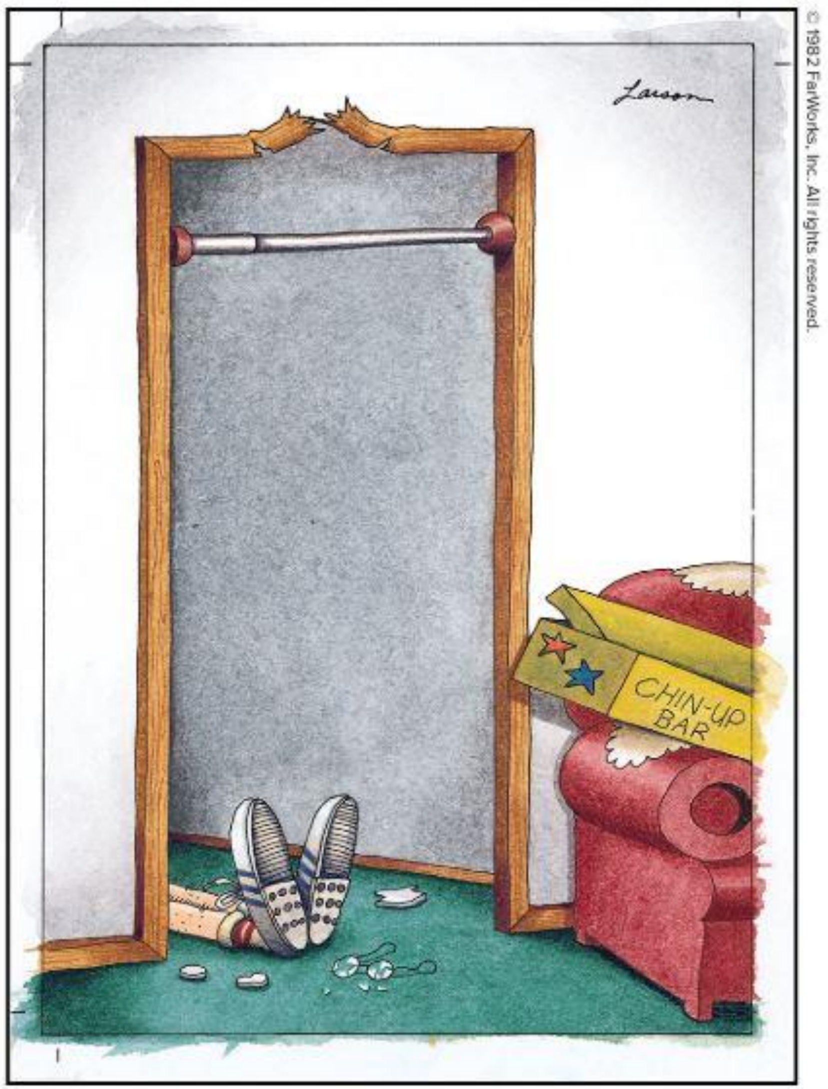
# Econ 312

## Monday, March 16
## Autoregressive Models and Integrated Processes

Reading: Last section of online time-series Chapter 1

# Today’s classic Far Side comic




Side note: Yes, my wife DID decide to install our 50-year-old chin-up bar JUST LAST WEEK. Thefarside.com must be spying on us!

But I assure you that my arm-strength-to-weight ratio is even lower than the stock market, so you needn’t worry about this happening to me.

# Context and preview

- **Last class:** We considered AR($p$) time-series data-generating processes, wrote them using polynomials in the lag operator $L$, and introduced the role of the roots of the autoregressive lag polynomial in determining the stationarity of the AR process

- **Today:** We discuss stationarity in more detail.
  - **Unit roots:** When one or more of the polynomial roots equal(s) one, the process is an **integrated process**
  - The **order of integration** $k$ is the number of unit roots; we denote an integrated process of order $k$ as $I(k)$
  - If we start with an AR process that is $I(k)$ and **difference** it $k$ times, we get a stationary process

# Polynomial roots

- The AR($p$) process is written

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$
$$\phi(L) y_t = \varepsilon_t$$

