

Modelli ad Equazioni Strutturali

Master Avanzato in Economia e Politica Agraria

Giovanbattista Califano

Università degli Studi di Napoli Federico II, giovanbattista.califano@unina.it

Scale Attitudinali per lo Studio
delle Preferenze del Consumatore
29 Maggio 2025

Tabella di marcia

15/05	Hello, Psychometrics!	✓
19/05	Affidabilità e Validità di una Misura	✓
22/05	Variabili Latenti: Riflessive o Formative?	✓
29/05	Modelli ad Equazioni Strutturali	○

Riferimenti consigliati:

- ▶ Capitoli 4 e 7 da Jhangiani et al. (2019)
- ▶ Capitoli 7, 10 e 14 da Olivero and Russo (2022)
- ▶ Capitolo 12 da Mehmetoglu and Jakobsen (2022)

Jhangiani, R. S., Chiang, I. A., Cuttler, C., and Leighton, D. C. (2019). *Research Methods in Psychology*. Kwantlen Polytechnic University, 4th edition.

Mehmetoglu, M. and Jakobsen, T. G. (2022). *Applied Statistics Using Stata: A Guide for the Social Sciences*. SAGE Publications Ltd, 2nd edition.

Olivero, N. and Russo, V. (2022). *Psicologia dei Consumi*. McGraw-Hill Education, 3rd edition.

A ciascuno di voi è stato assegnato un articolo scientifico. L'obiettivo dell'esame è analizzare criticamente l'uso degli strumenti psicometrici all'interno dello studio e proporre un'estensione motivata, fondata sulle vostre conoscenze teoriche e metodologiche. Il report va consegnato entro il **27/06/2025**.

Info e materiale:

<https://data.califano.xyz/esame>

Prima di iniziare...

<https://www.stata.com/manuals/sem.pdf>

Prima di iniziare...

Remarks and examples

Structural equation modeling encompasses a broad array of models from linear regression to measurement models to simultaneous equations, including along the way confirmatory factor analysis (CFA), correlated uniqueness models, latent growth models, multiple indicators and multiple causes (MIMIC) models, and item response theory (IRT) models.

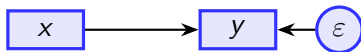
Structural equation modeling is not just an estimation method for a particular model in the way that Stata's `regress` and `probit` commands are, or even in the way that `stcox` and `mixed` are. Structural equation modeling is a way of thinking, a way of writing, and a way of estimating.

Prima di iniziare...

Remarks and examples

Structural equation modeling encompasses a broad array of models from linear regression to measurement models to simultaneous equations, including along the way confirmatory factor analysis (CFA), correlated uniqueness models, latent growth models, multiple indicators and multiple causes (MIMIC) models, and item response theory (IRT) models.

Structural equation modeling is not just an estimation method for a particular model in the way that Stata's `regress` and `probit` commands are, or even in the way that `stcox` and `mixed` are. Structural equation modeling is a way of thinking, a way of writing, and a way of estimating.

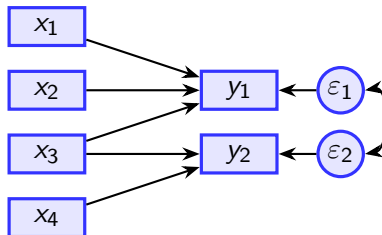


Prima di iniziare...

Remarks and examples

Structural equation modeling encompasses a broad array of models from linear regression to measurement models to simultaneous equations, including along the way confirmatory factor analysis (CFA), correlated uniqueness models, latent growth models, multiple indicators and multiple causes (MIMIC) models, and item response theory (IRT) models.

Structural equation modeling is not just an estimation method for a particular model in the way that Stata's `regress` and `probit` commands are, or even in the way that `stcox` and `mixed` are. Structural equation modeling is a way of thinking, a way of writing, and a way of estimating.



Passiamo a...

Un po' di terminologia

Un po' di storia

Un po' di teoria

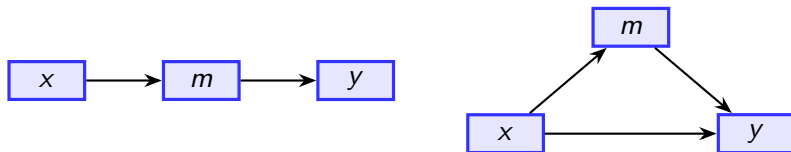
Un po' di Stata

Tipi di SEM

- ▶ La modellistica a equazioni strutturali permette di stimare le relazioni tra variabili osservate e/o latenti attraverso lo studio della matrice di varianza e covarianza. Per questo, è anche chiamata **Covariance Based (CB) SEM**¹.
- ▶ Poiché viene utilizzata per stimare più equazioni simultaneamente, i modelli sono descritti e riportati in forma di diagrammi a blocchi.
- ▶ Come notava anche il buon manuale di Stata, sotto il cappello della SEM rientrano molti modelli diversi...

