Structural equation models for quantitative genetic analysis

Genetics (2004) 167, 1407-1424

Daniel Gianola

Daniel Sorensen

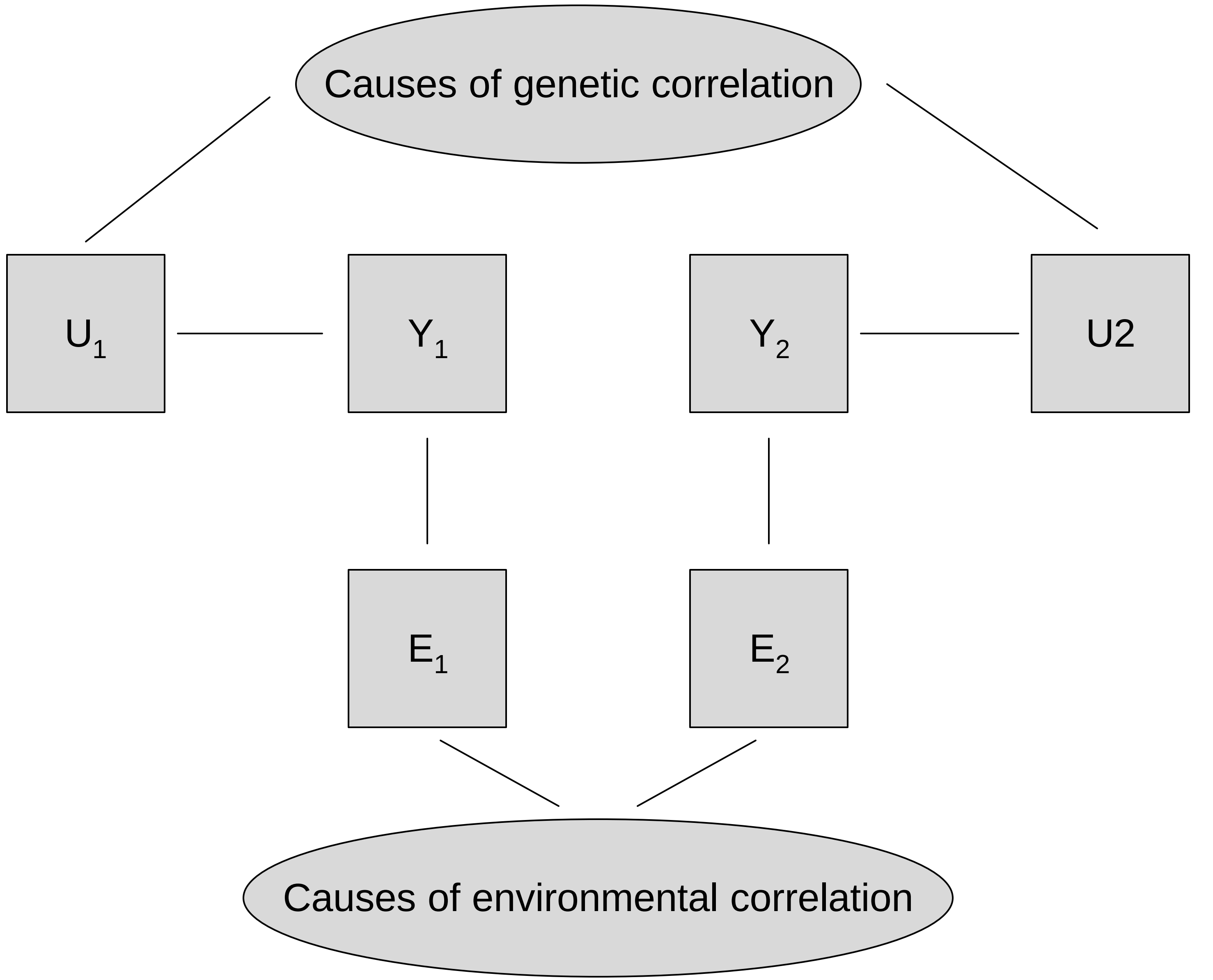
Ministry of Food, Agriculture and Fisheries
Danish Institute of Agricultural Sciences

Multivariate models important

- Multiple trait selection in plant and animal breeding
- Evolution of fitness depends on genetic variance-covariance parameters
- Correct statistical representation of multivariate models

