

(#163) . CRECEN LAS DUDAS SOBRE PLS (PARTIAL LEAST SQUARES)

[REVISIÓN DE ARTÍCULO] Partial Least Squares (PLS) es una técnica de análisis de datos cuyo uso **ha crecido muchísimo en los últimos años**, especialmente en las áreas de gestión de empresas o marketing. Se presenta típicamente como una alternativa a los modelos de ecuaciones estructurales (SEM) con una serie de supuestas ventajas sobre éstos, como la capacidad para manejar muestras pequeñas y distribuciones alejadas de la normalidad.

Sin embargo, existe un cuerpo de evidencia importante (los autores mencionan varios artículos recientes) que sugiere que muchas de esas supuestas ventajas de PLS no tienen un sustrato estadístico y empírico, y que además, el número de limitaciones de esta técnica es amplio y poco entendido.

En este artículo, los investigadores **explican el porqué PLS debería ser evitado como técnica de análisis de datos**.

Qué es PLS

Un análisis PLS consiste en dos etapas. En la primera de ellas, los indicadores de las variables latentes se combinan como sumas ponderadas; en la segunda de ellas esas sumas ponderadas se emplean en regresiones separadas, analizando su significación al comparar el ratio de ese coeficiente de regresión y su error estándar calculado por bootstrapping con una distribución t-Student. La idea es combinar esos indicadores con el fin de que esas ponderaciones sean más fiables que cualquiera de esas medidas por separado. Por tanto, esa suma ponderada es en realidad un proxy de la variable latente. De este modo, en realidad la diferencia con respecto a una regresión tradicional o un análisis de

componentes principales es la forma en que esos indicadores se ponderan.

La cuestión aquí es cómo genera PLS esas ponderaciones de ítemes, y si el procedimiento es más óptimo que el de técnicas clásicas de resumir información como el análisis factorial (reflectivo) o el análisis factorial de componentes principales (formativo)

Ponderaciones

Los autores critican varios de los argumentos por los cuales se defiende el procedimiento de ponderación de los ítemes. En primer lugar **no hay pruebas matemáticas de que ese procedimiento sea óptimo**. Si es para maximizar el R-cuadrado de las regresiones entre las ponderaciones, los autores indican que otros procedimientos lo superan ampliamente, y que además no está clara la razón por la cual se debe buscar esa maximización del R-cuadrado.

Otro argumento en defensa de PLS es que las ponderaciones reducen el impacto del error de medida, algo que los autores desmontan aludiendo a **varias simulaciones que demuestran que no es así**.

Además la dependencia del modelo de los indicadores ponderados conlleva la inestabilidad de esas ponderaciones, y la aparición del “**interpretational confounding**”, algo que (y esto no lo dicen los autores en el artículo) puede ser mejor gestionado con los modelos de ecuaciones estructurales **fijando el coeficiente del mejor indicador posible** de cada variable latente (indicador “*gold standar*” en terminología del profesor Leslie Hayduk).

Estimadores sesgados e inconsistentes

Aproximar variables latentes con ponderaciones de indicadores produce estimadores sesgados e inconsistentes. Recordemos que la consistencia significa que el estimador se acercará al

valor poblacional a medida que se incrementa el tamaño muestral. **La propia literatura sobre PLS admite que los estimadores no son consistentes**, lo cual se produce en mayor medida cuando los indicadores no son extremadamente fiables.

Capitalization on chance

Los autores indican que se ha puesto poco énfasis en la literatura sobre PLS en la “capitalization on chance”, que se refiere a la obtención de resultados sesgados determinados por la propia aleatoriedad muestral, es decir, el tamaño del error debido al análisis de una muestra que es en sí un caso extremo. **Este tipo de problemas se agudiza con muestras pequeñas, precisamente las que suele emplear PLS.**

Problemas cuando se testan modelos

Testar un modelo significa establecer una serie de restricciones en los parámetros y evaluar la probabilidad de los estadísticos observados dadas esas restricciones. En terminología SEM, sería realizar restricciones causales (por ejemplo, fijar varios parámetros a cero entre relaciones de variables latentes) y luego testar esas restricciones contra los datos empíricos a través de la discrepancia entre la matriz de covarianzas estimada y la restringida por el modelo propuesto. Esa discrepancia se mide en términos estadísticos con el test de la chi-cuadrado, que no es más que un test que nos dice si esa divergencia entre las dos matrices es estadísticamente relevante o no.

Pero en PLS no se pueden testar esas restricciones usando un test estadístico, como el de la chi-cuadrado. Por tanto, no hay ninguna forma de conocer si la teoría propuesta (el modelo del investigador) es congruente con los datos empíricos o no, es decir, no es un método para testar modelos.

Aunque algunos defensores de PLS argumentan que no se trata de buscar el ajuste del modelo sino maximizar su capacidad predictiva, pero si el modelo de partida no es correcto, esa

supuesta ventaja no es tal, ya que los estimadores serían sesgados, y probablemente esa capacidad predictiva alta mostrada en esa muestra específica no lo sería en una posible replicación.

Problemas para evaluar la calidad de las mediciones

Pese a que PLS emplea dos términos de SEM (fiabilidad compuesta y varianza extraída media), la forma de computarlos hace que su resultado sea cuestionable, ya que no están computados a través de la relación de análisis factorial, tal y como hace SEM, sino entre la correlación entre los indicadores y sus propias ponderaciones. Además, y como recalcan los autores, estos índices de calidad de la medición no están basados en ningún test estadístico.

