Advanced Quantitative Methods: Time-series analysis

Johan A. Elkink

University College Dublin

20 February 2014

Time-series processes

A time-series can have been generated by various different types of processes.

Which process generated the data of course affects which econometric model is more appropriate to estimate its parameters.

Linear model

The linear regression model looks like:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, or, if we have no explanatory variables, μ is a constant.

For now, we will look at the latter case, $\mu_t = \mu$.

In the linear model, we assume ε to be an IID variable, $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$.

Moving average process

In the moving average model, we replace the assumption of entirely independent residuals by assuming that the residual at time t is a weighted average between that residual and the one at $t-1$.

$$y_t = \mu + (\varepsilon_t + \phi\varepsilon_{t-1}) \qquad -1 < \phi < 1$$

Moving average process

The above is a so-called MA(1) process, a moving average process with one lag.

This model can be generalised to more lags, the MA(q) process:

$$y_t = \mu + (\varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2})$$

$$y_t = \mu + (\varepsilon_t + \sum_{l=1}^{q} \phi_l \varepsilon_{t-l})$$

Moving average process

Theoretically this model can be generalised to infinitely many lags:

$$y_t = \mu + (\varepsilon_t + \sum_{l=1}^{\infty} \phi_l \varepsilon_{t-l})$$

Now, we could assume that $\phi_l = \alpha^l$, for some $|\alpha| < 1$, thus an exponentially decreasing function of the lag.

Autoregressive process

$$y_t = \mu + \sum_{l=0}^{\infty} \alpha^l \varepsilon_{t-l}$$

This can be shown to be equivalent to:

$$y_t = (1 - \alpha)\mu + \alpha y_{t-1} + \varepsilon_t,$$

$$y_t = \delta + \alpha y_{t-1} + \varepsilon_t,$$

which is called the **autoregressive process**.

Autoregressive process

The AR(1) process can also be extended to the AR(p) process:

$$y_t = \delta + \sum_{l=1}^{p} \alpha_l y_{t-l} + \varepsilon_t$$

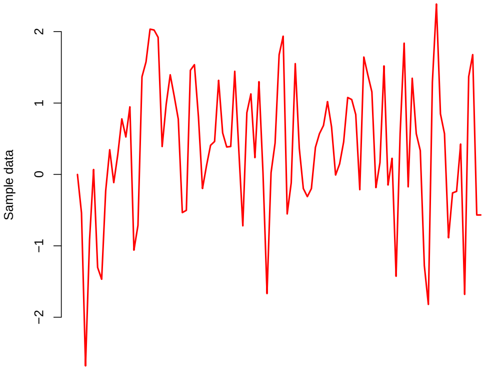
Whereby

$$y_t = \delta + \sum_{l=1}^{\infty} \alpha_l y_{t-l} + \varepsilon_t$$

would be equivalent to a MA(1) process.

