ECON0108 2022-23 Part 1

Slides for Lecture 2

Andrew Chesher

CeMMAP & UCL

Review

- We ask: what can be known of the *structure* of an economic process given knowledge of the probability distribution of observables it produces?

- Pertinent question: distinct structures can deliver the same distribution of observables: *observational equivalence* (OE)

- A model's restrictions can deliver a situation in which each structure delivers a distinct distribution of outcomes.

- A model and distribution of observables, F_{YZ}, identifies the value of a structural feature θ if in all model-admissible OE structures that deliver F_{YZ} the value of the structural feature is constant.

- A model and a distribution F_{YZ} identify the value of a structural feature θ if there exists a functional $\mathcal{G}(\cdot)$ such that for all the structures S admitted by the model
$$\theta(S) = a \implies \mathcal{G}(F_{YZ}^{S}) = a.$$

Example

- Let F_{YZ} be the distriibution of Y and Z delivered by a structure. Consider the incomplete model

$$\tilde{Y}_1 = \alpha_0 + \alpha_1 \tilde{Y}_2 + \tilde{U} \qquad\qquad Cov[\tilde{U}\tilde{Z}] = 0 = Cov[\tilde{U}\tilde{Z}]$$

Example

- Let F_{YZ} be the distriibution of Y and Z delivered by a structure. Consider the incomplete model

$$\tilde{Y}_1 = \alpha_0 + \alpha_1 \tilde{Y}_2 + \tilde{U} \qquad\qquad Cov[\tilde{U}\tilde{Z}] = 0 = Cov[\tilde{U}\tilde{Z}]$$

- Express in terms of variables which are *deviations about expected values*, e.g. $Y_1 = \tilde{Y}_1 - E[\tilde{Y}_1]$.

$$Y_1 = \alpha_1 Y_2 + U \qquad\qquad E[UZ] = 0$$

Example

- Let F_{YZ} be the distriibution of Y and Z delivered by a structure. Consider the incomplete model

$$\tilde{Y}_1 = \alpha_0 + \alpha_1 \tilde{Y}_2 + \tilde{U} \qquad Cov[\tilde{U}\tilde{Z}] = 0 = Cov[\tilde{U}\tilde{Z}]$$

- Express in terms of variables which are *deviations about expected values*, e.g. $Y_1 = \tilde{Y}_1 - E[\tilde{Y}_1]$.

$$Y_1 = \alpha_1 Y_2 + U \qquad E[UZ] = 0$$

- There is - Wright (1928)

$$E[Y_1 Z] = \alpha E[Y_2 Z]$$

and α may be *overidentified* because for every element Z_ℓ of Z such that $E[Y_2 Z_\ell] \neq 0$

$$\alpha = \frac{E[Y_1 Z_\ell]}{E[Y_2 Z_\ell]} = \mathcal{G}_\ell(F_{YZ})$$

Identification in a treatment effects model

- Let $D \in \{0, 1\}$ indicate whether (1) an unemployed person enrolls in a training programme.

- Let $U_0 \in \{0, 1\}$ if there is (1) return to work within 1 year if **not** on a training programme ($D = 0$).

- Let $U_1 \in \{0, 1\}$ if there is (1) return to work within 1 year if **on** a training programme ($D = 1$).

- We observe (Y, D) where

$$Y = DU_1 + (1 - D)U_0$$

- The **structural feature** of interest is the average treatment effect (ATE)

$$\text{ATE}: \quad E[U_1] - E[U_0]$$

Identification in a treatment effects model

- Define $\lambda = P[D = 1]$.
- By the Law of Total Probability

$$E[U_1] = E[U_1|D = 1] \times \lambda + E[U_1|D = 0] \times (1 - \lambda)$$

$$E[U_0] = E[U_0|D = 1] \times \lambda + E[U_0|D = 0] \times (1 - \lambda)$$

Identification in a treatment effects model

- Define $\lambda = P[D = 1]$.
- By the Law of Total Probability

$$E[U_1] = E[U_1|D = 1] \times \lambda + E[U_1|D = 0] \times (1 - \lambda)$$

$$E[U_0] = E[U_0|D = 1] \times \lambda + E[U_0|D = 0] \times (1 - \lambda)$$

- Objects in red cannot be deduced from F_{YD} but they are bounded, lying in $[0, 1]$ - Manski (1990).