➔ Quality of inferences

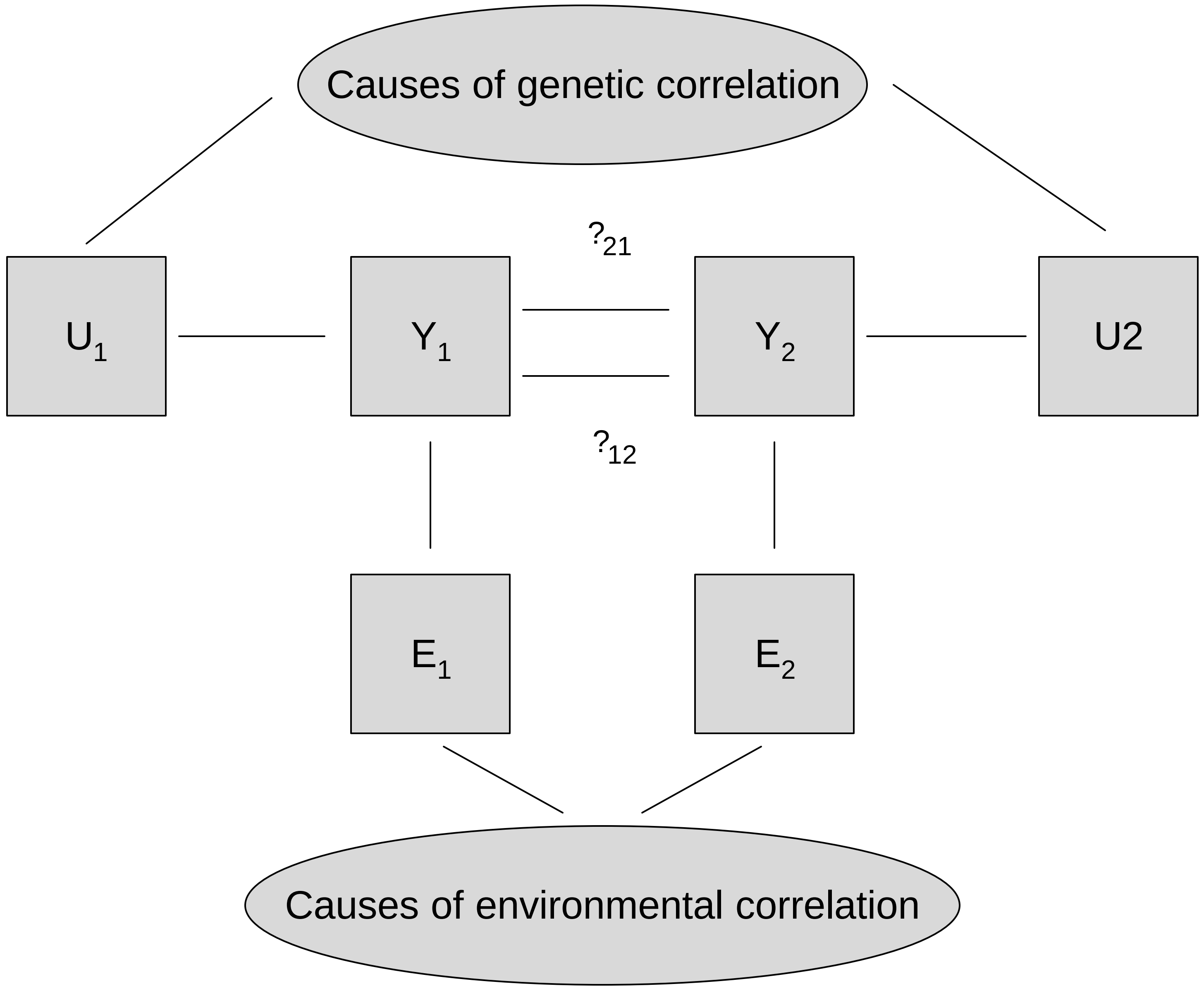
STANDARD BIVARIATE MODEL



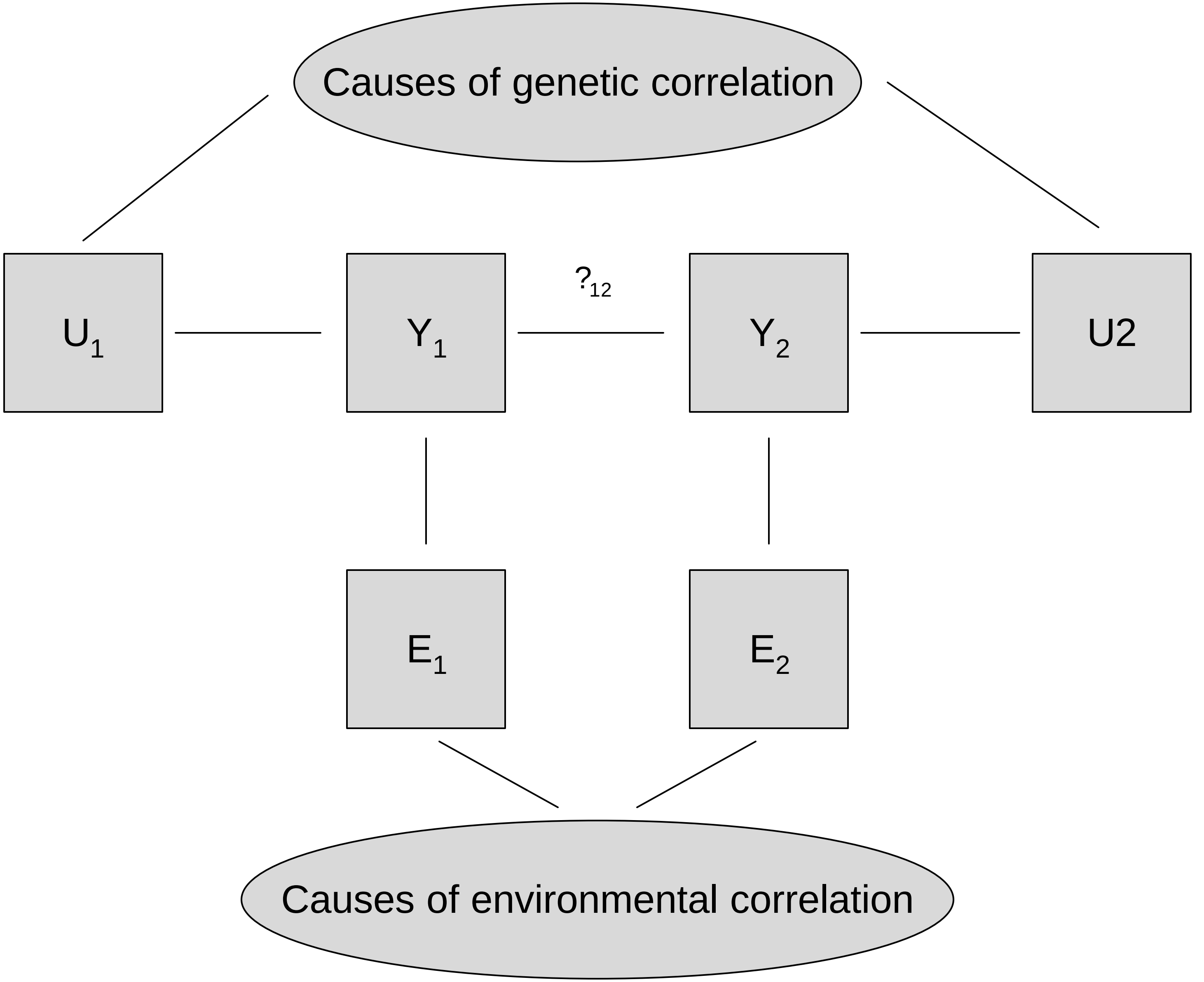
What if there are feedback or simultaneity relationships between variables?

