SIDE-EFFECTS OF IMPROVING FORECAST ACCURACY WITH MORE DATA

PROF. DR. ROBERT BOUTE
ARNOUD WELLENS, PH.D.


SAVE
SAVE
ORANGE DOT SALE
450
cheeses
SAVE
SAVE
PNP
GOLDEN DELICIOUS
R23 99
SAVE
ROYAL GALA
APPLES 1.5KG
R24 99
Fresh

Inaccurate forecasting is expensive

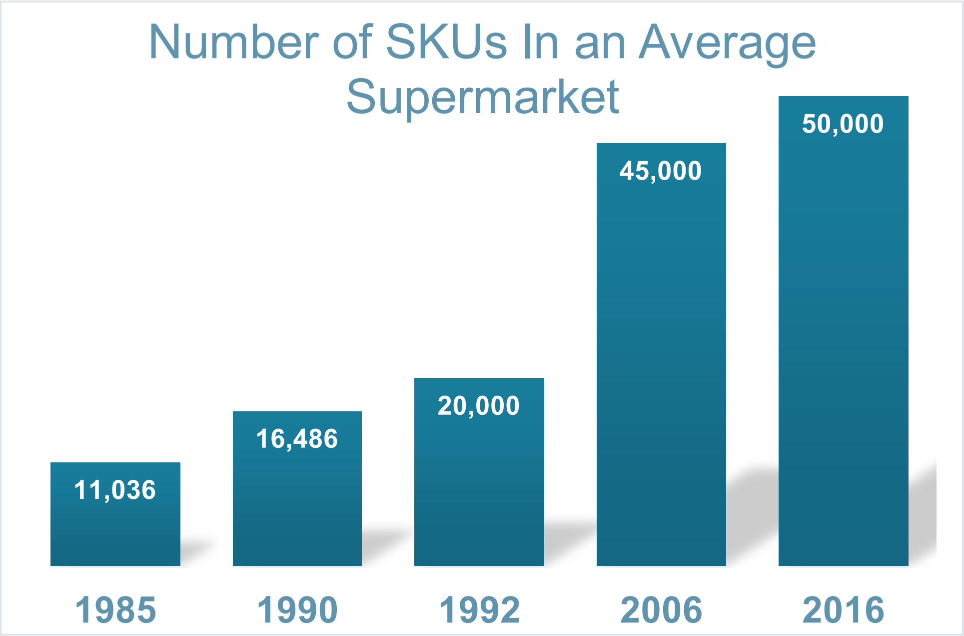
Retailers are losing $1.75 trillion over this

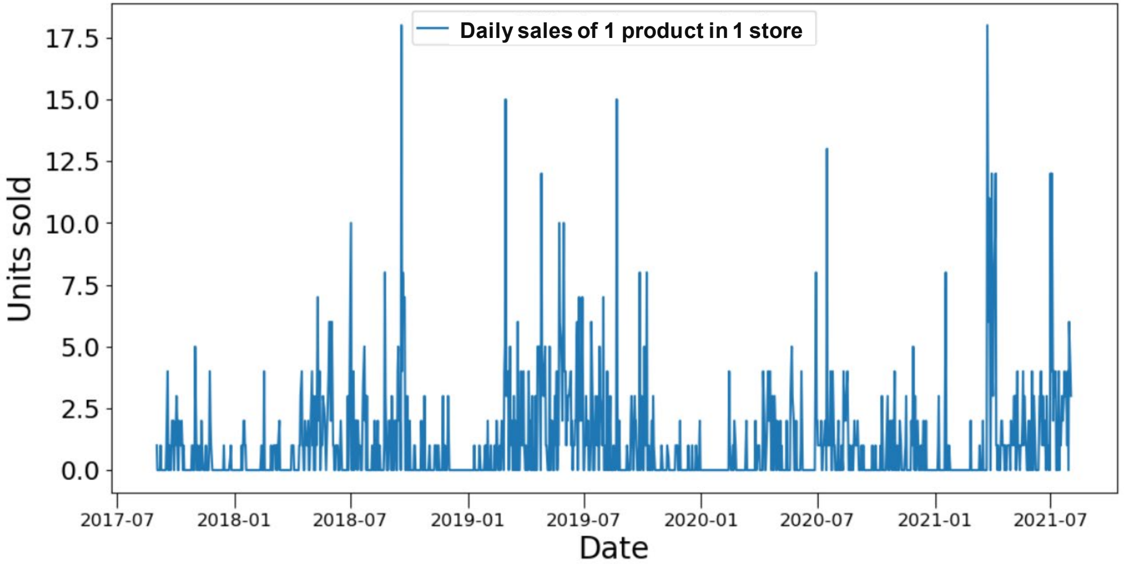
CNBC

- **Out-of-stocks** account for $634.1 billion in lost retail sales
- **Overstocks** contribute $471.9 billion in lost revenues

