1. Exercise

Consider the panel model

$$
\begin{aligned}
y &= X\beta + Z\mu + u \\
y &= X\beta + e.
\end{aligned} \tag{1}
$$

The GLS estimator is useful when the variance-covariance matrix of the error term abstracts from the case where the error variance-covariance matrix is diagonal with constant elements on the diagonal, i.e. when $E(ee') = \sigma^2 I$.

A frequent case of this abstraction is the random effects model where the individual effects are assumed to be random variables with $\mu_i \sim IID(0, \sigma_\mu^2)$ and independent of the idiosyncratic error u_{it}. In this case, the individual effects are subsumed by the error term which alters the variance-covariance matrix of the error term such that $E(ee') = \Omega \neq \sigma^2 I$. Its elements on the main diagonal are $\sigma_\mu^2 + \sigma_u^2$ and the off-diagonal elements of each block are σ_μ^2.[1] Therefore, in the random effects model, the errors are serially correlated within individuals and OLS would be consistent yet inefficient.

In contrast, the GLS estimator

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \tag{2}$$

is both consistent *and* efficient, provided that X_{it} is also independent of μ_i. The efficiency follows because the GLS estimator is equivalent to applying OLS on the transformed model whose error variance-covariance matrix is no longer correlated for observations of the same individual after the transformation. The technical procedure of this transformation relies on Ω being positive definite. In this case there exists a transformation matrix P such that $P'P = \Omega^{-1}$.[2] Transformation of the model (1) with P yields

$$Py = PX\beta + Pe \tag{3}$$

and applying OLS on this transformed model gives the GLS estimator for β

$$
\begin{aligned}
\hat{\beta}_{OLS} &= (X'P'PX)^{-1}X'P'Py \\
&= (X'\Omega^{-1}X)X'\Omega^{-1}y.
\end{aligned} \tag{4}
$$

In contrast to OLS on the untransformed model, GLS is efficient because the transformed error is no longer serially correlated within individuals, i.e.,

$$E(Pee'P') = E(P\Omega P') = I.^{3,4} \tag{5}$$

[1] Lecture slides 2, pp. 26-27 and Baltagi (2005) p.15

[2] Exercise sheet 4, exercise 5.

[3] Using the fact that $P\Omega P' = I$. See exercise sheet 4, exercise 5.

[4] Cameron and Trivedi (2005) pp.81-82

In practice, a number of problems arise with the implementation of the GLS estimator. The first one follows directly from the above assumption of the random effects model that the regressors are independent of the individual effects. This is a strong assumption and is often only warranted when the sample units are drawn from a large population. If this assumption turned out to be false, then the estimates obtained from GLS will be inconsistent.[5]

The second problem is the fact that Ω has to be consistently estimated. Therefore, one needs estimates of the variance components σ_μ^2 and σ_u^2. There are a handful of propositions to estimate them, the most prominent being the methods suggested by Nerlove, Swamy-Arora, Wallace and Hussain, or Amemiya.

A common intercept can be estimated in the same way as the coefficients of the independent variables, assuming it is included in X. Otherwise α can be estimated by $\hat{\alpha}_{GLS} = \bar{y} - \bar{x}'\hat{\beta}_{GLS}$, where $\bar{x}$ is the vector of independent variable means. Since the individual effects are assumed to be random variables and thus are specifically modeled as a component of the error term, they cannot be estimated as parameters of the model in the classical sense. They can, however, be estimated as realizations from their distribution by

$$\mu = \left(1-(1-\theta)^2\right)\frac{1}{T}ZZ'(y-X\beta), \tag{6}$$

where $(1-\theta)^2 = \frac{\sigma_u^2}{T\sigma_\mu^2+\sigma_u^2}$ and σ_μ^2, σ_u^2, and β are substituted for the estimates of the random effects model.[6,7]

Besides the random effects models, other abstractions from the diagonal and constant structure of the error variance matrix are possible. These are heteroskedasticity and serial correlation. In case of heteroskedasticity, the diagonal elements are no longer identical and in case of serial correlation, like in the random effects model, off-diagonal elements of each block are non-zero. Both heteroskedasticity and serial correlation make the OLS estimator inefficient and can be observed frequently in real world data samples, even more so serial correlation with panel data. Therefore, GLS can also be used to remedy these complications in panel data as opposed to pooled OLS.

The GLS estimator offers gains in efficiency whenever the variance-covariance structure of the error term is such that the Gauss-Markov conditions are violated. The goal of GLS is to correct this issue by transforming the data so that the transformed error structure satisfies the Gauss-Markov conditions again and thereby allowing for correct parameter inference.

[5] Baltagi (2005) p.14 and Greene (2003) pp.293-294

[6] Croissant and Millo (2008); see the source code of the `plm::ranef()` function. Retrieved from https://rdrr.io/rforge/plm/src/R/tool_ranfixef.R.

[7] Greene (2003) p.296

2. Exercise

Assume the model with individual and time fixed effects

$$y = X\beta + Z_\mu\mu + Z_\lambda\lambda + u. \tag{7}$$

The two-way within estimator is given by

$$\hat{\beta} = (X'QX)^{-1}X'Qy, \tag{8}$$

with $Q = I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T$.[8]

For the first within transformation define $Q_\mu = I_{NT} - I_N \otimes \bar{J}_T$. $\bar{J}_T$ is a matrix of $1/T$ so that Q_μ yields deviations from individual means. Similarly, define $Q_\lambda = I_{NT} - \bar{J}_N \otimes I_T$ for the second within transformation. $\bar{J}_N$ is a matrix of $1/N$ so that Q_λ yields deviations from time period means.

