

Análisis Factorial y Estimación PLS de modelos SEM

Modelos de Ecuaciones Estructurales con PLS (Partial Least Squares)

29 y 30 de noviembre de 2017

José Mondéjar Jiménez

**Área de Estadística Económica y Empresarial
Universidad de Castilla-La Mancha**

- Los modelos de ecuaciones estructurales pretenden modelar las relaciones estadísticas entre constructos latentes que no son directamente observables.
- En 1960, Karl Jöreskog diseñó un algoritmo de máxima verosimilitud para la estimación de modelos de estructuras de covarianzas, y en 1970, desarrolló el primer software de amplia difusión: LISREL.
- Su director de tesis, Herman Wold, criticó la dependencia de los supuestos distribucionales, que afecta a la validez de los resultados empíricos. Por ello, propuso un enfoque alternativo, Partial Least Squares (PLS), y en 1977 desarrolló el algoritmo NIPALS (Nonlinear Iterative Partial Least Squares).
- Ambos enfoques han seguido desarrollándose y, en la actualidad, son instrumentos cuantitativos complementarios de medición, evaluación y contraste ampliamente utilizados.

¿Por qué dos metodologías? Básicamente por los supuestos y los objetivos buscados:

A. La metodología SEM está basada en el supuesto de normalidad de las variables y utiliza una estimación máximo verosímil.

1. Ajusta las varianzas-covarianzas entre las variables observadas.
2. Necesita una muestra generalmente grande.
3. Está enfocada a la “reproducción” de la estructura de relaciones entre variables.

¿Por qué dos metodologías? Básicamente por los supuestos y los objetivos buscados:

B. La metodología PLS no asume la normalidad y estima por mínimos cuadrados de forma recursiva.

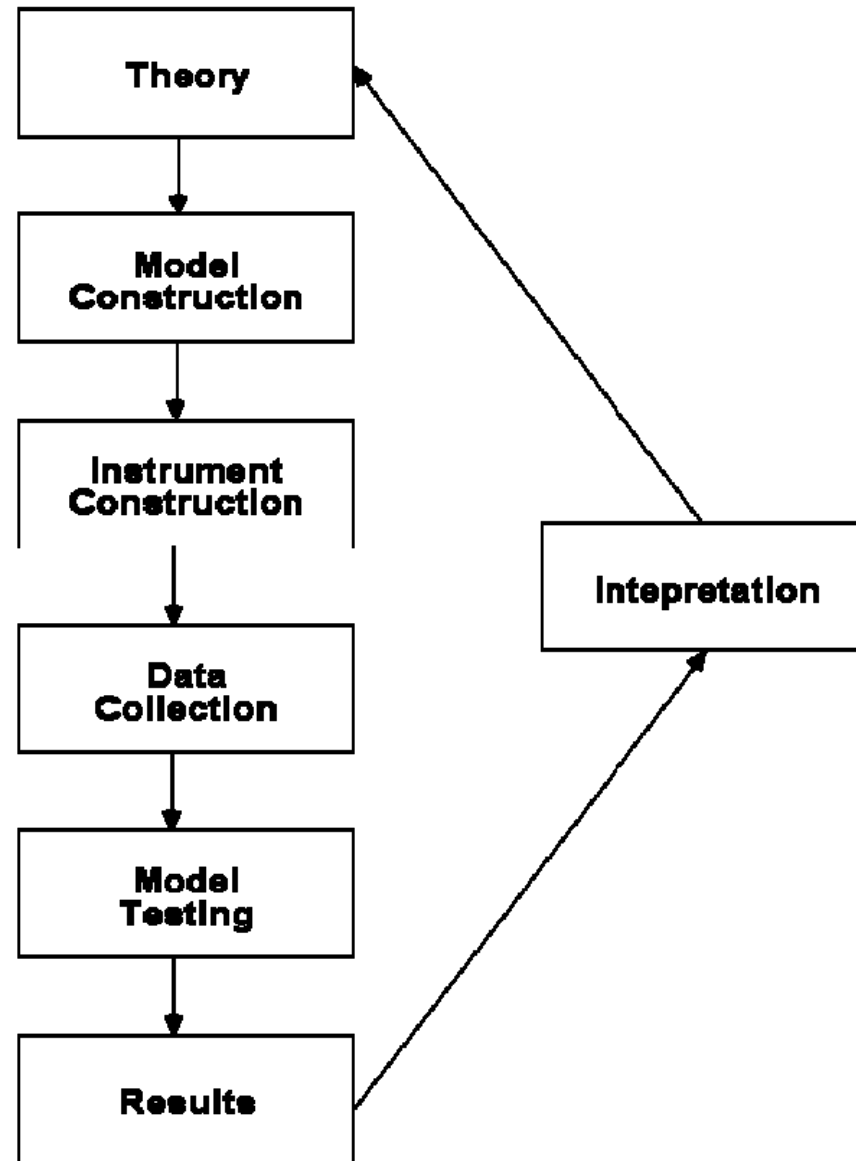
1. Utiliza un esquema de regresión con supuestos “débiles”. Válido con escalas numéricas, ordinales e incluso nominales (con precaución).
2. Aplicable incluso con muestras pequeñas.
3. Enfocada hacia la predicción, generando combinaciones lineales de indicadores que maximicen la covarianza.

