Departamento de Matemática
Universidad Técnica Federico Santa María

Una crítica a los modelos de panel cruzados para la estimación del efecto recíproco entre felicidad y desempeño laboral

Francisco Alfaro Medina
Orientador: Marcos Gómez, Wenceslao Unanue (UAI)
Co-Orientador: Ronny Vallejos (UTFSM)

04 de Octubre, 2018

Contenidos

Motivación

- Problemas con el indicador PIB para mostrar el progreso de un pais.

Motivación

- Problemas con el indicador PIB para mostrar el progreso de un pais.
- Buscar nuevas alternativas de indicadores sociales y económicos.

Motivación

- Problemas con el indicador PIB para mostrar el progreso de un pais.
- Buscar nuevas alternativas de indicadores sociales y económicos.
- Tipo de relación entre la felicidad y el desempeño laboral.
    1. Tipo de relación causal
    2. Manifestaciones a lo largo del tiempo.

Objetivos

- **Objetivo general**: Establecer tipo de relación entre la felicidad y el desempeño laboral a lo largo del tiempo.

Objetivos

- **Objetivo general**: Establecer tipo de relación entre la felicidad y el desempeño laboral a lo largo del tiempo.
- **Objetivos específicos:**
  1. Estudiar modelos SEM para datos longitudinales.
  2. Estudiar las ventajas y desventajas de los modelos de panel cruzados (CLPM).
  3. Alternativas para los modelos CLPM.
  4. Realizar comparaciones entre las distintas alternativas.

Preliminares - Análisis Factorial

En general, el modelo lo podemos escribir por

$$Y = \mu + \alpha\Phi + \epsilon \tag{1}$$

o de la forma

$$X = \alpha\Phi + \epsilon \tag{2}$$

donde $Y = (y_1, ..., y_p)^\top$ es un vector de variables aleatorias observadas, $\Phi = (\phi_1, ..., \phi_r)^\top$ es el vector de de $r < p$ variables aleatorias no observadas o variables latentes llamados factores, α es la matriz $(r \times p)$ matriz de coeficientes fijos (cargas), y $\epsilon = (\epsilon_1, ..., \epsilon_p)^\top$ es el vector de error aleatorio.

Preliminares - Análisis Factorial

Los siguientes supuestos son usualmente hecho en los modelos factoriales (2).

$$
\begin{aligned}
&\text{(i)}\ \rho(\alpha) = r < p \\
&\text{(ii)}\ \mathbb{E}(X|\Phi) = \alpha\Phi \\
&\text{(iii)}\ \mathbb{E}(XX^\top) = \Sigma, \mathbb{E}(\Phi\Phi^\top) = \Omega, \mathrm{y}\mathbb{E}(\epsilon\epsilon^\top) = \Psi = \mathrm{diag}(\sigma_1^2, ..., \sigma_p^2) \\
&\text{(iv)}\ \mathbb{E}(\Phi\epsilon^\top) = 0
\end{aligned}
\tag{3}
$$

Sea Ω la matriz de correlación de los factores. El número total de parámetros estimados en $\alpha.\Omega$ y Ψ es

$$(pr + \frac{1}{2}r)(r+1) + p = \frac{1}{2}(2p + r)(r+1) \tag{4}$$

Sea n_α,n_Ω y n_Ψ el número de parámetros en α, Ω y Ψ respectivamente, donde $m = n_\alpha + n_\Omega + n_\Psi$. Entonces el número de parámetros no restringidos es

$$\frac{1}{2}(2p + r)(r+1) - m \tag{5}$$

el cual debe ser un número más pequeño que $p(p+1)/2$, el número de elementos distintos de Σ, para reproducir una solución no trivial. Así

$$r^2 + m > \frac{1}{2}(p+r)(p+r+1) \tag{6}$$

y una condición necesaria para los parámetros de los factores es determinado unicamente por

$$n_\alpha + n_\Omega \geq r^2 \tag{7}$$

Preliminares - Modelos Ecuaciones Estructurales

- El modelo de ecuaciones estructurales (Structural Equation Modeling, SEM) es una técnica que combina tanto la regresión múltiple como el análisis factorial.
- Permite al investigador no solo evaluar las muy complejas interrelaciones de dependencia sino también incorporar los efectos del error de medida sobre los coeficientes estructurales al mismo tiempo.
- Se separa en tres etapas: Tipos de Modelos, Análisis de Rutas, Relación de Causalidad.