Retailers struggle to utilize the mounds of customer data they've acquired over the past few years and accurately forecast demand.

Especially the product-store level is challenging





Grocery stores carry 40,000 more items than they did in the 1990s

Machine learning to the rescue?
They outperform in the recent forecast competitions

Makridakis & Hibon (1979)

M-competition (1982)

M2 (1993)

M3 (2000)

NN3 (2006)

Walmart Store Sales (2014)

Rossmann Store Sales (2015)

Wikipedia (2017)

M4 (2018)

KDD Cup (2018)

M5 (2020)

Simple statistical methods

Rise of machine learning

Hyndman, R. J. (2020). A brief history of forecasting competitions.International Journal of Forecasting, 36(1):7–14.
Bojer, C. S. and Meldgaard, J. P. (2021). Kaggle forecasting competitions: An overlooked learning opportunity.International Journal of Forecasting, 37(2):587–603.

But the winning ML methods are hard to implement

- **Computational requirements** are high
  - Small product portfolio: 2,000 SKUs x 15 stores x 14-day horizon = **420,000 forecasts/ day**
  - Large portfolio: 60,000 SKUs x 15 stores x 14-day horizon = **12.6 million forecasts/day**

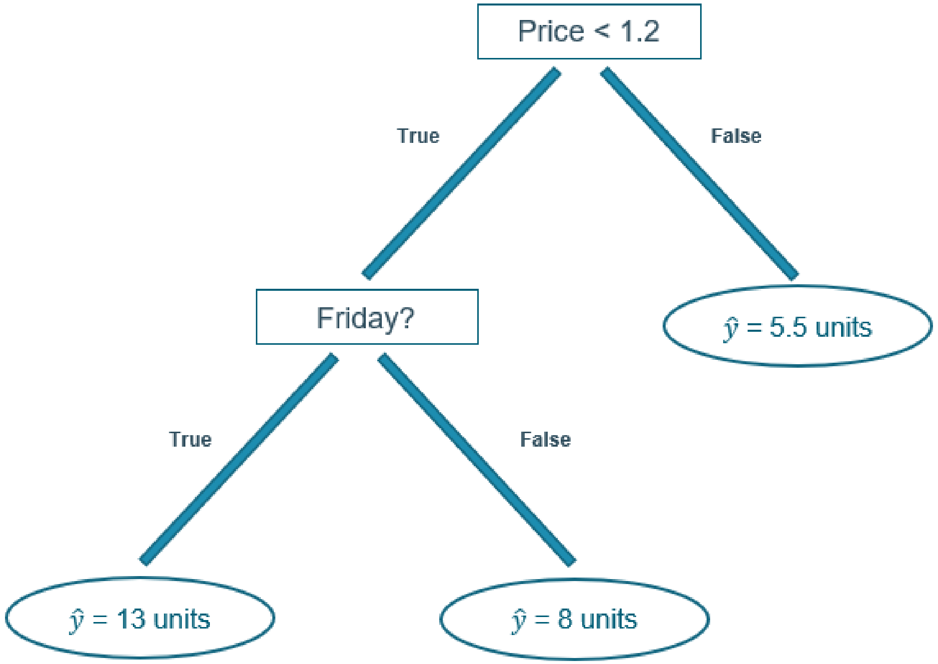
- **Model complexity** requires expert knowledge

- Are we ready to **trust a black box?**

- **Result: Most retailers still use a simple statistical method**

Simple.

