An Introduction to SEM

DATA 695 Research Capstone Project

David Errington
Master of Data Science & Analytics Student
Faculty of Graduate Studies

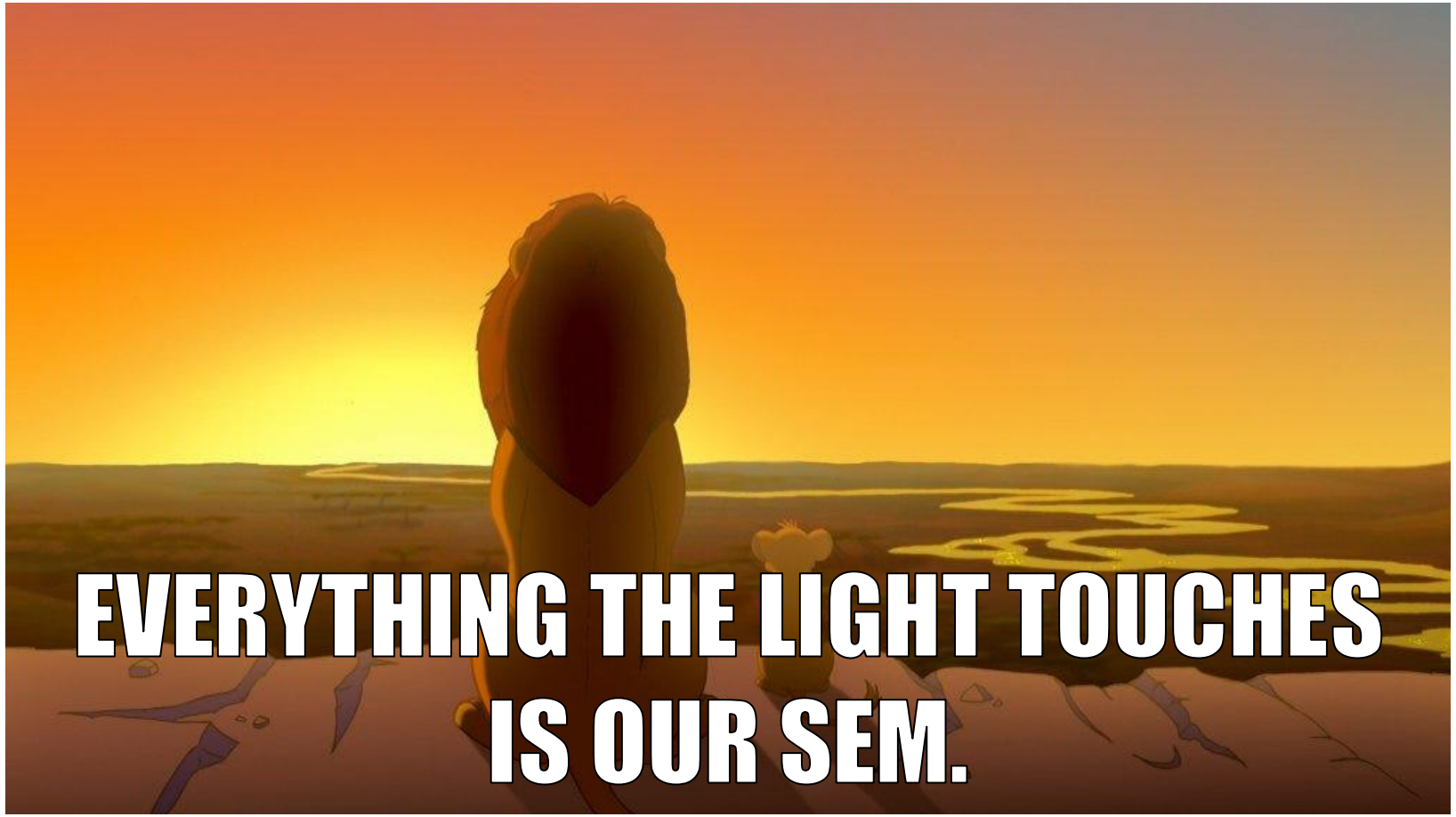
June 17, 2025

Table of Contents

The University of Calgary, located in the heart of Southern Alberta, both acknowledges and pays tribute to the traditional territories of the peoples of Treaty 7, which include the Blackfoot Confederacy (comprised of the Siksika, the Piikani, and the Kainai First Nations), the Tsuut'ina First Nation, and the Stoney Nakoda (including Chiniki, Bearspaw, and Goodstoney First Nations). The City of Calgary is also home to the Métis Nation of Alberta (Districts 5 and 6).

Chapter 1: Introduction, Background, & Review

What is Structural Equation Modelling?



