

# **Discussion of “Learning Equivalence Classes of Acyclic Models with Latent and Selection Variables from Multiple Datasets with Overlapping Variables”**

**Jiji Zhang and Ricardo Silva**

Lingnan University/University College London

AISTATS 2011 – Fort Lauderdale, FL

# On overlapping variables and partial information

---

- ▶ Three main issues in statistical learning: estimation, computation and identification
- ▶ The link to Spirtes (2001): “An anytime algorithm for causal inference”
  - ▶ The problem, then: estimating Markov equivalence classes when independence assessments stop at a particular order
  - ▶ T&S can be seen as a generalization in some directions: from incomplete independence assessments to more general equivalence classes



# Built-in robustness

---

- ▶ Incorrect decisions on qualitative information might be less likely by pooling data from different sources

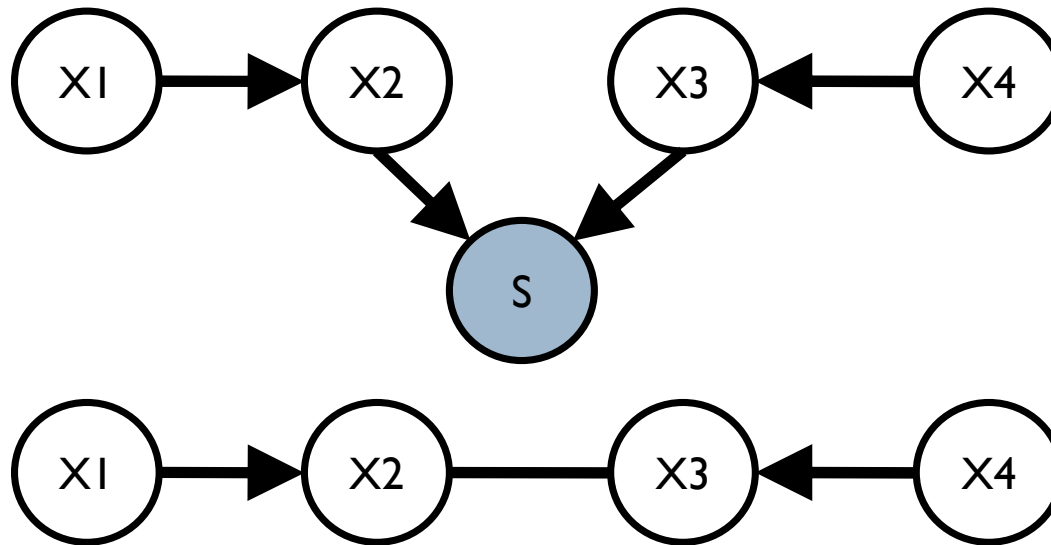
“[It] has been suggested that causal discovery methods based solely on associations will find their greatest potential in longitudinal studies conducted under slightly varying conditions, where accidental independencies are destroyed and only structural independencies are preserved.” (Pearl, 2009, p.63)



# On selection bias

---

- ▶ The assumption of common structure on separate studies also implies common selection bias