En general, el modelo de variable latente lo podemos escribir como

$$\eta = B\eta + \Gamma\xi + \zeta \tag{8}$$

donde $\eta = (\eta_1, ..., \eta_m)^\top$ es un vector de variables endógenas latentes , $\xi = (\xi_1, ..., \xi_n)^\top$ es el vector de variables exógenas latentes, B es la matriz $(m \times m)$ de coeficientes que relacionan las variables latentes endógenas entre sí, Γ es la matriz $(m \times n)$ de coeficientes que relacionan las variables latentes exógenas con las endógenas y $\zeta = (\zeta_1, ..., \zeta_m)^\top$ es el vector de error aleatorio.

Por conveniencia, suponemos que

$$\begin{aligned} &\text{(i) } \mathbb{E}(\eta) = \mathbb{E}(\xi) = \mathbb{E}(\zeta) = 0 \\ &\text{(ii) } \eta \text{ no esta correlacionado con } \xi \\ &\text{(iii) } (I - B) \text{ es una mmatriz no singular} \end{aligned} \tag{9}$$

En general, el modelo de medición lo podemos escribir como

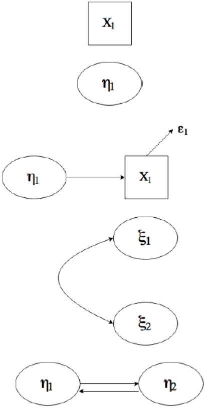
$$\begin{aligned} Y &= \Lambda_y \eta + \epsilon \\ X &= \Lambda_x \xi + \delta \end{aligned} \tag{10}$$

donde $Y = (y_1, ..., y_p)^\top$ es un vector de variables observadas que actúan como indicador de la variable η, $X = (x_1, ..., x_q)^\top$ es un vector de variables observadas que actúan como indicador de la variable ξ, Λ_y es la matriz $(p \times m)$ de coeficientes que relaciona Y con η, Λ_y es la matriz $(q \times n)$ de coeficientes que relaciona Y con η, $\epsilon = (\epsilon_1, ..., \epsilon_p)^\top$ es la medida de error para la variable Y y $\delta = (\delta_1, ..., \delta_q)^\top$ es la medida de error para la variable X.

Por conveniencia, suponemos que

$$\begin{aligned}(\text{i})\ & \mathbb{E}(\eta) = \mathbb{E}(\xi) = \mathbb{E}(\epsilon) = \mathbb{E}(\delta) = 0 \\ (\text{ii})\ & \epsilon \text{ no esta correlacionado con } \eta, \xi, \delta \\ (\text{iii})\ & \delta \text{ no esta correlacionado con } \xi, \eta, \epsilon \end{aligned} \tag{11}$$

Figura: Notación utilizada en el análisis de rutas.



una caja rectangular o cuadrada corresponde a una variable observada

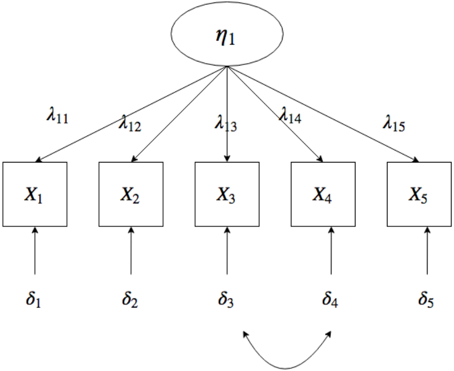
un circulo o elipse corresponde a una variable latente

una variable no cerrada significa un término de perturbación (error en la ecuación o medida)
una flecha recta significa la suposición de que la variable en la base de la flecha causa una variable en la punta de la flecha

una flecha curva de dos cabezas significa una asociación no analizada entre dos variables

dos flechas rectas de una sola cabeza que conectan dos variables significa una relación de retroalimentación o causalidad recíproca

Figura: Diagrama de rutas para una variable latente con cuatro indicadores.



La descomposición de la $COV(x_1, x_4)$ es

$$\begin{aligned} COV(x_1, x_5) &= COV(\lambda_{11}\eta_1 + \delta_1, \lambda_{51}\eta_1 + \delta_5) \\ &= \lambda_{11}\lambda_{51}\phi_{11} \end{aligned}$$

el lado derecho de la ecuación anterior, indica que la $COV(x_1, x_5)$ se encuentra en función de los efectos de η_1 sobre x_1 y x_5 (i.e. λ_{11} y λ_{51}) y la varianza de la variable latente η_1.

