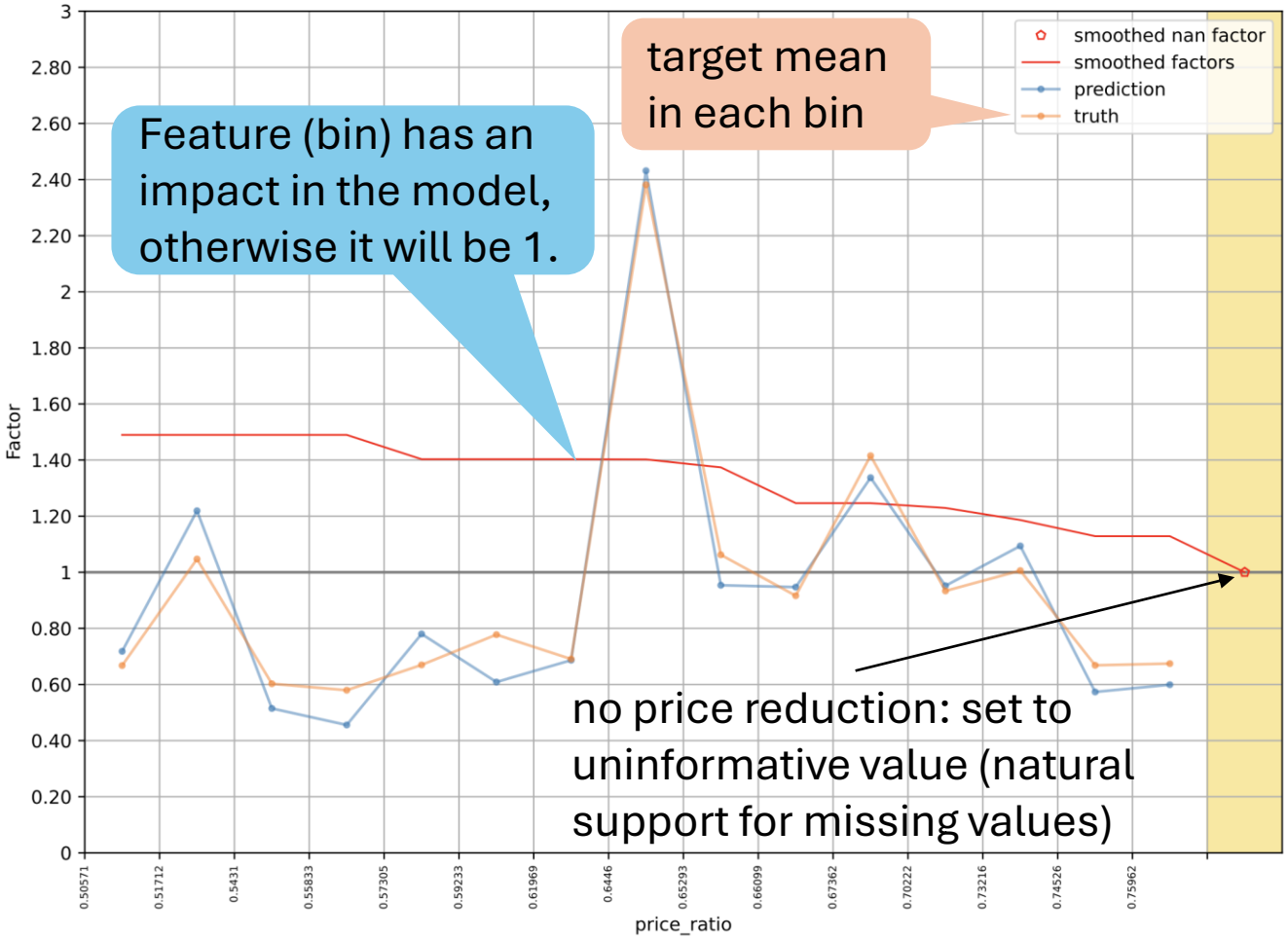
Cyclic Boosting

- binning → local optimization (low-bias method, capturing rare effects)
- smoothing over different bins (reduction of variance)
- forward-stagewise fitting (cyclic coordinate descent)
- some similarity to Generalized Additive Models (e.g., link function, interaction terms)

