Heteroscedasticity autocorrelation consistent estimator

Contents

1. GLS with known covariance matrix
   - Heteroskedasticity
   - Autocorrelation
   - GLS and maximum likelihood estimation

2. GLS with unknown covariance matrix
   - AR(1) errors with exogenous regressors

Consider the model

$$Y = X\beta + \epsilon \tag{1}$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \quad X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix}$$

where X is an $T \times p$ matrix of fixed or stochastic explanatory variables, such that $T \ll p$, $\beta \in \mathbb{R}^p$ is a vector of parameters, and

$$\epsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix} \quad \text{s.t.} \quad \epsilon \mid X \sim N(0, \sigma^2 \Sigma) \tag{2}$$

where

$$\sigma^2 \Sigma = \mathbb{E}\left[\epsilon\epsilon' \mid X\right]$$

Matrix Σ is symmetric and positive definite, and there exists a non-singular $T \times T$ matrix C, such that

$$\Sigma^{-1} = C'C$$

Proof.

We know that there exists a non-singular matrix L such that $\Sigma = LL'$, and so $\Sigma^{-1} = [L']^{-1}L^{-1}$. Let $C = L^{-1}$, which yields the earlier results. □

Imagine transforming the population residuals ϵ by C:

$$\tilde{\epsilon} = C\epsilon$$

which would generate a new set of residuals $\tilde{\epsilon}$ with zero mean and conditional covariance matrix, given by

$$\mathbb{E}[\tilde{\epsilon}\tilde{\epsilon}' \mid X] = C\mathbb{E}[\epsilon\epsilon' \mid X]C' = C\Sigma C'.$$

But $\Sigma = \left[\Sigma^{-1}\right]^{-1} = [C'C]^{-1}$; hence,

$$\mathbb{E}[\tilde{\epsilon}\tilde{\epsilon}' \mid X] = C[C'C]^{-1}C' = \sigma^2 I$$

We may now transform regression equation (1) by premultiplying both its sides by C, which yields

$$\tilde{Y} = \tilde{X}\beta + \tilde{\epsilon}$$

where

$$\tilde{Y} \equiv CY, \quad \tilde{X} \equiv CX, \quad \tilde{\epsilon} \equiv C\epsilon$$

and

$$\tilde{\epsilon} \mid X \sim N(0, \sigma^2 I).$$

The Lasso estimator on the transformed model is as follows

$$\tilde{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{2}\|\tilde{Y} - \tilde{X}\beta\|_2^2 + \lambda\|\beta\|_1$$

A simple case is when the variance of ε_t is presumed to be proportional to the square of the explanatory variable for that equation, say $x_{1,t}^2$:

$$\mathbb{E}[\epsilon\epsilon' \mid X] = \sigma^2 \begin{bmatrix} x_{1,1}^2 & 0 & \cdots & 0 \\ 0 & x_{1,2}^2 & \cdots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \cdots & x_{1,T}^2 \end{bmatrix} = \sigma^2\Sigma$$

Then, it is easy to see that

$$C = \begin{bmatrix} 1/|x_{1,1}| & 0 & \cdots & 0 \\ 0 & 1/|x_{1,2}| & \cdots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \cdots & 1/|x_{1,T}| \end{bmatrix}$$

Hence, the estimation of Lasso will be conducted by using the transformed variables $\tilde{y}_t = y_t/|x_{1,t}|$ and $x_t/|x_{1,t}|$.

Consider the case were

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, \quad u_t \sim N(0, \sigma^2) \tag{3}$$

and $|\rho| < 1$, then

$$\mathbb{E}[\epsilon\epsilon' \mid X] = \frac{\sigma^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix} = \sigma^2\Sigma$$

and subsequently,

$$C = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix} \tag{4}$$

In other words, the transformation must take the form $\tilde{y}_1 \equiv y_1\sqrt{1-\rho^2}$ and $\tilde{x}_1 \equiv x_1\sqrt{1-\rho^2}$ and $\tilde{y}_t \equiv y_t - \rho y_{t-1}$ and $x_t - \rho x_{t-1}$ for $t = 2, \cdots, T$.

