Participants(14): Shreyantan Chanda, Moustapha Diallo, Liam Hillis, Naomi Kim, Peter Kochek, Bowen Li, Diego De Los Santos, Josih Torres, Jingyao Wang Wu, Jefferey Zhang, Joseph Xia, Zesheng Yu, Alexander Yu, Hanzhang Li

Sales Modeling With Economic Indicators

Supervisors: Gabe Hart, Alex Iosevich, Brian McDonald, Will Burstein.

August 8, 2025

Data Summary

	A	B	C	D	E	F	G	H	I	J	K	L
1		product_ic	location_ic	state	departmei	_201405	_201406	_201407	_201408	_201409	_201410	_201411
2	0	7	516	NY	Clinique	0	0	0	0	0	0	0
3	1	10	353	NY	Kitchen El	1106	287.04	1334.88	692.16	406	1845.6	146.32
4	2	10	516	NY	Kitchen El	70	33.28	75.6	56	85.84	148.8	57.04
5	3	14	8	NJ	Clinique	26	45.76	86.4	38.08	238.96	93.6	47.12
6	4	14	10	NY	Clinique	800	468	503.28	665.28	450.08	444	133.92
7	5	14	32	CA	Clinique	62	85.28	51.84	1460.48	452.4	540	262.88
8	6	14	57	FL	Clinique	98	62.4	62.64	67.2	90.48	36	47.12
9	7	14	88	CA	Clinique	918	47.84	185.76	2943.36	403.68	314.4	32.24
10	8	14	141	CA	Clinique	322	322.4	142.56	174.72	53.36	249.6	86.8

- 999 product-location rows
- 155 weeks (Feb 2014–Jan 2017)
- Total cells = 154,845

Data Cleaning

- **Initial Cleaning Steps:**
  - Remove trailing-year zero → 77 rows deleted
  - Nullify leading pre-launch zeros → 14517 cells out of 154,845(155*999) = 9.3%
- **Anomaly Detection Approaches:**

Method	**What it does**	**Result**
Modified Z-Score	Detects trend outliers	14803 cells flagged (10.3%)
Seasonality-aware residual	Detects seasonal deviations	0 cells flagged
Rolling % Change	Guards against sudden shifts	72 clean rows(k=1, 265 k=3, 554; k=5, 875)

Data Aggregation

- **Aggregation Options:**
  - By Product (298 rows) – used for model
  - By Location (321 rows) – not used due to data sparsity
- **Reasons:**
  - Product-level patterns = cleaner trends
  - Location-level data = too sparse/erratic

Data Augmentation

- **Gaussian Noise:** Add random variation to data
- Purpose: Improve generalization and reduce overfitting
- **Accuracy:** Top10 nMAE 0.31-0.43 $\rightarrow$ 0.25-0.34

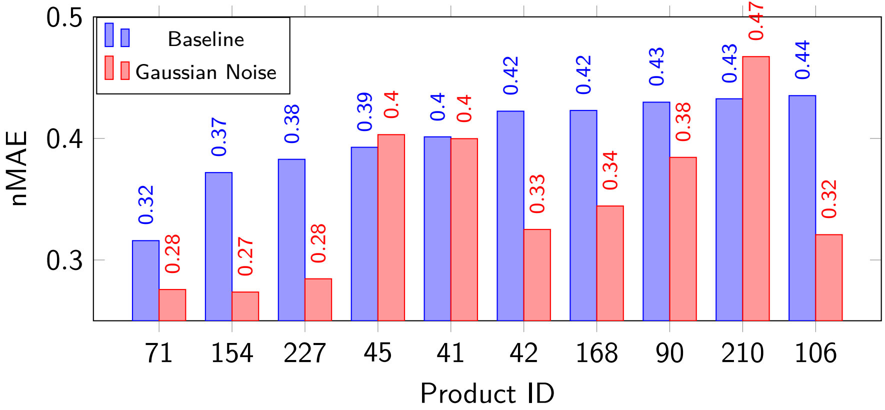


Figure: nMAE Comparison of Top 10 Products: Baseline vs. Gaussian Noise

Models

- XG Boost
  - A decision tree machine learning library.
  - Great for structured data. Easy to train and tune.