$$E[U_1|D = 1] \times \lambda \leq E[U_1] \leq E[U_1|D = 1] \times \lambda + (1 - \lambda)$$

Identification in a treatment effects model

- Define $\lambda = P[D = 1]$.
- By the Law of Total Probability

$$E[U_1] = E[U_1|D = 1] \times \lambda + E[U_1|D = 0] \times (1 - \lambda)$$

$$E[U_0] = E[U_0|D = 1] \times \lambda + E[U_0|D = 0] \times (1 - \lambda)$$

- Objects in red cannot be deduced from F_{YD} but they are bounded, lying in $[0, 1]$ - Manski (1990).

$$E[U_1|D = 1] \times \lambda \leq E[U_1] \leq E[U_1|D = 1] \times \lambda + (1 - \lambda)$$

$$E[U_0|D = 0] \times (1 - \lambda) \leq E[U_0] \leq \lambda + E[U_0|D = 0] \times (1 - \lambda)$$

Identification in a treatment effects model

- Define $\lambda = P[D = 1]$.
- By the Law of Total Probability

$$E[U_1] = E[U_1|D = 1] \times \lambda + E[U_1|D = 0] \times (1 - \lambda)$$

$$E[U_0] = E[U_0|D = 1] \times \lambda + E[U_0|D = 0] \times (1 - \lambda)$$

- Objects in red cannot be deduced from F_{YD} but they are bounded, lying in $[0, 1]$ - Manski (1990).

$$E[U_1|D = 1] \times \lambda \leq E[U_1] \leq E[U_1|D = 1] \times \lambda + (1 - \lambda)$$

$$-\lambda - E[U_0|D = 0] \times (1 - \lambda) \leq -E[U_0] \leq -E[U_0|D = 0] \times (1 - \lambda)$$

Identification in a treatment effects model

- There are bounds on the ATT.

$$
\begin{aligned}
E[U_1|D=1] \times \lambda - \lambda - E[U_0|D=0] \times (1-\lambda) \\
\leq \text{ATT} \leq \\
E[U_1|D=1] \times \lambda + (1-\lambda) - E[U_0|D=0] \times (1-\lambda)
\end{aligned}
$$

Identification in a treatment effects model

- There are bounds on the ATT.

$$
\begin{aligned}
E[Y|D=1]\times\lambda-\lambda-E[Y|D=0]\times(1-\lambda) \\
\leq \text{ATT} \leq \\
E[Y|D=1]\times\lambda+(1-\lambda)-E[Y|D=0]\times(1-\lambda)
\end{aligned}
$$

Identification in a treatment effects model

- There are bounds on the ATT.

$$
\begin{aligned}
&E[Y|D=1] \times \lambda - \lambda - E[Y|D=0] \times (1-\lambda) \\
&\qquad \leq \text{ATT} \leq \\
&\qquad E[Y|D=1] \times \lambda + (1-\lambda) - E[Y|D=0] \times (1-\lambda)
\end{aligned}
$$

- Bounds can be estimated. Confidence regions can be constructed separately for the upper and lower bounds, Manski et al (1992).

- A confidence region for the bounding set can be calculated, Horowitz and Manski (2000), Chernozhukov, Hong and Tamer (2003).

- A confidence region for the value of the ATT can be calculated, Imbens and Manski (2004).

Linear simultaneous equations models

- Equilibrium market model: Y_1 is quantity traded at price Y_2. M (here 2) outcomes, K exogenous variables

$$\begin{aligned} \text{demand} &: \quad Y_1 = \alpha_1 Y_2 + Z'\beta_1 + U_1 \\ \text{supply} &: \quad Y_2 = \alpha_2 Y_1 + Z'\beta_2 + U_2 \end{aligned}$$

- Notation:

$$\underset{1\times M}{Y}' \equiv \begin{bmatrix} Y_1 & Y_2 \end{bmatrix} \qquad \underset{1\times M}{U}' \equiv \begin{bmatrix} U_1 & U_2 \end{bmatrix}$$

