ECON2125/4021/8013

Lecture 23

John Stachurski

Semester 1, 2015

Announcements

SELT feedback is live

- Criticism is welcome — constructive preferred

More solved practice questions coming next week

Linear Models

When studying economic systems we often use linear models

- more correctly, affine models — see below

The advantage of linear systems

- Simple dynamics

The disadvantage of linear systems

- Simple dynamics

Ideal <u>if</u> they can replicate the phenomenon you wish to study

Often used as a building block for more complex models

A generic (deterministic) linear model on $\mathbb{R}^N$ takes the form

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{b}$$

where

- $\mathbf{x}_t$ is $N \times 1$, a vector of "state" variables
- $\mathbf{A}$ is $N \times N$, $\mathbf{b}$ is $N \times 1$, contain parameters
- A dynamical system $(\mathbb{R}^N, g)$ with $g(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$
- Despite the terminology, g is actually affine

When $N = 1$ this becomes

$$x_{t+1} = ax_t + b$$

Example. A simple linear macroeconomic model might look like

$$
\begin{aligned}
\pi_{t+1} &= a_{11}\pi_t + a_{12}i_t + a_{13}y_t + b_1 \\
i_{t+1} &= a_{21}\pi_t + a_{22}i_t + a_{23}y_t + b_2 \\
y_{t+1} &= a_{31}\pi_t + a_{32}i_t + a_{33}y_t + b_3
\end{aligned}
$$

where

- π is inflation
- i is the interest rate
- y is an "output gap"

In general we know that for any (S, g) we have $\mathbf{x}_t = g^t(\mathbf{x}_0)$

For linear systems we can write this out explicitly:

$$
\begin{aligned}
\mathbf{x}_t &= \mathbf{A}\mathbf{x}_{t-1} + \mathbf{b} \\
&= \mathbf{A}(\mathbf{A}\mathbf{x}_{t-2} + \mathbf{b}) + \mathbf{b} \\
&= \mathbf{A}^2\mathbf{x}_{t-2} + \mathbf{A}\mathbf{b} + \mathbf{b} \\
&= \mathbf{A}^2(\mathbf{A}\mathbf{x}_{t-3} + \mathbf{b}) + \mathbf{A}\mathbf{b} + \mathbf{b} \\
&= \mathbf{A}^3\mathbf{x}_{t-3} + \mathbf{A}^2\mathbf{b} + \mathbf{A}\mathbf{b} + \mathbf{b} \\
&= \cdots
\end{aligned}
$$

More generally,

$$\mathbf{x}_t = \mathbf{A}^j \mathbf{x}_{t-j} + \mathbf{A}^{j-1}\mathbf{b} + \mathbf{A}^{j-2}\mathbf{b} + \cdots + \mathbf{A}\mathbf{b} + \mathbf{b}$$

Setting $j = t$

$$\mathbf{x}_t = \mathbf{A}^t \mathbf{x}_0 + \mathbf{A}^{t-1}\mathbf{b} + \mathbf{A}^{t-2}\mathbf{b} + \cdots + \mathbf{A}\mathbf{b} + \mathbf{b}$$

In short,

$$g^t(\mathbf{x}_0) = \mathbf{A}^t \mathbf{x}_0 + \sum_{i=0}^{t-1} \mathbf{A}^i \mathbf{b}$$

In the scalar case this is

$$g^t(x_0) = a^t x_0 + b \sum_{i=0}^{t-1} a^i$$

Stability of Linear Models

Let's consider existence / uniqueness / stability of steady states of linear systems

In particular we study properties of the dynamical system $(\mathbb{R}^N, g)$ with

$$g(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$$

Even existence of a steady state is not guaranteed — consider

$$x_{t+1} = x_t + 1$$

It turns out that existence / uniqueness / stability etc. all depend on the spectral radius of $\mathbf{A}$

Fact. If $\rho(\mathbf{A}) < 1$, then $(\mathbb{R}^N, g)$ is globally stable, with unique steady state

$$\mathbf{x}^* = \sum_{i=0}^{\infty} \mathbf{A}^i \mathbf{b}$$

Proof: A steady state is a solution to $\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b}$, or

$$(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b} \qquad (\star)$$

Recall that $\rho(\mathbf{A}) < 1$ implies $\|\mathbf{A}^k\| < 1$ for some $k \in \mathbb{N}$