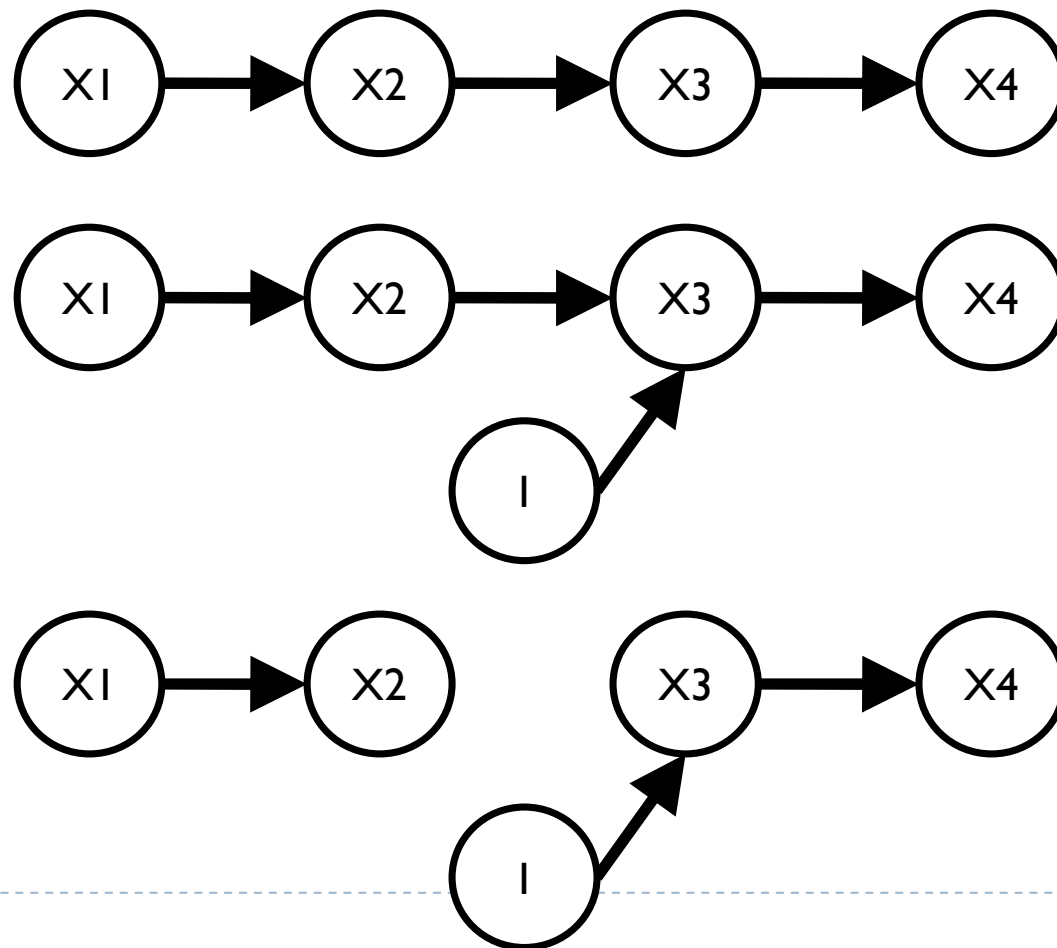
- ▶ T&S make the role of structural assumptions very clear, but the one concerning common selection bias structure might not be as believable depending on the domain
- 



# Combining different interventional studies

---

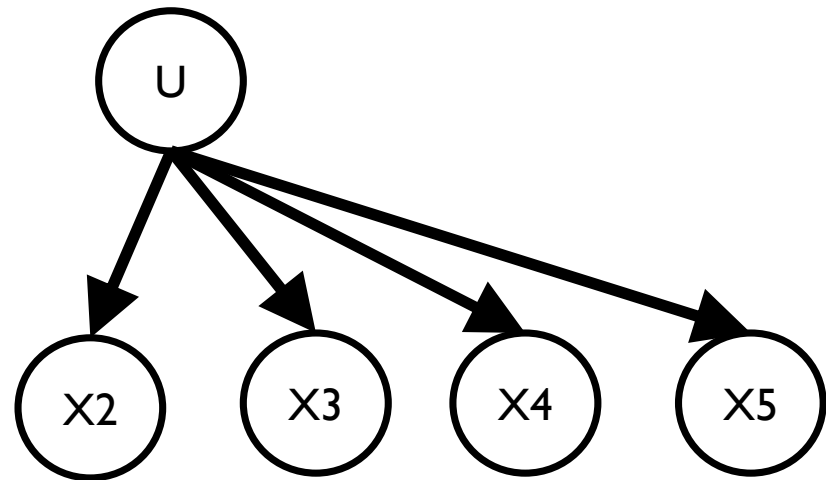
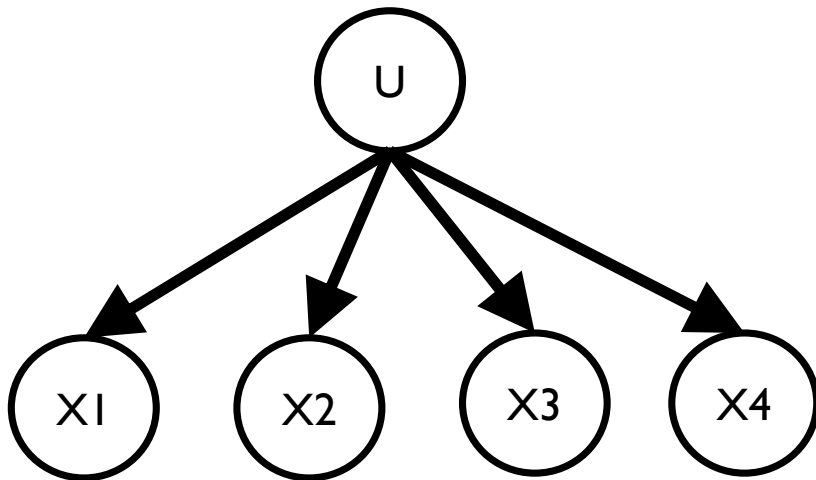
- ▶ Structures might change given different interventional studies



