Advanced time-series analysis (University of Lund, Economic History Department)

30 Jan-3 February and 26-30 March 2012

Lecture 1 Fundamentals of time-series, serial correlation, lag operators, stationarity. Long- and short-run multipliers, impulse-response functions.

1.a. Fundamentals of time-series, **serial correlation*:***

Time series contain observations of random variable Y at certain points of time.
$\{y_1, y_2, ..., y_T\} = \{y_t\}_{t=1}^{T}$

Y is an unknown process (Data Generating Process - DGP) that we wish to understand, uncover, but all we have are just some realizations of it (a sample).

Observations in time series has a fixed order, you cannot reorder the observations as you like.

Serial correlation must be given attention.

We can characterize time series (y) by three fundamental statistics:

Expected value ($E(y_t)$), variance ($\gamma_0 = Var(y_t) = E(y_t - E(y_t))^2$ and autocovariance:
$\gamma_s = Cov(y_t, y_{t-s}) = E\left((y_t - E(y_t))(y_{t-s} - E(y_{t-s}))\right)$

Autocorrelation at lag k (Autocorrelation function at lag k: ACF_k) is defined as:

$$ACF_k = \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov(y_t, y_{t-k})}{Var(y_t)}$$

It is worthwhile to compare this with the standard formula for linear correlation:

$Cor(y_t, y_{t-k}) = \dfrac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_t) \cdot Var(y_{t-k})}}$ They are not the same, but they should be reasonably close provided y is homoscedastic.

Useful to understand ACF(k) is to see it as a regression coefficient. Let us take the following regression:

$y_{t-k} = \beta_0 + \beta_1 y_t + \varepsilon_t$ with $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$

The OLS estimator for the slope coefficient is: $\hat{\beta}_1 = \dfrac{Cov(y_t, y_{t-k})}{Var(y_t)} = ACF(k)$

How to test is autocorrelation is present?

Q-tests:

Box-Pierce: $Q_{BP}(k) = T\sum_{i=1}^{k} q_i^2$ and Ljung-Box: $Q_{LB}(k) = T(T+2)\sum_{i=1}^{k}\frac{q_i^2}{T-k}$.

In case of no autocorrelation (this H0), both statistics should follow a chi-squared distribution with k degrees-of-freedom.

Partial Autocorrelation Function (PACF): the autocorrelation between y_t and y_{t-k} with the correlation between y_t and all lags $y_{t-1}...y_{t-k-1}$ removed.

This is the easiest to see as a coefficient of a multivariate regression:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_k y_{t-k} + \varepsilon_t$$

Then β_k is the PACF(k). That is, the relationship between y_t and y_{t-k} with all lower order autocorrelation captured.

The graphical tool that combines ACF and PACF is called the correlogram.

1.b. Lag-operators

This might seem useless for some, but it helps a lot to be able to correctly read the literature and to understand fundamental concepts:

Let *L* be an operator with the following feature:

$Ly_t = y_{t-1}$ that is multiplying by *L* creates a time series lagged by one period.

$L(Ly_t) = L^2 y_t = y_{t-2}$. Generally: $L^k y_t = y_{t-k}$

The lag operator is commutative, distributive and associative:

For example:

$(L^1 + L^2)y_t = y_{t-1} + y_{t-2}$, $L^1 L^2 y_t = L^3 y_t = y_{t-3}$, $L^{-1} y_t = y_{t+1}$

The operation: $(1-L)y_t = y_t - y_{t-1}$ is called first-differencing.

An important equality is:

$\left(1-\theta L\right)^{-1} y_t = \left(1+\theta L + \theta^2 L^2 + \theta^3 L^3 + ...\right) y_t = \theta(L) y_t$ provided $|\theta| < 1$ (the series at the right-hand side is absolute summable).

Proof: multiply both sides by $\left(1-\theta L\right)$:

$$y_t = (1-\theta L)\left(1+\theta L+\theta^2L^2+\theta^3L^3+...\right)y_t = \left(1+\theta L+\theta^2L^2+\theta^3L^3+...\right)-\left(\theta L+\theta^2L^2+\theta^3L^3+\theta^4L^4+...\right)$$
$$y_t = \left(1-\theta^n L^n\right)y_t$$

Since $\lim_{n\to\infty}\left(\theta^n L^n y_t\right)=0$ if $|\theta|<1$ so the equality holds.

1.c. Stationarity

Stationarity in the strict sense: A process is called strictly stationary if its probability distribution is independent of time. In other words: if we have a time series observed between time t and T, then the joins probability distribution of these observations should be the same as those observed between any t+k and T+k.

