covariance between observations

Generalized Least Squares

- So far, we have focused on modeling the ***expected value of the dependent variable*** and calculated its standard error. Now we concentrate on the estimating covariance matrix of the ***distribution of the random component*** and use it to adjust the least squares estimator
- General Least Squares (GLS) allows addressing violations of the assumption that the regression disturbances $\boldsymbol{\varepsilon}$ are independently identical distributed $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \cdot \mathbf{I})$.
- If we assume that the disturbances are normally distributed with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \cdot \boldsymbol{\Omega})$ then regression models under the violation of i.i.d. can be estimated by identifying $\boldsymbol{\Omega}$ with the maximum likelihood method.
no longer diagonal matrix
- Let us assume that the disturbances $\boldsymbol{\varepsilon}$ have a covariance matrix $\mathrm{E}(\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}^T) = \sigma^2 \cdot \boldsymbol{\Omega}$.
- The generalized least squares estimator accounting for the covariance structure $\boldsymbol{\Omega}$ in the disturbance $\boldsymbol{\varepsilon}$ becomes:

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T \cdot \boldsymbol{\Omega}^{-1} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \boldsymbol{\Omega}^{-1} \cdot \mathbf{y} \ .$$

- Alternatively, but equivalently (see Hamilton pp 188-189), after transforming the dependent variable and independent variables by $\mathbf{y}^* = \boldsymbol{\Omega}^{-\frac{1}{2}} \cdot \mathbf{y}$ and $\mathbf{X}^* = \boldsymbol{\Omega}^{-\frac{1}{2}} \cdot \mathbf{X}$ we get the standard OLS equation:
after this transformation, all observation becomes independent

$$\boxed{\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^{*T} \cdot \mathbf{X}^*)^{-1} \cdot \mathbf{X}^{*T} \cdot \mathbf{y}^*}.$$

- For simple covariance structures $\mathbf{\Omega}$, like diagonal matrices $\mathbf{\Omega} = \begin{pmatrix} \omega_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \omega_n \end{pmatrix}$, the inverse $\mathbf{\Omega}^{-1} = \begin{pmatrix} \frac{1}{\omega_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\omega_n} \end{pmatrix}$ or its inverse square root $\mathbf{\Omega}^{-\frac{1}{2}}$ are easily calculated.

  For general covariance structures $\mathbf{\Omega}$ more advanced methods such as the eigen-decomposition or Cholesky decomposition must be applied.
- The covariance matrix of the estimated generalized least squares regression coefficients $\hat{\boldsymbol{\beta}}_{GLS}$ becomes:

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{GLS}\right) = \sigma^2 \cdot (\mathbf{X}^T \cdot \mathbf{\Omega}^{-1} \cdot \mathbf{X})^{-1} \ .$$

- Why does generalized least squares work?
  - The transformed regression model becomes: $\mathbf{\Omega}^{-\frac{1}{2}} \cdot \mathbf{y} = \mathbf{\Omega}^{-\frac{1}{2}} \cdot \mathbf{X} \cdot \boldsymbol{\beta} + \mathbf{\Omega}^{-\frac{1}{2}} \cdot \boldsymbol{\varepsilon}$.
  - Therefore the transformed disturbances $\mathbf{\Omega}^{-\frac{1}{2}} \cdot \boldsymbol{\varepsilon}$ will follow an i.i.d normal distribution:

$$\mathrm{E}\left(\mathbf{\Omega}^{-\frac{1}{2}} \cdot \boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}^T \cdot \mathbf{\Omega}^{-\frac{1}{2}^T}\right) = \mathbf{\Omega}^{-\frac{1}{2}} \cdot \mathrm{E}(\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}^T) \cdot \mathbf{\Omega}^{-\frac{1}{2}^T}$$

$$\boxed{\begin{aligned} &= \mathbf{\Omega}^{-\frac{1}{2}} \cdot \sigma^2 \cdot \mathbf{\Omega} \cdot \mathbf{\Omega}^{-\frac{1}{2}^T} \\ &= \sigma^2 \cdot \mathbf{I} \end{aligned}}$$

it is a fixed matrix

- The ***main problem*** is that the covariance structure $\sigma^2 \cdot \mathbf{\Omega}$ of the disturbances is usually unknown. However, it may be specified by function that just depends on a set of unknown parameters. These parameters must be estimated from the data.
- If GLS is used with a ML estimated covariance matrices $\widehat{\mathbf{\Omega}}$ then it is called ***feasible*** GLS.