# Beyond independence constraints

---

- ▶ Cudeck (2000): overlapping factor analysis, finding  $\sigma_{15}$

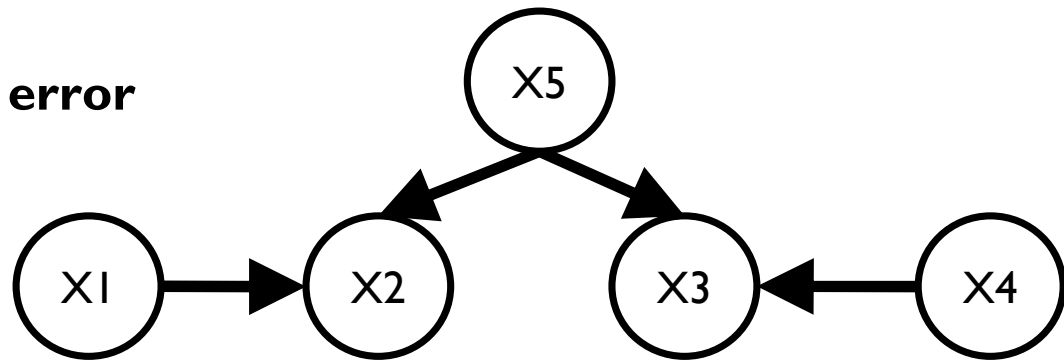


# Beyond independence constraints

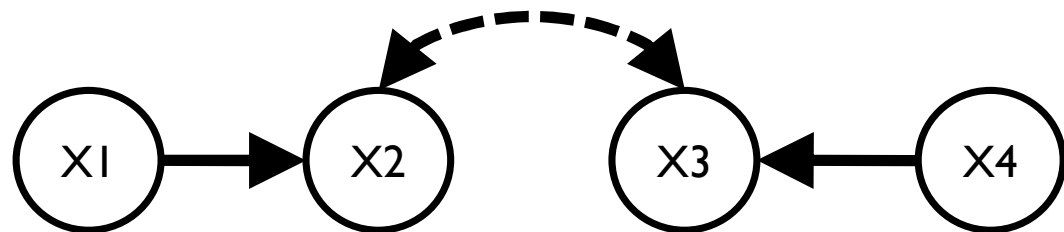
---

- ▶ Less obvious: what should be done with, e.g., additive error functional constraints

**Dataset I:  $X_2 = f(X_1, X_5) + \text{error}$**



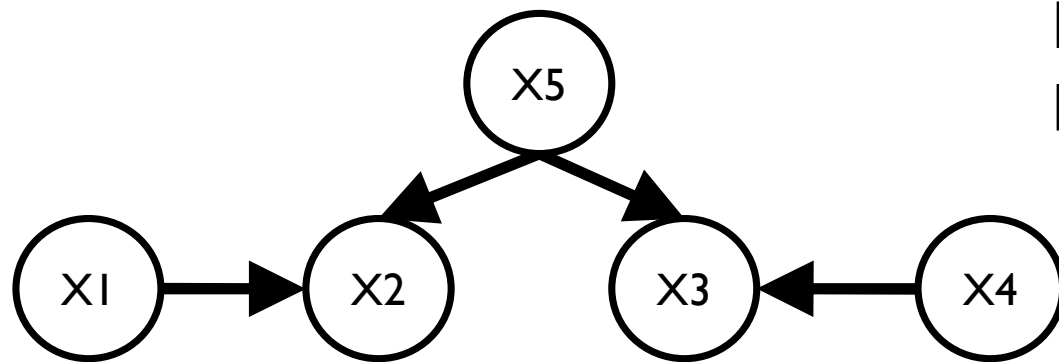
**Dataset II:  $X_2 = g(X_1) + \text{error}$ ?**



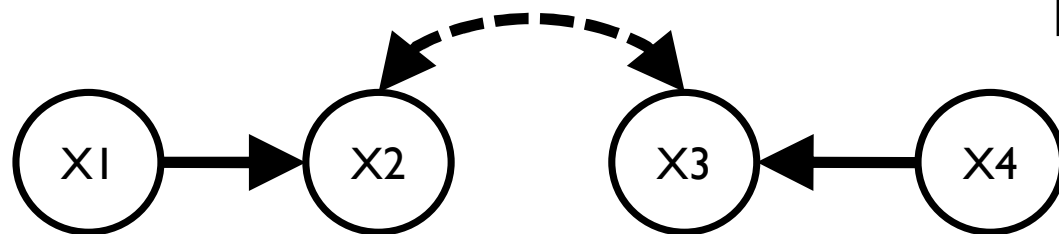
# The Bayesian approach

---

- ▶ Latent variable models might not be necessary, neither is the iid assumption.