➔Feedback inhibition well known in biology

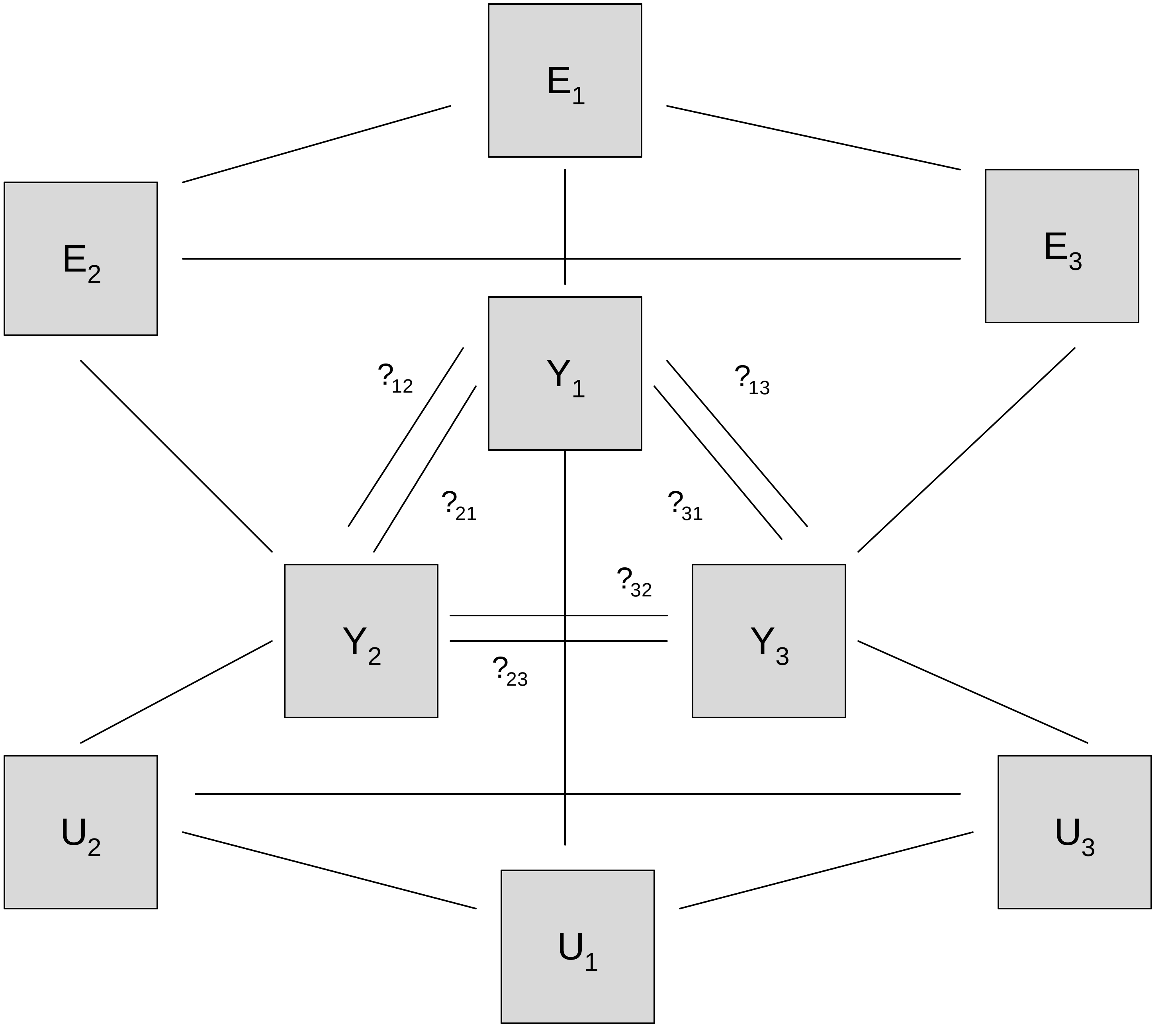
A SIMPLE SIMULTANEOUS EQUATION MODEL



