Identifying School Climate Variables Associated with Students' Financial Literacy Outcomes

A Cross-Country Comparison Using PISA 2018 Data

Tony C. A. Tan

Thesis submitted for the degree of
Master in Assessment, Measurement and Evaluation
30 credits

Centre for Educational Measurement
Faculty of Educational Sciences

UNIVERSITY OF OSLO

Spring 2021

Identifying School Climate Variables Associated with Students' Financial Literacy Outcomes

A Cross-Country Comparison Using PISA 2018 Data

Tony C. A. Tan

Identifying School Climate Variables Associated with Students' Financial Literacy Outcomes

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

敬致父母

To my parents

Study hard what interests you the most in the most undisciplined, irreverent and original manner possible.

Richard P. Feynman

Contents

List of Tables

List of Figures

Acknowledgement

Thank-you goes to

Popular Abstract

This is a press release style abstract.

Abstract

Repeated economic crises in recent memory have exposed the harsh consequences of financial *illiteracy* shared by high proportions of the general population. Policy makers experienced little resistance when identifying youth as the most effective group for bringing about improvement in citizens' ability to engage with economic and financial matters, but opinions quickly diverge over the optimal approaches for achieving such targeted outcome. Existing literature frequently reports the importance of family environment in cultivating students' financial literacy through the process of "financial socialisation" – [definition goes here] (reference). Such practice, however, encounters interrogation by educators over equity concerns should families remain the main arena for financial literacy development. Schools play vital roles in alleviating inequality in accessing education and training in general but scarce research so far has been devoted into identifying the specific classroom factors that are most effective in advancing students' financial literacy outcomes. The current study therefore attempts to contribute to this enquiry by investigating the relationship between school climate variables and students' financial literacy achievement with an aim of stimulating policy debate about the levers and instruments available to education interventionists for the purpose of improving young people's financial literacy and preparedness as they step into an increasingly uncertain world. Using the 2018 PISA dataset, this paper employs a three-level hierarchical model to conduct cross-country comparisons to highlight school climate variables that are most strongly associated with high financial literacy outcomes.

Appendices

Appendix A Review of Matrix Calculus

A.1 Notations

Let us first establish the notation. This is important because bad notation is a serious obstacle to elegant mathematics and coherent exposition and it can be misleading.
Unless specified otherwise, φ denotes a scalar function; $\boldsymbol{f}$ a vector function and $\boldsymbol{F}$ a matrix function. Also, x denotes a scalar argument, $\boldsymbol{x}$ a vector argument and $\boldsymbol{X}$ a matrix argument. For example, we write

$$\begin{array}{lll} \varphi(x) = x^2 & \varphi(\boldsymbol{x}) = \boldsymbol{a}^\intercal \boldsymbol{x} & \varphi(\boldsymbol{X}) = \operatorname{tr}\left(\boldsymbol{X}^\intercal \boldsymbol{X}\right) \\ \boldsymbol{f}(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix} & \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} & \boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{a} \\ \boldsymbol{F}(x) = x^2 \boldsymbol{I}_m & \boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{x}^\intercal & \boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{X}^\intercal \end{array}$$

Since the prime notation $'$ may easily cause confusion between derivatives and transposes, preference is given to the Leibniz notation $\frac{\mathrm{d}}{\mathrm{d}x}$ for derivatives and $^\intercal$ for transposes—unless this system becomes too cumbersome, in which case $\boldsymbol{f}'(\boldsymbol{x})$ will denote derivatives and $\boldsymbol{f}(\boldsymbol{x})'$ for transposes.

