Economics Letters MS#EL39358
An equicorrelation Moulton factor in the presence of arbitrary intra-cluster correlation*

Gabriel Montes-Rojas

CONICET-Universidad de San Andrés and City University London

January 20, 2016

Abstract

This note highlights the potential pitfalls of using an equicorrelation model to estimate standard errors when the true model has arbitrary intra-cluster correlation. It derives a generalized equicorrelation Moulton factor that quantifies the potential biases in standard errors for OLS estimators. As with the famous Moulton factor, the key role is not played by the correlation of the error terms but rather by the intra-correlation of the covariates themselves.

Keywords: Microeconometrics; Clusters; Aggregate variables; Moulton factor.

JEL Classification: C2; C12.

Word count: 1992

---

*The author expresses his gratitude to the constructive comments of the Editor Prof. Badi Baltagi and an anonymous reviewer which have helped greatly to improve the paper, and to Prof. Walter Sosa-Escudero for productive discussions on the topic. All remaining errors are mine.

1 Introduction

Statistical inference when data are grouped into clusters is an important issue in empirical work, and failure to control for within-cluster correlation can lead to misleadingly small standard errors (see the discussion in Cameron and Miller, 2015). This is especially important when using aggregate variables on micro units in which OLS standard errors are seriously underestimated. The seminal work of Moulton (1986, 1987, 1990) allows for a quantification of this potential pitfall, a fact that has been emphasized in the Angrist and Pischke (2009, ch.8) textbook among many others.

The most obvious type of intra-group correlation arises when all observations within a group share an unobserved common factor, hence all observations in a group are 'equicorrelated' in the sense that all pairwise correlations are the same. Beyond equicorrelation little can be said if observations within a group do not follow a relevant ordering (i.e. time, spatial).

The goal of this note is to evaluate potential misspecification in estimating the OLS standard errors using an equicorrelation model when the true underlying data generating process has arbitrary intra-cluster correlation, i.e. not necessarily constant among intra-cluster observations, and where there is no intra-cluster obvious ordering (i.e. students within a class). We then define the *equicorrelation Moulton factor* as the difference between the true variance-covariance matrix of the OLS estimator and that of an assumed equicorrelation model.

As with the famous Moulton factor, the key role is given by the joint consideration of the intra-cluster correlation of the error term and the covariates. More formally, given an intra-cluster covariance structure of the error term and one of the covariates, the comparison of the equicorrelation and an arbitrary intra-cluster correlation model will depend on the sample intra-cluster covariance between the covariance factors of the error term and the covariates.

In most empirical settings, both covariance factors are positively correlated (i.e. a high correlation between two unobservables usually corresponds to a high correlation between the covariates), and thus this determines that the equicorrelation model would underestimate the true variance, thus acting in the same way as the OLS Moulton factor.

The OLS Moulton factor shows that in the special case of covariates with no intra-cluster correlation, the standard OLS variance is correct. Our analysis also shows that in the special case of constant intra-cluster covariates (eg. aggregate variables), the equicorrelation model is also correct. In practical terms, if the within cluster correlation of the covariates is small, OLS standard errors are approximately correct, while if the correlation is large, random-effects equicorrelation standard errors are appropriate.

The results determine that in an OLS model with arbitrary intra-cluster correlation Liang and Zeger (1986) and Arellano (1987) extension of White (1980) variance estimate for heteroskedasticity to the cluster set-up, defined as White's cluster-robust standard errors, should be used rather than an equicorrelation model. In fact, as noted by Angrist and Pischke (2009, ch.8) "The clustered variance estimator [...] is consistent as the number of groups gets large under any within-group correlation structure." (p.313) Wooldridge (2010) recommends to implement the random-effects estimator, which is likely to be more efficient than pooled OLS, even when the intra-cluster error structure model does not follow equicorrelation and is unknown, but "to make the variance estimator of the random effects robust to arbitrary heteroskedasticity and within-group correlation" (p.867).