What we actually say here is that the main characteristics of the time series should remain the same whenever we observe it. This is however a theoretical concept.

Stationarity in the weak sense or covariance stationarity defines stationary processes in term of their moments. This makes this definition easier to work with and testable.

Time series y is called *covariance stationary* if:

1. It has a finite mean: $E(y_t)=\mu<\infty$. (This means that the expected value should be independent of time – no trend!)
2. It has a finite variance: $Var(y_t)=\sigma_y^2<\infty$
3. Its autocovariance depends on the difference of the observations (k) only and is independent of time: $Cov(y_t, y_{t-s}) = Cov(y_{t+k}, y_{t-s+k})$

Let us assume that we have an AR(p) process as follows:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + ... + \theta_p y_{t-p} + \varepsilon_t$$

This can be rewritten with lag operators as follows:

$y_t = \left(\theta_1 L + \theta_2 L^2 + ... + \theta_p L^p\right) y_t + \varepsilon_t$ (this we call a polynomial form, denoted as $\theta(L)$)

or

$\left(1-\theta_1 L - \theta_2 L^2 - \theta_p L^p\right) y_t = \varepsilon_t$ (the left-hand side is called the *inverse polynomial* form)

The above series is stationary if **all** the roots (solutions) of the following inverse polynomial:

$1-\theta_1 z - \theta_2 z^2 - \theta_p z^p = 0$ lie outside the unit circle (if the roots are real (not complex) it means that all roots should be larger than one in absolute value).

Beware: this rule is only correct if we have the inverse polynomial form. If we solve the standard polynomial form then the roots should be within the unit-circle so that the process is stationary.

Some examples:

Ex 1.1: $y_t = 0.5y_{t-1} + \varepsilon_t$

The inverse polynomial form is:

$(1-0.5L)y_t = \varepsilon_t$

The characteristic equation is:

$1-0.5z=0$, where the root is: $z=2$ so this AR(1) model is stationary.

Ex 1.2: $y_t = 0.5y_{t-1} + 0.3y_{t-2} + \varepsilon_t$

The inverse polynomial form is:

$(1-0.5L-0.3L^2)y_t = \varepsilon_t$

The characteristic equation is:

$1-0.5z-0.3z^2=0$, where the roots are: $z_{1,2} = \dfrac{0.5 \pm \sqrt{0.25-4\cdot(-0.3)\cdot 1}}{-0.6}$ $z_1 = -2.84, z_2 = 1.17$

Both roots exceed 1 in absolute terms, the above AR(2) model is stationary.

Ex 1.3: $y_t = 0.4y_{t-1} - 0.2y_{t-2} + \varepsilon_t$

The inverse polynomial form is:

$(1-0.4L+0.2L^2)y_t = \varepsilon_t$

The characteristic equation is:

$1-0.4z+0.2z^2=0$, where the roots are: $z_{1,2} = \dfrac{0.4 \pm \sqrt{0.16-4\cdot 0.2\cdot 1}}{0.4}$.

There is a problem now because the expression under the square root is negative. The roots are complex in this case. Do not worry, using what we now about complex numbers helps:

$\sqrt{0.16-4\cdot 0.2\cdot 1} = \sqrt{-0.64} = \sqrt{0.64}i = 0.8i$ ($i=\sqrt{-1}$)

So: $z_{1,2} = \dfrac{0.4 \pm 0.8i}{0.4} = 1 \pm 2i$ The absolute value of these complex numbers is its modulus:

$|1+2i| = |1-2i| = \sqrt{1^2 + 2^2} = \sqrt{5} = 2.236$, which is larger than one. The process is stationary.

This means that the process will return to its mean in an oscillating way.

It is because of potentially complex roots that we mention unit circle in the definition. Simply, the modulus looks like an equation of a circle. What the definition required was that if the complex root was a+bi or a-bi then $a^2+b^2>1$. That is the modulus lies outside of a circle with unit radius.

Let us find a non-stationary series!

This requires that the roots are on or inside the unit circle:

The series: $y_t = y_{t-1} + \varepsilon_t$ or $(1-L)y_t = \varepsilon_t$ is non-stationary, since the solution is:

$1 - z = 0 \rightarrow z = 1$ This is the case when we say that the above process contains a **unit-root**.

If the roots are inside the unit circle, the process is called explosive, but this case does not occur in social sciences so we do not pay extra attention to this case.