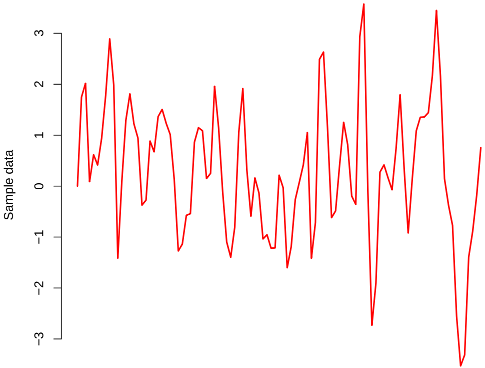
Moving average process

Simulated data, MA(1), $\phi = .5$



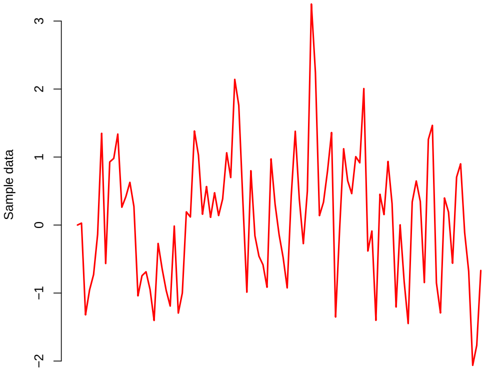
Moving average process

Simulated data, MA(1), $\phi = .9$



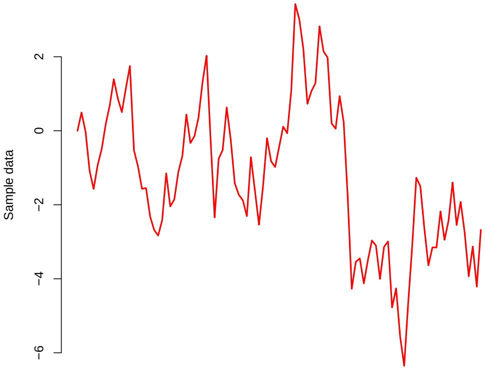
Autoregressive process

Simulated data, AR(1), $\alpha = .5$



Autoregressive process

Simulated data, AR(1), $\alpha = .9$



ARMA(p,q)

The moving average process, MA(q), and the autoregressive process, AR(p), can be combined in the ARMA(p,q) process.

$$y_t = \mu + \sum_{l=1}^{p} y_{t-l}\alpha_l + \sum_{l=1}^{q} \varepsilon_{t-l}\phi_l + \varepsilon_t$$

Autocorrelation function

The autocorrelation function (**ACF**), or correlogram, is the correlation between y_t and y_{t-k}, as a function of k.

If we define

$$\begin{aligned} Var(y_t) = Var(y_{t-k}) &= \gamma_0 \\ Cov(y_t, y_{t-k}) &= \gamma_k, \end{aligned}$$

then

$$\rho_k = \frac{\gamma_k}{\sqrt{\gamma_0 \gamma_0}} = \frac{\gamma_k}{\gamma_0}$$

Autocorrelation function

For the moving average model:

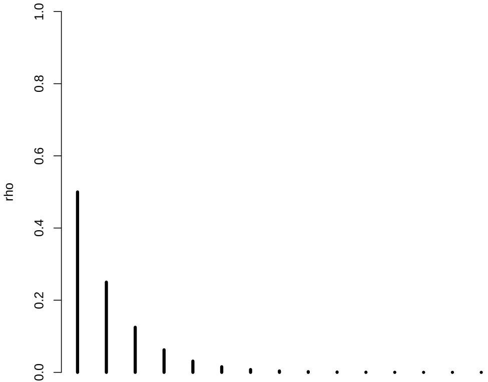
$$\rho_1 = \frac{\phi}{1+\phi^2}, \qquad \rho_k = 0 \quad \forall \quad k > 0$$

For the autoregressive model:

$$\rho_k = \alpha^k$$

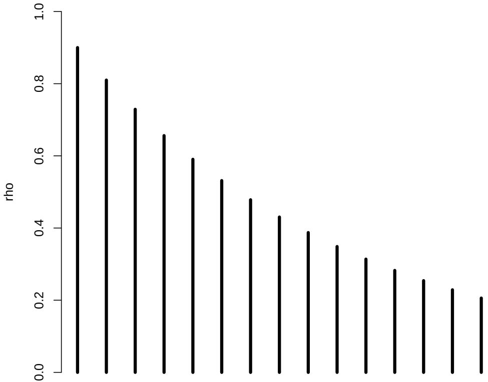
Autocorrelation function

Theoretical ACF, AR(1) process, $\alpha = .5$



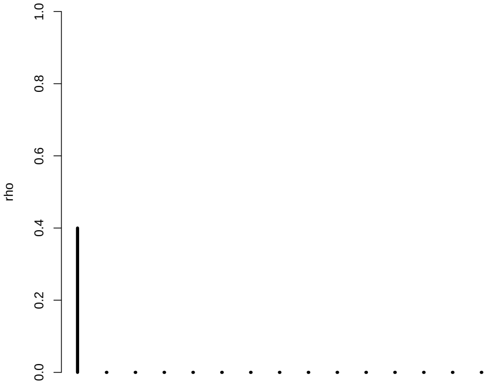
Autocorrelation function

Theoretical ACF, AR(1) process, $\alpha = .9$



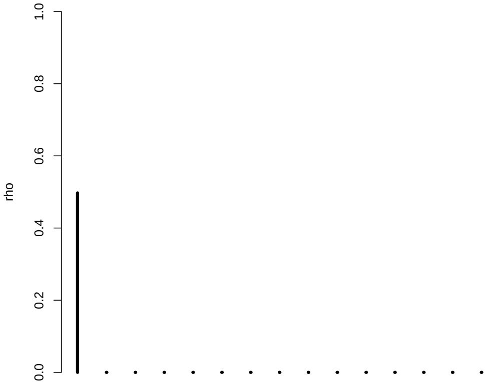
Autocorrelation function

Theoretical ACF, MA(1) process, $\phi = .5$



Autocorrelation function

Theoretical ACF, MA(1) process, $\phi = .9$



Partial autocorrelation

Instead of looking at the autocorrelation function, one can look at the **partial autocorrelation function (PACF)**. This describes the correlation between y_t and y_{t-k}, given all values of y in between.

This can quite simply be calculated by looking at $\hat{\alpha}_k$, the coefficient on the kth coefficient of the AR(k) model.

An AR(p) has an exponentially decreasing ACF and a sharp cut-off point in the PACF. The cut-off point suggests the proper value for p. A very slow (linear) decline in the ACF suggests a unit root. An MA(q) has a sharp cut-off point in the ACF. The cut-off point suggests the proper value for q.

Exercise: unemployment

Using `unemployment.dta`, look at plots of unemployment rates and exports over time and produce ACF and PACF plots for those two variables.

Stationarity

A process is **strictly stationary** if the underlying probability distribution is constant over time.

A process is **weakly stationary** if the following conditions hold:

$$E(y_t) = \mu \quad \forall \quad t$$
$$Var(y_t) = \sigma^2 \quad \forall \quad t$$
$$Cov(y_t, y_{t-k}) = Cov(y_{t+j}, y_{t+j-k}) \quad \forall \quad t, k, j$$

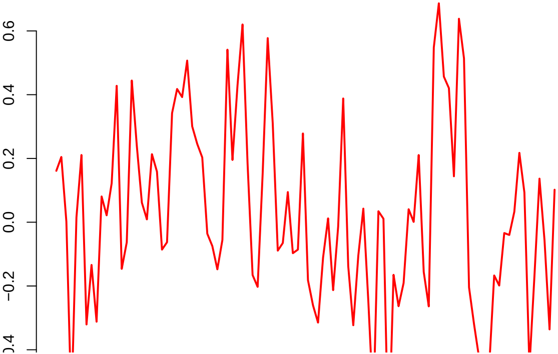
and it follows that the autocorrelations will depend on the lag length only:

$$Cor(y_t, y_{t-k}) = \frac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_t)Var(y_{t-k})}} = \rho_k.$$

(Harrison 2009: 3)

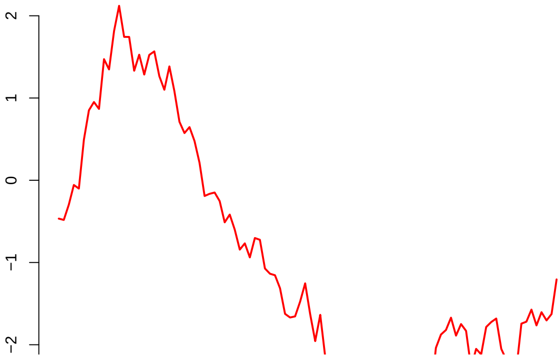
Stationarity: example

$$y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$$



Nonstationarity: example

$$y_t = y_{t-1} + \varepsilon_t$$



Unit root

An AR(1) process where $|\alpha| = 1$ (i.e., $y_t = \delta + y_{t-1} + \varepsilon_t$) is said to have a **unit root**.

Unit roots can be much harder to detect. E.g.
$y_t = \delta + 0.8y_{t-1} + 0.2y_{t-2} + \varepsilon_t$ also has a unit root.

Consequences:

- Consistency and asymptotical normality proofs of OLS, GLS, ML, IV are invalid.
- Regressing two variables with unit roots on each other leads to spurious regression.

(Harrison 2009: 42-46)

Dickey-Fuller test

Subtracting y_{t-1} from both sides of $y_t = \alpha y_{t-1} + \varepsilon_t$ gives:

$$\Delta y_t = (\alpha - 1)y_{t-1} + \varepsilon_t = \beta y_{t-1} + \varepsilon_t$$

so we can regress $(y_t - y_{t-1})$ on y_{t-1} to test whether there is a unit root.