However, while cluster-robust standard errors is a safe approach, it should be noted that its asymptotic validity crucially depends on a large number of clusters (i.e. $N \to \infty$). If the equicorrelation model were true, asymptotic valid inference and efficiency can be achieved for fixed N and $T \to \infty$ (eg. consider the random-effects GLS estimator, see Hsiao, 2003, p.38).

2 One-way error components model

Consider the one-way error components regression model (see Baltagi, 2013, ch.2)

$$y_{it} = x_{it}\beta + e_{it}, \tag{1}$$

$$e_{it} = \mu_i + \nu_{it},$$

$$i = 1, 2, ..., N, \ t = 1, 2, ..., T.$$

Assume that the t-ordering cannot be used to evaluate intra-cluster correlation because the ordering is not known by the econometrician, that is, we do not know the structure of network relationships among observations within a cluster.

In matrix notation the model above can be written as $\boldsymbol{y} = \boldsymbol{x}\beta + \boldsymbol{e}$, where $\boldsymbol{y}$ and $\boldsymbol{e}$ are $NT \times 1$ matrices, $\boldsymbol{x}(= [x_1', ..., x_N']' = [x_{11}', ..., x_{1T}', ..., x_{N1}', ..., x_{NT}']'$, x_i $T \times K$ matrices, x_{it} $1 \times K$ vectors) and β are matrices of dimensions $NT \times K$ and $K \times 1$, respectively. Moreover consider the NT-dimensional vector $\boldsymbol{\nu}(= [\nu_1', \nu_2', ..., \nu_N']'$, ν_i $T \times 1$ vectors) and the N-dimensional vector $\boldsymbol{\mu}$ such that $\boldsymbol{e} = \boldsymbol{\mu} \otimes \iota_T + \boldsymbol{\nu}$, where ι_T is a $T \times 1$ vector of 1s and $\otimes$ is the Kronecker product. Consider the OLS estimator $\hat{\boldsymbol{\beta}} = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y}$, and consider the goal of estimating $Var[\hat{\boldsymbol{\beta}}|\boldsymbol{x}]$.

A natural concern in such models is the possibility of intra-group correlation in the error term e_{it}. Naturally, the presence of μ_i induces correlation for all observations corresponding to a certain 'group' (class, school, country, industry) i. As a matter of fact, due to this factor all correlations among error terms within a group are the same, this correlation induced by the presence of the random effect μ_i is labelled as *equicorrelation*. A second source of intra-group correlation is the possibility that the ν_{it} terms are correlated among themselves within the group.

Consider the following assumptions.

Assumptions

(i) $E[\nu_{it}|x_i] = E[\mu_i|x_i] = 0, \forall i, t;$

(ii) $Var[\mu_i|x_i] = \sigma_\mu^2$, $Var[\nu_{it}|x_i] = \sigma_\nu^2$, $\forall i, t;$

(iii) $Cov[\mu_i, \nu_{it}|x_i] = 0, \forall i, t.$

(iv) $Cov[\nu_{it}, \nu_{ij}|x_i] = \rho_i(t,j)\sigma_\nu^2, \rho_i(t,j) = \rho_i(j,t), \forall i, t \neq j.$

For simplicity we assume homoskedastic models, that the intra-cluster correlation is the same across groups, and a balanced panel. We allow for arbitrary within group correlation structure. We do not impose a structure to the function $\rho(.,.)$, other than symmetry and other requirements for positive-definiteness of the variance-covariance matrix. In the case of time-series or spatial correlation we have additional information about the intra-cluster correlation structure, which in turn, can be used to identify the relevant parameters (eg. AR(1) serial correlation in which $\rho(t,j) = \rho^{|t-j|}, 0 \leq |\rho| < 1$ or spatial correlation in which $\rho(t,j) = f(dist(t,j))$).

