FORECASTING CONSTRUCTION MATERIAL PRICE INDICES USING PRODUCER PRICE INDEX (PPI) DATA

Sebastian Robalino-Ordonez | AI in the Built Environment | University of Florida

PROBLEM STATEMENT

Volatility in construction material costs poses a recurring challenge in budgeting and procurement. Sudden changes in commodities such as lumber, steel, and cement can inflate project costs or delay schedules.

This project uses Producer Price Index (PPI) data from the U.S. Bureau of Labor Statistics to forecast future price-index values for core building materials, helping planners anticipate cost movement.

Input: Year-Month (YYYY-MM) and Material Type (Lumber, Iron & Steel, Construction Materials)

Output: Predicted PPI value (base 1982 = 100)

This forecast supports cost-planning automation in the built environment.

DATA

Data sourced from the Federal Reserve Economic Data (FRED) portal, published by the U.S. Bureau of Labor Statistics (BLS).

Includes three key series:

- Construction Materials – WPUSI012011 (1947–2025)
- Lumber – WPU081 (1926–2025)
- Iron & Steel – WPU101 (1926–2025)

Each series is monthly and non-seasonally adjusted.

Combined, they form a multi-material time series of nearly 3,000 monthly observations.

```
           date      ppi               material  year  month  ppi_last_month
0    1947-01-01   22.200  Construction Materials  1947      1             NaN
1    1947-02-01   22.500  Construction Materials  1947      2          22.200
2    1947-03-01   22.900  Construction Materials  1947      3          22.500
3    1947-04-01   23.200  Construction Materials  1947      4          22.900
4    1947-05-01   23.300  Construction Materials  1947      5          23.200
...         ...      ...                    ...   ...    ...             ...
3333 2025-05-01  269.699                 Lumber  2025      5         274.511
3334 2025-06-01  266.567                 Lumber  2025      6         269.699
3335 2025-07-01  264.248                 Lumber  2025      7         266.567
3336 2025-08-01  266.308                 Lumber  2025      8         264.248
3337 2025-09-01  259.721                 Lumber  2025      9         266.308

      ppi_last_year  rolling_12
0               NaN         NaN
1               NaN         NaN
2               NaN         NaN
3               NaN         NaN
4               NaN         NaN
...             ...         ...
...
```

FEATURE ENGINEERING

Data pre-processing in Python using Pandas:

- Added Year, Month, and Material columns
- Created lag features (ppi_last_month, ppi_last_year)
- Added a 12-month rolling average
- Dropped missing or incomplete records

```python
data["year"] = data["date"].dt.year
data["month"] = data["date"].dt.month
data["ppi_last_month"] = data.groupby("material")["ppi"].shift(1)
data["ppi_last_year"] = data.groupby("material")["ppi"].shift(12)
data["rolling_12"] = data.groupby("material")["ppi"].rolling(window=1
```

EVALUATION METRIC

Performance measured using Mean Absolute Error (MAE), expressed in index points.

MAE represents the average absolute deviation between actual and predicted PPI values.

It is robust to outliers and interpretable for economic forecasting.

$$MAE = \frac{1}{n}\Sigma|y - \hat{y}|$$

BASELINES

Naïve Forecast: Predicts the next month's PPI as the last observed value.

12-Month Rolling Average: Smooths short-term fluctuations for a slower-moving forecast.

The baselines establish lower-bound performance for future models.

	Material	MAE_Naive	MAE_Rolling12
0	Construction Materials	0.718	2.458
1	Iron & Steel	1.527	5.013
2	Lumber	2.309	5.286

MAIN MODELS

Linear Regression

- Captures long-term inflation trends.
- Fast, interpretable, and ideal for small datasets.
- Uses: year, month, lagged PPI, rolling average.
- Train/test split: 80/20 by time order.

Decision Tree Regressor

- Non-linear comparison (oracle).
- Captures short-term spikes but struggles with trending data.
- Often overfits when data show strong upward inflation patterns.

MODEL RESULTS

Linear Regression maintained low errors across all materials, while the Decision Tree model overfit, leading to high MAE values.

Key takeaway: simple linear models outperform non-linear models for monotonic, inflationary economic data.

	Material	MAE_LinearRegression	MAE_DecisionTree
0	Construction Materials	1.706	62.029
1	Iron & Steel	5.577	77.011
2	Lumber	6.765	32.430

INTERPRETATION

Linear Regression models the slow, consistent price rise effectively and generalizes to unseen months.

Decision Tree Regressors fail to extrapolate, confirming the hypothesis that interpretable, linear methods are more reliable for price index forecasting.

CHALLENGES

- Missing or delayed monthly reports required imputation.
- Post-2020 inflation spikes increased volatility.
- Material variance differences demanded normalization.
- Decision Tree instability on non-stationary trends.

CONCLUSION

- Integrate macroeconomic indicators (energy costs, interest rates).
- Experiment with LSTM time-series models.
- Automate monthly dataset updates via the FRED API.
- Expand visualization dashboard for planners and researchers.

FUTURE WORKS

- Integrate macroeconomic indicators (energy costs, interest rates).
- Experiment with LSTM time-series models.
- Automate monthly dataset updates via the FRED API.
- Expand visualization dashboard for planners and researchers.

REFERENCES

- Aslam, B. et al. (2023). AI forecasting of construction cost index. Societies, 13(10), 219.
- Lederer, L. et al. (2024). Comparative models for construction price forecasting. Buildings, 14(5), 1325.
- Liu, Q. et al. (2024). Data-driven construction material-price prediction. Buildings, 14(10), 3156.
- Ma, M. et al. (2024). Review of construction price-forecasting methods. Int. J. Constr. Manag., 24(11), 1191–1200.
- Tang, X. et al. (2019). GA-SVR + ARIMA model for PPI forecasting. Comput. Mater. Continua, 58(2), 463–491.
- FRED / BLS (1926–2025). Producer Price Index datasets (WPUSI012011, WPU081, WPU101).

VISUALIZATION