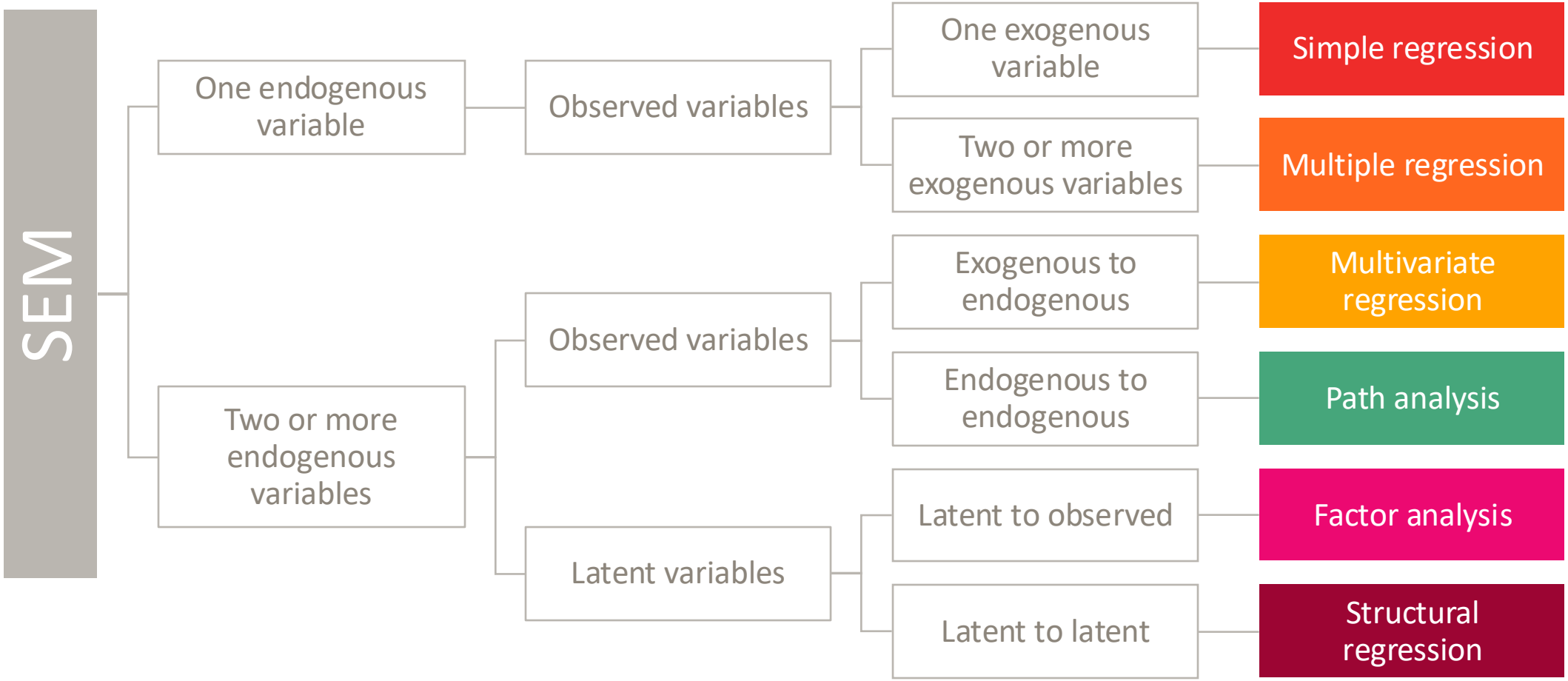
UNIVERSITY OF CALGARY

What is Structural Equation Modelling?

"…a linear model framework that models both *simultaneous regression equations* with *latent variables*."

— Johnny Lin, PhD (2024)

What is Structural Equation Modelling?



UNIVERSITY OF CALGARY

Terminology: Exogenous vs. Endogenous Variables

- *Exogenous variables*:
  - Variables that are not expressed as a function of other variables; they exist "outside" the system of variables under study.
  - Often referred to as "*independent*" variables (denoted as x or x_i).

- *Endogenous variables*:
  - Variables that are expressed as a function of one or more other variables; they exist "*inside*" the system of variables under study.
  - Often referred to as "*dependent*" variables (denoted as y or y_i).

Terminology: Observed vs. Latent Variables

- *Observed variable(s)*:
  - Variables that can be directly measured or "observed".
  - Examples: Height, weight, age, etc.

- *Latent variable(s)*:
  - Variables that (usually) cannot be directly measured; instead, they are often "inferred" (denoted as lowercase "eta" η or η_i)
  - Example: Intelligence.

Review of Linear Regression

- Given a data set $\{y_i, x_{i1}, \ldots, x_{ip}\}$ of n samples,

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

for $i = 1, \ldots, n$

where $p =$ number of "independent" variables/predictors.

Review of Linear Regression

- Given a data set $\{y_i, x_{i1}, \ldots, x_{ip}\}$ of n samples,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}$$

for $i = 1, \ldots, n$

where $p =$ number of "independent" variables/predictors.

Review of Linear Regression

- In R (or RStudio):

```
> model <- lm(y ~ x1 + ... + xp, data = ...)
```

Linear Regression as a Structural Equation Model

- Step 1: Start with simple linear regression equation.

