Stationary Time Series, Conditional Heteroscedasticity, Random Walk, Test for a Unit Root, Endogenity, Causality and IV Estimation

Chapter 1

Financial Econometrics

Michael Hauser

WS17/18

Content

- Notation
- Conditional heteroscedasticity, GARCH, estimation, forecasting
- Random walk, test for unit roots
- Granger causality
- Assumptions of the regression model
- Endogeneity, exogeneity
- Instrumental variable, IV, estimator
- Testing for exogeneity:
  Hausman test and Hausman-Wu test
- Appendix: OLS, Weak stationarity, Wold representation, ACF, AR(1), MA(1), tests for WN, prediction of an AR(1)

Notation

$\boldsymbol{X}$, $\boldsymbol{\Pi}$ . . . matrices
$\boldsymbol{x}$, $\boldsymbol{\beta}$, $\boldsymbol{\epsilon}$ . . . column vectors
x, β, ϵ . . . real values, single variables
i, j, t . . . indices
L . . . lag operator. We do not use the backward shift operator B.
WN, RW . . . white noise, random walk

asy . . . asymptotically
df . . . degrees of freedom
e.g. . . . exempli gratia, for example
i.e. . . . id est, that is
i.g. . . . in general
lhs, rhs . . . left, right hand side
rv . . . random variable
wrt . . . with respect to

Conditional heteroscedasticity

GARCH models

Conditional heteroscedasticity

A **stylized fact** of stock returns and other financial return series are periods of high volatility followed by periods of low volatility. The increase and decrease of this volatility pattern can be captured by GARCH models.
The idea is that the approach of new information increases the uncertainty in the market and so the variance. After some while the market participants find to a new consensus and the variance decreases.

We assume - for the beginning - the (log)returns, r_t, are WN

$$r_t = \mu + \epsilon_t \qquad \text{with} \qquad \epsilon_t \sim \mathrm{N}(0, \sigma_t^2)$$

$$\sigma_t^2 = \mathrm{E}(\epsilon_t^2 | \mathrm{I}_{t-1})$$

σ_t^2 is the predictor for ϵ_t^2 given the information at period $(t-1)$. σ_t^2 is the **conditional variance** of ϵ_t^2 given I_{t-1}. $\quad \mathrm{I}_{t-1} = \{\epsilon_{t-1}^2, \epsilon_{t-2}^2, \ldots, \sigma_{t-1}^2, \sigma_{t-2}^2, \ldots\}$.

Conditional heteroscedasticity

Volatility is commonly measured either by

- ϵ_t^2 the 'local' variance (r_t^2 contains the same information), or
- $|\epsilon_t|$, the modulus of ϵ_t.

Generalized autoregressive conditional heteroscedasticity, GARCH

The generalized autoregressive conditional heteroscedasticity, GARCH, model of order $(1,1)$, **GARCH(1,1)**, uses as set of information $\quad I_{t-1} = \{\epsilon^2_{t-1}, \sigma^2_{t-1}\}$

$$\sigma^2_t = a_0 + a_1 \epsilon^2_{t-1} + b_1 \sigma^2_{t-1}$$

The **unconditional variance** of ϵ_t is then
[We set $\mathrm{E}(\sigma^2_t) = \mathrm{E}(\epsilon^2_t) = \sigma^2$. So $\sigma^2 = a_0 + a_1\sigma^2 + b_1\sigma^2$ and we solve wrt σ^2.]

$$\mathrm{E}(\epsilon^2_t) = \frac{a_0}{1 - (a_1 + b_1)}$$

The unconditional variance is constant and exists if

$$a_1 + b_1 < 1$$

Further as σ^2_t is a variance and so positive, a_0, a_1 and b_1 have to be positive.

$$a_0, a_1, b_1 > 0$$

GARCH(r, s)

A GARCH model of order (r, s), **GARCH**(r, s), $r \geq 0, s > 0$, is

$$\sigma_t^2 = a_0 + \sum_1^s a_j \epsilon_{t-j}^2 + \sum_1^r b_j \sigma_{t-j}^2 \quad \text{with} \quad \sum_1^s a_j + \sum_1^r b_j < 1, \;\; a_j, b_j > 0$$

An autoregressive conditional heteroscedasticity, **ARCH**, model of order s is a GARCH$(0, s)$. E.g. ARCH(1) is

$$\sigma_t^2 = a_0 + a_1 \epsilon_{t-1}^2 \quad \text{with} \quad a_1 < 1, \;\; a_0, a_1 > 0$$

Comment 1:
A GARCH(1,0), $\sigma_t^2 = a_0 + b_1 \sigma_{t-1}^2$, is *not useful* as it describes a deterministic decay once a starting value for some past σ_{t-j}^2 is given.
Comment 2:
The order s refers to the number of a_j coefficients, r to the b_j coefficients.

IGARCH(1,1)

An empirically relevant version is the integrated GARCH, **IGARCH**(1,1) model,

$$\sigma_t^2 = a_0 + a_1 \epsilon_{t-1}^2 + b_1 \sigma_{t-1}^2 \qquad \text{with} \qquad a_1 + b_1 = 1, \;\; a_0, a_1, b_1 > 0$$

where the conditional variance exists, but not the unconditional one.

GARCH(1,1): ARMA and ARCH representations

We can define an *innovation in the variance* as

$$\nu_t = \epsilon_t^2 - \mathrm{E}(\epsilon_t^2 | \mathrm{I}_{t-1}) = \epsilon_t^2 - \sigma_t^2$$

Replacing σ_t^2 by $\sigma_t^2 = \epsilon_t^2 - \nu_t$ in the GARCH(1,1) model we obtain

$$\epsilon_t^2 = a_0 + (a_1 + b_1)\epsilon_{t-1}^2 + \nu_t - b_1 \nu_{t-1}$$

This model is an *ARMA(1,1)* for ϵ_t^2. So the ACF of r_t^2 can be inspected for getting an impression of the dynamics. The model is stationary, if $(a_1 + b_1) < 1$.

