Professor Yanai and Multivariate Analysis

Yoshio Takane
University of Victoria/McGill University

The 24th International Workshop on Matrices and Statistics (IWMS)
Haikou, China, May 2015

Professor Yanai in 1992 (Puntanen, Styan, and Isotalo, 2011, p. 307)

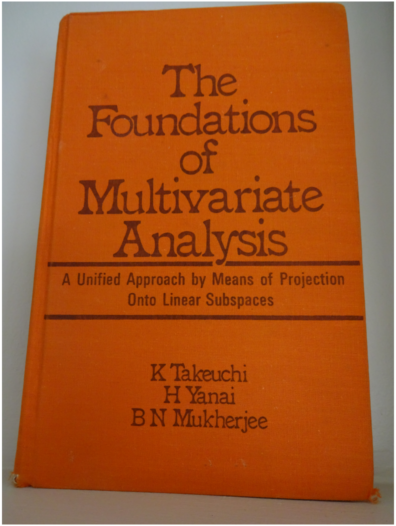
Projectors and SVD

Common threads running through them are:

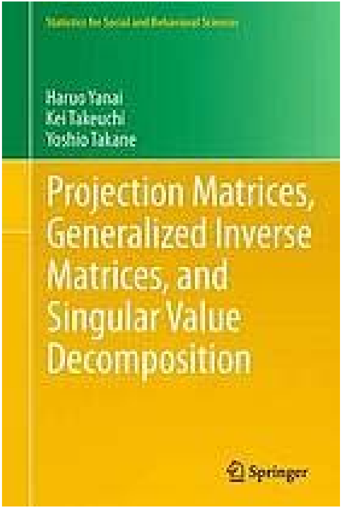
- **projectors**,
- **singular value decomposition (SVD)**,

which are main subject matters of Yanai, Takeuchi, and Takane (2011).

Takeuchi, Yanai, and Mukherjee (1982): The Foundations of Multivariate Analysis



Yanai, Takeuchi, and Takane (2011): Projection matrices, generalized inverse matrices, and singular value decomposition



Topics Covered

- (1) Constrained principal component analysis (CPCA)
- (2) Khatri's lemma
- (3) The Wedderburn-Guttman theorem
- (4) Ridge operators
- (5) Constrained canonical correlation analysis
- (6) Causal inferences

Orthogonal Projectors

- Sp($\mathbf{X}$): The space spanned by column vectors of $\mathbf{X}$.
- Ker($\mathbf{X}'$): The orthogonal complement subspace to Sp($\mathbf{X}$).
- Orthogonal projectors onto Sp($\mathbf{X}$): $\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$.
- Orthogonal projectors onto Ker($\mathbf{X}'$): $\mathbf{Q}_X = \mathbf{I} - \mathbf{P}_X$.
- Basic properties:

$$\mathbf{P}_X' = \mathbf{P}_X,\ \mathbf{Q}_X' = \mathbf{Q}_X \text{ (symmetric).}$$
$$\mathbf{P}_X^2 = \mathbf{P}_X,\ \mathbf{Q}_X^2 = \mathbf{Q}_X \text{ (idempotent).}$$
$$\mathbf{P}_X\mathbf{Q}_X = \mathbf{Q}_X\mathbf{P}_X = \mathbf{O} \text{ (orthogonal).}$$

K-Orthogonal Projectors

- Let $\mathbf{K}$ be an *nnd* matrix such that $\text{rank}(\mathbf{KX}) = \text{rank}(\mathbf{X})$.
- K-orthogonal projectors: $\mathbf{P}_{X/K} = \mathbf{X}(\mathbf{X}'\mathbf{KX})^{-}\mathbf{X}'\mathbf{K}$, and $\mathbf{Q}_{X/K} = \mathbf{I} - \mathbf{P}_{X/K}$.
- Basic properties:
  $(\mathbf{KP}_{X/K})' = \mathbf{KP}_{X/K}$, $(\mathbf{KQ}_{X/K})' = \mathbf{KQ}_{X/K}$ (K-symmetric).
  $\mathbf{P}^2_{X/K} = \mathbf{P}_{X/K}$, $\mathbf{Q}^2_{X/K} = \mathbf{Q}_{X/K}$ (idempotent).
  $\mathbf{P}'_{X/K}\mathbf{KQ}_{X/K} = \mathbf{Q}'_{X/K}\mathbf{KP}_{X/K} = \mathbf{O}$ (K-orthogonal).