- Define

$$\underset{M\times M}{\Gamma} \equiv \begin{bmatrix} 0 & \alpha_2 \\ \alpha_1 & 0 \end{bmatrix} \qquad \underset{M\times M}{I_M - \Gamma} \equiv \begin{bmatrix} 1 & -\alpha_2 \\ -\alpha_1 & 1 \end{bmatrix} \qquad \underset{K\times M}{B} \equiv \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix}$$

- There is

$$Y'(I_M - \Gamma) = Z'B + U'$$

Linear simultaneous equations models

- There is

$$Y'(I_M - \Gamma) = Z'B + U'$$

and so

$$Y' = Z'\Pi + V \qquad \Pi \equiv B(I_M - \Gamma)^{-1} \qquad V = U'(I_M - \Gamma)^{-1}$$

- Normalizing $E[U] = 0$ a zern uncorrelatedness restriction: $E[ZU'] = 0 \implies E[ZV'] = 0$ implies

$$E[ZY'] = E[ZZ']\Pi$$

and the model identifies Π when $\operatorname{rank} E[ZZ'] = K$.

$$\Pi = E[ZZ']^{-1}E[ZY'] = \mathcal{G}(F_{YZ})$$

- Elements of Γ and B are identified by this model if there are restrictions such that their values can be uniquely deduced from Π.

Linear simultaneous equations models

- When can the unknown elements of Γ and B be deduced from $\Pi = B(I_M - \Gamma)^{-1}$?

- There is

$$\Pi(I_M - \Gamma) = B$$

equivalently using, $vec(DEF) = (F' \otimes D)vec(E)$

$$(I_M \otimes \Pi)vec(I_M - \Gamma) - vec(B) = 0$$

- With N_R *linear* restrictions and known constant matrices R_1 and R_2 and vector r:

$$R_1 vec(I_M - \Gamma) + R_2 vec(B) = r$$

there is:

$$\begin{bmatrix} I_M \otimes \Pi & -I_{MK} \\ R_1 & R_2 \end{bmatrix} \begin{bmatrix} vec(I_M - \Gamma) \\ vec(B) \end{bmatrix} = \begin{bmatrix} 0 \\ r \end{bmatrix}$$

Linear simultaneous equations models

- There is:

$$\begin{bmatrix} I_M \otimes \Pi & -I_{MK} \\ R_1 & R_2 \end{bmatrix} \begin{bmatrix} vec(I_M - \Gamma) \\ vec(B) \end{bmatrix} = \begin{bmatrix} 0 \\ r \end{bmatrix}$$

which can be solved for $vec(I_M - \Gamma)$ and $vec(B)$ under the:

  - **Rank** condition

$$rank \begin{bmatrix} I_M \otimes \Pi & -I_{MK} \\ R_1 & R_2 \end{bmatrix} = M^2 + MK \qquad \begin{bmatrix} MK \times M^2 & MK \times MK \\ N_R \times M^2 & N_R \times MK \end{bmatrix}$$

  which can hold only under the:

  - **Order** condition:

$$MK + N_R \geq M^2 + MK$$

$$\Rightarrow \qquad N_R \geq M^2 \qquad \Rightarrow \qquad N_R - M \geq M(M-1)$$

  so e.g. in each equation $M - 1$ exclusion restrictions,

Conditional moment restrictions

- The linear model for scalar Y and $k \times 1$ vector X with a conditional mean independence restriction is.

$$Y = X'\beta + U \qquad E[U|X = x] = 0 \qquad \forall x \in \mathcal{R}_X$$

which implies

$$E[Y|X = x] = x'\beta$$

- *Suppose there are n values of x in R_X.* Define arrays:

$$\underset{(n\times k)}{X_n} \equiv \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix} \qquad \underset{(n\times 1)}{\bar{Y}_n} \equiv \begin{bmatrix} E[Y|X = x_1] \\ E[Y|X = x_2] \\ \vdots \\ E[Y|X = x_n] \end{bmatrix} = X_n\beta$$

that is:

$$\bar{Y}_n = X_n\beta \qquad \Longrightarrow \qquad X_n'\bar{Y}_n = X_n'X_n\beta$$