Models

- XG Boost
  - A decision tree machine learning library.
  - Great for structured data. Easy to train and tune.
- Long short term memory neural network
  - A recurrent neural network designed for time-series data
  - Can capture week-week changes + seasonality. Auto learns sequential relationships.

Models

- XG Boost
  - A decision tree machine learning library.
  - Great for structured data. Easy to train and tune.
- Long short term memory neural network
  - A recurrent neural network designed for time-series data
  - Can capture week-week changes + seasonality. Auto learns sequential relationships.
- Rolling Average
  - Works by computing avg of past 3 weeks. Doesn't learn / train. Benchmark

Models

- XG Boost
  - A decision tree machine learning library.
  - Great for structured data. Easy to train and tune.
- Long short term memory neural network
  - A recurrent neural network designed for time-series data
  - Can capture week-week changes + seasonality. Auto learns sequential relationships.
- Rolling Average
  - Works by computing avg of past 3 weeks. Doesn't learn / train. Benchmark
- Feed forward network
  - Type of neural network where information flows in one direction, from input to output. No cycles or feedback loops

LSTM/XG Boost Parameters

- LSTM
  - Window size of 24, hidden size of 16, 1 layer, and 100 epochs
- XG Boost
  - N estimators is 500, learning rate of 0.01, and max depth of 9
- Baseline Performances (Product)
  - XG Boost Model: 0.275 nMAE
  - LSTM: 0.316 nMAE

Feature	Importance
353_lag_1	0.135890
1429_rolling_std_5	0.127123
1263_rolling_mean 3	0.100379
623_lag_1	0.070093
752_rolling_std 5	0.069776
1436_momentum	0.053268
566_momentum	0.035335
997_momentum	0.034656
1008_momentum	0.034055
961_rolling_mean_3	0.032660

Measuring Accuracy

- Definitions
  - Our forecast predictions are represented by $F = (f_1, f_2, ...f_n)$
  - The actual values are represented by $V = (v_1, v_2, ...v_n)$

Measuring Accuracy

- Definitions
  - Our forecast predictions are represented by $F = (f_1, f_2, ... f_n)$
  - The actual values are represented by $V = (v_1, v_2, ... v_n)$
- MAE
  - $MAE = \frac{1}{n} \sum_{i=1}^{n} |f_i - v_i|$
  - Does not account for scale differences on its own

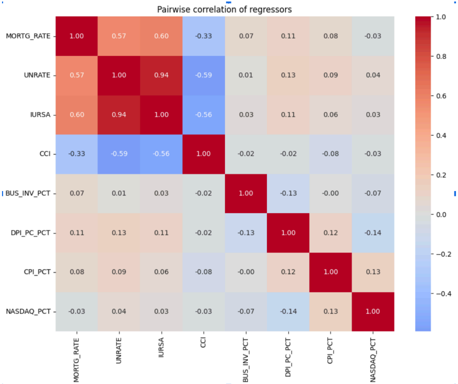
Measuring Accuracy

- Definitions
  - Our forecast predictions are represented by $F = (f_1, f_2, ...f_n)$
  - The actual values are represented by $V = (v_1, v_2, ...v_n)$
- MAE
  - $MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - v_i|$
  - Does not account for scale differences on its own
- MSE
  - $MSE = \frac{1}{n}\sum_{i=1}^{n}(f_i - v_i)^2$
  - Penalizes large misses more
  - Also does not scale difference

Measuring Accuracy

- Definitions
  - Our forecast predictions are represented by $F = (f_1, f_2, ...f_n)$
  - The actual values are represented by $V = (v_1, v_2, ...v_n)$
- MAE
  - $MAE = \frac{1}{n}\sum_{i=1}^{n} |f_i - v_i|$
  - Does not account for scale differences on its own
- MSE
  - $MSE = \frac{1}{n}\sum_{i=1}^{n} (f_i - v_i)^2$
  - Penalizes large misses more
  - Also does not scale difference
- **Normalized MAE**
  - Normalized MAE $= \frac{\frac{1}{n}\sum_{i=1}^{n} |f_i - v_i|}{\frac{1}{n}\sum_{i=1}^{n} |v_i|}$
  - Allows for fair comparison across products

- **Pairwise Pearson Correlation Heatmap**
  - High-correlation pairs ($|r| \geq 0.8$): UNRATE IURSA $\rightarrow$ r $= + 0.9377$



Multicollinearity Check(2)