1. GLS with known covariance matrix
   - Heteroskedasticity
   - Autocorrelation
   - GLS and maximum likelihood estimation
2. GLS with unknown covariance matrix
   - AR(1) errors with exogenous regressors

Under the earlier assumptions that

- x_t is stochastic;
- Conditional on the full matrix X, the vector $\epsilon \sim N(0, \sigma^2\Sigma)$;
- Σ is a known positive definite matrix.

it is known that

$$Y \mid X \sim N(X\beta, \sigma^2\Sigma)$$

Hence,

$$\begin{aligned} L(Y \mid X; \beta, \sigma^2\Sigma) &= (2\pi)^{-T/2}|\det(\sigma^2\Sigma)|^{-1/2} \\ &\quad \times \exp\left(\left(-\frac{1}{2}\right)(Y - X\beta)'(\sigma^2\Sigma)^{-1}(Y - X\beta)\right) \end{aligned}$$

and subsequently

$$\begin{aligned} \log\{L(Y \mid X; \beta, \sigma^2\Sigma)\} &\equiv l(Y \mid X; \beta, \sigma^2\Sigma) \\ &= \left(-\frac{T}{2}\right)\log(2\pi) - \left(\frac{1}{2}\right)\log\left|\det\left(\sigma^2\Sigma\right)\right| \end{aligned}$$

$$-\left(\frac{1}{2}\right)(Y-X\beta)'(\sigma^2\Sigma)^{-1}(Y-X\beta)$$

From earlier recall that $(\Sigma)^{-1} = C'C$. Hence, we may express the right hand terms of the above log-likelihood function as

$$\begin{aligned}
\left(\frac{1}{2}\right)(Y-X\beta)'(\sigma^2\Sigma)^{-1}(Y-X\beta) &= \left(\frac{1}{2}\right)(Y-X\beta)'(\sigma^2)^{-1}(C'C)(Y-X\beta) \\
&= \left(\frac{1}{2\sigma^2}\right)(Y-X\beta)'(C'C)(Y-X\beta) \\
&= \left(\frac{1}{2\sigma^2}\right)(CY-CX\beta)'(CY-CX\beta) \\
&= \left(\frac{1}{2\sigma^2}\right)(\tilde{Y}-\tilde{X}\beta)'(\tilde{Y}-\tilde{X}\beta)
\end{aligned}$$

Moreover,

$$-\left(\frac{1}{2}\right)\log\left|\det\left(\sigma^2\Sigma\right)\right| = -\left(\frac{1}{2}\right)\log\left|\sigma^{2T}\det\left(\Sigma\right)\right|$$

$$= -\left(\frac{1}{2}\right)\log\left|\sigma^{2T}\det\left\{(C'C)^{-1}\right\}\right|$$

$$= -\left(\frac{1}{2}\right)\log\left|\sigma^{2T}\det\{(C'C)\}^{-1}\right|$$

$$= -\left(\frac{1}{2}\right)\log\sigma^{2T} - \left(\frac{1}{2}\right)\log\left|\det\{(C'C)\}^{-1}\right|$$

$$= -\left(\frac{T}{2}\right)\log\sigma^{2} + \left(\frac{1}{2}\right)\log\left|\det\{(C'C)\}\right|$$

$$= -\left(\frac{T}{2}\right)\log\sigma^{2} + \log\left|\det(C)\right|$$

Therefore, the conditional log-likelihood function can be expressed as

$$l(Y \mid X; \beta, \sigma^2\Sigma) = -\left(\frac{T}{2}\right)\log(2\pi) - \left(\frac{T}{2}\right)\log(\sigma^2) \quad (5)$$

$$+\log|\det(C)| - \left(\frac{1}{2\sigma^2}\right)(\tilde{Y} - \tilde{X}\beta)'(\tilde{Y} - \tilde{X}\beta) \quad (6)$$

Thus, the likelihood is maximized with respect to β by an OLS regression of $\tilde{Y}$ on $\tilde{X}$.

