Can a neural network beat 'no change' on Treasury yields? Walk-forward forecasts, 2006–2026

Gaurav Gupta · Research note · 4 November 2026 · US Treasury par yields, daily, Jan 1990 – Oct 2026; 249 monthly forecasts

26.0 bp
Typical one-month 10-year surprise under no change (LSTM 26.1, Diebold-Li 27.0)

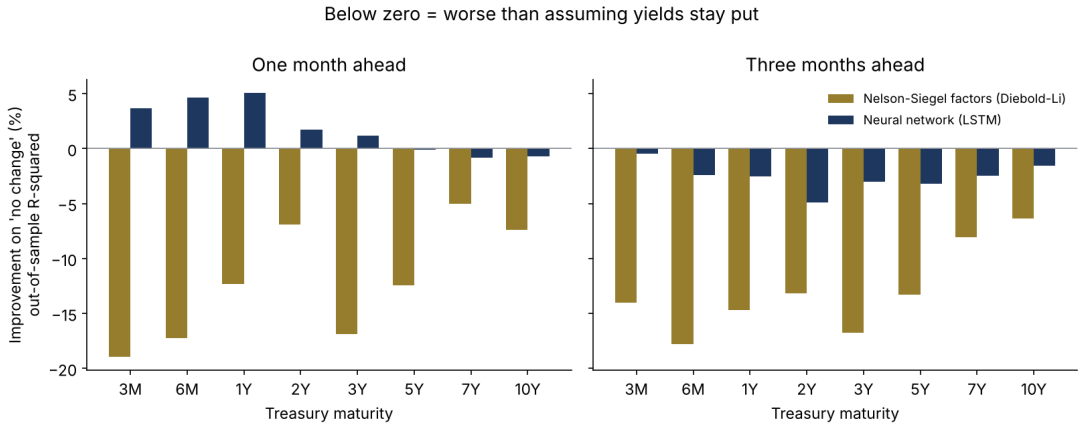
+5.1%
Best LSTM gain on no change (1Y); best DM t = −1.45, not significant

−5% to −19%
Diebold-Li vs no change, one month ahead: worse at every maturity

2–4 bp
Typical LSTM forecast move vs 21–27 bp actual moves

Maturity	No change (bp)	LSTM (bp)	Diebold-Li (bp)	LSTM R^2 vs no change	DL R^2 vs no change
3M	21.8	21.4	23.7	+3.7%	−18.9%
1Y	21.1	20.5	22.3	+5.1%	−12.3%
2Y	23.8	23.6	24.6	+1.7%	−6.9%
5Y	26.8	26.8	28.5	−0.1%	−12.4%
10Y	26.0	26.1	27.0	−0.7%	−7.4%

Root-mean-square error of one-month-ahead forecasts at month-ends, Jan 2006 – Sep 2026. R^2 = share of no-change squared error removed (Campbell-Thompson). Three months ahead both models lose at every maturity.



Out-of-sample R^2 against no change, by maturity.

SOURCES

- US Treasury daily par yield curve via defeatbeta/yahoo-finance-data (ODC-BY), sha256-pinned
- Diebold & Li (2006), J. Econometrics 130(2); Nelson & Siegel (1987)
- Diebold & Mariano (1995); Campbell & Thompson (2008)

METHOD

- Eight maturities 3M–10Y; forecasts at month-ends from Jan 2006
- No change; Nelson-Siegel factors + direct AR per factor; LSTM (32 units, 21-day window, 3 seeds)
- Annual refits; training targets on or before the refit date; early stopping on the latest 15%
- RMSE, out-of-sample R^2, Diebold-Mariano (Newey-West), hit rate, duration overlay

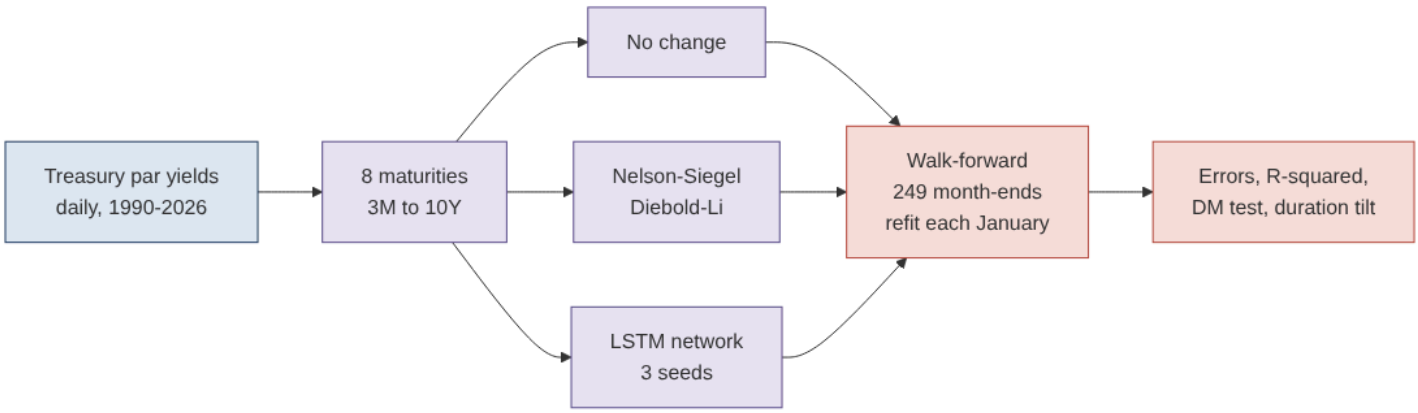
KEY CALCULATIONS

- Typical 10Y surprise: 26 bp × \$100k/bp ≈ \$2.6M a month; 2022: 44 bp ≈ \$4.4M
- Duration overlay, \$100k/bp, 249 months: LSTM +\$41.0M (54% right), DL −\$51.8M (44%)
- Sharpe-like 0.22 and −0.28 a year: t ≈ 1.0 and −1.3 over 21 years

RECOMMENDATIONS

- Require any curve forecast to beat no change out of sample, by maturity and horizon, before it drives hedges.
- Test with walk-forward refits, significance tests and a 2022 sub-period.
- Size hedges on the distribution of surprises, not on point forecasts.
- Look for value at the front end first; it is the only place either model helped.
- Re-test forecasts annually; retire any below no change over three years.

Independent research on public data; not deployed; not investment advice. Dollar figures are illustrative. Analysis, code, charts and text were produced with AI assistance and reviewed by the author. Code: github.com/guptagaurav-info/yield-curve-deep-learning



Pipeline. Blue: sources. Lavender: code. Red: checks.