Heteroscedasticity no constant variance

- Remember, we are now explicitly modeling not only the expected value of the dependent variable $E(\mathbf{y}) = \boldsymbol{\mu} = \mathbf{X} \cdot \boldsymbol{\beta}$ but we also focus on the covariance structure among the disturbances $\sigma^2 \cdot \mathbf{\Omega}$.

  Furthermore, we need the explicit assumption that the dependent variable is normally distributed $\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2 \cdot \mathbf{\Omega})$ in order to perform maximum likelihood estimation.
- If heteroscedasticity is present, that is, $\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 \cdot \omega_i$, with ω_i being the ***individual variation*** of the disturbances around the ***global variance*** σ^2, the covariance structure takes the special diagonal form of

$$\mathbf{\Omega} = \begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_n \end{bmatrix}$$

with its inverse and square root of its inverse:

$$\mathbf{\Omega}^{-1} = \begin{bmatrix} \frac{1}{\omega_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\omega_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\omega_n} \end{bmatrix} \text{ and } \mathbf{\Omega}^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\omega_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\omega_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\omega_n}} \end{bmatrix}$$

- The challenge becomes estimating the n unknown variances $\mathrm{VAR}(\varepsilon_i) = \sigma_i^2 = \sigma^2 \cdot \omega_i$. It is impossible to estimate this many unknown variances since we have only n observations. Therefore, one needs to builds a ***functional model*** for the individual variances, which depends only on a few unknown parameters $\gamma_0, \cdots, \gamma_P$.
- A linear model for the unknown variances could be specified using P ***known*** external variables $\mathbf{z}_0, \cdots, \mathbf{z}_P$.
- All $\mathbf{z}_p$ must be positive ***positive*** so that the estimated variance remains positive.
- The underlying ***multiplicative*** model structure becomes, which can be transformed into a linear model structure:

$$\begin{aligned} &\sigma_i^2 = \sigma^2 \cdot {z_{i1}}^{\gamma_1} \cdot \cdots \cdot {z_{iP}}^{\gamma_P} \quad \text{and with } \boxed{\sigma^2 = \exp(\gamma_0)} \\ &= \exp(\gamma_0 \cdot 1 + \ln({z_{i1}}^{\gamma_1}) + \cdots + \ln({z_{iP}}^{\gamma_P})) \\ &= \exp(\gamma_0 \cdot 1 + \gamma_1 \cdot \ln(z_{i1}) + \cdots + \gamma_P \cdot \ln(z_{iP})) \end{aligned}$$

- Therefore, $\sigma^2 = \exp(\gamma_0)$ provides an estimate for the global variance.
- It is important that all variable $\mathbf{z}_p$ ***must be*** entered in their ***logarithmic*** form into the model specification for the heteroscedastic error variances (see the help of the function `lmHetero( )` in the R package `TexMix`).
- For the simple model with just one weights variable $\mathbf{z}_1$, that is,

$$\sigma_i^2 = \exp(\gamma_0 \cdot 1 + \gamma_1 \cdot \ln(z_{i1})) \,,$$

  we get the relationships

$$\begin{array}{ll} \gamma_1 > 0 & \sigma_i^2 \text{ is increasing with increasing } z_{i1} \\ \gamma_1 \cong 0 & \sigma_i^2 \text{ is not affected by } z_{i1} \;\Rightarrow\; \text{homoscedasticity} \\ \gamma_1 < 0 & \sigma_i^2 \text{ is decreasing with increasing } z_{i1} \end{array}$$

  spatial condition

- If the estimate coefficient γ_1 does not differ significantly for zero then the model is homoscedatistic with regards to the weights variable $\mathbf{z}_1$.
  Consequently, an adjustment for heteroscedasticity is ***not*** requires.
- Assuming the model is estimated using the maximum likelihood approach, a likelihood ratio test can be performed with

$$-2 \cdot \big(\mathcal{LL}(OLS\ model) - \mathcal{LL}(weighted\ model)\big) \sim \chi^2_{df=P-1}$$

- The -script **`MultiWeightedMaxLike.r`** demonstrates the maximum likelihood estimation of the unknown parameters γ and the estimation of the weights vector $\mathbf{w} = \left[\frac{1}{\omega_1}, \dots, \frac{1}{\omega_n}\right]^T$.

