DDBI Lecture 3: Multivariate Time Series

Relationships and Forecasting

Dr. Stavros K. Stavroglou

December 23, 2024

Outline

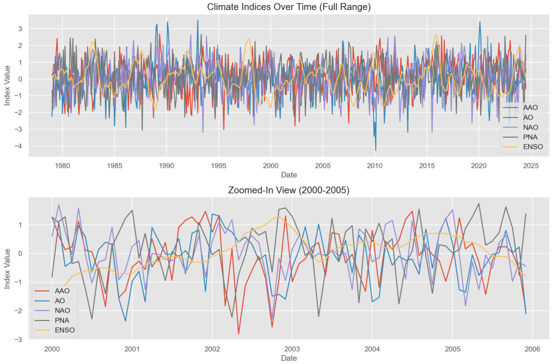
Section 1 Goals

After this section, you should be able to:

- Load a multivariate climate dataset (`combined_climate_indices_2024.csv`)
- Visualize and interpret the basic characteristics of **multiple** time series
- Understand the concept of **correlation** and discuss its pros & cons
- Assess linear relationships among **ENSO**, **AAO**, **AO**, **NAO**, and **PNA**

Dataset Description

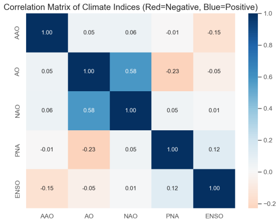
- **Columns:** `AAO`, `AO`, `NAO`, `PNA`, `ENSO`
- **Monthly frequency** data (missing values interpolated)
- Each index measures different atmospheric or oceanic oscillation



(All climate indices over time.)

Correlation Matrix

- **Correlation:** linear association between pairs of variables
- **Pros:**
  - Quick and interpretable measure
  - Indicates potential predictive relationships
- **Cons:**
  - Only captures linear relationships
  - Does not imply causation



(Heatmap of correlation matrix.)

What is Granger Causality?

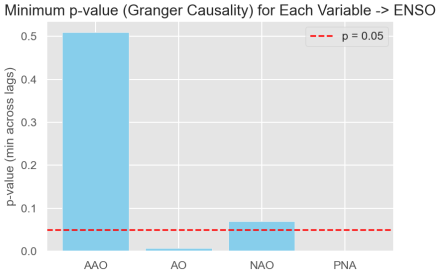
Definition: A variable X *Granger-causes* Y if *past values* of X help predict future values of Y in a linear sense.

Key Points:

- Incorporates **time-lagged** effects
- Statistical test (F-test) to assess improvement in predictability
- Does not necessarily imply true physical causation
- Proper lag selection (e.g., 1 to 6 months) can change conclusions

Practical Interpretation

- If AO Granger-causes ENSO, then including $\mathrm{AO}_{t-1}, \mathrm{AO}_{t-2}, \ldots$ improves the forecast of ENSO.
- If no causality is found, it may be that:
  - The relationship is non-linear and not captured in a linear test.
  - The chosen lags are insufficient.
- **Stationarity** often assumed; differencing or transformations might be needed.



(Granger causality interpretation diagram.)

Why Transfer Entropy?

Transfer Entropy (TE):

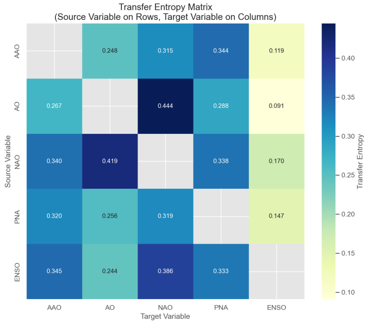
- A **non-linear** measure of directional information flow
- Captures how knowledge of X reduces uncertainty about future Y
- Great for complex or non-linear relationships

Challenges:

- More computationally intensive
- Requires careful binning or kernels for probability estimates
- Interpretation less straightforward than correlation or Granger

Practical Insights from TE

- **High** TE from $X \to Y$ suggests strong directional influence of X on Y
- TE can be **asymmetric**: $X \to Y$ might differ from $Y \to X$
- Especially useful in climate applications where non-linear feedback loops exist



(Possible TE matrix for climate indices.)

