Economics 468 – Midterm Exam

1. Consider the following linear regression model:

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}, \tag{1}$$

where the dependent variable $\boldsymbol{y}$ is represented by an $n \times 1$ vector, and the independent explanatory variables are grouped into two matrices, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, of dimensions $n \times k_1$ and $n \times k_2$ respectively, with $k_1 + k_2 = k$.

Write down the form of the FWL (Frisch-Waugh-Lovell) regression corresponding to (1) in which only the parameters $\boldsymbol{\beta}_2$ appear. Write down the FWL regression in which only the parameters $\boldsymbol{\beta}_1$ appear. Give explicit algebraic expressions for the OLS (ordinary least squares) estimates $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ obtained by running regression (1).

The FWL regression with only $\boldsymbol{\beta}_2$ is

$$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{M}_1\boldsymbol{u},$$

and that with only $\boldsymbol{\beta}_1$ is

$$\boldsymbol{M}_2\boldsymbol{y} = \boldsymbol{M}_2\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{M}_2\boldsymbol{u}.$$

Here $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ project on to $\mathcal{S}^\perp(\boldsymbol{X}_1)$ and $\mathcal{S}^\perp(\boldsymbol{X}_2)$ respectively.

The explicit expressions, derived directly from the FWL regressions, are:

$$\hat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1^\top\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top\boldsymbol{M}_2\boldsymbol{y} \text{ and } \hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y}. \tag{S1}$$

State the FWL theorem as it applies to regression (1). Prove algebraically or geometrically that the estimate $\hat{\boldsymbol{\beta}}_2$ from (1) is the same as that from the FWL regression from which $\boldsymbol{\beta}_1$ has been eliminated.

The FWL theorem states two things about regression (1) and one or other of the FWL regressions: First, the OLS estimates of $\boldsymbol{\beta}_2$ are the same from either (1) or the FWL regression from which $\boldsymbol{\beta}_1$ is eliminated. Second, the residuals from these two regressions are the same.

OLS estimation of (1) gives

$$\boldsymbol{y} = \boldsymbol{P_X}\boldsymbol{y} + \boldsymbol{M_X}\boldsymbol{y} = \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2 + \boldsymbol{M_X}\boldsymbol{y}.$$

Premultiplying by $\boldsymbol{X}_2^\top\boldsymbol{M}_1$ gives

$$\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2,$$

since $\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{M_X}\boldsymbol{y} = \boldsymbol{X}_2^\top\boldsymbol{M_X}\boldsymbol{y} = \boldsymbol{0}$. Premultiplying by $(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}$ gives

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y},$$

which is what is given by the FWL regression, by (S1). (Many other proofs are possible.)

Suppose now that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ contain mutually orthogonal sets of regressors, that is, that $\boldsymbol{X}_1^\top\boldsymbol{X}_2 = \boldsymbol{0}$. Which of the following regressions will give the same estimate of $\boldsymbol{\beta}_2$ as (1)? Why?

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u};\\
\boldsymbol{P}_1\boldsymbol{y} &= \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u};\\
\boldsymbol{P}_1\boldsymbol{y} &= \boldsymbol{P}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u};\\
\boldsymbol{P}_X\boldsymbol{y} &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u};\\
\boldsymbol{P}_X\boldsymbol{y} &= \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u};\\
\boldsymbol{M}_1\boldsymbol{y} &= \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u};\\
\boldsymbol{M}_1\boldsymbol{y} &= \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u};
\end{aligned}$$

where, as usual, $\boldsymbol{P}_1$ is the orthogonal projection on to $\mathcal{S}(\boldsymbol{X}_1)$, $\boldsymbol{P_X}$ is the orthogonal projection onto $\mathcal{S}(\boldsymbol{X}_1, \boldsymbol{X}_2)$, and $\boldsymbol{M}_1 = \mathbf{I} - \boldsymbol{P}_1$.

The estimate of $\boldsymbol{\beta}_2$ from (1) is

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{y} = (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{y},$$

since the orthogonality gives $\boldsymbol{M}_1 \boldsymbol{X}_2 = \boldsymbol{X}_2$. This is just the estimate from the first equation above, which therefore gives the same estimate as (1).