The process $y_t = y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$ is called a random-walk. This is a fundamental time-series model type. The idea is that the change of the process is completely random since:

$\Delta y_t = \varepsilon_t$ where $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$ that is it has no autocorrelation, i.e., it is not predictable.

Another version is the random-walk with drift:

$y_t = \alpha + y_{t-1} + \varepsilon_t, \varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$ which is also a non-stationary series but has a trend with slope α.

In order to understand why these processes are non-stationary let us express their value from their starting value y_0.

$$
\begin{aligned}
y_0 &= y_0 \\
y_1 &= y_0 + \varepsilon_1 \\
y_2 &= y_1 + \varepsilon_2 = y_0 + \varepsilon_1 + \varepsilon_2 \\
y_3 &= y_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \\
y_t &= y_0 + \sum_{i=1}^{t} \varepsilon_i
\end{aligned}
$$

We can observe that even after t periods, the initial value is the expected value of the series ($E(y_t) = y_0$), and it remembers all shocks that occurred. The process has a “long-memory”. What about its volatility?

if: $y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$ then $\sigma_y^2 = \sum_{i=1}^{t} \sigma_{\varepsilon_i}^2$ that is the variance of a random-walk equals the sum of the variance of all previous and the current innovations. If we assume homogeneity of the innovations, that is: $\sigma_\varepsilon^2 = \sigma_{\varepsilon_i}^2$ for all i, then $\sigma_y^2 = \sigma_\varepsilon^2 t$. Obviously the variance of y depends on time, that is, it cannot be stationary, since as $\lim_{t\to\infty} \sigma_y^2 = \lim_{t\to\infty} \sigma_\varepsilon^2 t = \infty$.

If we have a random-walk with drift:

$$y_0 = y_0$$
$$y_1 = y_0 + \alpha + \varepsilon_1$$
$$y_2 = y_1 + \alpha + \varepsilon_2 = y_0 + \alpha + \alpha + \varepsilon_1 + \varepsilon_2$$
$$y_3 = y_0 + \alpha + \alpha + \alpha + \varepsilon_1 + \varepsilon_2 + \varepsilon_3$$
$$y_t = y_0 + t\alpha + \sum_{i=1}^{t} \varepsilon_i$$

That is: $E(y_t) = y_0 + t\alpha$ and $\sigma_y^2 = \sum_{i=1}^{t} \sigma_{\varepsilon_i}^2 = \sigma_\varepsilon^2 t$

This process is non-stationary for two reasons: firstly, its expected value depends on time (has a trend); secondly its variance also depends on time.

The non-stationary processes are also known as **non-invertible**.

The reason is that as you remember the expression:

$(1-\theta L)^{-1} y_t = (1 + \theta L + \theta^2 L^2 + \theta^3 L^3 + ...) y_t$ was only true if $|\theta| < 1$, that is, the infinite series at the right hand side was convergent (also called absolute summable): $\lim_{n\to\infty} (1-\theta^n L^n) y_t = y_t$ if $|\theta| < 1$. If $|\theta| = 1$, $\lim_{n\to\infty} (1-\theta^n L^n) y_t = 0$ that is: $(1-\theta L)^{-1} y_t = 0$ which cannot be unless all observation are zero. In short, the random-walk process is non-invertible.

Wold's representation theorem says that every covariance-stationary time series Y_t can be written as an infinite moving average (MA(∞)) process of its innovation process.

That is, all covariance stationary series can be also seen as a linear combination of all the random shocks, innovation that occurred to it.

1.d. Short- and long-run multipliers.

An important and useful was to look at time-series by the effect of some innovations on their value.

Obviously, if we had a series like this:

$y_t = \alpha_0 + \alpha_1 y_{t-1}$ we would find a constant value for all observations sooner or later, equaling its expected value, $Ey_t = \frac{\alpha_0}{1-\alpha_1}$, independently of the initial value. This is a deterministic process.

There is some process behind the time series that drives the series, it is called the **forcing process (or input variable)**.

The forcing process may contain exogenous variables (denoted by X) that might be observed or not, or purely random shocks. In the simplest case the forcing process is a purely random innovation $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$, making the series a stochastic time series.