A este respecto, vuelvo a remitirme a la perspectiva explicada por Hayduk en sus libros y en sus artículos sobre la estrategia de modelización y la calidad de las mediciones; es muy sencillo: **primero ver si el modelo se ajusta vía chi-cuadrado, y luego valorar la fiabilidad de los ítems de cada variable latente en función de las restricciones del indicador *gold standard*.**

Uso del test t en situaciones alejadas de la normalidad

Los autores enfatizan una de las paradojas de PLS, y es el de usar remuestreo para calcular los errores estándar de los parámetros, pero luego emplear el t-test para estimar la significatividad de los parámetros en situaciones de no normalidad, lo que incrementa los falsos positivos. Además, y como bien referencian los autores, **el uso de bootstrapping en muestras pequeñas (PLS es usado en muchas ocasiones en muestras muy reducidas) es muy arriesgado.**

Muestras pequeñas

Incluso los defensores de PLS tienen problemas para justificar el uso de esta herramienta en muestras pequeñas, es como **una**

especie de mantra sin evidencia sólida detrás. Los estudios que comparan PLS y SEM en muestras pequeñas generalmente encuentran que el sesgo es menor en SEM que en PLS. Los autores abogan en ese tipo de muestras por usar mínimos cuadrados en dos etapas.

Datos alejados de la normalidad

PLS usa estimación por mínimos cuadrados ordinarios para la estimación de los parámetros, lo que conlleva la asunción de homocedasticidad de los errores y también normalidad. Es cierto que la normalidad no es indispensable para que los estimadores sean consistentes, insesgados y eficientes, pero sí que se asume para los test estadísticos inferenciales.

Predicción frente explicación

Es cierto que otro mantra sobre PLS es que predice más que explica. Cuando hablamos de predicción estamos penalizando el sustento teórico de los modelos y su correcta especificación por la construcción de algoritmos que proveen de buena capacidad predictiva /clasificatoria, pero cuya interpretabilidad es una quimera. Esas **formas de predicción de caja negra**, con algoritmos como las redes neuronales y otros similares son poco útiles para explicar.

Si PLS se enfatiza en la predicción frente a la explicación está enfatizando su débil capacidad para explicar teorías, pero al mismo tiempo debería de justificar que es superior algorítmicamente a otros métodos de predicción (redes neuronales, máquinas de vectores soporte, etc.)

Exploratorio frente a confirmatorio

Si PLS no puede detectar malas especificaciones de los modelos, difícilmente puede ser de ayuda en tareas exploratorias (teoría en sus comienzos). En el caso de buscar patrones desde los datos, las técnicas comentadas anteriormente de *data mining* pueden ser empleadas, por lo que

realmente PLS no añade “valor” como técnica a las ya existentes. En cualquier caso, como comentan los autores, parece incongruente que si se quiere explorar al estilo *data mining* se requiere que con PLS se plantee un modelo teórico para el modelo de medida y el modelo estructural.

Mediciones formativas

Esta es otra de las supuestas ventajas de PLS, aunque como bien indican los autores hace falta una **discusión más profunda sobre en qué medida las mediciones formativas de un constructo no puede re-especificarse como variables latentes y un ítem reflectivo cada una de ellas**. Esa re-especificación estaría en consonancia con la representación de las mediciones como reflejos de la variable que se quiere medir, y no como causa de ellas, algo que es, desde muchos puntos de vista, altamente cuestionable.

Conclusión/Comentarios

A pesar de las recientes aportaciones en la literatura que tratan de cubrir algunas de las limitaciones mencionadas, es evidente que PLS es una herramienta que debería ofrecer bastante desconfianza para ser usada por investigadores aplicados, al menos con la “alegría” que se suele hacer en las áreas de gestión de empresas o marketing, donde es común su uso para testar modelos, en el caso de usar muestras pequeñas, etc. Es como una especie de vía de escape para no emplear modelos de ecuaciones estructurales (SEM) y aplicar esta alternativa que promete ser más adecuada para teorías débiles, muestras pequeñas y datos alejados de la normalidad.

Los autores destacan el editorial de 2015 de la principal revista científica en el campo de la gestión de empresas, Journal of Operations Management, que argumentaba que las investigaciones que emplearan esta técnica serían muy probablemente rechazados de inmediato.

Concluyen comentando el hecho de que **PLS no debería de**

enseñarse en las universidad a los estudiantes de doctorado, al menos, no de la forma que ahora se hace, donde cabrían priorizar otras herramientas con un soporte mucho más sólido desde el punto de vista estadístico.

En definitiva, como investigador llevo años desconfiando de esta herramienta, nunca la he usado y no he aconsejado nunca su empleo a mis estudiantes. Como revisor he rechazado los artículos que usaban PLS como alternativa a SEM, y como profesor he desaconsejado que se organizaran cursos de doctorado o de formación de PLS si los presupuestos son limitados y se pueden priorizar otro tipo de cursos. Sin embargo, puedo cambiar de opinión si la investigación en PLS muestra evidencias sólidas de mejora que cubran las limitaciones expuestas. Pero, de momento, **parece prudente alinearse con las conclusiones derivadas de este artículo.**

Rönkkö, M. et al. (2016). Partial Least Squares Path Modeling: Time for Some Serious Second Thoughts. *Journal of Operations Management*, doi: [10.1016/j.jom.2016.05.002](https://doi.org/10.1016/j.jom.2016.05.002)

Indicadores de calidad de la revista*

	Impact Factor (2015)	Cuartil	Categoría
Thomson-Reuters (JCR)	6.807	Q1	OPERATIONS RESEARCH & MANAGEMENT SCIENCE
Scimago (SJR)	5.05	Q1	STRATEGY & MANAGEMENT

* *Es simplemente un indicador aproximado para valorar la calidad de la publicación*