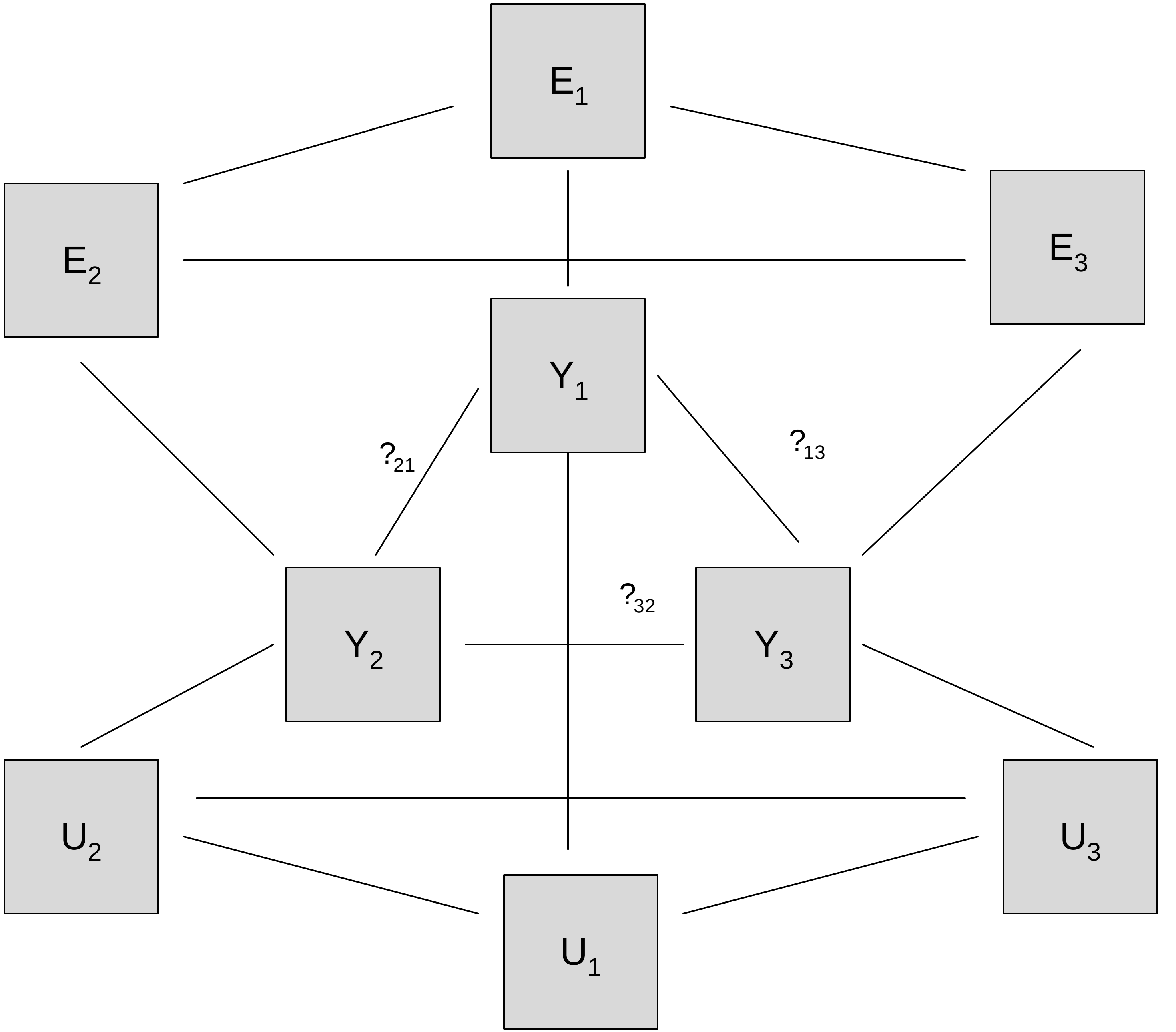
A SIMPLER MODEL



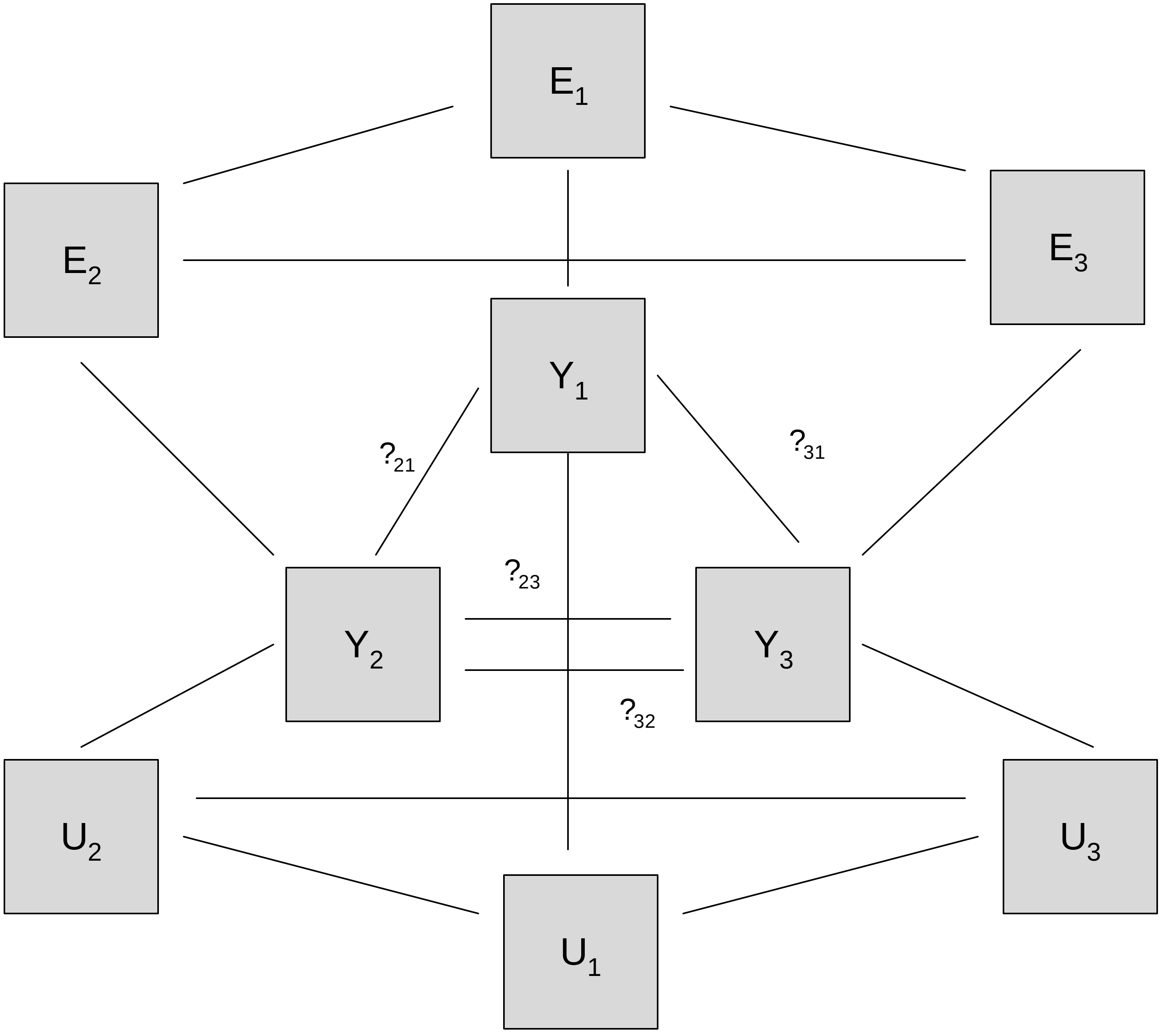