1. **NIPALS (Nonlinear Iterative Partial Least Squares)**
2. **PLS Regression (Partial Least Squares Regression)**
3. **PLS Discriminant Analysis**
4. **SIMCA (Soft Independent Modeling by Class Analogy)**
5. **PLS Approach to Structural Equation Modeling**
6. **N-way PLS**
7. **PLS Logistic Regression**
8. **PLS Generalized Linear Model**

2

Lógica interna de la metodología PLS

DIAGRAMA DE PASOS:



2

Lógica interna de la metodología PLS

1. **Especificación del modelo teórico.** *Cuidado: Un modelo puede estar teóricamente bien especificado pero ser estadística o empíricamente no identificable por efecto de los datos.*
2. **Contrastar el ajuste del modelo de medida.**
3. **Re-especificar y re-estimar el modelo si los datos indican que es necesario.**
4. **Cuando el modelo de medida tenga un buen ajuste, contrastar el modelo estructural.**
5. **Re-especificar y reestimar el modelo estructural, si fuese necesario.**

1. La interpretación “clásica” de la regresión está referida a variables medidas en escalas continuas.
2. Cuando se utilizan escalas discretas (tipo Likert) se debe tener en cuenta la limitación de la varianza.
3. Este tipo de escalas se adaptan mejor a la metodología PLS que las escalas Guttman o las escalas Mokken, por lo que se ha generalizado su uso.
4. Cuando la escala es de pocos puntos, se debe cuidar especialmente la interpretación del estadístico alfa de Cronbach y la estructura de correlación.

2

Fiabilidad y validez del modelo de medida

FIABILIDAD: consistencia entre los diversos indicadores de cada variable latente, indicando unidimensionalidad de la misma. En escalas Likert, se suele utilizar la correlación entre el ítem y el total de la escala.

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_T^2} \right) = \frac{k\rho}{1 + (k-1)\rho}$$

FIABILIDAD COMPUESTA: no es individual para cada constructo, sino que considera el modelo en su conjunto. Sería recomendable analizarla después de comprobar las cargas cruzadas.

$$IFC_i = \frac{(\sum_j L_{ij})^2}{(\sum_j L_{ij})^2 + \sum_j (1 - L_{ij}^2)}$$

VARIANZA EXTRAÍDA MEDIA (AVE): Proporción de varianza del constructo “explicada” por los indicadores.

