Econometrics: Time Series

James Aketch Economics BSc.
University College Dublin

December 2, 2025

Econometrics: Time Series

Difference Operator

The difference operator measures how a variable changes from one period to the next. It helps remove trending from a time series and helps achieve **stationarity** in ARIMA models.

The first difference operator is denoted by $(1 - L)$, where L is the lag operator discussed later. Basically, when gets applied to a variable y_t, it subtracts the previous value of the series from the current/present one in lay speak its the **first difference** that shows how much something went up or down since last time.

$$(1 - L)y_t = y_t - y_{t-1}$$

If the data still shows a trend after the first difference, try and apply the operator again to get the second difference. This shows whether the rate of change is **speeding up or slowing down**.

$$(1 - L)^2 y_t = (1 - 2L + L^2)y_t = y_t - 2y_{t-1} + y_{t-2}$$

In the general case, applying the operator d times is as simple as:

$$(1 - L)^d y_t$$

0.0.1 Definitions:

- $(1-L)$: first difference operator shows the change in y_t from one period to the next.

- $(1-L)^2$: second difference operator, measures how the rate of change changes itself over time.

Lag Operator Notation

Earlier, I discussed the lag operator briefly; however, now I will get into the weeds of the notation. Essentially, it is a shorthand way of showing how past values of a time series relate to their present value. Thankfully, it tries to shorten lengthy autoregressive and moving average equations by using exponents and the operator L, instead of repeatedly writing lagged terms.

For example, instead of writing:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \varepsilon_t$$

we can write the same $AR(4)$ process more compactly as:

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4) y_t = \alpha + \varepsilon_t$$

The lag operator, L, takes a time series variable one period back in time:

$$Ly_t = y_{t-1}$$

Again. Applying it twice moves the variable two periods back:

$$L^2 y_t = y_{t-2}$$

In general, applying the operator n times gives the n-th lag:

$$L^n y_t = y_{t-n}$$

The lag operator makes it easier to represent and combine AR and MA models, particularly when making ARIMA and VAR models, as it groups lags into simple polynomial functions of L.

0.0.2 Definitions:

- L: lag operator, shifts a variable one time period back, such that $Ly_t = y_{t-1}$.

- L^n: the n-th power of the lag operator, shifts the variable n periods back $(L^n y_t = y_{t-n})$.

- $(1 - L)$: the first difference operator, showing the change from one period to the next $(y_t - y_{t-1})$.

- $(1 - L)^d$: differencing operator of order d, used to make a series stationary by removing trends.

Autoregressive (AR) Process

An autoregressive process (AR) models how the current value of a variable depends on its own past values and a random disturbance (white noise). It captures how past fluctuations influence the present. Note: An AR process can extend infinitely into the future.

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

0.0.3 Definitions:

- y_t: dependent variable, the current value of the series being explained (e.g., GDP, price, or demand).

- α: Alpha is the intercept/constant term, representing the baseline level when the past effects are zero.

- ϕ_i: The variable phi denotes the autoregressive parameters $(i = 1, \ldots, p)$, showing how much the past values influence the current one.

- y_{t-i} The lagged dependent variables, *earlier values used to predict the present values.*

- p: is the number of lags in the process (order of the autoregressive process).

- ε_t: is the error term or stochastic disturbance that represents random shocks that are not explained by past data.

0.0.4 In lag operator form:

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)y_t = \varepsilon_t$$

0.1 Understanding AR(1), AR(2), AR(3) and AR(p) Numbering

An AR(1) process means the current value of the series is dependent only on its value in the previous period. e.g, if we were looking at GDP, this year's GDP would depend on last year's GDP and some random change (the error term).

Generally, an AR(p) process means the model includes p a number of lagged terms i.e. It looks back p time-steps into the past. Each lag shows how much influence some earlier values have on the current one. The higher the order p, the more memory the model has about the past like a computer with more RAM

Moving Average (MA) Process

A moving average (MA) process explains how the current value of a time series depends on past random shocks (errors) and the average level of the series. It represents how previous disturbances continue to influence the present.

Basically, how the current value of a time series depends on past random shocks. It serves like a memory of past disturbances in the data. An even simpler explanation is it looks at past shocks to the present with some noise in the mix.

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

0.1.1 Definitions:

- y_t: is the dependent variable, the current observed value of the series (e.g., GDP, price, or demand).