However, under the H_0 of a unit root, $\Delta y_t \sim I(0)$ and $y_t \sim I(1)$, so t-test is invalid. Critical values τ_{nc}, τ_c and τ_{ct} have been published for processes without constant, with constant, and with constant and trend, respectively.

The test assumes no autocorrelation in ε.

(Harrison 2009: 46-47)

Augmented Dickey-Fuller test

The DF test only works for AR(1) processes without autocorrelation in ε. For AR(p) processes or AR(1) processes with autocorrelated errors, we can use the ADF test.

$$\Delta y_t = \beta^* y_{t-1} + \sum_{k=1}^{p} \delta_k \Delta y_{t-k} + \varepsilon_t$$

and the same τ's can be used as critical values for t_{β^*}.

The power of this test is low (i.e. detects unit root too easily). The power depends on the type and strength of the autocorrelation.

(Davidson & MacKinnon 1999: 610-613; Harrison 2009: 48)

Dickey-Fuller test

Dickey-Fuller test:

```
library(tseries)
adf.test(x, k = 0)
```

Augmented Dickey-Fuller test:

```
adf.test(x)
adf.test(x, k = 2)
```

Exercise: unemployment

Using `unemployment.dta`, test for stationarity in unemployment rates and exports.

Integrated

If $E(y_t)$, $Var(y_t)$ and $Cov(y_t, y_{t-k})$ converge to limits μ^*, σ^{*2} and ρ^*_k, respectively, as $t \rightarrow \infty$, then the process is called **asymptotically stationary**, or **integrated of order zero**, or $I(0)$.

A stationary process is thus $I(0)$, but an $I(0)$ process not necessarily stationary.

(Harrison 2009: 40)

Dynamic models

If a regression model depends on lagged values of (one or more) dependent variables, it is called a **dynamic model**.

(Davidson & MacKinnon 1993: 680)

Partial adjustment model

A common model dynamic model is the **partial adjustment model**. Imagine that theory suggests that $y_t^* = \mathbf{X}_t\beta^* + \varepsilon_t$. We can imagine that it takes time for **y** to adjust when **X** changes. Thus, Δy_t depends on the distance of y_{t-1} from it's expected (or equilibrium) value:

$$\begin{aligned} y_t - y_{t-1} &= (1-\delta)(y_t^* - y_{t-1}) + v_t \\ y_t &= y_{t-1} - (1-\delta)y_{t-1} + (1-\delta)\mathbf{X}_t\boldsymbol{\beta}^* + (1-\delta)\varepsilon_t + v_t \\ &= \mathbf{X}_t\boldsymbol{\beta} + \delta y_{t-1} + u_t, \end{aligned}$$

with $\boldsymbol{\beta} = (1-\delta)\boldsymbol{\beta}^*$ and $u_t = (1-\delta)\varepsilon_t + v_t$.

In political science, most time-dependence is assumed to be captured simply by adding a lagged dependent variable to the model. The partial adjustment model is thus ubiquitous in the field.

(Davidson & MacKinnon 1993: 680)

Partial adjustment model

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + \delta y_{t-1} + u_t$$

Because y_{t-1} will be correlated with u_t, OLS estimates are biased, but they are still consistent.

(Davidson & MacKinnon 1993: 681)

Autoregressive distributed lag

A generalization of the partial adjustment model is the ADL(p,q) model:

$$y_t = \alpha + \sum_{l=1}^{p} \beta_l y_{t-l} + \sum_{l=0}^{q} \gamma_l x_{t-l} + u_t$$

Note that many models are special cases of ADL(1,1):

Static model	$\beta_1 = \gamma_1 = 0$
AR(1)	$\gamma_0 = \gamma_1 = 0$
Partial adjustment	$\gamma_1 = 0$
AR(1) errors	$\gamma_1 = -\beta_1 \gamma_0$
First differences	$\beta_1 = 1$ and $\gamma_1 = -\gamma_0$

(Davidson & MacKinnon 1993: 682)

ADL(1,1)

$$y_t = \alpha + \beta_1 y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + u_t$$

