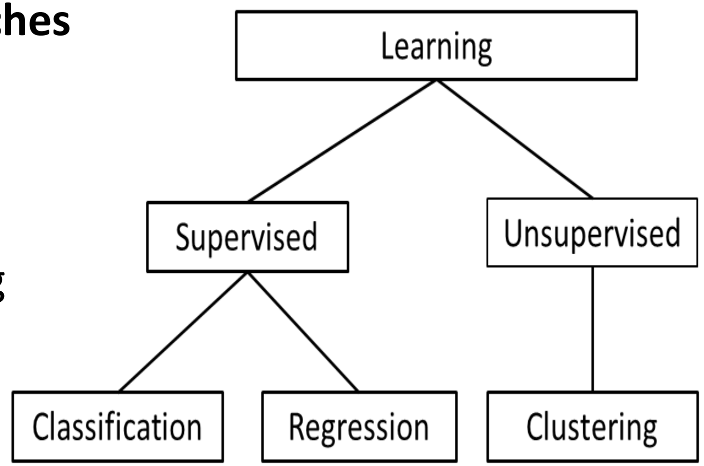
Forecasting Approaches

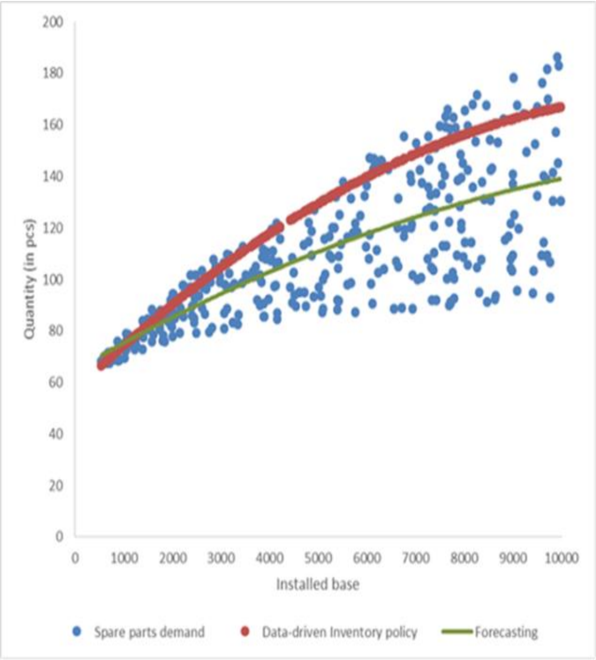
•Time Series

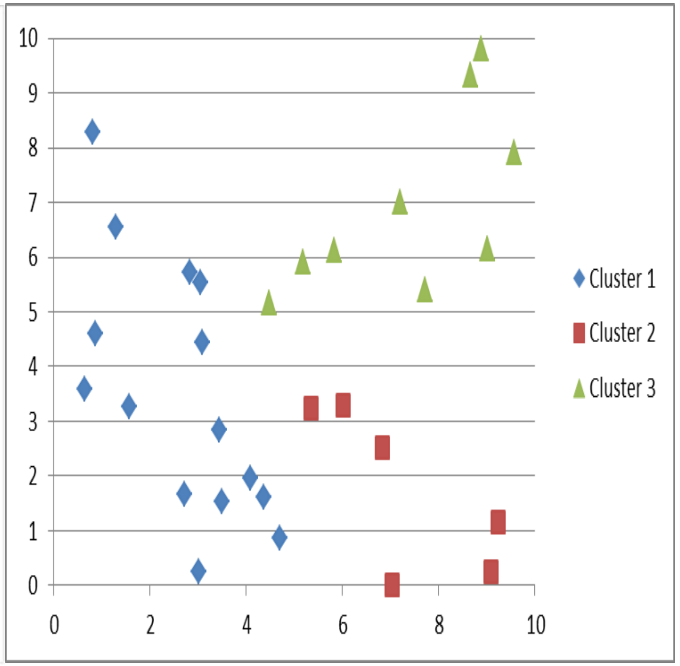
•Linear Regression

•Feature Engineering

•Random Forest







Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, *13*(3).

