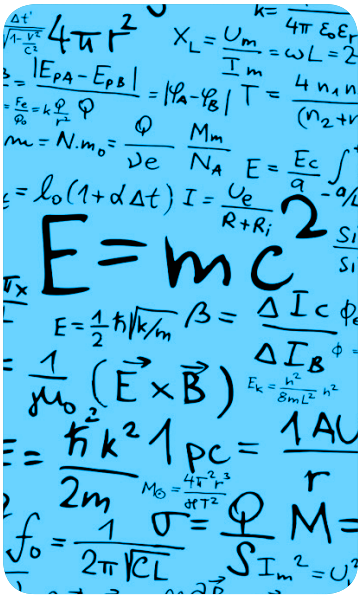
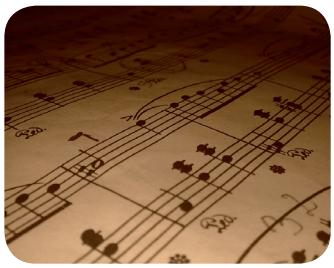




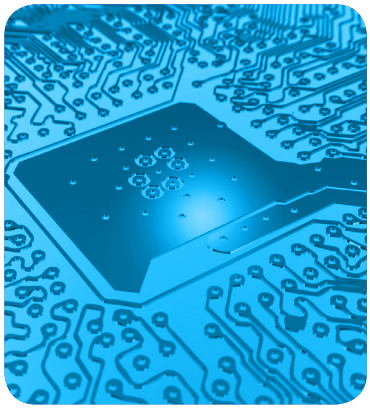
MÁSTERES de la UAM

Facultad de Psicología /12-13

Máster en Metodología de las Ciencias del Comportamiento y de la Salud

UNIVERSIDAD AUTONOMA DE MADRID

Campus Internacional excelencia UAM CSIC



UAM EDICIONES

Evaluación de la significación práctica de los parámetros estimados en modelos con factores débiles mediante Análisis Factorial Confirmatorio (AFC) y simulación Monte Carlo

Daniel Ondé Pérez

Resumen

En el presente trabajo se analiza la calidad en la recuperación de los denominados factores débiles (RFD) con objeto de mostrar que la interpretación directa de las saturaciones que pertenecen al factor débil debe realizarse a partir de niveles más exigentes que los utilizados al evaluar la dimensionalidad de un modelo, test, escala, etc. Se han realizado dos estudios de simulación Monte Carlo mediante AFC, el primero elaborado a partir del modelo de MacCallum y Tucker (1991), compuesto por 3 factores ortogonales y 12 indicadores distribuidos normalmente, con *lambda* = 0,50 en los 3 indicadores que representan al factor definido como débil. En el segundo estudio se utilizó la misma estructura pero con algunas variaciones en el nivel de debilidad, la correlación entre factores, y el nivel de carga del resto de factores. Se compara el nivel promedio de recuperación por condición experimental utilizado habitualmente en la literatura revisada (*nivel promedio de RFD*), con la evaluación de la precisión de las estimaciones solución a solución desde un enfoque más aplicado (*tipos de RFD*). Los resultados indican que los niveles promedio de RFD no garantizan la posibilidad de interpretar las saturaciones de factores débiles, resultando necesaria la evaluación de los tipos de RFD.

Índice

Introducción

En el contexto del Análisis Factorial Confirmatorio (AFC), el punto de partida es la especificación correcta de los modelos y esta fase debe elaborarse en relación con la teoría. En esta fase de especificación es necesario resolver lo que denominaremos el *problema de representación* ya que en psicología (y más concretamente en psicometría), es habitual encontrar situaciones en las que existe más de un posible modelo para el constructo objeto de estudio. Esta pluralidad de modelos implica una competencia en el intento de explicar la estructura y dimensionalidad de los datos, por lo que el problema de representación puede expresarse de forma genérica a partir del siguiente interrogante: ¿en qué medida las dimensiones latentes de un modelo factorial se encuentran adecuadamente representadas por las variables, indicadores o ítems utilizados para su evaluación?

Esta situación es de especial interés en constructos formados por dimensiones en las que aparece un factor débil (en contextos aplicados los modelos factoriales no suelen ser equipotenciales), lo que puede llevar a dos tipos de prácticas inadecuadas en el intento de mejorar la bondad de ajuste: a) reducir la dimensionalidad real del modelo y b) eliminar los indicadores o ítems que generan el factor débil (FD). Para evitar este tipo de prácticas se hace preciso evaluar el funcionamiento de los procedimientos de estimación en estas condiciones, lo que se conoce como *estudio de recuperación de factores débiles* (RFD).

Puesto que la capacidad para determinar correctamente la estructura y dimensionalidad de los datos es un tema de gran relevancia para la medición en psicología, se hace preciso disponer de herramientas para detectar incluso pequeños errores de especificación en los modelos factoriales. Entre estas herramientas juega un papel muy importante la simulación Monte Carlo siendo de especial relevancia el estudio de constructos multidimensionales en los que se identifica un FD.

Desde la óptica de la recuperación de parámetros, varias son las temáticas que se derivan de la cuestión del problema de representación, tales como el número mínimo de indicadores que debe tener un factor, el tamaño muestral necesario para estimar los parámetros del modelo con precisión, el tipo de error presente en el modelo, el número de factores a representar, la adecuación de los distintos métodos de estimación existentes, el nivel de medida de los indicadores, la existencia de covariación entre factores, o el nivel de carga o saturación mínimo que debe tener un factor para explicar los indicadores que lo representan. Todas estas temáticas se encuentran estrechamente relacionadas ya que ninguna de ellas da cuenta por sí sola del problema de representación.

Hasta ahora, en los denominados estudios RFD el interés se ha centrado especialmente en la comparación entre métodos de estimación ante determinadas fuentes de variación, como el tamaño muestral, la existencia de factores correlacionados o el nivel de carga factorial de los indicadores del FD. Además, los estudios RFD han evaluado la calidad de la recuperación a partir de índices que promedian las desviaciones entre parámetros estimados y poblacionales por condición experimental, promediando a su vez todos los indicadores que pertenecen al FD. Sin embargo, estos estudios no han prestado suficiente atención a la significación práctica de los parámetros estimados. En otras palabras, los índices utilizados habitualmente obtienen valores informativos de cada condición simulada que resultan demasiado genéricos, desatendiendo el riesgo relativo para la investigación aplicada de interpretar estimaciones del FD poco precisas en cada caso concreto.

El presente trabajo se sitúa dentro del ámbito de los estudios RFD en situaciones en las que la debilidad del factor viene determinada por niveles bajos de carga factorial. Su objetivo ha sido abordar la evaluación de la calidad de la recuperación introduciendo el análisis de patrones de estimación entre los indicadores del FD *solución a solución*. Para ello, se ha elaborado un procedimiento en el que se han establecido distintos puntos de corte en las saturaciones estimadas, lo que ha permitido clasificar en una primera fase estimaciones consideradas como óptimas, aceptables o imprecisas desde un punto de vista aplicado en cada una de las soluciones generadas. En base a esta primera clasificación, se han identificado distintos tipos de RFD a partir de los patrones de estimación presentes en cada solución factorial generada, y se han calculado porcentajes de soluciones que pueden considerarse como aceptables e inaceptables.

En este trabajo se compara el enfoque relativo a la identificación de patrones y tipos de RFD con el enfoque desarrollado en estudios previos sobre RFD. La evaluación de los tipos de RFD es complementaria al análisis por condición experimental realizado en estudios previos, si bien permite profundizar en la capacidad para interpretar los modelos factoriales más allá de su dimensionalidad subyacente, evaluando las garantías que tiene la interpretación directa de la relación entre indicadores y FD ante distintas condiciones.

La organización de este documento se articula a partir de un primer capítulo teórico en el que se reflexiona sobre el problema de representación con FD y la validez de los modelos factoriales, con el fin de resaltar la importancia de este tipo de investigación. A continuación, se formalizan los aspectos básicos del Análisis Factorial (AF), distinguiendo entre Análisis Factorial Exploratorio (AFE) y AFC, y se desarrollan de manera sintética las aproximaciones más frecuentes al proceso de estimación de parámetros. Este capítulo concluye con una

revisión de distintos estudios sobre recuperación de parámetros en AF, de la que se han extraído las principales recomendaciones y variables objeto de estudio que son de interés dentro de los estudios RFD, y que hemos denominado como tópicos de investigación.

En el capítulo 2 se define el diseño metodológico de la investigación, asentado en dos estudios de RFD mediante simulación Monte Carlo elaborados con los programas PRELIS 2.0 y LISREL 8.8, en los que se manipuló el tamaño muestral sobre el que se elaboran los modelos, el nivel de debilidad de los factores y la ausencia o presencia de correlación entre factores, con 1.000 réplicas muestrales por condición experimental.

En el ESTUDIO 1 se replicó y reanalizó el modelo factorial inicialmente propuesto por MacCallum y Tucker (1991), posteriormente ampliado por Briggs y MacCallum (2003) y por Ximénez (2006), manipulando el tamaño muestral sobre el que se realizaban las estimaciones. Este modelo factorial se define a partir de 12 indicadores con distribución normal univariada y 3 factores, siendo el tercero de ellos relativamente débil (λ = 0,5). Este estudio ha permitido una re-evaluación de los resultados a la luz de su significación práctica comparando entre la calidad de la recuperación basada en promedios por condición y la evaluación de los tipos de RFD identificados solución a solución.

En el ESTUDIO 2 de simulación se analizó el efecto de variables no contempladas en el primer estudio, incluyéndose variables que han mostrado tener un potencial efecto sobre la calidad de la recuperación de parámetros en otros estudios RFD revisados, como es la variación a la baja de las saturaciones de los indicadores en el FD (en este caso, λ = 0,2, λ = 0,3 y λ = 0,4) o la correlación entre factores (ρ = 0,3). Nuevamente, se compara entre la recuperación promedio por condición experimental y la evaluación de los tipos de RFD identificados.

En ambos estudios los parámetros se han recuperado a partir de ML y ULS, al tratarse de dos de los métodos de estimación más utilizados en AFC, presentes en los trabajos previos revisados, y al existir cierto grado de discrepancia sobre la adecuación y las ventajas de usar uno u otro en diferentes condiciones simuladas. En el capítulo 3 se exponen los resultados encontrados, seguidos del epígrafe que versa sobre la discusión de dichos resultados y las conclusiones de la investigación. Finalmente, en los Anexos se aporta información adicional a los resultados presentados.

1. Marco teórico

Al aplicar AF en un contexto en el que tanto el buen conocimiento de las dimensiones subyacentes a los datos como la adecuada selección de sus indicadores observables son aspectos cubiertos en la investigación, nos encontramos ante una situación en la que la recuperación de parámetros es óptima. En estos casos, los modelos factoriales elaborados no solo tienen relevancia desde el punto de vista teórico, sino que también se alcanza la suficiente cantidad de relevancia estadística.

Si los factores de un modelo presentan saturaciones altas (lo que implica un error de medición bajo), la falta de ajuste solamente puede deberse a limitaciones en el tamaño de la muestra, siempre y cuando los modelos estén bien especificados. Entonces, resulta lógico pensar que al aumentar el tamaño de la muestra tanto el ajuste de los datos a los modelos como la calidad de la recuperación de los parámetros alcanzará los niveles deseados, debido a las propiedades asintóticas de los estimadores.

¿Qué ocurre cuando en determinadas ocasiones el modelo factorial presenta FD? Ya no se trata solamente de una cuestión de estabilidad o de precisión de los parámetros estimados, sino que la investigación puede derivar en un delicado equilibrio que es el resultado de contraponer la relevancia teórica de los modelos con el adecuado ajuste a los datos empíricos y con el principio de parsimonia. Esto es, en ocasiones el investigador puede verse ante la necesidad de eliminar varianza relevante de su modelo en aras de un mayor ajuste o, a la inversa, aportar a la comunidad científica modelos más comprehensivos aunque débilmente sustentados sobre los datos empíricos.

La presencia de FD en los modelos factoriales puede suponer un problema para el investigador al verse forzado a tomar algún tipo de decisión *no deseada a priori*, lo que implica a su vez ciertas consideraciones sobre la posibilidad de evaluar la validez, especialmente en lo referente a la validez de constructo o *análisis de la estructura interna* (Messick, 1989, 1995, Elosua, 2003). El análisis de la estructura interna evalúa el grado en el que las relaciones entre los ítems y los componentes del test conforman el constructo (o constructos) que se pretenden medir y sobre el que debe basarse la interpretación de sus puntuaciones.

Según los últimos estándares (AERA, APA y NCME, 1999[1]), el análisis de la estructura interna se centra en la evaluación de la *dimensionalidad* del test y de la *invarianza factorial* al comparar grupos, aunque esto último excede los objetivos de la presente investigación.

[1] Actualización prevista en 2013.

Actualmente el AFC es probablemente la técnica más empleada para evaluar la validez de constructo, con el fin de buscar evidencias que garanticen la fidelidad entre la estructura de las puntuaciones y la estructura del constructo (o constructos) objeto de estudio. Su aplicación no obstante, requiere de una teoría sobre el constructo medido que a veces puede verse comprometida por la debilidad de alguna de las dimensiones hipotetizadas (esto es, la presencia de un FD).

En principio, *si la inclusión del FD en el modelo factorial potencia la capacidad para obtener evidencias respecto a la estructura interna del constructo objeto de estudio, la recomendación desde el punto de vista de la validez del modelo debe ser la de mantenerlo, aunque esto conlleve cierta pérdida en su ajuste global*. Pero, ¿hasta qué punto es admisible dicha pérdida?

Que las saturaciones presentes en un factor sean bajas depende ciertamente del error con el que están medidas, aunque también conviene señalar que dichas saturaciones son también el reflejo de indicadores débilmente correlacionados entre sí. Es labor del investigador decidir cuándo las correlaciones entre indicadores son lo suficientemente elevadas como para tener algún tipo de relevancia teórica y de significación práctica, si bien la ocurrencia de este tipo de situaciones, entre otras, son las que justifican los estudios RFD. Mediante estos estudios se trata de conocer e identificar en qué medida la presencia de un FD puede ser asumida estadísticamente bajo ciertas condiciones o escenarios dentro del modelo factorial elaborado. De manera complementaria, los estudios RFD tienen la utilidad de proporcionar al investigador un conocimiento más preciso sobre los riesgos de interpretar directamente las relaciones numéricas entre factores e indicadores ante determinadas condiciones de aplicación. En este sentido, los estudios RFD son útiles para indicar cuándo puede ser apropiado realizar *interpretaciones directas con garantías* o, por el contrario, cuándo el contexto de aplicación no permite ir más allá de la *evaluación de la dimensionalidad de los datos*.

Por otro lado, dentro del campo de la Psicología, los estudios RFD mediante AFC se encuentran insertos en un ámbito de investigación más amplio, relacionado con la identificación e interpretación de dimensiones subyacentes en los datos mediante aplicación de AF. Además, la calidad con la que se recuperan las cargas o saturaciones que representan las distintas dimensiones subyacentes en AF es un problema de investigación que ha recibido una gran cantidad de atención dentro del campo de la Psicología, lo cual se refleja en el elevado número de publicaciones y monográficos existentes sobre este tema. Por esta razón,

en este capítulo se exponen de manera sintética cuestiones básicas sobre la elaboración de modelos factoriales que van a permitir contextualizar el presente trabajo de investigación.

Las cuestiones que se abordan en esta investigación afectan especialmente a la fase de especificación de los modelos y al problema de representación, así como a su interpretación. Partiendo del modelo de factor común de Spearman, se desarrolla formalmente el AF y se exponen los métodos más importantes de estimación de parámetros. Se describen cuales son las diferencias más destacadas entre AFE y AFC, justificando el uso de AFC en los estudios de simulación Monte Carlo elaborados. Para cerrar este capítulo, se describen los principales tópicos de investigación en AF aplicables a los estudios RFD.

1.1. Formalización del AF a partir del modelo de factor común

El AF suele utilizarse para desvelar o identificar conjuntos de relaciones entre variables que son observables (también denominadas *indicadores*), y variables hipotéticas tales como rasgos, constructos teóricos, dimensiones subyacentes, etc. Estas variables hipotéticas se definen habitualmente como *variables latentes* (o *rasgos latentes* desde la tradición más psicométrica). Si un conjunto de variables observables o indicadores presenta un grado de relación estadística sustancial, resulta razonable suponer que dicha relación mutua surge de un conjunto más pequeño de variables no observables o variables latentes.

El propósito del AF es, por tanto, describir e identificar un conjunto de datos multivariados observables en términos de dichas variables no observables o latentes, comúnmente denominadas desde este entorno de trabajo como *factores*.

El modelo más antiguo y mejor conocido dentro de la familia de los AF es el *modelo de factor común* que Spearman publicó en 1904 (actualizado en 1927), y que Thurstone generalizó en 1947 (McDonald, 1982; Ferrando y Lorenzo-Seva, 1993; Brown, 2006). Este modelo puede expresarse en su forma generalizada de la siguiente manera:

$$x_j = \lambda_{j1}\xi_1 + \lambda_{j2}\xi_2 + \ldots + \lambda_{jm}\xi_{m1} + \delta_j \qquad [1]$$

En donde:

- x_j corresponde a una de las j variables observables, obtenida en una muestra de n observaciones independientes.
- λ_{jm} representa el peso o carga factorial relativa a la variable observable j en el factor ξ_m. En el caso en que m = 1 nos encontramos ante el modelo de un factor común de Spearman (cuando m > 1 se trata de un modelo multidimensional).

- δ_j es el término error, es decir, la parte de la puntuación observada del indicador x_j que no puede explicarse mediante la influencia de ninguna variable latente ni de ningún otro término error asociado a otra variable observable. Este término también se denomina *unicidad*.

El modelo de factor común, como es sabido, impone ciertas restricciones: el modelo asume que las puntuaciones factoriales se distribuyen N~(0,1), los términos error son independientes entre sí e independientes del factor común y su distribución tiene media cero, las distribuciones marginales tanto de las variables como del factor son normales, y se asume que las variables observables se distribuyen conjuntamente en forma normal multivariada (supuesto de normalidad multivariante). En notación matricial el modelo generalizado de factor común se expresa (Brown, 2006):

$$\mathbf{x} = \mathbf{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta} \qquad [2]$$

(Notación expandida) $$\mathbf{\Sigma} = \mathbf{\Lambda}_x\boldsymbol{\phi}\mathbf{\Lambda}'_x + \mathbf{\Theta}_\delta \qquad [3]$$

En la ecuación [2], **x** representa un vector p x 1 de variables observables aleatorias, $\mathbf{\Lambda}_x$ es una matriz p x n que contiene los pesos o cargas de los factores, $\boldsymbol{\xi}$ es un vector n x 1 que contiene las variables latentes o factores (sus varianzas, más concretamente), y $\boldsymbol{\delta}$ es un vector p x 1 que contiene los errores de medida o unicidades de las variables observables. En la ecuación [3] $\mathbf{\Sigma}$ es la matriz simétrica p x p de correlaciones entre las variables observables o indicadores, $\mathbf{\Lambda}_x$ es la matriz p x m de las saturaciones o cargas factoriales λ, $\boldsymbol{\phi}$ es la matriz simétrica m x m de las correlaciones entre factores, y $\mathbf{\Theta}_\delta$ es la matriz diagonal p x p de los errores o unicidades δ.