$$AVE_i = \frac{\sum_j L_{ij}^2}{\sum_j L_{ij}^2 + \sum_j (1 - L_{ij}^2)}$$

2

Fiabilidad y validez del modelo de medida

VALIDEZ: de contenido (teórica) y de construcción, tanto convergente (tamaño de las cargas L_{ij} y AVE, cargas cruzadas) y divergente:

$$\min(AVE_i, AVE_j) > \rho_{ij}^2$$

La fiabilidad está relacionada con la consistencia interna de las escalas, es decir, que todos los indicadores “miden la misma dimensión”. La validez hace referencia a que se “mide lo que se quiere medir”, por lo que se debe evaluar en los tres ámbitos comentados: contenido, convergente (capacidad de captar aspectos del mismo constructo) y divergente (discrimina entre constructos, de forma que no hay indicadores “generales” para todos).

Tras comprobar que el modelo de medida está correctamente especificado (y sólo tras este paso), podemos evaluar el modelo estructural.

3

Medidas de ajuste del modelo estructural

Se analizan, básicamente, tres aspectos: el grado de ajuste, la significación de la estimación de los parámetros y la capacidad predictiva del modelo.

GRADO DE AJUSTE: se mide mediante el coeficiente R^2 de Nagelkerke, de interpretación similar al de Pearson.

$$R_N^2 = \frac{1 - \left[\frac{L(M_{intercep})}{L(M_{full})} \right]^{\frac{N}{2}}}{1 - L(M_{intercep})^{\frac{N}{2}}}$$

SIGNIFICACIÓN DE LOS PARÁMETROS: al disponer sólo de una muestra para estimar los parámetros, es preciso simular otras para poder estimar la variabilidad de las estimaciones y construir un estadístico t-Student.

BOOTSTRAPPING

CAPACIDAD PREDICTIVA: capacidad del modelo para “reconstruir” las observaciones iniciales a partir de la información contenida en el resto de la muestra.

BLINDFOLDING

SEGMENTACIÓN DE LA MUESTRA: en Ciencias Sociales es difícil garantizar que toda la muestra se ajusta a la misma distribución probabilística. Con PLS, es posible abordar la segmentación en función del modelo estructural, es decir, agrupando a los individuos de la muestra de forma que presenten parámetros lo más diferentes posibles.

FIMIX

Heterogeneidad no observable: FIMIX

En la práctica, es usual encontrar cierto grado de heterogeneidad, que puede invalidar los resultados. Cuando se utilizan variables latentes, se puede producir una heterogeneidad no observable en los indicadores, sino que se produce en las relaciones entre constructos.

En PLS, existen distintos algoritmos para tratarla:

- PLS-TPM
- REBUS-PLS
- FIMIX-PLS

La heterogeneidad no observable afecta a los resultados del modelo:

- Resultados difíciles de explicar, no esperados
- Ajuste pobre (R^2 , AVE, IFC, etc.)
- Coeficientes no significativos por el efecto de la composición

El enfoque FIMIX-PLS es una técnica de segmentación basada en el modelo ajustado, que produce grupos en función de las respuestas obtenidas (y los valores de los parámetros estimados)

**Veamos algunos ejemplos
prácticos
(learning by doing)**

Ejemplo 1: Motivaciones del turista en Cuenca.



**TURISMO CULTURAL EN LA
CIUDAD DE CUENCA,
PATRIMONIO DE LA
HUMANIDAD**



- ▮ **Introducción**

- ▮ **Turismo Cultural y Ciudades Patrimonio de la Humanidad**

- ▮ **El Turista Cultural**

- ▮ **Metodología de investigación:**

- ▮ **Diseño de la encuesta**

- ▮ **Muestra y procedimiento de obtención de la información**

- ▮ **Motivaciones del turista cultural**

- ▮ **Conclusiones e implicaciones**

Introducción

- ▮ **Notable crecimiento del sector del turismo cultural en nuestro país.**
- ▮ **Ciudades o conjuntos históricos declarados como “Patrimonio de la Humanidad” por la UNESCO.**
- ▮ **España es el país del mundo que goza de una mayor riqueza artística de interés universal.**




