# HW1: Introduction to Deep Forecasting

Luke Shumway - A02268065

13 September 2024

## 1 The Problem

A good example of a problem that can be solved with time-series forecasting is Disease Outbreak/Spread Prediction.

## 2 Why it Fits

**Econometrics:** Disease outbreaks typically follow certain seasonal trends and are influenced by structured variables like population density, vaccination rates, weather, and mobility patterns. Econometrics can model these relationships linearly, making predictions based on historical trends.

**Machine Learning:** ML models can automatically learn patterns from historical outbreaks, capturing complex and non-linear relationships. ML models could incorporate a wider variety of features, such as demographic data, local healthcare capacity, and travel patterns.

**Deep Learning:** DL models excel at capturing long-term dependencies and complex time-based interactions. This is useful when multiple time series (like the number of reported cases, mobility data, and temperature) are combined to model outbreak progression. DL can also integrate unstructured data sources, such as social media chatter, news reports, or public health announcements, to improve predictions.

## 3 Uniqueness

Since disease outbreak data can be both structured and unstructured, both types of data will be reviewed.

**Structured Data:** includes Timestamped disease case counts, population data, vaccination rates, hospitalizations, weather data.

**Unstructured Data:** Social media reports of symptoms or mentions of the disease. News articles or reports from health authorities.

# 4 Econometrics Model

For simplicity, we will only have three variables, the tempurature, vaccination rate, and (ironically enough) airline passengers (which is a stand-in for mobility data at large). The model is then

$$y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 T_t + \beta_3 V_t + \beta_4 P_t + \varepsilon_t$$

where $y_t$ is the number of new cases at time $t$, $T_t$ is the temperature at time $t$, $V_t$ is the vaccination rate at time $t$, $P_t$ is the number of airline passengers at time $t$, and $\varepsilon_t$ is the error term for time $t$.