Multivariate Methods

BSAD 8310: Business Forecasting — Lecture 5

Department of Economics

University of Nebraska at Omaha

Spring 2026

Beyond Univariate Forecasting

Sometimes other series carry information about y's future that y's own past cannot provide.

All models in Lectures 1–4 use only y_t's own history:

$$\hat{y}_{T+h|T} = f(y_T, y_{T-1}, \dots)$$

But many series are linked:

- Consumer sentiment today $\rightarrow$ retail spending next month
- Interest rates today $\rightarrow$ housing starts in 3–6 months
- Unemployment $\rightarrow$ consumer credit defaults with a lag

If series x_t contains information about y_{t+h} not already in $y_t, y_{t-1}, \dots$, then using x_t can reduce forecast error. **Granger causality** formalizes this idea.

Socratic: can using x_t ever hurt forecast accuracy? When?

Model	Role of x_t	Use case
VAR(p)	Symmetric: y and x forecast each other	Macro, finance; unknown direction
ARIMAX	Exogenous: x_t drives y_t, not vice versa	Clear causal direction
ECM	Error-correction: long-run equilibrium	Cointegrated I(1) pairs

Supporting tools:

- **Granger causality test:** does x_t help predict y_t?
- **Cross-correlation function (CCF):** at which lag does x_t predict y_{t+k}?
- **Cointegration test:** do two I(1) series share a long-run trend?

Vector Autoregression (VAR)

A VAR treats all variables symmetrically: every variable can predict every other.

For two stationary series y_{1t} and y_{2t}, the VAR(1) is:

$$\begin{cases} y_{1t} = c_1 + a_{11}\, y_{1,t-1} + a_{12}\, y_{2,t-1} + \varepsilon_{1t} \\ y_{2t} = c_2 + a_{21}\, y_{1,t-1} + a_{22}\, y_{2,t-1} + \varepsilon_{2t} \end{cases}$$

1980

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

$\mathbf{y}_t$ is $k \times 1$; each $\mathbf{A}_i$ is $k \times k$; $\boldsymbol{\Sigma}$ is the $k \times k$ error covariance matrix.

Key feature: a_{12} captures the effect of $y_{2,t-1}$ on y_{1t} — the cross-variable coefficient that univariate AR misses.

Estimation: each equation is estimated separately by OLS (equivalent to GLS since all equations share the same regressors).

Order selection:

- Fit VAR(p) for $p = 0, 1, \dots, p_{\max}$
- Select p minimizing AIC or BIC:

$$\mathrm{BIC} = -2\hat{\ell} + \kappa \ln T, \quad \kappa = k^2 p + k \text{ parameters}$$

- BIC preferred (penalizes overparameterization)
- Use *same* p for all equations

Parameter explosion: VAR(p) with k variables has $k^2 p + k$ parameters total.

$k = 5$, $p = 4$: $5^2 \times 4 + 5 = \mathbf{105}$ parameters!

Keep k small (≤ 4) or use LASSO-VAR for large systems.

1969

x_t **Granger-causes** y_t if past values of x_t help predict y_t beyond what y_t's own past can explain.

Test procedure (in a VAR):

1. Fit full VAR (with x-lags) and restricted VAR (without x-lags)
2. F-test: H_0: all $a_{12,\ell} = 0$ ($\ell = 1, \ldots, p$; no Granger causality)

Lab 5 example: Does consumer sentiment (UMCSENT) Granger-cause retail sales (RSXFS)? H_0: all sentiment lags $= 0$ in the retail equation.

Granger causality $\neq$ structural causality:
tests *predictive content*, not mechanism.

Question: what is the dynamic effect of a one-unit shock to y_{it} on y_{jt} over future horizons $h = 0, 1, \ldots$?

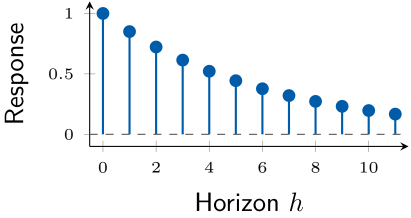
The **impulse response** at horizon h:

$$\mathrm{IRF}_{ji}(h) = \frac{\partial y_{j,T+h}}{\partial \varepsilon_{i,T}}$$

Trace the propagation of a shock through the VAR system.