Recursive substitution of lags of $\sigma_t^2 = a_0 + a_1 \epsilon_{t-1}^2 + b_1 \sigma_{t-1}^2$ in σ_t^2 gives an *infinite ARCH representation* with a geometric decay

$$\sigma_t^2 = \frac{a_0}{1 - b_1} + a_1 \sum_1^\infty b_1^{j-1} \epsilon_{t-j}^2$$

Variants of the GARCH model

- ***t*-GARCH**:
  As the unconditional distribution of r_t can be seen as a mixture of normal distributions with different variances, GARCH is able to model fat tails. However empirically, the reduction in the kurtosis of r_t by a normal GARCH model is often not sufficient.
  A simple solution is to consider an already fat tailed distribution for ϵ_t instead of a normal one. Candidates are e.g. the t-distribution with $df > 2$, or the GED (generalized error distr) with tail parameter $\kappa > 0$.
- **ARCH-in-mean**, **ARCH-M**:
  If market participants are risk avers, they want a higher average return in uncertain periods than in normal periods. So the mean return should be higher when σ_t^2 is high.

$$r_t = \mu_0 + \mu_1 \sigma_t^2 + \epsilon_t \qquad \text{with} \qquad \mu_1 > 0, \ \epsilon_t \sim \mathrm{N}(0, \sigma_t^2)$$

Variants of the GARCH model

- **asymmetric GARCH, threshold GARCH, GJR model**:
  As the assumption that both good and bad new information has the same absolute (symmetric) effect might not hold. A useful variant is

  $$\sigma_t^2 = a_0 + a_1 \epsilon_{t-1}^2 + b_1 \sigma_{t-1}^2 + \gamma D_{t-1} \epsilon_{t-1}^2$$

  where the dummy variable D indicates a positive shock. So

  $$D_{t-1} = 1, \text{ if } \epsilon_{t-1} > 0 \qquad D_{t-1} = 0, \text{ if } \epsilon_{t-1} \leq 0$$

  The contribution of ϵ_{t-1}^2 to σ^2 is

  $$(a_1 + \gamma), \text{ if } \epsilon_{t-1} > 0 \qquad a_1, \text{ if } \epsilon_{t-1} \leq 0$$

  If $\gamma < 0$, negative shocks have a larger impact on future volatility than positive shocks. (GJR stands for Glosten, Jagannathan, Runkle.)

Variants of the GARCH model

- **Exponential GARCH, EGARCH**:
  By modeling the log of the conditional variance, $\log(\sigma_t^2)$, the EGARCH guarantees positive variances, independent on the choice of the parameters. It can be formulated also in an asymmetric way ($\gamma \neq 0$).

  $$\log(\sigma_t^2) = a_0 + a_1 \frac{|\epsilon_{t-1}|}{\sigma_{t-1}^2} + b_1 \log(\sigma_{t-1}^2) + \gamma \frac{\epsilon_{t-1}}{\sigma_{t-1}^2}$$

  If $\gamma < 0$, positive shocks generate less volatility than negative shocks.

Estimation of the GARCH(r, s) model

Say r_t can be described more generally as

$$r_t = \mathbf{x}_t'\boldsymbol{\theta} + \epsilon_t$$

with vector $\mathbf{x_t}'$ of some third variables, $\mathbf{x}_t' = (x_{1t}, \ldots, x_{mt})'$, possibly including lagged r_t's, seasonal dummies, etc.

ϵ_t with conditional heteroscedasticity can be written as

$$\epsilon_t = \sigma_t \xi_t \quad \text{with} \quad \xi_t \text{ iid } \mathrm{N}(0, 1)$$

σ_t^2 has the form $\sigma_t^2 = a_0 + a_1\epsilon_{t-1}^2 + \ldots + a_s\,\epsilon_{t-s}^2 + b_1\sigma_{t-1}^2 + \ldots + b_r\sigma_{t-r}^2$.

So the conditional density of $r_t|\mathbf{x}_t, \mathrm{I}_{t-1}$ is given by

$$f_{\mathrm{N}}(r_t|\mathbf{x}_t, \mathrm{I}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp(-\frac{1}{2}\frac{\epsilon_t^2}{\sigma_t^2}), \quad t = \max(r, s) + 1, \ldots, n$$

Comments: $\mathrm{V}(\sigma_t\xi_t|\mathbf{x}_t, \mathrm{I}_{t-1}) = \sigma_t^2\mathrm{V}(\xi_t|\mathbf{x}_t, \mathrm{I}_{t-1}) = \sigma_t^2$,

starting values $\sigma_1^2, \ldots, \sigma_r^2$ appropriately chosen,

$\epsilon_t = \epsilon_t(\boldsymbol{\theta}), \quad \sigma_t^2 = \sigma_t^2(a_0, a_1, \ldots, a_s, b_1, \ldots, b_r)$

Estimation of the GARCH model

The ML, maximum likelihood, estimator of the parameter vector $\boldsymbol{\theta}, a_0, a_1, \ldots, a_s$, $b_1, \ldots, b_r$ is given by

$$\max_{\boldsymbol{\theta}, a_0, a_1, \ldots, a_s, b_1, \ldots, b_r} \prod_{\max(r,s)+1}^{n} f_{\mathrm{N}}(r_t | \mathbf{x}_t, \mathrm{I}_{t-1})$$

as the ϵ's are uncorrelated.

The estimates are asy normal distributed. Standard t-tests, etc. apply.

Note the requirement of uncorrelated ϵ's. In case autocorrelation in the returns r_t is ignored, the model is misspecified. And, you will detect ARCH effects although they might not exist. So in a first step, model the level of r_t by ARMA, e.g., and then fit GARCH models to the residuals.

Remark: If r_t is autocorrelated, so also r_t^2 is autocorrelated i.g.

Forecasting a GARCH(1,1) process

Using $\sigma^2 = a_0/(1-(a_1+b_1))$ in $\sigma_t^2 = a_0 + a_1\epsilon_{t-1}^2 + b_1\sigma_{t-1}^2$ we rewrite the model as

$$\sigma_t^2 - \sigma^2 = a_1[\epsilon_{t-1}^2 - \sigma^2] + b_1[\sigma_{t-1}^2 - \sigma^2]$$

So the 1-step ahead forecast $\mathrm{E}(\epsilon_{t+1}^2|\mathrm{I}_t)$ is

$$\sigma_{t+1|t}^2 = (\sigma_{t+1}^2) = \mathrm{E}(\epsilon_{t+1}^2|\mathrm{I}_t) = \sigma^2 + a_1[\epsilon_t^2 - \sigma^2] + b_1[\sigma_t^2 - \sigma^2]$$

with the 1-step ahead forecast error $\epsilon_{t+1}^2 - \sigma_{t+1|t}^2 = \nu_{t+1}$.