The immediate effect of x on y would be

$$\frac{dy_t}{dx_t} = \gamma_0$$

The effect after one time period would be

$$\frac{dy_{t+1}}{dx_t} = \beta_1 (\frac{dy_t}{dx_t}) + \gamma_1 = \beta_1 \gamma_0 + \gamma_1$$

The **long-run multiplier** would be

$$\frac{\gamma_0 + \gamma_1}{1 - \beta_1}$$

ADL(1,1) and error correction

$$\begin{aligned} y_t &= \alpha + \beta_1 y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + u_t \\ \Delta y_t &= \alpha + (\beta_1 - 1) y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + u_t \\ &= \alpha + (\beta_1 - 1)(y_{t-1} + \frac{\gamma_0 + \gamma_1}{\beta_1 - 1} x_{t-1}) + \gamma_0 \Delta x_t + u_t \\ &= \alpha + (\beta_1 - 1)(y_{t-1} - \lambda x_{t-1}) + \gamma_0 \Delta x_t + u_t, \end{aligned}$$

with $\lambda = \frac{\gamma_0 + \gamma_1}{1 - \beta_1}$, the long-run multiplier.

(Davidson & MacKinnon 1993: 683; Greene 2003: 579)

Error correction model

$$\Delta y_t = \alpha + (\beta_1 - 1)(y_{t-1} - \lambda x_{t-1}) + \gamma_0 \Delta x_t + u_t$$

- $\Delta y_t = \alpha + \gamma_0 \Delta x_t + u_t$ is the **equilibrium relationship**;
- $(\beta_1 - 1)(y_{t-1} - \lambda x_{t-1})$ the **equilibrium error**;
- $y_{t-1} - \lambda x_{t-1}$ measures how much y_{t-1} deviates from its long-run equilibrium with x;
- $\beta_1 - 1$ is the amount of adjustment in time y_t as a result of this disequilibrium.

(Davidson & MacKinnon 1993: 683; Greene 2003: 579)

Dynamic models

"The key fact to remember when attempting to specify dynamic regression models is that there are generally a great many a priori plausible ways to do so. It is a serious mistake to limit attention to one particular type of model, such as (...) partial adjustment models. (...) the ADL(p,q) family of models will often provide a good place to start."

I.e. do not just throw in y_{t-1} as a variable to fix time dependence.

(Davison & MacKinnon 1993: 684)

Exercise: unemployment

Using `unemployment.dta`, study the extend to which unemployment rates can be blamed on the party affiliation of the president in office.

Inspired by (Hibbs 1977)

Integrated

If $E(y_t)$, $Var(y_t)$ and $Cov(y_t, y_{t-k})$ converge to limits μ^*, σ^{*2} and ρ_k^*, respectively, as $t \to \infty$, then the process is called **asymptotically stationary**, or **integrated of order zero**, or $I(0)$.

A stationary process is thus $I(0)$, but an $I(0)$ process not necessarily stationary.

(Harrison 2009: 40)

Integrated

We can also have higher orders of integration, such as $I(1)$ or $I(2)$. $I(1)$ means that the series will be $I(0)$ when taking first differences. E.g.

$$\begin{aligned} y_t &= \delta + y_{t-1} + \varepsilon_t \\ y_t - y_{t-1} &= \delta + y_{t-1} + \varepsilon_t - y_{t-1} \\ \Delta y_t &= \delta + \varepsilon_t, \end{aligned}$$

which is a stationary process.

Example: demand for money

According to economic theory, demand for money m is proportional to price levels p (if money is worth less, people want to hold more in their accounts) and as real income y and transactions increase, people will want to hold more. The cost of holding money is based on the interest rate r.

Equilibrium: $m = py^{\beta_y} e^{\beta_r r}$.

(Enders 2004: 320)

$$m_t = p_t^{\beta_p} y_t^{\beta_y} e^{\beta_r r_t} \varepsilon_t^*$$
$$\log m_t = \beta_0 + \beta_p \log p_t + \beta_y \log y_t + \beta_r r_t + \varepsilon_t,$$

with theory implying $\beta_p = 1$, $\beta_y > 0$, $\beta_r < 0$ and ε_t stationary.

p_t, y_t and r_t are all taken to be nonstationary $I(1)$ variables.