The second gives

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{P}_1 \boldsymbol{y} = \boldsymbol{0},$$

since $\boldsymbol{P}_1 \boldsymbol{X}_2 = \boldsymbol{0}$ by orthogonality. Thus the estimates are different.

The third cannot even be run, since the regressors are $\boldsymbol{P}_1 \boldsymbol{X}_2 = \boldsymbol{0}$. Obviously different.

The fourth must be transformed into an FWL regression by premultiplication by $\boldsymbol{M}_1$. This gives

$$\boldsymbol{M}_1 \boldsymbol{P}_{\boldsymbol{X}} \boldsymbol{y} = \boldsymbol{M}_1 \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{M}_1 \boldsymbol{u} = \boldsymbol{X}_2 \boldsymbol{\beta} + \boldsymbol{M}_1 \boldsymbol{u},$$

and estimates

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{P}_{\boldsymbol{X}} \boldsymbol{y}.$$

Now $\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}_2^\top \boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}_2^\top$, and so the estimates are the same.

The fifth gives

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{P}_{\boldsymbol{X}} \boldsymbol{y}.$$

Since, as above, $\boldsymbol{X}_2^\top \boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}_2^\top$, the estimates are the same.

For the sixth,

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{y}.$$

Since $\boldsymbol{X}_2^\top \boldsymbol{M}_1 = \boldsymbol{X}_2^\top$, the estimates are the same.

Lastly, the seventh is the FWL regression from (1), and so always gives the same estimates.

> **2.** Suppose that the $n \times k$ matrix $\boldsymbol{X}$ of explanatory variables is partitioned into two submatrices, as follows: $\boldsymbol{X} = [\boldsymbol{X}_1 \;\; \boldsymbol{X}_2]$, where $\boldsymbol{X}_1$ has k_1 columns and $\boldsymbol{X}_2$ has k_2 columns. If the orthogonal projections on to the spans of the columns of $\boldsymbol{X}$ and $\boldsymbol{X}_1$ are denoted by $\boldsymbol{P}_{\boldsymbol{X}}$ and $\boldsymbol{P}_1$ respectively, show that $\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1$ is an orthogonal projection matrix.

The key to proving this is that $\boldsymbol{P}_{\boldsymbol{X}} \boldsymbol{P}_1 = \boldsymbol{P}_1 \boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_1$. Students may choose to prove this, although they have not been explicitly asked to do so. The proof is:

$$\boldsymbol{P}_{\boldsymbol{X}} \boldsymbol{P}_1 = \boldsymbol{P}_{\boldsymbol{X}} \boldsymbol{X}_1 (\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^\top = \boldsymbol{X}_1 (\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^\top = \boldsymbol{P}_1, \qquad \text{(S2)}$$

since $\boldsymbol{P}_{\boldsymbol{X}} \boldsymbol{X}_1 = \boldsymbol{X}_1$. The other equality is just the transpose of this one. If this is done correctly, bonus marks are given.

A square matrix is an orthogonal projection if and only if it is symmetric and idempotent. Both $\boldsymbol{P}_{\boldsymbol{X}}$ and $\boldsymbol{P}_1$ are orthogonal projection matrices, and so are symmetric. So therefore is $\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1$.

The square of the matrix is, using (S2) and the idempotency of $\boldsymbol{P}_{\boldsymbol{X}}$ and $\boldsymbol{P}_1$,

$$(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1)(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1) = \boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1 - \boldsymbol{P}_1 + \boldsymbol{P}_1 = \boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1.$$

Thus $\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1$ is an orthogonal projection matrix.

> Show that the trace (that is, the sum of the diagonal elements) of $\boldsymbol{P}_{\boldsymbol{X}}$ is equal to k. What is the trace of $\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1$?

By the property that $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA})$, we have that

$$\operatorname{tr}(\boldsymbol{P}_{\boldsymbol{X}}) = \operatorname{tr}\big(\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top\big) = \operatorname{tr}\big(\boldsymbol{X}^\top \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\big) = \operatorname{tr}(\mathbf{I}).$$

Here the identity matrix is of dimension k, and so the trace equals k.

By the same argument, $\operatorname{tr}(\boldsymbol{P}_1) = k_1$, and so, since the trace is a linear operation, $\operatorname{tr}(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_1) = k - k_1 = k_2$.