Machine Learing for Demand Prediction

Example: Forecast the demand at the upstream from the distorted demand signals

Demand Data
1) Simulated Data
2) Received Data from Industry

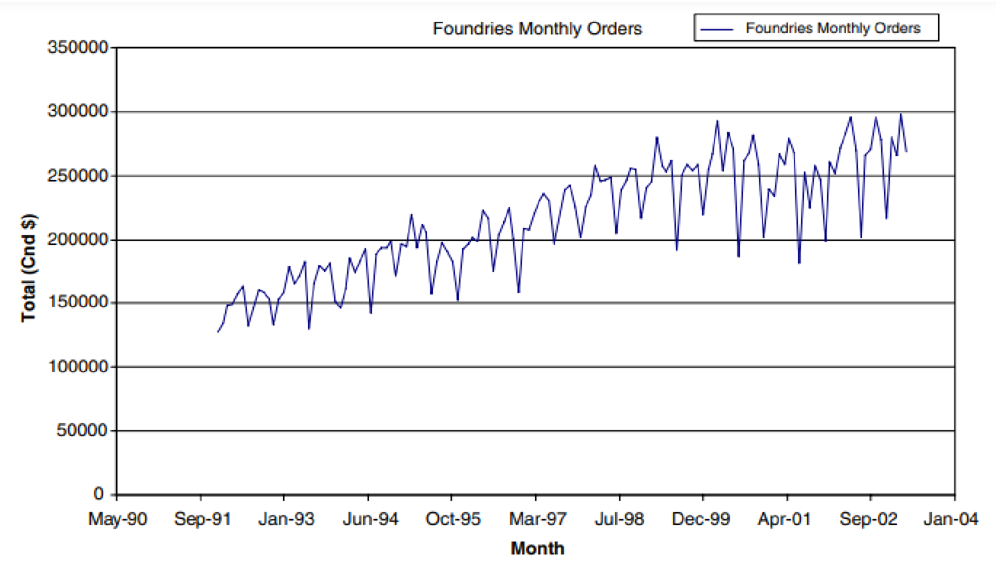


Fig. 7. Foundries monthly orders.

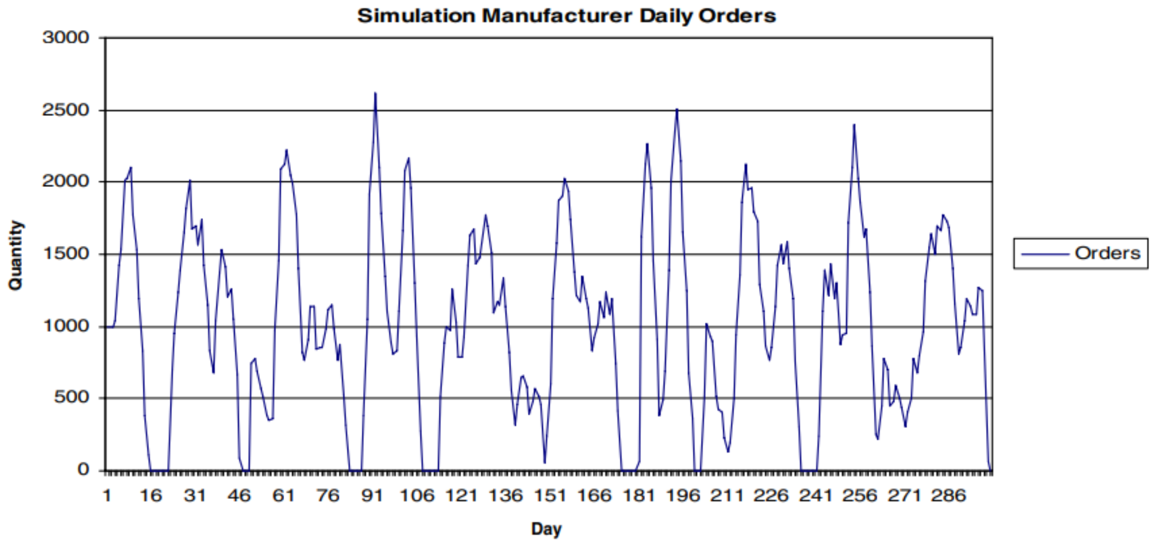


Fig. 8. Data obtained from simulation.

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140-1154.