The h-step ahead forecast is, $h \geq 2$,

[replacing both ϵ_t^2, σ_t^2 by their $(h-1)$-step ahead forecast $\sigma_{t+h-1|t}^2$]

$$\sigma_{t+h|t}^2 = \sigma^2 + (a_1+b_1)[\sigma_{t+h-1|t}^2 - \sigma^2] =$$

$$= \sigma^2 + (a_1+b_1)^{h-1}[a_1(\epsilon_t^2 - \sigma^2) + b_1(\sigma_t^2 - \sigma^2)]$$

Random walk and unit root test

Random walk, I(1), Dickey Fuller test

Random walk, RW

A process $y_t = \alpha y_{t-1} + \epsilon_t$ with $\alpha = 1$ is called random walk.

$$y_t = y_{t-1} + \epsilon_t \quad \text{with} \quad \epsilon_t \text{ WN}$$

Taking the variances on both sides gives $\mathrm{V}(y_t) = \mathrm{V}(y_{t-1}) + \sigma^2$. This has only a solution $\mathrm{V}(y)$, if $\sigma^2 = 0$. So *no unconditional variance of y_t exists*.

Starting the process at $t = 0$ its explicit form is $y_t = y_0 + \sum_1^t \epsilon_j$.
The **conditional expectation** and **conditional variance** are

$$\mathrm{E}(y_t|y_0) = y_0, \qquad \mathrm{V}(y_t|y_0) = t\,\sigma^2$$

The conditional variance of a RW increases with t. The process is **nonstationary**.

The associated characteristic polynomial $1 - z = 0$ [as $(1 - \mathrm{L})y_t = \epsilon_t$] has a root on the unit circle, $|z| = 1$. Moreover it has a **unit root**.

$$z = 1$$

Random walks

There are 3 standard types of a random walk. ϵ_t is WN.

- Random walk (without a drift):

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

- Random walk with a drift. c is the drift parameter.

$$y_t = c + y_{t-1} + \epsilon_t \quad \text{or} \quad y_t = y_0 + c\,t + \sum_1^t \epsilon_j$$

- Random walk with drift and trend.

$$y_t = c + b\,t + y_{t-1} + \epsilon_t$$

ARIMA(1,1,1)

Say we have a process $y_t = 1.2y_{t-1} - 0.2y_{t-2} + \epsilon_t - 0.5\epsilon_{t-1}$ written in lag polynomials $(1 - 1.2\mathrm{L} + 0.2\mathrm{L}^2)y_t = (1 - 0.5\mathrm{L})\epsilon_t$. The characteristic AR polynomial $1 - 1.2z + 0.2z^2 = 0$ has the roots $z_1 = 1/0.2$, $z_2 = 1$. So we can factorize as

$$(1 - 0.2\mathrm{L})(1 - \mathrm{L})y_t = (1 - 0.5\mathrm{L})\epsilon_t$$

Replacing $(1 - \mathrm{L})y_t$ by $\Delta y_t = y_t - y_{t-1}$ the process may be written as

$$(1 - 0.2\mathrm{L})\Delta y_t = (1 - 0.5\mathrm{L})\epsilon_t \quad \text{or} \quad \Delta y_t = 0.2\Delta y_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}$$

So Δy_t, the **differenced** y_t, is a stationary ARMA(1,1) process. The original process y_t (the level) is integrated of order 1, an ARIMA(1,1,1).

Taking simple differences removes the nonstationarity and brings us in the world of stationary processes, where elaborate techniques are available.

Sometimes differencing twice is necessary.
Root for $(1 - \alpha\mathrm{L})$: $1 - \alpha z = 0$ with $z_1 = (1/\alpha)$. So $1 - (1/z_1)z = 0$.

Test for a unit root, Dickey-Fuller test

We start with the model

$$y_t = c + \alpha y_{t-1} + \epsilon_t, \qquad \epsilon_t \ WN$$

and want to know whether

- $\alpha = 1$, the process is a random walk, or
- $\alpha < 1$, the process is stationary.

Estimation of α with OLS gives consistent estimates in both cases. For RW's they are even super-consistent, and converge faster than with $\sqrt{n}$.

For the Dickey-Fuller test we subtract on both sides y_{t-1} and get

$$\Delta y_t = c + (\alpha - 1)y_{t-1} + \epsilon_t.$$

We estimate with OLS and calculate the standard t-statistic for $(\alpha - 1)$

$$\tau = \frac{\hat{\alpha} - 1}{se(\hat{\alpha})}.$$

Dickey-Fuller test, DF test

Under the null hypothesis

$$H_0 : \alpha = 1 \quad \text{or} \quad (\alpha - 1) = 0$$

the

$$\tau\text{-statistic}$$

is distributed according to the distribution tabulated in Dickey-Fuller(1976).
It has fatter tails than the *t*-distribution and is somewhat skewed to the left.

The distribution depends on the sample size, and on the type of the RW model, without drift, with drift, with a linear trend:

$$\text{A} : y_t = y_{t-1} + \epsilon_t, \quad \text{B} : y_t = c + y_{t-1} + \epsilon_t, \quad \text{C} : y_t = c + b\,t + y_{t-1} + \epsilon_t$$

The alternative hypothesis is

$$H_A : \alpha < 1 \quad \text{or} \quad (\alpha - 1) < 0$$

Dickey-Fuller test, deterministic trend

Remark:
Model C, $y_t = c + b\,t + y_{t-1} + \epsilon_t$, is, used to distinguish between a RW and a deterministic (linear) trend.

$$\Delta y_t = c + b\,t + (\alpha - 1)y_{t-1} + \epsilon_t$$

H_0: If the DF test leads to an acceptance of the H_0 of a RW, then y_t is a RW, and not a deterministic trend. (It might be that the conditional mean has a quadratic trend.)

