Path Analysis

- Types of Associations
- Recursive vs. Non-recursive
- Tracing rules
- Counting df
- Sample size recommendations

Additional readings

- Oct 26 Path analysis
  - » Chemers, Hu, & Garcia (2001)
- Nov 2 CFA
  - » Geary et al. (1996)
- Nov 9 Hybrid model
  - » Nunez (2009)

- Please read before class
- We may come back to these later in the quarter, as well

Types of Associations

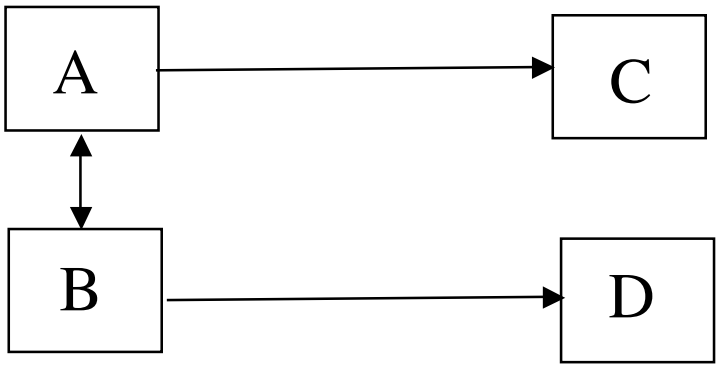
- Unanalyzed Association
  - » non-directional relationship
  - » the type evaluated by Pearson correlation
  - » represented by double-headed arrow
- Direct
  - » a directional relationship between variables
  - » the type of association evaluated in multiple regression or ANOVA
  - » the building block of SEM models
  - » represented by a single-headed arrow

Types of Associations

- Indirect
  - » Two (or more) directional relationships, in series
  - » V1 affects V2 which in turns affects V3
  - » relationship between V1 and V3 is mediated by V2
  - » represented by a series of single-headed arrows
- Spurious
  - » V2 and V3 are both caused by V1
  - » V2 and V3 are correlated because of their common cause
  - » represented by two single-headed arrows
- Total
  - » sum of all direct and indirect effects

Relationships between Variables

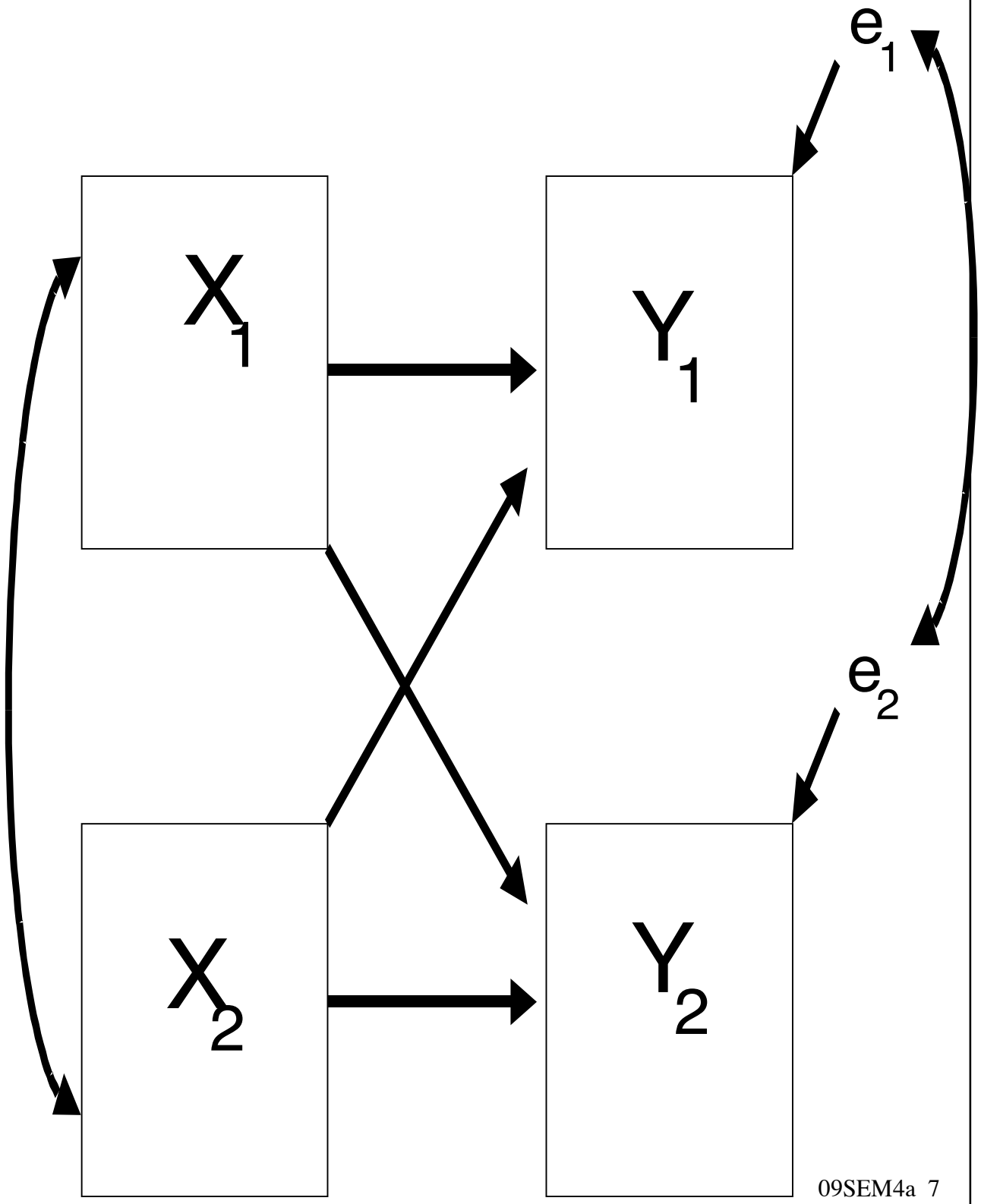
- Absence of a direct effect of X on Y does not mean that X is an unimportant predictor of Y
  - » X may have important indirect effects
- Absence of direct and indirect effects does not mean 2 variables are uncorrelated
  - » they may be correlated due to common causal antecedents