Show that any $n \times 1$ vector $\boldsymbol{z}$ of the form $\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma}$, for an arbitrary $k_2 \times 1$ vector $\boldsymbol{\gamma}$, is left unchanged when premultiplied by $\boldsymbol{P_X} - \boldsymbol{P}_1$; that is, show that $(\boldsymbol{P_X} - \boldsymbol{P}_1)\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma}$.

We have

$$\begin{aligned}(\boldsymbol{P_X} - \boldsymbol{P}_1)\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma} &= \boldsymbol{P_X}\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma} \quad \text{since } \boldsymbol{P}_1\boldsymbol{M}_1 = \mathbf{0} \\ &= \boldsymbol{P_X}(\mathbf{I} - \boldsymbol{P}_1)\boldsymbol{X}_2\boldsymbol{\gamma} \\ &= \boldsymbol{P_X}\boldsymbol{X}_2\boldsymbol{\gamma} - \boldsymbol{P}_1\boldsymbol{X}_2\boldsymbol{\gamma} \quad \text{since } \boldsymbol{P_X}\boldsymbol{P}_1 = \boldsymbol{P}_1 \\ &= \boldsymbol{X}_2\boldsymbol{\gamma} - \boldsymbol{P}_1\boldsymbol{X}_2\boldsymbol{\gamma} \quad \text{since } \boldsymbol{P_X}\boldsymbol{X}_2 = \boldsymbol{X}_2 \\ &= (\mathbf{I} - \boldsymbol{P}_1)\boldsymbol{X}_2\boldsymbol{\gamma} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\gamma},\end{aligned}$$

as required.

Why do the above results prove that $\boldsymbol{P_X} - \boldsymbol{P}_1 = \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$, where this last matrix denotes the orthogonal projection on to $\mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$?

The dimension of the range of the orthogonal projection matrix $\boldsymbol{P_X} - \boldsymbol{P}_1$ equals its trace, which, as we have seen, is k_2. The matrix $\boldsymbol{M}_1\boldsymbol{X}_2$ has k_2 columns, all linearly independent if those of $\boldsymbol{X}$ are. Thus $\mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$, which we have shown to lie in the range of $\boldsymbol{P_X} - \boldsymbol{P}_1$, is of dimension k_2, and thus constitutes the entire range of $\boldsymbol{P_X} - \boldsymbol{P}_1$. This proves the result.

Suppose now that one of the columns of $\boldsymbol{X}$ is $\boldsymbol{e}_t$, the dummy variable for observation t. Give the algebraic expression of the $n \times 1$ vector $\boldsymbol{e}_t$. Show that, if a dependent variable $\boldsymbol{y}$ is regressed on $\boldsymbol{X}$, the fitted value for observation t is equal to y_t.

The algebraic expression is

$$\boldsymbol{e}_t = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where the 1 is the component in position t.

The residual for observation t is the t^{th} component of $\boldsymbol{M_X y}$, that is, $\boldsymbol{e}_t^\top \boldsymbol{M_X y}$. Since $\boldsymbol{e}_t$ is one of the regressors, $\boldsymbol{M_X e}_t = \mathbf{0}$, and so the residual is 0. Thus, since the fitted value plus the residual equals y_t, the fitted value is y_t.

3. The following printout gives the results of estimating the following autoregressive distributed lag model of the consumption function for Canada, using quarterly data for 1953:1 until 1996:4:

$$c_t = \alpha + \beta c_{t-1} + \gamma_0 y_t + \gamma_1 y_{t-1} + u_t. \tag{2}$$

Here c_t is the logarithm of consumption in period t, y_t the logarithm of personal disposable income, and c_{t-1} and y_{t-1} the first lags of these variables.

(printout omitted)

Such models are often expressed in first-difference form, that is, as

$$\Delta c_t = \delta + \theta \Delta y_t + \phi c_{t-1} + \psi y_{t-1} + u_t, \tag{3}$$

where the first-difference operator Δ is defined so that $\Delta c_t = c_t - c_{t-1}$. Obtain the algebraic relation between the parameters α, β, γ_1, and γ_2 of (2), and the parameters δ, θ, ϕ, and ψ of (3). Give the numerical values of the OLS estimates of δ, θ, ϕ, and ψ.