By the Neumann series lemma, $(\star)$ has the unique solution

$$(\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \sum_{i=0}^{\infty} \mathbf{A}^i \mathbf{b}$$

It remains to show that

$$\mathbf{x}_t = \mathbf{A}^t \mathbf{x}_0 + \sum_{i=0}^{t-1} \mathbf{A}^i \mathbf{b} \to \sum_{i=0}^{\infty} \mathbf{A}^i \mathbf{b} =: \mathbf{x}^*$$

By definition, we have $\sum_{i=0}^{t-1} \mathbf{A}^i \mathbf{b} \to \sum_{i=0}^{\infty} \mathbf{A}^i \mathbf{b} = \mathbf{x}^*$

Hence if $\mathbf{A}^t \mathbf{x}_0 \to \mathbf{0}$, then

$$\mathbf{x}_t = \mathbf{A}^t \mathbf{x}_0 + \sum_{i=0}^{t-1} \mathbf{A}^i \mathbf{b} \to \mathbf{0} + \mathbf{x}^* = \mathbf{x}^*$$

To see that $\mathbf{A}^t \mathbf{x}_0 \to \mathbf{0}$, note (see the rules for matrix norms) that

$$\|\mathbf{A}^t \mathbf{x}_0 - \mathbf{0}\| = \|\mathbf{A}^t \mathbf{x}_0\| \leq \|\mathbf{A}^t\| \, \|\mathbf{x}_0\| \to 0$$

How exactly do we show that $\|\mathbf{A}^t\| \, \|\mathbf{x}_0\| \to 0$ as $t \to \infty$?

Since $\rho(\mathbf{A}) < 1$, there exists a $k \in \mathbb{N}$ with $\|\mathbf{A}^k\| < 1$

For any t we can write $t = nk + j$ for some $j \in \{0, \ldots, k-1\}$

Using the submultiplicative property of the matrix norm, we have

$$\|\mathbf{A}^t\| = \|\mathbf{A}^{nk+j}\| = \|\mathbf{A}^{nk}\mathbf{A}^j\| \leq \|\mathbf{A}^{nk}\| \, \|\mathbf{A}^j\|$$

Let $L := \max_{0 \leq j \leq k-1} \|\mathbf{A}^j\|$

We then have

$$\|\mathbf{A}^t\| \leq L\|\mathbf{A}^{nk}\| \leq L\|\mathbf{A}^k\|^n$$

Now observe that $t \to \infty$ means $n \to \infty$, and $\|\mathbf{A}^k\| < 1$

There's another way we can show $\mathbf{A}^t \mathbf{x}_0 \to \mathbf{0}$ if $\mathbf{A}$ is diagonalizable

Recall this means that we can write $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ where

$$\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N), \quad \lambda_n = n\text{-th eigenvalue of } \mathbf{A}$$

Recall further that $\mathbf{A}^t = \mathbf{P}\mathbf{D}^t\mathbf{P}^{-1}$

That is,

$$\mathbf{A}^t = \mathbf{P} \begin{pmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_N^t \end{pmatrix} \mathbf{P}^{-1}$$

Since $\rho(\mathbf{A}) < 1$ we have $|\lambda_n| < 1$ for all n

Hence $\mathbf{A}^t \to \mathbf{0}$

Example. Let

$$\mathbf{A} = \begin{pmatrix} 0.6 & -0.7 \\ 0.6 & 0.65 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

```
In [1]: import numpy as np
In [2]: from scipy.linalg import eig
In [3]: A = np.asarray([[0.6, -0.7], [0.6, 0.65] ])
In [4]: evals, evecs = eig(A)
In [5]: evals
Out[5]: array([0.625+0.64759169j,  0.625-0.64759169j])
In [6]: np.abs(evals)
Out[6]: array([0.9,  0.9])  # Implies globally stable
```

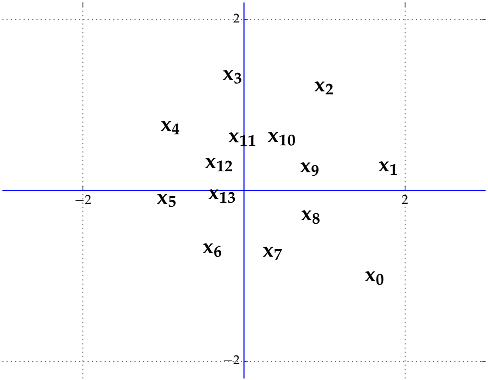

Figure : Convergence towards the origin for $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t$

Stochastic Dynamics

Now it's time to add shocks to our model

As discussed earlier, the data in econ / finance tends to be "noisy" relative to models

- humans are hard to model...