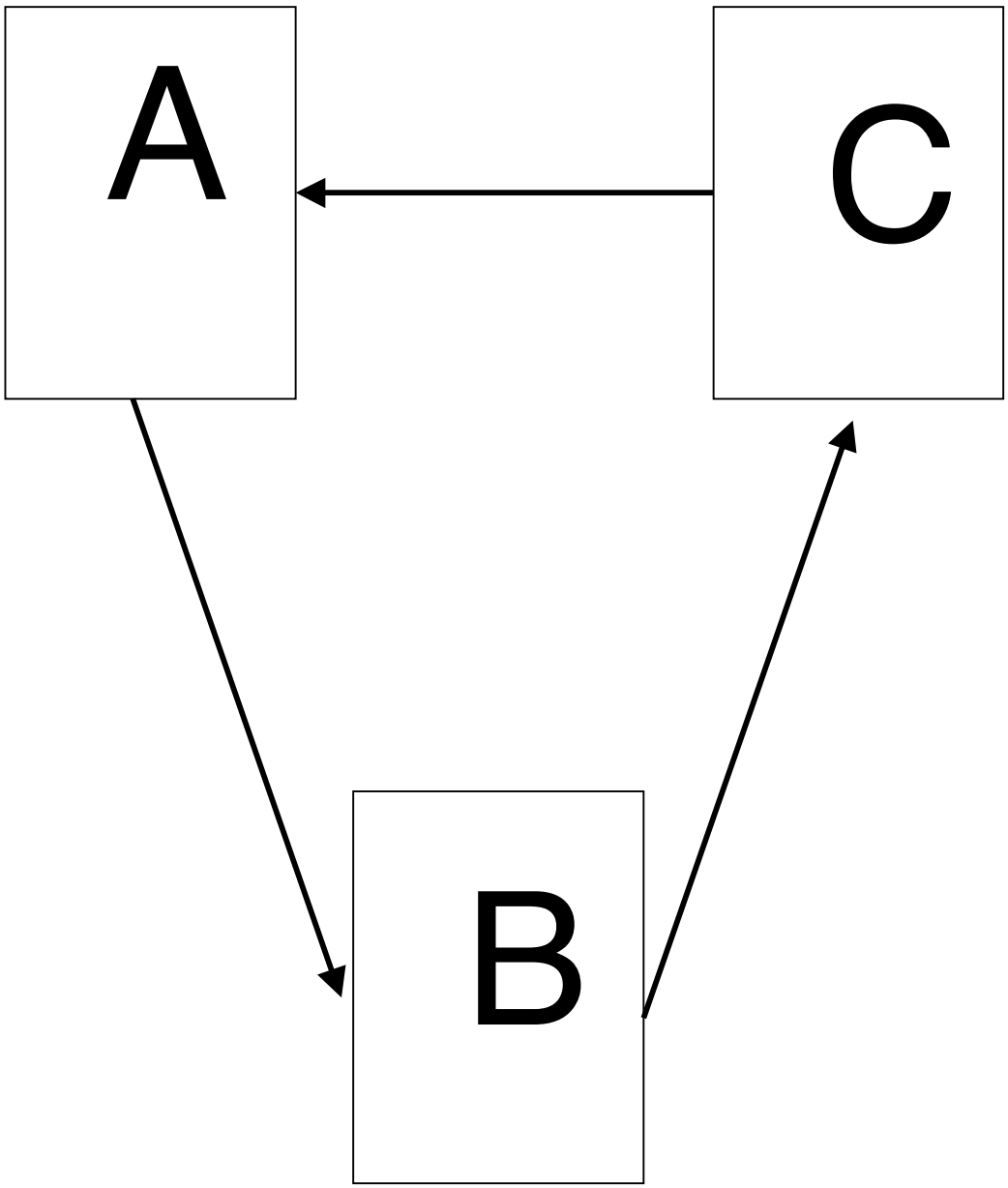
Recursive vs. Non-recursive Models

- In a **recursive** model, causation flows in only one direction
  - » a consequent variable never exerts causal influence (either directly or indirectly) on an antecedent variable that first exerts causal influence on it
- In a **non-recursive** model, causation may flow in more than one direction
  - » a variable may have a direct or indirect effect on another variable that preceded it in the causal chain