Cuando las variables latentes o factores no guardan relación con otras variables, más allá de los indicadores propuestos, se dice que las variables latentes son *exógenas*, y se representan normalmente mediante el parámetro ξ_i (con varianza ϕ_{ii}). Los indicadores se representan mediante la letra "*x*", las cargas o pesos factoriales mediante λ_x y los errores mediante δ.

Cuando los factores guardan relación con otras variables distintas a los indicadores (por ejemplo, un factor de segundo orden), se denominan variables latentes *endógenas* y se representan mediante el parámetro η_i (con varianza Ψ_{ii}). Los indicadores se representan mediante la letra "*y*", las cargas o pesos factoriales como λ_y o β_y y los errores mediante ε.

Las ecuaciones [2] y [3] son referidas por Brown para expresar en notación matricial la ecuación expuesta en [1], en donde $\mathbf{\Sigma}$ y $\boldsymbol{\phi}$ son matrices de correlaciones[2].

[2] En realidad, Brown utiliza una notación basada en variables "y" que suele ser habitual cuando se elaboran modelos a partir del programa LISREL. En el presente trabajo, para mantener la coherencia de la notación científica, se van a utilizar notaciones basadas en "*x*" ya que los factores de los modelos especificados no guardan relación estructural con otras variables o factores distintos de sus indicadores.

Así, $\boldsymbol{\Sigma}$ es la matriz que se utiliza cuando se elaboran modelos factoriales mediante AFE, ya que la fuente de información principal son las correlaciones entre las variables observables. La matriz $\boldsymbol{\phi}$ se utiliza en AFE cuando se asume correlación entre factores, derivando la solución en algún tipo de rotación oblicua.

En AFC lo habitual es referirse a $\boldsymbol{\Sigma}$ y $\boldsymbol{\phi}$ como matrices de varianzas-covarianzas. En esta línea, el uso de matrices de varianzas y covarianzas entre-indicadores y entre-factores es habitual en la notación que aparece en la literatura sobre AFC (por ejemplo, Anderson y Gerbing, 1988; Muthén, 1993; Skrondal y Rabe-Hesketh, 2004; Mulaik, 2009).

En ocasiones, la expresión $\boldsymbol{\Sigma}$ aparece reflejada como $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, en donde $\boldsymbol{\theta}$ se refiere genéricamente a todos los parámetros del modelo (λ, δ, etc.). Más concretamente, $\boldsymbol{\theta}$ es un vector $\boldsymbol{\theta} = (\hat{\boldsymbol{\theta}},\boldsymbol{\theta}^*)$, en donde $\hat{\boldsymbol{\theta}}$ contiene los parámetros libres y $\boldsymbol{\theta}^*$ contiene los parámetros fijos del modelo estructural (Mulaik, 2009; pág. 155).

En cualquier caso, lo importante aquí es que la expresión formal descrita en [3] es válida tanto para AFE como para AFC, ya que el principal propósito de ambos modelos de AF es identificar los factores latentes que reflejan la variabilidad de un conjunto de indicadores o variables observables, así como el patrón de covariación o de interrelación de dichos factores. No obstante, aunque varios de los conceptos que fundamentan el AF son comunes en ambos tipos de modelos, estas similitudes obedecen en mayor medida a la terminología que proviene del modelo de factor común (o factores comunes) y que se viene utilizado desde hace varias décadas, ya que las diferencias entre ambos modelos son bastante importantes.

Actualmente, se puede considerar el AFC como un método de análisis mucho más flexible que el AFE ya que puede generalizarse a un mayor número de contextos de investigación caracterizados por profundizar en los patrones de variación conjunta entre variables observables y factores latentes.

Una de las diferencias fundamentales entre el AFE y el AFC, como es bien conocido, es que el primero es un tipo de modelo de AF que no impone restricciones en los patrones de relación entre las variables observables (de ahí su carácter descriptivo-exploratorio), mientras que el segundo es un tipo de modelo en el que el investigador debe decidir y definir *a priori* cómo será la estructura a evaluar[3].

En función de dichas decisiones, el modelo planteado se formalizará imponiendo algún tipo de restricción sobre los parámetros a estimar (número de factores, ortogonalidad vs. covariación entre factores, métricas, etc.). En términos de Bollen (1989, pág. 228), en AFE el

[3] Otra distinción entre ambos tipos de modelos es la de *soluciones restringidas* (AFE) y *no restringidas* (AFC). Ver Jöreskog (1979) y Mulaik (1972), citado en Bollen, 1989 (pág. 227).

número de variables latentes no se encuentra determinado antes del análisis, todas las variables latentes normalmente tienen influencia sobre todas las variables observables, los errores de medida no pueden estar correlacionados, y no se suelen identificar los parámetros.

Por el contrario, en AFC el modelo está definido de antemano por el investigador al igual que el número de variables latentes, la influencia de cada variable latente sobre cada variable observada está identificada (no existe influencia de todos los factores sobre todos los indicadores), los errores de medida pueden estar correlacionados, y se requiere la identificación del modelo en términos del número de parámetros a estimar. En otras palabras, a diferencia de lo que ocurre en AFE, en AFC el investigador contrasta su teoría a partir del modelo propuesto. Por esta razón, mientras que el AFE *explora* las relaciones entre variables observables y variables latentes, el AFC las *contrasta* ya que está dirigido por el planteamiento de hipótesis. De hecho, dado que en la práctica el AFE se utiliza en ocasiones como guía en la identificación del número de factores subyacentes, la especificación previa del número de factores en los modelos de AFC implica un marcado movimiento desde la investigación exploratoria hacia la confirmatoria (Bollen, 1989; pág. 230).

Siguiendo a Brown (2006, cáp. 3), se exponen a continuación y de manera sintética cuáles son las principales similitudes conceptuales y terminológicas entre ambos modelos de AF y qué diferencias sustantivas se pueden establecer entre los mismos.

El primer concepto a destacar es el de *unicidad* (δ) o variabilidad específica de cada variable observable, esto es, la que no puede explicarse a partir del modelo factorial propuesto (ver ecuación [1]). Se expresa en términos de varianza error o de error de medición. Equivale para cada variable observable a la diferencia entre el total de la varianza que resulta posible explicar menos su *comunalidad* ($1 - h^2$). Además, la unicidad está compuesta por dos elementos, la varianza específica asociada a cada indicador (s) y su correspondiente error aleatorio (e).

$$\delta = s + e \qquad [4]$$

Tanto el AFE como el AFC diferencian entre comunalidad (varianza común) y unicidad (varianza específica), aunque una distinción importante entre ambos modelos de AF es que en AFE no se realizan especificaciones sobre la naturaleza de la relación entre los errores de medición de los indicadores, asumiendo que dichos errores son aleatorios. En AFE no se pueden establecer las restricciones necesarias en el modelo para contrastar empíricamente dichas asunciones teóricas sobre el error, lo que suele desembocar en la aparición de factores adicionales, de carácter espurio, que tienen muy poco significado teórico ya que están más

relacionados con efectos derivados del error de medición que con predicciones de la teoría sustantiva.

Por su parte, los modelos de AFC permiten establecer relaciones entre términos error (correlaciones entre parámetros δ), lo que implica la posibilidad (no siempre explícita) de asumir teóricamente cierto modelo relacionado con el error. Pero debe tenerse en cuenta que existen ciertos riesgos, a pesar de suponer una clara ventaja del AFC sobre el AFE. Cuando correlacionamos errores sin incluir explicaciones plausibles sobre las causas de dichas correlaciones, estamos enfocando la importancia de nuestro modelo solo desde el punto de vista de la comunalidad. Si el error no es aleatorio, nuestro modelo debería incluir consideraciones sobre las fuentes de unicidad que no son explicadas de manera directa.

Esta cuestión es especialmente importante desde el punto de vista de los estudios RFD, ya que en AFE pueden aparecer factores débiles de manera aleatoria, como resultado del error de medición, mientras que en AFC se puede contrastar de una manera más precisa hasta qué punto dichos FD pueden ser recuperados en aras de una mejor representatividad del contenido del constructo que se evalúa y de su dimensionalidad. El AFC, al estar orientado por la teoría, permite un mayor grado de control y de precisión sobre la representación de las dimensiones objeto de estudio.

Otro concepto que permite diferenciar claramente entre AFE y AFC es el relacionado con el *tipo de soluciones* que se pueden obtener con uno u otro modelo. Una solución completamente estandarizada (TOTEST) debe cumplir que la varianza de las variables latentes o factores sea igual a 1 (en ese caso, los pesos o cargas factoriales de los indicadores se pueden interpretar como correlaciones o coeficientes de regresión estandarizados). Tradicionalmente, en AFE la solución es de tipo TOTEST[4], mientras que en AFC la solución puede ser TOTEST, estandarizada (EST) y no estandarizada (NOEST).

Las soluciones EST son aquellas que reflejan relaciones entre variables observables que están en su métrica original y factores o variables latentes que están estandarizados, mientras que las soluciones NOEST producen una estimación de parámetros basada únicamente en la métrica original de las variables observables. Para soluciones EST, y en términos de regresión, por cada unidad estandarizada que aumenta la puntuación de un factor, aumenta una unidad no estandarizada cada indicador relacionado con dicho factor (esto es, en su métrica original).

[4] En ocasiones, se utiliza la matriz de covarianzas cuando se aplica ML en AFE, aunque esta situación produce diferencias en los valores de χ^2 respecto al uso de la matriz de correlaciones (Brown, 2006: pág. 88).

En las soluciones NOEST el cambio en cada unidad de medida se establece en las métricas originales tanto de los indicadores como de los factores. Por esta razón, muchos investigadores prefieren informar de las soluciones NOEST entendiendo que la relación entre variables observables y latentes puede estar sesgada en las soluciones estandarizadas (TOTEST y EST), al existir diferencias en la escala de medida de los indicadores, como ocurre con cierta frecuencia en la práctica.

La principal diferencia respecto al AFE en cuanto al tipo de solución es que la potencialidad de la solución NOEST es mucho mayor, convirtiendo al AFC en una herramienta más flexible. Dado que las soluciones NOEST implican varias de las potencialidades y de los aspectos técnicos del AFC, el procedimiento de estimación normalmente utiliza la matriz de varianzas-covarianzas (información no estandarizada), bien sea calculando dicha matriz desde los datos brutos, bien sea calculándola a partir de la matriz de correlaciones si es la matriz que se utiliza como input en el análisis. Sobre esta cuestión Cudeck (1989)[5] observa que las estimaciones obtenidas en AFC a partir de matrices de correlaciones producto-momento de Pearson son solo correctas si se cumple simultáneamente que (1) las únicas restricciones sobre los parámetros del modelo son las necesarias para fijar la escala de las variables latentes o consisten en fijar parámetros a cero, y (2) la matriz de residuos contiene ceros en la diagonal[6].

En el presente trabajo se utilizan matrices de correlaciones como input, dado que se trabaja con ULS además de con ML (ver apartado 1.2), pero las estructuras factoriales definidas en la parte empírica cumplen simultáneamente con ambos criterios.

Por último, otro concepto relevante para comparar entre AFE y AFC es el de *estimación de parámetros en relación a la parsimonia de los modelos*. Aunque los argumentos relativos a los métodos de estimación que se incorporaron en el apartado 1.2 provienen en su mayoría de fuentes relacionadas con el AFC, un aspecto que comparten el AFE y el AFC es la utilización de un mismo conjunto de métodos y procedimientos de estimación. Es el caso, por ejemplo, de los métodos ML, GLS, y ULS implementados en SPSS para AFE (ML y ULS aparecen también implementados en el programa FACTOR de Lorenzo-Seva y Ferrando, 2007)[7]. Estos estimadores se utilizan de manera equivalente también en AFC para evaluar el grado en el que la solución estimada reproduce adecuadamente la matriz observada (**S**).

[5] Citado en Batista y Coenders (2000).

[6] Esta situación no resulta extrapolable al caso de las correlaciones policóricas o tetracóricas ya que el procedimiento en tres etapas desarrollado por Muthén permite obtener estimaciones consistentes y asintóticamente eficientes (Forero, Maydeu-Olivares y Gallardo-Pujol, 2009).

[7] Los métodos de estimación son también compartidos por los modelos de ecuaciones estructurales (MEE), que se caracterizan por estudiar relaciones entre variables latentes endógenas y exógenas. El AFC es un caso particular de modelo estructural, también denominado modelo de medida.

En el caso del AFE se suele hacer referencia a estos métodos como métodos de extracción de factores, mientras que en AFC nos referimos a ellos como métodos de estimación de parámetros.

La diferencia fundamental respecto a la estimación de parámetros entre ambos modelos de AF es que los modelos de AFC, generalmente, son más parsimoniosos al tratar de reproducir las relaciones observadas entre indicadores a partir de la estimación de un número más reducido de parámetros que en AFE. Esto es debido a que, cuando los modelos incluyen más de un factor o variable latente[8], todos los indicadores en AFE saturan en todos los factores una cierta cantidad o carga, mientras que en AFC los indicadores tienen saturaciones nulas en los factores latentes con los que no guardan relación teórica ($\lambda_{ik} = 0$). Por esta razón, en AFE se utiliza la rotación para favorecer una interpretación más clara de qué indicadores saturan en cada factor, mientras que en AFC la rotación no resulta aplicable, puesto que los indicadores se relacionan estructuralmente con un solo factor[9]. Además, el hecho de que en AFC los modelos estén pre-especificados teóricamente permite a los investigadores imponer varios tipos de restricciones, modulando así el grado de parsimonia de los modelos, lo que permite a su vez realizar cierto tipo de comparaciones o de evaluaciones que no son viables mediante AFE. Es decir, en AFC el tipo de restricciones impuestas sobre los modelos se pueden evaluar estadísticamente comparando el ajuste y la recuperación de parámetros de los modelos más restrictivos con el ajuste y la recuperación de los modelos más generales.

Algunos tópicos de investigación relacionados con lo anterior son la comparación de modelos anidados, la evaluación de *tau* equivalencias o tests paralelos, la evaluación de la fiabilidad de las escalas, el análisis multigrupo o de invarianza factorial, o el análisis de la matriz multitratamiento-multimétodo.

La capacidad para imponer distintos tipos de restricciones convierte al AFC en un modelo sensiblemente más flexible que el AFE, como ya se ha comentado. Por esta razón parece más justificado actualmente el análisis de la RFD mediante simulación a partir de modelos de AFC ya que el grado de control sobre la representación de factores en la fase de especificación, así como la potencialidad de esta herramienta de análisis, es mayor que en AFE.

[8] Cuando los modelos incluyen un solo factor (por ejemplo, cuando se evalúa la unidimensionalidad), el número de pesos o cargas factoriales a estimar es el mismo tanto en AFE como en AFC.

[9] En AFC puede resultar de interés representar modelos en los que alguno de los indicadores tiene relación estructural con más de un factor latente. En cualquier caso, de existir otros factores distintos, las saturaciones de estos indicadores respecto a dichos factores serán igualmente nulas ($\lambda_{ik} = 0$).

1.2. Métodos de estimación de parámetros

La lógica de la estimación de parámetros se articula, fundamentalmente, en torno al proceso de búsqueda de aquel conjunto de parámetros que minimice las diferencias entre los elementos de la matriz de varianzas-covarianzas *observadas* y los elementos de la matriz de varianzas-covarianzas *reproducidas* por el modelo factorial (por ejemplo, Bollen, 1989; Brown, 2006; Abad, Olea, Ponsoda y García, 2011). La matriz observada se denomina **S** y la matriz reproducida o pronosticada por el modelo se denomina $\hat{\boldsymbol{\Sigma}}$.

En la práctica, el ajuste perfecto no resulta plausible ya que en Psicología la medición de variables introduce siempre alguna cantidad de error. Por lo tanto, la situación de partida suele ser aquella en la que $\hat{\boldsymbol{\Sigma}} \neq \mathbf{S}$, existiendo diferencias o residuos entre los parámetros reproducidos por el modelo y su valor equivalente en la matriz observada[10].

Estas diferencias se utilizan para expresar la cantidad de discrepancia entre ambas matrices, aplicando una función lineal cuyo rango de valores oscila teóricamente entre cero cuando lo reproducido por el modelo equivale a lo observado empíricamente, y hasta $+\infty$ cuando las matrices a comparar son de varianzas-covarianzas.

Esta función lineal se denomina *función de discrepancia* entre **S** y $\hat{\boldsymbol{\Sigma}}$, o $F(\mathbf{S}, \hat{\boldsymbol{\Sigma}})$, y su expresión matemática varía según sea el método de estimación empleado. Esto es, la forma concreta que adopta esta función, según los desarrollos de distintos autores, es la que conforma los distintos métodos de estimación utilizados.

A partir de la matriz **S** los métodos de estimación calculan de manera iterativa distintos conjuntos de parámetros. Para cada uno de estos conjuntos se reproduce la matriz $\hat{\boldsymbol{\Sigma}}$ a partir del modelo especificado y se calcula después la diferencia entre **S** y $\hat{\boldsymbol{\Sigma}}$. Estas diferencias se pueden formalizar en un vector $\mathbf{d}=\{d_1, d_2, \ldots, d_L\}$ que refleja los residuos resultantes al comparar la matriz observada con la pronosticada[11].

Con los residuos obtenidos (**d**) se calcula el valor correspondiente de la función de discrepancia (F) en cada paso del proceso, seleccionando finalmente aquel conjunto de residuos que produzca el valor más próximo a cero, esto es, que reproduzca más fielmente la matriz **S** original.

Mediante dicho proceso iterativo, cada método pretende alcanzar una solución final que minimice el valor de F. El proceso consta de tres pasos: (1) selección de valores iniciales, (2)

[10] Cuando se realizan estudios Monte Carlo, al partir de situaciones simuladas, la matriz observada corresponde a los parámetros poblacionales, por lo que la matriz debe expresarse como **Σ** en lugar de **S**.

[11] El número de residuos presentes en un modelo, y que serán tratados en F, equivale al número de datos disponibles en **S**, el cual se obtiene mediante la expresión $L = (J(J + 1))/2$, en donde J es el número de variables observables o indicadores. La expresión anterior se utiliza también para calcular el número de grados de libertad de los modelos. Batista y Coenders (2000) indican que el vector de residuos **d** contiene únicamente los elementos no duplicados o no redundantes de la matriz **S**, al ser simétrica.

reglas que establecen cuándo cambiar al paso siguiente, y (3) reglas que establecen cuándo finalizar el proceso iterativo (la formalización matemática de este proceso puede consultarse en Bollen, 1989; apéndice 4C). Cuando se alcanza una solución final, entonces decimos que la solución es convergente, o que el modelo es convergente.

Los métodos de estimación convergen cuando, bien la diferencia en el valor de *F* entre un paso y el siguiente es muy pequeña, o bien cuando las diferencias en los parámetros estimados son muy pequeñas entre ambos pasos.