Machine Learing for Demand Prediction

Example cont.:
Neural networks (NN)

Simulations data set

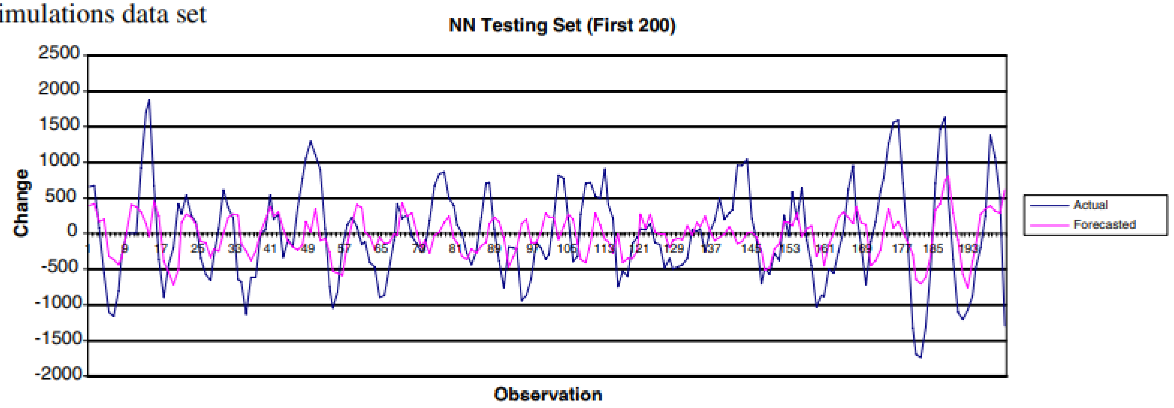


Fig. 9. Simulations testing data set results.

Foundries data set

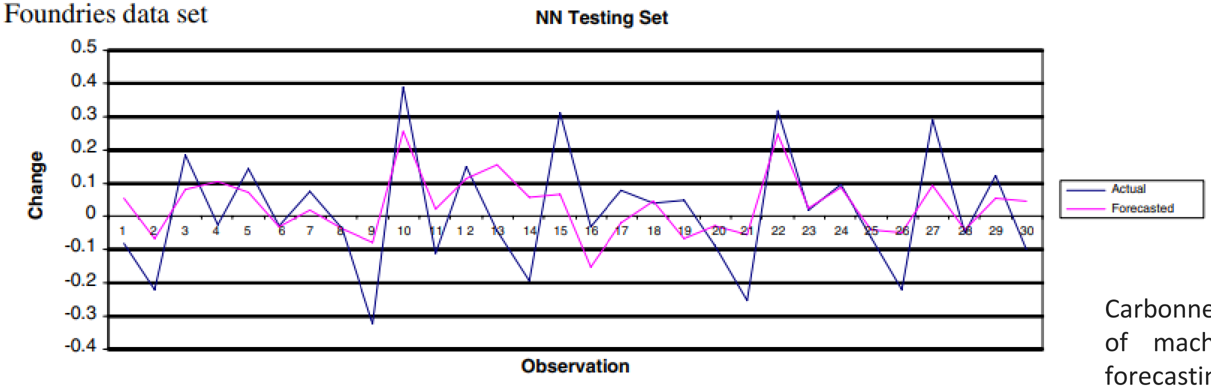


Fig. 10. Foundries testing data set results.

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140-1154.

Machine Learing for Demand Prediction

Example cont.:
Recurrent neural networks (RNN)

Simulations data set

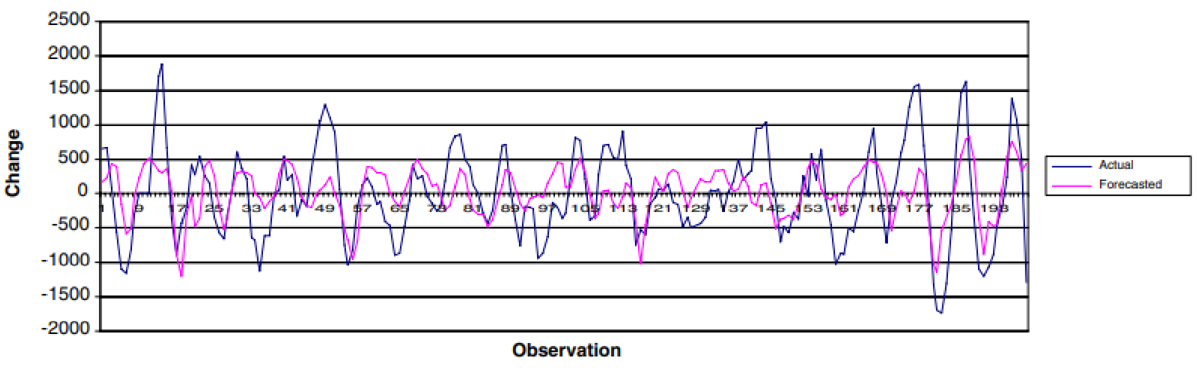


Fig. 11. Simulations testing data set results.