- So, an identifying correspondence for β is:

$$\beta = [X_n'X_n]^{-1}[X_n'\bar{Y}_n]$$

Conditional moment restrictions: GLS

- The linear model for scalar Y and $k \times 1$ vector X

$$Y = X'\beta + U \qquad E[U|X = x] = 0 \qquad x \in \mathcal{R}_X$$

which implies

$$E[Y|X = x] = x'\beta$$

- There is:

$$\bar{Y}_n = X_n\beta$$

which implies for any $k \times n$ matrix $H(X_n)$:

$$\underset{k\times n}{H(X_n)}\,\underset{n\times 1}{\bar{Y}_n} = \underset{k\times n}{H(X_n)}\,\underset{n\times k}{X_n}\,\underset{k\times 1}{\beta}$$

- If $k \times k$ matrix $H(X_n)X_n$ has rank k then

$$\beta = [H(X_n)X_n]^{-1}[H(X_n)\bar{Y}_n]$$

- For example let $H(X_n) = X_n'\Omega_n^{-1}$ then

$$\beta = [X_n'\Omega_n^{-1}X_n']^{-1}[X_n'\Omega_n^{-1}\bar{Y}_n]$$

Identification via extremum conditions

- For $n \times n$ positive definite Ω_n define

$$b_* \equiv \arg\min_b (\bar{Y}_n - X_n b)' \Omega_n^{-1} (\bar{Y}_n - X_n b)$$

- The solution is $b_* = [X_n' \Omega_n^{-1} X_n']^{-1} [X_n' \Omega_n^{-1} \bar{Y}_n]$ and we just showed

$$\beta = [X_n' \Omega_n^{-1} X_n']^{-1} [X_n' \Omega_n^{-1} \bar{Y}_n]$$

- So there is the identifying correspondence:

$$\beta = \arg\min_b (\bar{Y}_n - X_n b)' \Omega_n^{-1} (\bar{Y}_n - X_n b)$$

Maximum likelihood

- Discrete random variable Y with support $\mathcal{R}_Y$ has probability mass function $p(y, \theta)$ for some value θ_0 of θ.

$$P[Y = y] = p(y, \theta)$$

  where $p(y, \theta) \neq 0$ is a twice differentiable function of θ, $\mathcal{R}_Y$ does not vary with θ.

- With independent realizations of $Y_1, Y_2, \ldots, Y_N$ with

$$P[Y_i = y] = p(y, \theta)$$

  the maximum likelihood estimator

$$\hat{\theta} \equiv \arg\max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \log(p(Y_i, \theta))$$

  is, under a concavity restriction, an analogue estimator built on the identifying correspondence

$$\arg\max_{\theta} E[\log p(Y, \theta)] = \theta_0 \text{ when } P[Y = y] = p(y, \theta_0)$$

Maximum likelihood

- We now show why is there this identifying correspondence.

$$\arg\max_{\theta} E[\log p(Y,\theta)] = \theta_0 \text{ when } P[Y=y] = p(y,\theta_0)$$

- When $P[Y=y] = p(y,\theta_0)$

$$E[\log p(Y,\theta)] = \sum_{y \in \mathcal{R}_Y} \log p(y,\theta)\, p(y,\theta_0)$$

so we must show θ_0 satisfies

$$\nabla_\theta E[\log p(Y,\theta)] = \sum_{y \in \mathcal{R}_Y} \nabla_\theta \log p(y,\theta)|_{\theta=\theta_0}\, p(y,\theta_0) = 0$$

- This is true because

$$\begin{aligned}\sum_{y \in \mathcal{R}_Y} \nabla_\theta \log p(y,\theta)|_{\theta=\theta_0}\, p(y,\theta_0) &= \sum_{y \in \mathcal{R}_Y} \left.\frac{\nabla_\theta p(y,\theta)}{p(y,\theta)}\right|_{\theta=\theta_0} p(y,\theta_0) \\ &= \textstyle\sum_{y \in \mathcal{R}_Y} \nabla_\theta p(y,\theta)|_{\theta_0} \\ &= \nabla_\theta \textstyle\sum_{y \in \mathcal{R}_Y} p(y,\theta)|_{\theta_0} = 0\end{aligned}$$

Maximum likelihood

- We have shown that, when $P[Y = y] = p(y, \theta_0)$, θ_0 satisfies the first order condition
$$\nabla_\theta \, E[\log p(Y, \theta)] = 0$$

- If the Hessian $\nabla_{\theta\theta'} E[\log p(Y, \theta)]$ is negative definite for all θ there is the identifying correspondence
$$\theta_0 = \arg\max_\theta E[L(Y, \theta) | Y \sim p(y, \theta_0)]$$

- We have shown that the maximum likelihood estimator
$$\hat{\theta} \equiv \arg\max_\theta \frac{1}{n} \sum_{i=1}^{N} \log(p(Y_i, \theta))$$

  is an analogue estimator.

