

Adjusting for Missing Information in Pooled Model through the Error Term

Dr. Abdalla Ahmed Alkhalifa Abdalla

Department of statistics-faculty of Science – Tabuk University (KSA)

Abstract: Griffiths and Anderson (1982) used nonlinear model to adjust for missing information through the error term, this paper adjust for missing information in linear model through the error term.

Keywords: firm effect; time effect; qualitative variable; error term; linear model

1. Introduction

If time series data and cross section data are combined "Pooled data", it's expected that the explanatory variables will not explain all variations in the dependent variable, because there may be variations due to time (time-series) or due to firm (cross-section) or due to qualitative variables in the model. The linear model for pooled data is characterized by the problem of missing information due to that, the explanatory variables are not account for all variations in the dependent variable.

Griffiths and Anderson (1982), used pooled data to estimate a production function of pastoral zone of eastern Australia, they used **nonlinear** model and adjusting for missing information through the error term u_{it} for firm ".i" and time "t", i.e. The error term as a function of time effect "g_t" and firm effect "f_i" and qualitative variables "h(x)", in two cases:

1. $u_{it} = (f_i + g_t + e_{it}) h_{it}$
2. $u_{it} = f_i + g_t + e_{it} h_{it}$

The objective of this paper is to find out how the **linear** model estimated when data are pooled, putting in consideration time and firm effects and the variables that are qualitative to adjust for missing information in the linear model using the first case of Griffiths and Anderson.

2. The Model

$$y_{it} = \alpha + \sum_{k=1}^k b_k x_{kit} + u_{it}$$

Where:

y_{it} is the observations of the dependent variable for firm .i at time t.

X_{kit} is the observation of the independent variable for firm .i at time t for the k^{th} regressor.

α_i and b_k are parameters.

u_{it} is the error term associated with y_{it} .

To adjust for missing information through the error term we let:

$$u_{it} = (f_i + g_t + e_{it}) h_{it}$$

Properties of u_{it} are:

1. $E(u_{it}) = 0$

$$\begin{aligned} 2. E(u_{it}u_{js}) &= E[(f_i + g_t + e_{it}) h_{it}][(f_j + g_s + e_{js}) h_{js}] \\ &= E(f_i f_j h_{it} h_{js}) + E(f_i g_s h_{it} h_{js}) + E(f_i e_{js} h_{it} h_{js}) \\ &\quad + E(g_t f_j h_{it} h_{js}) + E(g_t g_s h_{it} h_{js}) + E(g_t e_{js} h_{it} h_{js}) \\ &\quad + E(e_{it} f_j h_{it} h_{js}) + E(e_{it} g_s h_{it} h_{js}) + E(e_{it} e_{js} h_{it} h_{js}) \\ &= (\sigma_f^2 + \sigma_g^2 + \sigma_e^2) h_{it}^2 \dots \text{for } i = j, t = s \\ &= \sigma_f^2 h_{it} h_{js} \dots \text{for } i = j, t \neq s \\ &= \sigma_g^2 h_{it} h_{js} \dots \text{for } i \neq j, t = s \\ &= 0 \dots \text{otherwise} \end{aligned}$$

The variance covariance matrix is:

$$v = E(uu')$$

$$= E \begin{pmatrix} u_1 u_1' & u_1 u_2' & \dots & u_1 u_N' \\ u_2 u_1' & u_2 u_2' & \dots & u_2 u_N' \\ \cdot & \cdot & \dots & \cdot \\ u_N u_1' & u_N u_2' & \dots & u_N u_N' \end{pmatrix}$$

Let:

$$E(u_i u_i') = c_{ii}$$

$$= \begin{pmatrix} u_{i1}^2 & u_{i1} u_{i2} & \dots & u_{i1} u_{iT} \\ u_{i2} u_{i1} & u_{i2}^2 & \dots & u_{i2} u_{iT} \\ \cdot & \cdot & \dots & \cdot \\ u_{iT} u_{i1} & u_{iT} u_{i2} & \dots & u_{iT}^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_f^2 h_{i1}^2 & \sigma_f^2 h_{i1} h_{i2} & \dots & \sigma_f^2 h_{i1} h_{iT} \\ \sigma_f^2 h_{i1} h_{i2} & \sigma_f^2 h_{i2}^2 & \dots & \sigma_f^2 h_{i2} h_{iT} \\ \cdot & \cdot & \dots & \cdot \\ \sigma_f^2 h_{iT} h_{i1} & \sigma_f^2 h_{iT} h_{i2} & \dots & \sigma_f^2 h_{iT}^2 \end{pmatrix}$$

And:

$$E(u_i u_j') = c_{ij}$$

$$= \begin{pmatrix} u_{i1} u_{j1} & u_{i1} u_{j2} & \dots & u_{i1} u_{jT} \\ u_{i2} u_{j1} & u_{i2} u_{j2} & \dots & u_{i2} u_{jT} \\ \cdot & \cdot & \dots & \cdot \\ u_{iT} u_{j1} & u_{iT} u_{j2} & \dots & u_{iT} u_{jT} \end{pmatrix}$$

And:

$$= \sigma^2 \begin{pmatrix} h_{i1}h_{j1} & 0 & 0 & \dots & 0 \\ 0 & h_{i2}h_{j2} & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 0 & h_{iT}h_{jT} \end{pmatrix}$$

Therefore:

$$v = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1N} \\ c_{21} & c_{22} & \dots & c_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ c_{N1} & c_{N2} & \dots & c_{NN} \end{pmatrix}$$

Which can be written as :

$$V = H \phi H$$

Where,

$$\phi = \sigma_f^2(I_N \otimes J_T) + \sigma_g^2(J_N \otimes I_T) + \sigma_e^2 I_{NT}$$

$$v^{-1} = H^{-1} \phi^{-1} H^{-1}$$

Then v^{-1} can be written as:

$$v^{-1} = H^{-1} [I - z(z'z + D^{-1})^{-1} z'] H^{-1}$$

Where,

$$z = [(I_N \otimes I_T)(I_N \otimes I_T)']$$

$$D = \begin{pmatrix} \sigma_f^2 & 0 \\ 0 & \sigma_g^2 \end{pmatrix}$$

And if the variance components are known, the generalized least square estimates are:

$$\hat{\beta} = (x'v^{-1}x)^{-1} x'v^{-1}y$$

3. Conclusion

It can be concluded that missing information can be captured through the error term.

4. Recommendation

Other papers can be done in the same topic to cover other methods used by Griffiths and Anderson.

References:

- [1] Griffiths William E. and Anderson Jock R-using time series and cross section data to estimate a production function with positive and negative marginal risk, journal of the American statistical association, vol 77, September 1982.
- [2] Judge et.al ,introduction to the theory and practice of econometrics, second edition, 1988.
- [3] Mandlackyair, on pooling time series and cross section data, Econometrica, vol 46, January 1978.
- [4] Terry E. Dielman, Pooled data for financial markets, second edition, 1980.
- [5] Wonnacott Ronald J. and Wonnacott Thomas H., Econometrics, second edition, 1979.