For ε_t to be stationary, $m_t - \beta_0 - \beta_p \log p_t - \beta_y \log y_t - \beta_r r_t$ has to be stationary. Thus the evolvement over

Cointegration

A set of independent variables $\mathbf{X}$ are said to be cointegrated of order d, b, $\mathbf{X} \sim CI(d, b)$, iff

- all variables in $\mathbf{X}$ are $I(d)$,
- and there exists a vector $\boldsymbol{\beta}$ such that $\mathbf{X}\boldsymbol{\beta}$ is $I(d-b)$, $b > 0$.

The most common scenario is $CI(1, 1)$, such as the example of money demand.

Taking this example,

- $\boldsymbol{\beta}$ is the **cointegrating vector**,
- ε_t is the **equilibrium error**.

(Enders 2004: 322)

Cointegration

- If **x** and **y** are $\sim CI(1,1)$, regressing **y** on **x** is **super-consistent**.
- If **x** and **y** are each $\sim I(1)$, but not $\sim CI(1,1)$, regressing **y** on **x** is spurious.

Determining whether the two are really cointegrated is thus crucial.

Testing for cointegration

Engle-Granger testing procedure (for $\sim CI(1,1)$):

1. Use ADF-test to determine order of integration of **x** and **y** - must be both $\sim I(1)$.
2. Regress **y** on **x**: $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$.
   - If the two variables are cointegrated, ε_t should be $\sim I(0)$.
   - We use the (A)DF-test on $\Delta e_t = \alpha e_{t-1} + u_t$.
   - Because e is an estimate, the usual (A)DF tables do not apply. The corrected version is called the **(augmented) Engle-Granger, or (A)EG, test**.

     The Phillips-Perron test is similar to the Engle-Granger version, but because it is resistant to heteroscedastic and autocorrelated residuals, it has slightly weaker power. This test is implemented in R, while AEG is not.

(Enders 2004: 335-339; Davidson & MacKinnon 1993: 720-722)

Exercise: unemployment

Using `unemployment.dta`, regress *unemployment* on the log of the ratio of *money* and *deflator*, the log of *purchases*, the log of *exports*, and *year*. Test whether these variables are cointegrated.

Cointegration and error correction

When two variables are cointegrated, there is a long-term equilibrium relation between the two, but on the short term there might be deviations.

We can thus expect that there is an adjustment at time t for any deviation from equilibrium at $t-1$ towards the equilibrium relationship. Thus cointegration and error correction are closely related.

$$\Delta y_t = \alpha + (\beta_1 - 1)(y_{t-1} - \lambda x_{t-1}) + \gamma_0 \Delta x_t + u_t$$
$$\Delta y_t = \alpha + (\beta_1 - 1)\varepsilon_{t-1} + \gamma_0 \Delta x_t + u_t$$

(Gujarati & Porter 2009: 764-765)

Cointegration and error correction

Taking the same example of money demand:

$$\log m_t = \beta_0 + \beta_p \log p_t + \beta_y \log y_t + \beta_r r_t + \varepsilon_t$$

In error correction formulation:

$$\Delta \log m_t = \alpha + (\beta_1 - 1)\hat{\varepsilon}_{t-1} + \gamma_p \Delta \log p_t + \gamma_y \Delta \log y_t + \gamma_r \Delta r_t + u_t$$

Thus, $(\hat{\beta}_1 - 1)$ is the estimated short term adjustment to disequilibrium; $\hat{\gamma}_y$ the short-term elacticity of y on m and $\hat{\beta}_y$ the long-run elasticity.

(Gujarati & Porter 2009: 763-765)

Estimation

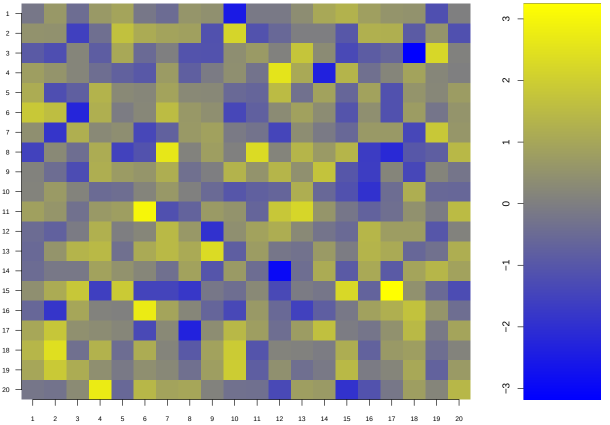
- As long as the errors are not autocorrelated, OLS is consistent and efficients.
- If the errors have an MA(q) component, maximum likelihood is generally required. In R: arima().
- If any of the variables is not stationary, the model needs to be reformulated first.