Complete models

- First study some **complete models**.
  - Complete models admit only structures which deliver a **unique** value of outcomes given values of observed and unobserved exogenous variables.

- Examples -
  - single equation models with a single endogenous variable,
  - simultaneous equations models delivering unique solutions for endogenous variables,
  - models of strategic interaction with a unique solution.

- We will consider,
  - recursive, "triangular" equation systems - popular in microeconometrics - control function methods.
  - closely related models incorporating conditional independence restrictions, and treatment effect models,

Triangular models

- Consider **complete** simultaneous equations models for **endogenous** outcomes X and Y with **triangular** structure

$$\begin{aligned} Y &= h(X, U) \\ X &= g(Z, V) \end{aligned}$$

"triangular" because endogenous Y does not appear in the equation for endogenous X.

Triangular models

- Consider **complete** simultaneous equations models for **endogenous** outcomes X and Y with **triangular** structure

$$\begin{aligned} Y &= h(X, U) \\ X &= g(Z, V) \end{aligned}$$

"triangular" because endogenous Y does not appear in the equation for endogenous X.

- In a linear simultaneous equations model with this recursive structure

$$\begin{aligned} Y &= \alpha + \beta X + U \\ X &= \gamma + \delta Z + V \end{aligned}$$

the matrix of coefficients on the endogenous variables is **triangular**.

$$\begin{bmatrix} Y & X \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\beta & 1 \end{bmatrix} = \begin{bmatrix} 1 & Z \end{bmatrix} \begin{bmatrix} \alpha & \gamma \\ 0 & \delta \end{bmatrix} + \begin{bmatrix} U & V \end{bmatrix}$$

Triangular models

- We will study models for **endogenous** outcomes X and Y with structural equations as follows:
  - parametric Gaussian model,
  $$Y = \alpha_0 + \alpha_1 X + U$$
  - nonparametric, non-Gaussian additive latent variate model,
  $$Y = h(X) + U$$
  - nonparametric, nonadditive latent variate model,
  $$Y = h(X, U)$$
  with h monotone in scalar U

Triangular Gaussian model

- Y and X are generated by a triangular Gaussian model:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 X + U \\ X &= g(Z) + V \end{aligned}$$

$$\left[\begin{array}{c} U \\ V \end{array}\right] \sim N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} \sigma_{uu} & \sigma_{uv} \\ \sigma_{uv} & \sigma_{vv} \end{array}\right]\right) \qquad \left[\begin{array}{c} U \\ V \end{array}\right] \perp\!\!\!\perp Z$$

Triangular Gaussian model

- Y and X are generated by a triangular Gaussian model:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 X + U \\ X &= g(Z) + V \end{aligned}$$

$$\left[\begin{array}{c} U \\ V \end{array}\right] \sim N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} \sigma_{uu} & \sigma_{uv} \\ \sigma_{uv} & \sigma_{vv} \end{array}\right]\right) \qquad \left[\begin{array}{c} U \\ V \end{array}\right] \perp\!\!\!\perp Z$$

- Conditional distribution of U given V and Z

$$U|V = v, Z = z \sim N\left(\frac{\sigma_{uv}}{\sigma_{vv}} v, \sigma_{uu} - \frac{\sigma_{uv}^2}{\sigma_{vv}}\right)$$

Triangular Gaussian model

- Y and X are generated by a triangular Gaussian model:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 X + U \\ X &= g(Z) + V \end{aligned}$$

$$\begin{bmatrix} U \\ V \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{uu} & \sigma_{uv} \\ \sigma_{uv} & \sigma_{vv} \end{bmatrix}\right) \qquad \begin{bmatrix} U \\ V \end{bmatrix} \perp\!\!\!\perp Z$$