Foundries data set

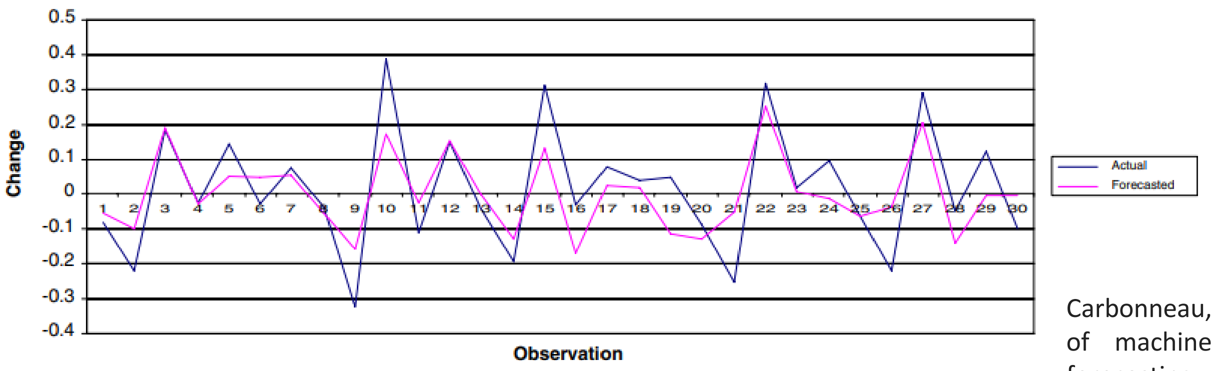


Fig. 12. Foundries testing data set results.

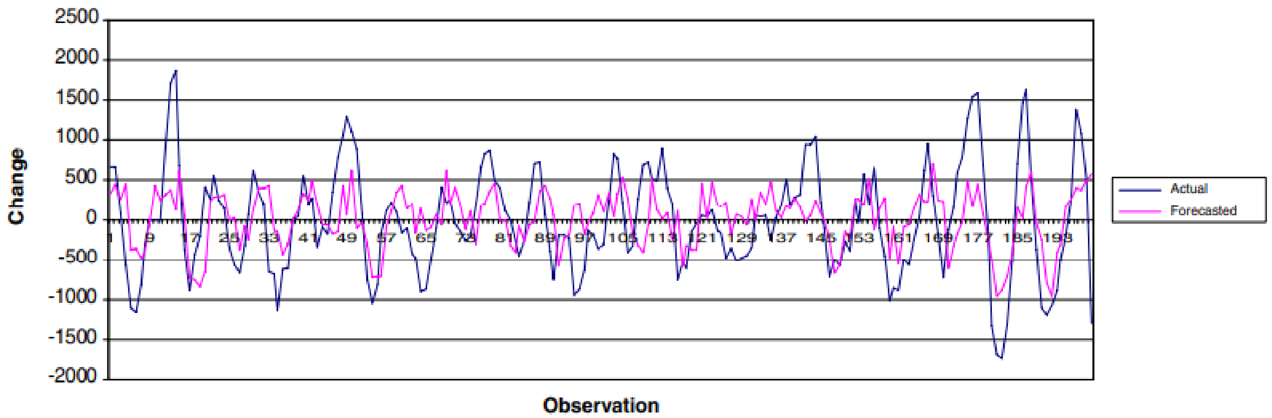
Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140-1154.

Machine Learing for Demand Prediction

Example cont.:

Support vector machines (SVM)