CPCA: Two Phases

- External Analysis and Internal Analysis.
- External Analysis: Decomposes the main data matrix according to the external information abut the row and columns of the data matrix $\Longrightarrow$ projection.
- Internal Analysis: Further analyses of decomposed matrices into components $\Longrightarrow$ SVD (singular value decomposition

External Analysis

- **Y**: The main data matrix.
- **G**: The row (left-hand) side information matrix.
- **H**: The column (right-hand) side information matrix.
- The basic decomposition:

$$\mathbf{Y} = \mathbf{P}_G\mathbf{Y}\mathbf{P}_H + \mathbf{Q}_G\mathbf{Y}\mathbf{P}_H + \mathbf{P}_G\mathbf{Y}\mathbf{Q}_H + \mathbf{Q}_G\mathbf{Y}\mathbf{Q}_H.$$

- A similar decomposition with K-orthogonal projectors.

Finer Decompositions (1)

- $\mathbf{G} = [\mathbf{M}, \mathbf{N}]$.
- (1) $\mathbf{P}_G = \mathbf{P}_M + \mathbf{P}_N \Leftrightarrow \mathbf{M}'\mathbf{N} = \mathbf{O}$.
- (2) $\mathbf{P}_G = \mathbf{P}_M + \mathbf{P}_N - \mathbf{P}_M\mathbf{P}_N \Leftrightarrow \mathbf{P}_M\mathbf{P}_N = \mathbf{P}_N\mathbf{P}_M$.
- (3) $\mathbf{P}_G = \mathbf{P}_M + \mathbf{P}_{Q_M N} = \mathbf{P}_N + \mathbf{P}_{Q_N M}$.
- (4) $\mathbf{P}_G = \mathbf{P}_{M/Q_N} + \mathbf{P}_{N/Q_M} \Leftrightarrow \text{rank}(\mathbf{G}) = \text{rank}(\mathbf{M}) + \text{rank}(\mathbf{N})$.
- (5) $\mathbf{P}_G = \mathbf{P}_{GA} + \mathbf{P}_{G(G'G)^- B} \Leftrightarrow \mathbf{A}'\mathbf{B} = \mathbf{O}$,
  $\text{Sp}(\mathbf{A}) \oplus \text{Sp}(\mathbf{B}) = \text{Sp}(\mathbf{G}')$.
- Analogous decompositions for $\mathbf{P}_H$, $\mathbf{P}_{G/K}$, and $\mathbf{P}_{H/L}$.

Finer Decompositions (2): Explanations

- (1) **M** and **N** are mutually orthogonal.
- (2) **M** and **N** are mutually orthogonal, except their common space. (ANOVA w/o interactions).
- (3) Fit one first and the other to the residuals.
- (4) **M** and **N** are disjoint. Fit both simultaneously.
- (5) A matrix of regression coefficients **C** constrained by $\mathbf{C} = \mathbf{AC}^*$ or by $\mathbf{B}'\mathbf{C} = \mathbf{O}$.

Internal Analysis

- PCA of terms obtained by the external analysis of $\mathbf{Y}$, e.g., $\mathbf{P}_G\mathbf{Y}\mathbf{P}_H$, which amounts to SVD($\mathbf{P}_G\mathbf{Y}\mathbf{P}_H$).

Khatri's Lemma (1)

- Constrained Correspondence Analysis (CCA).
- $\mathbf{U}$: The row representation matrix. (We consider only the row side.)
- Two ways of constraining $\mathbf{U}$: (1) $\mathbf{U} = \mathbf{A}\mathbf{U}^*$, and (2) $\mathbf{B}'\mathbf{U} = \mathbf{O}$.
- $\mathbf{P}_A = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}' = \mathbf{I} - \mathbf{B}(\mathbf{B}'\mathbf{B})^-\mathbf{B}' = \mathbf{Q}_B$.
- What happens if non-identity metric $\mathbf{K}$ is used?
- Let $\mathbf{A}$ $(p \times r)$ and $\mathbf{B}$ $(p \times (p - r))$ be matrices such that rank$(\mathbf{A}) = r$, rank$(\mathbf{B}) = p - r$, and $\mathbf{A}'\mathbf{B} = \mathbf{O}$. Then $\mathbf{I} = \mathbf{A}(\mathbf{A}'\mathbf{K}\mathbf{A})^{-1}\mathbf{A}\mathbf{K} + \mathbf{K}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{K}^{-1}\mathbf{B})^{-1}\mathbf{B}'$ (Khatri, 1966).