Objetivos Específicos... (Doble):

- ▮ **Analizar las motivaciones del turista cultural.**
- ▮ **Contrastar los cambios en la demanda turística al tiempo que otras motivaciones de ocio.**

El entorno cultural y el turista

TURISMO CULTURAL:

 Organizar y realizar un viaje con una propuesta de contenido territorial o temático para llevar a cabo actividades que permiten experimentar la cultura y las diferentes formas de vida de otras gentes y, como consecuencia, conocer y comprender sus costumbres, tradiciones, entorno físico, ideas intelectuales y lugares históricos, arqueológicos, arquitectónicos o de otra significación cultural.

(Vázquez Casielles, 2005).

El entorno cultural y el turista

▮ **Gran oferta turística cultural** en las “Ciudades Patrimonio”.

▮ Se deben tener en cuenta la serie de oportunidades motivadas por el turismo que suponen para la recuperación del patrimonio cultural y la dinamización de las economías.

▮ España es el país del mundo en el que más ciudades ostentan el título de pertenencia a la Lista del Patrimonio Cultural de la Humanidad de la UNESCO.

▮ Además, se encuentran organizadas en el “**Grupo de Ciudades Patrimonio de la Humanidad de España**”.

Metodología: Diseño de la encuesta

🏛️ Encuesta a visitantes enmarcada dentro del proyecto “Observatorio Turístico de Cuenca”.



Estructura del cuestionario:

🏛️ **Bloque 1:** Número de visitas a la ciudad de Cuenca, opinión general de la ciudad, valoración respecto a otras ciudades similares, motivo principal de su visita, duración prevista de su estancia y lugar de alojamiento.

🏛️ **Bloque 2:** Calidad percibida de los diferentes servicios turísticos: 15 ítems.

🏛️ **Bloque 3:** Motivaciones de su visita turística a la ciudad: 34 ítems.

🏛️ **Bloque 4:** Preguntas de clasificación.

Ficha técnica de la investigación

UNIVERSO	Turistas mayores de 18 años
UNIDAD MUESTRAL	Individuos que solicitan información en oficinas de turismo
ÁMBITO GEOGRÁFICO	Ciudad declarada "Patrimonio de la Humanidad"
METODOLOGÍA	Encuesta personal mediante cuestionario estructurado
TAMAÑO DE LA MUESTRA	1075 encuestas válidas
ERROR DE MUESTREO	± 4,19%
NIVEL DE SIGNIFICACIÓN	95% (p=q=0,5)
FECHAS DEL TRABAJO DE CAMPO	Enero a diciembre de 2006

Resultados preliminares y estado actual de la investigación

Validación del cuestionario final:

Bloque	Ítems	Alfa de Cronbach
Relación calidad-precio	3	0,774
Motivación	14	0,992
Valoración	9	0,891

Análisis factorial:

Medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO)		,933
Test de esfericidad de Barlett	Chi cuadrado aprox.	16395,001
	g.l.	, 325
	Significación	,000

Resultados preliminares y estado actual de la investigación

Resultados de la matriz de componentes rotados:



Conclusiones

▮ El análisis factorial de los datos nos permite definir tres factores que explican más del 55 % de la varianza.

▮ Turistas interesados culturalmente.

▮ Diferencias entre la demanda de turismo cultural internacional y nacional.

▮ Variables específicas de motivación del turista.

Limitaciones

▮ Carácter estático y puntual.

▮ Ampliación muestral: turistas que no demandan información en oficinas.

Líneas de investigación futura

▄ Nuevos análisis estadísticos (e.g. SEM).

▄ Ampliación del estudio.

▄ Estudio de nuevas variables.

▄ Posibilidades de análisis longitudinal.



Muchas gracias por su atención

Jose.Mondejar@uclm.es