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$$

Linear Regression as a Structural Equation Model

- Step 2: Take the intercept…

$$y_i = \boxed{\beta_0} + \beta_1 x_{i1} + \varepsilon_i$$

Intercept

Linear Regression as a Structural Equation Model

- Step 2: And replace it with “alpha”.

$$y_i = \boxed{\alpha} + \beta_1 x_{i1} + \varepsilon_i$$

Intercept

Linear Regression as a Structural Equation Model

- Step 3: Take the regression coefficient(s)...

$$y_i = \alpha + \beta_1 x_{i1} + \varepsilon_i$$

Intercept

Coefficient

Linear Regression as a Structural Equation Model

- Step 3: And replace it(them) with “gamma”.

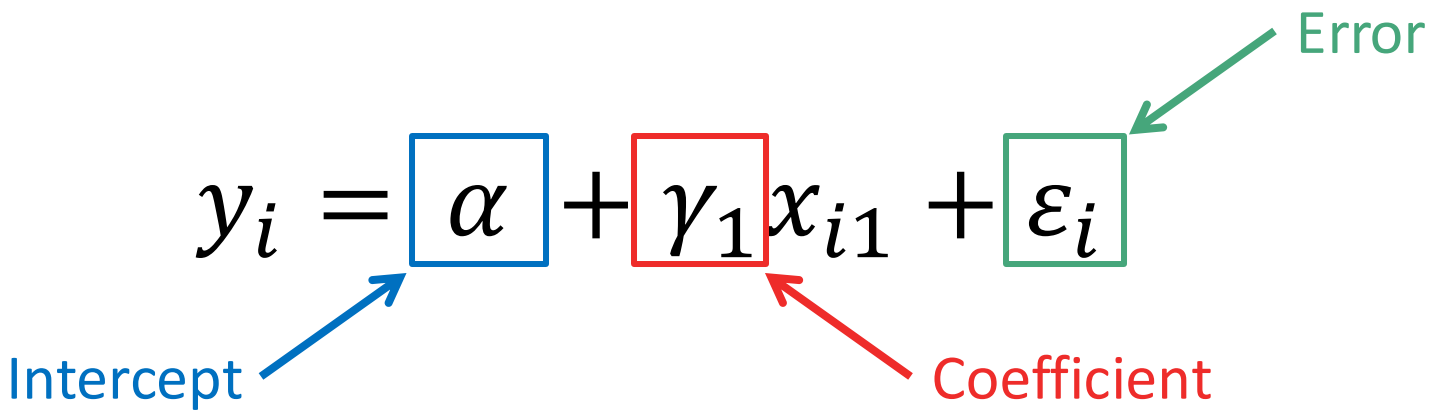
$$y_i = \alpha + \gamma_1 x_{i1} + \varepsilon_i$$

Intercept

Coefficient

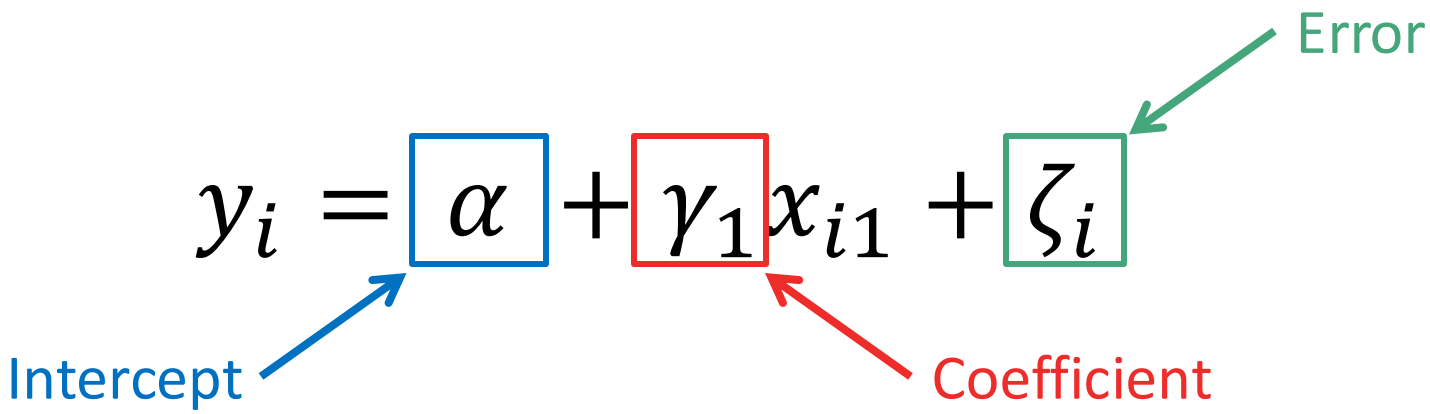
Linear Regression as a Structural Equation Model

- Step 5: Take the residual error term...



Linear Regression as a Structural Equation Model

- Step 5: And replace it with “zeta”.



Example:

References

- Lin, J. (2024). *Introduction to Structural Equation Modeling (SEM) in R with lavaan.* OARC Stats – Statistical Consulting Web Resources | UCLA. https://stats.oarc.ucla.edu/r/seminars/rsem/