
Regression-Based Forecasting

BSAD 8310: Business Forecasting — Lecture 2

Department of Economics
University of Nebraska at Omaha
Spring 2026

-
- 1 From Benchmarks to Regression**
 - 2 The Regression Model for Forecasting**
 - 3 Trend and Seasonality as Regressors**
 - 4 Prediction Intervals**
 - 5 Autoregressive Models**
 - 6 Pitfalls**
 - 7 Lecture 2: Key Takeaways**

From Benchmarks to Regression

Naïve models ignore information. Regression lets us use it.

Recall the four benchmarks from Lecture 1:

| Benchmark | Information used |
|---------------------|-----------------------------|
| Naïve | Last observed value only |
| Seasonal naïve | Same season, last year only |
| Historical mean | Unconditional average |
| Random walk + drift | Last value + average change |

None of these benchmarks can use **leading indicators** — variables that move *before* y_t does.

Examples: consumer confidence → retail sales next quarter; interest rates → housing starts.

Information we can exploit:

- **Leading indicators:** variables observable at T that predict y_{T+h}
- **Structural relationships:** economic theory says x_t affects y_t
- **Deterministic patterns:** known trend, calendar effects, holidays
- **Own-lags:** past y values predict future y

- Retailer: use *ad spend, price, and calendar flags* to forecast demand
- Macro analyst: use *leading index* to forecast next quarter's GDP
- Bank: use *yield curve* to forecast loan defaults 6 months ahead

The Regression Model for Forecasting

OLS is not just for causal inference — it is also a forecasting tool.

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + \varepsilon_t, \quad t = 1, \dots, T$$

- y_t : scalar response (what we want to forecast)
- $\mathbf{x}_t = (1, x_{1t}, \dots, x_{kt})'$: $(k + 1)$ -vector of predictors (including intercept)
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)'$: unknown coefficients
- $\varepsilon_t \sim WN(0, \sigma^2)$: white-noise errors

Notation reminder (from Lecture 1): ε_t denotes the true model innovation; $e_t = y_t - \hat{y}_{t|t-1}$ denotes the realized forecast error. These are conceptually distinct.

Stack observations into matrices: $\mathbf{y} = (y_1, \dots, y_T)', \mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_T)' (T \times (k + 1))$.

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \sum_{t=1}^T (y_t - \mathbf{x}'_t \boldsymbol{\beta})^2 = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$$

Gauss-Markov theorem: Under classical assumptions, $\hat{\boldsymbol{\beta}}$ is the **Best Linear Unbiased Estimator** (BLUE) (Hamilton 1994, Ch. 10).

Classical assumptions for valid OLS inference:

1. $\mathbb{E}[\varepsilon_t | \mathbf{X}] = 0$ (zero conditional mean)
2. $\mathbb{E}[\varepsilon_t^2 | \mathbf{X}] = \sigma^2$ (homoscedasticity)
3. $\mathbb{E}[\varepsilon_t \varepsilon_s | \mathbf{X}] = 0$ for $t \neq s$ (no autocorrelation)

The forecasting constraint: to use \mathbf{x}_{T+h} in the forecast $\hat{y}_{T+h|T}$, every element of \mathbf{x}_{T+h} must be *known* or *forecastable* at time T .

Type 1: Deterministic

Always known at T :

- Time index: t, t^2
- Seasonal dummies
- Holiday indicators
- Trend functions

Type 2: Lagged values

Known when $h \leq \text{lag}$:

- y_{t-1}, y_{t-2}
- $x_{1,t-1}$ (lagged)
- Economic releases with delay

Type 3: External regressors

Must be forecast first:

- GDP, CPI forecasts
- Competitor prices
- Weather forecasts
- Budget scenarios

Given $\hat{\beta}$ estimated on the training set $\{1, \dots, T\}$:

$$\hat{y}_{T+h|T} = \mathbf{x}'_{T+h} \hat{\beta}$$

Plug the *known or forecasted* predictor vector \mathbf{x}_{T+h} into the fitted model.

Important nuances:

- $\hat{\beta}$ is estimated only on training data — never on the test set
- If \mathbf{x}_{T+h} contains forecasted values, forecast uncertainty propagates (widens intervals)
- For $h > 1$ with lagged- y predictors, use *recursive substitution* (see Section 5)

In Lab 02, \mathbf{x}_{T+h} will contain only deterministic terms (*trend index + seasonal dummies*), so \mathbf{x}_{T+h} is always known exactly.

Trend and Seasonality as Regressors

Deterministic components require no forecasting: their values are always known at the forecast origin.

The simplest structural model for a series with an upward or downward drift:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad t = 1, 2, \dots, T$$

Forecast: $\hat{y}_{T+h|T} = \hat{\beta}_0 + \hat{\beta}_1(T + h)$ — always known.