El análisis de ruta distingue tres tipos de efectos: efectos directos, indirectos y totales. El efecto directo es la influencia de una variable sobre otra que no está mediada por ninguna otra variable en un modelo de ruta. Los efectos indirectos de una variable están mediados por al menos una variable intermedia. La suma de los efectos directos e indirectos son los efectos totales:

$$\text{efectos totales} = \text{efecto directo} + \text{efectos indirectos}$$

Estudiemos el efecto total de la variable ξ_1 sobre η_2:

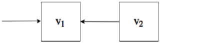
$$\begin{aligned} \text{efecto total} &= \text{efecto directo} + \text{efecto indireco} \\ &= \gamma_{21} + \gamma_{11}\beta_{21} \end{aligned}$$

El efecto total de ξ_1 sobre y_6 es:

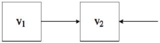
$$\begin{aligned} \text{efecto total} &= \text{efecto directo} + \text{efecto indireco} \\ &= 0 + (\gamma_{21}\lambda_8 + \gamma_{11}\beta_{21}\lambda_8) \end{aligned}$$

en este caso, el efecto total de ξ_1 sobre y_6 se compone de efectos indirectos.

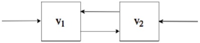
Figura: Relaciones causales entre variables. Formas de covariación.



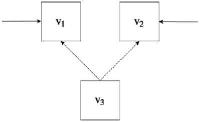
(a) Figura 3.a



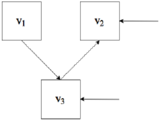
(b) Figura 3.b

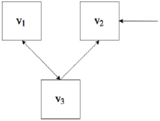


(c) Figura 3.c



(d) Figura 3.d



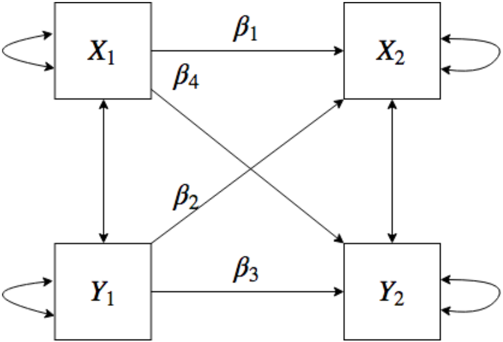


El modelo más simple de panel cruzado incluye dos constructos medidos en dos puntos de tiempo, donde cada constructo medido corresponderá a una variable. La Figura 4 muestra un diagrama de ruta para un modelo de panel de dos olas sobre las variables X e Y. El modelo se lo podemos describir mediante el siguiente sistema de ecuaciones

$$\begin{aligned} X_2 &= \beta_1 X_1 + \beta_2 Y_1 + \epsilon_x \\ Y_2 &= \beta_3 Y_1 + \beta_4 X_1 + \epsilon_y \end{aligned} \tag{12}$$

donde X_i, Y_i son las variables medidas en el tiempo i (con $i = 1, 2$), β_1 y β_3 son los efectos autorregresivos y β_2 y β_4 son los efectos cruzados.

Figura: Modelo CLMP de dos olas para las variables X e Y.



En el año 1980, David Rogosa público el artículo denominado " Una crítica de la correlación cruzada". Tal publicación dejo en claro que las correlaciones cruzadas entre los datos del panel longitudinal es una base inadecuada para hacer inferencias causales. Desde entonces, los investigadores se han respaldado en los modelos de panel cruzado (CLPM) para estudiar causalidad, sin embargo, este tipo de modelos presenta varias dificultades, tanto en el muestreo realizado como en los supuestos que se deben cumplir para que los modelos CLPM sean interpretables.

a) **Sincronicidad**
b) **Confiabilidad de la medición**
c) **Selectividad de la muestra**
d) **Desgaste selectivo**
e) **Efectos de reprueba**
f) **Efectos del tiempo**

a) **Teoría del cambio**
b) **Estacionariedad**
c) **Comparación de coeficientes cruzados**
d) **Variables omitidas**

a) Modelo de trayectorias autorregresivas (ALT) y modelo RI-CLPM
b) Modelo de rasgo autorregresivo (STARTS)
c) Modelo del puntaje de cambio latente (LCT)
d) Modelo de estado latente modificado (LST)

Los modelos ALT, desarrollado por Curran y Bollen (2001; ver también Bollen & Curran, 2006), permiten que las personas se caractericen por su propia trayectoria en el tiempo, mientras que sus observaciones también puedan exhibir algún efecto de arrastre de un instante al siguiente. En la extensión bivariada , los efectos aleatorios que describen las trayectorias individuales pueden correlacionarse entre sí a través de las variables, y también puede haber ser influencias cruzadas entre las observaciones (como en el CLPM).