Once again consider the conditional covariance matrix of the error terms in (2). Up to this point, it was assumed that the elements of Σ are known a priori. In reality, Σ is of a particular form $\Sigma(\theta)$, where θ is a vector of parameters that must be estimated from the data, - i.e.

$$\epsilon \mid X \sim N(0, \sigma^2 \Sigma(\theta))$$

For instance, for the autocorrelated residual case (3), θ is the scalar ρ. Our task is then to estimate ρ and β jointly from the data. One approach is to estimate ρ and β jointly from the data and find the values that maximize (5). The latter can be formed and maximized numerically and has the appeal of offering a single rule to follow whenever $\mathbb{E}[\epsilon\epsilon' \mid X]$ is not of the simple form $\sigma^2 I$. However, quite often simple estimator can have desirable properties. In a classical asymptotics setting, its turn out that

$$\sqrt{T}(X'[\Sigma(\hat{\rho})]^{-1}X)^{-1}(X'[\Sigma(\hat{\rho})]^{-1}Y) \xrightarrow{p} \sqrt{T}(X'[\Sigma(\rho_0)]^{-1}X)^{-1}(X'[\Sigma(\rho_0)]^{-1}Y)$$

Let us maintain the earlier assumption that

$$\epsilon \mid X \sim N(0, \sigma^2 \Sigma(\rho))$$

which rules out endogenous variables; in other words it is assumed that x_t is uncorrelated with ε_{t-s}.

Recall that the determinant of a lower triangular matrix is simply the product of the terms on the principal diagonal. From (4), it is evident that

$$\det(C) = \sqrt{1-\rho^2}.$$

Thus, the log-likelihood function (5) is expressed as

$$\begin{aligned} l(Y \mid X; \beta, \rho, \sigma) = &-\left(\frac{T}{2}\right)\log(2\pi) - \left(\frac{T}{2}\right)\log(\sigma^2) + \left(\frac{1}{2}\right)\log(1-\rho^2) \\ &- \left[\frac{(1-\rho^2)}{2\sigma^2}\right](y_1 - x_1'\beta)^2 \\ &- \left(\frac{1}{2\sigma^2}\right)\sum_{t=2}^{T}[(y_t - x_t'\beta) - \rho(y_{t-1} - x_{t-1}'\beta)]^2 \end{aligned} \quad (7)$$

One approach as mentioned in the earlier section, is to maximize (7) with respect to β, ρ and σ^2.

- If we knew the value of ρ, then the value of β that maximizes (7), could be found by an OLS regression of $y_t - \rho y_{t-1}$ on $x_t - \rho x_{t-1}$) for $t = 2, \cdots, T$ [let us call this regression A];
- Conversely, if we knew the value of ρ that maximizes (7) would be found by an OLS regression of $(y_t - x_t'\beta)$ on $(y_{t-1} - x_{t-1}'\beta)$ [let us call this regression B].
- We can thus, start by an initial guess of ρ, say $\rho = 0$. Perform regression A to get an initial estimate of β.
- This estimate β can then be used in regression B to get an updated estimate of ρ, for example, by regressing the OLS residuals $\hat{\varepsilon}_t = y_t - x_t'\beta$ on its own lagged value.
- The new estimate of ρ can be used to repeat the two regressions. Zigzagging back and forth between A and B is known as the iterated Cochrane-Orcutt method, and will converge to a local maximum of (7).

