Feasible GLS method

Exercise 1

Consider a linear model

$$sav = \beta_0 + \beta_1 inc + u$$

and use Feasible GLS method to correct heteroskedasticity.

In general, for linear model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

which exhibits heteroskedasticity (variance changes across the data), we model the function h and use the data to estimate $\hat{h}_i$. Using $\hat{h}_i$ instead of h_i in the GLS transformation yields an estimator called **the feasible GLS (FGLS)** estimator.

In this approach we model the heteroskedasticity in a following way; we write a variance in a form

$$Var(u|x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k),$$

with

$$h(x) = \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k)$$

where $x_1, \cdots, x_k$ are explanatory variables and δ_j are uknown parameters. Under this hypothesis we write

$$u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k)\nu$$

with $\mathbb{E}(\nu|x_1, \cdots, x_k) = 1$ and switch to the model (which is a linear regression model):

$$\log(u^2) = \alpha_0 + \delta_1 x_1 + \cdots + \delta_k x_k + e,$$

where e is zero mean random variable and independent of $x_1, \cdots, x_k$. The intercept in this equation is different from δ_0, but this is not important. Now we run the regression of $log(u^2)$ on $x_1, x_2, \cdots, x_k$. We need from this regression the fitted values; call these $\hat{g}_i$. Then, the estimates of h_i (of the initial model) are simply $\hat{h}_i = \exp(\hat{g}_i)$.

Essentially, FGLS method consists of 5 steps:

1. Run the regression of y on $x_1, x_2, \cdots, x_k$ and obtain the residuals $\hat{u}^2$.

2. Create $log(\hat{u}^2)$ by first squaring the OLS residuals and then taking the natural log.

3. Run the regression $log(u^2)$ on $x_1, x_2, \cdots, x_k$ and obtain the fitted values $\hat{g}$.
4. Exponentiate the fitted values from, i.e. $\hat{h} = exp(\hat{g})$.
5. Estimate the equation
$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u.$$
by WLS, using weights $1/\hat{h}$.

Code:

$$load \ \ saving.raw$$
$$y = saving(:, 1);$$
$$inc = saving(:, 2);$$
$$[n, k] = size(saving);$$
$$X = [ones(n, 1), inc];$$
$$[n, k] = size(X)$$
$$beta = inv(X' * X) * X' * y$$
$$u = y - X * beta;$$

From now on, we consider the model

$$log(u^2) = \delta_0 + \delta_1 inc + e$$
$$lu2 = log(u.^2);$$
$$y = lu2;$$
$$beta = inv(X' * X) * X' * y$$

and compute its residuals

$$u = y - X * beta;$$

We obtain estimated weights $\hat{g}_i$, and obtain estimated weights $\hat{h_i} = exp(\hat{g}_i)$ of the initial model.

$$g = X * beta$$
$$new\_weight = sqrt(exp(g))$$

Finally, we perform FGLS correction of the initial model:

$$y = saving(:, 1);$$
$$inc = saving(:, 2);$$
$$X = [ones(n, 1), inc];$$
$$ys = y./new_weight;$$
$$Xs = [ones(n, 1)./new\_weightinc./new\_weight]$$

Exercise 2

We consider a model

$cigs = \beta_0 + \beta_1 log(income) + \beta_2 log(cigpric) + \beta_3 educ + \beta_4 age + \beta_5 age^2 + \beta_6 restaurn + u$

We conduct FGLS method to correct heteroskedasticity. Code:

$$loadsmoke.raw$$

$$[n, k] = size(smoke)$$
$$y = smoke(:, 6);$$
$$X = [ones(n, 1) smoke(:, [1, 4, 7, 8, 9, 10])];$$
$$beta = inv(X' * X) * X' * y$$
$$u = y - X * beta;$$

We consider logarithm of residuals:

$$lu2 = log(u.^2);$$

$$y = lu2;$$
$$beta = inv(X' * X) * X' * y$$
$$g = X * beta$$
$$new\_weight = sqrt(exp(g))$$

and correct the initial model with new estimated weights:

$$y = smoke(:, 6);$$

$$X = [ones(n, 1) smoke(:, [1, 4, 7, 8, 9, 10])];$$
$$[n, k] = size(X);$$
$$ys = y./new\_weight;$$
$$Xs = X;$$
$$for \ \ i = 1 : k$$
$$Xs(:, i) = Xs(:, i)./new\_weight$$
$$end$$

Time series

Exercises 3-6

- The autoregressive process AR(p) is of the form:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t$$

  with $\mathbb{E}[u_t] = 0, \mathbb{E}[u_t u_s] = \sigma_u^2$ if $t = s$ and 0 else

- The moving average process Ma(q) is of the form:

$$y_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \cdots + \psi_q \epsilon_{t-q}$$