Multiplying with Q_μ and Q_λ in (7) gives

$$Q_\mu Q_\lambda y = Q_\mu Q_\lambda X\beta + Q_\mu Q_\lambda Z_\mu\mu + Q_\mu Q_\lambda Z_\lambda\lambda + Q_\mu Q_\lambda u. \tag{9}$$

Assuming for now that

$$Q_\mu Q_\lambda Z_\mu\mu = Q_\mu Q_\lambda Z_\lambda\lambda = 0, \tag{10}$$

yields

$$Q_\mu Q_\lambda y = Q_\mu Q_\lambda X\beta + Q_\mu Q_\lambda u. \tag{11}$$

OLS on this equation gives

$$\begin{aligned}
\hat{\beta} &= (X'Q_\lambda' Q_\mu' Q_\mu Q_\lambda X)^{-1}X'Q_\lambda' Q_\mu' Q_\mu Q_\lambda y \\
&= (X'(Q_\mu Q_\lambda)' Q_\mu Q_\lambda X)^{-1}X'(Q_\mu Q_\lambda)' Q_\mu Q_\lambda y \\
&= (X'Q'QX)^{-1}X'Q'Qy \\
&= (X'QX)^{-1}X'Qy,
\end{aligned}$$

where I used the fact that Q is symmetric and idempotent[9] as well as

$$Q_\mu Q_\lambda = Q. \tag{12}$$

8 Lecture slides 3, p.8

9 Lecture slides 3, p.9

Thus, it suffices to show that (12) and (10) hold and that Q is symmetric and idempotent.

$$\begin{aligned}
Q_\mu Q_\lambda &= (I_{NT} - I_N \otimes \bar{J}_T)(I_{NT} - \bar{J}_N \otimes I_T) \\
&= I_{NT} - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + (I_N \otimes \bar{J}_T)(\bar{J}_N \otimes I_T) \\
&= I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + I_N \bar{J}_N \otimes \bar{J}_T I_T \\
&= I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T \\
&= Q
\end{aligned} \tag{12}$$

Using the definitions $Z_\mu = I_N \otimes \iota_T$ and $Z_\lambda = \iota_N \otimes I_T$, then (10) holds because

$$\begin{aligned}
Q_\mu Q_\lambda Z_\mu = Q Z_\mu &= (I_{NT} - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T)(I_N \otimes \iota_T) \\
&= I_N \otimes \iota_T - (I_N \otimes \bar{J}_T)(I_N \otimes \iota_T) - (\bar{J}_N \otimes I_T)(I_N \otimes \iota_T) + (\bar{J}_N \otimes \bar{J}_T)(I_N \otimes \iota_T) \\
&= I_N \otimes \iota_T - I_N \otimes \iota_T - \bar{J}_N \otimes \iota_T + \bar{J}_N \otimes \iota_T \\
&= 0
\end{aligned} \tag{10}$$

and

$$\begin{aligned}
Q_\lambda Z_\lambda &= (I_{NT} - \bar{J}_N \otimes I_T)(\iota_N \otimes I_T) \\
&= \iota_N \otimes I_T - (\bar{J}_N \otimes I_T)(\iota_N \otimes I_T) \\
&= \iota_N \otimes I_T - \iota_N \otimes I_T \\
&= 0.
\end{aligned} \tag{10}$$

Since the transpose of a sum of matrices is the sum of the transposed matrices ($(A+B)' = A' + B'$) and Q is the sum of four symmetric matrices, Q must be symmetric as well, i.e., $Q' = Q$. Or formally, given that $(A \otimes B)' = A' \otimes B'$ and $I_N, I_T, \bar{J}_N, \bar{J}_T$ are all symmetric:

$$\begin{aligned}
Q' &= (I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T)' \\
&= (I_N \otimes I_T)' - (I_N \otimes \bar{J}_T)' - (\bar{J}_N \otimes I_T)' + (\bar{J}_N \otimes \bar{J}_T)' \\
&= I_N' \otimes I_T' - I_N' \otimes \bar{J}_T{}' - \bar{J}_N{}' \otimes I_T' + \bar{J}_N{}' \otimes \bar{J}_T{}' \\
&= I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T \\
&= Q
\end{aligned} \tag{13}$$

Finally, Q is idempotent because

$$\begin{aligned}
Q^2 &= (I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T)(I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T) \\
&= I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T - I_N \otimes \bar{J}_T + I_N \otimes \bar{J}_T + \bar{J}_N \otimes \bar{J}_T - \bar{J}_N \otimes \bar{J}_T \\
&\quad - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T + \bar{J}_N \otimes I_T - \bar{J}_N \otimes \bar{J}_T + \bar{J}_N \otimes \bar{J}_T - \bar{J}_N \otimes \bar{J}_T - \bar{J}_N \otimes \bar{J}_T + \bar{J}_N \otimes \bar{J}_T \\
&= I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T \\
&= Q.
\end{aligned} \tag{14}$$

References

Baltagi, Badi (2005). *Econometric Analysis of Panel Data*. 3rd Edition. Wiley.

Cameron, A Colin and Pravin K Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Croissant, Yves and Giovanni Millo (2008). "Panel Data Econometrics in R: The plm Package". In: *Journal of Statistical Software* 27.2, pp. 1–43. DOI: 10.18637/jss.v027.i02.

Greene, William H (2003). *Econometric Analysis*. 5th Edition. Prentice Hall.