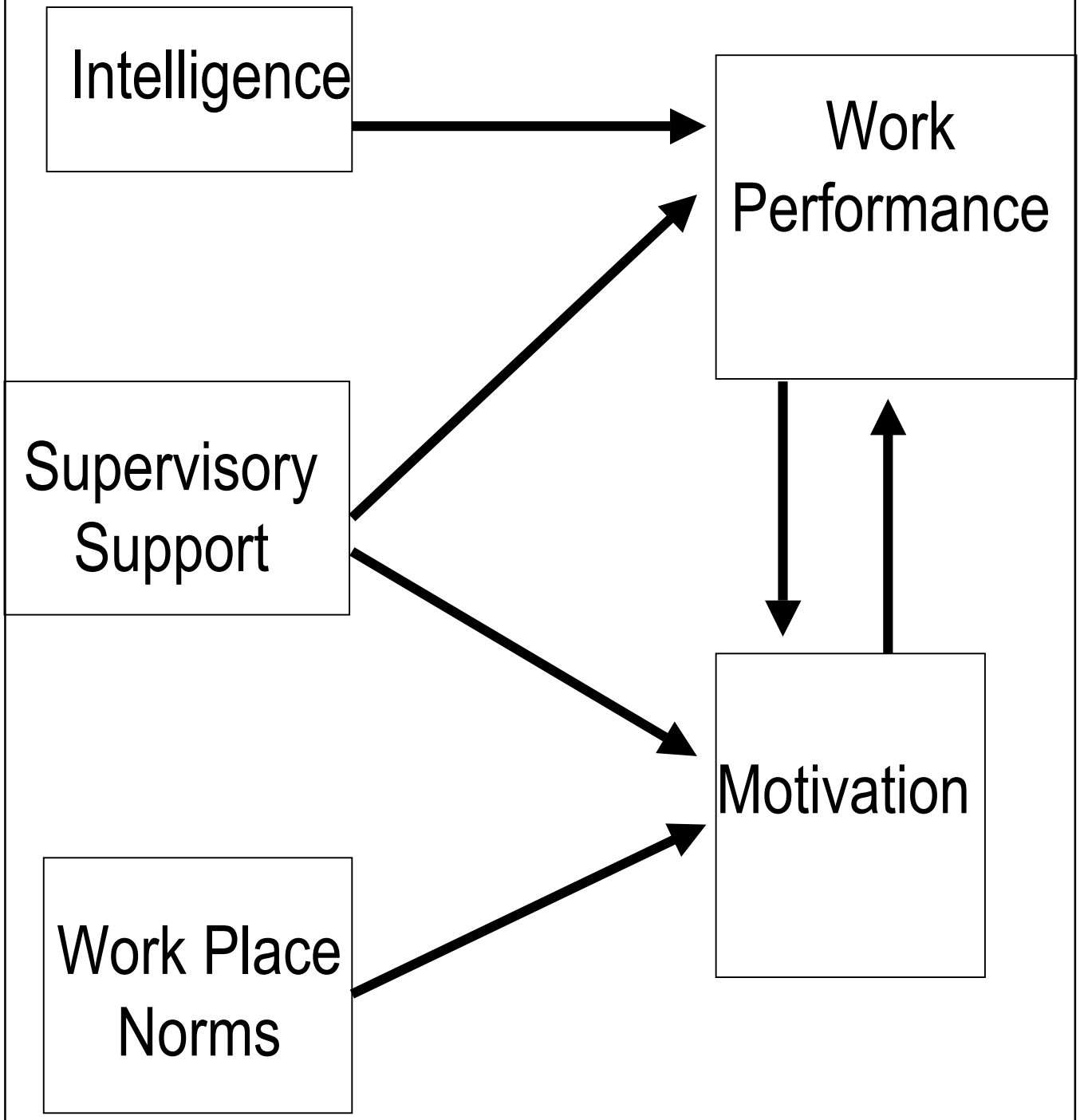
Recursive



Nonrecursive



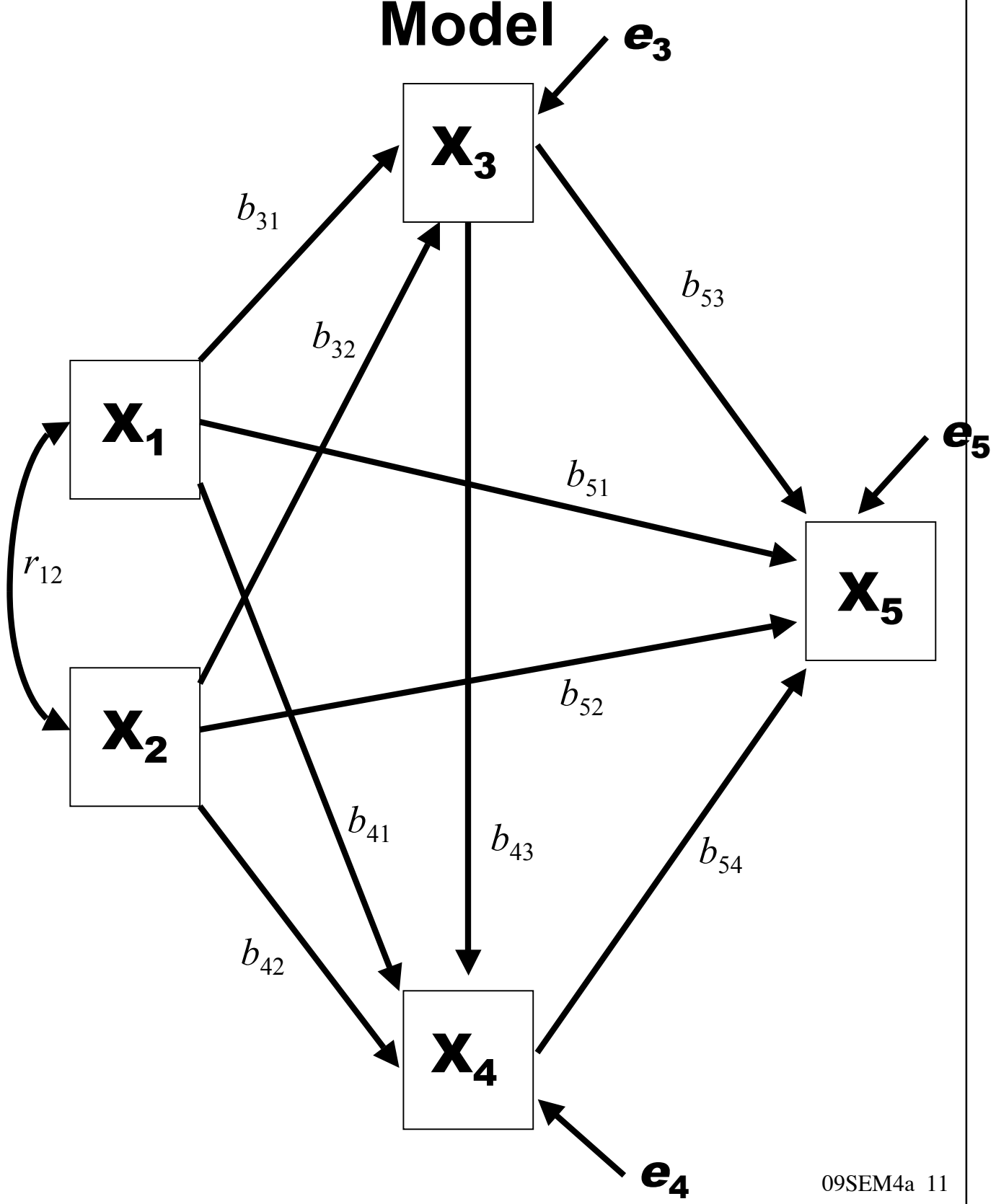
Nonrecursive



Diagrams and Equations

- Path models are usually represented in journal articles via diagrams
  - » they can be specified this way in most SEM programs, as well
- Equivalently, they can be represented as regression-like equations
- You should be able to specify a model with either method

Five-Variable Path Model



Equations for the Previous Path Diagram

$$X_3 = b_{31}*X_1 + b_{32}*X_2 + e_3$$

$$X_4 = b_{41}*X_1 + b_{42}*X_2 + b_{43}*X_3 + e_4$$

$$X_5 = b_{51}*X_1 + b_{52}*X_2 + b_{53}*X_3 + b_{54}*X_4 + e_5$$

Computing Model-Implied Correlations

- Recall that SEM will minimize the difference between the actual covariance (or correlation) matrix and the model-implied covariance (or correlation) matrix
- For recursive models, can compute the model-implied correlation by hand
  - » using the tracing rule

Rules for Computing Tracing

1. If one variable causes another, always start with the one that is the effect. If they are not directly causally related, then the starting point is arbitrary. But once a start variable is selected, always start there.
2. Start against an arrow (that is, go from effect to cause). Remember, the goal at this point is to go from the start variable to the other variable.
3. Each particular tracing of paths between the two variables can go thru only one noncausal path.

Rules for Computing Tracing

4. For each particular tracing of paths, any intermediate variable can be included only once.
5. The tracing can go back against paths (from effect to cause) for as far as possible, but, regardless of how far back, once the tracing goes forward causally, it cannot turn back against an arrow.