- Conditional distribution of U given V and Z

$$U|V = v, Z = z \sim N\left(\frac{\sigma_{uv}}{\sigma_{vv}} v, \sigma_{uu} - \frac{\sigma_{uv}^2}{\sigma_{vv}}\right)$$

- The expected value of Y given V and Z.

$$E[Y|V = v, Z = z] = \alpha_0 + \alpha_1 (g(z) + v) + \frac{\sigma_{uv}}{\sigma_{vv}} v$$

Given $Z = z$, there is $V = v$ if and only if $X = x \equiv g(z) + v$, so:

$$E[Y|X = x, Z = z] = \alpha_0 + \alpha_1 x + \frac{\sigma_{uv}}{\sigma_{vv}} (x - g(z))$$

Triangular Gaussian model

- Structures admitted by this model have:

$$E[Y|X = x, Z = z] = \alpha_0 + \alpha_1 x + \frac{\sigma_{uv}}{\sigma_{vv}} (x - g(z)) \qquad E[X|Z = z] = g(z)$$

so

$$\begin{aligned} E[Y|X = x, Z = z] &= \alpha_0 + \alpha_1 x + \frac{\sigma_{uv}}{\sigma_{vv}} (x - E[X|Z = z]) \\ &= \alpha_0 + \left(\alpha_1 + \frac{\sigma_{uv}}{\sigma_{vv}}\right) x - \frac{\sigma_{uv}}{\sigma_{vv}} E[X|Z = z] \end{aligned}$$

Triangular Gaussian model

- Structures admitted by this model have:

$$E[Y|X=x,Z=z] = \alpha_0 + \alpha_1 x + \frac{\sigma_{uv}}{\sigma_{vv}}(x - g(z)) \qquad E[X|Z=z] = g(z)$$

so

$$\begin{aligned} E[Y|X=x,Z=z] &= \alpha_0 + \alpha_1 x + \frac{\sigma_{uv}}{\sigma_{vv}}(x - E[X|Z=z]) \\ &= \alpha_0 + \left(\alpha_1 + \frac{\sigma_{uv}}{\sigma_{vv}}\right)x - \frac{\sigma_{uv}}{\sigma_{vv}}E[X|Z=z] \end{aligned}$$

- Consider two values z' and z'' and define

$$x' = E[X|Z=z'] \qquad x'' = E[X|Z=z'']$$

and note that

$$\begin{aligned} E[Y|X &= x', Z=z'] = \alpha_0 + \alpha_1 x' \\ E[Y|X &= x'', Z=z''] = \alpha_0 + \alpha_1 x'' \end{aligned}$$

and so

$$\alpha_1 = \frac{E[Y|X=x'',Z=z''] - E[Y|X=x',Z=z']}{E[X|Z=z''] - E[X|Z=z']}$$

Triangular Gaussian model: analogue estimation

- Structures admitted by this model have:

$$
\begin{aligned}
E[Y|X=x,Z=z] &= \alpha_0 + \alpha_1 x + \frac{\sigma_{uv}}{\sigma_{vv}}\left(x - E[X|Z=z]\right) \\
&= \alpha_0 + \left(\alpha_1 + \frac{\sigma_{uv}}{\sigma_{vv}}\right) x - \frac{\sigma_{uv}}{\sigma_{vv}} E[X|Z=z]
\end{aligned}
$$

- Analogue estimation involves estimating $E[X|Z=z]$ then:

- Estimating e.g. by OLS the coefficients in

$$
Y_i = \alpha_0 + \left(\alpha_1 + \frac{\sigma_{uv}}{\sigma_{vv}}\right) X_i - \frac{\sigma_{uv}}{\sigma_{vv}} E[\widehat{X|Z=}Z_i] + \varepsilon_i, \quad i \in \{1,\ldots,N\}
$$

or

$$
Y_i = \alpha_0 + \alpha_1 X_i + \frac{\sigma_{uv}}{\sigma_{vv}}\left(X_i - E[\widehat{X|Z=}Z_i]\right) + \varepsilon_i, \quad i \in \{1,\ldots,N\}
$$