MODEL WITH 3 VARIABLES: FULL SIMULTANEITY



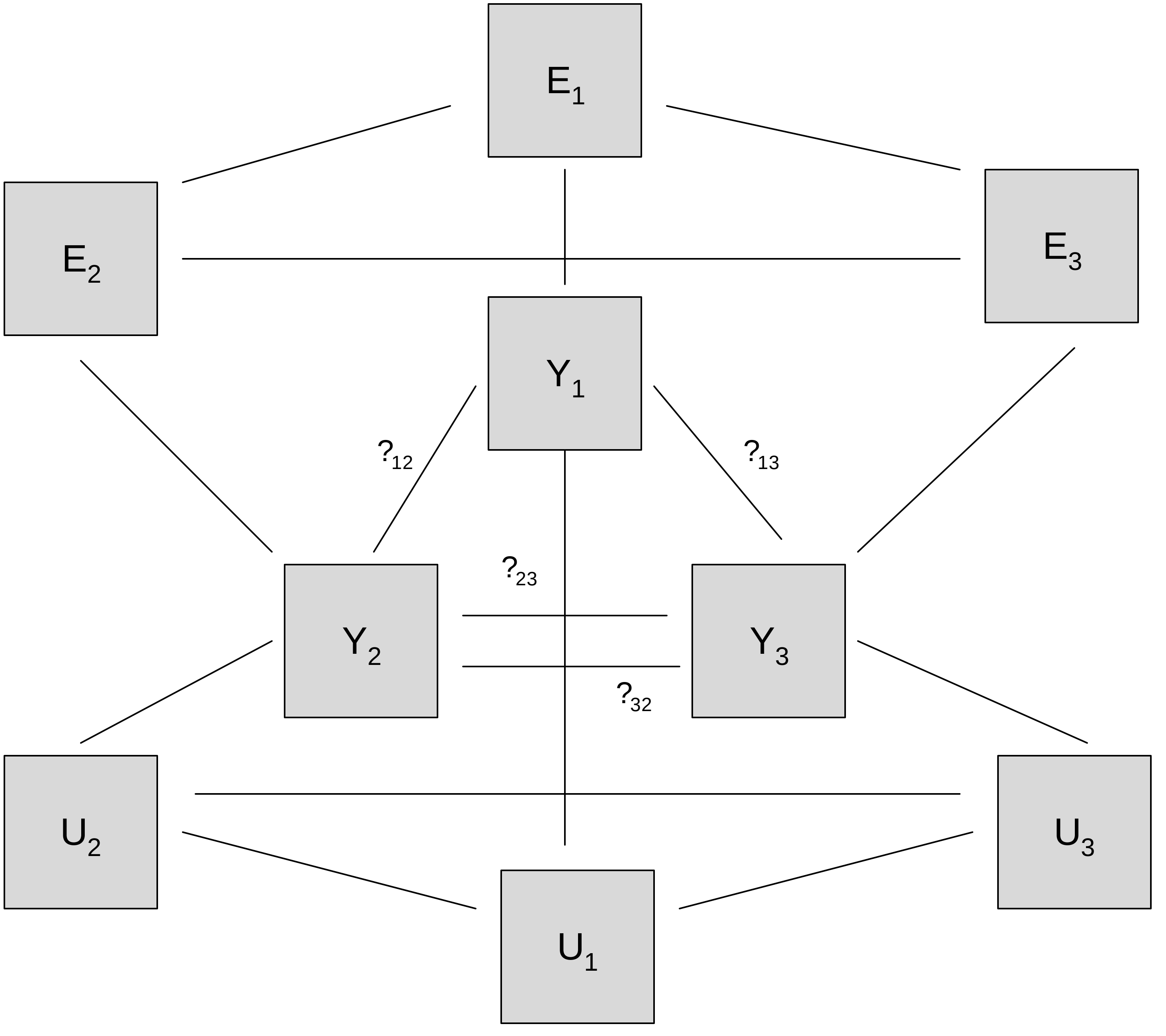
MODEL WITH 3 VARIABLES: FULL RECURSIVENESS