→ individual explainability

multiplicative mode → factors

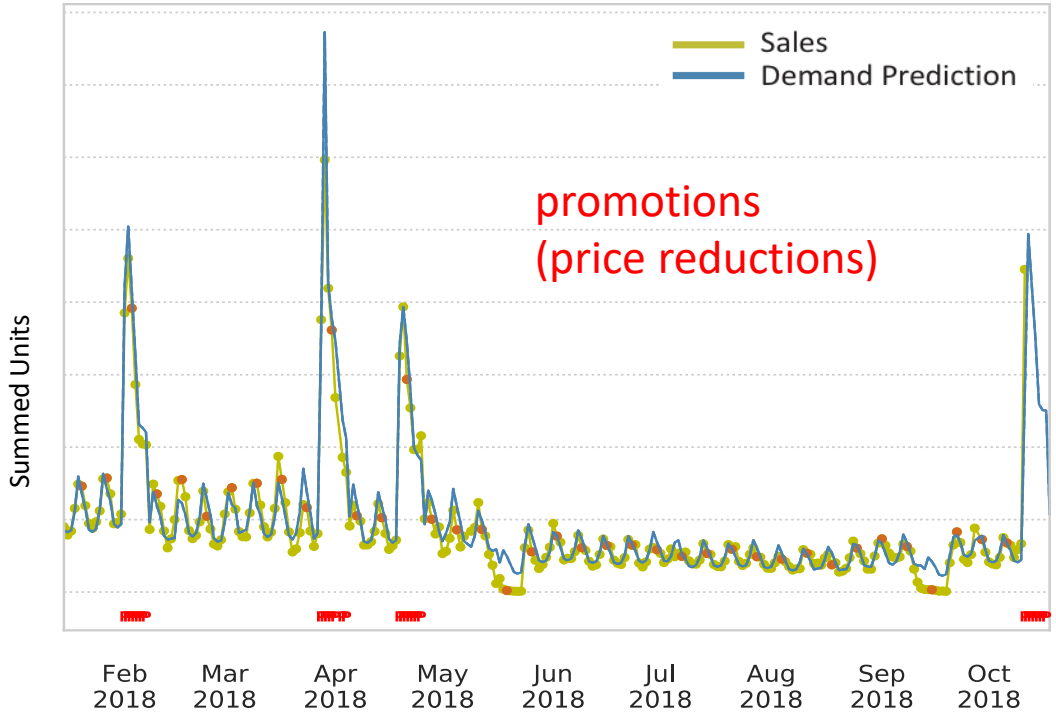


Retail Demand Forecasting & Replenishment

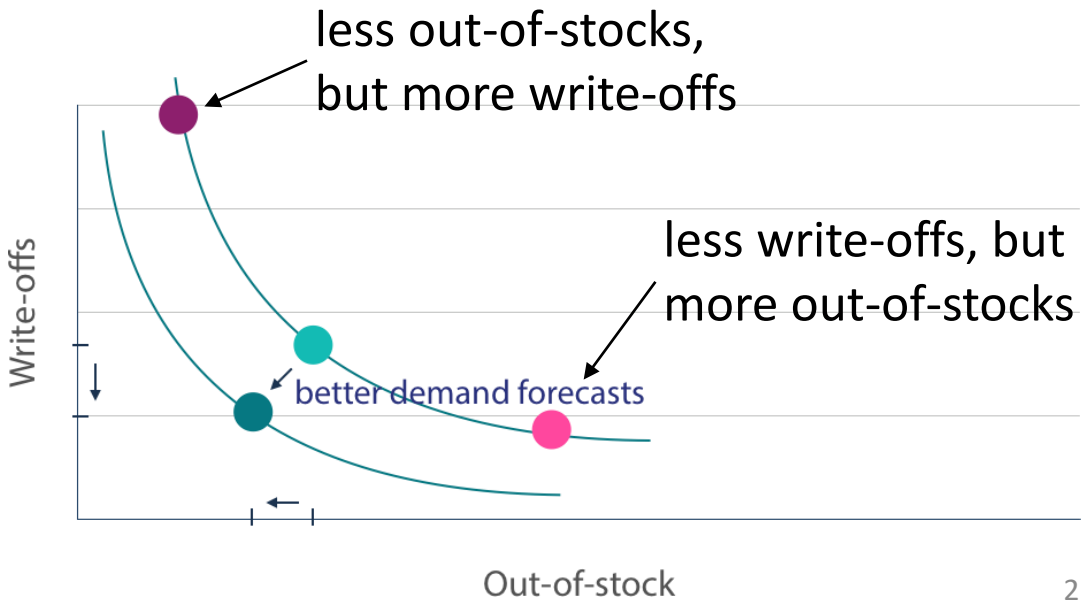
many individual time series to consider

typical retail grocery chain:
- products (items): ~20k
- locations (stores): ~500
- daily/hourly aggregated sales

categorical features important:
products and locations → high cardinality

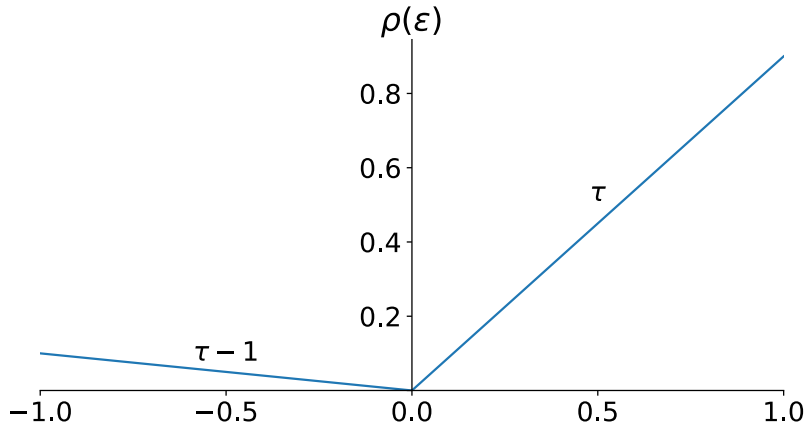


order optimization → choose demand quantile:



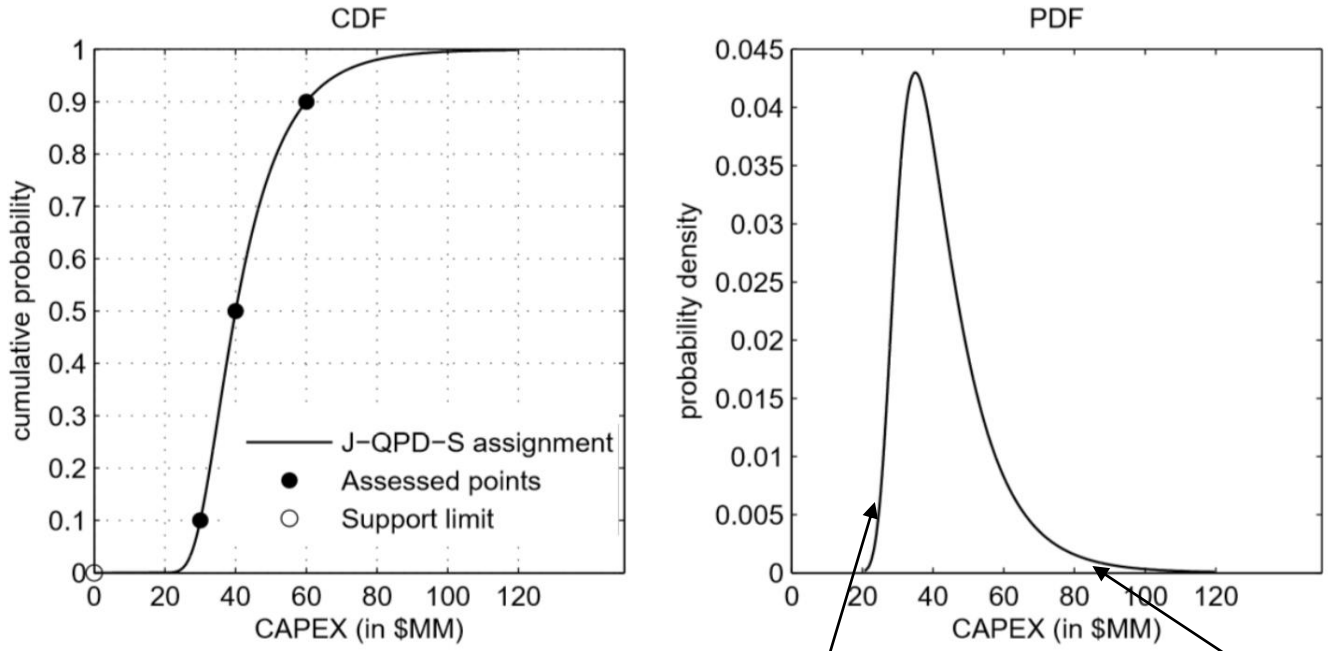
Prediction of Probability Distributions

pinball loss for quantile predictions:



$$(1-\tau)\sum_{y_i<\hat{q}_i}(\hat{q}_i - y_i) + \tau\sum_{y_i\geq\hat{q}_i}(y_i - \hat{q}_i)$$

use in quantile-parameterized distributions:



J-QPD

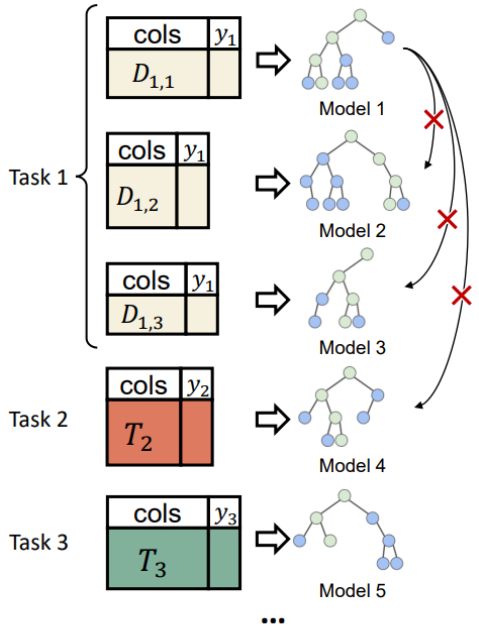
choose low demand quantile for ordering: high understock risk

choose high demand quantile for ordering: high overstock risk

Idea of Tabular Foundation Models

foundation models prevalent in vision and language (unstructured, homogenous data)

but not (yet) in structured/tabular, **heterogenous** data



Existing works:
- **one** model, **one** dataset;
- not transferable across datasets
- if transferable, needs finetuning on each dataset