Define the average ν-correlation as

$$\bar{\rho}_\nu := \frac{1}{\sigma_\nu^2}\frac{2}{T(T-1)}\sum_{t=1}^{T-1}\sum_{j=t+1}^{T} E[\nu_{it}\nu_{ij}] = \frac{2}{T(T-1)}\sum_{t=1}^{T-1}\sum_{j=t+1}^{T} \rho_i(t,j), \tag{2}$$

and let

$$\lambda_\nu^2 := \sigma_\nu^2(1-\bar{\rho}_\nu), \tag{3}$$

and

$$\lambda_\mu^2 := \sigma_\mu^2 + \sigma_\nu^2\bar{\rho}_\nu. \tag{4}$$

Finally define the intra-group correlation as

$$IC := \frac{2}{T(T-1)}\sum_{t=1}^{T-1}\sum_{j=t+1}^{T} \frac{E[e_{it}e_{ij}]}{\sqrt{Var(e_{it})}\sqrt{Var(e_{ij})}} = \frac{\sigma_\mu^2 + \sigma_\nu^2\bar{\rho}_\nu}{\sigma_\mu^2 + \sigma_\nu^2} = \frac{\lambda_\mu^2}{\lambda_\mu^2 + \lambda_\nu^2}. \tag{5}$$

The key point is that without a known error structure nothing can be learned beyond equicorrelation. Note that an equicorrelation model with $\sigma^2_\mu > 0$ and $\bar{\rho}_\nu = 0$ may have the same IC as one with $\sigma^2_\mu = 0$ and $\bar{\rho}_\nu \neq 0$. In fact, $(\sigma^2_\nu, \sigma^2_\mu, \bar{\rho}_\nu)$ cannot be identified as separate objects: only linear combinations of λ^2_ν and λ^2_μ can be estimated.[1]

3 Equicorrelation Moulton factor

Given the assumptions of the model, then consider

$$\begin{aligned} \Omega := E[\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{x}] &= E[diag(\boldsymbol{\nu}_i\boldsymbol{\nu}_i') + diag(\mu_i^2(\iota_T\iota_T'))|\boldsymbol{x}] \\ &= E[(\boldsymbol{\nu}_i\boldsymbol{\nu}_i') \otimes I_N|\boldsymbol{x}] + E[\mu_i^2(\iota_T\iota_T') \otimes I_N|\boldsymbol{x}]. \end{aligned} \tag{6}$$

Then

$$Var[\hat{\boldsymbol{\beta}}|\boldsymbol{x}] = (\boldsymbol{x}'\boldsymbol{x})^{-1}(\boldsymbol{x}'\Omega\boldsymbol{x})(\boldsymbol{x}'\boldsymbol{x})^{-1}.$$

Note that Ω acts as a selector and weighting matrix, which selects which row and columns of $\boldsymbol{x}$ should be considered and weights them accordingly.

- In the $i.i.d.$ case, $\Omega_0 := (\lambda^2_\nu + \lambda^2_\mu)I_{NT}$, and thus only the xs that correspond to the same values of i and t are considered.

- The random-effects equicorrelation matrix would consider a different weight for those observations (i, t) but would also weight all observations *within* the same i, thus producing $\Omega_e := \lambda^2_\nu I_{NT} + \lambda^2_\mu(\iota_T\iota_T') \otimes I_N$.