Exercise: unemployment

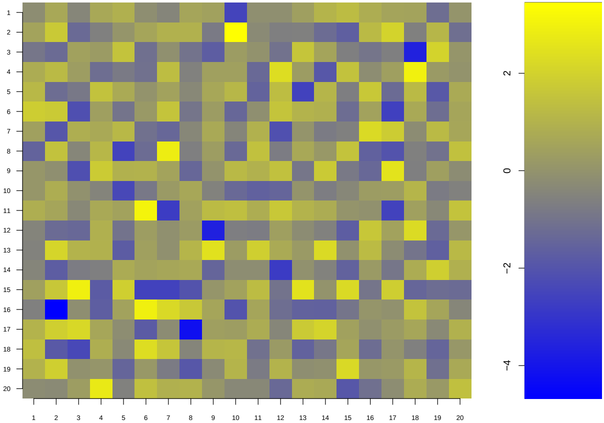
Using `unemployment.dta`, regress *unemployment* on the log of the ratio of *money* and *deflator*, the log of *purchases*, the log of *exports*, and *year*.

Calculate the short-term and long-term elasticities of the effect of exports on unemployment.

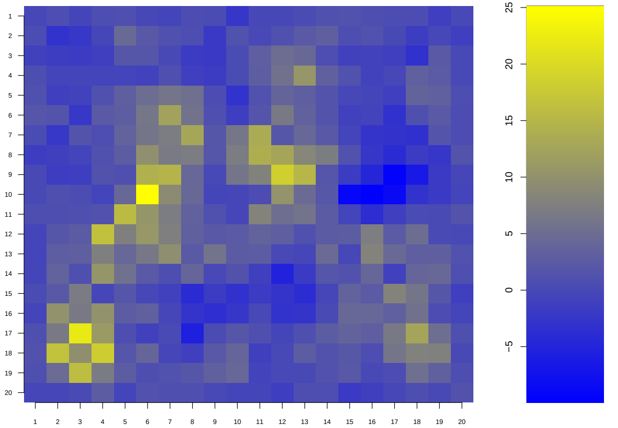
No spatial autocorrelation



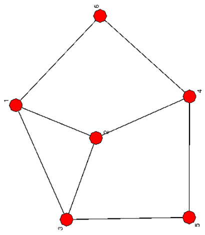
Negative spatial autocorrelation



Positive spatial autocorrelation



Connection matrix



$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\tilde{\mathbf{W}} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

Spatial processes

Spatial autocorrelation has processes somewhat analogous to serial autocorrelation.

Spatial error process: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}$.

Spatial lag process: $\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

(Anselin 1988)

Moran's I

$$I = \frac{\sum_i \sum_j \tilde{w}_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_i \sum_j \tilde{w}_{ij}} \cdot \frac{n}{\sum_i (x_i - \overline{x})^2} \sim N(\mu_I, \sigma_I^2)$$

$$\mu_I = E(I) = \frac{-1}{n-1}$$
$$\sigma_I^2 = Var(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{S_0^2 (n^2 - 1)},$$

where

$$S_0 = \sum_i \sum_j (w_{ij} + w_{ji}), S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2, S_2 = \sum_i \sum_j (\tilde{w}_{ij} + \tilde{w}_{ji})^2$$

Moran's I

```
library(ape)
Moran.I(y, W)
Moran.I(residuals(lm(y ~ x1 + x2)), W)
```

Moran's I can only be calculated with a known $\mathbf{W}$ matrix. Higher order lags are also possible, e.g. $\mathbf{W}^2$.