- μ: mean or constant term, representing the average level of the process.

- θ_j: moving average (MA) parameters for lagged error terms ($j = 1, \ldots, q$), essentially, it shows how the magnitude of past shocks influences the present value.

- ε_{t-j}: represents the lagged error terms; simply put, it represents the shocks from earlier periods that continue to affect the present.

- q: shows the order of the moving average process, basically, it indicates how many past shocks are included.

- ε_t: white-noise error term, the new random disturbance in the current period.

In lag operator form:

$$y_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t$$

ARIMA Model

By combining AR and MA models, it gives us the ARIMA(p, d, q) model, where differencing of order d is applied to achieve **stationarity.**

$$\Delta^d y_t = \beta_0 + \beta_1 \Delta^d y_{t-1} + \cdots + \beta_p \Delta^d y_{t-p} + u_t + \gamma_1 u_{t-1} + \gamma_2 u_{t-2} + \cdots + \gamma_q u_{t-q}$$

Using lag operator notation:

$$\Phi(L)(1 - L)^d y_t = \Theta(L) u_t$$

where

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p, \quad \Theta(L) = 1 + \gamma_1 L + \gamma_2 L^2 + \cdots + \gamma_q L^q$$

0.1.2 Definitions:

β_i: autoregressive parameters after differencing, showing how much past values influence the current one.

- Δ^d: differencing operator of order d, used to make the time series stationary.

- y_t: observed time series at time t.

- β_0: constant term or intercept.

- u_t: white noise error term.

- γ_j: moving average (MA) parameters for lagged error terms ($j = 1, \ldots, q$); they represent the magnitude and direction of how past shocks influence the current value of y_t.

- p: number of autoregressive lags.

- d: number of differences applied to the series.

- q: number of moving average lags.

Concepts

Lags: $L^n y_t = y_{t-n}$ determine how far back we look to see if the process is affected by past values.

Granger Causality: If including lagged values of X improves the prediction of Y beyond what Y's own lags can do, then X is said to Granger-cause Y. Basically, if there are smaller forecast errors from adding 's lags in addition to or with 's own lone lags can accomplish its ***Granger causal***

Wald Test: Tests whether parameters are equal to specific values (often zero); used for both linear and non-linear models.

F-Test: Checks whether two or more variables jointly improve the model's explanatory power.

Impulse Response Function (IRF): Traces the effect of a one-time shock on current and future values of the variables in a dynamic system. Basically, it is like a path of how earlier shocks applied to the system. Did every variable within the system react over the next period etc.

$$y_t = \varphi y_{t-1} + \varepsilon_t, \quad IRF(0) = 1, \ IRF(1) = \varphi, \ IRF(2) = \varphi^2, \ \ldots, \ IRF(h) = \varphi^h$$

Autocorrelation Measures how much a series is correlated with its own past values over time ***Autocorrelation is like it insists upon itself over time...***

1 Stationarity:

Simply put, it implies that the series *trends* around some fixed average value(s) by definition, the process oscillates around a fixed mean μ, with constant variance $\text{Var}(y_t)$ and auto-covariance that depends only on lag k, not on time t.

Again, it depends on K and not actually on time t; e.g., how far back you look but not the time itself which makes the covariates consistent over time

$$|\phi| < 1 \Rightarrow \text{stationary process}, \quad |\phi| = 1 \Rightarrow \text{unit root (stochastic) process.}$$

The unit root process is essentially a random walk of shocks that accumulate forever the new value is the previous value with an added shock . *Note: If differencing makes a process stationary, it has a unit root (e.g., a random walk). ADF and KPSS tests are common tests used to check for stationarity.*

VAR and SVAR Models

A Vector Autoregression (VAR) is a system of (AR) autoregressive equations where each variable depends on its own past and the pasts (s) of others.

Again, basically, it shows how a range of variables see some or any kind of effect by assessing how each variable responds to past movements within itself and the others.