H_A : If the DF test rejects the H_0 of a RW, then y_t does not obey to a RW. If b is significant, y_t might be modeled more adequately by a deterministic trend.

The augmented Dickey-Fuller test

In case ϵ_t is not WN, but (stationary and) autocorrelated, the Dickey-Fuller test is *not reliable*.

Therefore we estimate the model augmented with lagged Δy_t's to capture this autocorrelation, and obtain approximately uncorrelated residuals.

$$\Delta y_t = c + (\alpha - 1) y_{t-1} + \sum_{1}^{p} \phi_j \Delta y_{t-j} + \epsilon_t$$

The maximal lag order p is chosen automatically by an information criterion of your choice, e.g. AIC, SBC.

Remark: Tests for bubbles ($\alpha > 1$) are found in Phillips, Shi and Yu(2015).

Granger causality

Granger causality

Granger causality is actually a concept for classification of predictability in stationary dynamic (AR(p)-type) models.

Restricted model, model 0:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \epsilon_{0,t}$$

We ask whether a stationary x_t can help to forecast y_t. We consider the model augmented by lagged x's:

Unrestricted model, model 1:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \beta_1 x_{t-1} + \ldots + \beta_q x_{t-q} + \epsilon_{1,t}$$

X **Granger-causes** Y, if at least one of the $\beta_j \neq 0$.

Testing via LR-test

The hypothesis of no Granger causation of X wrt Y

$$H_0: \ \beta_1 = \ldots = \beta_q = 0$$

can be tested for fixed p, q via a LR-test:

$$n \log(\frac{s_0^2}{s_1^2}) \sim \chi^2(q)$$

The test statistic is distributed χ^2 with q degrees of freedom.

$\log(\sigma_0^2/\sigma_1^2)$ may be interpreted as the *strength of the causality*.

Testing via F-test, instantaneous causality

Or we test via a Wald F-statistic

$$\frac{(RSS_0 - RSS_1)/q}{RSS_1/(n-p-q)} \sim F(q, n-p-q)$$

- If in the augmented model also the *current* x_t plays a role, then this is called **instantaneous causality** between X and Y.
- *Both* Y **Granger-causes** X and X **Granger-causes** Y might exist. This is called **feedback relationship.**

[RSS . . . residual sum of squares]

Example, deficiencies

Example: Daily returns of various European stock indices were slightly Granger-caused by the New York stock exchange before 9/11 2001. However during the months after, adjustment speeded up, instantaneous causality rose and lagged reaction vanished.

Granger causation is subject to the same deficiencies as the bivariate correlation coefficient:

- Causation may *vanish* when including a third variable (spurious correlation), or
- Causation may *turn out* only after adding a third variable.

The space within causation is searched for is relevant.

Assumptions of classical regression model

The linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

The dependent variable $\boldsymbol{y}$ is $(n \times 1)$, a column of length n.
The regressor matrix $\boldsymbol{X}$ is $(n \times K)$, $\boldsymbol{X} = [x_{tk}]_{n \times K}$.
The parameter vector $\boldsymbol{\beta}$ is $(K \times 1)$ and
the error term $\boldsymbol{\epsilon}$ is $(n \times 1)$.

More explicitly

$$\boldsymbol{y} = \boldsymbol{x}_1 \beta_1 + \ldots + \boldsymbol{x}_K \beta_K + \boldsymbol{\epsilon}$$

$\boldsymbol{x}_k$ is the k-th column in $\boldsymbol{X}$, $\boldsymbol{x}_k = [x_{\cdot k}]_{n \times 1}$.

In terms of single time points or individuals

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta} + \epsilon_t$$

where $\boldsymbol{x}_t'$, $\boldsymbol{x}_i'$, are the t-th/i-th row in $\boldsymbol{X}$.

Assumptions: review

A0. True model: We know the true model and the model in question is specified correctly in the given form.

A1. Linear model: The model is linear. (I.e. linear in the parameters.)

A2. Full rank of $\boldsymbol{X}$: No variable can be expressed as linear combination of the others. All parameters are identified. No multicollinearity.

As the regressors are stochastic the condition is:
$\mathrm{plim}(\boldsymbol{X}'\boldsymbol{X}/n) = \boldsymbol{Q}$ is finite, constant and non singular.
If the x-variables have all mean zero, $\boldsymbol{Q}$ is the asymptotic covariance matrix.

In the special case of a deterministic $\boldsymbol{X}$ A2 reduces to the full rank, nonsingularity, respectively, of $(\boldsymbol{X}'\boldsymbol{X})$.

Assumptions: review

A3. Exogeneity of the explanatory variables:
The explanatory variables at any point of time, i, are independent with the current, and all past and future errors.
$E[\epsilon_i | x_{j1}, ..., x_{jK}] = 0 \qquad i, j = 1, \ldots, n$.

A4. Homoscedasticity and uncorrelated errors:
The disturbances ϵ_t have finite and constant variance, σ^2.
They are uncorrelated with each other.

A5. Exogenously generated data: The generating process of the x's is outside the model. The analysis is done conditionally on the observed $\boldsymbol{X}$.

A6. Normal distribution: The disturbances are (approximately) normally distributed (for convenience). $\boldsymbol{y}$ is multivariate normal conditional on $\boldsymbol{X}$,
$\boldsymbol{y}|\boldsymbol{X} \sim \mathrm{N}(.,.)$

Endogeneity

Endogeneity: definition

Endogeneity exists if explanatory variables are correlated with the disturbance term. We start with

$$\boldsymbol{y} = \boldsymbol{X}\beta + \epsilon$$

and estimate by OLS, with $\boldsymbol{b} = \hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$. Then $\boldsymbol{b}$ may be written as

$$\boldsymbol{b} = \beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\epsilon$$

If $\mathrm{E}(\boldsymbol{X}'\epsilon)$ or $\mathrm{plim}(\boldsymbol{X}'\epsilon/n)$ is *not zero*, then the estimates are biased or inconsistent.