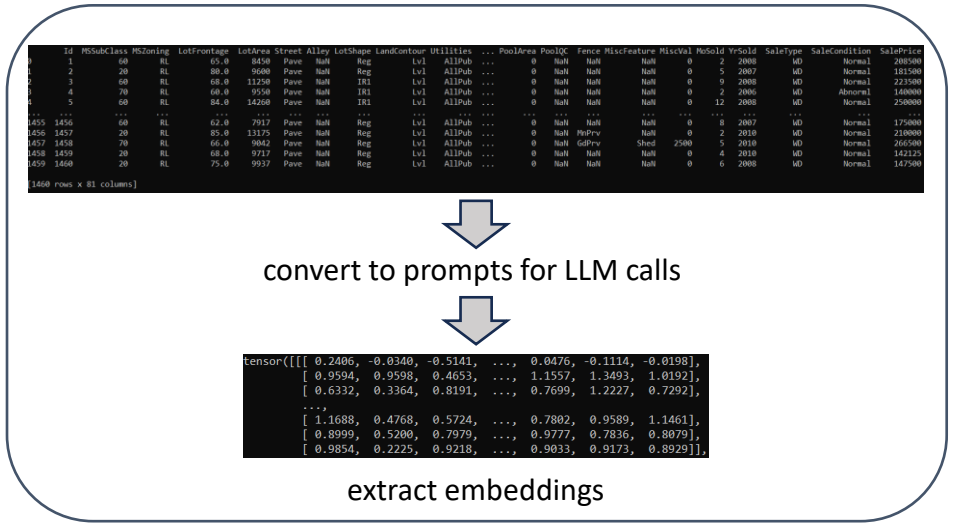
source

goal:

- pre-training across data sets and even different tasks
- finetuning on small data sets
- benefit from world knowledge in LLMs, for example in terms of data imputation

Concept Model: tabGPT

to overcome data integration challenge:



MLP head

transformer encoder

embeddings:

+

×

numerical value

column name or description

categorical value

column name or description

Transformer for Numerical Data

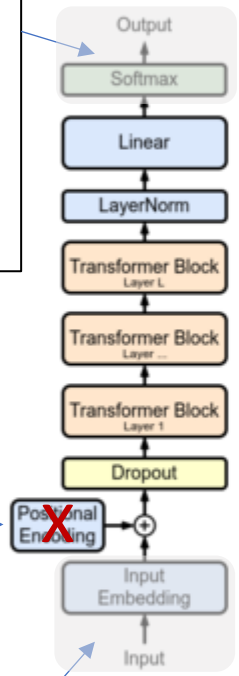
adaptions to GPT architecture:

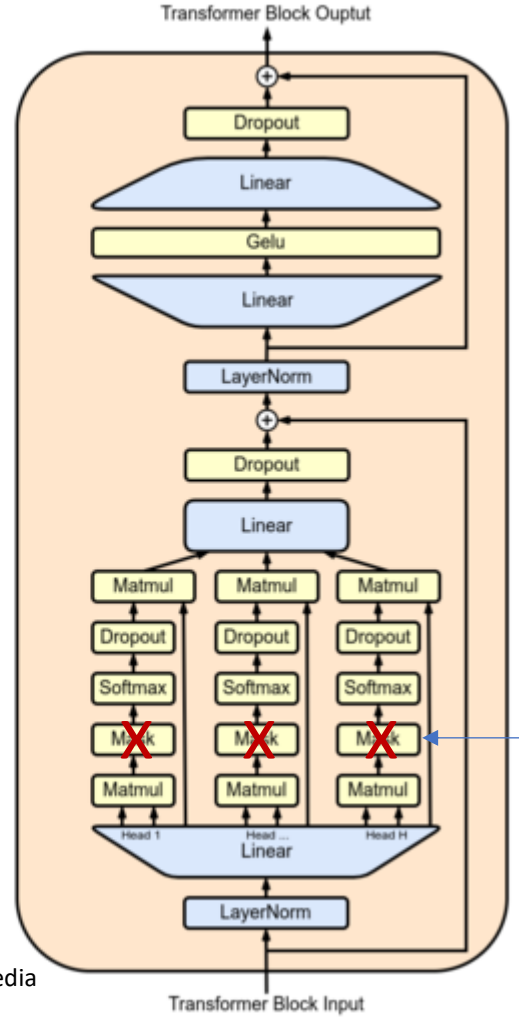
replace language model head and loss function to reflect tabular prediction setup

also use average pooling over sequence instead of last element for prediction

drop positional encoding (permutation invariance)

input previously generated column embeddings directly





drop causal masking (permutation invariance)

from wikipedia