A continuación, se describen las funciones de discrepancia de cuatro de los métodos más frecuentes en la literatura sobre AF, y que se diferencian en la forma en la que tratan matemáticamente el vector **d**.

Estos métodos son Mínimos cuadrados no Ponderados (*Unweighted Least Squares* - ULS), Máxima Verosimilitud (*Maximum Likelihood* - ML), Mínimos Cuadrados Ponderados (*Weighted Least Squares* - WLS)[12], y Mínimos Cuadrados Generalizados (*Generalized Least Squares* - GLS)[13].

$$F_{ULS} = \sum_{l}^{L} d_l^2 = \mathbf{dd'} \qquad [5]$$

$$F_{WLS} = F_{GLS} = F_{ML} = \sum_{l}^{L}\sum_{l'}^{L} w_{ll'} d_l d_{l'} = \mathbf{d'Wd} \qquad [6]$$

Los métodos de estimación difieren en el tratamiento que realizan de la matriz de pesos o ponderaciones **W**[14] [15]. El caso más general, aquel que cumplen los cuatro métodos, hace referencia a la *consistencia* de las estimaciones (Batista y Coenders, 2000). Es decir, las estimaciones obtenidas minimizando [5] y [6] son *consistentes*, ya que convergen hacia los valores poblacionales cuando las muestras son infinitas[16].

El caso más sencillo es el de ULS, en donde *F* obtiene su valor calculando la suma de cuadrados de los residuos. Matricialmente, **W** se sustituye en este método por una matriz

[12] Este método fue propuesto por Browne (1984) bajo la denominación Método Asintóticamente Libre de Distribución (*Asymptotically Free Distribution* - ADF). No obstante, suele denominarse también como WLS, siguiendo la terminología propuesta por Jöreskog para el programa LISREL.

[13] En realidad, como se verá más adelante, el método GLS es un caso particular del método WLS en el que se asume la normalidad multivariante de las variables observables. Batista y Coenders (2000) denominan a este método NT-WLS, de *Normal Theory Weighted Least Squares*, denominado GLS por otros autores como Jöreskog y por el programa LISREL.

[14] Se han simplificado las expresiones matemáticas de las distintas funciones de discrepancia descritas en este trabajo. La expresión completa hace referencia a **S** y a $\hat{\boldsymbol{\Sigma}}$: por ejemplo $F_{WLS}(\mathbf{S},\hat{\boldsymbol{\Sigma}})$.

[15] Las expresiones anteriores suelen denominarse *métodos de información completa*, ya que todos los residuos se tienen en cuenta simultáneamente para estimar todos los parámetros.

[16] La *consistencia* es una propiedad *asintótica* de los estimadores. En realidad, esta afirmación hace referencia a ciertas condiciones generales que debe cumplir **W** para que las estimaciones sean consistentes, y supone una generalización del Teorema 2.5 de Newey y McFadden (1995) aplicado a la estimación mediante ML.

identidad (**I**), obteniendo estimaciones consistentes con tamaños muestrales suficientes, simplificando la ecuación [6] al producto de los residuos al cuadrado (ecuación [5]).

Este método no resulta adecuado cuando se utilizan matrices de varianzas-covarianzas ya que el tamaño de los residuos dependerá de la escala de medida de las variables observables. Su uso está recomendado cuando se analizan matrices de correlaciones, con soluciones estandarizadas o totalmente estandarizadas[17]. De lo anterior se deriva que el método ULS no asume ningún tipo de distribución de las variables observadas por lo que no resulta posible obtener índices de ajuste con distribución estadística conocida (Abad, Olea, Ponsoda y García, 2011).

Utilizando matrices **W** más complejas se pueden obtener *estimaciones asintóticamente eficientes*, esto es, estimadores que convergen más rápidamente sobre los valores de los parámetros poblacionales a medida que aumenta el tamaño muestral (estimaciones en las que la varianza del parámetro estimado converge más rápidamente hacia cero). Los métodos ML, WLS, y GLS son más eficientes que otros métodos de estimación de parámetros. Pueden ser expresados formalmente de manera idéntica ya que la diferencia entre estos tres métodos se encuentra en el tipo de matriz utilizada para sustituir la matriz de pesos **W** de la ecuación [6].

El método ML fue desarrollado originalmente por R. A. Fisher en la década de los 20. Más adelante, importantes contribuciones fueron realizadas por H. Cramer, C. R. Rao y A. Wald (citado en Magnus y Neudecker, 1999 pp. 370). Asumiendo que la distribución de las variables observables es normal multivariante, el método ML sustituye en F_{ML} la matriz **W** por la matriz $\mathbf{\Gamma}^{-1}$.

La matriz **Γ** es la denominada *matriz de varianzas-covarianzas asintótica*, y se refiere a la variabilidad que tendrían los elementos de la matriz **S** a través de distintas muestras, lo que permite obtener (tras varios cálculos) una estimación asintótica para cada uno de estos elementos a partir de sus valores esperados. Al invertir **Γ**, la ponderación de F_{ML} asigna valores más pequeños al vector **d** cuanto mayores son los elementos correspondientes en **S**, puesto que en estas condiciones se asume que su variabilidad *entre-muestras* será también mayor. De esta manera, se relativizan los residuos potencialmente mayores permitiendo realizar la estimación más eficientemente[18].

A diferencia de ULS, el método ML permite obtener medidas estadísticas de ajuste y contrastes sobre la significación de los parámetros, siempre que el supuesto sobre la

[17] Recordemos que en este tipo de soluciones (ver apartado 1.1), los pesos o cargas factoriales de los indicadores se pueden interpretar como correlaciones o coeficientes de regresión estandarizados. El análisis de la matriz de varianzas-covarianzas produce soluciones no estandarizadas en las que la estimación de parámetros se basa únicamente en la métrica original de las variables observables (Brown, 2006).

[18] La formulación matemática extensa de la distribución asintótica de las estimaciones se puede consultar, por ejemplo, en Maydeu-Olivares, 2006;. pp. 60-62.

distribución de las variables observables se cumpla. Su principal desventaja es que tiende a subestimar los errores típicos a partir de cierto grado de desviación de la distribución normal, lo que aumenta la probabilidad de producir falsos positivos en la estimación (error tipo I). En definitiva, se considera que el método ML permite alcanzar estimaciones robustas en situaciones en las que existen pequeñas desviaciones respecto a la distribución normal.

Los métodos GLS y WLS fueron desarrollados por Browne[19] [20]. El método GLS sustituye también la matriz de pesos **W** por la inversa de la matriz de varianzas-covarianzas asintóticas ($\mathbf{\Gamma}^{-1}$) bajo el supuesto de normalidad multivariante. La principal diferencia con ML es que GLS solamente toma en consideración en **Γ** los momentos de segundo orden de las variables observables. Por su parte, en ausencia de normalidad multivariante, el método WLS utiliza la matriz **H,** en lugar de la matriz **Γ,** que combina los momentos de segundo orden y los de cuarto orden de las variables observables, sustituyendo en F_{WLS} la matriz **W** por la matriz $\mathbf{H}^{-1}$ [21].

El método GLS mantiene las mismas ventajas y desventajas que el método ML, al sostenerse en el supuesto de normalidad multivariante. No obstante, es un método menos extendido en la práctica ya que, en parte, su desarrollo en el contexto del análisis de estructuras de covarianza es algunos años posterior al de los desarrollos de ML que realizó Jöreskog a finales de los años 70. Además, en el caso de ML la matriz de pesos **W** es función del modelo especificado, mientras que en GLS solamente es función de los momentos de segundo orden, independientemente del modelo sobre el que se efectúe la recuperación de parámetros. Por este motivo, la estimación máximo-verosímil proporciona mejores índices de ajuste y estimaciones de parámetros más precisas que GLS en condiciones en las que los modelos no están adecuadamente especificados.

Como decíamos, tanto ML como GLS y WLS son considerados métodos asintóticamente eficientes (a diferencia de ULS), alcanzando soluciones bastante similares entre sí, aunque con ligeras diferencias en cuanto al ajuste y a los parámetros estimados. En comparación con los anteriores, el método WLS es asintóticamente eficiente para cualquier distribución de las variables observables, por lo que es el método más correcto matemáticamente.

A pesar de ello, la matriz $\mathbf{H}^{-1}$ tiende a ser inestable con muestras que no sean muy numerosas o con modelos que incluyan muchas variables (Batista y Coenders, 2000). Siguiendo a

[19] Browne, M. W. (1974). Generalized least-squares estimators in the analysis of covariance structures. South African Statistical Journal, 8,1–24.

[20] Browne, M. W. (1984). Asymptotically distribution-free methods for analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62–83.

[21] Para una revisión exhaustiva de las diferencias en la formulación matemática de ML, GLS y WLS, así como de las diferencias en el tipo de solución proporcionada por cada método, se recomienda consultar el trabajo de Olsson, Foss, Troye y Howell (2000).

Jöreskog (en Cudeck y MacCallum, 2007; pp. 47-77), los métodos de estimación ULS, ML y GLS pueden ser considerados como casos particulares del método general de WLS. En función de **S** y $\hat{\boldsymbol{\Sigma}}$ la ecuación [6] puede formularse de la siguiente manera:

$$F_{WLS} = \frac{1}{2} tr\left[(\mathbf{S}-\hat{\boldsymbol{\Sigma}})\mathbf{W}\right]^2 \qquad [7]$$

En la ecuación anterior **W** representa cualquier matriz de ponderación definida positiva ya que WLS es un método asintóticamente libre de distribución, y el resto de métodos tratan de minimizar *F* sustituyendo en la ecuación [7] las matrices de ponderación correspondientes mencionadas más arriba[22].

En resumen, el método ULS es adecuado para estimar parámetros a partir de las matrices de correlaciones cuando las variables son continuas, y no necesita asumir distribuciones normales, aunque precisamente por ello no permite realizar inferencias estadísticas sobre los resultados. El método WLS no parece recomendable dadas las fuertes restricciones que impone sobre el número de variables observables y sobre la muestra. Finalmente, los dos métodos *Normal Theory* (NT) son bastante similares aunque en la práctica se suele preferir ML en lugar de GLS.

En situaciones en las que las variables observables a analizar son de tipo ordinal o categórico, diversos autores (Bock y Lieberman (1970), Christoffersson (1975), y especialmente, Muthén (1978, 1984) y Jöreskog (1990, 1994)), han desarrollado métodos de estimación que suponen una generalización de los métodos utilizados con variables continuas. En los modelos factoriales ordinales y categóricos se asume que para cada variable observable categórica (x), subyace una variable de respuesta latente (x^*) que es continua. La relación entre la variable observable y su continuo de respuesta latente se establece definiendo la existencia teórica de puntos de corte en el continuo latente o *umbrales*:

$$x_i = q \quad si \quad \tau_{i,q} < x_i^* < \tau_{i,q+1} \qquad [8]$$

En donde q es la categoría de respuesta de la variable observable i, y los umbrales (τ) definen los márgenes en el continuo latente que permiten establecer correspondencias entre los valores de x^* y los valores de x[23]. Dentro de estos modelos se asume, por tanto, que cuando un individuo selecciona una categoría de respuesta, el valor de su respuesta latente x^* se

[22] El tratamiento de ML, GLS y ULS como casos particulares de WLS también puede consultarse en Bentler, P. M. (2006). EQS 6 Structural Equations Program Manual. Encino, CA: Multivariate Software, Inc. pp. 130.

[23] En el caso de variables observables ordinales se debe cumplir adicionalmente la siguiente condición: $-\infty = \tau_0 < \tau_1 < \tau_2 \ldots < \tau_{m-1} < \tau_m = \infty$, en donde m es el número de categorías ordenadas de las variables x.

encontrará dentro de un rango de valores establecido por el umbral inmediatamente inferior asociado al valor observado en *x* y por el umbral inmediatamente superior.

En estos modelos, además, se asume que los factores están vinculados con las respuestas latentes *x**, que tanto los factores como los errores se distribuyen normalmente (términos **ξ** y **δ** de la ecuación [2]), que la media de los factores y de los errores es cero y que no existe correlación entre ambos.

El proceso de estimación que siguen estos modelos factoriales establecen tres etapas sucesivas: (1) se estiman los umbrales para cada variable observable *x*, (2) se estiman las correlaciones policóricas (o tetracóricas si las variables observables son dicotómicas) para cada par de variables utilizando los umbrales iniciales, y (3) se estiman los parámetros que minimizan *F* solo que los residuos se calculan en base a la matriz de correlaciones estimada en el paso 2. El proceso de minimización de la función de discrepancia sigue la misma lógica expuesta anteriormente (ecuaciones [5] y [6]), dando lugar a los distintos métodos de estimación para modelos ordinales y categóricos dependiendo del tipo de matriz de pesos **W** que se utilice:

$$ULS: \quad \mathbf{W}=\mathbf{I} \qquad [9]$$

$$GLS: \quad \mathbf{W}=\hat{\boldsymbol{\Gamma}}^{-1} \qquad [10]$$

$$DWLS: \quad \mathbf{W}=\left(diag\left(\hat{\boldsymbol{\Gamma}}\right)\right)^{-1/2} \qquad [11]$$

En el caso de ULS, la diferencia entre la ecuación [5] y la [9] es que en la segunda el vector de residuos **d** se calcula comparando la matriz de correlaciones policóricas-tetracóricas estimadas en el paso 2 en lugar de la matriz observada.

El término $\hat{\boldsymbol{\Gamma}}$ hace referencia a la matriz de varianzas-covarianzas asintótica que se estima a partir de la matriz de correlaciones policóricas-tetracóricas, lo que da origen al método WLS generalizado por Muthén (1978, 1984) para el análisis de variables observables ordinales y dicotómicas. Por su parte, Muthén, du Toit y Spisic (1997) desarrollaron un método WLS robusto (RWLS, también denominado *Diagonal Weighted Least Squares* - DWLS), que utiliza solamente las varianzas estimadas de la matriz de correlaciones para minimizar *F* (la diagonal de la matriz). Su funcionamiento, a diferencia de WLS para variables continuas (WLS completo o *full WLS*), es adecuado incluso con muestras pequeñas, ya que permite obtener soluciones convergentes incluso cuando la matriz **W** es definida no positiva (Flora y Curran, 2004)[24].

[24] Esto último puede aplicarse también a ULS.

Para finalizar, los métodos de estimación para modelos ordinales o categóricos plantean la necesidad de introducir el debate que existe en torno a la asunción de continuos latentes lineales frente a la aplicación de modelos no lineales. Este debate gira en torno a la legitimidad de utilizar AF cuando las variables observables proveen a la investigación de datos de carácter categórico (más concretamente, cuando dichas variables son de carácter dicotómico).

La primera perspectiva considera que el AF resulta un modelo legítimo para obtener las saturaciones de las variables continuas subyacentes en los factores cuando se utilizan correlaciones policóricas o tetracóricas, ya que se asume que la discretización de dichas variables refleja un continuo de respuesta subyacente. Es decir, a partir de datos o de respuestas ordinales o categóricas, se estima una matriz compuesta por todos los pares de correlaciones policóricas o tetracóricas *bajo la asunción de un continuo latente* (*x**), y a partir de dicha matriz de correlaciones se realiza el AF. Esta perspectiva aparece cristalizada cuando se trabaja con datos categóricos en el programa LISREL de Jöreskog y Sörbom (2006) y en el programa MPLUS de Muthén[25], principalmente. La segunda perspectiva se fundamenta en la inadecuación del AF cuando las variables observables son de carácter discreto y cuyo principal representante es McDonald (1982). Dentro de este enfoque se considera que, aunque el factor o variable latente constituye un continuo ilimitado, una dicotomía tiene un rango de valores necesariamente limitado entre “0” y “1”, por lo que inevitablemente en muchos casos las relaciones entre el factor o factores y las variables observables serán no lineales.

Dentro de la perspectiva no lineal, el desarrollo de los modelos se encuentra más estrechamente relacionado con las funciones de distribución acumulada que se representan habitualmente mediante las curvas características de los ítems (CCI). Más concretamente, la función de ojiva normal (Lawley, 194·; Lord, 1952) y la función logística (Birbaum, 1968)[26]. Ambas funciones de distribución acumulada son las contrapartidas *no normales* del modelo de un factor común de Spearman, al tiempo que suponen una transformación no lineal del mismo. Esta perspectiva aparece cristalizada principalmente en el programa NOHARM de Fraser y McDonald[27].

En este apartado se han mostrado los principales métodos de estimación de parámetros que se utilizan en los modelos factoriales, siguiendo el desarrollo de distintos autores posicionados dentro del primer enfoque, aquel que asume la existencia de continuos subyacentes de tipo

[25] Muthén, L.K. and Muthén, B.O. (1998-2012). Mplus User’s Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
[26] Tomado de McDonald (1982), pp. 382.
[27] Fraser, C., & McDonald, R. P. (2003). NOHARM version 3.0: A windows program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory. Welland, ON: Niagara College.

lineal. Por tanto, la lógica de la estimación de parámetros que se desarrolla en la parte empírica de este trabajo a partir de variables continuas resulta extrapolable al contexto en el que los datos son de carácter ordinal o categórico, incluyendo los datos dicotómicos. No obstante, de cara a investigaciones futuras con este tipo de datos parece necesario mencionar que dichos métodos de estimación no son los únicos métodos existentes, encontrando en el enfoque no lineal una importante vía de trabajo alternativa.

1.3. Tópicos de investigación en los estudios sobre recuperación de factores débiles

Varios de los tópicos de investigación en este contexto son comunes a los de los estudios que aplican AF. Estos tópicos comunes, como el análisis del tamaño muestral, la simulación de distintos tipos de distribución de los indicadores, o la comparación entre métodos de estimación tienen, por tanto, un carácter transversal, y los resultados previos obtenidos en los estudios que aplican AF suponen una importante base para los estudios RFD.

En este sentido, muchas de las recomendaciones que ofrecen los investigadores en AF son extensibles al contexto de los estudios RFD, ya que estos últimos son un caso particular de los primeros. Sobre los tópicos de los estudios mediante AF se pueden revisar los trabajos de Muthén (1978, 1984 y 1993), Muthén y Kaplan (1985 y 1992), Flora y Curran (2004) y Forero, Maydeu-Olivares, y Gallardo-Pujol (2009), así como las recomendaciones de Fabrigar *et al.* (1999), Batista y Coenders (2000), y Ferrando y Anguiano-Carrasco (2010).

Especialmente relevantes para el presente trabajo son las recomendaciones técnicas relacionadas con el tamaño muestral. El AF se sustenta en el grado de covariación o de correlación entre las variables observables, y estos estadísticos presentan fluctuaciones muestrales altas, por lo que las recetas habituales del tipo al menos 10 veces más sujetos que variables no tienen una base sólida, sugiriéndose muestras de 200 observaciones (como mínimo) para elaborar modelos factoriales, incluso en circunstancias óptimas.