Linear Regression Setup

Goal: Predict ENSO using other indices as regressors:

$$\widehat{\text{ENSO}}_t = \beta_0 + \beta_1(\text{AAO}_t) + \beta_2(\text{AO}_t) + \beta_3(\text{NAO}_t) + \beta_4(\text{PNA}_t) + \epsilon_t.$$

Procedure:

- Train/test split (e.g. 80% training, 20% testing)
- Fit model on training set
- Forecast on test set
- Compare predictions vs. actuals

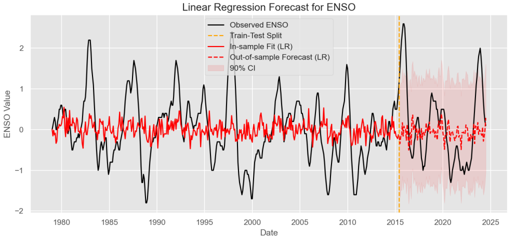
Advantages & Disadvantages of Linear Regression

Pros:

- Simple & easy to interpret
- Quick to implement
- Acts as a good baseline

Cons:

- Only *linear* relationships
- No automatic handling of **lags** unless explicitly included
- Potentially misses time-dependent patterns



(LR forecast for ENSO, with train/test split.)

VAR: Why Use It?

Vector Autoregression:

- Considers **all** variables in a system *together*
- Each variable regressed on lagged values of *all* variables
- Captures potential feedback loops (e.g. ENSO $\leftrightarrow$ AO)

Pros:

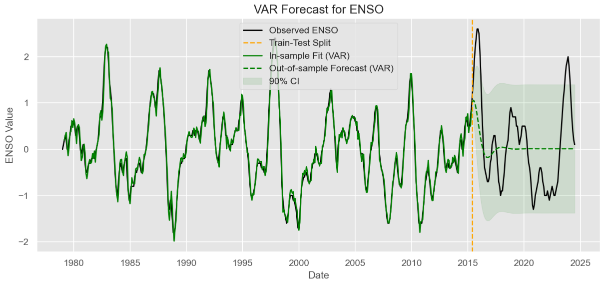
- Can improve forecasts if cross-variable information is valuable
- More holistic approach for multi-series data

Cons:

- **Linear** model; might miss non-linear effects
- Potentially large parameter space (p lags $\times$ number of variables)
- Assumes stationarity (data may need differencing/transformations)

VAR Forecast for ENSO

- **Steps:**
  - Fit a VAR model on training portion of the entire dataset
  - Select an appropriate lag order (using AIC/BIC)
  - Forecast multiple steps ahead into the test period
- **Compare** the VAR-generated ENSO forecast to actual ENSO

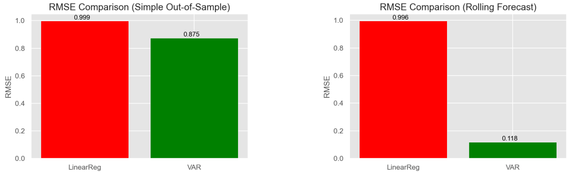


(VAR-based ENSO forecast in-sample vs. out-of-sample.)

Forecast Accuracy Metrics

Common Metrics:

- **MAE** (Mean Absolute Error)
- **MSE** (Mean Squared Error)
- **RMSE** (Root MSE)
- **MAPE** (Mean Absolute Percentage Error)
- **MASE** (Mean Absolute Scaled Error)
- **Max Error** (largest single deviation)



(Tables comparing LR vs. VAR, simple vs. rolling.)

Rolling Forecasts: A More Realistic Evaluation

Rolling Process:

1. Fit on the training set
2. Forecast *one-step* ahead
3. **Incorporate** the new actual into your training data
4. Re-fit model, forecast the next step
5. Continue until the test set is fully forecast

Pros:

- Simulates real-time updates
- Shows how model evolves as new data arrives

Cons:

- Re-fitting can be expensive (especially VAR)

Key Takeaways

- **Correlation, Granger Causality, Transfer Entropy** each reveal different aspects of multivariate relationships
- **Linear Regression & VAR** are baseline forecasting approaches for multiple time series
- **Rolling forecasts** provide a more realistic performance measure
- In complex real-world applications (e.g. climate), **non-linear** or machine learning models may be beneficial

Next Steps

- **Practical Lab Session:**
  - Explore the dataset and generate the visuals shown here (time-series plots, correlation heatmap, etc.)
  - Conduct Granger tests and Transfer Entropy (with appropriate libraries or custom scripts)
  - Implement **Linear Regression** and **VAR** forecasting, compare with simple vs. rolling approaches
- **Further Investigation:**
  - Stationarity checks, transformations
  - Non-linear models or ML-based models for more complex interactions
  - Multi-step ahead forecasting and cross-validation

Thank you!