A.2 Derivatives and differentials

A.2.1 Derivative

Definition A.2.1 (Derivatives)**.** If $\boldsymbol{f}$ is an $m \times 1$ vector function of an $n \times 1$ vector $\boldsymbol{x}$, then the *derivative* (or *Jacobian matrix*) of $\boldsymbol{f}$ is the $m \times n$ matrix

$$\mathrm{D}\boldsymbol{f}(\boldsymbol{x}) := \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^\intercal}, \tag{A.1}$$

whose elements are the partial derivatives

$$\frac{\partial f_i(\boldsymbol{x})}{\partial x_j}, \text{ for } \begin{array}{l} i = 1, \cdots, m, \\ j = 1, \cdots, n. \end{array}$$

A.2.2 Differential

In the one dimensional case, the equation

$$\lim_{u \to 0} \frac{\varphi(x+u) - \varphi(x)}{u} = \varphi'(x) \tag{A.2}$$

defines the derivative of φ at x. Rewriting Equation (A.2) gives

$$\varphi(x+u) = \varphi(x) + \varphi'(x)u + O(u), \tag{A.3}$$

where the remainder term $O(u)$ quickly vanishes as u approaches 0.

Definition A.2.2 (Differential)**.** We define the (first) *differential* of φ at x (with increment u) as

$$\mathrm{d}\varphi(x; u) = \varphi'(x)u. \tag{A.4}$$

For example, for $\varphi(x) = x^2$, we have $\mathrm{d}\varphi(x; u) = 2xu$. In practice, we write $\mathrm{d}x$ instead of u, so that $\mathrm{d}\varphi(x) = \varphi'(x)\mathrm{d}x = 2x\mathrm{d}x$.
In the vector case, similar to Equation (A.3), we have

$$\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{u}) = \boldsymbol{f}(\boldsymbol{x}) + [\mathrm{D}\boldsymbol{f}(x)]\boldsymbol{u} + O(\boldsymbol{u}), \tag{A.5}$$

and the (first) differential is defined as

$$\mathrm{d}\boldsymbol{f}(\boldsymbol{x}; \boldsymbol{u}) = [\mathrm{D}\boldsymbol{f}(x)]\boldsymbol{u}. \tag{A.6}$$

Although rarely used in econometrics, for completeness, the matrix case can be obtained from the vector case by writing $\boldsymbol{f} := \mathrm{vec}\,(\boldsymbol{F})$ and $\boldsymbol{x} := \mathrm{vec}\,(\boldsymbol{X})$.

A.2.3 Which to use?

For practical rather than theoretical reasons, the treatment of matrix calculus is based on differentials ($\mathrm{d}\boldsymbol{f}$) rather than derivatives ($\mathrm{D}\boldsymbol{f}$) because the former yields a result with the same dimension as $\boldsymbol{f}$. For example, consider $\underset{m\times 1}{\boldsymbol{f}}(\underset{n\times 1}{\boldsymbol{x}})$ (reading "$\boldsymbol{f}$ being an $m \times 1$ vector function of an $n \times 1$ vector $\boldsymbol{x}$"), $\mathrm{D}\boldsymbol{f}(\boldsymbol{x})$ is an $m \times n$ matrix (due to Definition A.2.1) whereas $\mathrm{d}\boldsymbol{f}(\boldsymbol{x})$ remains an $m \times 1$ vector (same as $\boldsymbol{f}$). The advantage is even larger for matrices: for $\underset{m\times p}{\boldsymbol{F}}(\underset{n\times q}{\boldsymbol{X}})$, $\mathrm{d}\boldsymbol{F}(\boldsymbol{X})$ has the same dimension as $\boldsymbol{F}$ irrespective of the dimension of $\boldsymbol{X}$, but $\mathrm{D}\boldsymbol{F}(\boldsymbol{X})$ is going to be a horrendous $mp \times nq$ matrix.

A.3 Layout convention

Under the *numerator layout*, when we differentiate a scalar function φ with respect to a column vector $\underset{n\times 1}{\boldsymbol{x}}$, we get a *row* vector of dimension $1 \times n$. If we want our result to be in the column form, we must differentiate φ with respect to a row vector to start with. This is why the denominator in Equation (A.1) contains a transpose.