Further Remarks

- An alternative expression:
  $\mathbf{K} = \mathbf{KA}(\mathbf{A}'\mathbf{KA})^{-1}\mathbf{AK} + \mathbf{B}(\mathbf{B}'\mathbf{K}^{-1}\mathbf{B})^{-1}\mathbf{B}'$.
- Useful for rewriting Q-type projectors into P-type.

Some Extensions

- Let $\mathbf{A}$ $(p \times r)$ and $\mathbf{B}$ $(p \times (p - r))$ be matrices such that $\text{rank}(\mathbf{A}) = r$ and $\text{rank}(\mathbf{B}) = p - r$, and let $\mathbf{M}$ and $\mathbf{N}$ be *nnd* matrices such that
  (i) $\mathbf{A}'\mathbf{M}\mathbf{N}\mathbf{B} = \mathbf{O}$,
  (ii) $\text{rank}(\mathbf{M}\mathbf{A}) = \text{rank}(\mathbf{A})$,
  (iii) $\text{rank}(\mathbf{N}\mathbf{B}) = \text{rank}(\mathbf{B})$.
  Then,
  $$\mathbf{I} = \mathbf{A}(\mathbf{A}'\mathbf{M}\mathbf{A})^{-}\mathbf{A}'\mathbf{M} + \mathbf{N}\mathbf{B}(\mathbf{B}'\mathbf{N}\mathbf{B})^{-}\mathbf{B}'.$$
- Reduces to the original lemma when $\mathbf{M} = \mathbf{K}$ and $\mathbf{N} = \mathbf{K}^{-1}$.

The WG Theorem

- Let $\mathbf{Y}$ $(n \times p)$ be of rank r, and let $\mathbf{A}$ $(n \times s)$ and $\mathbf{B}$ $(p \times s)$ be such that $\mathbf{A}'\mathbf{Y}\mathbf{B}$ is invertible.
- Then,

$$\begin{aligned}\text{rank}(\mathbf{Y}_1) &= \text{rank}(\mathbf{Y}) - \text{rank}(\mathbf{Y}\mathbf{B}(\mathbf{A}'\mathbf{Y}\mathbf{B})^{-1}\mathbf{A}'\mathbf{Y}) \\ &= \text{rank}(\mathbf{Y}) - \text{rank}(\mathbf{A}'\mathbf{Y}\mathbf{B}) = r - s,\end{aligned}$$

  where

$$\mathbf{Y}_1 = \mathbf{Y} - \mathbf{Y}\mathbf{B}(\mathbf{A}'\mathbf{Y}\mathbf{B})^{-1}\mathbf{A}'\mathbf{Y}.$$

- Wedderburn (1934) for $s = 1$. Guttman (1944) for $s > 1$. Guttman (1957) reverse.

The Generalized WG Theorem

- When $\mathbf{A}'\mathbf{YB}$ is not invertible, can we replace it by a generalized inverse?
- Yes, but it requires a condition.
- A rank additivity (subtractivity) problem?

$$\text{rank}(\mathbf{Y} - \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}) = \text{rank}(\mathbf{Y}) - \text{rank}(\mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}). \quad (1)$$

- Does the following always hold?

$$\text{rank}(\mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}) = \text{rank}(\mathbf{A}'\mathbf{YB}) \quad (2)$$

- No. Tian and Styan (2009) showed the following always holds:

$$\text{rank}(\mathbf{Y} - \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}) = \text{rank}(\mathbf{Y}) - \text{rank}(\mathbf{A}'\mathbf{YB}). \quad (3)$$

- (2) requires a condition, as does (1).