Example: democracy

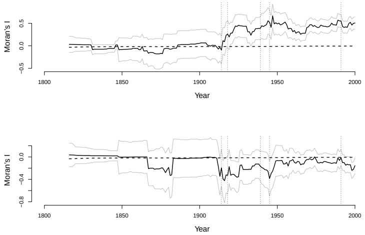


Figure : Spatial clustering, Polity IV dichotomized, $I(y_t)$ and $I(\Delta y_t)$, 1800-2003

(Elkink 2011)

Checking residuals

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\mathbf{e}'\mathbf{W}\mathbf{e}}{\mathbf{e}'\mathbf{e}} \sim N(\mu_I, \sigma_I^2)$$

$$LM_{err} = \frac{n^2(\frac{\mathbf{e}'\mathbf{W}\mathbf{e}}{\mathbf{e}'\mathbf{e}})^2}{tr(\mathbf{W}'\mathbf{W} + \mathbf{W}^2)} \sim \chi^2(1)$$

$$LM_{lag} = \frac{n^2(\frac{\mathbf{e}'\mathbf{W}\mathbf{y}}{\mathbf{e}'\mathbf{e}})^2}{(\mathbf{W}\mathbf{X}\hat{\beta}^{OLS})'\mathbf{M}\mathbf{W}\mathbf{X}\hat{\beta}^{OLS}/\sigma^2 + tr(\mathbf{W}'\mathbf{W} + \mathbf{W}^2)} \sim \chi^2(1),$$

with LM_{err} and LM_{lag} referring to tests for spatial error and spatial lag processes, respectively.

(Anselin & Hudak 1992: 520)

Appendix

MA(1) properties

$$y_t = \mu + \varepsilon_t + \phi\varepsilon_{t-1}$$

$$\begin{aligned} E(y_t) &= E(\mu + \varepsilon_t + \phi\varepsilon_{t-1}) \\ &= E(\mu) + E(\varepsilon_t) + \phi E(\varepsilon_{t-1}) \\ &= \mu + 0 + 0 = \mu \end{aligned}$$

$$\begin{aligned} Var(y_t) &= Var(\mu + \varepsilon_t + \phi\varepsilon_{t-1}) \\ &= Var(\varepsilon_t) + \phi^2 Var(\varepsilon_{t-1}) - 2Cov(\varepsilon_t, \phi\varepsilon_{t-1}) \\ &= \sigma^2 + \phi^2\sigma^2 \end{aligned}$$

AR(1) properties

$$y_t = \delta + \alpha y_{t-1} + \varepsilon_t$$

$$\begin{aligned}
E(y_t) &= E(\delta + \alpha y_{t-1} + \varepsilon_t) \\
&= E(\delta) + \alpha E(y_{t-1}) + E(\varepsilon_t) \\
&= \delta + \alpha E(y_t) + 0 \\
(1-\alpha)E(y_t) &= \delta \\
E(y_t) &= \frac{\delta}{1-\alpha}
\end{aligned}$$

Note that stating that $E(y_t) = E(y_{t-1})$ assumes stationary process!

Markov chains

When the dependent variable is categorical, there are different ways of modeling this (e.g. survival analysis). One approach is that of a Markov chain.

Markov chains

$$Pr(y_t = 1) = \mathcal{F}(y_{t-1}(\mathbf{X}_t\boldsymbol{\beta}) + (1 - y_{t-1})(\mathbf{X}_t\boldsymbol{\alpha}))$$

$\mathcal{F}()$ can be a logistic or probit transformation.

For examples:

- Adam Przeworski & Fernando Limongi (1997), "Modernization: theories and facts." *World Politics* 49:2, 155-183.
- Kristian S. Gleditsch & Michael D. Ward (2006), "Diffusion and the international context of democratization." *International Organization* 60:4, 911-933.

Other topics

A number of important topics have not been covered, among which:

- Vector autoregression (VAR) models, e.g.

$$y_t = \alpha_1 + \beta_1 y_{t-1} + \gamma_1 x_{t-1} + u_{1,t}$$
$$x_t = \alpha_2 + \beta_2 y_{t-1} + \gamma_2 x_{t-1} + u_{2,t}$$

- Autoregressive Conditional Heteroscedasticity models, e.g.

$$\sigma_t^2 = \alpha + \gamma_1 u_{t-1}^2 \qquad \text{ARCH}$$
$$\sigma_t^2 = \alpha + \sum_{i=1}^{p} \gamma_i u_{t-i}^2 + \sum_{i=1}^{q} \delta_i \sigma_{t-i}^2 \qquad \text{GARCH(p,q)}$$