- Example: Find the 5 unique tracings between X3 and X4

a) X_4 to its cause X_3; or b_{43}

b) X_4 to its cause X_1 and then with an arrow to X_3, $b_{41}*b_{31}$ (noncausal due to shared antecedent X_1)

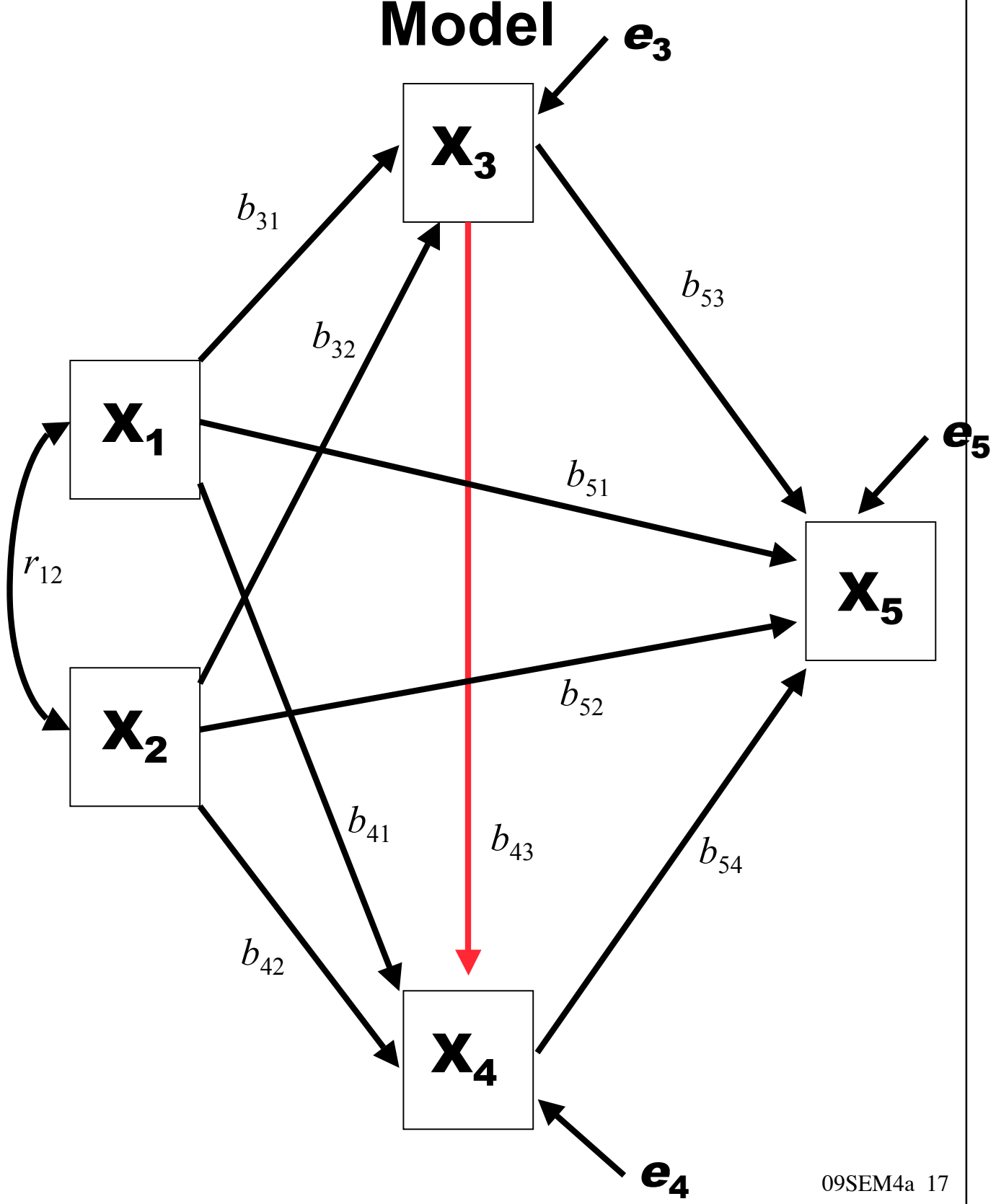
c) X_4 to its cause X_2 and then to X_3, $b_{42}*b_{32}$ (noncausal due to shared antecedent X_2)

d) X_4 to its cause X_1 thru a noncausal path to X_2 and to X_3, $b_{41}*r_{12}*b_{32}$

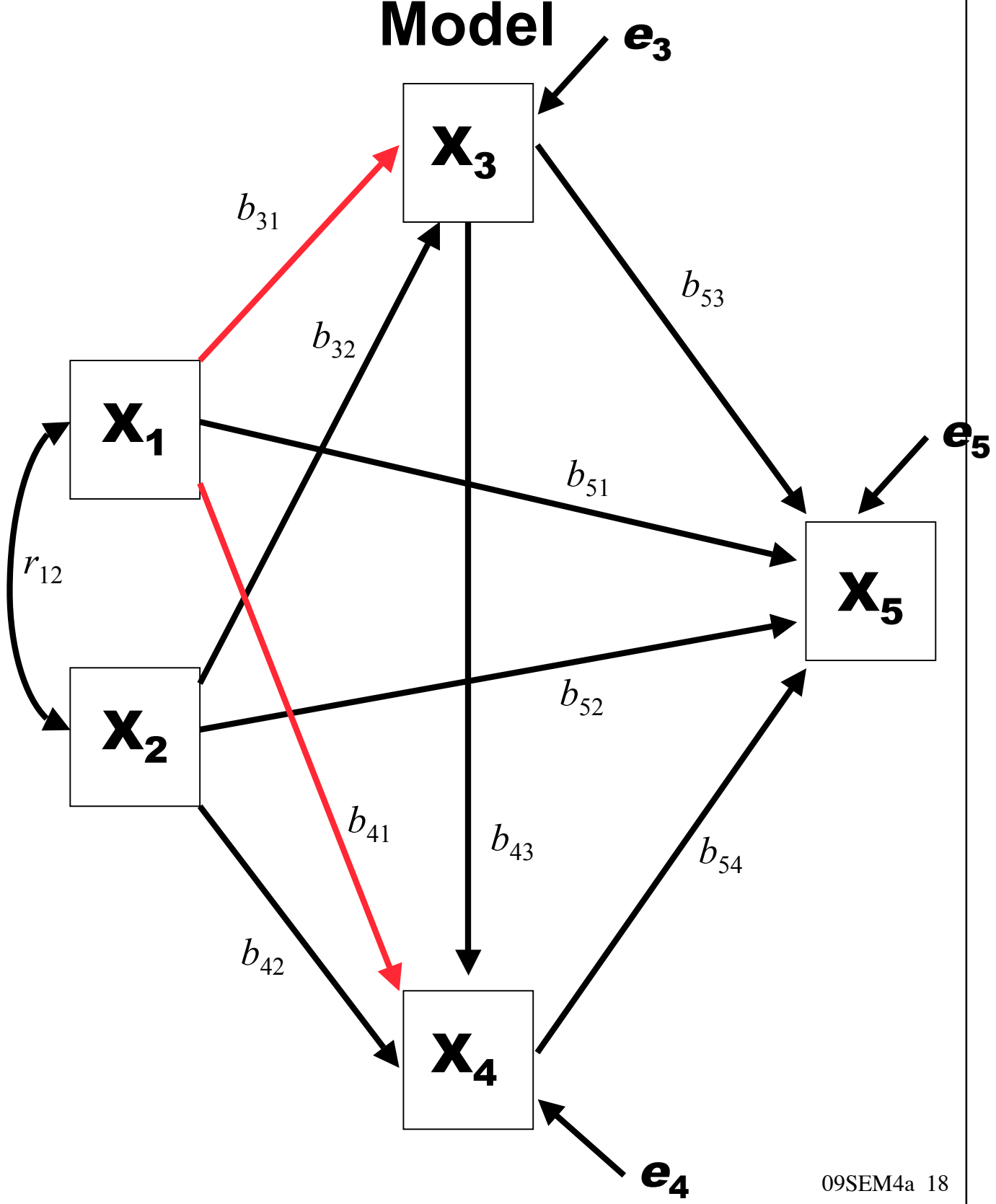
e) X_4 to its cause X_2 thru a noncausal path to X_1 and to X_3, $b_{42}*r_{12}*b_{31}$