The *ns* Condition

- Let $\mathbf{C} = \mathbf{B}(\mathbf{A}'\mathbf{Y}\mathbf{B})^{-}\mathbf{A}'$.
- The *ns* condition for (1)to hold is:

$$\mathbf{YCYCY} = \mathbf{YCY}.$$

- Equivalent conditions:
  $(\mathbf{YCYY}^{-})^2 = \mathbf{YCYY}^{-} \Leftrightarrow (\mathbf{Y}^{-}\mathbf{YCY})^2 = \mathbf{Y}^{-}\mathbf{YCY}$.
- $(\mathbf{YC})^2 = \mathbf{YC}$ or $(\mathbf{CY})^2 = \mathbf{CY}$ (sufficient but not necessary).
- $\mathbf{CYC} = \mathbf{C}$ (sufficient but not necessary). Even stronger than idempotency of $\mathbf{YC}$ or $\mathbf{CY}$.

The WG Decomposition

- $\mathbf{Y} = \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y} + (\mathbf{Y} - \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y})$.
- Let $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ be matrices such that
  (i) $\mathrm{Sp}(\tilde{\mathbf{A}}) \subset \mathrm{Sp}(\mathbf{Y})$,
  (ii) $\mathrm{Sp}(\tilde{\mathbf{B}}) \subset \mathrm{Sp}(\mathbf{Y}')$,
  (iii) $\mathrm{rank}(\mathbf{A}'\mathbf{YB}) + \mathrm{rank}(\tilde{\mathbf{B}}'\mathbf{Y}^{-}\tilde{\mathbf{A}}) = \mathrm{rank}(\mathbf{Y})$,
  (iv) $\mathbf{A}'\mathbf{YY}^{-}\tilde{\mathbf{A}} = \mathbf{A}'\tilde{\mathbf{A}} = \mathbf{O}$,
  (v) $\tilde{\mathbf{B}}'\mathbf{Y}^{-}\mathbf{YB} = \tilde{\mathbf{B}}'\mathbf{B} = \mathbf{O}$.
- Then, $\mathbf{Y} = \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y} + \tilde{\mathbf{A}}(\tilde{\mathbf{B}}'\mathbf{Y}^{-}\tilde{\mathbf{A}})^{-}\tilde{\mathbf{B}}'$.

Ridge Operator: Definition

- $\mathbf{R}_X(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda \mathbf{P}_{X'})^{-}\mathbf{X}'$, where $\mathbf{P}_{X'} = \mathbf{X}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{X}$ is the orthogonal projector onto $\text{Sp}(\mathbf{X}')$. ($\mathbf{P}_{X'} = \mathbf{I}$ if $\mathbf{X}$ is columnwise nonsingular.)
- The ridge LS estimation $\min_{\mathbf{c}} = \phi_\lambda(\mathbf{c})$, where $\phi_\lambda(\mathbf{c}) = \text{SS}(\mathbf{e}) + \lambda \text{SS}(\mathbf{c})_{P_{X'}}$ and $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{c}$. (We assume, w/o loss of generality, that $\text{Sp}(\mathbf{c}) \subset \text{Sp}(\mathbf{X}')$.)

Ridge Operator: Some Properties

- Let $\mathbf{S}_X(\lambda) = \mathbf{I} - \mathbf{R}_X(\lambda)$.
- $\mathbf{R}_X(\lambda)$ and $\mathbf{S}_X(\lambda)$ have properties similar to those of $\mathbf{P}_X$ and $\mathbf{Q}_X$.
- For example:
  $\mathbf{R}_X(\lambda)\mathbf{K}_X(\lambda)\mathbf{R}_X(\lambda) = \mathbf{R}_X(\lambda)$ (i.e., $\mathbf{K}_X(\lambda) = \mathbf{R}_X(\lambda)^+$.),
  $\mathbf{R}_X(\lambda) - \mathbf{R}_X(\lambda)^2 = \mathbf{R}_X(\lambda)\mathbf{S}_X(\lambda) = \mathbf{S}_X(\lambda)\mathbf{R}_X(\lambda) \geq \mathbf{O}$.
  $\mathbf{R}_X(\lambda)\mathbf{K}_X(\lambda) = \mathbf{P}_X$, etc.
- Similar decompositions of $\mathbf{R}_X(\lambda)$ to those of $\mathbf{P}_X$.