A.4 Application in OLS

A.4.1 Background

Imagine we are interested in learning the return on education. We might propose a rather simple model

$$\texttt{inc} = \beta_0 + \beta_1\texttt{edu} + \beta_2\texttt{exp} + \varepsilon \tag{A.7}$$

where inc is one's income, edu and exp denote years of formal education and years spent in the labour market, respectively.
We managed to collect survey data from n respondents and organised this information in the following system of equations:

$$\begin{cases} \texttt{inc}_1 = \beta_0 + \beta_1\texttt{edu}_1 + \beta_2\texttt{exp}_1 + \varepsilon_1 \\ \texttt{inc}_2 = \beta_0 + \beta_1\texttt{edu}_2 + \beta_2\texttt{exp}_2 + \varepsilon_2 \\ \quad\cdots \\ \texttt{inc}_n = \beta_0 + \beta_1\texttt{edu}_n + \beta_2\texttt{exp}_n + \varepsilon_n \end{cases} \tag{A.8}$$

This system of linear equations can be represented in the matrix notation using

$$\underset{n\times 1}{\boldsymbol{y}} = \begin{pmatrix} \texttt{inc}_1 \\ \texttt{inc}_2 \\ \cdots \\ \texttt{inc}_n \end{pmatrix}, \quad \underset{n\times 3}{\boldsymbol{X}} = \begin{pmatrix} 1 & \texttt{edu}_1 & \texttt{exp}_1 \\ 1 & \texttt{edu}_2 & \texttt{exp}_2 \\ \cdots & & \\ 1 & \texttt{edu}_n & \texttt{exp}_n \end{pmatrix}, \quad \underset{3\times 1}{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \text{ and } \underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix} \tag{A.9}$$

as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{A.10}$$

A.4.2 Ordinary least squares

The objective of OLS is to minimise the *sum of squared* error terms. A handy way of representing sum of squared ε is

$$\text{SSE} = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_n^2 = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix} = \boldsymbol{\varepsilon}^\intercal \boldsymbol{\varepsilon}. \tag{A.11}$$

In fact, $\boldsymbol{x}^\intercal \boldsymbol{x}$ is the mathematical translation of "sum of squared" of $\boldsymbol{x}$.
Now we are ready to continue. We want to carefully choose a combination of β_0, β_1 and β_2 in order to make SSE as small as possible, ie

$$\min_{\boldsymbol{\beta}} \{\boldsymbol{\varepsilon}^\intercal \boldsymbol{\varepsilon}\} = \min_{\boldsymbol{\beta}} \{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\intercal (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\} \tag{A.12}$$

(the equal sign is due to Equation (A.10)).
Two observations can be made from the minimisation problem in Equation (A.12):

1. both $\boldsymbol{y}$ and $\boldsymbol{X}$ are collected data therefore can no longer be changed by the researcher; but we are free to adjust $\boldsymbol{\beta}$ in whatever way we want, meaning $\boldsymbol{\beta}$ is the "independent variable" and SSE is a function of $\boldsymbol{\beta}$, and

2. $\boldsymbol{\varepsilon}^\intercal \boldsymbol{\varepsilon}$ is a scalar function (please verify).

Then,

$$\begin{aligned} \varphi(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}^\intercal \boldsymbol{\varepsilon} &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\intercal (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \left(\boldsymbol{y}^\intercal - \boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal\right) (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \boldsymbol{y}^\intercal \boldsymbol{y} - \boldsymbol{y}^\intercal \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{y} + \boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X}\boldsymbol{\beta} \end{aligned} \tag{A.13}$$

We now differentiate $\varphi(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$:

$$\begin{aligned} \frac{\mathrm{d}\varphi(\boldsymbol{\beta})}{\mathrm{d}\boldsymbol{\beta}} &= -\boldsymbol{y}^\intercal \boldsymbol{X} - \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}} \left[\left(\boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{y}\right)^\intercal\right] + \boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X} + \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}} \left[\left(\boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X}\boldsymbol{\beta}\right)^\intercal\right] \\ &= -\boldsymbol{y}^\intercal \boldsymbol{X} - \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}} \left[\boldsymbol{y}^\intercal \boldsymbol{X}\boldsymbol{\beta}\right] + \boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X} + \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}} \left[\boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X}\boldsymbol{\beta}\right] \\ &= -\boldsymbol{y}^\intercal \boldsymbol{X} - \boldsymbol{y}^\intercal \boldsymbol{X} + \boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X} + \boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X} \\ &= -2\boldsymbol{y}^\intercal \boldsymbol{X} + 2\boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X} \end{aligned} \tag{A.14}$$