	Causal		Noncausal	
	Direct	Indirect	Shared Antecedent	Unanalyzed Prior Association
a	b_{43}			
b			$b_{41}*b_{31}$	
c			$b_{42}*b_{32}$	
d				$b_{41}*r_{12}*b_{32}$
e				$b_{42}*r_{12}*b_{31}$

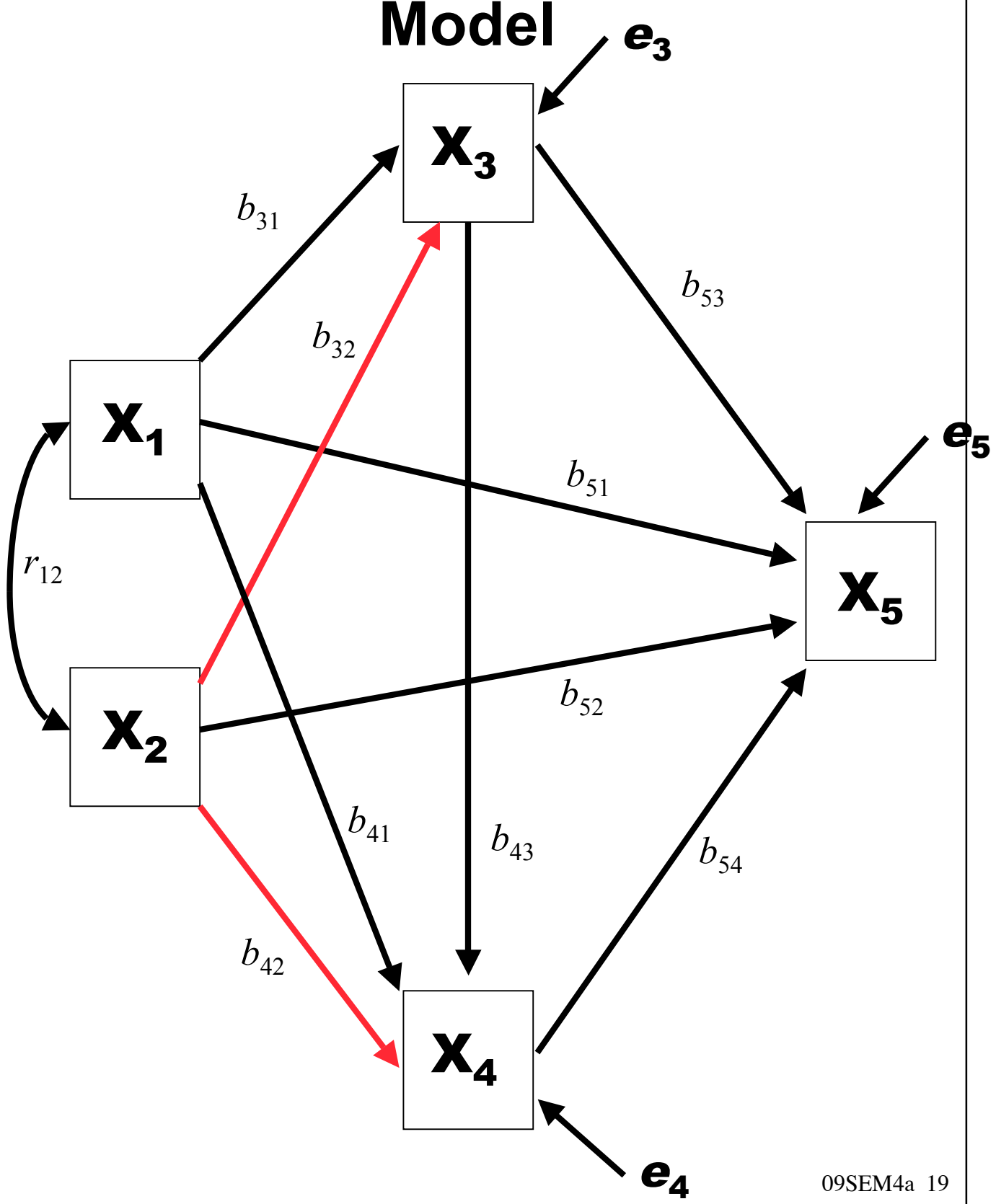
Five-Variable Path Model



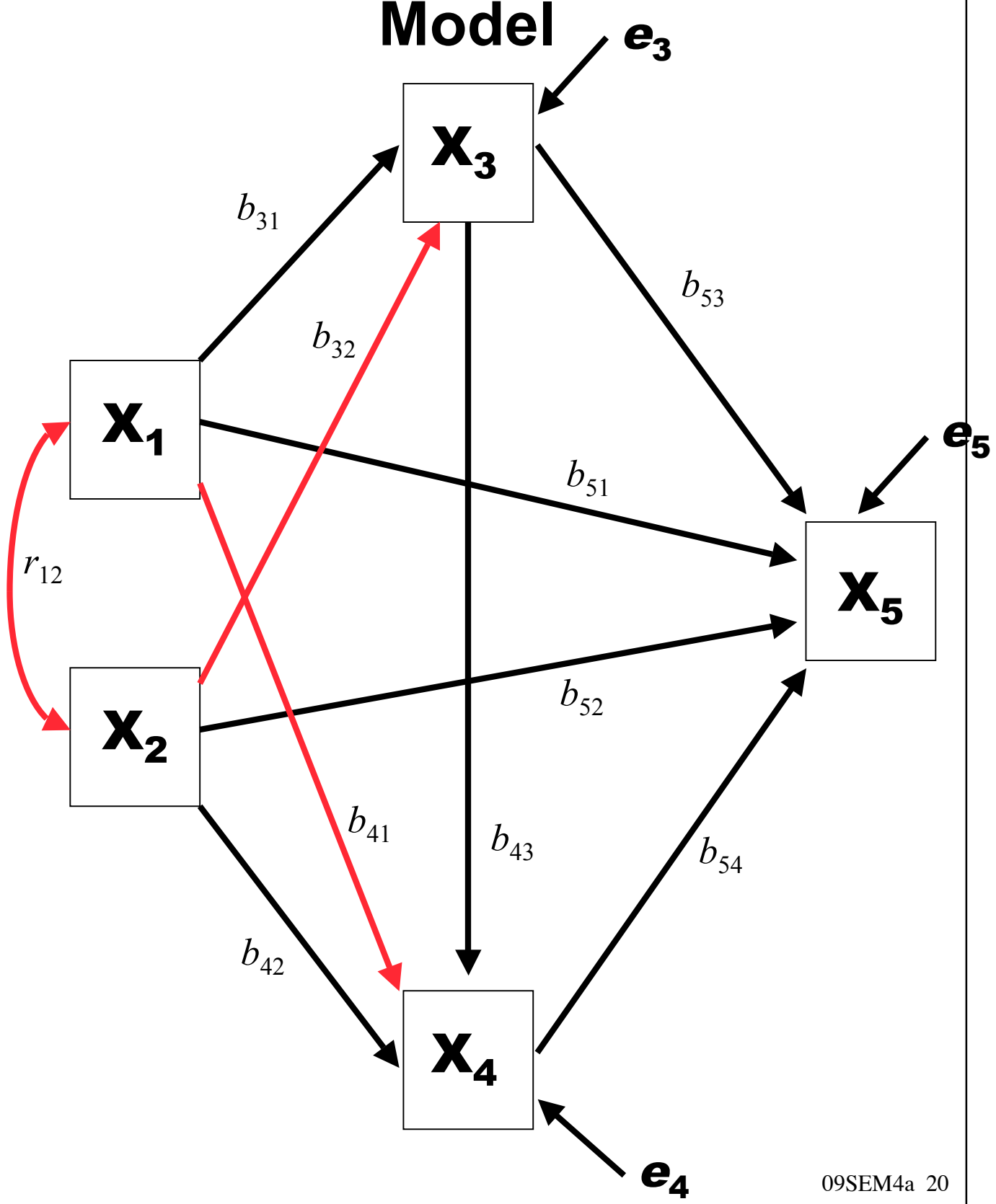
Five-Variable Path Model



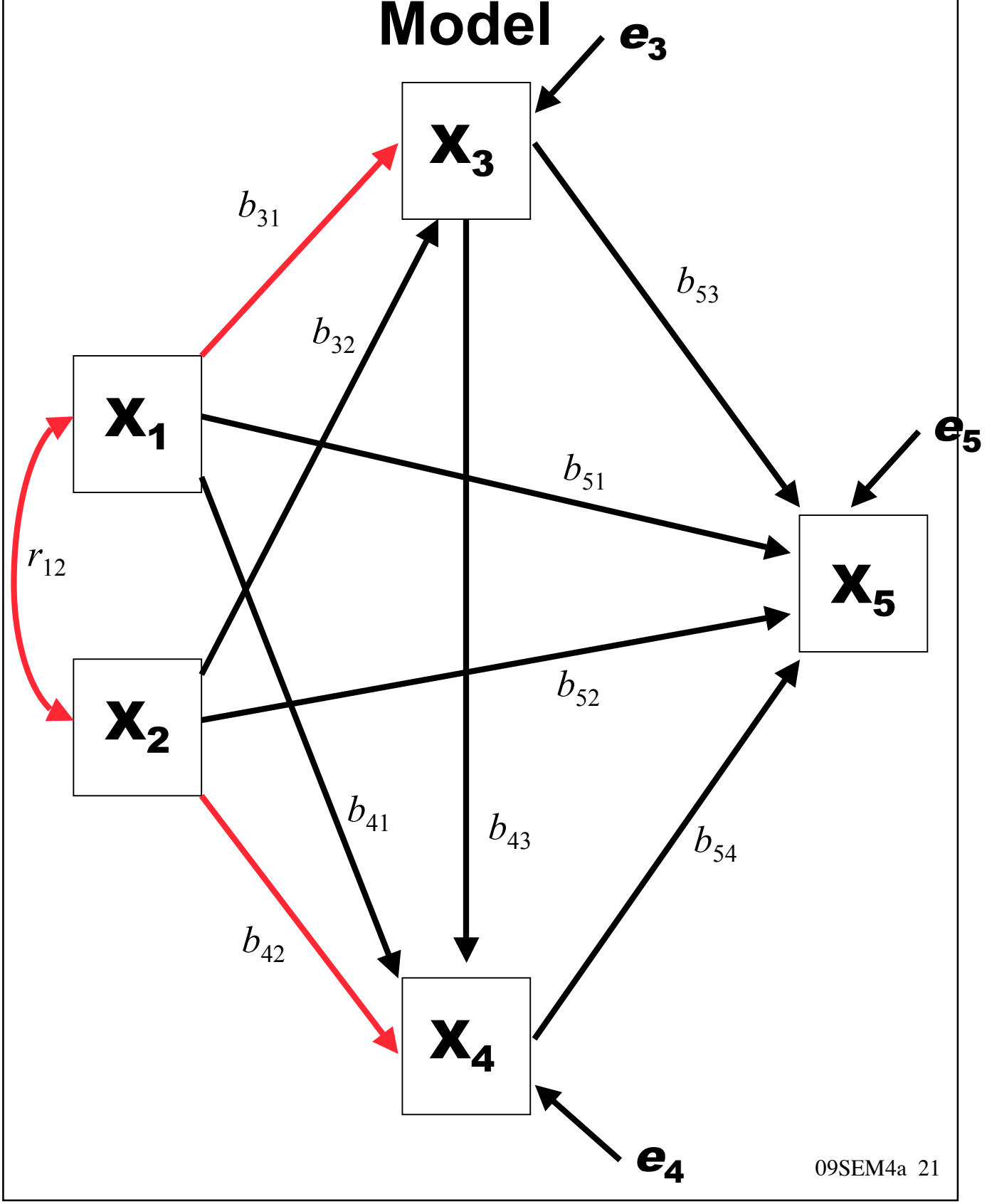
Five-Variable Path Model



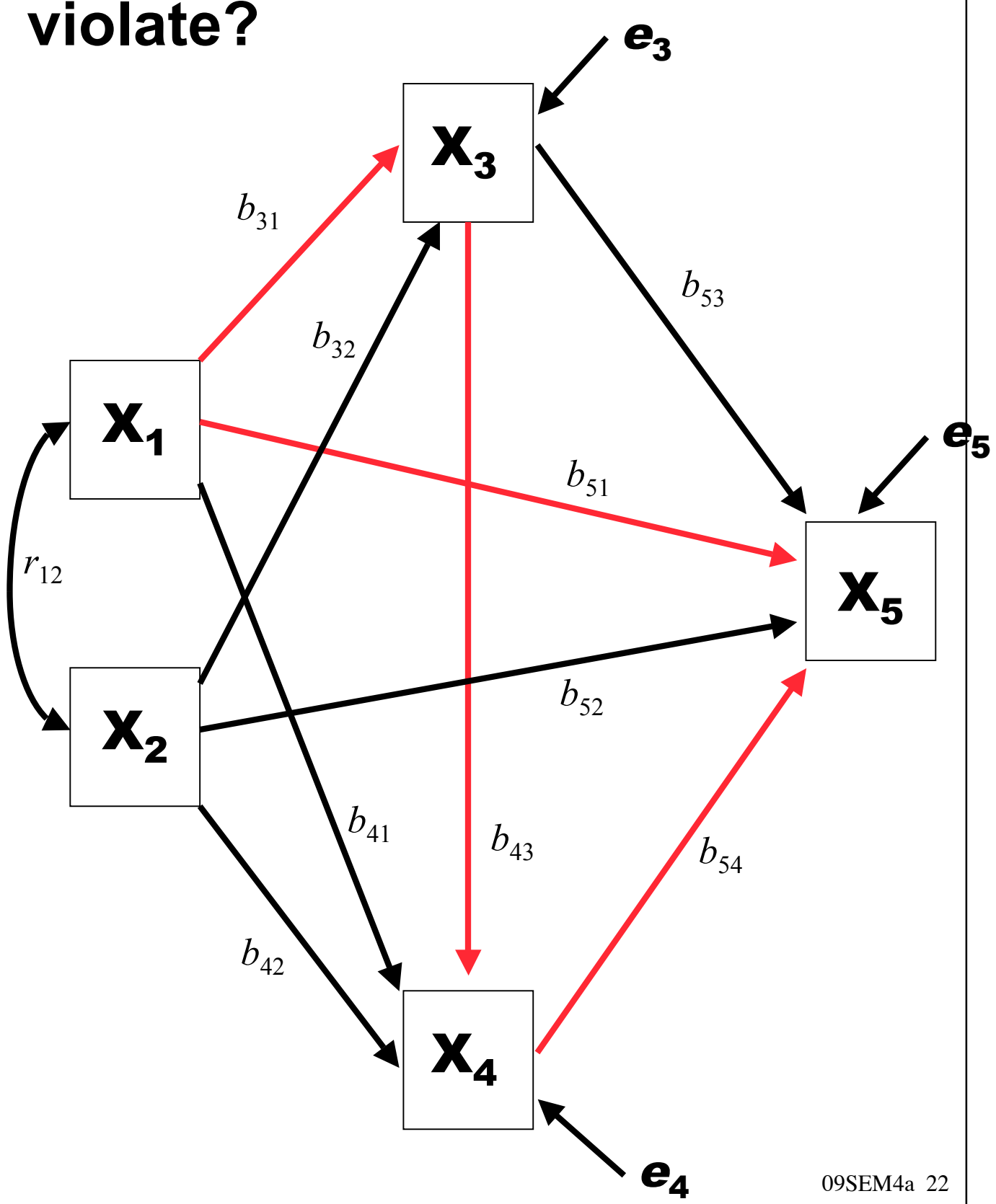
Five-Variable Path Model



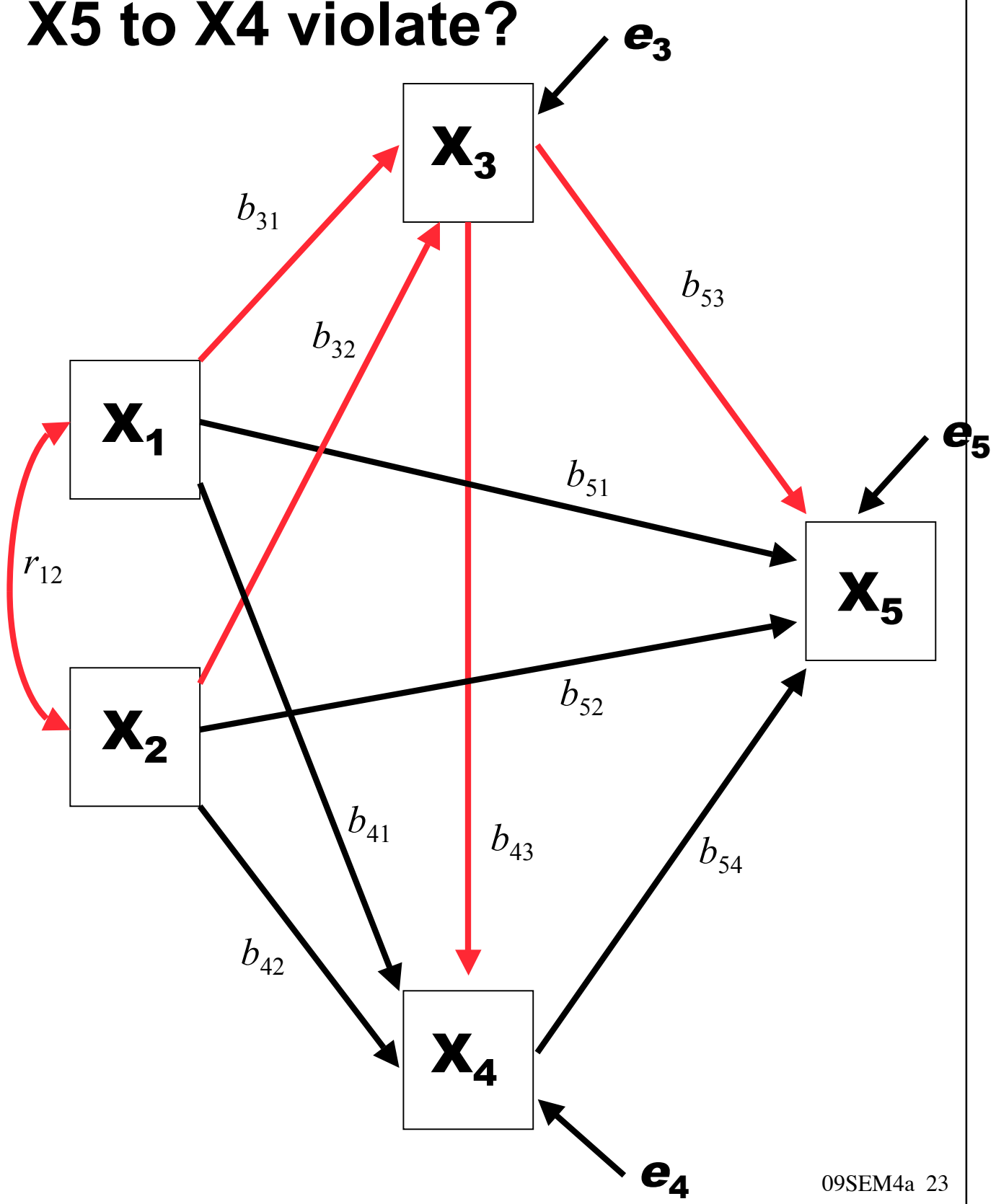
Five-Variable Path Model



What rule does this tracing violate?



What rule does this tracing from X5 to X4 violate?