MODEL WITH 3 VARIABLES: RECURSIVENESS AND SIMULTANEITY

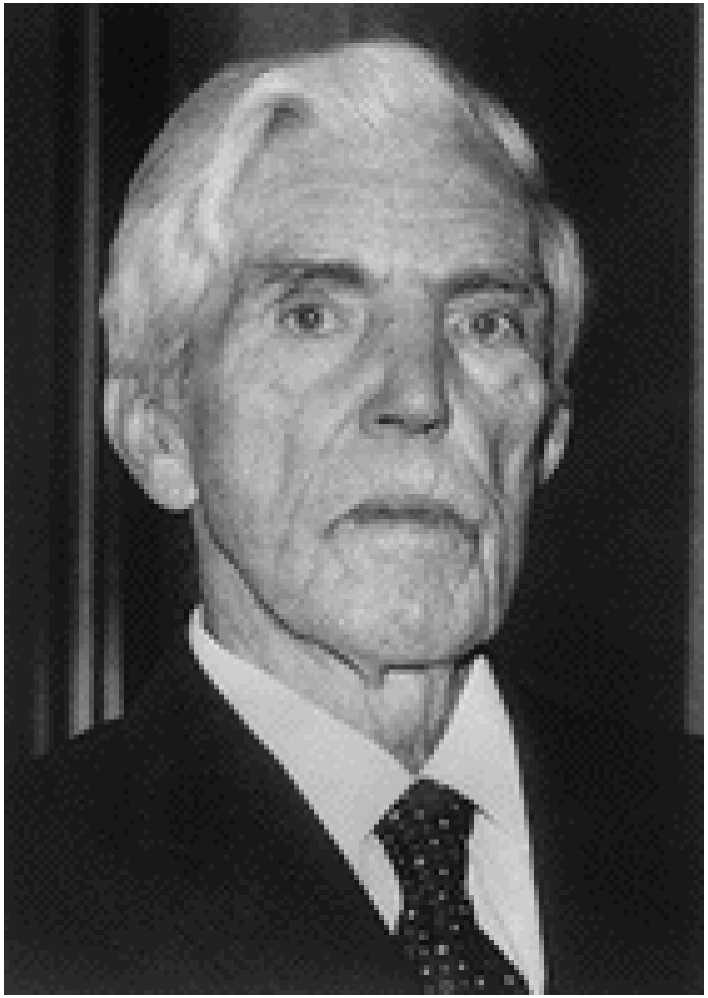


MODEL WITH 3 VARIABLES: SIMULTANEITY AND RECURSIVENESS



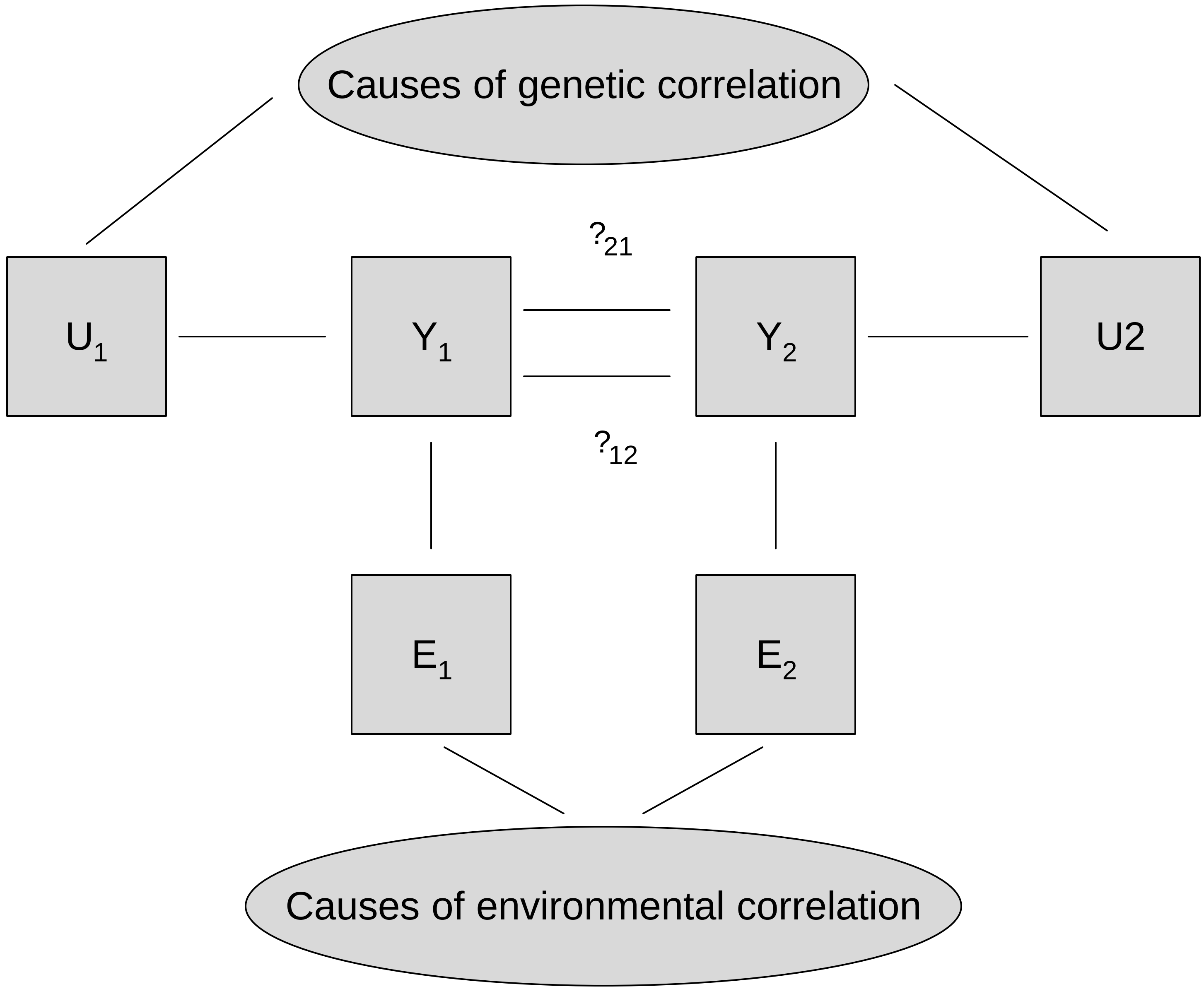
Sewall Wright

Corn and Hog Correlations [Wright, 1925]
Path analysis with reciprocal interaction [Wright, 1960]



Trygve Haavelmo: 1989 Nobel Prize in Economics

A SIMPLE SIMULTANEOUS EQUATION MODEL



PUTTING THE SYSTEM IN EQUATIONS

$$y_{i1} = \lambda_{12} y_{i2} + \mathbf{x}'_{i1} \boldsymbol{\beta}_1 + u_{i1} + e_{i1},$$

Effect of disease on production

$$y_{i2} = \lambda_{21} y_{i1} + \mathbf{x}'_{i2} \boldsymbol{\beta}_2 + u_{i2} + e_{i2}$$

Effect of production on disease

HOW TO INTERPRET THE GENETIC AND RESIDUAL EFFECTS?

$$y_{i1} - \lambda_{12} y_{i2} = \mathbf{x}'_{i1} \boldsymbol{\beta}_1 + u_{i1} + e_{i1}$$

$$y_{i2} - \lambda_{21} y_{i1} = \mathbf{x}'_{i2} \boldsymbol{\beta}_2 + u_{i2} + e_{i2}$$

OFFSPRING-PARENT REGRESSION

$$b_{OP} = \frac{\frac{1}{2}\left(\sigma_{u1}^{2}+\lambda_{12}^{2}\sigma_{u2}^{2}\right)+\lambda_{12}\sigma_{u12}}{\sigma_{u1}^{2}+\sigma_{e1}^{2}+2\lambda_{12}(\sigma_{u12}+\sigma_{e12})+\lambda_{12}^{2}\left(\sigma_{u2}^{2}+\sigma_{e2}^{2}\right)}$$

GENETIC CORRELATION

$$Corr(u_{i1}^*, u_{i2}^*) = \frac{(1+\lambda_{12}\lambda_{21})\sigma_{u12}+\lambda_{21}\sigma_{u1}^2+\lambda_{12}\sigma_{u2}^2}{\sqrt{\left(\sigma_{u1}^2+2\lambda_{12}\sigma_{u12}+\lambda_{12}^2\sigma_{u2}^2\right)\left(\sigma_{u2}^2+2\lambda_{21}\sigma_{u12}+\lambda_{21}^2\sigma_{u1}^2\right)}}$$