We simplified the M5 winning LightGBM method



Decision-tree framework

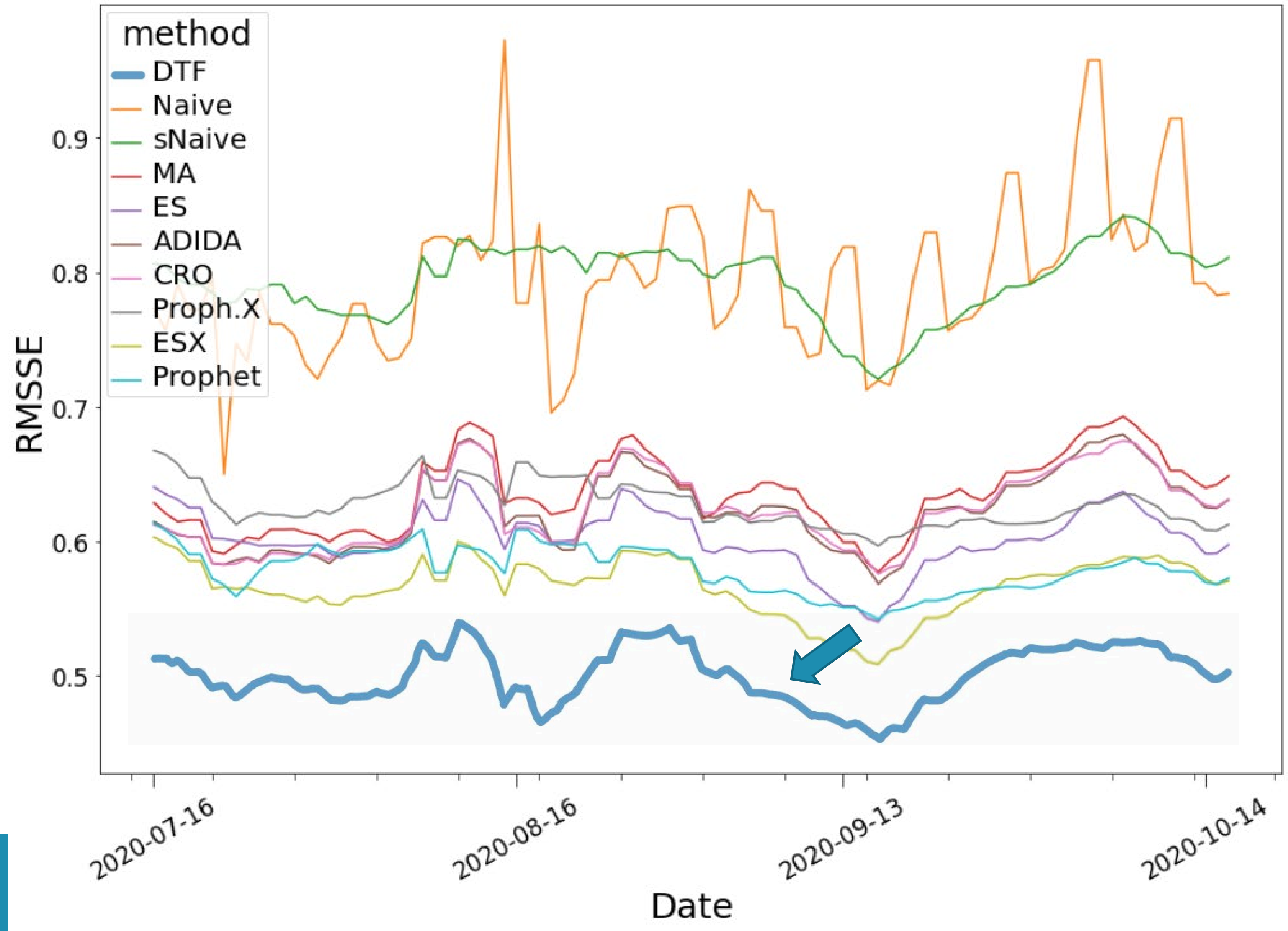
- **Train a single global tree-based method**
- **Use all available data ‘as is’**
- **Perform ‘basic’ feature engineering**
- **Automate the hyperparameter tuning**