Ridge Metric Matrix

- Ridge metric matrix: $\mathbf{K}_X(\lambda) = \mathbf{P}_X + \lambda(\mathbf{XX}')^+$.
- Then, $\mathbf{R}_X(\lambda)$ can be rewritten as:

$$\mathbf{R}_X(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{K}_X(\lambda)\mathbf{X})^-\mathbf{X}'.$$

Generalized Ridge Operator

- Generalized ridge operator:
  $\mathbf{R}_X^{(W,L)}(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X} + \lambda\mathbf{L})^{-}\mathbf{X}'\mathbf{W}$, where $\mathbf{L}$ is an *nnd* matrix such that $\text{Sp}(\mathbf{L}) \subset \text{Sp}(\mathbf{X}')$, and $\mathbf{W}$ is an *nnd* matrix such that $\text{rank}(\mathbf{W}\mathbf{X}) = \text{rank}(\mathbf{X})$.
- Generalized ridge metric matrix:
  $\mathbf{K}_X^{(W,L)}(\lambda) = \mathbf{P}_X + \lambda\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{L}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}$.
- Then, $\mathbf{R}_X^{(W,L)}(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{K}_X^{(W,L)}(\lambda)\mathbf{X})^{-}\mathbf{X}'\mathbf{W}$.

Decompositions of Total Association

- Total association between $\mathbf{X}$ and $\mathbf{Y}$: $\text{tr}(\mathbf{P}_X\mathbf{P}_Y)$.
- $\mathbf{X} = \mathbf{M} + \mathbf{N}$, $\mathbf{M}'\mathbf{N} = \mathbf{O}$ does not guarantee $\mathbf{P}_X = \mathbf{P}_M + \mathbf{P}_N$.
- cf. $\mathbf{X} = [\mathbf{M}, \mathbf{N}]$, $\mathbf{M}'\mathbf{N} = \mathbf{O}$ leads to $\mathbf{P}_X = \mathbf{P}_M + \mathbf{P}_N$.
- We need orthogonal decompositions of $\mathbf{P}_X$ and $\mathbf{P}_Y$ to derive additive decompositions of the total association.

Two Orthogonal Decompositions of Projectors

- (1) Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{W}$ be matrices such that $\mathrm{Sp}(\mathbf{A}) = \mathrm{Ker}(\mathbf{H}'\mathbf{X}'\mathbf{P}_G\mathbf{X})$, $\mathrm{Sp}(\mathbf{B}) = \mathrm{Ker}(\mathbf{H}'\mathbf{X}'\mathbf{Q}_G\mathbf{X})$, and $\mathrm{Sp}(\mathbf{W}) = \mathrm{Ker}(\mathbf{X}'\mathbf{G})$. Then,

$$\mathbf{P}_{[X,G]} = \mathbf{P}_{P_G XH} + \mathbf{P}_{P_G XA} + \mathbf{P}_{Q_G XH} + \mathbf{P}_{Q_G XB} + \mathbf{P}_{GW}.$$

- (2) Let $\mathbf{K}$, $\mathbf{U}$, and $\mathbf{V}$ be matrices such that $\mathrm{Sp}(\mathbf{K}) = \mathrm{Ker}(\mathbf{H}'\mathbf{X}'\mathbf{X})$, $\mathrm{Sp}(\mathbf{U}) = \mathrm{Ker}(\mathbf{G}'\mathbf{X}\mathbf{H})$, and $\mathrm{Sp}(\mathbf{V}) = \mathrm{Ker}(\mathbf{G}'\mathbf{X}\mathbf{K})$. Then,

$$\mathbf{P}_{[X,G]} = \mathbf{P}_{P_{XH} G} + \mathbf{P}_{XHU} + \mathbf{P}_{P_{XK} G} + \mathbf{P}_{XKV} + \mathbf{P}_{Q_X G}.$$

Constrained Canonical Correlation Analysis

- Similar decompositions of $\mathbf{P}_{[\mathbf{Y},\mathbf{G}_Y]}$.
- Take one term each from a decomposition of $\mathbf{P}_{[\mathbf{X},\mathbf{G}_X]}$ and that of $\mathbf{P}_{[\mathbf{Y},\mathbf{G}_Y]}$, apply SVD to the product of the two, e.g.,

$$\mathrm{SVD}(\mathbf{P}_{\mathbf{Q}_{G_X} XH_X}\mathbf{P}_{YH_Y U_Y}).$$

Confounding Variables

- Causal inferences without randomization. How to eliminate the effects of confounding variables.
- $\mathbf{y}$: The dependent variable.
- $\mathbf{x}$: The independent variable.
- $\mathbf{U}$: The confounding variables.
- Regression analysis (1): $\mathbf{y} = \mathbf{x}a_1 + \mathbf{Uc} + \mathbf{e}_1$. The OLS estimate of $\mathbf{x}a_1$ is given by

$$\mathbf{x}\hat{a}_1 = \mathbf{P}_{x/Q_u}\mathbf{y} \tag{4}$$

.