[1]Consider a list of within cluster transformations of the residuals. Define $\bar{e}_i = \frac{1}{T}\sum_{t=1}^T e_{it}$ as the group-average transformation and $\tilde{e}_{it} = e_{it} - \bar{e}_i$ as the within-group deviations. Moreover, let $\overline{e^2_{it}} = \frac{1}{T}\sum_{t=1}^T e^2_{it}$, $\overline{\tilde{e}^2_{it}} = \frac{1}{T}\sum_{t=1}^T \tilde{e}^2_{it}$, $\check{e}_i = \frac{2}{T(T-1)}\sum_{t=1}^{T-1}\sum_{j=t+1}^T e_{it}e_{ij}$ and $\tilde{e}_i = \frac{2}{T(T-1)}\sum_{t=1}^{T-1}\sum_{j=t+1}^T \tilde{e}_{it}\tilde{e}_{ij}$. Simple calculations determine that $\phi_0 = E[\overline{e^2_{it}}] = E[e^2_{it}] = \sigma^2_\nu + \sigma^2_\mu$, $\phi_1 = E[\bar{e}^2_i] = \frac{1}{T}[\sigma^2_\nu(1-\bar{\rho}_\nu)] + (\sigma^2_\mu + \sigma^2_\nu\bar{\rho}_\nu)$, $\phi_2 = E[\overline{\tilde{e}^2_{it}}] = \frac{T-1}{T}[\sigma^2_\nu(1-\bar{\rho}_\nu)]$, $\phi_3 = E[\check{e}_i] = (\sigma^2_\mu + \sigma^2_\nu\bar{\rho}_\nu)$ and $\phi_4 = E[\tilde{e}_i] = -\frac{1}{T}(\sigma^2_\nu(1-\bar{\rho}_\nu))$. Note that by analysis of variance decompositions $\phi_0 = \phi_1 + \phi_2$ and $\phi_3 - \phi_1 = -\frac{1}{(T-1)}\phi_2 = \phi_4$. Thus we can only obtain linear functions of λ^2_ν and λ^2_μ using ANOVA type analysis.

- In an arbitrary intra-cluster correlation $(\iota_T\iota_T')$ should be changed by an arbitrary $T\times T$ symmetric matrix, say P_T, with typical element $\{\rho_{th}\}_{t=1,h=1}^{T,T}$ with $\rho_{tt}=0, t=1,2,...,T$ and $\rho_{th}=\rho_{ht}, t,h=1,2,...,T$, such that $\Omega_w := \sigma_\nu^2 I_{NT}+\sigma_\nu^2 P_T\otimes I_N+\sigma_\mu^2(\iota_T\iota_T')\otimes I_N$.

Then for the equicorrelation model

$$\begin{aligned}
\boldsymbol{x}'\Omega_e\boldsymbol{x} &= \sum_{i=1}^{N}\left(\lambda_\nu^2 x_i'x_i+\lambda_\mu^2 x_i'(\iota_T\iota_T')x_i\right)=\sum_{i=1}^{N}\left(\lambda_\nu^2\sum_{t=1}^{T}x_{it}'x_{it}+\lambda_\mu^2\sum_{t=1}^{T}\sum_{h=1}^{T}x_{it}'x_{ih}\right)\\
&= \sum_{i=1}^{N}\left((\lambda_\nu^2+\lambda_\mu^2)\sum_{t=1}^{T}x_{it}'x_{it}+\lambda_\mu^2\sum_{t=1}^{T}\sum_{h=1,h\neq t}^{T}x_{it}'x_{ih}\right),
\end{aligned}$$

and for the arbitrary intra-cluster model

$$\begin{aligned}
\boldsymbol{x}'\Omega_w\boldsymbol{x} &= \sum_{i=1}^{N}\left(\sigma_\nu^2 x_i'x_i+\sigma_\nu^2 x_i'P_Tx_i+\sigma_\mu^2 x_i'(\iota_T\iota_T')x_i\right)\\
&= \sum_{i=1}^{N}\left(\sigma_\nu^2\sum_{t=1}^{T}x_{it}'x_{it}+\sigma_\nu^2\sum_{t=1}^{T}\sum_{h=1,h\neq t}^{T}\rho_{th}x_{it}'x_{ih}+\sigma_\mu^2\sum_{t=1}^{T}\sum_{h=1}^{T}x_{it}'x_{ih}\right)\\
&= \sum_{i=1}^{N}\left((\sigma_\nu^2+\sigma_\mu^2)\sum_{t=1}^{T}x_{it}'x_{it}+\sum_{t=1}^{T}\sum_{h=1,h\neq t}^{T}(\sigma_\mu^2+\sigma_\nu^2\rho_{th})x_{it}'x_{ih}\right)\\
&= \sum_{i=1}^{N}\left((\lambda_\nu^2+\lambda_\mu^2)\sum_{t=1}^{T}x_{it}'x_{it}+\sum_{t=1}^{T}\sum_{h=1,h\neq t}^{T}(\sigma_\mu^2+\sigma_\nu^2\rho_{th})x_{it}'x_{ih}\right).
\end{aligned}$$

The main difference between the two models is that not all intra-cluster pairs in x_i are weighted the same. That is, in an equicorrelation model $x_{it}'x_{ih}$ will receive the same weight for all $t\neq h$, while in an arbitrary intra-cluster model the weights will depend on ρ_{th}.