Simulations data set



Foundries data set

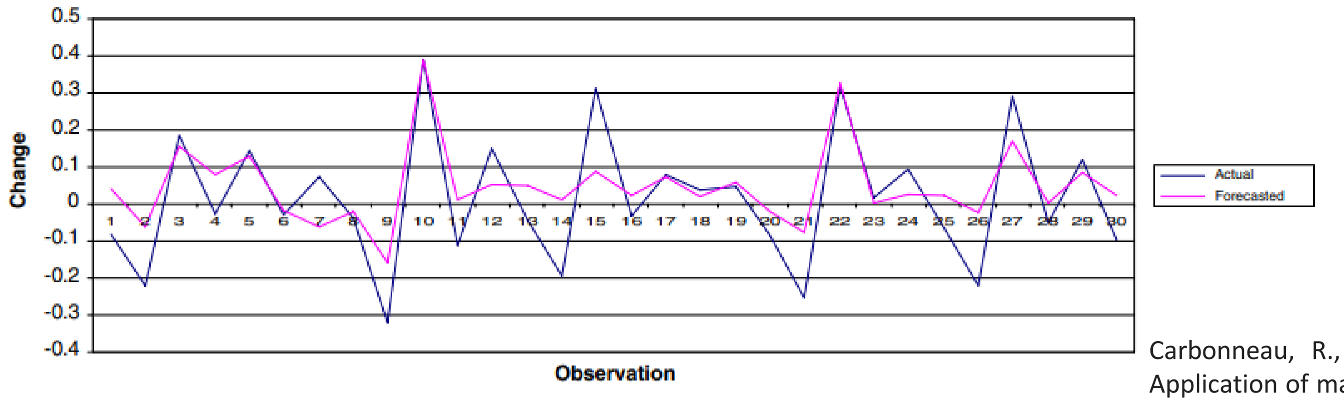


Fig. 14. Foundries testing data set results.

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140-1154.

Machine Learing for Demand Prediction

Example cont.:

Machine Learning Techniques

- Recurrent neural networks (RNN)
- Support vector machines (SVM)
- Neural networks (NN)

Traditional Approaches

- Multiple Linear Regression (MLR) (= Auto Regressive Model)
- Naïve Forecast
- Moving Average
- Trend-based Forecast (= Simple Regression)

Table 3
Comparison of the performance (MAE) of forecasting techniques on the simulation data set

Forecasting technique	Testing set		Training set	
	MAE	Std. dev.	MAE	Std. dev.
RNN	447.72	328.23	461.66	350.35
LS-SVM	453.04	341.88	449.32	365.01
MLR	453.22	343.65	464.62	375.62
NN	455.41	354.40	471.03	383.29
Naïve	520.53	407.29	536.45	435.05
Moving average	526.61	370.35	558.82	400.11
Trend	618.02	487.42	674.68	490.50

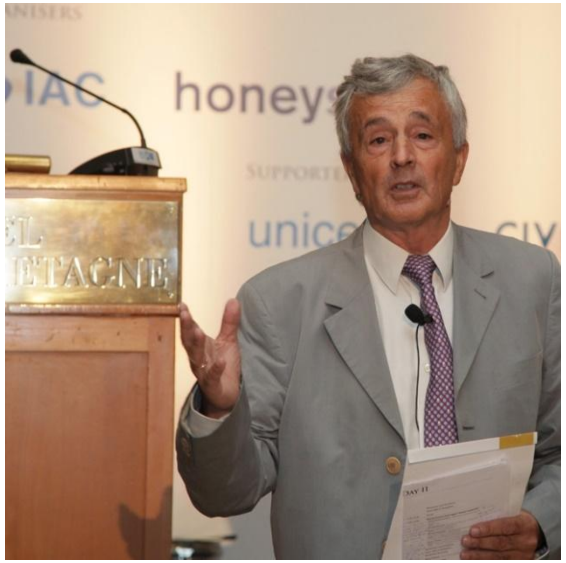
Table 4
Comparison of the performance of forecasting techniques on the foundries data set

Forecasting technique	Testing set		Training set	
	MAE	Std. dev.	MAE	Std. dev.
RNN	20.352	16.203	15.521	12.334
LS-SVM	20.485	17.304	3.665	3.722
MLR	21.396	19.705	15.007	15.041
NN	25.260	19.733	12.855	12.057
Moving average	25.481	19.253	18.205	13.028
Trend	27.323	24.198	17.995	17.292
Naïve	32.591	23.485	20.263	17.380

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140-1154.

The M-competitions

Spyros Makridakis



Four competitions between various forecasting methods, each competition with a different focus point

M1 (1982):

- Statistically sophisticated or complex methods do not necessarily provide more accurate forecasts than simpler ones.
- The relative ranking of the performance of the various methods varies according to the accuracy measure being used.
- The accuracy when various methods are combined outperforms, on average, the individual methods.
- The length of the forecasting horizon involved is important.