Subtracting c_{t-1} from (2) gives

$$\begin{aligned} c_t - c_{t-1} \equiv \Delta c_t &= \alpha + (\beta - 1)c_{t-1} + \gamma_0 y_t + \gamma_1 y_{t-1} + u_t \\ &= \alpha + \gamma_0(y_t - y_{t-1}) + (\beta - 1)c_{t-1} + (\gamma_0 + \gamma_1)y_{t-1} + u_t. \end{aligned}$$

Thus the algebraic relation is

$$\delta = \alpha, \quad \theta = \gamma_0, \quad \phi = \beta - 1, \quad \psi = \gamma_0 + \gamma_1.$$

Using the information in the printout, we see that

$$\begin{aligned} \hat{\delta} = \hat{\alpha} = 0.063936, \quad \hat{\theta} = \hat{\gamma}_0 = 0.290988, \\ \hat{\phi} = \hat{\beta} - 1 = 0.969225 - 1 = -0.030775, \\ \hat{\psi} = \hat{\gamma}_0 + \hat{\gamma}_1 = 0.290988 - 0.265151 = 0.025837. \end{aligned}$$

> How would you compute the standard errors for δ, θ, ϕ, and ψ? In particular, what information would you need over and above that provided in the printout in order to do so? (This information would ordinarily be supplied by a regression package.)

Clearly, the standard errors of $\hat{\delta}$, $\hat{\theta}$, and $\hat{\phi}$ are exactly the same as those of $\hat{\alpha}$, $\hat{\gamma}_0$, and $\hat{\beta}$ respectively, and are thus equal to 0.021660, 0.055112, and 0.022310 respectively.

The only problem is for $\hat{\psi} = \hat{\gamma}_0 + \hat{\gamma}_1$. Since

$$\text{Var}(\hat{\psi}) = \text{Var}(\hat{\gamma}_0) + \text{Var}(\hat{\gamma}_1) + 2\,\text{cov}(\hat{\gamma}_0, \hat{\gamma}_1),$$

we would need to know the value of $\text{cov}(\hat{\gamma}_), \hat{\gamma}_1)$ in order to compute this. The estimated covariance matrix, $\hat{\sigma}^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$, would supply this information.

> **4.** State the Gauss-Markov theorem, paying special attention to how to formulate the notion of efficiency used in the theorem, and to the conditions under which the theorem is true.

The Gauss-Markov theorem states that, for the linear regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u},$$

where the explanatory variables $\boldsymbol{X}$ are exogenous, that is, independent of the disturbances $\boldsymbol{u}$, and where the disturbances are of expectation 0 and have covariance matrix $\text{Var}(\boldsymbol{u}) = \text{E}(\boldsymbol{uu}^\top)$ equal to $\sigma^2\mathbf{I}$ for some scalar variance σ^2, then the ordinary least squares estimator is "best" in the class of linear unbiased estimators of the parameter vector $\boldsymbol{\beta}$. By "best", it is meant that the difference between the covariance matrix of an arbitrary estimator in this class and the covariance of the OLS estimator is a positive semidefinite matrix.

> Show that all estimators of the class covered by the Gauss-Markov theorem are solutions to estimating equations of the form
>
> $$\boldsymbol{W}^\top(\boldsymbol{y} - \boldsymbol{X\beta}) = \mathbf{0}. \tag{4}$$
>
> What are the requirements on the matrix $\boldsymbol{W}$ for the estimator defined by (4) to exist and to belong to the Gauss-Markov class of estimators?

A linear estimator is by definition an estimator of the form $\boldsymbol{Ay}$ for some exogenous $k \times n$ matrix $\boldsymbol{A}$ (n is the sample size, k is the number of explanatory variables, the dimension of $\boldsymbol{\beta}$). For this estimator to be unbiased, we know that $\boldsymbol{AX} = \mathbf{I}$, the $k \times k$ identity matrix. If we set $\boldsymbol{W} = \boldsymbol{A}^\top$, then we have

$$\boldsymbol{Ay} = \boldsymbol{W}^\top\boldsymbol{y} = (\boldsymbol{W}^\top\boldsymbol{X})^{-1}\boldsymbol{W}^\top\boldsymbol{y} \tag{S3}$$

because

$$\boldsymbol{W}^\top\boldsymbol{X} = \boldsymbol{AX} = \mathbf{I}.$$

The estimator (S3) is the solution of the estimating equations

$$\boldsymbol{W}^\top(\boldsymbol{y} - \boldsymbol{X\beta}) = \mathbf{0},$$

as required.