The equicorrelation Moulton factor is defined as the $K\times K$ matrix

$$M^{w-e} := (\boldsymbol{x}'\boldsymbol{x})^{-1}(\boldsymbol{x}'\Omega_w\boldsymbol{x}-\boldsymbol{x}'\Omega_e\boldsymbol{x})(\boldsymbol{x}'\boldsymbol{x})^{-1}.$$

Since $x_{it}'x_{ih}$ is a $K\times K$ matrix, an element-by-element analysis is necessary. Its diagonal elements correspond to the difference in the variance of each

$\hat{\beta}_k,\ k = 1, 2..., K$, while the off-diagonal terms to the covariances of the β's parameter estimates. It is possible that the sign of different diagonal elements in M^{w-e} are different, and then, the equicorrelation may underestimate the variance for some coefficient estimators and overestimate for others.

Note that the sample covariance of $(\sigma_\mu^2 + \sigma_\nu^2 \rho_{th})$ and $x'_{it}x_{ih}$ across $t, h = 1, 2, ..., T, t \neq h$, provides additional information that is not captured by the equicorrelation model. The following proposition compares the equicorrelation and the arbitrary intra-cluster correlation variances in terms of the covariance between these two elements.

Proposition 1. *(i) If the sum (or average) over i of the sample covariances of $(\sigma_\mu^2 + \sigma_\nu^2 \rho_{th})$ and $x_{it}^{k\prime} x_{ih}^j$, where $k, j = 1, 2, ..., K$ correspond to the k and j columns of x_{it}, is positive, negative or zero, then $\boldsymbol{x}^{k\prime}(\Omega_w - \Omega_e)\boldsymbol{x}^j$ is positive, negative or zero, respectively.*
(ii) If the sum (or average) over i of the sample covariances is positive, negative or zero, for all $k, j = 1, 2, ..., K$, then M^{w-e} is positive definite, negative definite or zero, respectively.

Proof. Note that by definition, the sample covariance is

$$
\begin{aligned}
&= \frac{\sum_{t=1}^T \sum_{h=1,h\neq t}^T (\sigma_\mu^2+\sigma_\nu^2\rho_{th}) x_{it}^{k\prime} x_{ih}^j}{\frac{1}{2}T(T-1)} - \frac{\left(\sum_{t=1}^T \sum_{h=1,h\neq t}^T (\sigma_\mu^2+\sigma_\nu^2\rho_{th})\right)\left(\sum_{t=1}^T \sum_{h=1,h\neq t}^T x_{it}^{k\prime} x_{ih}^j\right)}{\frac{1}{4}T^2(T-1)^2} \\
&= \frac{\sum_{t=1}^T \sum_{h=1,h\neq t}^T (\sigma_\mu^2+\sigma_\nu^2\rho_{th}) x_{it}^{k\prime} x_{ih}^j}{\frac{1}{2}T(T-1)} - \frac{(\sigma_\mu^2+\sigma_\nu^2\bar{\rho}_\nu)\left(\sum_{t=1}^T \sum_{h=1,h\neq t}^T x_{it}^{k\prime} x_{ih}^j\right)}{\frac{1}{2}T(T-1)} \\
&= \frac{\sum_{t=1}^T \sum_{h=1,h\neq t}^T (\sigma_\mu^2+\sigma_\nu^2\rho_{th}) x_{it}^{k\prime} x_{ih}^j}{\frac{1}{2}T(T-1)} - \frac{\lambda_\mu^2\left(\sum_{t=1}^T \sum_{h=1,h\neq t}^T x_{it}^{k\prime} x_{ih}^j\right)}{\frac{1}{2}T(T-1)} \\
&= \frac{2}{T(T-1)}\left(x_i^{k\prime}\Omega_w x_i^j - x_i^{k\prime}\Omega_e x_i^j\right).
\end{aligned}
$$

