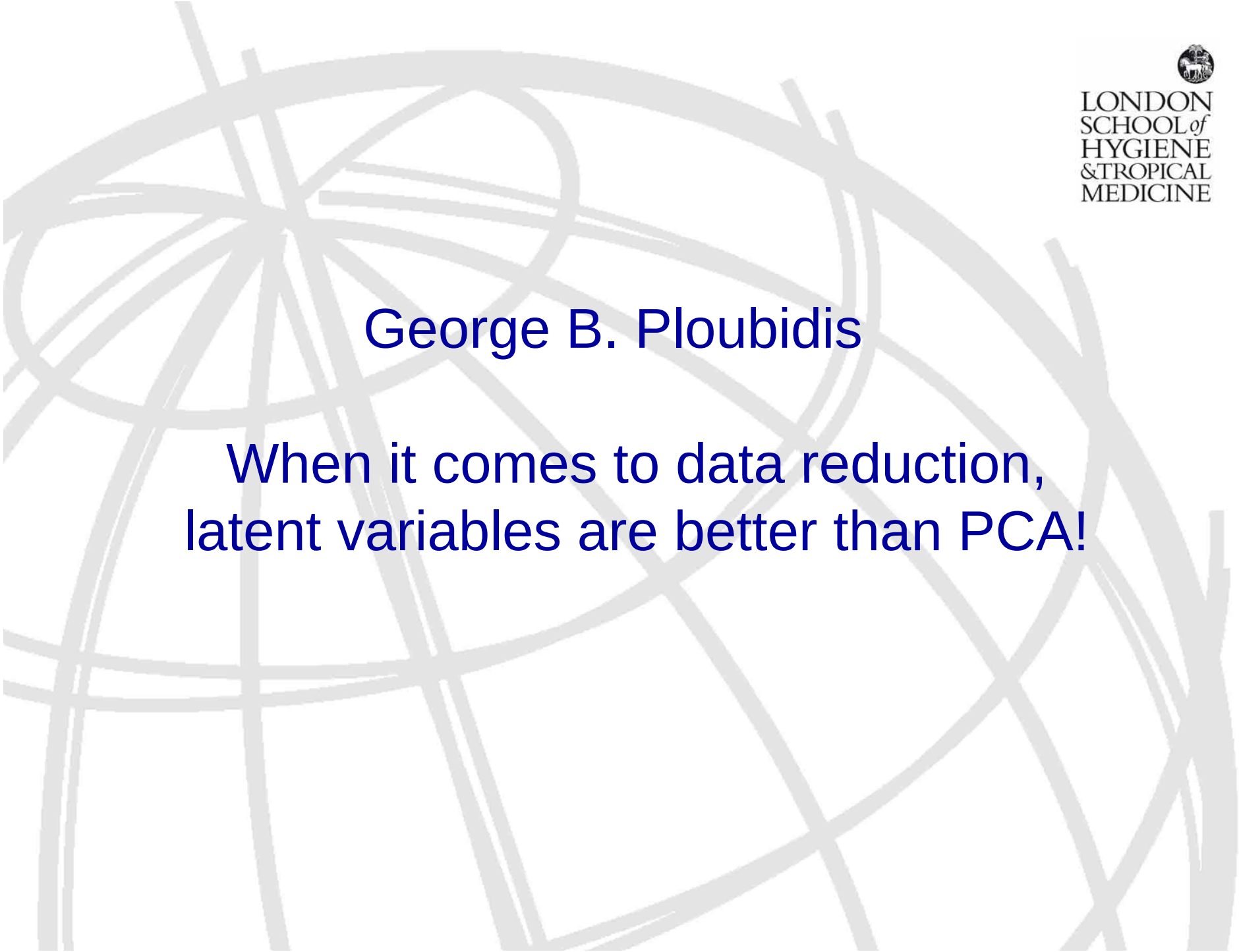


LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

George B. Ploubidis

When it comes to data reduction, latent variables are better than PCA!

Overview

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

- A little bit on data reduction
- Latent variable measurement models
- Latent variables VS Traditional methods
- Some other things you may want to do with latent variables

Data reduction

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

- A useful “first step”
- We have too many observations
  - need to “reduce” them to a smaller set of variable(s)
- The goal therefore is to reduce a large number of variables to meaningful summaries that can be used in further analyses

Traditional data reduction

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

- Principal Components Analysis (PCA)
  - Mathematically determined empirical solution
  - Extracts the continuous component(s) underlying a set of variables
- Cluster Analysis
  - Creates groupings (subsets) based on distance (similarity) between observations