Pilot in one store

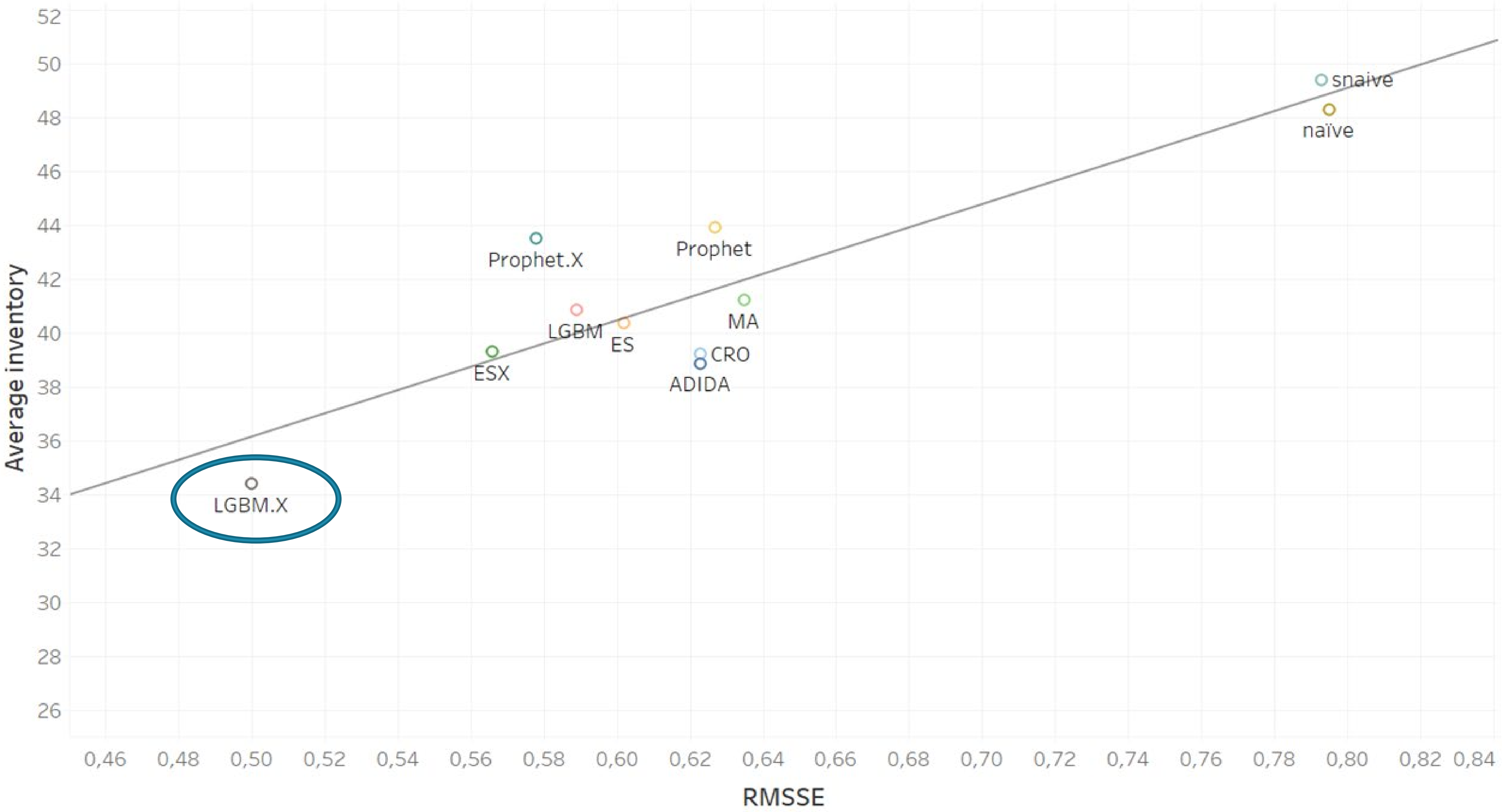
- **4,523 food products**
- Forecast the **daily sales**
- **104 different inputs**
- **Total data:** >500 million training data points

Our simplified decision-tree framework (DTF) outperforms the best benchmark by 11.48%

- (Seasonal) Naive
- Moving Average
- Exponential Smoothing
- Croston's method
- ADIDA (temporal aggregation)
- Prophet
- Exponential Smoothing with Explanatory variables
- Prophet with Explanatory variables



Superior forecast accuracy requires 12.5% less inventory to achieve a 95% target service level