Orthogonalization: Cholesky decomposition of $\boldsymbol{\Sigma}$ to identify structural shocks; *ordering matters* — place the most exogenous variable first.

IRF: sentiment $\rightarrow$ retail (simulated):



Effect decays geometrically (ρ^h, here $\rho = 0.85$); fades after ≈ 6 periods.

1-step ahead:

$$\hat{\mathbf{y}}_{T+1|T} = \hat{\mathbf{c}} + \hat{\mathbf{A}}_1 \mathbf{y}_T + \cdots + \hat{\mathbf{A}}_p \mathbf{y}_{T-p+1}$$

Multi-step (recursive substitution):

$$\hat{\mathbf{y}}_{T+h|T} = \hat{\mathbf{c}} + \hat{\mathbf{A}}_1 \hat{\mathbf{y}}_{T+h-1|T} + \cdots + \hat{\mathbf{A}}_p \hat{\mathbf{y}}_{T+h-p|T}$$

using $\hat{\mathbf{y}}_{T+k|T} = \mathbf{y}_{T+k}$ for $k \leq 0$.

VAR forecasts are particularly useful at short horizons and when cross-variable dynamics are strong. At long horizons the VAR reverts toward its unconditional mean, matching the theoretical property of stationary systems.

Socratic: if $a_{12} = a_{21} = 0$ in a VAR(1), what model does each equation reduce to?

ARIMAX Models

When the causal direction is clear, augment ARIMA with an exogenous predictor.

Standard OLS forecasting (Lecture 2) assumed independent errors. In practice, residuals are often autocorrelated.

$$y_t = \underbrace{\mathbf{x}_t'\boldsymbol{\beta}}_{\text{regression}} + \underbrace{\eta_t}_{\text{ARIMA error}}, \qquad \eta_t \sim \mathrm{ARIMA}(p,d,q)$$

The error η_t captures the autocorrelation *not* explained by the regressors $\mathbf{x}_t$.

Why not use OLS?

- OLS is unbiased but *inefficient* (ignores error covariance)
- Standard errors are wrong $\Rightarrow$ invalid t-tests and prediction intervals (PIs)
- ARIMAX/SARIMAX jointly estimates $\boldsymbol{\beta}$ and the ARIMA parameters (ϕ, θ) by MLE $\Rightarrow$ correct inference

A **contemporaneous** regressor captures only the same-period effect:

$$y_t = \beta_0 + \beta_1 x_t + \eta_t$$

A **distributed lag** (DL) model captures delayed effects:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \cdots + \beta_{L+1} x_{t-L} + \eta_t$$

Example: advertising spend x_t affects sales y_t with a lagged carryover effect:

- β_1: immediate effect (same month)
- β_2: carryover from last month's ads
- Diminishing returns: $|\beta_1| > |\beta_2| > |\beta_3|$

Cross-correlation function (CCF): $r_k = \mathrm{Corr}(x_t, y_{t+k})$. Bars outside $\pm 1.96/\sqrt{T}$ indicate significant lag k. Select L at the largest significant k; refine by AIC/BIC.

Differencing rule: if both $y_t \sim I(1)$ and $x_t \sim I(1)$, difference *both* before fitting (unless cointegrated — see below).

Python: `statsmodels.tsa.statespace.sarimax.SARIMAX` accepts an `exog` argument for exogenous regressors.

```
SARIMAX(y, exog=X, order=(p,d,q), seasonal_order=(P,D,Q,m)).fit()
```

Model selection workflow:

1. Unit-root test y_t and x_t separately $\rightarrow$ choose d
2. Plot CCF of $(\Delta^d y_t)$ vs. $(\Delta^d x_t)$ $\rightarrow$ choose lag L
3. Identify ARIMA(p, d, q) for residuals via ACF/PACF
4. Select (p, q) by AIC; check residuals with Ljung-Box

Socratic: if x_t is I(1) but y_t is I(0), should you difference x_t before including it as a regressor?