Counting Degrees of Freedom

- In regression, the number of df for error = # observations - # of parameters estimated
  - » 100 participants, 3 predictors
  - » 100 observations
  - » estimating 4 parameters (B0, B1, B2, B3)
  - » df for error = 96
- In SEM, it's the same idea
- Except here, the number of observations = number of observed variances and covariances
- Formula: $v(v+1)/2$
  - » where v = number of observed variables
  - » e.g., in 5-variable model
    - $5*6/2 = 30/2 = 15$

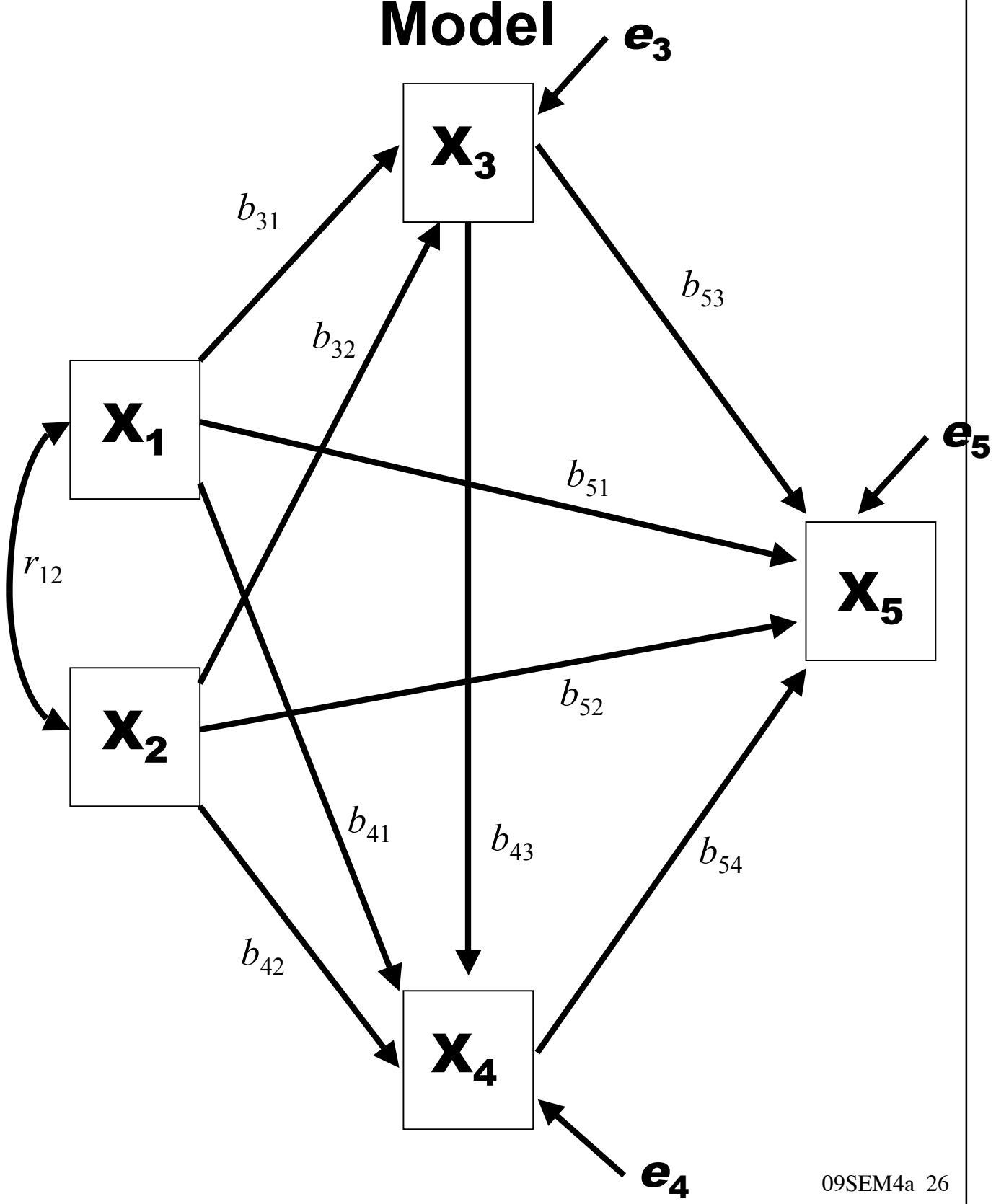
Counting Parameters

- Variances for all exogenous variables. observed and unobserved (e.g., disturbances)
- Any covariances between exogenous variables, observed and unobserved
- Direct effects on endogenous variables

OR

- Count 1 for each exogenous variable (for its variance)
- + 1 for each arrow (single or double headed)
- except for arrows where the path coefficient is fixed to a constant
  - » as when setting the scale of a latent variable

Five-Variable Path Model



DF for error

- When # of parameters to estimate = # of observations
  - » model is just identified
  - » also called a "saturated" model
  - » fit is perfect
- The more df for error, the more chance of bad fit
  - » therefore, the more impressive to get good fit

Sample Size

- SEM is a large sample technique
  - » here, we refer to the number of people or cases, not the number of variances and covariances
- In addition to issues of power (minimizing Type II errors), estimates may be unstable if small samples
  - » < 100 considered small sample
    - Klein says: "untenable" except for extremely simple models
  - » 100-200 considered medium sample
  - » > 200 considered large sample
- More complicated models require large samples for stable estimates
- Ratio of cases to parameters
  - » 20:1 best, 10:1 ok, 5:1 absolute minimum