There are several models where endogeneity arises:

- The errors-in-the-variables model
- Simultaneity
- Dynamic regression

Endogeneity: the errors-in-the-(explanotary)variables model

The true model is: $y = \beta x + \epsilon$

However, we observe only x^o, i.e. x with a **measurement** (random) **error** η, which we cannot control for: $x^o = x + \eta$

The model we estimate is: $y = \beta x^o + \epsilon^o$
So $\epsilon^o = \epsilon - \beta\eta$, and $\mathrm{Cov}(x^o, \epsilon^o) \neq 0$.

$$\mathrm{Cov}(x^o, \epsilon^o) = \mathrm{Cov}(x + \eta, \epsilon) - \beta\,\mathrm{Cov}(x + \eta, \eta) \neq 0$$

The first term can be assumed to be zero, $\mathrm{Corr}(\epsilon, \eta) = 0$, the second one is different from zero.

The measurement error causes inconsistency of the estimates. The bias in β does not vanish even with $n \to \infty$.

Endogeneity: simultaneity

Simultaneity is a general phenomenon. It says that individual decisions in (economic) systems are interdependent. Single independent decisions can hardly be separated out.

Example: In macroeconomics the simple consumption function describes the relationship between private consumption, C, and disposable income, Y.

$$C = \alpha + \beta Y + \epsilon$$

However, income is generated by consumption expenditure, and investment and government consumption, etc. Say, the latter is aggregated and denoted by Z.

$$Y = C + Z$$

Both equations give a simultaneous equation system.

Endogeneity: simultaneity

LS of the consumption function yields

$$(a\ b)' = (\alpha\ \beta)' + [(\mathbf{1}\ \boldsymbol{Y})'(\mathbf{1}\ \boldsymbol{Y})]^{-1}(\mathbf{1}\ \boldsymbol{Y})'\boldsymbol{\epsilon}$$

Since $\quad Y = C + Z$,

$$\mathrm{E}(\mathbf{1}\ \boldsymbol{Y})'\boldsymbol{\epsilon} = \mathrm{E}[\mathbf{1}\ (\boldsymbol{C} + \boldsymbol{Z})]'\boldsymbol{\epsilon} \neq \mathbf{0}$$

$\mathrm{E}(C'\epsilon) \neq 0$ as $\quad C = \alpha + \beta Y + \epsilon, \quad$ and so $(a\ b)'$ is biased.
This is a simple example for the **simultaneous equation bias.**

Example: It is difficult to distinguish empirically between pure univariate expectations and expectations influenced by the overall economic state. Further, consider expectations about the expectations of other subjects,. . ..

Strict / weak exogeneity

Assumption A3 says that current, past and future disturbances are uncorrelated with current explanatory variables.

$$E[\epsilon_i | x_{j1}, ..., x_{jK}] = 0 \quad i, j = 1, \ldots, n$$

or

$$E[x_{jk}\, \epsilon_i] = 0 \quad i, j = 1, \ldots, n \ \text{ for all } k$$

We say the x variables are **strictly exogenous.**

For **weak exogeneity** only

$$E[x_{ik}\, \epsilon_i] = 0 \quad i = 1, \ldots, n \ \text{ for all } k$$

is required.

Predetermined variables, dynamic regression

Variable x is called **predetermined** if

$$\mathrm{E}[x_t \, \epsilon_{t+s}] = 0 \quad s > 0$$

Current x_t is uncorrelated with all *future* ϵ's, but possibly correlated with past.

Consider the simple dynamic regression with $|\gamma| < 1$

$$y_t = \gamma y_{t-1} + \epsilon_t$$

Repeated substitution gives

$$y_t = \sum_{i=0}^{t-1} \gamma^i \epsilon_{t-i} + \gamma^t y_0 = \epsilon_t + \gamma \epsilon_{t-1} + \ldots + \gamma^{t-1} \epsilon_1 + \gamma^t y_0$$

y_{t-1} does not dependent on $\epsilon_t, \epsilon_{t+1}, \ldots$. So y_{t-1} is predetermined, and weakly exogenous.

The lack of contemporary correlation of y_t and ϵ_t is essential. If $\mathrm{E}(x_t \, \epsilon_t) = 0$ the consequences of the dynamic dependencies vanish at least asymptotically and we obtain consistent estimates, which might, however, be biased in small samples.

Instrumental variables

Instrumental variables, IV

A solution for the edogeneity problem (biases/inconsistencies) are IV estimators. **Instrumental variables**, short **instruments**, are essentially *exogenous* variables - they do not correlate with the disturbances by assumption - used as *proxies* for endogenous explanatories.

- The art is to find variables which are good proxies (and do not show endogeneity), so that the fit of the regression will become acceptable.
- Sometimes the loss of information due to the use of IVs (bad proxies) is larger than the bias induced by endogenity. Then it is preferable to stay with the simple LS estimator.

IV estimation: a 2-step procedure

Say the model is

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$$

$\boldsymbol{X}_1$ contains the variables leading to the endogeneity problem.

- We choose *at least* as many instruments, the columns in $\boldsymbol{Z}$, as there are variables in $\boldsymbol{X}_1$,
- regress (LS) $\boldsymbol{X}_1$ on $\boldsymbol{Z}$: $\boldsymbol{X}_1 = \boldsymbol{Z}\gamma + \boldsymbol{\eta}$.
  We get $\widehat{\boldsymbol{X}}_1 = \boldsymbol{Z}\widehat{\gamma}$, and
- replace $\boldsymbol{X}_1$ in the original model by $\widehat{\boldsymbol{X}}_1$:

$$\boldsymbol{y} = \widehat{\boldsymbol{X}}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$$

- The resulting LS estimate $\boldsymbol{b} = \boldsymbol{b}_{IV}$ is the instrumental variable estimator.

IV estimator: more formally

Formally we relax assumption A3 to AI3 (weak exogeneity) with

$$\textbf{AI 3.} \quad E(\epsilon_i | x_{i,1}, \ldots, x_{i,K}) = 0, \ \ i = 1, \ldots, n$$

Only the contemporaneous correlation has to vanish.