From the above, it is clear that $\boldsymbol{W}$ must be, first, such that $\boldsymbol{W}^\top\boldsymbol{X}$ is nonsingular, in order for the inverse matrix to exist, and, second, exogenous, in order to comply with that requirement imposed on the matrix $\boldsymbol{A}$.

"Efficiency" for the Gauss-Markov theorem is defined by a property of covariance matrices. What is this property? Explain clearly why this way of defining efficiency is appropriate for estimators of the parameters of a linear regression model.

The efficiency property is that stated above, namely that the difference between the covariance matrix of an arbitrary linear unbiased estimator and that of the OLS estimator is a positive semidefinite matrix. This is a sensible way to define efficiency, because, if we consider any linear combination of the elements of the estimated parameter vector, say $\boldsymbol{a}^\top\hat{\boldsymbol{\beta}}$, with exogenous $\boldsymbol{a}$, then the variance of this linear combination is $\boldsymbol{a}^\top\boldsymbol{V}\boldsymbol{a}$, where $\boldsymbol{V}$ is the covariance matrix of $\hat{\boldsymbol{\beta}}$. If we compare this variance for $\boldsymbol{V}_1$, the covariance matrix of an arbitrary linear unbiased estimator, and $\boldsymbol{V}_0$, the OLS covariance matrix, we have for the difference between the two variances

$$\boldsymbol{a}^\top\boldsymbol{V}_1\boldsymbol{a} - \boldsymbol{a}^\top\boldsymbol{V}_0\boldsymbol{a} = \boldsymbol{a}^\top(\boldsymbol{V}_1 - \boldsymbol{V}_0)\boldsymbol{a}. \tag{S4}$$

Now the efficiency criterion says that the difference $\boldsymbol{V}_1 - \boldsymbol{V}_0$ is positive semidefinite, which means that, for any $\boldsymbol{a}$, the quadratic form (S4) is nonnegative. Therefore the variance of the linear combination as estimated by OLS can never be greater than that estimated with another linear unbiased estimator.

If $\boldsymbol{A}$ is a positive definite matrix, show that $\boldsymbol{A}^{-1}$ is also positive definite. Does the same result hold if $\boldsymbol{A}$ is positive semi-definite without being positive definite? Why or why not?

Consider the quadratic form $\boldsymbol{a}^\top\boldsymbol{A}^{-1}\boldsymbol{a}$. If $\boldsymbol{A}^{-1}$ is positive definite, then this quadratic form is positive for all nonzero vectors $\boldsymbol{a}$. Now

$$\boldsymbol{a}^\top\boldsymbol{A}^{-1}\boldsymbol{a} = \boldsymbol{a}^\top\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{a} = (\boldsymbol{a}^\top\boldsymbol{A}^{-1})\boldsymbol{A}(\boldsymbol{A}^{-1}\boldsymbol{a}). \tag{S5}$$

Since $\boldsymbol{a} \neq \boldsymbol{0}$, and since $\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$ are nonsingular (or else the inverse would not exist!), it follows that $\boldsymbol{A}^{-1}\boldsymbol{a} \neq \boldsymbol{0}$. But it then follows that the rightmost expression in (S5) is positive, because $\boldsymbol{A}$ is positive definite. This proves that $\boldsymbol{A}^{-1}$ is also positive definite.

A matrix $\boldsymbol{A}$ that is positive semidefinite without being positive definite is singular, because there exists a nonzero vector $\boldsymbol{a}$ such that $\boldsymbol{a}^\top\boldsymbol{A}\boldsymbol{a} = 0$, whence $\boldsymbol{A}\boldsymbol{a} = \boldsymbol{0}$. A nonsingular matrix is not invertible, so that $\boldsymbol{A}^{-1}$ does not exist. Thus the result does not hold for a matrix that is positive semidefinite without being positive definite.