- **Variance Inflation Factor (VIF)**

Variable	**VIF**
UNRATE	8.962530
IURSA	8.916796
MORTG_RATE	1.584676
CCI	1.550863
DPI_PC_PCT	1.086566
NASDAQ_PCT	1.057646
CPI_PCT	1.054268
BUS_INV_PCT	1.033329

Table: Rule of Thumb for VIF Interpretation

VIF Range	**Interpretation**	**Action**
$\text{VIF} \leq 5$	Low/moderate overlap	Safe
$5 < \text{VIF} \leq 10$	Noticeable	Consider thinning
$\text{VIF} > 10$	Severe multicollinearity	Drop or combine

Multicollinearity Check(3)

- **Condition Number of the correlation matrix:**
- $7.0 < 30 - 35 \rightarrow$ the whole regressor block is numerically well-conditioned.
- **Conclusion:**
  - Most regressors are safe to use
  - But IURSA and UNRATE carry near-duplicate information - additional testing to choose which one to retain

Regressor Results

- Economic Indicators had both positive and negative effects.
- IURSA, CCI, DPI, and NASDAQ closing prices all improved the regression model.
- CPI, Gas Prices, and Average Business Inventory all created similar performance as the baseline.

	Regressor	NMAE	MAPE	CVRMSE
1	Baseline	0.32-0.44	24.67-42.56	93.41-1037.18
2	Gaussian	0.25-0.34	21.61-34.31	81.15-533.59
3	DPI	0.24-0.37	26.22-56.09	42.868-265.08
4	Gas Prices	0.30-0.41	31.31-54.27	49.21-983.84
5	Business_inventories	0.32-0.41	30.54-56.13	66.61-844.35
6	CCI	0.23-0.33	25-28	62-320
7	CPI	0.32-0.40	27.53-46.80	57.82-967.14
8	NASDAQ	0.29-0.35	25.68-36.65	56.98-396.89
9	IURSA	0.24-0.4	24.72-51.21	47-249

Multiple Regressors

- CCI and NASDAQ closing yielded worse performance
- CCI, IURSA, and DPI together (our 3 helpful regressors) led to better performance than the baseline

	Regressor	NMAE	MAPE	CVRMSE
10	CPI and NASDAQ	0.34-0.41	24.69-45.75	41.48-404.90
11	CPI and IURSA and DPI	0.24-0.36	25.31-53.56	42.67-249.63
12	CCI and IURSA and Gaussian	0.25-0.36	25.06-43.49	41.47-248.83

Fourier Norms Definition

- Let $f : \mathbb{Z}_N \to \mathbb{R}$ and $\widehat{f}$ be its fourier transform

$$\|\widehat{f}\|_{L^1(\mu)} = \frac{1}{N} \sum_{m \in \mathbb{Z}_N} |\widehat{f}(m)|$$

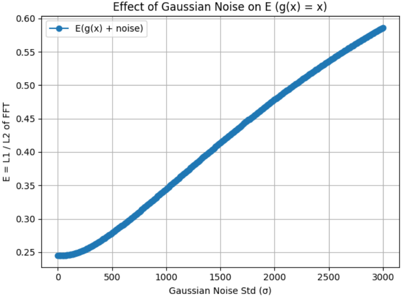
$$\|\widehat{f}\|_{L^2(\mu)} = \left[\frac{1}{N} \sum_{m \in \mathbb{Z}_N} |\widehat{f}(m)|^2 \right]^{1/2}$$

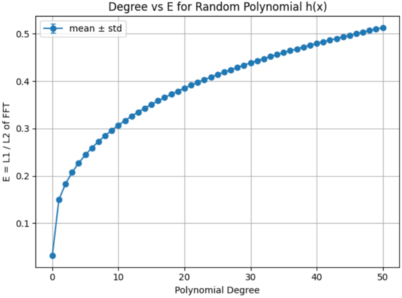
Fourier Ratio

$$\mathcal{E} = \frac{\|\widehat{f}\|_{L^1(\mu)}}{\|\widehat{f}\|_{L^2(\mu)}}, \quad \frac{1}{\sqrt{N}} \leq \mathcal{E} \leq 1$$

- Low $\mathcal{E}$: spectrum concentrated in few frequencies $\rightarrow$ well-approximated by a low-degree trigonometric polynomial $\rightarrow$ more forecastable
- High $\mathcal{E}$: spectrum spread across many frequencies $\rightarrow$ noise-like $\rightarrow$ less forecastable