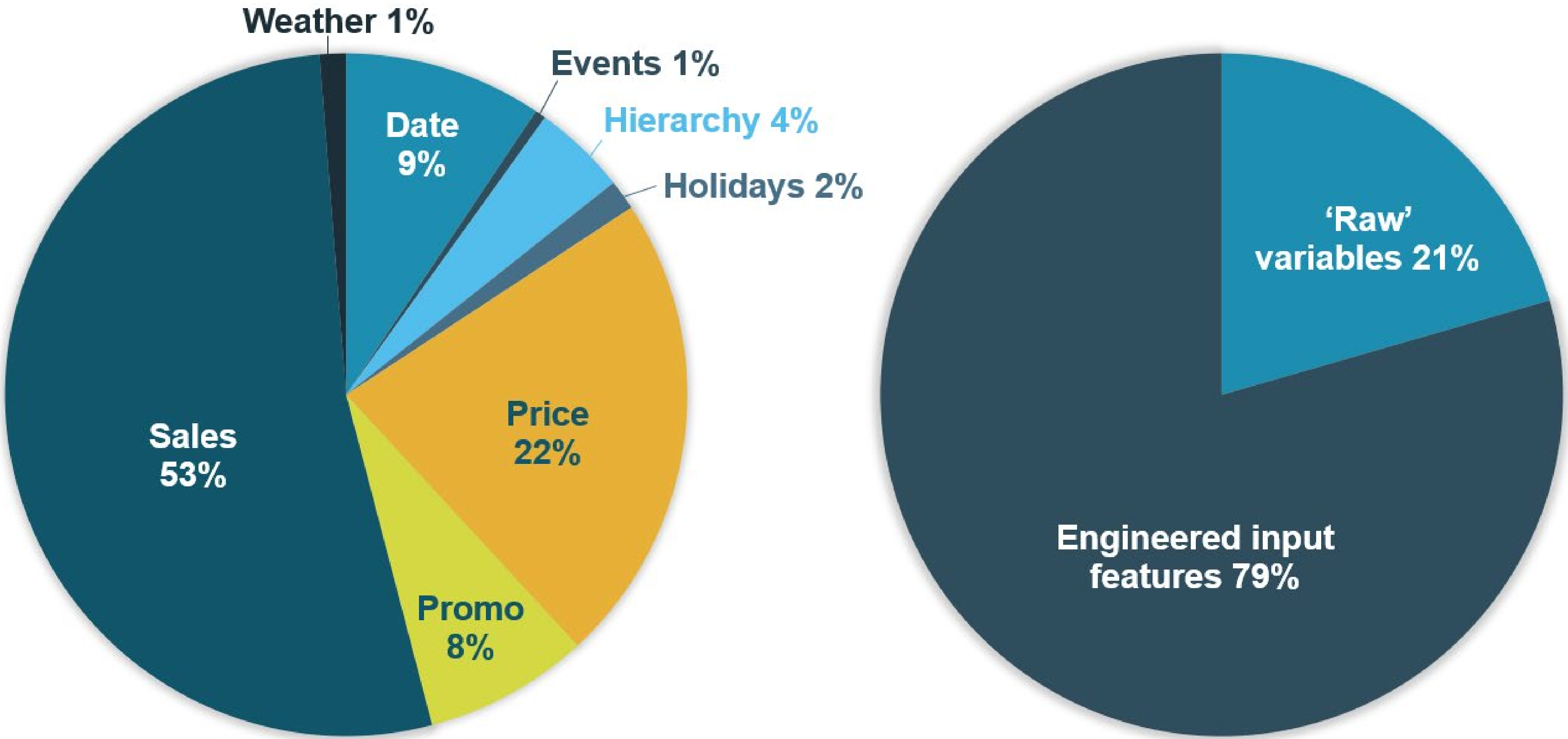


Only marginal gains with more sophisticated versions

Models	RMSSE	Improvement over ESX
Decision-tree framework (DTF)	0.501 ± 0.01	11.48 %
Recursive DTF (RDTF)	0.499 ± 0.01	11.84 %
Ensemble of DTF and RDTF	0.497 ± 0.01	12.19 %
DTF-l-1	0.503 ± 0.002	11.13%
DTF-l-2	0.514 ± 0.005	9.19%
DTF-sl-70	0.502 ± 0.003	11.31%
DTF-fs-sl-70	0.502 ± 0.002	11.31%
DTF-m5	0.491 ± 0.001	13.25%

Opening the Black Box.