Modelling an Autoregressive Spatial Process with FGLS

- Let $\mathbf{V}$ be the a $n \times n$ coded spatial link matrix among the n spatial objects. For a precise definition of the coded link matrix $\mathbf{V}$ see the lecture notes on *Spatial Autocorrelation*.
- The error structure of disturbances for a Gausian autoregressive spatial process is

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \cdot \boldsymbol{\Omega}(\rho))$$

with $\boldsymbol{\Omega}(\rho) = (\mathbf{I} - \rho \cdot \mathbf{V})^{-1} \cdot (\mathbf{I} - \rho \cdot \mathbf{V}^T)^{-1}$. measures the autocorrelation level

- Therefore, $\boldsymbol{\Omega}(\rho)^{-\frac{1}{2}} = [\boldsymbol{I} - \rho \cdot \mathbf{V}]$ and i.i.d. disturbances can be obtained with the transformation $\boldsymbol{\eta} = \boldsymbol{\Omega}(\rho)^{-\frac{1}{2}} \cdot \boldsymbol{\varepsilon}$ where $\boldsymbol{\eta} = \mathbf{N}(\mathbf{0}, \sigma^2 \cdot \mathbf{I})$.
- The generalized least squares model becomes

$$\boldsymbol{\Omega}(\rho)^{-\frac{1}{2}} \cdot \mathbf{y} = \boldsymbol{\Omega}(\rho)^{-\frac{1}{2}} \cdot \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\eta}$$

which those unknown parameters $\hat{\sigma}^2$, $\hat{\rho}$ and $\widehat{\boldsymbol{\beta}}$ with a maximum likelihood estimator using the feasible general least squares approach.

- The -script **`SpatialACItaly.rmd`** demonstrates the estimation of a spatial autoregressive process.

Modelling Serial Autocorrelation with GLS (skipped, not test relevant)

- Assuming an autoregressive a stochastic process in the random component ε_t becomes
$$\varepsilon_t = \underbrace{\varphi \cdot \varepsilon_{t-1}}_{AR-\text{process}} + \mu_t \text{ with iid } \mu_t \sim N(0, \sigma^2)$$
- Underlying covariance structure $\mathbf{\Omega}$ depends on the AR parameter φ.
- For instance, a process with $|\varphi| < 1$ becomes a first order AR process.
  Its underlying covariance structure for a process of an equally spaced temporal sequence of observation of length T is
$$\mathbf{\Omega}_{(\varphi)} = \frac{1}{1-\varphi^2} \cdot \begin{bmatrix} 1 & \varphi & \varphi^2 & \cdots & \varphi^{T-2} & \varphi^{T-1} \\ \varphi & 1 & \varphi & \cdots & \varphi^{T-3} & \varphi^{T-2} \\ \varphi^2 & \varphi & 1 & \cdots & \varphi^{T-4} & \varphi^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \varphi^{T-2} & \varphi^{T-3} & \varphi^{T-4} & \cdots & 1 & \varphi \\ \varphi^{T-1} & \varphi^{T-2} & \varphi^{T-3} & \cdots & \varphi & 1 \end{bmatrix}$$
- Notice, the more time periods Δt two observations are apart the less their stochastic dependence becomes, because for $|\varphi| < 1$ the autocorrelation effect $\left|\varphi^{\Delta t}\right| < |\varphi|$ for $\Delta t > 1$.
- Insert $\mathbf{\Omega}_{(\varphi)}^{-1/2}$ here.

- Assuming this covariance structure $\mathbf{\Omega}_{(\varphi)}$, the unknown parameters $\{\boldsymbol{\beta}, \sigma^2, \varphi\}$ can be estimated by maximum likelihood assume a distribution of the endogenous variable is $\mathbf{y} \sim N\left(\mathbf{X} \cdot \boldsymbol{\beta}, \sigma^2 \cdot \mathbf{\Omega}_{(\varphi)}\right)$.
- The R-script **`GLSArmaConcord.rmd`** demonstrates the estimation of temporally autocorrelated stochastic processes.