Batista y Coenders (2000) indican que en último término el tamaño de la muestra dependerá del modelo concreto que se pretenda ajustar, aunque normalmente tamaños muestrales entre 200 y 500 suelen ser suficientes. En general, para estos autores el tamaño de la muestra deberá ser mayor cuanto menores sean los porcentajes de varianza explicada, mayor sea la multicolinealidad, menor sea el número de indicadores por factor, y mayor sea el número total de variables y parámetros (pp. 74).

Por otra parte, existen algunos tópicos que son más específicos (aunque no exclusivos) de los estudios RFD, como el análisis del nivel de carga factorial o saturación, o el número de indicadores con saturaciones bajas presentes en el modelo. Sobre el nivel de carga, Forero, Maydeu-Olivares, y Gallardo-Pujol (2009), llegan a utilizar saturaciones de 0,40 como forma de representar la aparición de FD, apoyándose en los trabajos de Briggs y MacCallum (2003) y Ximénez (2006). Este es el nivel de saturación más bajo con el que trabajaron Briggs y MacCallum en su estudio, si bien estos mismos autores comentan en sus conclusiones que, en términos aplicados, los investigadores podrían considerar que las saturaciones bajas que permiten identificar un FD pueden oscilar entre 0,20 y 0,30 (pp. 53). Por otra parte, Catell (1988) propone no interpretar cargas factoriales o saturaciones inferiores a 0,20 - 0,30. Más restrictivo es MacDonald (1985), que propone no interpretar factores que no tengan, al menos, 3 indicadores con saturaciones de 0,30 o superiores (tomado de Ferrando y Anguiano-Carrasco, 2010; pág. 27).

Los trabajos realizados por Ximénez (2006, 2007 y 2009) son más esclarecedores en este sentido puesto que esta autora llega a trabajar con saturaciones de 0,25 y 0,35. Desde la óptica con la que se ha elaborado la presente investigación, parece más lógico entender que los valores de 0,40 deberían considerarse como un techo aproximado de los niveles de saturación que deben presentar los FD, un límite entre lo que hemos definido como FD y un nivel de carga factorial intermedio o elevado.

El estudio de Briggs y MacCallum de 2003 consistió en comparar el comportamiento de ML con el método de Mínimos Cuadrados Ordinarios (*Ordinary Least Squares* - OLS), en el contexto del AFE con variables continuas y distribución normal multivariante. OLS no es propiamente un método de estimación, sino un conjunto de estimadores basados en un criterio común: minimizar la suma de cuadrados de las diferencias entre las correlaciones observadas y las reproducidas por el modelo[28]. Briggs y MacCallum mostraron que OLS supera a ML en la calidad de la recuperación de parámetros que pertenecen al FD. Estos autores, basándose en investigación previa (MacCallum y Tucker, 1991; MacCallum, Tucker y Briggs, 2001), plantearon que las diferencias en la recuperación del FD entre ambos estimadores se debe a la naturaleza del error presente en los datos.

El argumento es que cuando se utiliza F_{ML} se opera desde una asunción implícita, la de que los factores comunes ajustan exactamente en la población, por lo que todas las discrepancias

[28] Tomado de Ferrando y Anguiano-Carrasco (2010; pp. 27). Los principales métodos basados en el criterio OLS, en el contexto del AFE, son AF de ejes principales, MINRES (Harman y Jones, 1966), ULS (Jöreskog, 1977), y Residual Mínimo (Comrey, 1962). En el estudio de Briggs y MacCallum utilizan MINRES, y en términos de AFC el método de estimación coincide con el formulado en la ecuación [5] para el método ULS.

presentes en el vector **d** son error muestral. Dado que las saturaciones bajas en la población son menos estables a través de distintas muestras que las saturaciones altas (es decir, son más sensibles al error muestral), suelen generar discrepancias más elevadas al ser comparadas con los valores de la matriz empírica **S** y, como hemos visto, ML invierte en *F* los pesos resultantes de la matriz **W**, por lo que este método otorga un mayor peso a los residuos asociados a las saturaciones más altas. Esto no ocurre con OLS ya que, al no asumir ningún tipo de distribución de los datos no diferencia en la ponderación entre saturaciones altas y bajas, por lo que trata de recuperar todos estos parámetros con la misma precisión.

Los autores pusieron a prueba los resultados encontrados por Browne (1968) en los que ML superaba a OLS en la recuperación de las saturaciones poblacionales, solo que introduciendo la variante de factores débiles presentes en los modelos. La simulación llevada a cabo por Briggs y MacCallum partía de tres escenarios distintos, dependiendo de la naturaleza del error: el primero era aquel en el que los datos presentaban error en el modelo (representado por cierta falta de ajuste), el segundo presentaba error muestral, y el tercero era una combinación de ambos tipos de error[29] [30].

Para la representación de FD se utilizaron tres tipos de matrices de correlaciones, dos de ellas correspondientes a dos modelos con un factor débil (3 factores y 12 indicadores: FD compuesto por 3 indicadores con saturaciones de 0,50, y 4 factores y 16 indicadores: FD compuesto por 4 indicadores con saturaciones de 0,45), y la tercera matriz correspondiente a un modelo con dos factores débiles (4 factores y 16 indicadores: un FD compuesto por 4 indicadores con saturaciones de 0,40 y otro FD compuesto por 4 indicadores y saturaciones de 0,45). Las saturaciones del resto de factores oscilaron entre 0,70 y 0,95. El número de réplicas por condición fue de 1.000.

En la condición en la que solamente existía error en el modelo, OLS recuperó mejor los parámetros de los FD en las 3 matrices de correlaciones analizadas cuando existís una cantidad moderada de perturbación en los modelos analizados. Cuando el ajuste era adecuado, ambos estimadores recuperaron también el FD adecuadamente y de manera muy similar.

En la condición en la que solamente existía error muestral, ML y OLS recuperaron adecuadamente el FD cuando el tamaño de la muestra fue mayor (N = 500). No obstante, cuando el tamaño muestral era de N = 300 o N = 100 (es decir, cuando el error muestral era mayor), nuevamente OLS superaba a ML en la recuperación de los parámetros en los tres

[29] El error en el modelo se manipuló introduciendo el efecto de un número elevado de factores menores, con un grado de influencia pequeño sobre las variables observables. Al no identificar estos factores en el modelo final, su influencia introduce un nivel de perturbación en el modelo (MacCallum y Tucker, 1991), controlando el nivel de desajuste al establecer puntos de corte en los valores de RMSEA.

[30] El error muestral se introdujo a partir de muestras N = 100, 300 y 500, en donde se equiparó la diagonal de las matrices de correlaciones a 1 (asumiendo que no existe error en el modelo).

modelos factoriales simulados. En ambas condiciones, cuando la naturaleza del error se analizó aisladamente, los autores detectaron que el método ML producía un mayor número de soluciones con casos Heywood, lo que puede explicar en parte las diferencias encontradas en el comportamiento de ambos estimadores[31]. Cuando se comparaban resultados en aquellas situaciones en las que ni ML ni OLS producían casos Heywood, el mejor comportamiento de OLS ya no resultaba tan marcado.

En la condición mixta, la que incluye tanto error en el modelo como error muestral, los resultados mostraron un patrón similar al de las otras dos condiciones. En todas las situaciones, con o sin casos Heywood, OLS recuperó más adecuadamente los FD que ML. Por tanto, los resultados encontrados contradicen los resultados previos de Browne en situaciones en las que existe uno o varios FD en el modelo factorial, ya que OLS funciona mejor que ML recuperando factores débiles en situaciones en las que existe un nivel de error moderado, independientemente de la fuente de error.

Por su parte, los estudios de Ximénez, que son continuación del de Briggs y MacCallum, aunque dentro del contexto del AFC, se fundamentan en la comparación de ML y ULS, añadiendo nuevos tamaños muestrales (N = 1.000 y 2.000), añadiendo correlación entre factores (ρ = 0,50) y aumentando el nivel de debilidad del factor (λ = 0,25 y λ = 0,35), y en donde se elaboraron distintos modelos factoriales, incluido el de MacCallum y Tucker de 1991. El número de réplicas utilizadas también fue de 1.000 muestras por condición experimental y se simularon variables continuas con distribución normal univariada.

Los resultados de Ximénez indicaron que en la situación en la que N = 100 ULS tiende a recuperar adecuadamente el FD en muchos de los casos en los que ML falla. A medida que aumenta el tamaño muestral (a partir de N = 500) disminuye la mejoría de ULS respecto a ML, existiendo pocas diferencias entre ambos métodos. Al igual que en el estudio precedente, parece que el mejor comportamiento de ULS está relacionado con la presencia de casos Heywood, puesto que ML produce una mayor proporción de este tipo de casos (al tiempo que un mayor número de soluciones no convergentes).

La correlación entre factores mostró ser una variable que influye en la recuperación de los parámetros del FD: en aquellos modelos en los que los factores estaban correlacionados mejoraba la recuperación de FD. Además, se hace patente en estos estudios que a medida que aumenta la debilidad del FD, disminuye la calidad de la recuperación.

[31] Briggs y MacCallum detectaron que ML produce más casos Heywood en la condición de error muestral que en la de error en el modelo.

2. Método

El objetivo de la presente investigación ha sido evaluar la capacidad de recuperación de parámetros en AFC de los métodos de estimación ML y ULS, comparando entre distintas configuraciones estructurales simuladas en las que existe un FD. Además de los criterios habituales (ej. coeficiente de congruencia y RMSD), basados en los valores promedio por condición experimental, se ha tomado en cuenta como criterio la potencial significación práctica de las soluciones generadas. Para ello, se han realizado dos estudios de simulación Monte Carlo aplicando AFC en los que se han manipulado distintas fuentes de variación[32], prestando especial atención a variaciones en el nivel de debilidad del FD, variaciones del tamaño muestral, y variaciones en la correlación entre factores. Estos estudios se han elaborado con los programas PRELIS 2.0 y LISREL 8.8.

ESTUDIO 1: En el primer estudio se simuló el modelo de factor débil de MacCallum y Tucker (1991), el cual ha servido como punto de partida en diversas investigaciones posteriores como Briggs y MacCallum (2003) y Ximénez (2005, 2006). Este modelo está compuesto por 3 factores ortogonales con 12 indicadores con distribución normal univariante, tres de los cuales definen el factor débil con saturaciones de 0,50 (ver Figura 1).

La elección de este modelo se ha realizado por varios motivos: en primer lugar, se trata de un modelo que ha sido estudiado previamente, con resultados informados, lo que potencia su comparabilidad.

En segundo lugar, aun tratándose de soluciones simuladas, la estructura de este modelo puede encontrarse con frecuencia en contextos aplicados en Psicología, lo que potencia su generalización a aquellos contextos en los que se dispone de datos reales.

En tercer lugar, y más importante, los estudios previos solamente han indicado ligeras diferencias entre ML y ULS en las condiciones más desfavorables, como N = 100 o ante soluciones que presentan casos Heywood, obteniendo unos índices de recuperación aceptables en todas las condiciones analizadas. Ninguno de estos estudios se ha ocupado de analizar la significación práctica de las soluciones generadas lo que ha motivado en el presente trabajo su inclusión con el fin de profundizar en la calidad de la recuperación, así como comparar nuevamente el comportamiento de ML y ULS desde este enfoque aplicado.

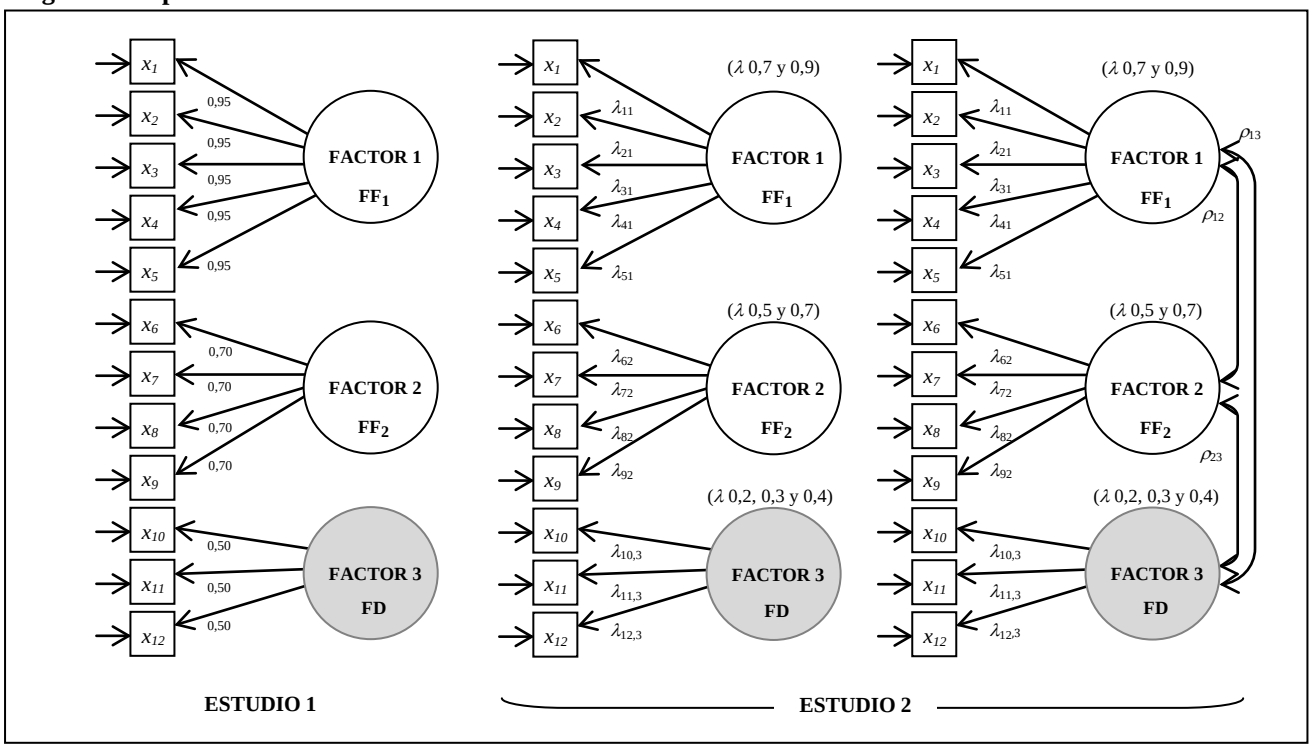
[32] Los estudios Monte Carlo, mediante la simulación de estructuras factoriales y sus soluciones en múltiples muestras replicadas, permiten al investigador la configuración controlada de escenarios complejos de análisis en los que, generalmente, interaccionan varias de las temáticas anteriores. En estos escenarios el investigador puede obtener resultados relevantes a partir de la variación conjunta y sistemática de los distintos niveles de las variables independientes manipuladas.

ESTUDIO 2: En el segundo estudio se han introducido nuevas condiciones de simulación basándonos tanto en los resultados del ESTUDIO 1 como en las investigaciones que se han realizado sobre el modelo de factor débil: niveles de carga factorial < 0,50 para los indicadores del FD (cargas más realistas desde un punto de vista aplicado), y considerar que los factores no son necesariamente ortogonales. Adicionalmente, en este estudio se analizó en profundidad la condición muestral N = 200 que parece ser crítica en la comparación entre los dos modelos de estimación propuestos, ya que existe cierto consenso en aceptar que es el número mínimo de observaciones recomendado para AF. Tanto el ESTUDIO 1 como el resto de investigaciones RFD coinciden en observar que en tamaños muestrales mayores ML y ULS alcanzan una gran convergencia en los valores de los parámetros estimados.

En relación a la falta de ortogonalidad entre factores, Ximénez (2006, 2009) utilizó correlaciones entre factores ρ = 0,50, nivel de relación que puede resultar elevado desde un punto de vista aplicado. Por esta razón, en el ESTUDIO 2 se utilizan valores de 0,30.

En relación con la carga factorial, se simuló un FD con valores de carga para sus indicadores a partir de tres condiciones: 0,20, 0,30 y 0,40. Por otra parte, respecto a los factores fuertes (FF), en los estudios previos el factor 1 (FF_1) se simuló mediante indicadores con valores λ = 0,95, y el factor 2 (FF_2) mediante indicadores con valores de λ = 0,70. Desde el punto de vista aplicado resulta poco realista considerar cargas factoriales tan elevadas, por lo que se introdujeron condiciones menos exigentes para su comparación.

Figura 1: Representación de las estructuras factoriales simuladas



El procedimiento para realizar los dos estudios de simulación expuestos más arriba se ha estructurado en torno a los siguientes pasos:

1. Las matrices de saturaciones reflejadas en la Figura 1 se han tomado como base para simular las matrices de correlaciones de cada condición experimental (10 condiciones en el ESTUDIO 1 y 48 condiciones en el ESTUDIO 2). Para cada una de estas condiciones se simularon 1.000 matrices de correlaciones o réplicas muestrales a partir de indicadores con distribución normal univariada simulados mediante el programa PRELIS 2 de Jöreskog y Sörbom (1996). Se utilizó siempre la misma semilla de aleatorización para generar las matrices de correlaciones.
2. Para cada matriz de correlaciones simulada ($\boldsymbol{\Sigma}$) se aplicó AFC con los métodos de estimación propuestos (ML y ULS), obteniendo la correspondiente matriz de parámetros estimados ($\hat{\boldsymbol{\Sigma}}$). En ambos casos se han utilizado las mismas matrices de correlaciones, esto es, la estimación de parámetros en cada condición se realizó siempre a partir del mismo conjunto de datos simulados. Esta técnica se denomina reducción de la varianza de los números aleatorios comunes (Revuelta y Ponsoda, 2003), y consiste en evitar introducir una fuente de variación no deseada debida al proceso de aleatorización de variables observables.
3. La aplicación de AFC se ha realizado mediante el programa LISREL 8.8 (Jöreskog y Sörbom, 2006), manteniendo el criterio de convergencia por defecto del programas (hasta 250 iteraciones). Se identificaron las soluciones no convergentes y las soluciones en las que se produjeron casos Heywood o valores aberrantes. Las soluciones no convergentes se eliminaron antes de proceder a analizar la RFD.

La selección de la distribución de los datos de las variables observables viene determinada por las simulaciones realizadas en los estudios RFD revisados, aunque consideramos que la lógica del diseño de investigación expuesto es extrapolable a otro tipo de distribuciones, y a comparaciones entre los métodos de estimación que se han propuesto en un caso u otro. Por otra parte, el uso de matrices de correlaciones viene determinado por la evaluación del comportamiento de ULS, ya que como se indicó en el apartado 1.2 se recomienda utilizar este método de estimación con este tipo de matrices cuando se trabaja con variables continuas.

Para evaluar la calidad en la recuperación de los parámetros del FD se han utilizado dos índices habituales en los estudios RFD, que se han operativizado como variables dependientes. Se trata del *coeficiente de congruencia* definido por Tucker (1951) y Tucker y Lewis (1973) y del índice RMSD (*Root Mean Square Deviation*), definido por Levine (1977).