(We were able to liberally apply transpose to terms containing $\boldsymbol{\beta}^\intercal$ and not to others because φ is a scalar function and each term in it must also be 1×1 in dimension, whose transpose must be equal to itself.)

Apply first order condition to Equation (A.14). An optimal $\widehat{\boldsymbol{\beta}}$ must satisfy

$$
\begin{aligned}
-2\boldsymbol{y}^{\mathsf{T}}\boldsymbol{X} + 2\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X} &= \boldsymbol{O} \\
2\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X} &= 2\boldsymbol{y}^{\mathsf{T}}\boldsymbol{X} \\
\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X} &= \boldsymbol{y}^{\mathsf{T}}\boldsymbol{X} \\
\left(\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}\right)^{\mathsf{T}} &= \left(\boldsymbol{y}^{\mathsf{T}}\boldsymbol{X}\right)^{\mathsf{T}} \\
\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}\widehat{\boldsymbol{\beta}} &= \boldsymbol{X}^{\mathsf{T}}\boldsymbol{y} \\
\widehat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}
\end{aligned}
\tag{A.15}
$$

Notice that another transpose was applied to Line 4 of Equation (A.15) in order to correct $\widehat{\boldsymbol{\beta}}^{\mathsf{T}}$ (due to Section A.3) back to its column form $\widehat{\boldsymbol{\beta}}$. In fact, it would be better to do $\frac{\mathrm{d}\varphi(\boldsymbol{\beta})}{\mathrm{d}\boldsymbol{\beta}^{\mathsf{T}}}$ in Equation (A.14) to avoid this later flipping. But the downside of this approach is a pedagogical one: most students would find differentiating with respect to $\boldsymbol{\beta}^{\mathsf{T}}$ out of blue while with respect to $\boldsymbol{\beta}$ is much more natural. In further derivations, $\frac{\mathrm{d}\varphi(\boldsymbol{\beta})}{\mathrm{d}\boldsymbol{\beta}^{\mathsf{T}}}$ will be used.

Derivative of quadratic forms

The derivative of a quadratic form $q(\boldsymbol{x}) = \boldsymbol{x}^\intercal \boldsymbol{A}\boldsymbol{x}$ is
$$\frac{\mathrm{d}q}{\mathrm{d}\boldsymbol{x}} = \boldsymbol{x}^\intercal \left(\boldsymbol{A} + \boldsymbol{A}^\intercal\right),$$
which can be further simplified to $\mathrm{d}q/\mathrm{d}\boldsymbol{x} = 2\boldsymbol{x}^\intercal \boldsymbol{A}$, if $\boldsymbol{A}$ is symmetric.

Name the expression in the facebook post φ, which is a function of $\boldsymbol{\beta}$:

$$\varphi(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\intercal \boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^\intercal\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\intercal\boldsymbol{\Omega}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right).$$

Typo

There is a typo in the original post: all $\boldsymbol{\Omega}$ should be in the inverse form $\boldsymbol{\Omega}^{-1}$, including the one sandwiched between $\boldsymbol{Z}^\intercal$ and $\boldsymbol{Z}$.