We show that the estimate $\boldsymbol{b}_{IV}$ is **asymptotically normal**.
For notational convenience the model is the following

$$\boldsymbol{y} = \boldsymbol{X}\beta + \epsilon, \quad \boldsymbol{X}\ (n \times K),\ \boldsymbol{Z}\ (n \times K)$$

There are - say K - instruments in $\boldsymbol{Z}$, so that $\boldsymbol{X}'\boldsymbol{Z}$ has rank K. LS of $\boldsymbol{X}$ on $\boldsymbol{Z}$ gives with

$$\boldsymbol{X} = \boldsymbol{Z}\gamma + \boldsymbol{\eta}$$

γ is $(K \times K)$ and $\boldsymbol{\eta}$ $(n \times K)$.

$$\widehat{\gamma} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X} \quad \text{and} \quad \widehat{\boldsymbol{X}} = \boldsymbol{Z}\widehat{\gamma} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$$

IV estimator: more formally

Now we have to estimate by LS

$$\boldsymbol{y} = \widehat{\boldsymbol{X}}\beta + \epsilon$$

$$\begin{aligned}
\boldsymbol{b} = \boldsymbol{b}_{IV} &= (\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1}\widehat{\boldsymbol{X}}'\boldsymbol{y} = \ldots \\
&= [\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}]^{-1}[\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}]'\boldsymbol{y} = \ldots \\
&= (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{y}
\end{aligned}$$

IV estimator: unbiasedness, consistency

We replace $\boldsymbol{y}$ by $\boldsymbol{y} = \boldsymbol{X}\beta + \epsilon$. So

$$\boldsymbol{b}_{IV} = \beta + (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\epsilon$$

For consistency we need that

$\text{plim}(\boldsymbol{Z}'\boldsymbol{Z}/n) = \boldsymbol{Q}_{ZZ}$ a positive definite matrix

$\text{plim}(\boldsymbol{Z}'\boldsymbol{X}/n) = \boldsymbol{Q}_{ZX} = \boldsymbol{Q}'_{XZ}$ a nonzero matrix

$\text{plim}(\boldsymbol{X}'\boldsymbol{X}/n) = \boldsymbol{Q}_{X}$ a positive definite matrix and

$\text{plim}(\boldsymbol{Z}'\epsilon/n) = 0$ as $\boldsymbol{Z}$ are valid instruments.

Then $\text{plim}\,\boldsymbol{b}_{IV} = \beta + \text{plim}(\boldsymbol{Z}'\boldsymbol{X}/n)^{-1}(\boldsymbol{Z}'\epsilon/n) = \beta + \boldsymbol{Q}_{ZX}^{-1}\,\text{plim}(\boldsymbol{Z}'\epsilon/n)$ and so

$\text{plim}\,\boldsymbol{b}_{IV} = \beta$.

The IV estimator is *consistent*.

IV estimator: asy normality

Application of a CLT (central limit theorem) gives the asymptotic distribution of $\boldsymbol{b}_{IV}$. The expectation is obtained by using (weak) AI3 $\mathrm{E}(\epsilon_t|\boldsymbol{z}_t') = 0$, the variance by considering $(\boldsymbol{Z}'\epsilon/\sqrt{n})$. Under very general conditions – moments up to order $(2+\epsilon)$ exist, ϵ weakly dependent – holds

$$\boldsymbol{Z}'\epsilon/\sqrt{n} \overset{d}{\to} \boldsymbol{N}[\mathbf{0}, \sigma^2 \boldsymbol{Q}_{ZZ}]$$

Remark: If the disturbances are heteroscedastic having covariance matrix $\Omega \neq \sigma^2 \boldsymbol{I}$, the asymptotic variance of $(\boldsymbol{Z}'\epsilon/\sqrt{n})$ would be $\mathrm{plim}(\boldsymbol{Z}'\Omega\boldsymbol{Z}/n)$ instead.

The IV estimator is asymptotically distributed as

$$\boldsymbol{b}_{IV} \sim \boldsymbol{N}[\beta, (\sigma^2/n)\boldsymbol{Q}_{ZX}^{-1}\boldsymbol{Q}_{ZZ}\boldsymbol{Q}_{XZ}^{-1}]$$

Remark: asyVar$(\boldsymbol{Z}'\epsilon/\sqrt{n}) = \mathrm{plim}(\boldsymbol{Z}'\epsilon/\sqrt{n})(\boldsymbol{Z}'\epsilon/\sqrt{n})' = \mathrm{plim}(\boldsymbol{Z}'\epsilon\epsilon'\boldsymbol{Z})/n = \sigma^2\boldsymbol{Q}_{ZZ}$

Hausman test: testing for endogeneity

If there is a suspicion of endogeneity about the $\boldsymbol{X}$'s, we want to check for it. The Hausman test offers a possibility.
We compare the estimates

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b}_{LS} + \widehat{\epsilon}_{t,LS}$$

and

$$\boldsymbol{y} = \widehat{\boldsymbol{X}}\boldsymbol{b}_{IV} + \widehat{\epsilon}_{t,IV}$$

Hausman test: testing for endogeneity

The idea is that *under the null hypothesis of no endogeneity* both OLS and IV are consistent, though IV is less efficient (because we use proxies as instruments).

$$Asy.Var[\boldsymbol{b}_{IV}] - Asy.Var[\boldsymbol{b}_{LS}] \ldots \text{nonnegative definite}$$

Further, under the null should hold

$$\mathrm{E}(\boldsymbol{d}) = \mathrm{E}(\boldsymbol{b}_{IV} - \boldsymbol{b}_{LS}) = \mathbf{0}$$

However *under the alternative of endogenity* LS is biased/inconsistent, and the estimates differ: $\mathrm{E}(\boldsymbol{d}) \neq \mathbf{0}$.
A Wald statistic based thereon is

$$H = \boldsymbol{d}'\{Est.Asy.Var[\boldsymbol{d}]\}^{-1}\boldsymbol{d}$$

with $Est.Asy.Var[\boldsymbol{d}]$ the estimated asymptotic variance of $\boldsymbol{d}$.