El primer índice evalúa la forma y la magnitud de la saturación teórica ($\boldsymbol{\Sigma}$) respecto de la estimada ($\hat{\boldsymbol{\Sigma}}$), mientras que el segundo calcula las diferencias en magnitud en todas las direcciones. Estos índices se han calculado utilizando exclusivamente los indicadores que representan al FD. El coeficiente de congruencia oscila entre ±1, y se interpreta igual que la correlación de Pearson, y RMSD oscila entre 0 y 2 (cuanto menor sea su valor mayor será la convergencia entre los valores de $\boldsymbol{\Sigma}$ y los de $\hat{\boldsymbol{\Sigma}}$).

Coeficiente de congruencia:
$$C_k = \frac{\sum_{i=1}^{p} \lambda_{ik(t)}\lambda_{ik(e)}}{\sqrt{\left(\sum_{i=1}^{p} \lambda_{ik(t)}^2\right)\left(\sum_{i=1}^{p} \lambda_{ik(e)}^2\right)}} \quad [12]$$

RMSD:
$$RMSD_k = \sqrt{\sum_{i=1}^{p} \left(\lambda_{ik(t)} - \lambda_{ik(e)}\right)^2 / p} \quad [13]$$

Donde:

- p es el número de variables que definen el factor k.
- $\lambda_{ik(t)}$ es la saturación teórica para la variable i en el factor k.
- $\lambda_{ik(e)}$ es la saturación estimada para la misma variable en el mismo factor.

El coeficiente de congruencia indica una recuperación excelente cuando obtiene valores por encima de 0,98, una recuperación aceptable cuando obtiene valores entre 0,92 y 0,98, y una recuperación en el límite de lo aceptable cuando obtiene valores entre 0,82 y 0,92. Por su parte, RMSD indica una recuperación satisfactoria de los parámetros del FD cuando obtiene valores por debajo de 0,20.

Siguiendo el criterio de Skrondal (2000), se ha realizado un ANOVA para cada conjunto de datos generados en ambos estudios de simulación, analizando los efectos principales y las interacciones dobles de las variables manipuladas o variables independientes sobre la RFD o variables dependientes (C_k y $RMSD_k$). Dado el elevado número de réplicas realizado, se ha calculado el tamaño del efecto asociado a los valores del estadístico de contraste F de los análisis de varianza (*eta* cuadrado parcial o η^2). Los modelos ANOVA analizados son los siguientes:

- ESTUDIO 1:
$$RFD = \mu + Método\ estimación\ (M) + N + M*N \quad [14]$$

- ESTUDIO 2:
$$RFD = \mu + M + \rho + \lambda_{FD} + \lambda_{FF1} + \lambda_{FF2} + M*\rho + M*\lambda_{FD} + M*\lambda_{FF1} + M*\lambda_{FF2} + \rho*\lambda_{FD} + \rho*\lambda_{FF1} + \rho*\lambda_{FF2} + \lambda_{FD}*\lambda_{FF1} + \lambda_{FD}*\lambda_{FF2} + \lambda_{FF1}*\lambda_{FF2} \quad [15]$$

Para la evaluación de la significación práctica de las soluciones generadas en los dos estudios realizados se han recodificado los valores de *lambda* estimados para cada indicador del FD (Tabla 1). El criterio de recodificación se ha establecido mediante puntos de corte en los valores de los parámetros estimados respecto a los valores teóricos de la matriz de saturaciones **Σ**, y mediante los cuales se ha diferenciado entre *estimaciones óptimas* (± 0,05 respecto al valor teórico), *estimaciones aceptables* (entre ± 0,05 y 0,10), y *estimaciones imprecisas* (por debajo o por encima de los puntos de corte anteriores). Para el establecimiento de estos puntos de corte se han seguido las recomendaciones de Muthén y Muthén[33]. Los valores recodificados como 1 y 5, tomados en conjunto, se han utilizado para calcular el número de estimaciones imprecisas en cada solución generada, así como para analizar la proporción de *sub* y *sobre* estimaciones tomadas por separado (en Anexos: Tabla 13 y Tablas 17-19). Los valores recodificados como 2 y 4, tomados en conjunto, se han utilizado para calcular el número de estimaciones aceptables en cada solución factorial.

Tabla 1: Recodificación de la estimación de parámetros (AFC) para la evaluación de la significación práctica de las soluciones factoriales

λ_{FD} **teórico**		**Valores recodificados de los λ_{FD} estimados**				
		Estimación imprecisa (1)	Estimación aceptable (2)	Estimación óptima (3)	Estimación aceptable (4)	Estimación imprecisa (5)
ESTUDIO 1	λ = 0,5	< 0,40	0,40 – 0,45	0,45 – 0,55	0,55 – 0,60	> 0,60
	λ = 0,2	< 0,10	0,10 – 0,15	0,15 – 0,25	0,25 – 0,30	> 0,30
ESTUDIO 2	λ = 0,3	< 0,20	0,20 – 0,25	0,25 – 0,35	0,35 – 0,40	> 0,40
	λ = 0,4	< 0,30	0,30 – 0,35	0,35 – 0,45	0,45 – 0,50	> 0,50

A partir de la recodificación anterior se analizó la presencia de los distintos patrones de estimación (*óptima* – O, *aceptable* – A, *imprecisa* – I), que aparecen en las soluciones generadas en las distintas condiciones manipuladas. Los patrones posibles corresponden a todas las combinaciones sin importar el orden de soluciones O, A e I en los tres indicadores del FD (x_{10}, x_{11} y x_{12}). Para identificar dichos patrones, se han generado 3 variables nuevas que recogen el número de veces que aparecen estimaciones O, A e I en estos tres indicadores.

Tabla 2: Recuento del número de veces que aparecen los 3 tipos de estimaciones definidos en los indicadores que representan al FD ($x_{10} - x_{12}$)

Recuento	**Variable O**	**Variable A**	**Variable I**
0	Sin RFD óptima	Solución con RFD mixta	RFD aceptable
1	Solución con RFD mixta	Solución con RFD mixta	Solución con RFD mixta
2	RFD aceptable	RFD aceptable	RFD inaceptable
3	RFD óptima	RFD aceptable	RFD inaceptable

[33] Muthén, L.K. and Muthén, B.O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling, 4, 599-620.

Los patrones oscilan desde la aparición de 3 estimaciones óptimas hasta la aparición de 3 estimaciones imprecisas (ver Tabla 2). En el primer caso, la calidad global de la RFD resulta óptima ya que todos los parámetros que representan al FD se han estimado al máximo nivel de precisión, lo que confiere garantías a la investigación para realizar interpretaciones directas de la relación entre los factores y sus indicadores, además de evaluar la dimensionalidad de los datos.

En el segundo caso la calidad global de la RFD resulta inaceptable puesto que ninguno de los parámetros que representa al FD se ha estimado con la suficiente significación práctica. En este tipo de situaciones, la recomendación es evaluar la dimensionalidad del modelo sin realizar interpretaciones directas de la relación entre factores y sus indicadores.

Los casos anteriores suponen los patrones más extremos, donde el tipo de RFD desde un punto de vista práctico resulta más evidente. En el resto, el tipo de RFD se compone de situaciones mixtas ya que el grado en el que las soluciones tomadas en su conjunto resultan aceptables puede adoptar distintas formas. Por esta razón, se elaboró un paso más en el proceso de recodificación de las estimaciones obtenidas (entre paréntesis aparece la recodificación final):

- Recuento estimaciones O: 3 (1); 1 – 2 (2); 0 (3)
- Recuento estimaciones A: 0 – 1 (1); 2 – 3 (2)
- Recuento estimaciones I: 0 (1); 1 (2); 2 – 3 (3)

Los patrones de estimación finalmente analizados (Tabla 3) son los que han permitido evaluar la calidad de la RFD desde un punto de vista práctico ya que reflejan de manera combinada la precisión de la estimación de los valores *lambda* de todos los indicadores que representan al FD, aplicados a cada solución factorial concreta.

Tabla 3: Tipología de RFD en función de los patrones de estimación de los indicadores que representan al FD

I	A	O	Tipo RFD	
1	1	1	**[1] RFD óptima**: los 3 indicadores que representan al FD se han estimado con una precisión que oscila entre ± 0,05 respecto al valor teórico. INTERPRETACIÓN DIRECTA	
1	1	2	RFD aceptable: 2 estimaciones O y 1 estimación A.	**[2] RFD precisa**: sin estimaciones I y con alguna A. INTERPRETACIÓN DIRECTA
1	2	2	RFD aceptable: 2 estimaciones A y 1 estimación O.	
1	2	3	RFD aceptable: 3 indicadores con estimaciones A.	
2	1	2	RFD imprecisa: 1 est. I, 1 est. A y 1 est. O.	**[3] RFD imprecisa**: 1 estimación I, con o sin 1 O. EVALUACIÓN DIMENSIONALIDAD
2	2	3	RFD imprecisa: 1 estimación. I y 2 estimaciones. A.	
3	1	2	RFD inaceptable: est. imprecisas con o sin 1 est. O.	**[4] RFD inaceptable**: 2 – 3 estimaciones I. EVALUACIÓN DIMENSIONALIDAD
3	1	3	RFD inaceptable: est. imprecisas con o sin 1 est. A.	

Solamente se muestran los patrones I, A y O que son viables. Por ejemplo, un patrón como I(1)-A(1)-O(3) no resulta viable ya que implicaría un patrón sin estimaciones imprecisas ni óptimas, y como máximo una sola estimación aceptable.

3. Resultados

3.1. ESTUDIO 1: Modelo de MacCallum y Tucker (1991)

El diseño del ESTUDIO 1 ha consistido en la manipulación de las siguientes condiciones:

- Tamaño muestral: N=100, N=300, N=500, N=1.000 y N=2.000.
- Método de estimación: ML y ULS.
- Sin correlación entre factores.
- Se trata de un estudio 5 x 2 con 1.000 réplicas muestrales por condición, lo que produce un total de 10.000 soluciones factoriales.

En los resultados previos (*r.p.*), el número de soluciones que no convergieron tanto con ML como con ULS fue de 12 y el número de soluciones que presentaron casos Heywood fue de 65 con cada método de estimación (24 y 130 en total, respectivamente). Esta situación se produjo siempre en el caso en el que N = 100. En la replicación de elaboración propia (*e.p.*), el número de soluciones no convergentes fue de 17 en el caso de ML y de 15 en el caso de ULS, y el número de soluciones con casos Heywood fue de 50 con ML y 52 con ULS (32 y 102 en total, respectivamente). Nuevamente, esta situación se produjo en *e.p.* cuando N = 100, salvo 2 soluciones con casos Heywood cuando N = 300, una con ML y otra con ULS. La Tabla 4 muestra los promedios de los índices de RFD definidos como variables dependientes así como su desviación típica. En general, se observa que la RFD es aceptable cuando N = 100 (C_k > 0,92 y $RMSD_k$ < 0,20) y excelente cuando N ≥ 300 (C_k > 0,98 y $RMSD_k$ < 0,10). Además, los resultados alcanzados por los dos métodos de estimación (ML y ULS) son muy similares en los dos estudios realizados (*r.p.* y *e.p.*).

Tabla 4: Calidad de la RFD en el modelo de MacCallum y Tucker (1991). ESTUDIO 1.

Tipo de resultados x Tamaño de muestra		C_k				$RMSD_k$			
		ML		ULS		ML		ULS	
		$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x
Resultados previos (*r.p.* Ximénez, 2005)	N=100	,955	,058	,955	,058	,156	,111	,156	,111
	N=300	,987	,016	,987	,016	,081	,040	,081	,040
	N=500	,992	,009	,992	,009	,062	,029	,062	,029
	N=1.000	,996	,004	,996	,004	,044	,019	,044	,019
	N=2.000	,998	,002	,998	,002	,030	,013	,030	,013
	Total	,986	,032	,986	,032	,075	,071	,075	,071
Elaboración propia (*e.p.*)	N=100	,952	,060	,945	,132	,164	,114	,168	,128
	N=300	,985	,021	,985	,021	,087	,047	,087	,047
	N=500	,992	,009	,992	,009	,066	,030	,066	,030
	N=1.000	,996	,004	,996	,004	,047	,021	,047	,021
	N=2.000	,998	,002	,998	,002	,033	,014	,033	,014
	Total	,985	,033	,983	,063	,079	,074	,080	,079

Al eliminar los casos Heywood del análisis, la calidad de la RFD cuando N = 100 aumentó aproximadamente en 0,01 puntos el promedio de C_k y disminuyó aproximadamente en 0,02 puntos el promedio de $RMSD_k$ (ver Tabla 5). En Anexos, Figuras 7 y 8, se muestran los valores promedio de estos índices y su dispersión.

Tabla 5: Calidad en la RFD para N = 100 al excluir los casos Heywood en el ESTUDIO 1.

Tipo de resultados x N = 100		C_k				$RMSD_k$			
		ML		ULS		ML		ULS	
		$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x
Resultados previos (*r.p.*)	N=100	,965	,039	,965	,039	,137	,066	,137	,066
	Total	,988	,022	,988	,022	,070	,053	,070	,053
Elaboración propia (*e.p.*)	N=100	,962	,041	,955	,127	,146	,070	,149	,083
	Total	,987	,024	,986	,058	,075	,056	,075	,060

En la Tabla 6 se muestran los resultados de aplicar ANOVA para el ESTUDIO 1, en donde se aprecia que son las variaciones del tamaño muestral las que provocan un efecto estadístico, frente a las variaciones del método de estimación (ML vs. ULS) que no provocan diferencias significativas. Más concretamente, a medida que aumenta el tamaño muestral aumenta significativamente el promedio de C_k y disminuye significativamente el promedio de $RMSD_k$, siendo claramente mayor el tamaño del efecto sobre este segundo índice RFD[34].

Tabla 6: Significación estadística y tamaño del efecto (ANOVA) de las condiciones manipuladas en el ESTUDIO 1.

Efectos principales e interacciones	Resultados previos (*r.p.*)				Elaboración propia (*e.p.*)			
	C_k		$RMSD_k$		C_k		$RMSD_k$	
	p	η^2	p	η^2	p	η^2	P	η^2
Método	,050	,000	,010	,000	,162	,000	,808	,000
Muestra	,000	,142	,000	,553	,000	,154	,000	,556
Método x Muestra	,202	,000	,002	,001	,001	,001	,472	,000
TOTAL		,142		,554		,155		,556

*RFD = μ + Método estimación (M) + N + M*N*

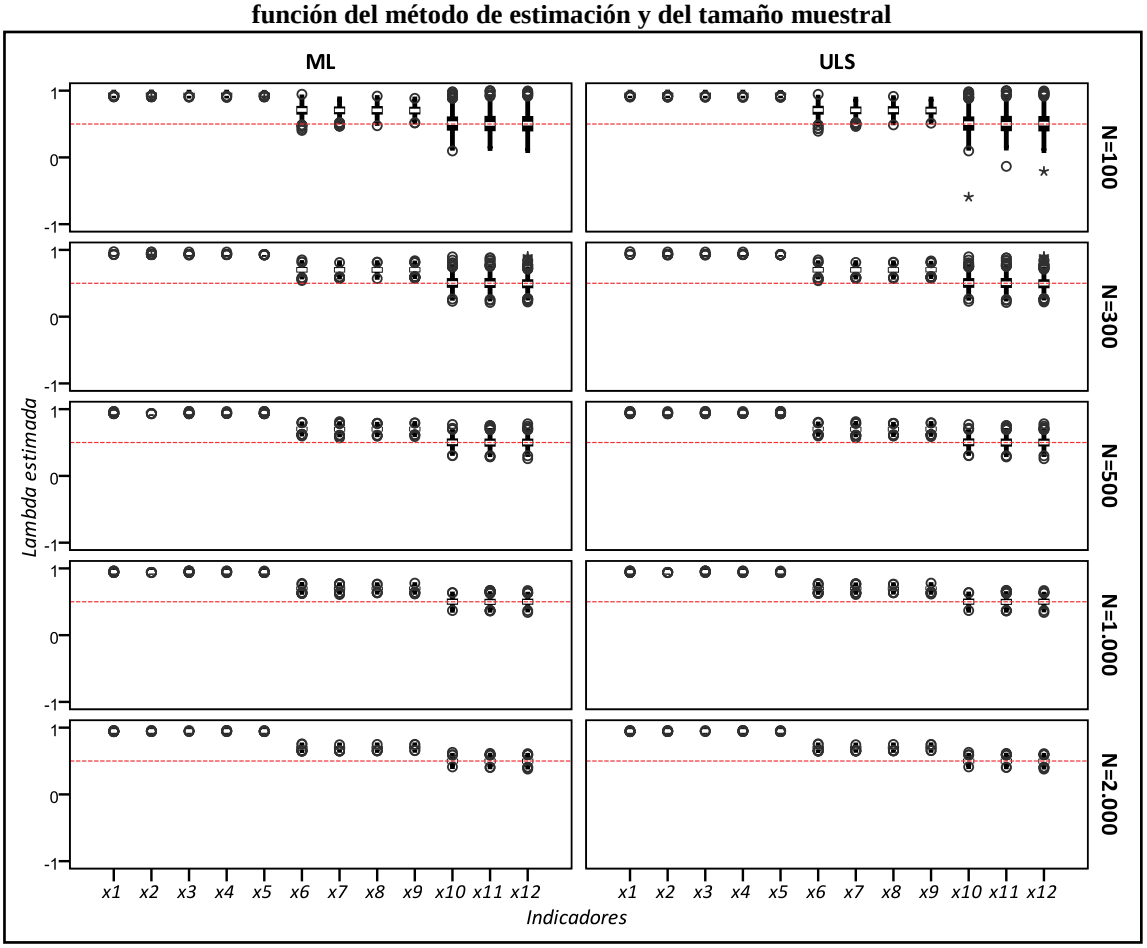
Siempre que el modelo esté correctamente especificado, con los resultados expuestos parece claro que ante un FD como el definido en el modelo de MacCallum y Tucker (3 indicadores continuos con distribución normal univariada, y con λ = 0,5), la calidad de la RFD es adecuada incluso con tamaños muestrales reducidos (N = 100), e independientemente de si el método de estimación empleado es ML o ULS. Además, el tamaño muestral influye

[34] En realidad, los *r.p.* surgen de un ANOVA que se elaboró con una VI adicional: la variable "*Modelo*", con 3 niveles: además del modelo de MacCallum y Tucker, se analizó un modelo con un solo factor y un modelo bidimensional (Ximénez, 2005). Aunque en el presente trabajo se evalúa solamente el modelo de MacCallum y Tucker, para permitir la comparación entre resultados, los resultados *e.p.* también incluyen las estimaciones para los otros dos modelos (uni y bidimensional), por lo que la Tabla 6 refleja los contrastes de efectos principales e interacción doble cuando existe una VI adicional. Por tanto, conviene señalar que los valores de significación y potencia estadística no deben tomarse directamente, sino que sirven para constatar que existen diferencias estadísticamente significativas en función del tamaño muestral y no en función del método de estimación.
Tomando en cuenta solamente estas 2 VI, el ANOVA sobre resultados *e.p.* indica una probabilidad asociada al estadístico *F* de "*Muestra*" < 0,0005 (η^2 = 0,100) y asociada al estadístico *F* de "*Método*" > 0,100 (η^2 = 0,000) para C_k y probabilidad asociada al estadístico *F* de "*Muestra*" < 0,0005 (η^2 = 0,359) y asociada al estadístico *F* de "*Método*" > 0,450 (η^2 = 0,000) para $RMSD_k$.

claramente en la recuperación puesto que a medida que aumenta éste, la precisión de las estimaciones resulta cada vez más óptima.