MATRIX REPRESENTATION OF MODELS NEEDED

$$\begin{bmatrix} 1 & -\lambda_{12} \\ -\lambda_{21} & 1 \end{bmatrix} \begin{bmatrix} y_{i1} \\ y_{i2} \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_{i1} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}'_{i2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i1} \end{bmatrix}$$

$$\boldsymbol{\Lambda}\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i + \mathbf{e}_i$$

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}'_{i1} & \mathbf{0} & . & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}'_{i2} & . & \mathbf{0} & \mathbf{0} \\ . & . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \mathbf{x}'_{i(K-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & . & \mathbf{0} & \mathbf{x}'_{iK} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ . \\ \boldsymbol{\beta}_{K-1} \\ \boldsymbol{\beta}_K \end{bmatrix}$$

REDUCED FORM OF MODEL

$$\begin{aligned}\mathbf{y}_i &= \boldsymbol{\Lambda}^{-1}\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Lambda}^{-1}\mathbf{u}_i + \boldsymbol{\Lambda}^{-1}\mathbf{e}_i \\ &= \boldsymbol{\mu}_i^* + \mathbf{u}_i^* + \mathbf{e}_i^*\end{aligned}$$

$$\mathbf{u}_i|\mathbf{G}_0 \sim N(\mathbf{0}, \mathbf{G}_0); \quad i = 1, 2, \ldots, N$$

$$\mathbf{e}_i|\mathbf{R}_0 \sim N(\mathbf{0}, \mathbf{R}_0)$$

$$\mathbf{u}_i^*|\boldsymbol{\Lambda}, \mathbf{G}_0 \sim N(\mathbf{0}, \boldsymbol{\Lambda}^{-1}\mathbf{G}_0\boldsymbol{\Lambda}'^{-1})$$