When appropriate:

- Steady linear growth or decline
- No pronounced seasonality
- Relatively stable variance

Linear trend extrapolates indefinitely. Long-horizon forecasts assume growth continues forever — verify this is substantively reasonable.

US retail sales grow by roughly 3% per year. Should you use linear or log-linear trend? What feature of the data would help you decide?

Define $D_{j,t} = 1$ if observation t is in season j , else 0.

$$y_t = \beta_0 + \sum_{j=2}^m \gamma_j D_{j,t} + \varepsilon_t$$

Include $m - 1$ dummies; drop one season as the **base category**.

Interpretation:

- β_0 : mean level in the base season (e.g., January)
- γ_j : seasonal deviation from the base in season j
- $\hat{y}_{T+h|T} = \hat{\beta}_0 + \hat{\gamma}_{s(T+h)}$

Dummy trap: Including all m dummies creates exact multicollinearity with the intercept. Always drop one season.

$$y_t = \underbrace{\beta_0 + \beta_1 t}_{\text{trend}} + \underbrace{\sum_{j=2}^m \gamma_j D_{j,t}}_{\text{seasonal}} + \varepsilon_t$$

Forecast: $\hat{y}_{T+h|T} = \hat{\beta}_0 + \hat{\beta}_1(T + h) + \hat{\gamma}_{s(T+h)}$ (always known)

Monthly retail sales exhibit both components:

- Upward trend (growing economy and population)
- Strong December spike, January dip (Lab 02 quantifies both)

Polynomial trend:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$$

Still OLS — just add t^2 as another predictor column.

Log-linear trend (when variance grows with level):

$$\ln y_t = \beta_0 + \beta_1 t + \varepsilon_t \Leftrightarrow y_t = e^{\beta_0} e^{\beta_1 t} e^{\varepsilon_t}$$

Forecasts: $\hat{y}_{T+h|T} = \exp(\hat{\beta}_0 + \hat{\beta}_1(T + h) + \frac{1}{2}\hat{\sigma}^2)$ (bias correction)

Rule of thumb: If the series grows at a roughly *constant percentage rate* (e.g., 3% per year), use log-linear trend. If it grows by a *constant dollar amount*, use linear trend.

Prediction Intervals

A point forecast without uncertainty bounds is an incomplete forecast.

$$\hat{y}_{T+h|T} \pm t_{0.025, T-k-1} \cdot \hat{\sigma}_e \sqrt{\underbrace{1}_{\text{irreducible error}} + \underbrace{\mathbf{x}'_{T+h}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{T+h}}_{\text{estimation uncertainty}}}$$

where $\hat{\sigma}_e^2 = \frac{1}{T-k-1} \sum_{t=1}^T e_t^2$ is the estimated residual variance.

Two sources of uncertainty in the PI:

- **Irreducible:** future shock ε_{T+h} has variance σ^2 (the “1” under the square root)
- **Estimation:** $\hat{\beta} \neq \beta$ (the $(X'X)^{-1}$ term; shrinks as $T \rightarrow \infty$)

PIs widen for predictions far from the center of the training data — including long-horizon forecasts where $t = T + h$ is far from \bar{t} .

The standard PI formula assumes:

1. $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$ (normality + homoscedasticity)
2. No autocorrelation: $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$
3. No model misspecification
4. Predictors \mathbf{x}_{T+h} known exactly

Violation of any assumption **invalidates** the nominal coverage. Always check: residual histogram, ACF of residuals, Breusch-Pagan test for heteroscedasticity. (Lecture 6 covers formal tests.)

Empirical rule of thumb: If ACF of residuals shows spikes at short lags (1–4), the PI is too narrow. Residual autocorrelation is exploitable information — switch to an AR or ARIMA model.

Autoregressive Models

The past values of a series often predict its future. AR models formalize this idea.

An **autoregressive model of order 1** is just OLS regression with one lag of y as the predictor:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} (0, \sigma^2) \text{ (normality not required for OLS)}$$

Stationarity condition: $|\phi_1| < 1$. If $\phi_1 = 1$, the model is a *random walk* (non-stationary, covered in Lecture 4).

One-step forecast:

$$\hat{y}_{T+1|T} = \hat{\phi}_0 + \hat{\phi}_1 y_T$$

As $h \rightarrow \infty$, multi-step AR(1) forecasts converge to the unconditional mean $\mu = \phi_0 / (1 - \phi_1)$ when $|\phi_1| < 1$.

Two-step forecast (recursive):

$$\hat{y}_{T+2|T} = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{T+1|T}$$

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t$$

Lag order selection: choose p to minimize an information criterion.

$$\text{AIC} = -2\hat{\ell} + 2(p + 2)$$

$$\text{BIC} = -2\hat{\ell} + (p + 2) \ln T$$

$\hat{\ell}$: maximized log-likelihood. $p + 2$ parameters: p AR coeffs + intercept + σ^2 .

BIC selects $p = 2$ but the residual ACF shows a significant spike at lag 12. What should you do?