Ahora bien, los resultados que conforman cada condición experimental se han fundamentado en los valores promedio de RFD obtenidos para todas las soluciones, sin tener en cuenta el número de dichas soluciones que resultan aceptables, lo que tiene relevancia desde un punto de vista práctico. Una simple inspección gráfica de los parámetros estimados sugiere la necesidad de incluir en los estudios RFD la evaluación de su significación práctica. A partir de los resultados *e.p.* se ha elaborado la Figura 2 en la que se muestra la importancia del error de medida o *unicidad* introducido por cada indicador analizado y la importancia del número de observaciones en el que se sustenta la estimación. La línea punteada indica el valor teórico de los parámetros *lambda* que representan al FD (valor λ = 0,5).

Figura 2: Diagrama de cajas para los valores de los parámetros *lambda* estimados en el ESTUDIO 1 en función del método de estimación y del tamaño muestral



Para los parámetros estimados del factor fuerte 1 (FF_1), con un valor de carga factorial muy elevado (λ = 0,95), se han obtenido valores muy precisos incluso con N = 100. Se observa claramente en la Figura 2 que la variabilidad de las estimaciones para $x_1 - x_5$ fue muy escasa,

manteniéndose su valor muy próximo al valor teórico utilizado para simular las matrices de correlaciones en todos los tamaños muestrales analizados. La estimación del resto de cargas factoriales no fue tan precisa, aunque mejora a medida que aumenta el tamaño muestral. Así, cuando N = 100 los parámetros estimados para los indicadores del factor 2 (FF_2) oscilaron entre 0,40 y 0,90 aproximadamente, empezando a obtener estimaciones bastante ajustadas al valor teórico a partir de N = 300. Por su parte, los parámetros estimados de los indicadores del FD tuvieron una dispersión bastante elevada incluso con N = 500, empezando a estabilizarse en torno a su valor teórico a partir de N =1.000. Y todo esto se refleja por igual en los dos métodos de estimación utilizados.

Aplicando las recodificaciones expuestas en el apartado sobre el método de investigación (ver Tablas 1, 2 y 3), se ha encontrado la siguiente distribución de los patrones de precisión entre los indicadores del FD ($x_{10} - x_{12}$):

Tabla 7: Distribución del tipo de RFD en función de los patrones de estimación presentes en $x_{10} - x_{12}$ (ESTUDIO 1)

I	**A**	**O**	Frecuencia	% TOTAL	**Tipo RFD**
1	1	1	2.590	26,0%	RFD óptima
1	1	2	2.396	24,0%	RFD precisa
1	2	2	1.074	10,8%	
1	2	3	214	2,1%	
2	1	2	1.786	17,9%	RFD imprecisa
2	2	3	466	4,7%	
3	1	2	534	5,4%	RFD inaceptable
3	1	3	908	9,1%	
TOTAL[a]			9.968	100%	

[a] Excluidas las soluciones no convergentes

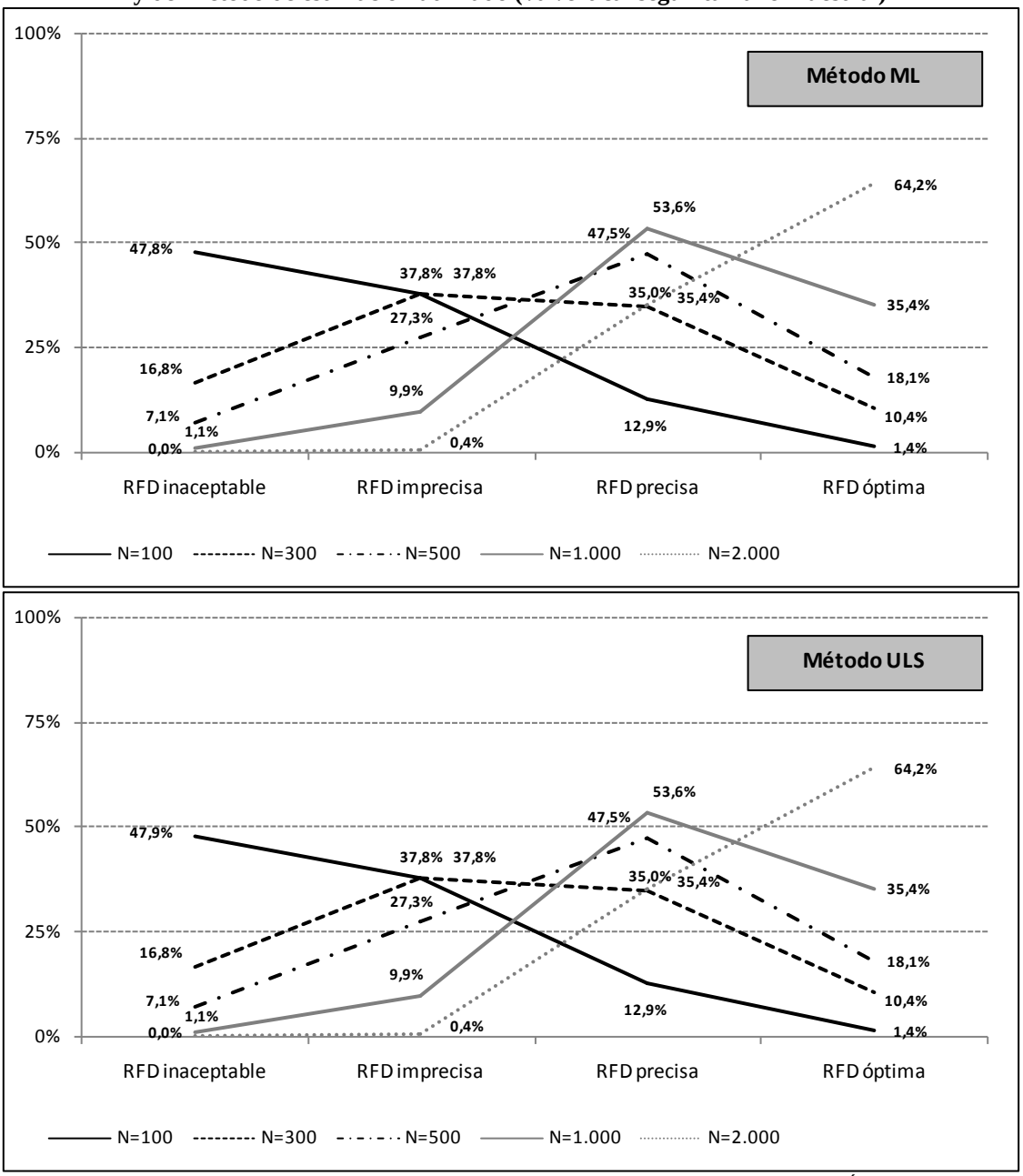
En la Tabla 7 se observa que, en total, se han ajustado 2.590 soluciones con RFD óptima (26%), 3.684 soluciones con RFD precisa (36,9%), 2.252 soluciones con RFD imprecisa (22,6%), y 1.442 soluciones con RFD inaceptable desde un punto de vista práctico (14,5%). En conjunto, en un 62,9% de las soluciones ajustadas se puede interpretar directamente la relación entre el FD y los indicadores que lo representan, frente al 37,1% de las soluciones restantes en las que solamente se debería evaluar la dimensionalidad de los datos. En el primer caso, la estimación del FD refleja con precisión la verdadera relación entre factor e indicadores, mientras que en el segundo caso la estimación no tiene significación práctica suficiente.

La distribución de estos patrones en función de las variables manipuladas en el ESTUDIO 1 (método de estimación y tamaño muestral), se representa en las Figuras 3.a y 3.b Se han incluido las soluciones con casos Heywood en la distribución del tipo de RFD, puesto que su

número es limitado y la distribución no varía sustancialmente cuando se eliminan aquellas que presentan este tipo de estimaciones (ver Anexos, Tabla 14).

Resulta evidente que a medida que aumenta el tamaño muestral de los datos analizados mejora también el tipo de RFD de las soluciones ajustadas. Así, mientras que casi el 48% de las soluciones cuando N = 100 conducen a un tipo de RFD inaceptable desde un punto de vista práctico, cuando N = 2.000 algo más del 64% de las soluciones obtuvieron un nivel de RFD óptimo.

Figura 3 (a y b): Distribución de los tipos de RFD del ESTUDIO 1 en función del tamaño muestral y del método de estimación utilizado (% vertical según tamaño muestral)



RFD óptima (26%): estimación conjunta del FD entre ± 0,05 respecto al valor teórico. INTETRPRETACIÓN DIRECTA.
RFD precisa (36,9%): sin estimaciones I, soluciones mixtas con estimaciones O y A. INTETRPRETACIÓN DIRECTA.
RFD imprecisa (22,6%): soluciones con alguna estimación I. EVALUACIÓN DIMENSIONALIDAD.
RFD inaceptable (14,5%): soluciones con 2 o 3 estimaciones I. EVALUACIÓN DIMENSIONALIDAD.

En la Figura 3 se aprecia que el tipo de RFD fue prácticamente idéntico tanto si se utiliza ML como si se utiliza ULS (la proporción de *sub* o *sobre* estimaciones fue muy similar también, ver Anexos, Tabla 13). Estos resultados son coherentes con los encontrados al analizar la calidad de la recuperación a partir de las dos VD clásicas (C_k y $RMSD_k$), si bien la evaluación de la significación práctica implementada en el presente estudio aporta información adicional.

Si atendemos a los valores promedio de los índices de RFD se observa que la recuperación de parámetros fue adecuada en todos los tamaños muestrales analizados, independientemente de si el método de estimación empleado es ML o es ULS, y mejora progresivamente a medida que aumenta el valor de N.

No obstante, las estimaciones realizadas no resultan interpretables directamente en más de un 37,1% de los casos, lo que implica un claro riesgo en contextos de investigación aplicada. Más concretamente, en torno al 85% de las soluciones generadas cuando N = 100 no deberían utilizarse para interpretar directamente las saturaciones presentes en el FD, debido a su falta de precisión (casi el 48% correspondía a un tipo de RFD inaceptable y aproximadamente el 38% restante correspondía a un tipo de RFD imprecisa). Cuando N = 300 la proporción de soluciones que no deberían interpretarse fue de casi el 55% (en torno a un 17% de soluciones con RFD inaceptable y en torno a un 38% de soluciones con RFD imprecisa). Incluso cuando N = 500 la proporción de este tipo de soluciones fue de un 24%, y cuando N = 1.000 fue de un 11%, aproximadamente.

Como puede apreciarse, la significación práctica de las estimaciones no está garantizada aunque los valores de C_k y $RMSD_k$ se encuentren dentro de los criterios considerados como aceptables. En otras palabras, aplicar los criterios anteriores no resulta un procedimiento suficientemente riguroso para detectar desviaciones en la recuperación del factores débiles, lo que puede tener una gran repercusión en estudios aplicados. De hecho, los valores promedio de estas dos VD reflejaron niveles adecuados incluso cuando el tipo de RFD es inaceptable (a excepción de C_k = 0,905 con ULS cuando N = 100, y de $RMSD_k$ > 0,20 con ML y ULS cuando N = 100). Estos valores promedios se pueden consultar en los Anexos, Tabla 15.

Los resultados encontrados indican que, al aumentar el tamaño muestral de los datos analizados se reduce el riesgo de interpretar saturaciones poco precisas, aunque los criterios de tamaño mínimo en torno a las 200 observaciones parecen válidos únicamente en modelos en los que no existen factores débilmente representados. Esta cuestión queda pendiente de una investigación más exhaustiva que será realizada en el ESTUDIO 2.

No obstante, puesto que la precisión de las estimaciones es sensible al tamaño muestral, conviene matizar los resultados anteriores. Las estimaciones definidas como óptimas (O) o

como aceptables (A) no presentan variaciones importantes entre un tamaño muestral y otro al encontrarse dentro de intervalos muy concretos, pero ¿las estimaciones definidas como imprecisas (I) reflejan el mismo grado de *imprecisión* de un tamaño muestral a otro? Lo esperable es encontrar un mayor nivel de imprecisión cuando el tamaño muestral sea más pequeño, lo que afectará a su vez a la *cualidad* del tipo de RFD. En la Tabla 8 se muestran las estimaciones promedio y su dispersión para el grupo de estimaciones definidas como imprecisas, diferenciando entre *sub* y *sobre* estimaciones.

Tabla 8: Estadísticos descriptivos de los valores *lambda* estimados para el FD a partir de los puntos de corte establecidos para estimaciones imprecisas (I). ESTUDIO 1

Precisión de las estimaciones		ML						ULS					
		x_{10}		x_{11}		x_{12}		x_{10}		x_{11}		x_{12}	
		$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x
N = 100	**[1]**	,298	,080	,311	,074	,309	,071	,287	,123	,301	,115	,298	,106
	[5]	,750	,161	,754	,173	,742	,225	,750	,161	,760	,201	,746	,235
N = 300	**[1]**	,357	,037	,352	,045	,355	,039	,357	,037	,352	,045	,355	,039
	[5]	,657	,053	,657	,056	,668	,098	,657	,053	,657	,056	,668	,098
N = 500	**[1]**	,374	,021	,369	,027	,372	,025	,374	,021	,369	,027	,372	,025
	[5]	,632	,031	,637	,035	,636	,037	,632	,031	,637	,035	,636	,037
N = 1.000	**[1]**	,388	,010	,384	,011	,384	,013	,388	,010	,384	,011	,384	,013
	[5]	,611	,011	,627	,023	,618	,019	,611	,011	,627	,023	,618	,019
N = 2.000	**[1]**	---	---	---	---	,379	---	---	---	---	---	,379	---
	[5]	,630	---	,614	---	,610	---	,630	---	,614	---	,610	---

[1] Estimación imprecisa subestimada: $\lambda < 0{,}40$.
[5] Estimación imprecisa sobreestimada: $\lambda > 0{,}60$.

Cuando N = 100, el conjunto de estimaciones I por debajo del valor teórico λ = 0,50 (en adelante I^-) se encontraba en torno al valor promedio 0,30. Dentro del mismo tamaño muestral, el conjunto de estimaciones I por encima del valor teórico (en adelante I^+) se encontraba en torno al valor promedio 0,75. Estos resultados sugieren realizar una evaluación del modelo exclusivamente desde el punto de vista de la dimensionalidad, evitando la interpretación directa de la relación entre indicadores y factores en este tipo de situaciones. Una estimación I^- puede inducir a pensar que la verdadera relación entre indicador y factor es bastante más débil de lo que es en realidad, y una estimación I^+, al contrario, puede inducir a una interpretación demasiado optimista. Nótese que el promedio de I^+ cuando N = 100 se encuentra por encima del valor teórico de las saturaciones del factor 2 (λ = 0,70).

Por otro lado, a medida que aumentó el tamaño muestral más se aproximó el valor promedio estimado al punto de corte establecido para la recodificación (0,40 para I^- y 0,60 para I^+), al tiempo que disminuyó considerablemente la dispersión de estas estimaciones. Por lo tanto, es necesario tomar con cautela los resultados anteriores relativos a la significación práctica de las estimaciones puesto que una RFD de tipo imprecisa o inaceptable no resulta igualmente inadecuada ante distintos tamaños muestrales.

3.2. ESTUDIO 2: Variación del factor débil y correlación entre factores

El diseño del ESTUDIO 2 ha consistido en la manipulación de las siguientes condiciones:

- Tamaño muestral: N=200.
- Método de estimación: ML y ULS.
- $\rho = 0$ y $\rho = 0{,}30$.
- Nivel de carga FD: 0,20, 0,30 y 0,40.
- Nivel de carga FF_1: 0,70 y 0,90.
- Nivel de carga FF_2: 0,50 y 0,70.
- Se trata de un estudio 2 x 2 x 3 x 2 x 2 con 1.000 réplicas muestrales por condición, lo que produce un total de 48.000 soluciones factoriales.

El número de soluciones no convergentes dentro del ESTUDIO 2 ascendía a 5.971 (12,4% del total de soluciones generadas mediante AFC, 2.993 con ML y 2.978 con ULS). El elevado número de este tipo de soluciones, en comparación con el ESTUDIO 1, se debe a que el tamaño muestral se ha fijado para todas las condiciones analizadas en N = 200, y a que el nivel de debilidad del factor 3 es mayor.

De hecho, cuanto mayor es el nivel de debilidad mayor ha sido el número de soluciones no convergentes que se producen al ajustar los modelos. Cuando los indicadores que representan al FD tienen saturaciones teóricas iguales a 0,20, el número de soluciones no convergentes fue de 3.587 (60,1%), 1.793 de las cuales se obtienen mediante ML y 1.794 mediante ULS. Para las saturaciones teóricas de 0,30, se contabilizaron 2.012 soluciones no convergentes (33,7%), 1.011 con ML y 1.001 con ULS, y para las saturaciones teóricas de 0,40 el número se redujo a 372 soluciones no convergentes (6,2%), 189 con ML y 183 con ULS. Todas estas soluciones se encontraban presentes cuando la correlación entre factores es nula ($\rho = 0$), y se distribuyeron de manera muy similar independientemente del nivel de carga factorial manipulado para el factor 1 y el factor 2.

Una vez eliminadas las soluciones no convergentes, se contabilizaron 1.226 casos Heywood, lo que supone un 2,9% de las soluciones restantes (418 cuando $\lambda_{FD} = 0{,}20$, 465 cuando $\lambda_{FD} = 0{,}30$, y 343 cuando $\lambda_{FD} = 0{,}40$). En función del método de estimación se contabilizaron 598 casos Heywood con ML y 628 con ULS. La aparición de casos Heywood, como ocurre con las soluciones no convergentes, solamente se reflejó cuando la correlación entre factores es nula, resultando independiente del nivel de carga de FF_1 y FF_2.

Respecto a la recuperación de parámetros alcanzada tanto por ML como por ULS, en la Tabla 9 se muestran los promedios y desviaciones típicas de los dos índices RFD analizados. Se puede apreciar que la RFD fue aceptable cuando λ_{FD} = 0,40 (C_k > 0,92 y $RMSD_k$ < 0,20). Nuevamente, la recuperación fue bastante similar al comparar las soluciones generadas mediante ML con las producidas mediante ULS.