First likelihood function:  
 $L(\theta_1 \mid \text{Dataset I})$



Second likelihood function:  
 $L(\theta_2 \mid \text{Dataset II})$

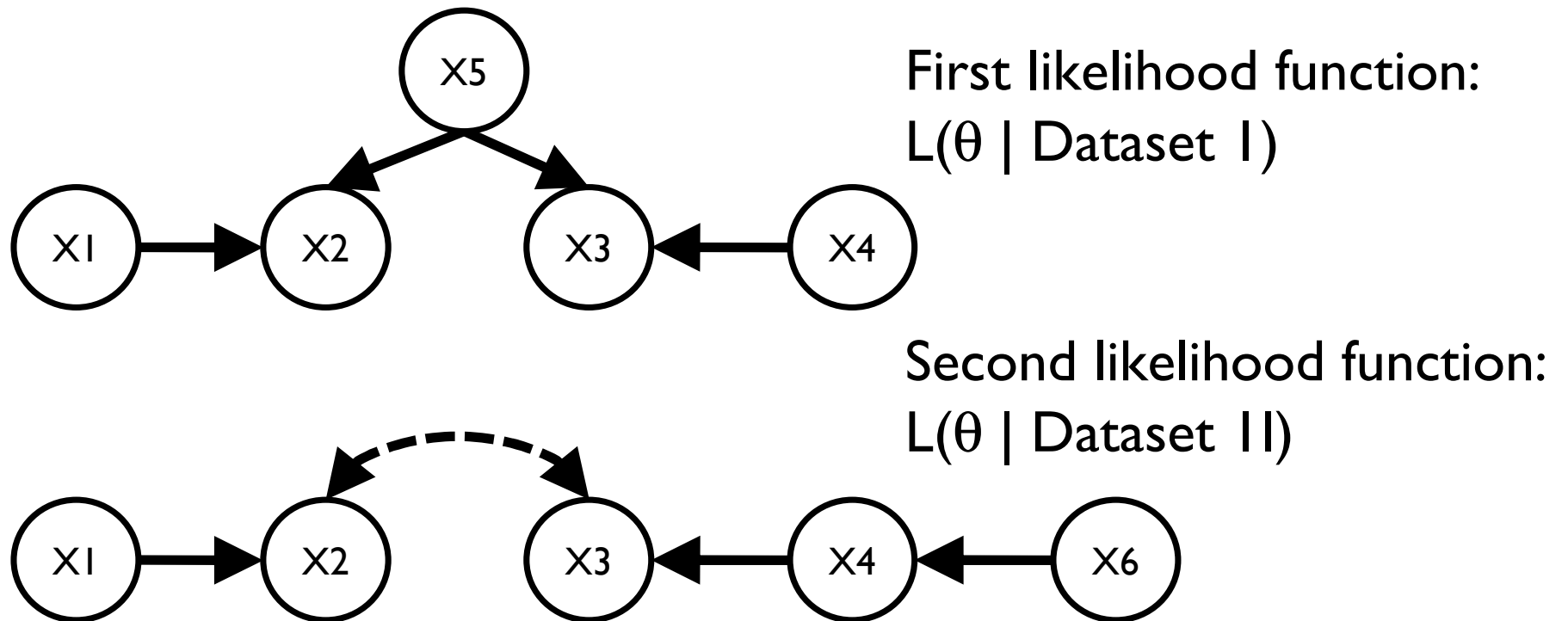




# Related problems: finding substructure by generalizing penalized composite likelihood?

---

- ▶ With/without same parameters, a single score function



- ▶  $S(\theta \mid \text{Data I, Data 2}) = \alpha_1 L(\theta_1 \mid \text{Data I}) + \alpha_2 L(\theta_2 \mid \text{Data 2}) + \Omega$
- 



# Other approaches: generalizing penalized composite likelihood?

---

- ▶ Enforcing constraints: link to constrained optimization

Maximize  $S(\theta \mid \text{Data 1}, \text{Data 2}) = \alpha_1 L(\theta_1 \mid \text{Data 1}) + \alpha_2 L(\theta_2 \mid \text{Data 2}) + \Omega$   
Subject to:

$$IM1_{(X1, X2, X3, X4)} = IM2_{(X1, X2, X3, X4)}$$

for the corresponding independent models IM on each subset