Now we can write: $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$

The **dynamic multiplicator** is the effect of a change in the input variable in period t on the value of y in period t+j. In this example: $\frac{\partial y_{t+2}}{\partial \varepsilon_t}$

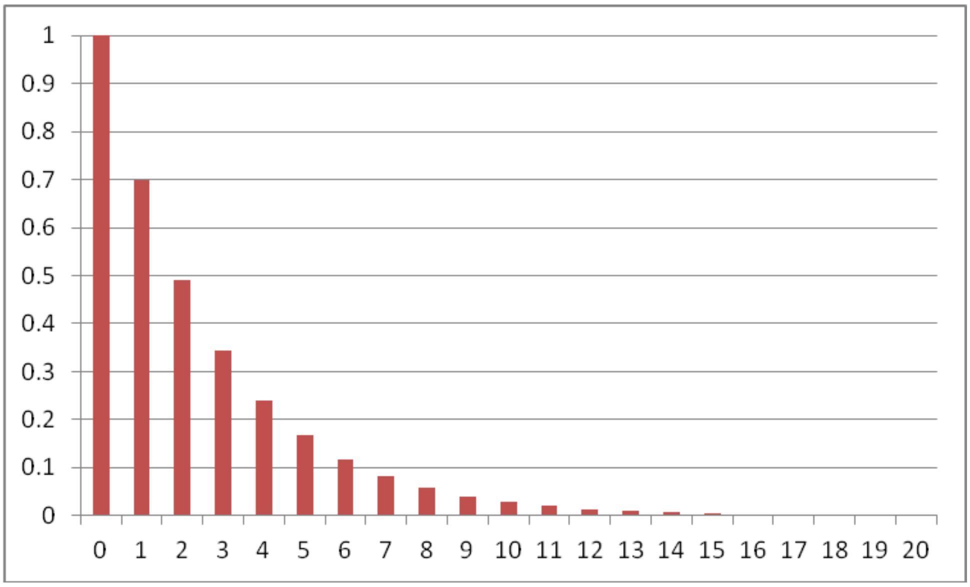
This is easy to find out now, since the process can be written as:

$y_t = (1-\alpha_1 L)^{-1}(\varepsilon_t + \alpha_0)$, hence, provided $|\alpha_1| < 1$, $y_t = (1+\alpha_1 L + \alpha_1^2 L^2 + \alpha_1^3 L^3 + \ldots)(\varepsilon_t + \alpha_0)$.

So the dynamic multiplicator after 2 periods is: $\frac{\partial y_{t+2}}{\partial \varepsilon_t} = \frac{\partial y_t}{\partial \varepsilon_{t-2}} = \alpha_1^2$

. It is obvious that as $|\alpha_1| < 1$, the more time passes after a shock, the lower its effect on current value of the y is going to be. The Impulse-Response Function (IRF) visualizes the effect of a change in an input variable in t on y in t+j for (j=0….k).

IRF of a unit impulse in ε_t if the process is $y_t = .7 \cdot y_{t-1} + \varepsilon_t$



The long-run effect of an input variable on y is the effect of a permanent change in the input variable on y with j→∞. This equals with the sum of the dynamic multiplicators for all j=0…∞, which is called the cumulative IRF.

$$\sum_{j=0}^{\infty}\frac{\partial y_{t+j}}{\partial \varepsilon_t}=1+\alpha_1+\alpha_1^2+\alpha_1^3+...=\frac{1}{1-\alpha_1}$$

Actually you can calculate the long-run effect (this can be called steady state as well) even simpler:

Let us assume we have the AR(1) process as above:

$y_t=\alpha_0+\alpha_1 y_{t-1}+\varepsilon_t$, we know that if the process is stable (stationary), the shock's total effect will converge to a finite value. So finally a new equilibrium value of y will be reached where:

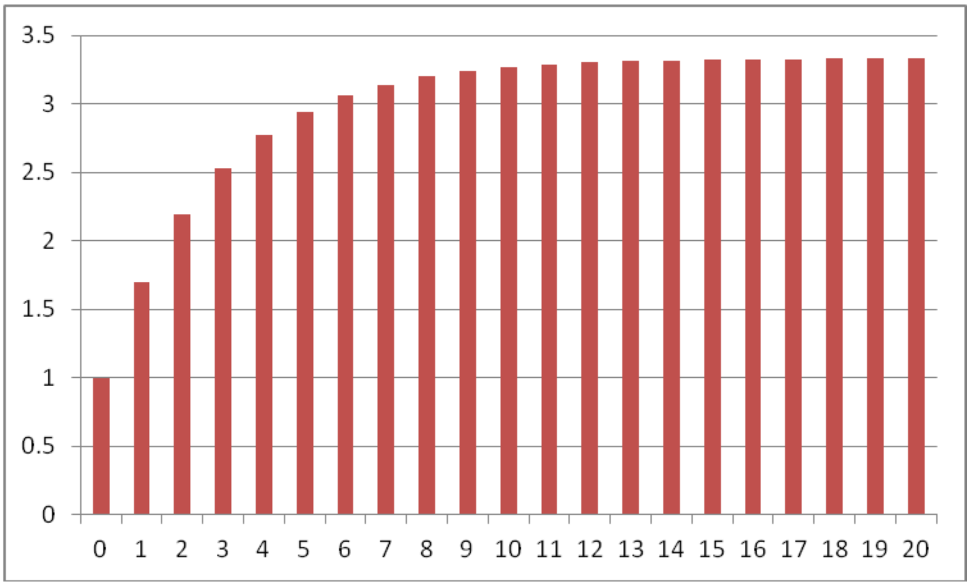
$y^*=y_t=y_{t-1}=y_{t-2}=...$. We just need to substitute this into the model:

$$y^*=\alpha_0+\alpha_1 y^*+\varepsilon_t$$

So: $y^*=\frac{\alpha_0}{1-\alpha_1}+\frac{\varepsilon_t}{1-\alpha_1}$ the long run effect is: $\frac{\partial y^*}{\partial \varepsilon_t}=\frac{1}{1-\alpha_1}$.

And here we go.

The cumulative IRF of **a unit impulse in** ε_t if the process is $y_t=.7\cdot y_{t-1}+\varepsilon_t$



Note: observe that after a few periods the total effect approaches 1/0.3=3.333.

What if we have a more complex forcing process?

Say: $y_t=\beta_0+\beta_1 x_t+\beta_2 x_{t-1}+\varepsilon_t$

The short-run multiplicators are:

$\frac{\partial y_t}{\partial x_t} = \beta_1$, so the immediate effect of a unit increase (impulse) in x on y equals β_1.

$\frac{\partial y_t}{\partial x_{t-1}} = \beta_2$, that is, after one period, the same unit increase (impulse) in x has an effect of β_2 on y.

$\frac{\partial y_t}{\partial x_{t-j}} = 0$ for j>1, that is after one period, an impulse in x has no effect on y at all.

But what was then the total (long-run) effect?

$$\frac{\partial y_t}{\partial x_t} + \frac{\partial y_t}{\partial x_{t-1}} = \beta_1 + \beta_2$$

You can also use the alternative way of thinking: at equilibrium (when all changes are complete) the system settles down at its steady-state (denoted by an asterisk).

$y^* = y_t = y_{t-1} = y_{t-2} = \ldots$ and $x^* = x_t = x_{t-1} = x_{t-2} = \ldots$ so:

$y^* = \beta_0 + \beta_1 x^* + \beta_2 x^* + \varepsilon_t = \beta_0 + (\beta_1 + \beta_2) x^* + \varepsilon_t$ which yields: $\frac{\partial y^*}{\partial x^*} = \beta_1 + \beta_2$.

Example: Determine the dynamic multipliers and the long-run effects of a change in x in the process (this is an ARX(1) process):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + \varepsilon_t$$

$$\frac{\partial y_t}{\partial x_t} = \beta_2 , \ \frac{\partial y_t}{\partial x_{t-1}} = \frac{\partial y_t}{\partial y_{t-1}} \cdot \frac{\partial y_{t-1}}{\partial x_{t-1}} = \beta_1 \beta_2 , \frac{\partial y_t}{\partial x_{t-2}} = \frac{\partial y_t}{\partial y_{t-1}} \cdot \frac{\partial y_{t-1}}{\partial y_{t-2}} \cdot \frac{\partial y_{t-2}}{\partial x_{t-2}} = \beta_1^2 \beta_2 , \frac{\partial y_t}{\partial x_{t-i}} = \beta_1^i \beta_2$$

Or, using lag operator makes our work much easier (for an AR(1) process at least..)!