AIC tends to select more lags (asymptotically inconsistent: does not recover the true order with probability 1). BIC penalizes more and is consistent. **Use BIC** for forecasting in most business applications (Box et al. 2015, Ch. 5).

For an AR(p) model, h -step-ahead forecasts use the **recursive substitution** method:

$$\hat{y}_{T+h|T} = \hat{\phi}_0 + \sum_{i=1}^p \hat{\phi}_i \hat{y}_{T+h-i|T}$$

where $\hat{y}_{T+j|T} = y_{T+j}$ for $j \leq 0$ (observed values).

$$\begin{aligned}\hat{y}_{T+1|T} &= \hat{\phi}_0 + \hat{\phi}_1 y_T + \hat{\phi}_2 y_{T-1} \\ \hat{y}_{T+2|T} &= \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{T+1|T} + \hat{\phi}_2 y_T \\ \hat{y}_{T+3|T} &= \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{T+2|T} + \hat{\phi}_2 \hat{y}_{T+1|T}\end{aligned}$$

Forecast uncertainty compounds with horizon: each substitution propagates estimation error from earlier steps.

How do we know if the AR order is adequate?

Residual ACF check:

- Compute residuals $e_t = y_t - \hat{y}_{t|t-1}$
- Plot autocorrelation function (ACF) of residuals
- All spikes should fall inside the 95% bands: $\pm 1.96/\sqrt{T}$
- Significant spike at lag $k \Rightarrow$ add lag k to the model

Residual autocorrelation signals information the model has not yet captured. A model with autocorrelated residuals is **suboptimal** and its prediction intervals are **invalid**. Full treatment: Lecture 4 (ARIMA).

Lab 02 plots the residual ACF for the AR(p) model and checks whether any spikes suggest a higher order is needed.

Pitfalls

Regression in time series requires extra care. Three failure modes to avoid.

Granger and Newbold (1974): Regressing two *independent* random-walk series y_t and x_t yields $R^2 \approx 0.7$ and apparently significant t -statistics — despite no true relationship between them.

Why this happens:

- Both y_t and x_t are $I(1)$ (non-stationary; require one difference to become stationary): their levels share a stochastic trend
- The usual t -distribution critical values are *invalid* under non-stationarity

Diagnostic heuristic: $R^2 > DW$ statistic

suggests spurious regression (Granger and Newbold 1974).

Fix: Difference the series (Δy_t , Δx_t) before regressing. Or test for cointegration (Lecture 5).

More predictors \Rightarrow better in-sample fit, but often worse out-of-sample forecast.

What happens:

- OLS fits noise along with signal
- High R^2 in-sample; large errors out-of-sample
- Worse with small T , large k
- Many lags: AR(20) on monthly data likely overfits

Rule of thumb: Keep $k \ll T$. A common bound: $k \leq T/10$.

A 12-month history and 11 seasonal dummies + intercept = 12 parameters. You are fitting the data, not the underlying pattern.

Regularization (LASSO, Ridge) addresses this formally in Lecture 8.

Critical constraint: To forecast y_{T+h} using a regressor x_{T+h} , you must know (or forecast) x_{T+h} at time T . Using data that won't be available at the forecast origin is **data leakage**.

Valid predictors for y_{T+1} :

- ✓ y_T, y_{T-1}, \dots (lagged y)
- ✓ x_{T-1} (lagged leading indicator)
- ✓ Seasonal dummy at $T + 1$
- ✓ Weather forecast for $T + 1$

Invalid predictors for y_{T+1} :

- ✗ x_{T+1} (contemporaneous, unknown)
- ✗ y_{T+1} (the target itself)
- ✗ Future CPI or GDP
- ✗ Any variable with publication lag $< h$

Regression extends benchmarks by incorporating **predictors**: deterministic trends, seasonal dummies, lagged values, and leading indicators.

The OLS forecast is $\hat{y}_{T+h|T} = \mathbf{x}'_{T+h}\hat{\beta}$, but \mathbf{x}_{T+h} must be **known or forecastable** at time T .

Prediction intervals have two components: irreducible error (σ^2) and estimation uncertainty ($((X'X)^{-1}$ term). Valid only under no autocorrelation and homoscedasticity.

AR(p) is OLS with lagged y . Use **BIC** to select p . Multi-step forecasts use recursive substitution.

Three pitfalls: **spurious regression** ($I(1)$ levels), **overfitting** (too many predictors), and the **regressor availability problem** (data leakage).

Next: *Exponential Smoothing (Lecture 3) — adaptive weighting without requiring explicit predictor specification.*

-
- Box, George E. P. et al. (2015). *Time Series Analysis: Forecasting and Control*. 5th. Hoboken, NJ: Wiley.
 - Granger, C. W. J. and Paul Newbold (1974). "Spurious Regressions in Econometrics". In: *Journal of Econometrics* 2.2, pp. 111–120.
 - Hamilton, James D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.