Kind of like a domino effect within the system

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \varepsilon_t$$

Vector Autoregression (VAR) Process

A Vector Autoregression (VAR) is a system of autoregressive equations that models how several variables move together over time. Each variable depends not only on its own past values but also on the past values of all other variables in the system. *This method is commonly used in macroeconometrics to study, for example, the effects of interest rates on output, inflation, or unemployment*

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \varepsilon_t$$

Matrix Form:

$$
\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Kt} \end{bmatrix}
=
\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{bmatrix}
+
\begin{bmatrix}
\phi_{11,1} & \phi_{12,1} & \cdots & \phi_{1K,1} \\
\phi_{21,1} & \phi_{22,1} & \cdots & \phi_{2K,1} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{K1,1} & \phi_{K2,1} & \cdots & \phi_{KK,1}
\end{bmatrix}
\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{K,t-1} \end{bmatrix}
+ \cdots +
\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{Kt} \end{bmatrix}
$$

1.0.1 Definitions:

- $\mathbf{y}_t$: $K \times 1$ vector of endogenous (inside) variables at time t , i.e, variables whose value(s) are determined by the model itself; they have a dependency on each other. (e.g., GDP, inflation, and interest rate).

- $\mathbf{c}$: $K \times 1$ vector of intercepts, like α for each equation.

- Φ_i: $K \times K$ coefficient matrices for the lag i $(i = 1, \ldots, p)$, show how a variable responds to the past values of all variables.

- p: order of the VAR process, indicating how many lags of each variable are included.

- ε_t: $K \times 1$ vector of error terms, assumed to be white noise again with mean zero and covariance matrix Σ_ε.

1.0.2 Further Explanation

Each row of the VAR represents one equation, where a variable (e.g., y_{1t}) is explained by its own lags and the lagged values of the other variables. This allows the model to capture how shocks in one part of the system can cascade through to others.

Structural Vector Autoregression (SVAR)

A **Structural VAR (SVAR)** model builds on the regular Vector Autoregression (VAR) . However. It applies additional restrictions to separate structural shocks using economic conventions

For instance. In a regular VAR, we can observe how variables move together, But we cannot tell which variable caused what?

Making it seem rather dismissible, being as we are looking at statistical data that can be interpreted *incorrectly*. However, the SVAR reduces any conjecture using the reduced-form VAR:

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \mathbf{u}_t$$

Here, $\mathbf{u}_t$ represents the residuals or *shocks*, which are correlated with each other.

The structural form of the model then goes to:

$$A_0 \mathbf{y}_t = \mathbf{a} + A_1 \mathbf{y}_{t-1} + A_2 \mathbf{y}_{t-2} + \cdots + A_p \mathbf{y}_{t-p} + \varepsilon_t$$

1.0.3 Definitions

- A_0: contains *weird* relationships among variables.

- ε_t Structural shocks (for example, a monetary policy shock or a supply shock) assumed to be independent.

The relationship between the reduced-form errors ($\mathbf{u}_t$) and the structural shocks (ε_t) is written as:
$$\mathbf{u}_t = A_0^{-1}\varepsilon_t$$
Since A_0 isn't directly observed, we can then use **identifying restrictions** These restrictions are derived from economic theory for example:

- *Certain variables might not respond immediately to others (short-run restrictions).*

- *Some effects may appear only after several periods (long-run restrictions).*

Once identified, the SVAR allows us to trace how a shock to one structural variable affects all others over time through the earlier **Impulse Response Functions (IRFs)** and **Variance Decompositions**.

1.0.4 Explanation :

The SVAR is essentially like taking the VAR model to solely capture the economic forces that are actually behind it e.g. It tells not just *what happened,* but *why it happened?*.

Co-integration and the Vector Error Correction Model (VECM)

Sometimes, two or more non-stationary $I(1)$ series may move together over time in such a way that a particular linear combination of them becomes stationary.

A **linear combination** means taking two or more non-stationary series and combining them using constants, for example $z_t = y_t - \theta x_t$. #

Even if y_t x_t both wander around over time, their combination z_t may actually be stationary, meaning it fluctuates around a constant mean. This

relationship is called **co-integration**. it implies there exists a long-run equilibrium relationship between the variables even if they drift over time on its own.

For example, if both y_t and x_t are $I(1)$ but $(y_t - \theta x_t)$ is in fact stationary, then the variables are deemed co-integrated.

2 The Engles–Granger Two-Step Approach:

1. Estimate the long-run equilibrium:

$$y_t = \alpha + \theta x_t + u_t$$

2. Test the residuals u_t for stationarity (using the ADF test). If u_t is stationary, y_t and x_t are co-integrated.

When co-integration is present, short-term and long-term dynamics can be captured using the ECM.