$$\mathbf{e}_i^*|\boldsymbol{\Lambda}, \mathbf{R}_0 \sim N(\mathbf{0}, \boldsymbol{\Lambda}^{-1}\mathbf{R}_0\boldsymbol{\Lambda}'^{-1})$$

$$
\begin{bmatrix} \Lambda\mathbf{y}_1 \\ \Lambda\mathbf{y}_2 \\ \cdot \\ \Lambda\mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \cdot \\ \mathbf{X}_N \end{bmatrix} \boldsymbol{\beta} + \mathbf{Z} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \cdot \\ \mathbf{u}_N \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \cdot \\ \mathbf{e}_N \end{bmatrix}
$$

$$
= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}
$$

$$
\mathbf{u}|\mathbf{G}_0 \sim N(\mathbf{0}, \mathbf{A} \otimes \mathbf{G}_0)
$$

$$
\mathbf{e}|\mathbf{R}_0 \sim N(\mathbf{0}, \mathbf{I} \otimes \mathbf{R}_0)
$$

LIKELIHOOD FUNCTION

$$l(\boldsymbol{\Lambda}, \boldsymbol{\beta}, \mathbf{R}_0, \mathbf{G}_0) \propto \int N(\mathbf{X}_\Lambda \boldsymbol{\beta} + \mathbf{Z}_\Lambda \mathbf{u}, \mathbf{R}_\Lambda) N(\mathbf{0}, \mathbf{A} \otimes \mathbf{G}_0) \mathbf{du}$$

$$\propto N(\mathbf{X}_\Lambda \boldsymbol{\beta} + \xcancel{\mathbf{Z}_\Lambda \mathbf{u}}, \mathbf{R}_\Lambda + \mathbf{Z}_\Lambda (\mathbf{A} \otimes \mathbf{G}_0) \mathbf{Z}'_\Lambda)$$

The structural parameters enter both in the incidence matrices and in the residual covariance matrix

➔DIFFICULT TO MAXIMIZE THE LIKELIHOOD

SYSTEM CONTAINS (POTENTIALLY)

➔ K^2 PARAMETERS (ELEMENTS OF **?**)

➔ $K\sum_{j=1}^{K} p_j$ (ELEMENTS OF B)

➔ $K(K+1)$ (ELEMENTS OF GENETIC AND RESIDUAL COVARIANCE MATRICES)

➔ $=K^2(2+p)+K$ (TOTAL PARAMETERS)

REDUCED MODEL CONTAINS

➔ $=K^2(1+p)+K$ (TOTAL PARAMETERS)

Need restrictions for uniqueness

$$\left[K^2(2+p)+K\right]-\left[K^2(1+p)+K\right]=K^2$$

RESTRICTIONS CAN BE

- “<u>Normalization</u>” (set diagonals of **?** to 1)
- “<u>Exclusion</u>” (elements of **?** or of **ß** may be null)
- Restrictions as linear combinations of parameters within and across the *K* equations
- Restrictions on the variance covariance matrices (typically not made)
- Rank checks used for assessing identification of system

IDENTIFIABILITY SITUATIONS

1) System just-identified: unique relationship between parameters of reduced and structural model (may be non-linear)
2) Overidentified: many ways in which parameters can be expressed from structural model
3) Underidentified: structural parameters cannot be solved from reduced model parameterization

IF (1) OR (2), PARAMETERS CAN BE INFERRED USING MAXIMUM LIKELIHOOD OR BAYESIAN PROCEDURES

BAYESIAN INFERENCE and MCMC

PRIOR

DATA

POSTERIOR

PRIOR DISTRIBUTIONS

- Structural parameters $\boldsymbol{\lambda}|\lambda_0, \tau^2 \sim N(\mathbf{1}\lambda_0, \mathbf{I}\tau^2)$
- Beta-coefficients $\boldsymbol{\beta}|\beta_0, b^2 \sim N(\mathbf{1}\beta_0, \mathbf{I}b^2)$
- Genetic effects $\mathbf{u}|\mathbf{G}_0 \sim N(\mathbf{0}, \mathbf{G}_0)$
- Residual covariance matrix $\mathbf{R}_0|\nu_R, \mathbf{V}_R \sim IW(\nu_R, \mathbf{V}_R)$
- Genetic covariance matrix $\mathbf{G}_0|\nu_G, \mathbf{V}_G \sim IW(\nu_G, \mathbf{V}_G)$

MCMC algorithm

- Metropolis within Gibbs

  ➔Gibbs proposals for “fixed” and random effects, and for covariance matrices

  ➔Metropolis for **?** (Gibbs if recursive system)
- After burn-in, draw *m* samples
- Estimate features using ergodic averages, e.g.

$$\widehat{b}_{OP} = \frac{1}{m}\sum_{j=1}^{m} \frac{\frac{1}{2}\left(\sigma_{u1}^{2(j)} + \lambda_{12}^{2(j)}\sigma_{u2}^{2(j)}\right) + \lambda_{12}^{(j)}\sigma_{u12}^{(j)}}{\sigma_{u1}^{2(j)} + \sigma_{e1}^{2(j)} + 2\lambda_{12}^{2(j)}\left(\sigma_{u12}^{(j)} + \sigma_{e12}^{(j)}\right) + \lambda_{12}^{2(j)}\left(\sigma_{u2}^{2(j)} + \sigma_{e2}^{2(j)}\right)}$$

CONCLUSIONS

- Extension of linear model to feedback and recursive relationships
- Implications on interpretation of genetic analysis
- Some extensions
  - ➔Non-linear models
  - ➔Discrete random variables
  - ➔Hierarchical modeling of **?**’s