Feature	VAR	ARIMAX
Direction of causality	Unknown / symmetric	Known (unidirectional)
Number of variables	Any k (keep $k \leq 4$)	1 outcome + regressors
Feedback loops	Allowed	Not modeled
Interpretability	IRF, FEVD†	Regression coefficients
Forecasting x_t	Joint (endogenous)	Needs external forecast
Preferred when...	Macro / finance; symmetric	Clear cause $\rightarrow$ effect

† *FEVD = forecast error variance decomposition: fraction of y_j's forecast variance explained by shocks to y_i at each horizon.*

ARIMAX treats x_t as **exogenous**: it assumes x_t does not respond to past y_t. If feedback exists, use VAR instead — otherwise forecasts of y_t will require an external forecast of x_t.

Cointegration and Error Correction

Two non-stationary series that share a long-run trend can be modeled together without differencing away their relationship.

Recall from Lecture 2, Granger and Newbold (1974): regressing independent random walks yields spuriously high R^2. But some I(1) pairs are *genuinely* linked.

Spurious: independent RWs diverge over time

Cointegrated: common trend, $y_t - \beta x_t \sim I(0)$

Engle–Granger test (Hamilton 1994, Ch. 19):

1. Regress y_t on x_t (levels)
2. Residuals $\hat{e}_t = y_t - \hat{\beta} x_t$
3. Augmented Dickey-Fuller (ADF) on $\hat{e}_t$: Reject unit root $\Rightarrow$ cointegrated

Example: income & consumption — both I(1), but $c_t - \beta y_t^d \sim I(0)$ (consumption tracks income in the long run).

If y_t and x_t are cointegrated with equilibrium $y_t = \beta x_t + u_t$ $(u_t \sim I(0))$:

$$\Delta y_t = \underbrace{\alpha(y_{t-1} - \beta x_{t-1})}_{\text{error correction}} + \underbrace{\gamma_1 \Delta y_{t-1} + \delta_1 \Delta x_{t-1} + \cdots}_{\text{short-run dynamics}} + \varepsilon_t$$

$\alpha < 0$: speed of adjustment; γ_1: own short-run lag; δ_1: cross short-run lag. If y_{t-1} is *above* equilibrium, the ECM term pushes $\Delta y_t < 0$ (mean-reverting).

- ECM captures *both* short-run dynamics and long-run equilibrium in one model
- Differenced variables $\Rightarrow$ no spurious regression
- Error correction term $\Rightarrow$ no loss of long-run info

Do not difference cointegrated series before regression — this discards the long-run relationship.

Example: $\hat{\alpha} = -0.3$, $y_{t-1} - \hat{\beta} x_{t-1} = +2.0$ *(10% above equilibrium)* $\Rightarrow$ *ECM contributes* $-0.3 \times 2.0 = -0.6$ *to* Δy_t *(partial correction).*

Granger causality: x_t Granger-causes y_t if x_t's past improves forecasts of y_t beyond y_t's own past.

VAR: symmetric multivariate model estimated equation-by-equation; use IRF to trace shock propagation; keep k small.

ARIMAX: augment ARIMA with exogenous regressors when causal direction is clear; use SARIMAX in statsmodels.

Cointegration: two I(1) series with a stable long-run relationship — model as ECM, not as levels or differences alone.

Model choice: symmetric dynamics $\rightarrow$ VAR; clear cause $\rightarrow$ ARIMAX; long-run equilibrium $\rightarrow$ ECM.

VAR(1) with $a_{12} = a_{21} = 0$ collapses to two independent AR(1) models — the multivariate framework nests the univariate special case.

We now have five model families:

- Benchmarks (naïve, seasonal naïve, mean, drift)
- Regression and AR models (Lecture 2)
- Exponential smoothing / ETS (Lecture 3)
- ARIMA / SARIMA (Lecture 4)
- VAR, ARIMAX, ECM (Lecture 5)

Key question: how do we rigorously compare these models out-of-sample?

Lecture 6: Forecast evaluation — walk-forward validation, the Diebold-Mariano test (Diebold and Mariano 1995), and forecast combination.

Lab 5: Granger causality tests, VAR fitting and IRF, ARIMAX on RSXFS + consumer sentiment.

- Diebold, Francis X. and Roberto S. Mariano (1995). "Comparing Predictive Accuracy". In: *Journal of Business & Economic Statistics* 13.3, pp. 253–263.
- Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods". In: *Econometrica* 37.3, pp. 424–438.
- 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.
- Sims, Christopher A. (1980). "Macroeconomics and Reality". In: *Econometrica* 48.1, pp. 1–48.