2.1 Error Correction Model (ECM)

This model shows how two or many variables that are linked in the long run adjust to short-term changes. In economic speak, *It connects the short-run of a relationship with its long-run equilibrium.*

In simplistic terms, the ECM says that even if two variables drift apart in the short run, They will gradually move back toward their long-run balance over time.

For example, if consumption and income are linked in the long run, a sudden increase in income might cause consumption to rise more slowly at first, But the model predicts that consumption will eventually "catch up" to income.

This sample model has two parts:

$$\Delta y_t = \alpha + \beta \Delta x_t + \lambda(y_{t-1} - \theta x_{t-1}) + \varepsilon_t$$

2.1.1 Definitions:

- Δy_t and Δx_t: show short-term "delta" changes.

- $(y_{t-1} - \theta x_{t-1})$: shows how far the system was from its long-run balance in the last period.

- λ: is the **speed of adjustment**, showing how quickly the system corrects any imbalance.

  If λ is negative and significant, it means the variables adjust back toward equilibrium over time.

Basically, the VECM captures short-run adjustments in $\Delta \mathbf{y}_t$ while making sure that the variables return to the long-run equilibrium.

3 Trend and Seasonality

Within time series, there are often systematic movements over time, known as **trends** and **seasonality**.

3.0.1 Trend:

A trend is a long-term upward or downward movement in the data.

- ***Deterministic Trend:*** follows a fixed path over time, e.g., $y_t = \alpha + \beta t + \varepsilon_t$.

- ***Stochastic Trend:*** driven by random shocks, e.g., a random walk process $y_t = y_{t-1} + \varepsilon_t$.

This separates the econometric boys from the time-series econometric **men**

3.0.2 Seasonality:

"They say people in your life are seasons and anything that happens is for a reason "– Kanye "Ye" West on Heard em Say

Often, we see repeating patterns that occur at regular intervals (monthly, quarterly, etc.). To remove seasonal effects, we apply a new **seasonal differencing operator**:

$$(1 - L^s)y_t = y_t - y_{t-s}$$

where s is the seasonal period (for example, $s = 12$ for monthly data).

De-trending and de-seasonalising a series helps achieve stationarity before estimating ARIMA or VAR models.

Forecasting

Forecasting uses a fitted model to predict future values of a series.

A one-step-ahead forecast for an AR(1) process is:

$$\hat{y}_{t+1|t} = \alpha + \phi_1 y_t$$

*The **forecast error** is the difference between the actual and predicted values:*

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1|t}$$

3.0.3 Evaluation of Forecasts:

We use the earlier tools of :

- Mean Squared Error (MSE): $\text{MSE} = \frac{1}{n} \sum e_t^2$

- Root Mean Squared Error (RMSE): $\text{RMSE} = \sqrt{\text{MSE}}$ But we introduce:

- Mean Absolute Error (MAE): $\text{MAE} = \frac{1}{n} \sum |e_t|$

Forecast Accuracy: When it comes to forecasting lower MSE, RMSE, and MAE values indicate more accurate forecasts.

Forecast Error Variance Decomposition (FEVD) in VAR models it measures how much of the forecast error var() of one variable is explained by shocks to itself and to others.

State-Space Models and the Kalman Filter

Now we introduce a *new* A state-space model which expresses a time series in terms of two equations: a *state (transition) equation* that describes how the unobserved state evolves over time, and a *measurement equation* that links the observed data to that state.
General Form:

$$\text{State equation:} \quad \mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t$$

$$\text{Measurement equation:} \quad \mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t$$

3.0.4 definition :

- $\mathbf{x}_t$: unobserved (latent) state vector.

- $\mathbf{y}_t$: observed data.

- $\mathbf{F}$: transition matrix describing how states evolve.

- $\mathbf{H}$: observation matrix linking states to observations.

- $\mathbf{w}_t$ and $\mathbf{v}_t$: process and measurement noise, usually assumed TO BE AGAIN! white noise.

The **Kalman Filter** is a recursive algorithm that estimates the unobserved states $(\mathbf{x}_t)$ over time as new observations become available.

3.1 Kalman Filter 3-Steps:

1. **Prediction:** Use the previous state estimate to predict the current state.

2. **Update:** adjust the prediction using the new observed data.

3. **Iteration:** repeat for each time step to produce updated state estimates.

3.2 Application(s) of Green's Function for IRFs

3.3 Brownian Motions and the Random Walk