Thus, adding shocks / noise to our models brings them closer to the data

- Prepares us to estimate our models
- Allows us to include patterns observed in the noise

Martingales

Stochastic models are often pieced together from simpler random components, such as IID sequences

Another such building block is martingales

A sequence of random vectors $\{\mathbf{w}_t\}_{t=1}^{\infty}$ is called a **martingale** if,

$$\forall t \geq 1, \quad \mathbb{E}\left[\mathbf{w}_{t+1} \mid \mathbf{w}_t, \mathbf{w}_{t-1}, \ldots, \mathbf{w}_1\right] = \mathbf{w}_t$$

For the rest of this lecture we use the abbreviated notation

$$\mathbb{E}_t[\,\cdot\,] := \mathbb{E}\left[\,\cdot\, \mid \mathbf{w}_t, \mathbf{w}_{t-1}, \ldots, \mathbf{w}_1\right]$$

so that the definition of a martingale becomes

$$\mathbb{E}_t[\mathbf{w}_{t+1}] = \mathbf{w}_t \quad \text{for all } t$$

Example. A player's wealth over a sequence of fair gambles follows a martingale

In particular, let w_t be wealth at time t, where

$$w_t = \sum_{i=1}^{t} \xi_i, \quad \{\xi_t\} \text{ is IID with } \mathbb{E}\,[\xi_t] = 0, \; \forall\, t$$

Then

$$\mathbb{E}_t[w_{t+1}] = \mathbb{E}_t[w_t + \xi_{t+1}] = \mathbb{E}_t[w_t] + \mathbb{E}_t[\xi_{t+1}]$$

The martingale property now follows:

- $\mathbb{E}_t[w_t] = w_t$ because w_t is known at t
- $\mathbb{E}_t[\xi_{t+1}] = \mathbb{E}\,[\xi_{t+1}] = 0$ by independence, zero mean of ξ_{t+1}

A sequence $\{\mathbf{w}_t\}_{t=1}^{\infty}$ is called a **martingale difference sequence** (**MDS**) if

$$\mathbb{E}_t[\mathbf{w}_{t+1}] = \mathbf{0}$$

for all t

Example. If $\{\mathbf{v}_t\}$ is a martingale then the first difference

$$\mathbf{w}_t := \mathbf{v}_t - \mathbf{v}_{t-1}$$

is a MDS because, for any t,

$$\begin{aligned}\mathbb{E}_t[\mathbf{w}_{t+1}] &= \mathbb{E}_t[\mathbf{v}_{t+1} - \mathbf{v}_t] \\ &= \mathbb{E}_t[\mathbf{v}_{t+1}] - \mathbb{E}_t[\mathbf{v}_t] = \mathbf{v}_t - \mathbf{v}_t = \mathbf{0}\end{aligned}$$

Example. Suppose that $\{\mathbf{w}_t\}$ is IID with $\mathbb{E}[\mathbf{w}_t] = \mathbf{0}$

Then $\{\mathbf{w}_t\}$ is an MDS

To see this observe that, by independence,

$$\mathbb{E}_t[\mathbf{w}_{t+1}] = \mathbb{E}[\mathbf{w}_{t+1}] \quad \text{for all } t$$

The conclusion follows

In fact a MDS is a generalization of the idea of a zero mean IID sequence

Often used in economics / finance / econometrics

- Nicely represents the idea of "unpredictable" sequence
- A more natural assumption than independence...?