Hausman test: testing for endogeneity

$Est.Asy.Var[\boldsymbol{d}]$ depends on the variance of $\boldsymbol{b}_{IV}$, $\boldsymbol{b}_{LS}$ and their covariances.
However, the covariance between an efficient estimator, $\boldsymbol{b}_E$, and its difference from an inefficient one, $\boldsymbol{b}_I$, is zero. (without proof)

$$Cov(\boldsymbol{b}_E, \boldsymbol{b}_I - \boldsymbol{b}_E) = Cov(\boldsymbol{b}_E, \boldsymbol{b}_I) - Cov(\boldsymbol{b}_E, \boldsymbol{b}_E) = 0$$

So $\quad Cov(\boldsymbol{b}_E, \boldsymbol{b}_I) = Var(\boldsymbol{b}_E)$. $\quad$ Applied to $\boldsymbol{b}_{IV}$ and $\boldsymbol{b}_{LS}$:

$$Asy.Var(\boldsymbol{b}_{IV} - \boldsymbol{b}_{LS}) = Asy.Var(\boldsymbol{b}_{IV}) - Asy.Var(\boldsymbol{b}_{LS})$$

This reduces H under the null of no endogeneity to

$$H = \frac{1}{\widehat{\sigma}^2} \boldsymbol{d}'[(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1} - (\boldsymbol{X}'\boldsymbol{X})^{-1}]^{-1}\boldsymbol{d} \ \sim \ \chi^2(K^*)$$

where $K^* = K - K_0$, and K_0 is the number of explanatories which are not under consideration wrt endogeneity. K^* ... no. of possibly endogenous variables.
[For $\boldsymbol{y} = \boldsymbol{X}\beta + \epsilon$ the variance of $\boldsymbol{b}_{LS}$ is $\mathrm{V}(\boldsymbol{b}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$]

Hausman test

The Hausman test is more generally applicable.

- If we compare an efficient and an inefficient estimator, which are both consistent, under the null.
- And, under the alternative the efficient one becomes inconsistent, and the inefficient remains consistent.

Hausman-Wu test: testing for endogeneity

A simple related test for endogeneity is the Hausman-Wu test. *Under the null* we consider the model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Under the alternative we consider the augmented model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \widehat{\boldsymbol{X}}^{*}\gamma + \boldsymbol{\epsilon}^{*}$$

where $\widehat{X}^{*}$ are the K^{*} explanatories under suspicion causing endogeneity approximated by their estimates using the instruments.

The idea is, that under the null hypothesis of no endogeneity, $\widehat{\boldsymbol{X}}^{*}$ represents an irrelevant additional variable, so $\gamma = 0$.

Hausman-Wu test: testing for endogeneity

Under the alternative the null model yields biased estimates.

The test statistics is a common F-test with K^* and $(n - K - K^*)$ degrees of freedom, where the restricted model ($\gamma = 0$) is tested against the unrestricted ($\gamma \neq 0$) one.

Important to note: The test depends essentially on the choice of appropriate instruments.

Exercises, references and appendix

Exercises

Choose 2 out of (2, 3, 4, 5) and 2 of (1G, 6G).

1G Use EViews and `Ex1_1_gspc_garch.wf1`. Find an appropriate EGARCH model for the return series of the S&P 500. For the period 2010-01-01 to 2013-12-31.
First test for a unit root in the stock prices, then specify and estimate the ARMA and GARCH models for the returns separately. Compare standard GARCH, GARCH-M, asymmetric GARCH, *t*-GARCH and EGARCH.
[For estimating ARMA-GARCH models in R see `Ex1_gspc_garch_R.txt` and `Ex1_interest_rates_R.txt`.]

2 Derive
(a) the ARMA(1,1) and
(b) the ARCH(∞) representation
of a GARCH(1,1) model.

Exercises

3 Has the model $y_t = 1.7y_{t-1} - 0.8y_{t-2} + \epsilon_t$ a unit root?

4 Derive the IV estimator and its asymptotic distribution for homoscedastic and uncorrelated errors.

5 Write down the F-statistic for the Hausman-Wu test.

6G Investigate the type and strength of causality in the US and European stock index returns around 9/11. Test for various types of causality: X causes Y, instantaneous causality, and feedback effects (X causes Y, and Y causes X). Use `Ex1_6_Granger_R.txt`.

References

Greene 13.2
Ramanathan 10
Verbeek 8
Phillips, Shi and Jun(2015): Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500, International Economic Review, 1043-77

Appendix

Useful formula for OLS

White noise, Wold representation, ACF, AR(1), MA(1), lag polynomial, roots

Some useful formula for OLS

Model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

The **normal equations** give the 1st order conditions for $\min_{\boldsymbol{\beta}}(\boldsymbol{\epsilon}'\boldsymbol{\epsilon})$

$$(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}$$

LS solution for $\boldsymbol{\beta}$

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

In order to see the (potential) unbiasedness of $\boldsymbol{b}$

$$\boldsymbol{b} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\epsilon}$$

The variance-covariance matrix of the estimate with $Var(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}$

$$Var(\boldsymbol{b}) = \mathrm{E}[(\boldsymbol{b}-\boldsymbol{\beta})(\boldsymbol{b}-\boldsymbol{\beta})'] = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

The residual vector $\boldsymbol{e}$

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b} = \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = [\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y} = \boldsymbol{M}\boldsymbol{y}$$

White noise, WN

A **stochastic process** $\{x_t\}$, $t = \ldots, -1, 0, 1, \ldots$, is a sequence of rv's, where the rv's have common properties, eg. for strictly stationarity with finite dimensional common distributions.

A stochastic process $\{x_t\}$ is a **white noise**, WN,

$$x_t = \epsilon_t$$

if for x_t (and so also for ϵ_t)

- the (unconditional) mean $\mathrm{E}(x_t) = \mathrm{E}(\epsilon_t) = 0$, for all t
- the (unconditional) variance $\mathrm{V}(x_t) = \mathrm{V}(\epsilon_t) = \sigma^2$ fixed (independent of t) and
- the autocorrelations $\mathrm{Corr}(x_t, x_{t-j}) = \mathrm{Corr}(\epsilon_t, \epsilon_{t-j}) = 0$ for $j \neq 0$.

More general, x_t may have a mean, which is different form zero

$$x_t = \mu + \epsilon_t$$

Weak stationarity, ACF

A process $\{x_t\}$ is **weakly stationary**, if

- $\mathrm{E}(x_t) = \mu$ fixed, independent of t,
- $\mathrm{V}(x_t) = \gamma_0 = \sigma_x^2$ fixed, independent of t
- $\mathrm{Cov}(x_t, x_{t-j}) = \gamma_j$ independent of t, and depends only on the time distance between the both periods t and $(t - j)$, $j \geq 0$.