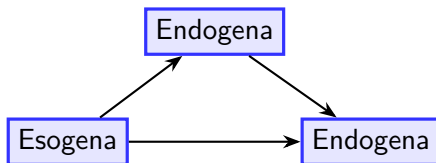
¹Anche per distinguerla dalla Partial Least Squares (PLS) SEM. 

Path Analysis (PA)



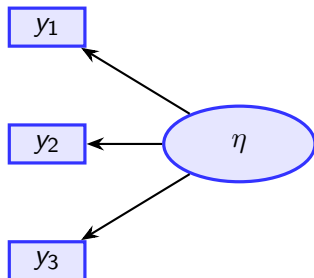
Si stimano relazioni tra variabili manifeste.

Path Analysis (PA)



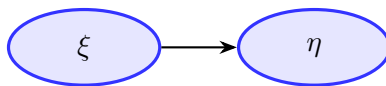
In questi modelli è complicato parlare di variabili dipendenti e indipendenti. Le variabili che non sono spiegate nel modello si definiscono esogene, mentre quelle che vengono spiegate si definiscono endogene.

Confirmatory Factor Analysis (CFA)



Si stimano relazioni tra variabili manifeste e latenti.

Latent Path Analysis (LPA)

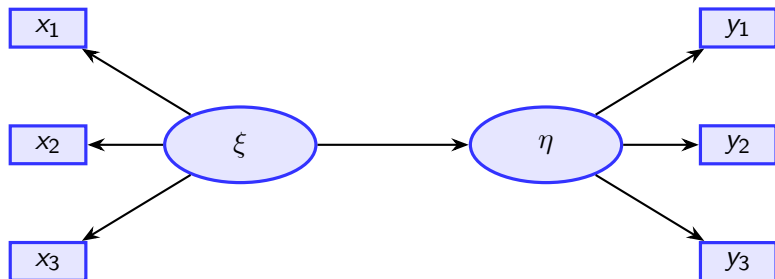


Si stimano relazioni tra variabili latenti.

Qualche esempio

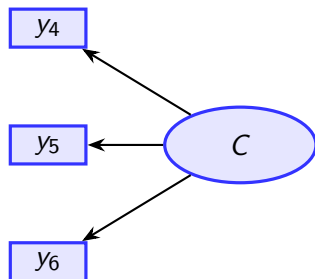
CFA + LPA

Ciò che si intende comunemente con SEM



*In questi casi la parte di CFA viene definita come **modello di misurazione** o **modello esterno** (outer). La parte di LPA (o PA), invece, come **modello strutturale** o **modello interno** (inner).*

Le variabili latenti sono solo continue?



No. Possiamo ipotizzare l'esistenza di classi non osservabili (o non osservate) e stimarle a partire da indicatori osservati. SEM a variabili latenti categoriali includono Latent Class Analysis, Latent Profile Analysis, e altri modelli di misture finite.

Passiamo a...

Un po' di terminologia

Un po' di storia

Un po' di teoria

Un po' di Stata

Breve storia della SEM

Origini

L'analisi SEM nasce dall'integrazione di:

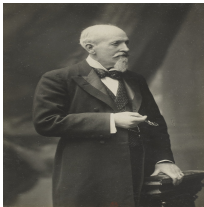
- ▶ **Path Analysis** (Sewall Wright, 1921)
- ▶ **Analisi Fattoriale** (Spearman, 1904; Thurstone, 1930)
- ▶ **Modelli a variabili latenti** (Jöreskog, 1969)

La SEM unifica approcci di tipo causale e misurativo in un'unica struttura statistica.

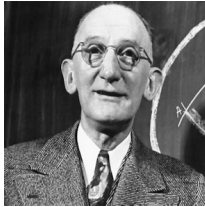
Sewall Wright



Charles Spearman



Louis Thurstone



Karl Jöreskog



Breve storia della SEM

CB-SEM: Covariance-Based SEM

- ▶ Formalizzato da Karl Jöreskog (1969) con LISREL;
- ▶ Obiettivo: riprodurre la **matrice di covarianze** osservata;
- ▶ Basato sulla **massima verosimiglianza**.

PLS-SEM: Partial Least Squares SEM

- ▶ Introdotto da Herman Wold (anni '70);
- ▶ Approccio **component-based**, iterativo;
- ▶ Focalizzato sulla **massimizzazione della varianza spiegata**;
- ▶ Preferibile in caso di: piccoli campioni, modelli formativi, non-normalità.

Passiamo a...