Shapley values unveil the main contributors

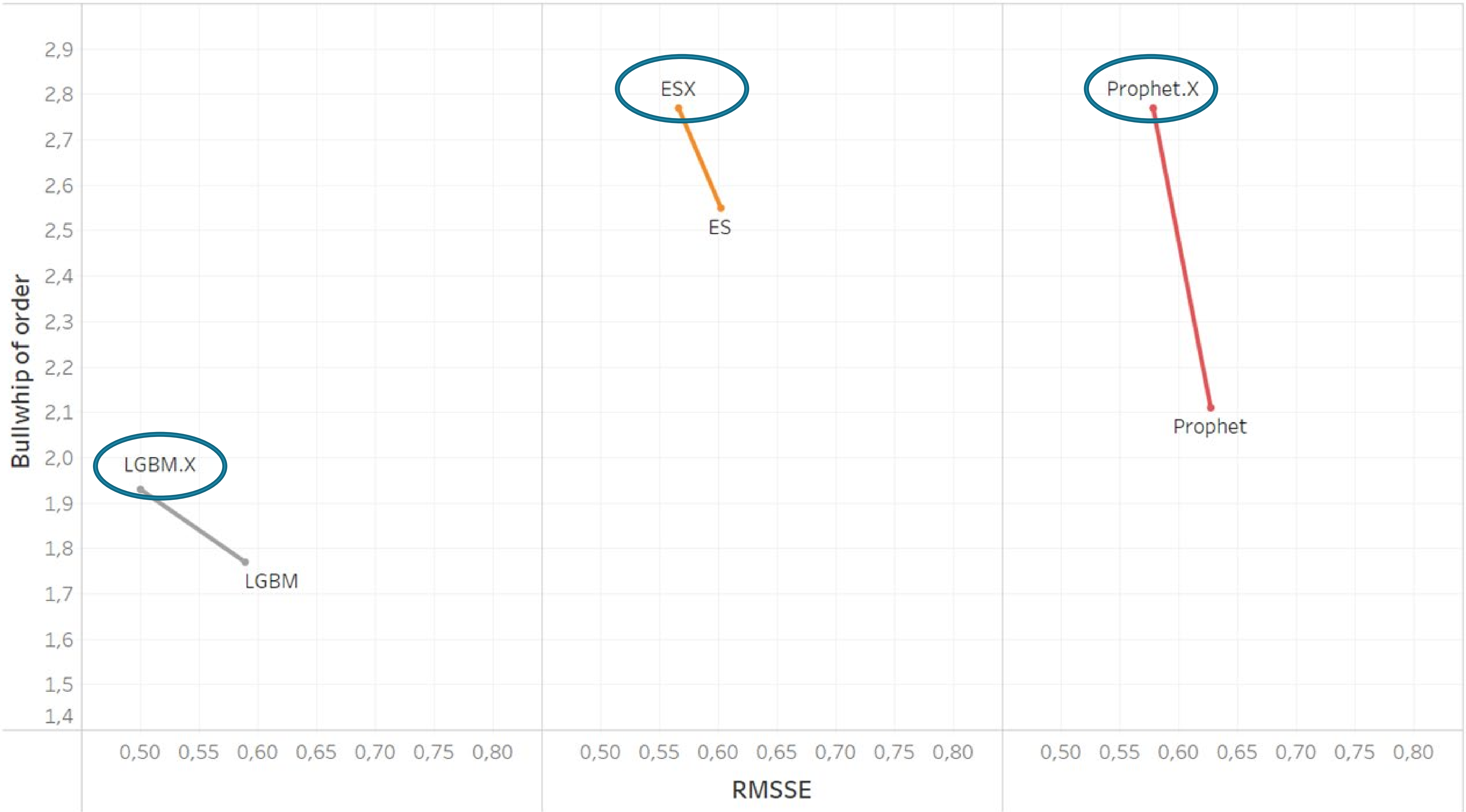


The use of Explanatory data & Feature engineering provide the biggest "bang for the buck"

Methods	**Accuracy gain** by adding **explanatory data & feature engineering**
Decision-tree framework	**+19.97%**
Exponential smoothing	**+5.98%**
Prophet	**+7.81%**