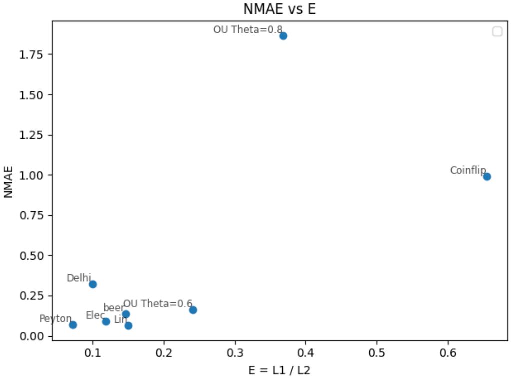
Synthetic Data: $\mathcal{E}$ Rises with Noise and Complexity





- Noise $\uparrow \Rightarrow \mathcal{E} \uparrow$ (approaches random-sequence level)
- Polynomial degree $\uparrow \Rightarrow \mathcal{E} \uparrow$ (complexity acts like noise)

Real-World Data: $\mathcal{E}$ Predicts Forecast Error



- Clear positive relationship between $\mathcal{E}$ and NMAE
- Higher $\mathcal{E} \Rightarrow$ higher forecast error
- Correlation: Pearson $r = 0.8577$ ($p = 0.0289$), Spearman $\rho = 0.6571$ ($p = 0.156$)

L_1 Imputation

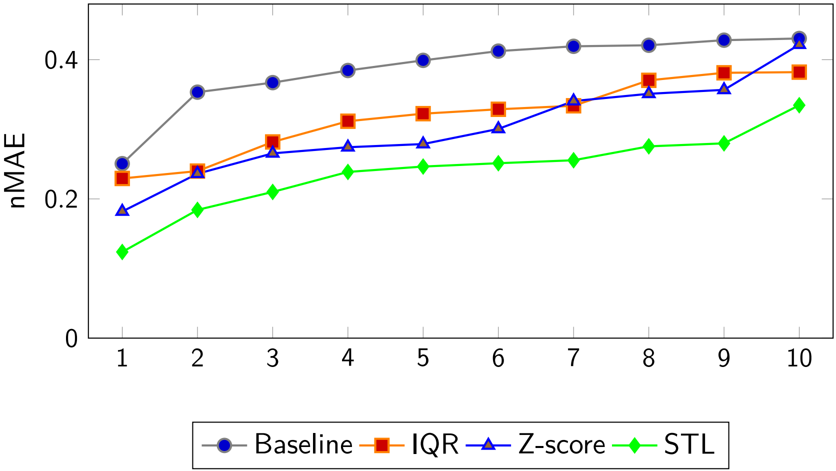


Figure: Top-10 nMAE by method (ranked) – Line Plot.

Next Steps

- Apply $\mathcal{E}$ to a larger and more diverse set of real-world time series

Next Steps

- Apply $\mathcal{E}$ to a larger and more diverse set of real-world time series
- Establish thresholds for “forecastable” vs. “non-forecastable” series

Next Steps

- Apply $\mathcal{E}$ to a larger and more diverse set of real-world time series
- Establish thresholds for "forecastable" vs. "non-forecastable" series
- Compare $\mathcal{E}$ with alternative forecastability measures(e.g., spectral entropy, Hurst exponent)

Next Steps

- Apply $\mathcal{E}$ to a larger and more diverse set of real-world time series
- Establish thresholds for “forecastable” vs. “non-forecastable” series
- Compare $\mathcal{E}$ with alternative forecastability measures(e.g., spectral entropy, Hurst exponent)
- Extend the analysis to multivariate settings with $f : \mathbb{Z}_N \to \mathbb{R}^n$

Next Steps

- Apply $\mathcal{E}$ to a larger and more diverse set of real-world time series
- Establish thresholds for "forecastable" vs. "non-forecastable" series
- Compare $\mathcal{E}$ with alternative forecastability measures(e.g., spectral entropy, Hurst exponent)
- Extend the analysis to multivariate settings with $f\colon \mathbb{Z}_N \to \mathbb{R}^n$
- Examine sensitivity of $\mathcal{E}$ to sampling rate, seasonality, and preprocessing choices