Un po' di terminologia

Un po' di storia

Un po' di teoria

Un po' di Stata

Processo di analisi

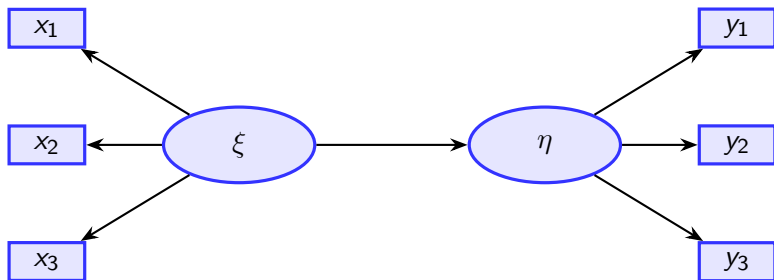
L'analisi SEM passa attraverso quattro fasi, uguali per **CB-SEM** e **PLS-SEM**. Tuttavia, i criteri specifici da utilizzare differiscono, in quanto questi modelli differiscono per funzionamento e per scopi di utilizzo (es. VL Riflessive o Formative). Ci soffermeremo sulle quattro fasi del processo con un focus sul **CB-SEM**:

1. Specifica del modello
2. Identificazione del modello
3. Stima del modello
4. Valutazione del modello

Specifica del modello

Fase 1:

La prima fase consiste nel definire il modello teorico e le relazioni ipotizzate. In molti software (compreso Stata) la specifica del modello può avvenire sia attraverso il disegno di diagrammi, sia attraverso la scrittura di codice.



```
. sem (Xi -> x1 x2 x3) (Eta -> y1 y2 y3) (Xi -> Eta)
```

Identificazione del modello

Fase 2:

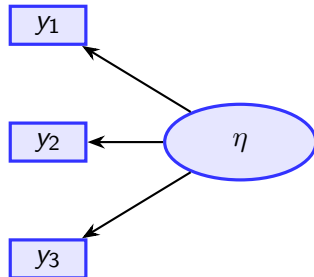
Purtroppo, non tutto quello che si può disegnare si può anche stimare. La **T-Rule**, ad esempio, è una condizione necessaria (ma non sufficiente) affinché un modello sia identificabile:

$$t \leq \frac{p(p+1)}{2}$$

- ▶ t è il numero di parametri da stimare (freccie nei diagrammi);
- ▶ p è il numero di variabili manifeste del modello.
- ▶ Se $t < \frac{p(p+1)}{2}$ il modello è *over-identified*: stimabile e testabile;
- ▶ Se $t = \frac{p(p+1)}{2}$ il modello è *just-identified*: stimabile ma non testabile (gradi di libertà = 0);
- ▶ Se $t > \frac{p(p+1)}{2}$ il modello è *under-identified*: non stimabile in modo univoco.

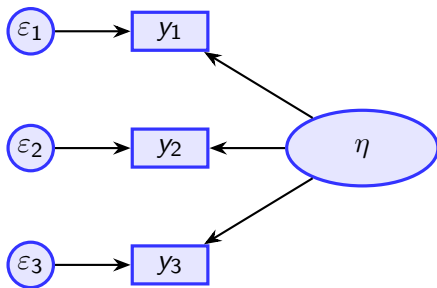
Identificazione del modello

$$t \leq \frac{p(p+1)}{2} \quad ?$$



Identificazione del modello

$$t \leq \frac{p(p+1)}{2} \quad ?$$



Fase 3:

La stima del modello consiste nel calcolare i valori dei parametri che meglio riproducono la matrice di covarianze osservata (tipicamente attraverso *massima verosimiglianza*).

Valutazione del modello

Fase 4:

La valutazione del modello avviene su tre livelli:

1. **Bontà di adattamento globale:**

- ▶ χ^2 test (meglio non significativo)
- ▶ $RMSEA < 0.06$
- ▶ $CFI, TLI > 0.95$
- ▶ $SRMR < 0.08$

2. **Modello di misurazione:**

- ▶ λ_i (carichi fattoriali) > 0.5 e significativi
- ▶ CR (Composite Reliability) > 0.7
- ▶ AVE (Average Variance Extracted) > 0.5
- ▶ Discriminant validity (es. criterio di Fornell-Larcker)

3. **Modello strutturale:**

- ▶ β (path coeff.) significativi
- ▶ R^2 accettabili per le endogene
- ▶ Effetti diretti, indiretti e totali significativi (se rilevante)

Cosa sono?

I **Modification Indices** (MI) segnalano potenziali miglioramenti al modello introducendo parametri non specificati (es. covarianze residue).

Attenzione: le modifiche suggerite vanno accettate solo se teoricamente giustificate, per evitare *overfitting*.

Introducendo nuovi parametri da stimare, ripartiamo dalla Fase 1!

Passiamo a...

Un po' di terminologia

Un po' di storia

Un po' di teoria

Un po' di Stata

clear all ;)

```
. webuse set data.califano.xyz  
. webuse mamme
```