Note that $\boldsymbol{x}^{k\prime}\Omega\boldsymbol{x}^j = \sum_{i=1}^N (x_i^{k\prime}\Omega x_i^j)$ and then the sign of the sum of the sample covariances determines the sign in *(i)*. For *(ii)* note that if the sign is the same for all k, j, then $(\boldsymbol{x}'\Omega_w\boldsymbol{x} - \boldsymbol{x}'\Omega_e\boldsymbol{x})$ is a matrix whose elements have the same corresponding sign. Then given that $(\boldsymbol{x}'\boldsymbol{x})^{-1}$ is positive definite the result follows. □

A special case is when the covariates have no intra-cluster variation, as in aggregate control variables or fixed characteristics of the individual (eg. gender, nationality, etc.). For two covariates k, j that are constant within cluster, the sample covariance of $(\sigma_\mu^2 + \sigma_\nu^2 \rho_{th})$ and $x_{it}^{k'} x_{ih}^{j}$ is zero, and therefore, the equicorrelation model is correct in the presence of any intra-cluster correlation structure.

4 Example with unknown intra-cluster correlation

In order to quantify the potential consequences of estimating an equicorrelation model when the underlying data generating process has arbitrary intra-cluster correlation, we consider a simple regression model of the form

$$y_{it} = x_{it}\beta + e_{it}, \tag{7}$$

$$e_{it} = \nu_{it} = \epsilon_{it} + \rho\epsilon_{it-1}, 0 < \rho < 1, \epsilon_{i0} = 0,$$

$$i = 1, 2, ..., N,\ t = 1, 2, ..., T.$$

Each observation (i, t) corresponds to the observation of individual t in group i. x is a scalar (assume for simplicity with mean 0), $\epsilon \sim i.i.d.(0, \sigma_\epsilon^2)$. This is of course a simple MA(1) model, but we assume we do not know the t-ordering.

Consider the objects defined in eqs. (2)-(4) $\lambda_\nu^2 = \sigma_\nu^2(1 - \bar{\rho}_\nu)$, $\lambda_\mu^2 = \sigma_\nu^2 \bar{\rho}_\nu = 2/T\rho\sigma_\epsilon^2$, $\bar{\rho}_\nu = \frac{2/T\rho}{1+\rho^2}$ (note the factor $1/T$), and $\lambda_\mu^2 + \lambda_\nu^2 = \sigma_\nu^2 = (1+\rho^2)\sigma_\epsilon^2$. Then define $IC = \frac{\lambda_\mu^2}{\lambda_\mu^2 + \lambda_\nu^2} = \frac{2/T\rho}{(1+\rho^2)}$ as the intra-cluster correlation (as in eq. (5)).

Consider now three different estimators of the variance of the OLS estimator $\hat{\beta} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}x_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2}$.

$V_0 := V_0(\hat{\beta}|x)$ is the standard OLS variance estimate given by

$$V_0 = \frac{(1+\rho^2)\sigma_\epsilon^2 \sum_{i=1}^N \sum_{t=1}^T x_{it}^2}{\left(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2\right)^2} = \frac{(\lambda_\mu^2 + \lambda_\nu^2) \sum_{i=1}^N \sum_{t=1}^T x_{it}^2}{\left(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2\right)^2}.$$

