

# Introduction to Forecasting

## Workshop on Gender Statistics and Analysis Module II – Topic 3

Providenciales, Turks and Caicos Islands  
16-19 February 2026



UNITED NATIONS

ECLAC

UNITED NATIONS ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN, SUBREGIONAL HEADQUARTERS FOR THE CARIBBEAN

# The topics covered are:

The purpose of forecasting

Time-series regressions vs ARIMA models



# **The purpose of forecasting**

# Forecasting allows for:

## Actionable Business/Policy Intelligence

Allows organizations to make sense of historical patterns and make decisions

## Quantifying Uncertainty

Rather than providing single point forecast, trend analysis helps estimate short-to-medium-term future behaviour from past performance, helping to mitigate risks

## Planning

Regression-based forecasting analyses provide a baseline for planning, a key area of public policy



# Past behaviour helps identify future trends

## But statistical tools are not a crystal ball

- Statistical inference methods, such as regression models, to identify trends continuing in the short- to medium-term
- But things can change
  - New shocks can occur
- Statistical method cannot see the future



UNITED NATIONS

ECLAC

# **Regression vs. ARIMA for Forecasting**

# Linear Time-Series Regression

Y=Female mortality rate, X=Fertility rate controlled by urbanization rate(1960-2024)

## Model Summary - Mortality rate, adult, female (per 1,000 female adults)

| Model          | R     | R <sup>2</sup> | Adjusted R <sup>2</sup> | RMSE  | Durbin-Watson   |           |        |
|----------------|-------|----------------|-------------------------|-------|-----------------|-----------|--------|
|                |       |                |                         |       | Autocorrelation | Statistic | p      |
| M <sub>1</sub> | 0.988 | 0.977          | 0.976                   | 9.490 | 0.968           | 0.060     | < .001 |

Note. M<sub>1</sub> includes Fertility rate, total (births per woman), Urban population (% of total population)

## Coefficients

| Model          |   | Unstandardized | Standard Error | Standardized | t      | p      |
|----------------|---|----------------|----------------|--------------|--------|--------|
| M <sub>1</sub> | (Intercept)                                 | 227.503        | 54.391         |              | 4.183  | < .001 |
|                | Fertility rate, total<br>(births per woman) | 19.252         | 5.777          | 0.457        | 3.333  | .001   |
|                | Urban population (%<br>of total population) | -1.894         | 0.486          | -0.534       | -3.895 | < .001 |



|                           | Time-Series Linear Regression   | ARIMA<br>(Autoregressive Integrated Moving Average)  |
|---------------------------|---|--|
| <b>Core Difference</b>    | Uses external predictor variables<br>Models how other factors influence the outcome and can work with smaller datasets  | Uses only the past values of the <i>same</i> variable<br>Models internal time dependence (autocorrelation).<br>However, it works best with larger data sets            |
| <b>Model</b>              | $Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + \varepsilon_t$   | $Y_t = f(Y_{t-1}, Y_{t-2}, \text{past errors})$  |
| <b>When to use which?</b> | <ul style="list-style-type: none"> <li>• Clear causal drivers exist</li> <li>• Scenario-based forecasting is needed</li> <li>• Policy variables influence outcomes</li> </ul> | <ul style="list-style-type: none"> <li>• Strong autocorrelation</li> <li>• No reliable external predictors</li> <li>• Pure time-series dynamics matter most</li> </ul> |

*ARIMA is essentially a regression on lagged values and lagged errors. It is a special case of time-series regression*



# ARIMA Model

Y=Female mortality rate, X=Fertility rate controlled by urbanization rate(1960-2024)

## Coefficients

|   | Estimate | Standard Error | t      | p      | 95% CI  |        |
|---|----------|----------------|--------|--------|---------|--------|
|   |          |                |        |        | Lower   | Upper  |
| Drift                                       | -2.317   | 0.917          | -2.528 | .014   | -4.153  | -0.481 |
| AR(1)                                       | 0.694    | 0.134          | 5.185  | < .001 | 0.426   | 0.962  |
| MA(1)                                       | -0.680   | 0.161          | -4.220 | < .001 | -1.003  | -0.357 |
| MA(2)                                       | 0.466    | 0.152          | 3.059  | .003   | 0.161   | 0.771  |
| Fertility rate, total<br>(births per woman) | 1.265    | 8.270          | 0.153  | .879   | -15.295 | 17.826 |
| Urban population (%<br>of total population) | -0.711   | 0.841          | -0.845 | .401   | -2.394  | 0.973  |

Note. An ARIMA(1, 1, 2) model was fitted.



UNITED NATIONS

ECLAC

# Interpreting outputs of Regression vs. ARIMA models

|                            | Time-Series Linear Regression  | ARIMA<br>(Autoregressive Integrated Moving Average)   |
|----------------------------|--|---|
| <b>Model results</b>       | Model fit is too good (R-squared = 0.98) and statistically very significant ( $p = 0.001$ ) but time-series highly autocorrelated (D-W = 0.06) | Statistically significant ( $p < 0.05$ ). On average, the process decreases by about 2.3 units per time period This suggests a systematic downward movement in the underlying series. Results from AR(1), MA(1) and MA(2) show persistence in the series and shocks carry one to two periods back |
| <b>Predictors</b>          | Fertility rate and urban population are statistically significant and the direction of the coefficients is what one would theoretically expect | Fertility rate and urban population are not statistically significant. There is no evidence that they affect the dependent variable   |
| <b>Big picture summary</b> | The model seems to have strong predictive power, <b>but</b> strong autocorrelation can inflate the statistical significance of the predictors  | The time-series is mainly driven by <b>its own past behaviour</b> , showing strong persistence, short-term shock adjustments and a significant downward drift. The demographic variables do not explain additional variation beyond the ARIMA structure.  |





**Thank you!**

Feel free to ask any questions



UNITED NATIONS

ECLAC