Tabla 9: Calidad en la RFD ante variaciones del factor débil, correlación entre factores y variaciones en FF_1 y FF_2

λ_{FD} x ρ x λ_{FF1} xλ_{FF2}			C_k				$RMSD_k$			
			ML		ULS		ML		ULS	
			$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x
λ_{FD} = 0,2	ρ = 0	λ_1=0,7 y λ_2=0,5	**,644**	,428	**,604**	,483	**,285**	,194	**,295**	,205
		λ_1=0,7 y λ_2=0,7	**,649**	,422	**,608**	,479	**,308**	,658	**,296**	,221
		λ_1=0,9 y λ_2=0,5	**,643**	,431	**,568**	,528	**,284**	,188	**,295**	,189
		λ_1=0,9 y λ_2=0,7	**,645**	,429	**,601**	,488	**,281**	,182	**,295**	,197
		Total	**,645**	,427	**,595**	,495	**,290**	,367	**,296**	,203
	ρ = 0,3	λ_1=0,7 y λ_2=0,5	**,681**	,401	**,688**	,387	,192	,091	,179	,078
		λ_1=0,7 y λ_2=0,7	**,701**	,387	**,717**	,351	,181	,086	,169	,072
		λ_1=0,9 y λ_2=0,5	**,712**	,364	**,693**	,379	,181	,082	,175	,074
		λ_1=0,9 y λ_2=0,7	**,731**	,342	**,719**	,349	,172	,077	,167	,070
		Total	**,706**	,374	**,704**	,367	,181	,085	,172	,074
λ_{FD} = 0,3	ρ = 0	λ_1=0,7 y λ_2=0,5	**,853**	,255	**,842**	,287	**,223**	,303	**,219**	,195
		λ_1=0,7 y λ_2=0,7	**,849**	,265	**,832**	,316	**,234**	,528	**,222**	,199
		λ_1=0,9 y λ_2=0,5	**,858**	,240	**,815**	,354	**,212**	,186	**,228**	,208
		λ_1=0,9 y λ_2=0,7	**,859**	,232	**,820**	,342	**,213**	,187	**,227**	,206
		Total	**,855**	,248	**,827**	,326	**,220**	,332	**,224**	,202
	ρ = 0,3	λ_1=0,7 y λ_2=0,5	**,863**	,268	**,868**	,202	,158	,092	,158	,074
		λ_1=0,7 y λ_2=0,7	**,879**	,233	**,881**	,173	,149	,083	,152	,068
		λ_1=0,9 y λ_2=0,5	**,896**	,148	**,868**	,192	,147	,069	,159	,071
		λ_1=0,9 y λ_2=0,7	**,903**	,130	**,882**	,164	,142	,065	,152	,067
		Total	**,885**	,203	**,875**	,183	,149	,078	,155	,070
λ_{FD} = 0,4	ρ = 0	λ_1=0,7 y λ_2=0,5	,928	,164	,935	,108	,161	,157	,159	,156
		λ_1=0,7 y λ_2=0,7	,937	,096	,928	,162	,156	,139	,163	,166
		λ_1=0,9 y λ_2=0,5	,921	,196	,928	,161	,163	,161	,161	,159
		λ_1=0,9 y λ_2=0,7	,933	,131	,925	,173	,160	,154	,164	,167
		Total	,930	,151	,929	,153	,160	,153	,162	,162
	ρ = 0,3	λ_1=0,7 y λ_2=0,5	,932	,225	,947	,099	,130	,098	,133	,064
		λ_1=0,7 y λ_2=0,7	,941	,195	,951	,075	,124	,089	,130	,060
		λ_1=0,9 y λ_2=0,5	,959	,047	,945	,084	,118	,057	,138	,063
		λ_1=0,9 y λ_2=0,7	,959	,074	,949	,074	,116	,056	,134	,060
		Total	,947	,156	,948	,084	,122	,077	,134	,062

Se han resaltado en negrita aquellos promedios RFD que se encuentran dentro de los valores considerados como inadecuados.

Cuando los factores están correlacionados mejora ligeramente la recuperación promedio de los parámetros estimados, especialmente en el caso del índice $RMSD_k$, con valores que cumplen el criterio de aceptabilidad (< 0,20) en todos los niveles de debilidad definidos. En Anexos, Figuras 11 a 16, se muestran los valores promedio de las dos VD y su dispersión.

El nivel de carga de los factores fuertes (FF_1 y FF_2), produjo ligeras diferencias en la recuperación, aunque bastante moderadas: se observa que cuando esta carga es mayor, mejora

ligeramente la recuperación en ambos índices en todas las condiciones combinadas entre nivel de FD y correlación entre factores.

Las Figuras 4 y 5 permiten una comparación más detallada de las diferencias existentes en los promedios de RFD por condición analizada. En estas figuras se aprecian las ligeras diferencias comentadas respecto a la mejor recuperación cuando los factores están correlacionados y, especialmente, cuando aumenta el valor teórico de λ_{FD}. Estos resultados se han elaborado incluyendo los casos Heywood presentes en las soluciones factoriales, si bien no difieren especialmente de los resultados cuando se eliminan este tipo de soluciones (ver Anexos, Tabla 16 y Figuras 9 y 10). La excepción se produce para el promedio de $RMSD_k$ cuando λ_{FD} = 0,30 y ρ = 0 que cumple el criterio < 0,20 cuando se eliminan los casos Heywood. Además, el método ULS parece tener un comportamiento menos adecuado que ML al recuperar los parámetros cuando λ_{FD} = 0,20 y ρ = 0.

Figura 4: Promedio de C_k por condición del ESTUDIO 2

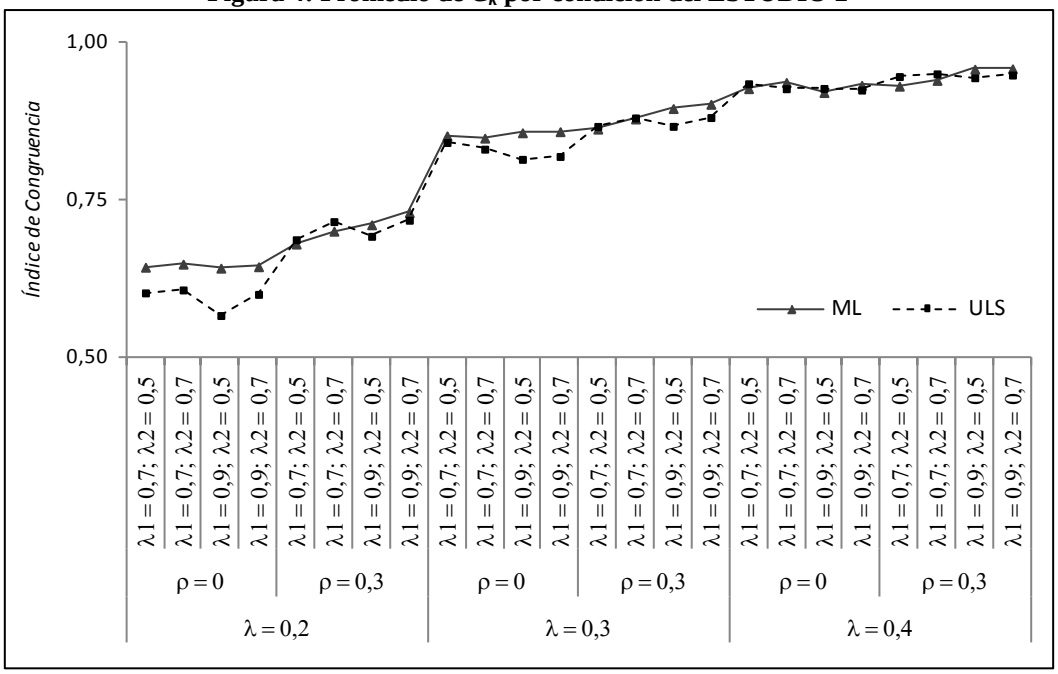
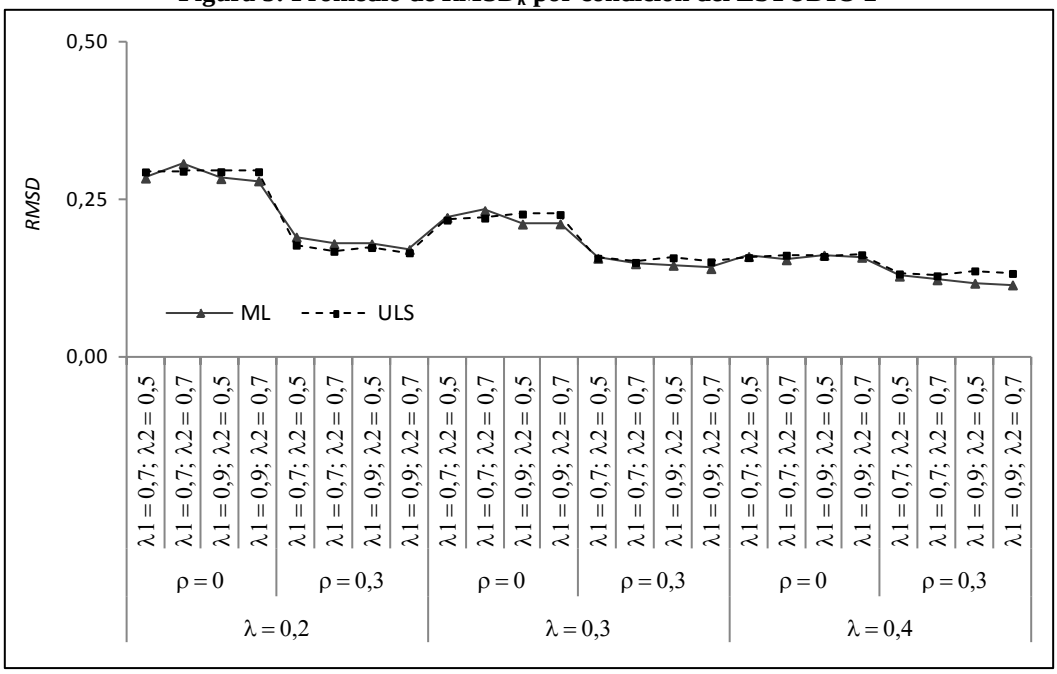


Figura 5: Promedio de $RMSD_k$ por condición del ESTUDIO 2



En la Tabla 10 se muestran los resultados de aplicar ANOVA para todos los efectos principales e interacciones dobles definidos en el ESTUDIO 2. Tanto si se incluyen las soluciones que contienen casos Heywood como si éstas se eliminan del análisis, el nivel de debilidad del FD es la VI que presenta un mayor grado de significación estadística (η^2 = 0,141 y η^2 = 0,140 para C_k, y η^2 = 0,044 y η^2 = 0,111 para $RMSD_k$, respectivamente).

Tabla 10: Significación estadística y tamaño del efecto (ANOVA) de las condiciones manipuladas en el ESTUDIO 2

Efectos principales e interacciones	**Con casos Heywood**				**Sin casos Heywood**			
	C_k		$RMSD_k$		C_k		$RMSD_k$	
	p	η^2	*p*	η^2	*p*	η^2	*p*	η^2
Método	,000	,001	,089	,000	,000	,001	,000	,000
ρ	,000	,007	,000	,044	,000	,003	,000	,033
λ_{FD}	,000	,141	,000	,044	,000	,140	,000	,111
λ_{FF1}	,203	,000	,027	,000	,290	,000	,145	,000
λ_{FF2}	,001	,000	,165	,000	,001	,000	,000	,000
Método x ρ	,000	,000	,861	,000	,000	,000	,733	,000
Método x λ_{FD}	,001	,000	,020	,000	,000	,000	,000	,001
Método x λ_{FF1}	,000	,000	,001	,000	,000	,000	,000	,000
Método x λ_{FF2}	,977	,000	,976	,000	,957	,000	,372	,000
ρ xλ_{FD}	,000	,002	,000	,010	,000	,002	,000	,018
ρ xλ_{FF1}	,000	,000	,594	,000	,000	,000	,000	,000
ρ xλ_{FF2}	,096	,000	,008	,000	,094	,000	,001	,000
λ_{FD} xλ_{FF1}	,971	,000	,289	,000	,986	,000	,666	,000
λ_{FD} xλ_{FF2}	,073	,000	,959	,000	,075	,000	,194	,000
λ_{FF1} xλ_{FF2}	,793	,000	,623	,000	,744	,000	,985	,000
TOTAL		,151		,098		,146		,163

$$RFD = \mu + M + \rho + \lambda_{FD} + \lambda_{FF1} + \lambda_{FF2} + M*\rho + M*\lambda_{FD} + M*\lambda_{FF1} + M*\lambda_{FF2} + \rho*\lambda_{FD} + \rho*\lambda_{FF1} + \rho*\lambda_{FF2} + \lambda_{FD}*\lambda_{FF1} + \lambda_{FD}*\lambda_{FF2} + \lambda_{FF1}*\lambda_{FF2}$$

Por su parte, la correlación entre factores también presenta un cierto efecto estadístico, aunque bastante más moderado (η^2 = 0,007 y η^2 = 0,003 para C_k, y η^2 = 0,044 y η^2 = 0,033 para $RMSD_k$, respectivamente). El resto de VI no presentaron un efecto estadísticamente suficiente sobre la RFD promedio. A la luz de estos resultados, al aumentar el grado de debilidad del factor 3 del modelo de MacCallum y Tucker, y para el tamaño muestral N = 200, se observa que la recuperación de parámetros disminuye en calidad independientemente del nivel de saturación de FF_1 y FF_2 y del método de estimación utilizado.

Aquellas soluciones en las que se analizan modelos con factores correlacionados (ρ = 0,3) producen una ligera mejoría en la RFD promedio, especialmente en el índice $RMSD_k$ y cuando los λ_{FD} teóricos corresponden a los valores 0,20 y 0,30. No obstante, el efecto principal más importante correspondía al nivel de debilidad del factor 3, relegando la correlación entre factores a un efecto menor sobre la variabilidad de los datos.

Continuando con el esquema de trabajo propuesto para el ESTUDIO 1, se comentan los resultados alcanzados tras evaluar la significación práctica de los parámetros estimados. Dado que la calidad de la recuperación no difiere sustancialmente al contemplar o no los casos Heywood[35], se ha decidido identificar los patrones de precisión y los tipos de RFD considerando las 42.029 soluciones totales (excluidas las soluciones no convergentes). Además, los valores aberrantes pueden producirse en los parámetros de cualquiera de los 12 indicadores presentes en los modelos (aunque con mayor probabilidad en los indicadores con mayor unicidad), y de coincidir con alguno de los indicadores del FD pueden ser considerados como una estimación imprecisa más, tanto I^- como I^+, con el propósito de analizar los patrones de precisión presentes en $x_{10} - x_{12}$.

Tabla 11: Distribución del tipo de RFD en función de los patrones de estimación presentes en $x_{10} - x_{12}$ (ESTUDIO 2)

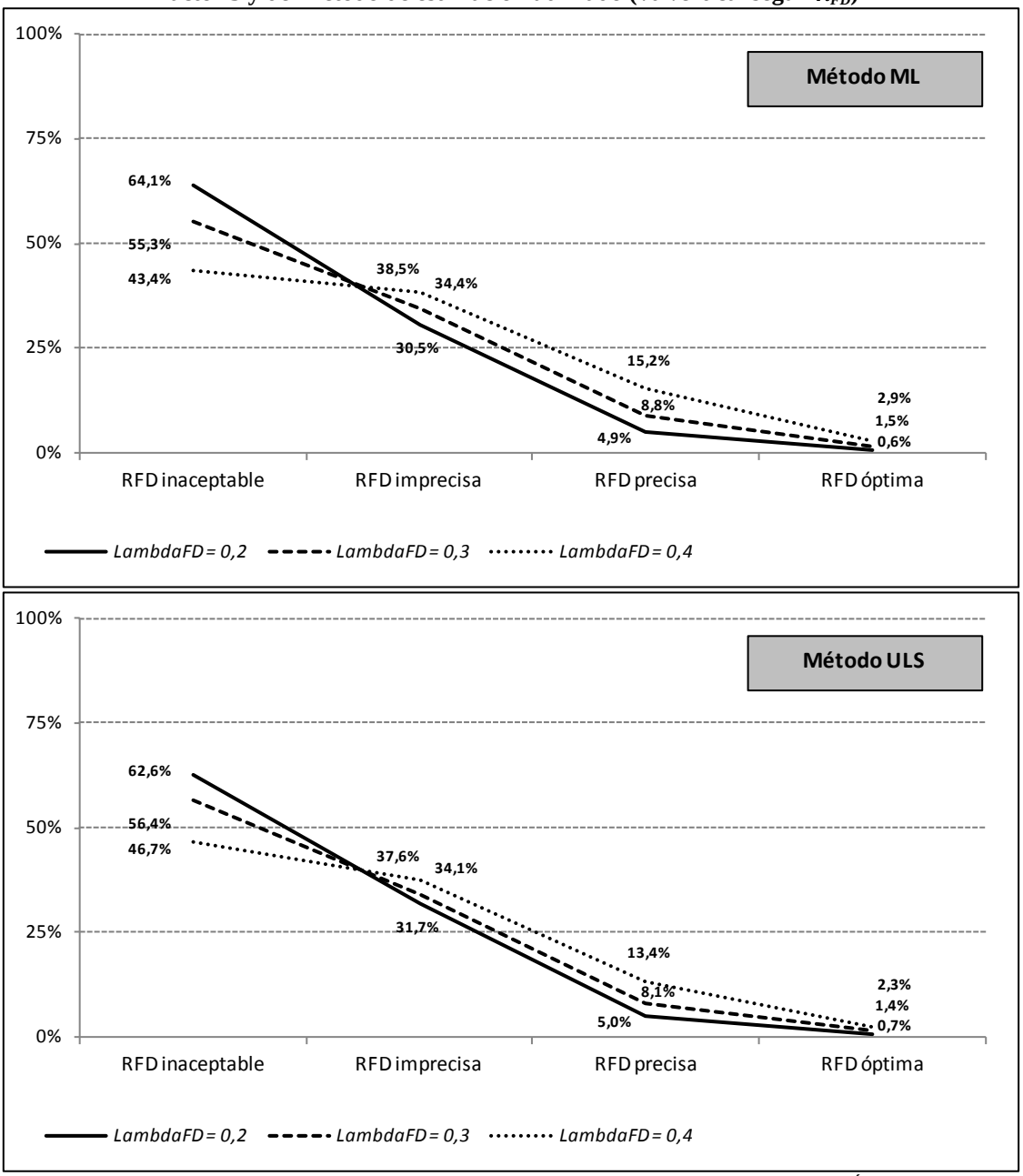
I	A	O	λ_{FD} = 0,20		λ_{FD} = 0,30		λ_{FD} = 0,40		Tipo RFD
			Frecuencia	% vertical	Frecuencia	% vertical	Frecuencia	% vertical	
1	1	1	78	0,6%	204	1,5%	408	2,6%	RFD óptima (1,7%)
1	1	2	277	2,2%	554	4,0%	1.147	7,3%	RFD precisa (9,6%)
1	2	2	254	2,0%	475	3,4%	788	5,0%	
1	2	3	84	0,7%	152	1,1%	297	1,9%	
2	1	2	2.965	23,9%	3.809	27,2%	4.624	29,6%	RFD imprecisa (34,6%)
2	2	3	891	7,2%	986	7,0%	1.319	8,4%	
3	1	2	2.621	21,1%	2.722	19,5%	2.414	15,4%	RFD inaceptable (54,1%)
3	1	3	5.243	42,2%	5.086	36,4%	4.631	29,6%	
TOTAL[a]			12.413	29,5%	13.988	33,3%	15.628	37,2%	
			42.029	100%					

[a] Excluidas las soluciones no convergentes.