Weak stationarity refers only to the first two moments of the process. A common distribution is not required.

$\{\gamma_j\}$, $j \geq 0$, is called the **autocovariance function**, ACF, of $\{x_t\}$.

The sample autocovariances are called **correlogram**.

Wold representation theorem

Every weakly stationary process may be represented as a (one-sided) weighted sum of a WN process $\{\epsilon_t\}$

$$x_t = \mu + \sum_0^{\infty} \delta_j \epsilon_{t-j} = \mu + \epsilon_t + \delta_1 \epsilon_{t-1} + \delta_2 \epsilon_{t-2} + \dots$$

with $\sum_0^{\infty} \delta_j^2 < \infty$, (square summable) and normalized with $\delta_0 = 1$.

As for all t $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{V}(\epsilon_t) = \sigma^2$ and $\mathrm{Corr}(\epsilon_t, \epsilon_{t-j}) = 0$ for $j \neq 0$

- $\mathrm{E}(x_t) = \mu + \sum_0^{\infty} \delta_j \mathrm{E}(\epsilon_{t-j}) = \mu$
- $\mathrm{V}(x_t) = \sum_0^{\infty} \delta_j^2 \mathrm{V}(\epsilon_{t-j}) = \sigma^2 \sum_0^{\infty} \delta_j^2 < \infty$ finite, fixed and independent of t.

Remark: $\mathrm{V}(X + Y) = \mathrm{V}(X) + \mathrm{V}(Y)$, if $\mathrm{Corr}(X, Y) = 0$.

AR(1)

An autoregressive process of order 1, AR(1), is given by

$$(x_t - \mu) - \alpha(x_{t-1} - \mu) = \epsilon_t$$

with ϵ_t WN and $|\alpha| < 1$.
Written as regression model

$$x_t = (1 - \alpha)\mu + \alpha x_{t-1} + \epsilon_t = c + \alpha x_{t-1} + \epsilon_t$$

Its Wold representation is

$$x_t = \mu + \sum_{0}^{\infty} \alpha^j \epsilon_{t-j}$$

Its ACF decays geometrically.

$$\mathrm{V}(x_t) = \gamma_0 = \sigma_x^2 = \sigma^2 \sum_{0}^{\infty} \alpha_j^2 = \sigma^2 \frac{1}{1 - \alpha^2}$$

$$\mathrm{Corr}(x_t, x_{t-j}) = \gamma_j / \gamma_0 = \alpha^j$$

Forecasting with AR(1)

For forecasting we take the conditional expectation of x_t based on available information at $(t-1)$, $\mathrm{I}_{t-1} = \{x_{t-1}, x_{t-2}, \ldots, \epsilon_{t-1}, \epsilon_{t-2}, \ldots\}$.

The predictor of an AR(1) process x_t for period t is

$$\mathrm{E}(x_t|\mathrm{I}_{t-1}) = (1-\alpha)\mu + \alpha x_{t-1}$$

The useful information reduces to $\mathrm{I}_{t-1} = \{x_{t-1}\}$.

Contrary, the unconditional expectation is $\mathrm{E}(x_t) = \mu$.

The innovation (stochastics) ϵ_t which is driving the AR(1)

$$x_t = (1-\alpha)\mu + \alpha x_{t-1} + \epsilon_t \qquad \text{is} \qquad \epsilon_t = x_t - \mathrm{E}(x_t|\mathrm{I}_{t-1})$$

AR(1), roots of the characteristic polynomial

The AR(1) can be written by means of the lag polynomial $(1 - \alpha \mathrm{L})$ as

$$(1 - \alpha \mathrm{L})(x_t - \mu) = \epsilon_t$$

The roots of the **characteristic polynomial** are given by

$$1 - \alpha z = 0$$

with the solution $z = 1/\alpha$. The stationarity condition $|\alpha| < 1$ translates for the root to $|z| = |1/\alpha| > 1$. The root lies *outside* the unit circle.

This stationarity condition generalizes to AR(p) processes of higher order p.

$$(1 - \alpha_1 \mathrm{L} - \ldots - \alpha_p \mathrm{L}^p)(x_t - \mu) = \epsilon_t$$

$$1 - \alpha_1 z - \ldots - \alpha_p z^p = 0, \qquad |z| > 1$$

The roots have to lie outside the unit circle.

MA(1)

A moving average process of order 1, MA(1), is given by

$$(x_t - \mu) = \epsilon_t + \beta \epsilon_{t-1}$$

with ϵ_t WN, and $|\beta| < 1$ for invertibility.
(Invertibility guarantees an unique AR representation. MAs are stationary independent of the choice of β.)

Its Wold representation is a finite sum of the ϵ's and is the model itself.

Its ACF vanishes after lag 1.

- $\mathrm{V}(x_t) = \gamma_0 = \sigma_x^2 = (1 + \beta^2)\sigma^2$
- $\mathrm{Corr}(x_t, x_{t-1}) = \gamma_1/\gamma_0 = \beta/(1 + \beta^2) < 1/2$
- $\mathrm{Corr}(x_t, x_{t-j}) = \gamma_j/\gamma_0 = 0 \qquad$ for $j > 1$.

Tests for white noise

The sample autocorrelations are denoted by $\hat{\rho}_j$, $\hat{\rho}_j = \hat{\gamma}_j/\hat{\gamma}_0$. Under WN the sample autocorrelations (for any $j > 0$) are asy distributed as

$$\hat{\rho}_j \overset{a}{\sim} \mathrm{N}(-\frac{1}{n}, \frac{1}{n})$$

So a 95% coverage interval for $\hat{\rho}_j$ under WN is $[-\frac{1}{n} - 1.96\sqrt{1/n},\ -\frac{1}{n} + 1.96\sqrt{1/n}]$.

A standard test for autocorrelation of zero up to a maximal order m, $1 \leq j \leq m$, is the **Box-Pierce test.** Using a correction for the *df*'s it is called **Ljung-Box test.**

$$Q = n(n+2)\sum_{1}^{m} \frac{\hat{\rho}_j^2}{n-j} \ \sim\ \chi^2_m$$

Under the

$$H_0 :\ \rho_1 = \ldots = \rho_m = 0$$

Q follows a chi-square distribution with m degrees of freedom.