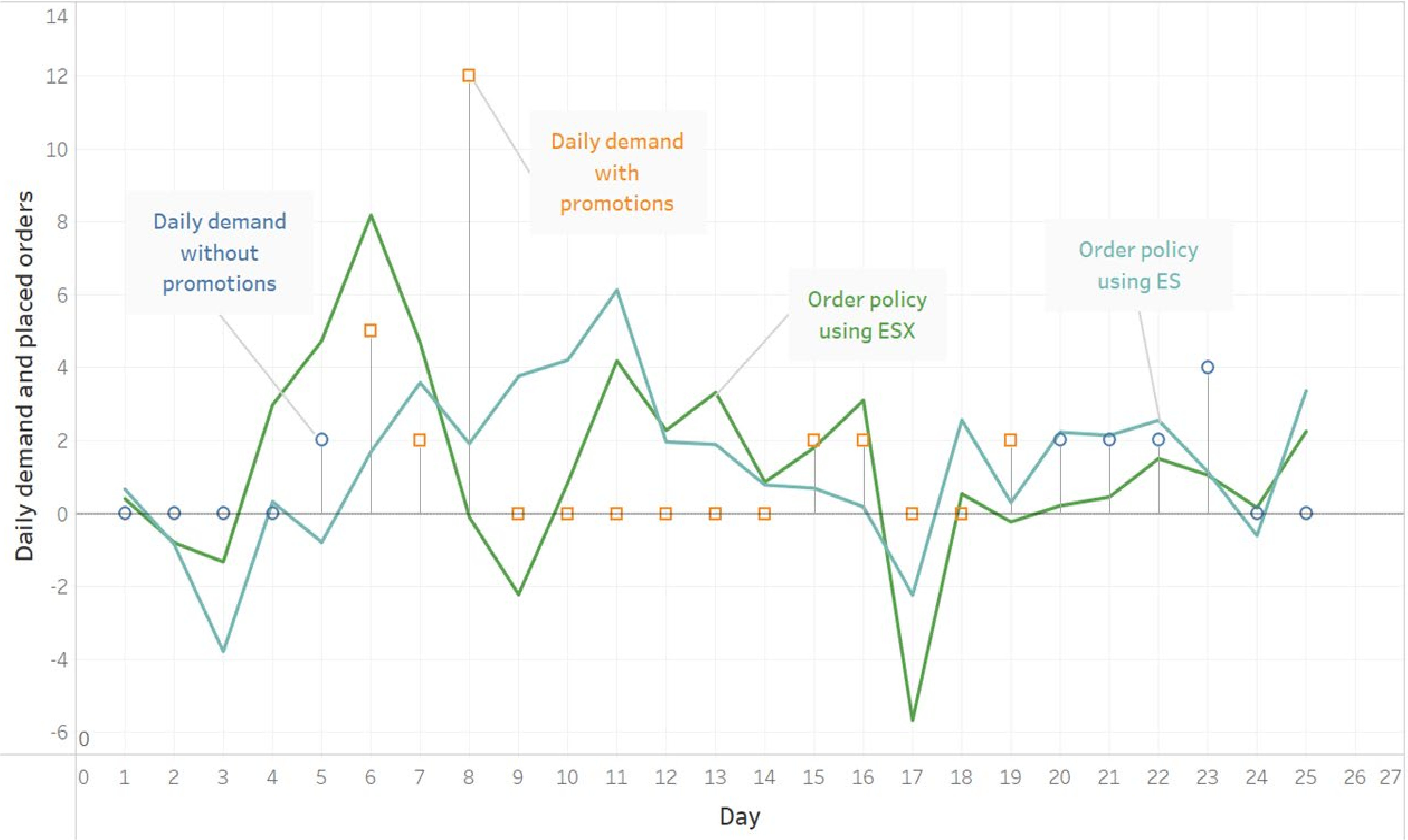
But there is a side-effect…

Using explanatory variables leads to a higher bullwhip



KU LEUVEN

Explanatory variables make the sales forecasts, and consequently the replenishment, more responsive



KU LEUVEN

Conclusion

Machine Learning can be simplified and still outperform as long as:

- We invest in **explanatory variables**
- We invest in **feature engineering**

But there is a side-effect: using explanatory variables creates a **higher bullwhip** effect in the replenishment orders.