The M-competitions

M4 (2018):

- Out of the 17 most accurate methods, 12 were “combinations” of mostly statistical approaches.
- The biggest surprise was a “hybrid” approach utilizing both Statistical and ML features.
- The second most accurate method was a combination of seven statistical methods and one ML one, with the weights for the averaging being calculated by a ML algorithm
- The first and the second most accurate methods also achieved an amazing success in specifying correctly the prediction intervals.
- The six pure ML methods submitted in the M4 performed poorly

Data-Driven Newsvendor

- **Limitations of the standard SAA formulation:**
  - Feature data (causal factors, covariates) not considered (e.g., temperature, weekday)
  - Historical observations equally weighted

- **Machine Learning (ML):**
  - “Learn“ patterns from historical data (e.g., causal relationships, auto-regression, trends, seasonality)
  - Objective: Improvement of out-of-sample generalization

- **Potential result of ML for newsvendor problem:**
  - Higher (lower) weight on more recent (less recent) data
  - Order decision as a function of features

Sequential forecasting and inventory control

- Sequential approach: first forecast demand/estimate parameters, then optimize inventories
- In the newsvendor model: how to find the demand distribution F? Then just take the desired quantile
- Normal distribution? Demand forecast + forecast error? Other distribution?
- How precise are parameter estimates?

General demand modelling in the big data era

- $D_i = d(\varphi, x_i) + error_i$
- d is the function that maps the available data to an estimate of demand
- Function can be linear (e.g. linear regression), non-linear, a decision tree, neural net…
- φ : parameters to be optimized, x_i available explanatory data (“features’’)
- But beware: not only the point forecast, also the accuracy is needed!

The Data Driven Newsvendor

- **Sample of demand scenarios**
  - Drawn from the underlying demand distribution
    - Uniformly (0,1)-distributed random number x
    - NORMDISTINV(x,mu,sigma,1)
  - d_i, i=1,…,n are realizations of demand random variable D
- **Overage and underage costs**
- **LP-formulation**
  - Second-stage problem (recourse)
    - Determine underage or overage cost for a given demand
  - First-stage problem
    - Determine initial capacity level before demand is known

Model

- Problem $\max_{y \geq 0} \Pi = -cy + E_D\left[p \cdot \min\{D, y\}\right]$

- Decision variables
  - First stage: initial capacity *y*
  - Second stage: overage and underage quantities
  - Model for g = 0

$$\max \Pi = -cy + \frac{1}{n}\sum_{i=1}^{n} p \cdot r_i$$

$$\begin{aligned} s.t. \quad & r_i \leq d_i \quad i = 1,\ldots,n \\ & r_i \leq y \quad i = 1,\ldots,n \\ & y, r_i \geq 0 \quad i = 1,\ldots,n \end{aligned}$$

SAA formulation limitation:

•Feature data (causal factors, covariates) not considered (e.g., temperature, weekday)

•Historical observations equally weighted

Example 1: Safety Stock Planning under Causal Demand Forecasting

$$\min C = \sum_{i=1}^{n} h y_i$$

$$\text{s.t. } y_i \geq \boxed{\sum_{j=0}^{m} \beta_j X_{ji}} - D_i \quad i = 1, \cdots, n$$

Order quantity s_i as a function of features

$$s_i \leq D_i \quad i = 1, \cdots, n$$

$$s_i \leq \sum_{j=0}^{m} \beta_j X_{ji} \quad i = 1, \cdots, n$$

Beutel, A. L., & Minner, S. (2012). Safety stock planning under causal demand forecasting. *International Journal of Production Economics*, *140*(2), 637-645.

Example 1: Cont.

$$D_i - \gamma_i M \leq \sum_{j=0}^{m} \beta_j X_{ji} \quad i = 1, \cdots, n$$

$$\sum_{i=1}^{n} \gamma_i \leq n(1 - P_1)$$

$$\sum_{i=1}^{n} s_i \geq P_2 \sum_{i=1}^{n} D_i$$

$$s_i, y_i \geq 0, \;\; \gamma_i \in \{0,1\}, \;\; \beta_j \in \Re \quad i = 1, \cdots, n \quad j = 1, \cdots, m$$

Beutel, A. L., & Minner, S. (2012). Safety stock planning under causal demand forecasting. *International Journal of Production Economics*, *140*(2), 637-645.