Fact. If $\{\mathbf{w}_t\}$ is a MDS, then $\mathbb{E}[\mathbf{w}_t] = \mathbf{0}$ for all t

Proof: By the law of iterated expectations,

$$\mathbb{E}[\mathbf{w}_t] = \mathbb{E}[\mathbb{E}_{t-1}[\mathbf{w}_t]] = \mathbb{E}[\mathbf{0}] = \mathbf{0}$$

Fact. If $\{\mathbf{w}_t\}$ is a martingale difference sequence, then

$$\mathbb{E}[\mathbf{w}_s \mathbf{w}_t'] = \mathbf{0} \quad \text{whenever} \quad s \neq t$$

We say that $\mathbf{w}_s$ and $\mathbf{w}_t$ are **orthogonal**

Proof: Supposing without loss of generality that $s < t$, we have

$$\mathbb{E}[\mathbf{w}_s \mathbf{w}_t'] = \mathbb{E}[\mathbb{E}_{t-1}[\mathbf{w}_s \mathbf{w}_t']] = \mathbb{E}[\mathbf{w}_s \mathbb{E}_{t-1}[\mathbf{w}_t']] = \mathbb{E}[\mathbf{0}] = \mathbf{0}$$

As an aside the term "orthogonal" is often used to indicate lack of correlation

To see the connection, let's suppose that $\{w_t\}$ is a scalar MDS

We know from the previous slide that

- $\mathbb{E}[w_t] = 0$ and
- $\mathbb{E}[w_s w_t] = 0$ when $s \neq t$ (orthogonality)

It follows that

$$\operatorname{cov}[w_s, w_t] = \mathbb{E}[(w_s - \mathbb{E}[w_s])(w_t - \mathbb{E}[w_t])] = \mathbb{E}[w_s w_t] = 0$$

Hence orthogonal $\implies$ zero correlation

Now consider the linear **stochastic difference equation**

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{b} + \mathbf{C}\mathbf{w}_{t+1}$$

We assume that

1. $\mathbf{A}$ is $N \times N$ and $\mathbf{b}$ is $N \times 1$
2. $\{\mathbf{w}_t\}$ is $M \times 1$, an MDS with $\mathbb{E}_t[\mathbf{w}_{t+1}\mathbf{w}'_{t+1}] = \mathbf{I}$ for all t
3. $\mathbf{C}$ is an $N \times M$ matrix called the **volatility matrix**
4. $\mathbf{x}_0$ is given

Note: 2 implies that $\mathbb{E}\left[\mathbf{w}_t\mathbf{w}'_t\right] = \mathbf{I}$ for any t because, by the law of iterated expectations,

$$\mathbb{E}\left[\mathbf{w}_t\mathbf{w}'_t\right] = \mathbb{E}\left[\mathbb{E}_{t-1}[\mathbf{w}_t\mathbf{w}'_t]\right] = \mathbb{E}\left[\mathbf{I}\right] = \mathbf{I}$$

Example. A simple linear macroeconomic model might look like

$$\pi_{t+1} = a_{11}\pi_t + a_{12}i_t + a_{13}y_t + b_1 + c_1 u_{t+1}$$

$$i_{t+1} = a_{21}\pi_t + a_{22}i_t + a_{23}y_t + b_2 + c_2 v_{t+1}$$

$$y_{t+1} = a_{31}\pi_t + a_{32}i_t + a_{33}y_t + b_3 + c_3 w_{t+1}$$

where

- π is inflation
- i is the interest rate
- y is an "output gap"
- u, v and w are shocks

Scalar Models

If we specialize to $N = 1$ then we get the scalar model

$$x_{t+1} = a x_t + b + c w_{t+1} \tag{1}$$

Let's look at some time series simulated on a computer

In each case we

1. assume that $\{w_t\} \overset{\text{IID}}{\sim} N(0, 1)$
2. draw $\{w_1, \ldots, w_T\}$ using a random number generator
3. fix $x_0 = 1$
4. update x_t via (1) until $t = T$

```python
import numpy as np
import matplotlib.pyplot as plt

T = 100  # Length of time series
a = 0.5  # Parameter
c = 1    # Parameter
b = 0    # Parameter

w = np.random.randn(T)  # T indep. standard normals
x = np.empty(T)         # Allocate memory
x[0]  = 1               # Initial condition
for t in range(T-1):
    x[t+1] = a * x[t] + b + c * w[t+1]
plt.plot(x)
plt.show()
```