Consider the estimator of ρ that results from the first iteration alone

$$\hat{\rho} = \frac{(1/T)\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{(1/T)\sum_{t=1}^{T} \hat{\varepsilon}_{t-1}^2} \tag{8}$$

where $\hat{\varepsilon}_t = y_t - \hat{\beta}' x_t$ and $\hat{\beta}$ is the OLS estimate of β. Notice that

$$\hat{\varepsilon_t} = (y_t - \beta' x_t + \beta' x_t - \hat{\beta}' x_t) = \varepsilon_t + (\beta - \hat{\beta})' x_t$$

allowing the numerator of (8) to be written

$$\begin{aligned}(1/T)\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} &= (1/T)\sum_{t=1}^{T} [\varepsilon_t + (\beta - \hat{\beta})' x_t][\varepsilon_{t-1} + (\beta - \hat{\beta})' x_{t-1}] \\ &= (1/T)\sum_{t=1}^{T} \varepsilon_t \varepsilon_{t-1} + (\beta - \hat{\beta})'(1/T)\sum_{t=1}^{T} (\varepsilon_t x_{t-1} + \varepsilon_{t-1} x_t) \\ &\quad + (\beta - \hat{\beta})' \left[(1/T)\sum_{t=1}^{T} x_t x_{t-1}' \right] (\beta - \hat{\beta}) \end{aligned} \tag{9}$$

As long as, $\hat{\beta}$ is a consistent estimate of β and boundedness conditions ensure that plims of

- $(1/T)\sum_{t=1}^{T} \varepsilon_t x_{t-1}$;
- $(1/T)\sum_{t=1}^{T} \varepsilon_{t-1} x_t$;
- $(1/T)\sum_{t=1}^{T} x_t x_{t-1}'$

exist, then

$$\begin{aligned}(1/T)\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1} &\xrightarrow{p} (1/T)\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1}\\ &= (1/T)\sum_{t=1}^{T}(u_t+\rho\varepsilon_{t-1})\varepsilon_{t-1}\\ &\xrightarrow{p} \rho\text{var}(\varepsilon)\end{aligned}$$

Similar analysis determines that the denominator of (8) converges in probability to var(ε), where they cancel each other out, and as such it can be established that

$$\hat{\rho} \xrightarrow{p} \rho.$$

Now if ε_t is uncorrelated with x_s for $s = t-1, t, t+1$ stronger claims can be made about the estimate of ρ based on an autoregression of the OLS residuals $\hat{\varepsilon}$. Specifically, if

$$\text{plim}\left[(1/T)\sum_{t=1}^{T}\varepsilon_t x_{t-1}\right] = \text{plim}\left[(1/T)\sum_{t=1}^{T}\varepsilon_{t-1} x_t\right] = 0$$

then multiplying (9) by $\sqrt{T}$, we find

$$
\begin{aligned}
(1/\sqrt{T})\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1} &= (1/\sqrt{T})\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1} \\
&\quad + \sqrt{T}(\beta-\hat{\beta})'(1/T)\sum_{t=1}^{T}(\varepsilon_t x_{t-1}+\varepsilon_{t-1}x_t) \\
&\quad + \sqrt{T}(\beta-\hat{\beta})'\left[(1/T)\sum_{t=1}^{T}x_t x_{t-1}'\right](\beta-\hat{\beta})' \\
&\xrightarrow{p} (1/\sqrt{T})\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1} + \sqrt{T}(\beta-\hat{\beta})'0 \\
&\quad + \sqrt{T}(\beta-\hat{\beta})'\,\mathrm{plim}\left[(1/T)\sum_{t=1}^{T}x_t x_{t-1}'\right]0 \\
&= (1/\sqrt{T})\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1}
\end{aligned} \tag{10}
$$

Hence,

$$\sqrt{T}\left[\frac{(1/T)\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{(1/T)\sum_{t=1}^{T}\hat{\varepsilon}_{t-1}^2}\right] \xrightarrow{p} \sqrt{T}\left[\frac{(1/T)\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1}}{(1/T)\sum_{t=1}^{T}\varepsilon_{t-1}^2}\right]$$

The OLS estimate of ρ based on the population residuals would have an asymptotic distribution given by

$$\sqrt{T}\left[\frac{(1/T)\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{(1/T)\sum_{t=1}^{T}\hat{\varepsilon}_{t-1}^2}\right] \xrightarrow{L} N(0, 1-\rho^2)$$