- The process ARMA(p,q) is of the form:

$$y_t - \sum_{k=1}^{p} \phi_k y_{t-k} = \epsilon_t + \sum_{j=1}^{q} \psi_j \epsilon_{t-j}$$

The autocorrelation (ACF) and partial autocorrelation (PACF) functions of these processes make it possible to identify them because they have the following properties:

- Autoregressive processes (**AR (p)**): The ACF decreases exponentially. Concerning the **PACF**, the coefficients are null or not significant for $|h| > p$.
- Moving average processes (**MA(q)**): The PACF decreases exponentially. Concerning the **ACF**, the coefficients are null or not significant for $|h| > q$.
- Processes (**ARMA(p,q)**): The ACF and the PACF are decreasing, but they do not necessarily become zero after a certain delay. It is therefore more difficult to identify an ARMA model than a pure autoregressive or a moving average model.

Simulate AR(1) model

$$y_n = 0.6 y_{n-1} + u_n$$

and plot its trajectory, compute ACF and PACF:

$$y = zeros(1000, 1);$$

$$for\ \ i = 2 : 1000;$$

$$y(i) = 0.6 * y(i-1) + randn;$$
$$end;$$
$$plot(y)$$
$$acf = autocorr(y, 20);$$
$$pacf = parcorr(y, 20);$$

Simulate MA(1) model

$$z_n = e_n + 0.8e_{n-1}$$

and plot its trajectory, compute ACF and PACF:

$$n = 1000; z = zeros(n, 1)$$
$$e = randn(n, 1);$$
$$for \;\; i = 2 : 1000;$$
$$z(i) = e(i) + 0.8 * e(i-1);$$
$$end;$$
$$plot(z);$$
$$acf = autocorr(z, 20);$$
$$pacf = parcorr(z, 20); ;$$

Exercise 7

The model of the ordinary least squares is considered under the following hypotheses:

- H1 : The model is linear in $x_{i,t}$
- H2 : the values $x_{i,t}$ are observed without error
- H3 : $\mathbb{E}[u] = 0$ the mean of the error is equal to zero
- H4: $\mathbb{E}[u^t u] = \sigma^2 I_n$, the variance of the error is constant
- H5 : $\mathbb{E}[u_t u_{t+1}]$, the errors are not correlated
- H6 : $Cov(x_{i,t}, u_t)$, the error is independent of the explanatory variable

The violation of hypothesis H5 concerns time series where the off-diagonal elements of the covariance matrix of the errors are nonzero. In this case, the obtained OLS estimators are unbiased but no longer have a minimal variance. Thus, we must identify new estimators and techniques for detecting possible autocorrelation of the errors.

The detection of a potential dependency of the errors can only be carried out through the analysis of the residuals.

Test of the autocorrelation

We start with regression

$$i3t = \beta_0 + \beta_1 inft + \beta_2 deft + u_t$$

for $t = 1, \cdots, n$.

$$y = intdef(:, 2);$$

$$[n, k] = size(intdef)$$

$$X = [ones(n, 1), intdef(:, [3, 6])];$$

$$[n, k] = size(X)$$

$$beta = inv(X' * X) * X' * y$$

$$u = y - X * beta;$$

$$sig2 = u' * u/(n - k)$$

$$std = sqrt(diag(sig2 * inv(X' * X)))$$

$$t = beta./std$$

The residuals of the regression can be correlated in series. The most popular and simple model to be tested is the AR (1) model. We will therefore test the presence of correlations in series of type AR (1).

We assume that the residuals are given by $u_t = \rho u_{t-1} + e_t$ We assume that $|\rho| < 1$ (statbility conditon), and that e_t are independent, zero mean random variables with variance σ_e^2.

In the model AR(1), the null hypothesis H_0 assumes that the errors are not correlated in the series

$$H_0 : \rho = 0.$$

Thus, we do a regression $\hat{u}_t = \rho \hat{u}_{t-1} + e_t$ for $t = 2$, recover $\hat{\rho}$,and calculate the statistic $t_{\hat{\rho}}$.

$$u\_ = [u(2 : n)]$$

$$u\_lag = [u(1 : n - 1)]$$

$$y = u\_$$

$$X = [u\_lag];\ \textit{without constant}$$

$$[n, k] = size(X)$$

$$rho = inv(X' * X) * X' * y$$

$$u = y - X * rho;$$

$$sig2 = u' * u/(n-k)$$
$$std = sqrt(diag(sig2 * inv(X' * X)))$$
$$t = rho./std$$

Finally, we use t-test
$$p = tdis\_prb(t, n-k)$$

We obtain $t_{\hat{\rho}} = 5.7295$ and $p = 4.6122 - 07$. Thus, we reject H_0.