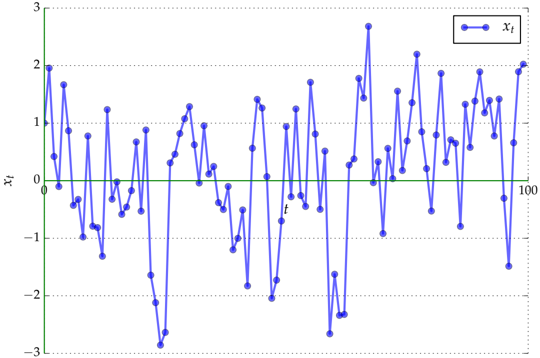

Figure : Linear Gaussian time series, $x_0 = 1$, $a = 0.5$, $b = 0$, $c = 1$

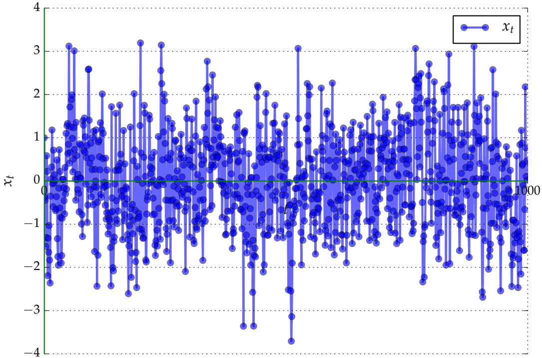

Figure : A longer time series, same parameters

Remarks:

The time series $\{x_t\}$ does not converge to a constant

- $x_{t+1} = ax_t + b + cw_{t+1}$
- since $c \neq 0$, each x_t is disturbed by a shock

Neither does it diverge to $+\infty$ or $-\infty$

- in this case $|a| < 1$, which leads to a kind of stability

We investigate these ideas in detail through the lecture

For starters let's see what happens when c gets small

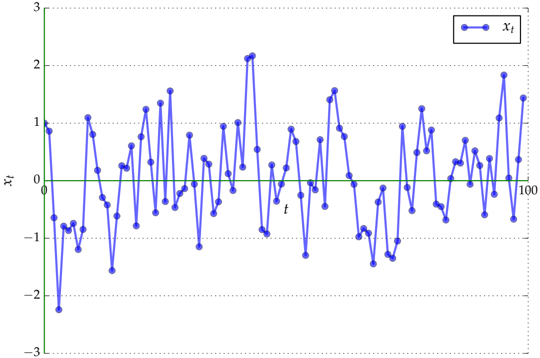

Figure : $x_0 = 1$, $a = 0.5$, $b = 0$, $c = 0.8$

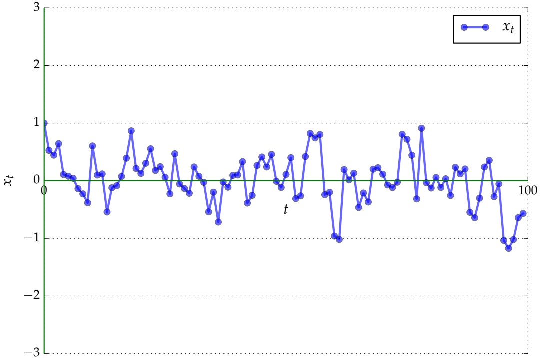

Figure : $x_0 = 1,\ a = 0.5,\ b = 0,\ c = 0.4$

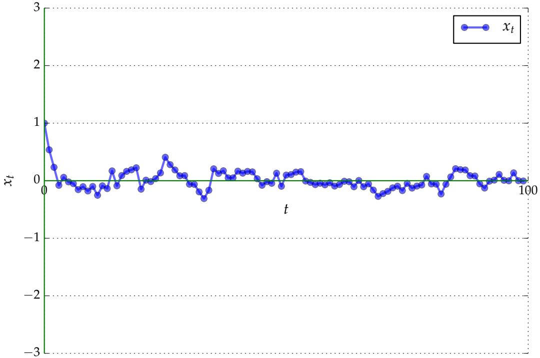

Figure : $x_0 = 1,\ a = 0.5,\ b = 0,\ c = 0.1$

Summary: Lower c means less volatility in the time series

- Hence c is often called the "volatility parameter"

Intuition: As c gets small, the model

$$x_{t+1} = ax_t + b + cw_{t+1}$$

becomes "more similar" to

$$x_{t+1} = ax_t + b$$

In the latter case, when $|a| < 1$, this series converges

What about if $|a| < 1$ does not hold?