$(1 - \beta_1 L) y_t = \beta_0 + \beta_2 x_t + \varepsilon_t \rightarrow$

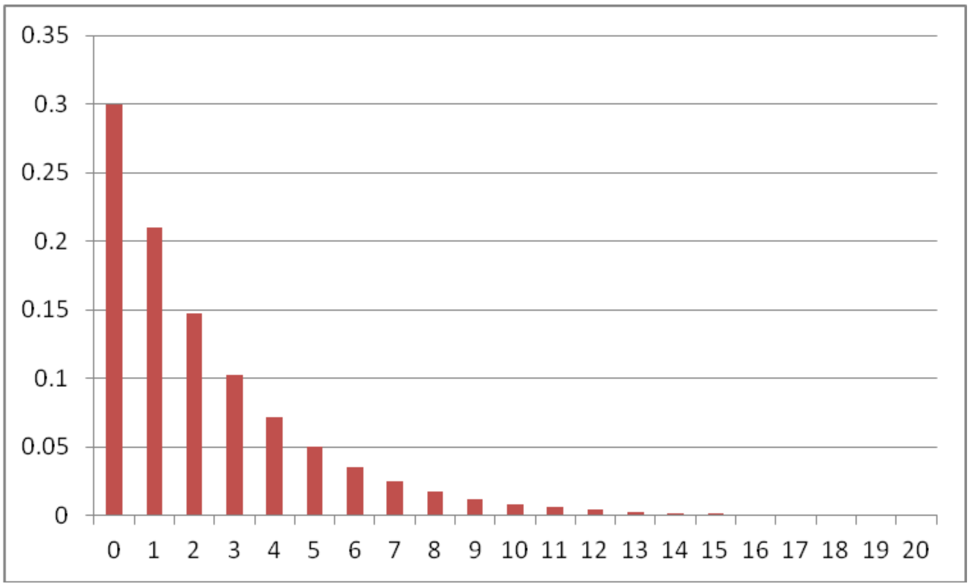
$y_t = (1 - \beta_1 L)^{-1} (\beta_0 + \beta_2 x_t + \varepsilon_t) = (1 + \beta_1 L + \beta_1^2 L^2 + \beta_1^3 L^3 + \ldots)(\beta_0 + \beta_2 x_t + \varepsilon_t)$ taking the derivatives of which with respect to different lags of x yields the same dynamic multipliers as above.

The long-run effect of actually much simpler to calculate. Let us use our simple method:

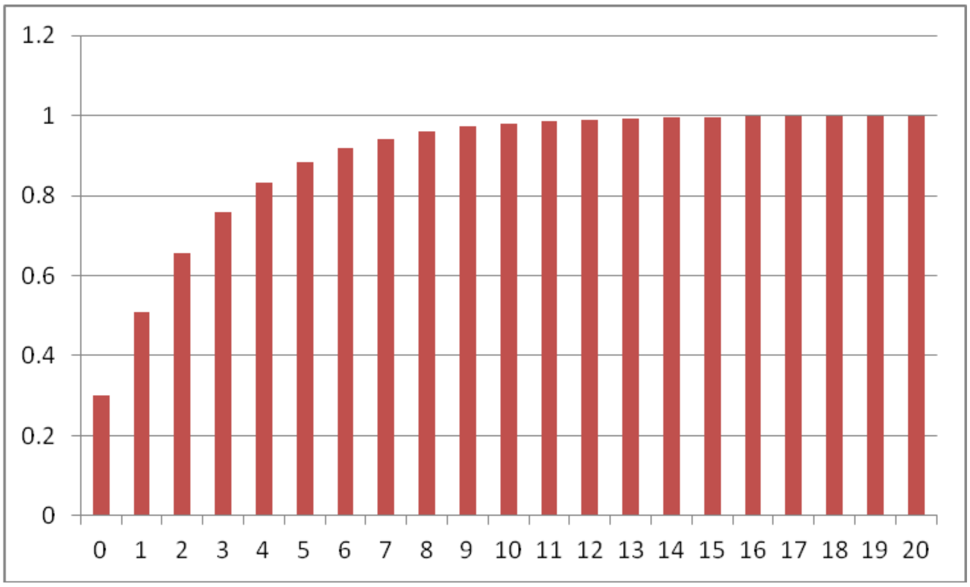
$$y^* = \beta_0 + \beta_1 y^* + \beta_2 x^* + \varepsilon_t \rightarrow y^* = \frac{\beta_0}{1-\beta_1} + \frac{\beta_2}{1-\beta_1} x^* + \frac{\varepsilon_t}{1-\beta_1} \rightarrow \frac{\partial y^*}{\partial x^*} = \frac{\beta_2}{1-\beta_1}$$

If $y_t = 0.7 y_{t-1} + 0.3 x_t + \varepsilon_t$ then

The IRF of **a unit impulse in** x_t if the process is $y_t = 0.7 y_{t-1} + 0.3 x_t + \varepsilon_t$



The cumulative IRF of **a unit impulse in** x_t if the process is $y_t = 0.7 y_{t-1} + 0.3 x_t + \varepsilon_t$



Note: the cumulative IRF nicely converges to the long-run effect: $\frac{.3}{1-.7} = 1$