- On the other hand, consider the regression of $\mathbf{x}$ onto $\mathbf{U}$, i.e., $\mathbf{x} = \mathbf{Ud} + \mathbf{e}_2$. The OLS estimate of $\mathbf{Ud}$ is given by

$$\mathbf{U}\hat{\mathbf{d}} = \mathbf{P}_U\mathbf{x}. \tag{5}$$

Linear Propensity Scores

- We call $\mathbf{P}_U\mathbf{x}$ linear propensity scores. Residuals from the above regression $\mathbf{Q}_U\mathbf{x}$ represent the portions of $\mathbf{x}$ left unaccounted for by $\mathbf{U}$.
- We next consider using $\mathbf{P}_U\mathbf{x}$ instead of $\mathbf{U}$ in the first regression, i.e., $\mathbf{y} = \mathbf{x}a_2 + \mathbf{P}_U\mathbf{x}b + \mathbf{e}_3$. the OLS estimate of $\mathbf{x}a_2$ is given by

$$\mathbf{x}\hat{a}_2 = \mathbf{P}_{x/Q_{P_U x}}\mathbf{y}, \tag{6}$$

  where $\mathbf{Q}_{P_U x} = \mathbf{I} - \mathbf{P}_U\mathbf{x}(\mathbf{x}'\mathbf{P}_U\mathbf{x})^{-1}\mathbf{x}'\mathbf{P}_U$.
- Since $\mathbf{Q}_{P_U x}\mathbf{x} = \mathbf{x} - \mathbf{P}_U\mathbf{x}(\mathbf{x}'\mathbf{P}_U\mathbf{x})^{-1}\mathbf{x}'\mathbf{P}_U\mathbf{x} = \mathbf{Q}_U\mathbf{x}$, we obtain

$$\mathbf{P}_{x/Q_{P_U x}}\mathbf{y} = \mathbf{P}_{x/Q_U}\mathbf{y}. \tag{7}$$

  This means (4) and (6) are equivalent.

Instrumental Variable (IV) Estimation

- Regression analysis: $\mathbf{y} = \mathbf{x}a_3 + \mathbf{e}_4$. The IV estimate of $\mathbf{x}a_3$ with $\mathbf{z} = \mathbf{Q}_U x$ as the IV is given by

$$\mathbf{x}\hat{a}_3 = \mathbf{P}_{x/P_z}\mathbf{y} = \mathbf{P}_{x/Q_U}\mathbf{y}. \tag{8}$$

- Since $\mathbf{P}_z = \mathbf{Q}_U\mathbf{x}(\mathbf{x}'\mathbf{Q}_U\mathbf{x})^{-1}\mathbf{x}'\mathbf{Q}_U$ and $\mathbf{x}'\mathbf{P}_z = \mathbf{x}'\mathbf{Q}_U$, this is identical to (4) and (6).

- It can also be easily verified that $\mathbf{z}$ defined above satisfies the following properties required of a IV:
  (i) $\mathbf{z}'\mathbf{U} = \mathbf{0}$ ($\mathbf{z}$ and $\mathbf{U}$ are uncorrelated),
  (ii) $\mathbf{z}'\mathbf{x} \neq 0$ ($\mathbf{z}$ and $\mathbf{x}$ are correlated),
  (iii) $\mathbf{z}'\mathbf{Q}_{[U,x]}\mathbf{y} = 0$ (*i.e.*, $\mathbf{z}$ has a predictive power on $\mathbf{y}$ only through $\mathbf{x}$).
- (i) and (ii) are trivial. That it also satisfies (3) can be seen from:

$$\mathbf{z}'\mathbf{Q}_{[U,x]}\mathbf{y} = \mathbf{x}'\mathbf{Q}_U\mathbf{Q}_{[U,x]}\mathbf{y} = \mathbf{x}'\mathbf{Q}_{[U,x]}\mathbf{y} = 0. \tag{9}$$

Thanks

Thanks for your attention.