In this case the time series tends to diverge

A property of some time series

- population (sometimes)
- GDP in developed countries
- value of a portfolio with compounded interest
- inflation during a hyperinflation

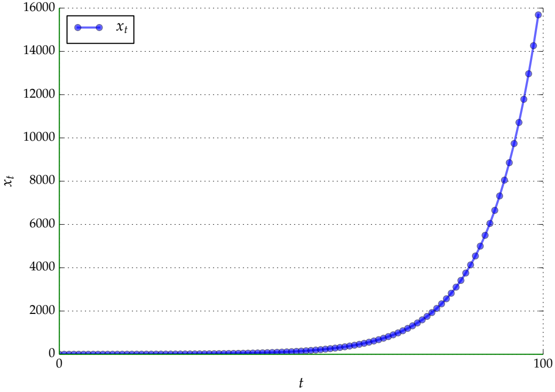

Figure : $x_0 = 1,\ a = 1.1,\ b = 0,\ c = 0.1$

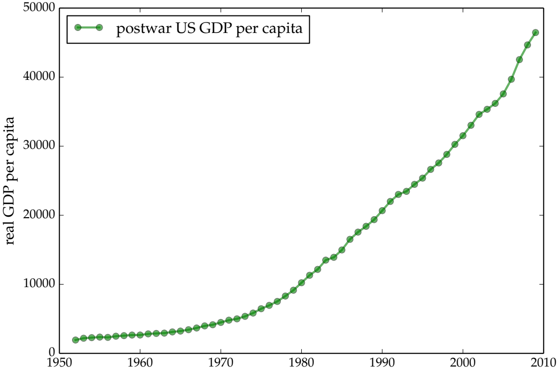

Figure : For comparison: US GDP per capita

For other kinds of time series, no divergence is observed

The assumption $|a| < 1$ is more reasonable

- returns on assets / portfolios
- GDP growth
- interest, inflation, unemployment rates