- We can factor the $p$-order polynomial $\phi(L)$ into

$$\phi(L) = \left(1 - \frac{1}{r_1} L\right)\left(1 - \frac{1}{r_2} L\right) \cdots \left(1 - \frac{1}{r_p} L\right)$$

- $r_1$ through $r_p$ are the $p$ roots of the polynomial

# Roots tell us about stationarity

1. If all of the roots are greater than 1 then $y$ is **stationary**

2. If $k$ roots are equal to 1 (**unit roots**) and the other $p - k$ roots are greater than one, then $y$ is **integrated of order *k*** and $\Delta^k y$ is stationary

3. If any root is less than 1, then $y$ is **non-stationary** and cannot be made stationary by differencing

Note: By greater than or less than 1, we mean in absolute value or, for complex roots, modulus. Roots with modulus greater than one are said to be "outside the unit circle" in the complex plane.

# Integrated processes

- A variable generated by an integrated process can be thought of as the accumulation of the underlying variable that is its difference
- Examples:
  1. Net investment $I_t$ is the change in the capital stock $K_t$ , so $I_t = \Delta K_t$. If $I_t$ is stationary [$I\,(0)$], then $K_t$ is integrated of order one [$I\,(1)$]. The capital stock is the accumulation of all past net investment.
  2. The inflation rate $\pi_t$ is the change in the log of the price level $\ln P_t$, so $\pi_t = \Delta P_t$. If $\pi_t$ is stationary, then $\ln P_t$ is $I\,(1)$. The current price level is the accumulation of all past inflation.
- In calculus, the **integral** is the inverse of the **derivative**; in time-series analysis, **integration** is the inverse of **differencing**

# Random walk as $I(1)$ process

- The **random walk** process is the simplest integrated process

$$y_t = y_{t-1} + \varepsilon_t$$
$$y_t - y_{t-1} = \varepsilon_t$$
$$(1 - L)\, y_t = \Delta y_t = \varepsilon_t$$

- It is a first-order process with one root that is **equal to one**
- The first difference of a random walk is not only stationary, it is white noise

# “Spurious regressions” with integrated variables

- Regressions in which both $y$ and $x$ are integrated processes will tend to look really good (high $t$ statistics and $R^2$ values) even if the underlying variables are unrelated
  - The American League attendance vs. Botswana GDP example in class
- Clive Granger and Paul Newbold (1974) called these “**spurious regressions**”
  - The next slide shows a table from their book, which is Table 4-1 in Chapter 4 of the instructor’s time-series chapters

# Granger and Newbold's Monte Carlo results

**Table 4-1. Granger and Newbold's spurious regression result**

| | Number of regressors | % rejection of $F$ test that all β coefficients are zero (@ 0.05) | Average $\bar{R}^2$ | % of $\bar{R}^2 > 0.7$ |
|---|---|---|---|---|
| Levels | 1 | 76 | 0.26 | 5 |
| | 2 | 78 | 0.34 | 8 |
| | 3 | 93 | 0.46 | 25 |
| | 4 | 95 | 0.55 | 34 |
| | 5 | 96 | 0.59 | 37 |
| Differences | 1 | 8 | 0.004 | 0 |
| | 2 | 4 | 0.001 | 0 |
| | 3 | 2 | –0.007 | 0 |
| | 4 | 10 | 0.006 | 0 |
| | 5 | 6 | 0.012 | 0 |

# Variables in a regression should be stationary

- If you have integrated variables in your model, you should generally difference them before running your regression

- There are some special-case exceptions to this when multiple variables are **cointegrated**, meaning that they follow a non-stationary path together

- We will study cointegration later

# Summary: Key points from this class

- Roots of an autoregressive polynomial are the key to assessing stationarity
- Autoregressive processes with unit roots are called integrated processes; order of integration = # of unit roots
- Random walk is the simplest integrated process; its first difference is white noise
- Regressions involving integrated processes are often spurious, with inflated $t$ statistics and $R$-square values
- We should only run regressions with stationary variables

# What's next

- In the next class (March 30), we will start to examine **regression with stationary variables**
- We will consider time-series regression assumptions **TS-1 through TS-6**
- Then we will discuss **autocorrelation in the error term** as one of the most common regression pathologies, considering its implications for OLS, how to detect it, and how to correct for it