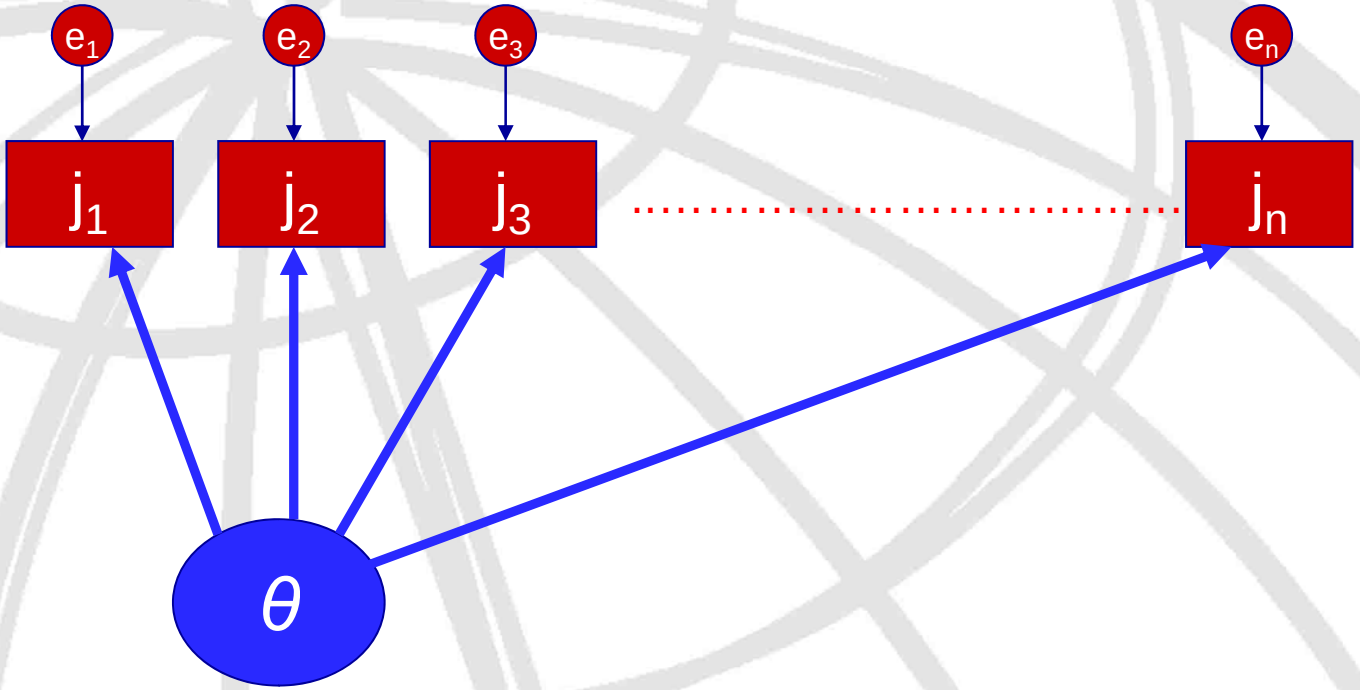
But, what is a latent variable model?

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

By latent variable model we mean any model that includes unobserved random variables which alternatively can be thought of as random parameters

Data reduction with latent variable measurement models

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE



The latent variable θ can be :

- Continuous, reducing the data to a dimension(s)
- Discrete, creating a typology – grouping – of individuals

But why LVMs are better than traditional methods?

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

	LVM	PCA
Correct for measurement error	Yes	No
Categorical/Ordinal indicators	Yes	Yes
Combinations of categorical/ordinal/continuous indicators	Yes	No
Measurement properties of the indicators	Yes	No
Fisher information	Yes	No
Higher order structures	Yes	No
	LVM	Cluster Analysis
Correct for measurement error	Yes	No
Model based criteria for the number of groups	Yes	No
Criteria for the classification quality	Yes	No
Measurement properties of the indicators	Yes	No
Higher order structures	Yes	No

LVM measurement models have been used to reduce data and derive indices or typologies for:

- Health status
- Alcohol use
- Smoking
- Wealth
- Mental health outcomes
- Frailty
- Grip strength
- Lung function
- Family functioning/environment

And there's more!

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

- The properties of the latent variable measurement models have led researchers to develop new applications, beyond data reduction

- Data reduction methods can be jointly modelled

Beyond data reduction

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

- Modeling longitudinal data
- MAR data – selection models
- Flexible time to event models (survival analysis etc)
- Controlling for unmeasured confounding

Software

- Mplus (the best)
- LISREL (the oldest, still good)
- EQS
- AMOS
- MX (freeware)
- R
- Statistica (path analysis only)
- WinBugs
- SAS, Stata, PASW, Splus

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

A paper to start with:

- **Rabe-Hesketh, S.** and Skrondal, A. (2008). Classical latent variable models for medical research. *Statistical Methods in Medical Research* **17**, 5-32.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Thank you for your attention