- El conjunto de datos consiste en un total de 1521 trabajadores para las distintas empresas en Chile, de los cuales 703 corresponden a trabajadores de nacionalidad chilena.
- A cada trabajador de se la hace un total de 15 preguntas relacionadas con la felicidad basadas en el sistema PERMA y otras 3 preguntas relacionadas con su trabajo en si.
- Cada pregunta tiene un sismte de puntación del 0 al 10.
- Este conjunto de preguntas fueron medidas en tres instantes de tiempos.

Figura: Box plot para las variables en estudio.

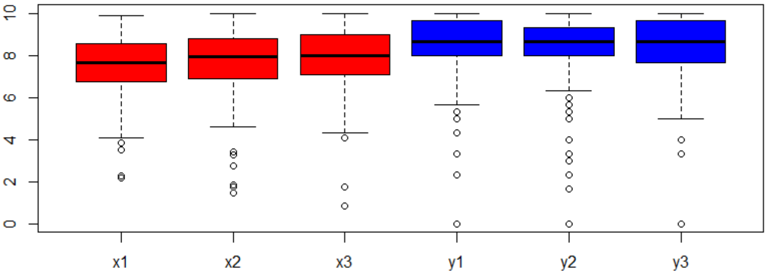


Figura: Esquema para el modelo CLPM.

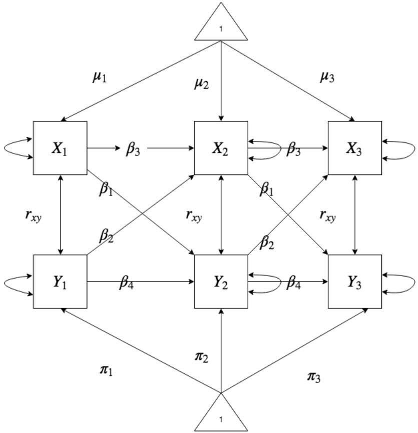
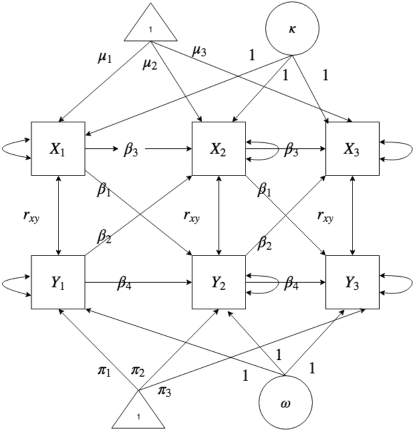


Figura: Esquema para el modelo RI-CLPM.



Cuadro: Test de diferencias chi cuadrado.

Ajuste	DF	AIC	BIC	Chisq	ChiDiff	DfDiff	Pr(>Chisq)
CLPM	1	8004	8122	0.48			
RI-CLPM	4	8044	8149	46.63	46.15	3	5.27e-10

Conclusiones

- Comprensión del estudio realizado, tanto del muestreo como la interacción entre las variables.
- Comprensión de las relaciones inter e inter clase.
- Alternativas a los modelos tradicionales CLPM.

Trabajos Futuros

- Buscar nuevos modelos que tengan como directrices las investigaciones realizadas.
- Dejar en claro los alcances del modelamiento matemático en las ciencias sociales.
- Realizar rutinas más sofisticadas para este tipo de problemas.

Referencias I

[1] Rogosa, D. (1980). *A Critique of Cross-Lagged Correlation.* Psychological Bulletin, Vol 88(2), Sep 1980, 245-258.

[2] Bollen, K & Curran P. (2004). *Autoregressive latent trajectory (ALT) models A synthesis of two traditions.* Sociological Methods Research, Vol 32, Issue 3, pp. 336 - 383.

[3] Selig, J. P., & Little, T. D. (2012). *Autoregressive and cross-lagged panel analysis for longitudinal data.* Handbook of developmental research methods (pp. 265-278). New York, NY, US: Guilford Press.

[4] Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). *A critique of the cross-lagged panel model.* Psychological Methods, 20(1), 102-116.

[5] Berry, D. & Willoughby, M. T. (2017). *On the Practical Interpretability of Cross-Lagged Panel Models: Rethinking a Developmental Workhorse.* Child Dev, 88: 1186-1206.

Referencias II

[6] Basilevsky , A. (1994). *Statistical Factor Analysis and Related Methods.* John Wiley & Sons.

[7] Bollen , K. (1989). *Structural Equations with Latent Variables.* John Wiley & Sons.

[8] Newsom , J. (2015). *Longitudinal Structural Equation Modeling.* Routledge.

Muchas Gracias.