Scalar function

Note that φ is a scalar function:
$$\underset{1\times 1}{\varphi}\left(\underset{k\times 1}{\boldsymbol{\beta}}\right) = \frac{1}{\sigma^2}\left(\underset{n\times 1}{\boldsymbol{y}} - \underset{n\times k}{\boldsymbol{X}}\underset{k\times 1}{\boldsymbol{\beta}}\right)^\intercal \underset{n\times n}{\boldsymbol{\Omega}^{-1}}\underset{n\times k}{\boldsymbol{Z}}\left(\underset{k\times n}{\boldsymbol{Z}^\intercal}\underset{n\times n}{\boldsymbol{\Omega}^{-1}}\underset{n\times k}{\boldsymbol{Z}}\right)^{-1}\underset{k\times n}{\boldsymbol{Z}^\intercal}\underset{n\times n}{\boldsymbol{\Omega}^{-1}}\left(\underset{n\times 1}{\boldsymbol{y}} - \underset{n\times k}{\boldsymbol{X}}\underset{k\times 1}{\boldsymbol{\beta}}\right).$$

When differentiating a scalar function φ with respect to a column vector $\boldsymbol{\beta}$, the result is a *row* vector. If this is undesirable, differentiate the scalar function φ with respect to the *transpose* of that vector $\boldsymbol{\beta}^\intercal$.

I want to know
$$\frac{\mathrm{d}\varphi}{\mathrm{d}\boldsymbol{\beta}} = \frac{\mathrm{d}\varphi}{\mathrm{d}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)}\frac{\mathrm{d}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)}{\mathrm{d}\boldsymbol{\beta}} = \frac{\mathrm{d}\varphi}{\mathrm{d}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)}\left(-\boldsymbol{X}\right),$$
so I first calculate (using the result from quadratic form derivatives stated at the beginning)

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)} = \frac{2}{\sigma^2}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\intercal\left[\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^\intercal\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\intercal\boldsymbol{\Omega}^{-1}\right].$$

Therefore,
$$\frac{\mathrm{d}\varphi}{\mathrm{d}\boldsymbol{\beta}} = -\frac{2}{\sigma^2}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\intercal\left[\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^\intercal\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\intercal\boldsymbol{\Omega}^{-1}\right]\boldsymbol{X}$$
$$\frac{\mathrm{d}\varphi}{\mathrm{d}\boldsymbol{\beta}^\intercal} = -\frac{2}{\sigma^2}\boldsymbol{X}^\intercal\left[\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^\intercal\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\intercal\boldsymbol{\Omega}^{-1}\right]\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)$$
(The second line is to avoid working with row vectors.)

Second derivative

The second derivative of φ is
$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}\boldsymbol{\beta}^\intercal\mathrm{d}\boldsymbol{\beta}} = \frac{2}{\sigma^2}\boldsymbol{X}^\intercal\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^\intercal\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\intercal\boldsymbol{\Omega}^{-1}\boldsymbol{X},$$
which is a positive definite $k\times k$ matrix (another quadratic form). This implies that the result from the first order condition below is a minimum.

Impose the first order condition:

$$\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\text{GLS-IV}}\right)=\boldsymbol{0}$$

$$\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{y}=\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\text{GLS-IV}}$$

Therefore:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\text{GLS-IV}}&=\left(\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{y}\\
&=\left(\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{X}}\right)^{-1}\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{y}}\\
&\neq\left(\widehat{\boldsymbol{X}}^{\intercal}\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{X}}\right)^{-1}\widehat{\boldsymbol{X}}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{y},\ \text{satisfying the claim in the facebook post.}
\end{aligned}$$

However, since $\widehat{\boldsymbol{X}}$ is the GLS-IV-estimator of $\boldsymbol{X}$ onto the $\boldsymbol{Z}$-space:

$$\begin{aligned}
\widehat{\boldsymbol{X}}&=\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{X},\ \text{and}\\
\widehat{\boldsymbol{X}}^{\intercal}&=\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}.
\end{aligned}$$

The last expression in the facebook post then becomes:

$$\begin{aligned}
&\left(\widehat{\boldsymbol{X}}^{\intercal}\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{X}}\right)^{-1}\widehat{\boldsymbol{X}}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{y}\\
=&\left(\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{y}\\
=&\left(\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{y}\\
=&\widehat{\boldsymbol{\beta}}_{\text{GLS-IV}}
\end{aligned}$$

After all, “one might have guessed” correctly!