This is the "stationary" case

- terminology defined more formally later

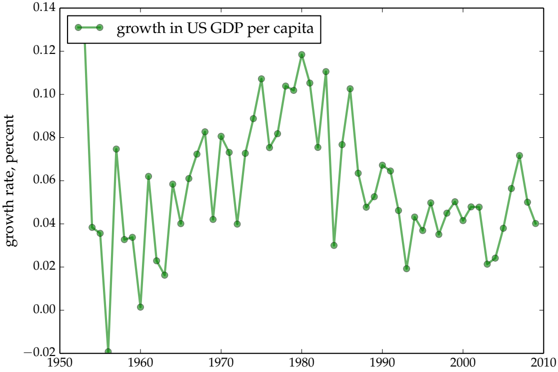

Figure : Growth rates are often stationary

Consider again the stationary case $|a| < 1$

The particular value a is still important as it governs the level of persistence

In the extreme case where $a = 0$, the $\{x_t\}$ process is IID

$$x_{t+1} = ax_t + b + cw_{t+1} = b + cw_{t+1}$$

$$\therefore \quad \{x_t\} \overset{\text{IID}}{\sim} N(b, c^2)$$

On the other hand, as $|a|$ gets close to 1, we see

- strong persistence / correlation
- long deviations from "average" values

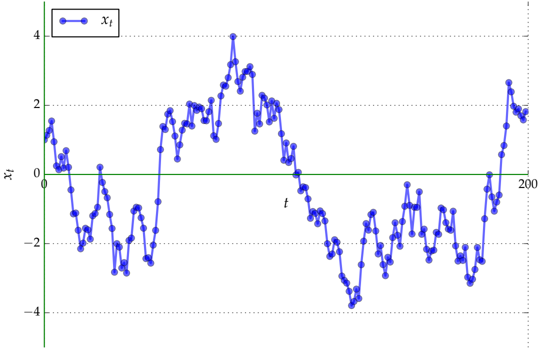

Figure : $x_0 = 1$, $a = 0.95$, $b = 0$, $c = 0.5$

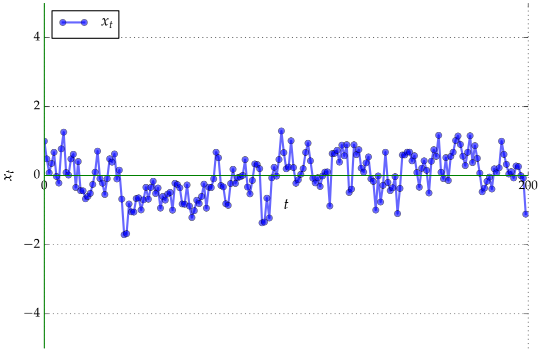

Figure : The same model but with $a = 0.75$

Moving Average Representations

Now let's return to the general case where

- $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{b} + \mathbf{C}\mathbf{w}_{t+1}$
- $\mathbf{w}_t$ is a MDS with $\mathbb{E}_t[\mathbf{w}_{t+1}\mathbf{w}'_{t+1}] = \mathbf{I}$
- $\mathbf{x}_0$ is a constant

In the deterministic case we expressed $\mathbf{x}_t$ in terms of $\mathbf{x}_0$

Here we can express $\mathbf{x}_t$ in terms of $\mathbf{x}_0$ and $\mathbf{w}_1, \ldots, \mathbf{w}_t$

Letting $\mathbf{v}_t := \mathbf{b} + \mathbf{C}\mathbf{w}_t$, we have

$$
\begin{aligned}
\mathbf{x}_t &= \mathbf{A}\mathbf{x}_{t-1} + \mathbf{v}_t \\
&= \mathbf{A}(\mathbf{A}\mathbf{x}_{t-2} + \mathbf{v}_{t-1}) + \mathbf{v}_t \\
&= \mathbf{A}^2\mathbf{x}_{t-2} + \mathbf{A}\mathbf{v}_{t-1} + \mathbf{v}_t \\
&= \mathbf{A}^2(\mathbf{A}\mathbf{x}_{t-3} + \mathbf{v}_{t-2}) + \mathbf{A}\mathbf{v}_{t-1} + \mathbf{v}_t \\
&= \mathbf{A}^3\mathbf{x}_{t-3} + \mathbf{A}^2\mathbf{v}_{t-2} + \mathbf{A}\mathbf{v}_{t-1} + \mathbf{v}_t \\
&= \cdots
\end{aligned}
$$

More generally,

$$\mathbf{x}_t = \mathbf{A}^j\mathbf{x}_{t-j} + \mathbf{A}^{j-1}\mathbf{v}_{t-(j-1)} + \mathbf{A}^{j-2}\mathbf{v}_{t-(j-2)} + \cdots + \mathbf{A}\mathbf{v}_{t-1} + \mathbf{v}_t$$

Setting $j = t$,

$$\begin{aligned}\mathbf{x}_t &= \mathbf{A}^t\mathbf{x}_0 + \mathbf{A}^{t-1}\mathbf{v}_1 + \mathbf{A}^{t-2}\mathbf{v}_2 + \cdots + \mathbf{A}\mathbf{v}_{t-1} + \mathbf{v}_t \\ &= \mathbf{A}^t\mathbf{x}_0 + \sum_{i=0}^{t-1}\mathbf{A}^i\mathbf{v}_{t-i}\end{aligned}$$

Making the substitution $\mathbf{v}_{t-i} = \mathbf{b} + \mathbf{C}\mathbf{w}_{t-i}$, we get

$$\mathbf{x}_t = \mathbf{A}^t\mathbf{x}_0 + \sum_{i=0}^{t-1}\mathbf{A}^i\mathbf{b} + \sum_{i=0}^{t-1}\mathbf{A}^i\mathbf{C}\mathbf{w}_{t-i}$$

The expression

$$\mathbf{x}_t = \mathbf{A}^t \mathbf{x}_0 + \sum_{i=0}^{t-1} \mathbf{A}^i \mathbf{b} + \sum_{i=0}^{t-1} \mathbf{A}^i \mathbf{C} \mathbf{w}_{t-i} \tag{2}$$

is called the **moving average** or **MA** representation of $\mathbf{x}_t$

Example. Consider the scalar case $x_t = a x_{t-1} + w_t$ with $|a| < 1$

The MA representation is

$$x_t = a^t x_0 + a^{t-1} w_1 + a^{t-2} w_2 + \cdots + a w_{t-1} + w_t$$

Since $|a| < 1$, earlier shocks (e.g., w_1) have less influence than later ones (e.g., w_t)

- Similar story in (2) when $\|A\| < 1$