$V_1 := V_1(\hat{\beta}|x)$ is the correct variance given by

$$\begin{aligned} V_1 &= \frac{(1+\rho^2)\sigma_\epsilon^2 \sum_{i=1}^N \sum_{t=1}^T x_{it}^2 + 2\rho\sigma_\epsilon^2 \sum_{i=1}^N \sum_{t=2}^T x_{it}x_{it-1}}{\left(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2\right)^2} \\ &= \frac{(\lambda_\mu^2 + \lambda_\nu^2) \sum_{i=1}^N \sum_{t=1}^T x_{it}^2 + \lambda_\mu^2 T \sum_{i=1}^N \sum_{t=2}^T x_{it}x_{it-1}}{\left(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2\right)^2}. \end{aligned}$$

Consider now the variance assuming equicorrelation, $V_2 := V_2(\hat{\beta}|x)$,

$$V_2 = \frac{(\lambda_\mu^2 + \lambda_\nu^2) \sum_{i=1}^N \sum_{t=1}^T x_{it}^2 + 2\lambda_\mu^2 \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{j=t+1}^T x_{it}x_{ij}}{\left(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2\right)^2}.$$

Define $\rho_x^{(1)} = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T x_{it}x_{it-1}$ as the average of the intra-cluster x sample autocovariance of order 1, and $\rho_x^{(T)} = \frac{2}{NT(T-1)} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{j=t+1}^T x_{it}x_{ij}$ is the average of the intra-cluster sample covariance of the xs that uses are intra-cluster observations.

The Moulton (1986,1987,1990) factor naturally arises as

$$\frac{V_1}{V_0} = 1 + IC(T-1)\rho_x^{(1)}.$$

If we assume that both $IC > 0$ and $\rho_x^{(1)} > 0$ then the standard OLS variance wrongly underestimates the true variance.

A generalization of the Moulton factor allows comparing it with other different models, such as the equicorrelation model. In this case,

$$\frac{V_1}{V_2} = \frac{1 + IC(T-1)\rho_x^{(1)}}{1 + IC(T-1)\rho_x^{(T)}}.$$

As a result, the only difference between the two is in the appropriate correlation of the xs that needs to be used. The equicorrelation model assumes that all potential interactions among the xs are needed to calculate the OLS variance, while the correct MA(1) uses only those that are one t apart. Given that we have a MA(1) structure in the error terms, such that the unobservable term in t is correlated with $t-1$ only, it is also likely that the x component follows a similar pattern of intra-cluster correlation. Then, we could assume that $\rho_x^{(1)} > \rho_x^{(T)}$, that is, the correlation between the t and $t-1$ xs is higher than the average correlation among all the xs within the cluster. In this case the equicorrelation model would be unde-estimating the true variance. Note that, although less likely, it may also be the case that $\rho_x^{(1)} < \rho_x^{(T)}$, in which case the equicorrelation model would be overestimating the true variance. Note that if aggregate covariates are used (i.e. with no intra-cluster variation), then $\rho_x^{(1)} = \rho_x^{(T)} = 1$, and thus, the equicorrelation model is appropriate.

References

- Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press, Princeton.

- Arellano, M. (1987). "Computing robust standard errors for within-group estimators," *Oxford Bulletin of Economic and Statistics* 49, 431-434.

- Baltagi, B.H. (2013) *Econometric Analysis of Panel Data* (fifth ed.) John Wiley & Sons.

- Cameron, C. and Miller, D.L. (2015). "A practicioner's guide to cluster-robust inference," *Journal of Human Resources* 50(2), 317-372.

- Hsiao, C. (2003) *Analysis of Panel Data* (second ed.) Cambridge University

Press.
- Moulton, B.R. (1986). “Random group effects and the precision of regression estimates,” *Journal of Econometrics* 32, 385-397.
- Moulton, B.R. (1987). “Diagnostics for group effects in regression analysis,” *Journal of Business & Economic Sataistics* 5(2), 275-282.
- Moulton, B.R. (1990). “An illustration of a pitfall in sstimating the effects of aggregate variables in micro units,” *Review of Economics and Statistics* 72, 334-338.
- White, H. (1980). “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica* 48, 817-838.
- Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data* (second ed.) MIT Press.