Example 2: Solving the Newsvendor Problem with Feature Data

1. Classical Newsvendor Problem

$$\min_{q\geq 0} \quad EC(q) := \mathbb{E}[C(q;D)],$$

$$C(q;D) := b(D-q)^{+} + h(q-D)^{+}$$

$$q^{*} = \inf\left\{y:\ F(y) \geq \frac{b}{b+h}\right\}.$$

2. Data-Driven Newsvendor Problem

In practice, the true distribution is unknown.

$$\min_{q\geq 0} \ \hat{R}(q;\mathbf{d}(n)) = \frac{1}{n}\sum_{i=1}^{n}[b(d_i-q)^{+} + h(q-d_i)^{+}], \quad \text{(SAA)}$$

$$\hat{q}_n = \inf\left\{y:\ \hat{F}_n(y) \geq \frac{b}{b+h}\right\},$$

SAA: Sample average expectation

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

Example 2: Cont.

3. Feature-Based Newsvendor Problem

In practice, the demand depends on many observable features such as location and weather.

$$\min_{q(\cdot)\in\mathcal{Q},\{q:\mathcal{X}\to\mathbb{R}\}} \mathbb{E}[C(q(\mathbf{x});D(\mathbf{x}))|\mathbf{x}],$$

Solution Approaches:

1. Empirical Risk Minimization (ERM) Algorithms
2. Kernel Optimization (KO) Method

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

Example 2: Cont.

3. Feature-Based Newsvendor Problem

1. Empirical Risk Minimization (ERM) Algorithms

$$\min_{q:q(\mathbf{x})=\sum_{j=1}^{p} q^j x^j} \hat{R}(q(\cdot); S_n)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left[b(d_i - q(\mathbf{x}_i))^+ + h(q(\mathbf{x}_i) - d_i)^+\right]$$
$$\equiv \min_{\mathbf{q}=[q^1,\ldots,q^p]} \frac{1}{n}\sum_{i=1}^{n}(bu_i + ho_i)$$
$$\text{s.t. } \forall\, i = 1, \ldots, n:$$
$$u_i \geq d_i - q^1 - \sum_{j=2}^{p} q^j x_i^j$$
$$o_i \geq q^1 + \sum_{j=2}^{p} q^j x_i^j - d_i$$
$$u_i, o_i \geq 0,$$

2. Kernel Optimization (KO) Method

$$\min_{q\geq 0} \quad \tilde{R}(q; S_n, \mathbf{x}_{n+1}) = \min_{q\geq 0} \frac{\sum_{i=1}^{n} K_w(\mathbf{x}_{n+1} - \mathbf{x}_i) C(q, d_i)}{\sum_{i=1}^{n} K_w(\mathbf{x}_{n+1} - \mathbf{x}_i)}.$$

$$\hat{q}_n^{\kappa} = \hat{q}_n^{\kappa}(\mathbf{x}_{n+1}) = \inf\left\{q : \frac{\sum_{i=1}^{n} \kappa_i \mathbb{I}(d_i \leq q)}{\sum_{i=1}^{n} \kappa_i} \geq \frac{b}{b+h}\right\},$$

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

Modern businesses

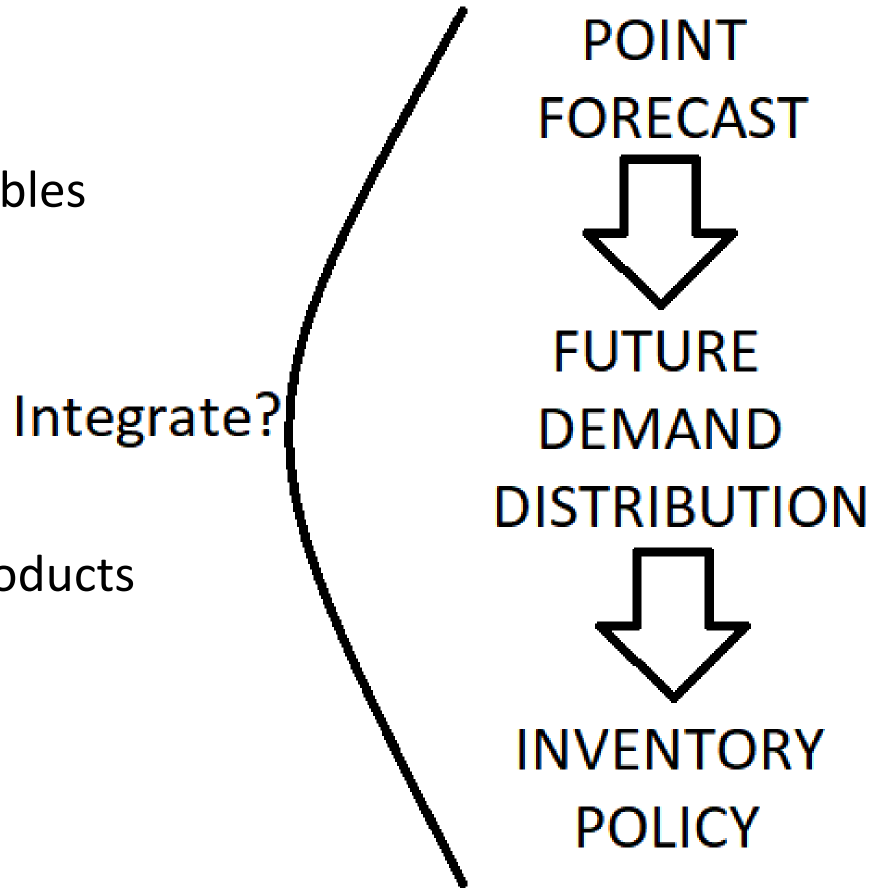
Abundance of data available:

- Possibilities for machine learning approaches
- Nonlinearities, complex dependencies, various explanatory variables

But also:

- New product introductions very frequently
- Short product life cycles
- Little data available – judgemental forecasting, finding similar products

- Pitfall: how to USE your forecast!
  - what is exactly needed for your inventory control model?
  - are your distributional assumptions valid?
  - how do you get an accurate measure of the forecast errors?



Classification

- **Stochastic OR/MS problems are characterized by 3 primitives:**
  - Data $\{y^1, \dots, y^N\}$ on stochastic parameters (e.g., demand, price, ...)
  - Feature data $\{x^1, \dots, x^N\}$ (e.g., temperature, economic indicators, ...)
  - Decision $z \in Z$ based on $X = x$ with objective $\min c(z, Y)$
- **Traditional stochastic optimization:**
  - $v^* = \min\limits_{z \in Z} \mathbb{E}[c(z; Y)]$, $z^* = \operatorname*{argmin}\limits_{z \in Z} \mathbb{E}[c(z; Y)]$
  - Estimation of a priori distribution ϕ_Y of Y
- **Predictive analytics (regression, machine learning, ...):**
  - Prediction $\mathrm{Y}=\mathrm{f}(\mathrm{X})$ based on $\{(\mathrm{x}^1,\mathrm{y}^1), \dots, (\mathrm{x}^N,\mathrm{y}^N)\}$
- **Data-Driven optimization (Prescriptive analytics):**
  - $v^* = \min\limits_{z \in Z} \mathbb{E}[c(z; Y) | \mathrm{X} = \mathrm{x}]$, $z^* = \operatorname*{argmin}\limits_{z \in Z} \mathbb{E}[c(z; Y) | \mathrm{X} = \mathrm{x}]$
  - No estimation of a-priori distribution, but use of data $\{(\mathrm{x}^1,\mathrm{y}^1), \dots, (\mathrm{x}^N,\mathrm{y}^N)\}$

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

Data-Driven Optimization + ML

- **Classical data-driven models optimize in-sample**
  - $v^* = \min_{z \in Z} \frac{1}{N} \sum_{i=1}^{N} c(z; \mathrm{y}^i),\ z^* = \operatorname{argmin}_{z \in Z} \frac{1}{N} \sum_{i=1}^{N} c(z; \mathrm{y}^i)$
  - see Data-driven newsvendor
- **ML: Consideration of out-of-sample generalization**
  - $v^* = \min_{z \in Z} \sum_{i=1}^{N} \mathrm{w}_{N,\mathrm{i}}(x) \cdot c(z; \mathrm{y}^i),\ z^* = \operatorname{argmin}_{z \in Z} \sum_{i=1}^{N} \mathrm{w}_{N,\mathrm{i}}(x) \cdot c(z; \mathrm{y}^i)$
  - $\mathrm{w}_{N,\mathrm{i}}(x)$: weight associated with y^i when observing feature $X = x$
- **$\boldsymbol{w_{N,i}(x)}$ is derived directly from the data:**
  - A-priori estimation using ML techniques: kNN regression, Kernel regression, ...
  - Integrated prediction-optimization framework