[35] La calidad RFD es bastante similar al comparar soluciones con y sin casos Heywood como consecuencia de que el número de soluciones que presentan este tipo de casos se distribuye de forma homogénea en los tres niveles de debilidad del FD.

Al variar la debilidad del FD el número de soluciones cuya RFD es imprecisa e, incluso, inaceptable, aumentó considerablemente respecto a los resultados observados en el ESTUDIO 1 (ver Tabla 11). Cuando λ_{FD} = 0,50 estos dos tipos de RFD se encontraban presentes en el 37,1% de las soluciones, mientras que en el ESTUDIO 2 ascendían a un 88,7%, lo que implica que cuanto mayor es la *unicidad* que introducen los indicadores de un factor en el modelo más difícil resulta estimar con precisión su *comunalidad*.

Figura 6 (a y b): Distribución de los tipos de RFD del ESTUDIO 2 en función del nivel de debilidad del factor 3 y del método de estimación utilizado (% vertical según λ_{FD})



RFD óptima (1,7%): estimación conjunta del FD entre ± 0,05 respecto al valor teórico. INTETRPRETACIÓN DIRECTA.
RFD precisa (9,6%): sin estimaciones I, soluciones mixtas con estimaciones O y A. INTETRPRETACIÓN DIRECTA.
RFD imprecisa (34,6%): soluciones con alguna estimación I. EVALUACIÓN DIMENSIONALIDAD.
RFD inaceptable (54,1%): soluciones con 2 o 3 estimaciones I. EVALUACIÓN DIMENSIONALIDAD.

Las Figuras 6.a y 6.b muestran la distribución del tipo de RFD en función del nivel de debilidad del FD. Cuando λ_{FD} = 0,20, en torno a un 62-64% de las soluciones generadas han obtenido una RFD de tipo inaceptable, frente al 55-57% cuando λ_{FD} = 0,30 y al 43-47% cuando λ_{FD} = 0,40. El tipo de RFD impreciso se encuentra presente en un intervalo entre el 30-32% cuando λ_{FD} = 0,20 y del 37-39% cuando λ_{FD} = 0,40. Las soluciones con RFD precisa obtuvieron su máximo entre un 13-15%, aproximadamente, cuando λ_{FD} = 0,40. Las soluciones con RFD óptima, aunque se produjeron solamente en un 1,7% del total de soluciones, parecen presentar una tendencia de aumento de tipo exponencial entre un tipo de λ_{FD} y otro. Recuérdese que en el ESTUDIO 1 este tipo de soluciones se encontraban presentes en un 26%.

Nuevamente, estos resultados divergen en cierta medida de los alcanzados a partir de los valores promedio de RFD (C_k y $RMSD_k$). Según los tipos de RFD analizados cuando N = 200, en el mejor de los escenarios (λ_{FD} = 0,40, C_k > 0,92 y $RMSD_k$ < 0,20) solamente el 16,8% de las soluciones podrían ser interpretadas directamente, resultando útiles para la evaluación de la dimensionalidad el 83,2% restante. En el resto de escenarios, cuando existe correlación entre factores, el valor promedio de $RMSD_k$ ha cumplido con el criterio utilizado en este tipo de estudios (ver Tabla 9), si bien la proporción de soluciones que sirven para realizar interpretaciones directas fue del 5,5% cuando λ_{FD} = 0,20 y del 10% cuando λ_{FD} = 0,30 (ver Anexos: Tablas 22, 23 y 24 para promedios VD por tipo de RFD). En el ESTUDIO 2 los modelos ajustados han producido un mayor número de sobreestimaciones que de subestimaciones. Por otra parte, la correlación entre factores y la variación en las cargas de los FF no produjeron variaciones respecto al tipo de RFD (ver Anexos: Tablas 17-21).

Tabla 12: Estadísticos descriptivos de los valores *lambda* estimados para el FD a partir de los puntos de corte establecidos para estimaciones imprecisas (I). ESTUDIO 2.

Precisión de las estimaciones			**ML**						**ULS**					
			x_{10}		x_{11}		x_{12}		x_{10}		x_{11}		x_{12}	
			$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x
λ_{FD} = 0,20	ρ = 0	**[1]**	,050	,075	-,116	,218	-,143	,266	-,010	,167	-,142	,235	-,149	,252
		[5]	,585	,912	,558	,280	,600	,348	,578	,372	,558	,289	,598	,340
	ρ = 0,3	**[1]**	-,032	,114	-,037	,116	-,039	,124	-,019	,102	-,027	,107	-,021	,098
		[5]	,427	,106	,417	,089	,423	,100	,421	,101	,408	,082	,419	,096
λ_{FD} = 0,30	ρ = 0	**[1]**	,072	,976	,080	,168	,089	,147	,093	,137	,058	,201	,065	,176
		[5]	,629	,293	,657	,392	,601	,234	,633	,302	,665	,409	,610	,256
	ρ = 0,3	**[1]**	,094	,110	,093	,105	,095	,103	,097	,093	,088	,094	,095	,086
		[5]	,504	,094	,493	,078	,501	,088	,509	,095	,498	,076	,507	,090
λ_{FD} = 0,40	ρ = 0	**[1]**	,213	,093	,207	,115	,206	,134	,212	,104	,208	,110	,204	,145
		[5]	,663	,256	,659	,230	,674	,305	,667	,271	,658	,230	,684	,339
	ρ = 0,3	**[1]**	,202	,128	,210	,116	,210	,126	,212	,079	,211	,080	,217	,070
		[5]	,578	,075	,572	,060	,577	,072	,593	,084	,582	,066	,591	,079

[1] Estimación imprecisa subestimada: ≥ 0,10 por debajo del valor teórico λ_{FD}.
[5] Estimación imprecisa sobreestimada: ≥ 0,10 por encima del valor teórico λ_{FD}.

Por último, en la Tabla 12 se muestran las estimaciones promedio y su dispersión para el grupo de estimaciones definidas como imprecisas dentro del ESTUDIO 2, diferenciando nuevamente entre *sub* y *sobre* estimaciones.

Cuando λ_{FD} = 0,20 y ρ = 0, las estimaciones I^- obtuvieron valores promedio próximos a cero en x_{10}, mientras que obtuvieron valores promedio negativos (< -0,10) en x_{11} y x_{12}. Para este mismo nivel del FD se obtuvieron promedios en las estimaciones I^+ entre 0,55 y 0,60 en los tres indicadores analizados. Por su parte, cuando λ_{FD} = 0,30 y ρ = 0, los promedios de I^- no superaron el valor 0,10 en ninguno de los indicadores del FD, y los promedios de I^+ se situaron entre los valores promedio 0,60-0,65. Y cuando λ_{FD} = 0,40 y ρ = 0 los promedios I^- se situaron ligeramente por encima del valor 0,20 mientras que los promedios I^+ se situaron ligeramente por encima del valor 0,65.

Cuando ρ = 0,30 los promedios I^- fueron bastante similares a los calculados para el escenario en el que los factores son ortogonales. Las diferencias se produjeron en los promedios de las estimaciones I^+, observándose cierta corrección de la sobreestimación cuando los factores están correlacionados ($I^+_{0,20}$ entre 0,40 y 0,43, $I^+_{0,30}$ en torno a 0,50, y $I^+_{0,40}$ entre 0,57 y 0,60).

En otras palabras, cuando se dispone de tamaños muestrales reducidos (N = 200), la imprecisión de las estimaciones es similar en los tres niveles de debilidad definidos para el FD en el ESTUDIO 2.

Discusión y conclusiones

El análisis de los resultados de RFD por condición experimental es de gran interés para el investigador aplicado, ya que informan de cuáles son los escenarios en los que se puede esperar que los distintos métodos de estimación funcionen adecuadamente, siempre que los modelos estén correctamente especificados. No obstante, la cuestión de fondo del presente trabajo es que en la investigación aplicada los modelos se asientan con bastante frecuencia sobre una sola muestra empírica. Tomada esta muestra individualmente, los parámetros estimados pueden presentar un mayor o menor grado de discrepancia respecto al parámetro poblacional, especialmente en aquellos indicadores con alto nivel de unicidad y en muestras no demasiado numerosas.

Desde las recomendaciones que surgen de los estudios que promedian la calidad de la recuperación, la investigación *una-muestra* puede disponer de garantías suficientes respecto a la adecuación de utilizar AF para analizar la estructura interna de los datos, con ciertas recomendaciones sobre cuál es el método de estimación que mejor comportamiento tiene en cada caso concreto. Sin embargo, en el presente trabajo se ha mostrado que estas recomendaciones deberían tomarse en cuenta *a priori* desde el punto de vista de la evaluación de la dimensionalidad de los datos y no desde la significación práctica de las saturaciones estimadas.

Que un determinado modelo de AFC haya sido elaborado dentro de un conjunto de condiciones en las que la RFD promedio resulta aceptable no garantiza la precisión de las estimaciones de aquellos indicadores o factores que presentan niveles bajos de carga factorial. La calidad de la recuperación se restringe a situaciones en las que o bien se dispone de saturaciones altas (por ejemplo, $\geq 0{,}70$), esto es, situaciones en las que el nivel de precisión es elevado incluso en las condiciones más desfavorables, o bien se dispone de un amplio número de observaciones. El problema es que estas situaciones no son muy frecuentes en la práctica.

Cuando el investigador dispone de un buen conocimiento de los valores teóricos o poblacionales de los parámetros, fundamentándose en la teoría existente y en investigaciones previas, dispone de información para diferenciar cuando los resultados del tipo *una-muestra* son adecuados o cuando se están produciendo errores importantes de estimación. Pero, en no pocas ocasiones, el investigador no dispone de este conocimiento, por lo que no tiene forma de profundizar en las relaciones entre indicadores y FD al no tener valor teórico o poblacional con el que comparar sus resultados. Esta circunstancia puede ser incluso más preocupante si se utiliza AFE como una primera aproximación al análisis de la estructura interna de los

datos, en donde es común encontrarse con factores adicionales con poca significación teórica, producto del error de medición.

Cuando una primera aproximación AFE se somete a un análisis AFC, estos *FD espurios* se estiman en ausencia de conocimiento sobre sus valores poblacionales, pudiendo incurrir en su interpretación directa sin garantías suficientes, otorgándoles naturaleza teórica a lo que solo es varianza irrelevante (Messick, 1993, 1995).

Los índices habitualmente utilizados en estudios RFD para evaluar la calidad de la recuperación de parámetros (C_k y $RMSD_k$) promedian las diferencias entre saturación teórica y estimada del factor definido como débil en *todas las soluciones* que pertenecen a una misma condición experimental. La lógica de estos índices, por tanto, se fundamenta en una medida global de la recuperación entre-soluciones dentro de la condición a la que pertenecen (por ejemplo, la calidad de la RFD para determinado tamaño muestral, para un modelo con factores correlacionados, o para un determinado método de estimación). Sobre estos índices, los resultados del ESTUDIO 1 replican los resultados obtenidos en investigación previa relacionada con RFD (Briggs y MacCallum, 2003; Ximénez, 2005, 2006). En términos generales, se añade una mayor evidencia de que la estimación de parámetros mejora notablemente a medida que aumenta el tamaño muestral, lo que era esperable dadas las propiedades matemáticas de los métodos de estimación empleados.

En el ESTUDIO 2 se observó que cuanto mayor es la unicidad presente en los indicadores que representan al FD peor es la recuperación promedio de los parámetros. En este segundo estudio también se mostró que la correlación entre factores mejoró la recuperación promedio, incluso considerando niveles más moderados (0,30) que los utilizados en anteriores estudios (0,50 en Ximénez, 2006), especialmente en el caso del índice $RMSD_k$.

La comparación entre ML y ULS no ha producido diferencias estadísticamente significativas en la RFD en ninguno de los dos estudios realizados. Es más, el tipo de RFD que puede identificarse mediante las estimaciones proporcionadas por ambos métodos es prácticamente idéntico, lo que deja a elección del investigador el uso de uno u otro método de estimación en las condiciones simuladas. En la investigación previa se comentaron algunas diferencias entre ambos métodos pero se fundamentaban principalmente en aspectos marginales dentro de las condiciones analizadas. Así, el mejor funcionamiento de ULS se relaciona con el análisis de aquellas soluciones en las que aparecen casos Heywood, o en condiciones concretas del diseño elaborado, como tamaños muestrales pequeños (N = 100). Los resultados obtenidos en este trabajo son congruentes con dichos resultados previos.

El aspecto fundamental del presente trabajo se ha centrado en la recodificación de los parámetros estimados en función de su grado de significación práctica, lo que ha permitido identificar distintas tipologías de RFD *solución a solución* y atendiendo a la información combinada de cada indicador del FD. Con este enfoque se ha tratado de profundizar y ampliar las recomendaciones hacia la investigación aplicada en aquellas situaciones en las que aparece FD.

El tipo de RFD definido como inaceptable es aquel en el que dos o tres indicadores de los analizados obtienen un nivel de precisión suficiente para la interpretación. Por su parte, una RFD definida como imprecisa es aquella en la que alguno de los 3 indicadores analizados obtiene una estimación imprecisa, pudiendo ser el resto estimaciones aceptables e, incluso, una estimación aceptable y otra óptima. Desde el punto de vista de la investigación aplicada, el problema de la interpretación de las saturaciones se produce cuando el investigador desconoce cuál de estos tres parámetros es el que ha obtenido una estimación I, una estimación A, o una estimación O, pudiendo incurrir en el error de interpretar las tres como si estuviesen estimadas con el mismo nivel de precisión (especialmente cuando el modelo se sustenta en AFE previo). Por esta razón, y a pesar de que alguna de las estimaciones pueda ser más precisa, lo recomendable en esta situación es no realizar interpretaciones del FD presente en el modelo más allá de su dimensionalidad.

De manera complementaria a las indicaciones fundamentadas en los promedios RFD por condición, la tipología RFD permite establecer cierto nivel de riesgo relativo ante distintos escenarios cuando se interpretan directamente la relación entre indicadores y FD. Este riesgo relativo es el que ha permitido diferenciar entre condiciones en las que se pueden realizar interpretaciones directas con garantías de aquellas en las que no debería evaluarse el modelo factorial más allá de su dimensionalidad.

El modelo propuesto por MacCallum y Tucker presenta un nivel de debilidad del FD que puede considerarse de nivel intermedio (λ_{FD} = 0,50), lo que se ha reflejado en unos niveles promedio de RFD adecuados en todas las condiciones del ESTUDIO 1. No obstante, la conclusión que puede derivarse a partir de los tipos de RFD analizados es que la interpretación directa de las saturaciones no puede realizarse con suficientes garantías hasta no disponer de tamaños muestrales N = 500 (65% aproximadamente de soluciones con RFD precisa u óptima; 18% de RFD inaceptable).

Siendo algo menos estrictos, podríamos considerar que las soluciones generadas con N = 300 permiten realizar interpretaciones con ciertas garantías para la investigación aplicada, ya que en torno al 45% de dichas soluciones fueron estimadas con RFD óptima o aceptable,

existiendo solamente un 16,8% de soluciones con RFD inaceptable. Las soluciones con RFD impreciso superaron el 37%, pudiendo argumentarse que al tener solamente un indicador estimado de manera imprecisa de los tres que representan al FD son soluciones que obtienen la suficiente significación práctica. Cuando N = 300 los valores promedio para las estimaciones de tipo I^- e I^+ se encontraban a una distancia en torno a los 0,15 puntos respecto al parámetro teórico, es decir, se obtuvieron estimaciones en torno a λ = 0,35 y en torno a λ = 0,65, estimaciones que podrían interpretarse como más débiles o más fuertes de lo que son realmente (λ_{FD} = 0,50). El problema surge, como se argumentaba más arriba, cuando el investigador no dispone de conocimiento previo sobre el valor poblacional de los parámetros que está estimando.

En resumen, con saturaciones intermedias o mayores la evaluación de la dimensionalidad del modelo resulta adecuada a partir de los tamaños mínimos recomendados, pero la interpretación directa de las saturaciones del factor más débil necesita de un tamaño muestral más exigente. Y esto en un modelo que, en la práctica, es un modelo que podríamos considerar como fuerte (recordemos que el factor 1 presentaba saturaciones de 0,95 y el factor 2 de 0,70)[36].

Respecto al ESTUDIO 2, si atendemos a la recuperación promedio y a las indicaciones que se desprenden de la investigación previa, en el mejor de los escenarios (λ_{FD} = 0,40), se podría llegar a la conclusión de que las saturaciones estimadas se pueden interpretar directamente incluso con tamaños muestrales pequeños, en torno a las 200 observaciones. No obstante, los resultados encontrados sugieren un nivel mayor de exigencia desde el punto de vista de la significación práctica de las estimaciones.

Al *estresar* el modelo utilizando un tamaño muestral reducido de manera constante (N = 200) y al aumentar el nivel de debilidad del FD, se ha tratado de reproducir situaciones de investigación aplicada que surgen con cierta frecuencia. Si con λ_{FD} = 0,50 y N = 100 la proporción de soluciones con RFD inaceptable era del 14,5%, con el doble de muestra esta proporción ascendía al 45% cuando λ_{FD} = 0,40 (llegando incluso, al 55,9% cuando λ_{FD} = 0,30, y hasta el 63,3% cuando λ_{FD} = 0,20). Y esta distribución del tipo de RFD es independiente de la correlación entre factores y de la carga factorial de los FF.

A partir de estos resultados, si con λ_{FD} = 0,50 el tamaño muestral necesario para permitir la interpretación directa de las estimaciones puede situarse entre las 300 y las 500

[36] El término fuerte se aplica a este modelo desde el punto de vista de sus niveles de carga. En términos del número de indicadores, el factor 3 puede considerarse como un factor débilmente representado ya que solamente tiene tres indicadores, frente a los cuatro indicadores del factor 2 y los cinco indicadores del factor 3. El modelo de MacCallum y Tucker se diseñó y analizó en el contexto del AFE, en donde la aparición de este tipo de factores, con menos indicadores que en el resto de factores, es bastante frecuente.

observaciones, con λ_{FD} = 0,40 el tamaño muestral necesario debería ser aún mayor, y más elevado aun cuando el nivel de debilidad es más pronunciado.

De cara a investigación futura, en términos de RFD el diseño ejecutado se ha centrado fundamentalmente en la comparación entre distintos niveles de debilidad variando el nivel de carga factorial de los indicadores que lo representan, reflejando así una medición débil de los indicadores (unicidad alta) y/o un nivel débil de correlación entre indicadores (comunalidad baja). Desde el punto de vista de la representación de dimensiones o factores, otro aspecto de especial relevancia para profundizar en el tipo de RFD evaluado en este estudio sería incluir escenarios controlados en los que varíe el número de indicadores que forman el FD. Los patrones de estimación serían más complejos, al incluir más indicadores, pudiendo incluso evaluar tipos de RFD más allá de los cuatro tipos propuestos en este trabajo para tres indicadores. Siguiendo este razonamiento, el procedimiento desarrollado aquí resulta de interés para escenarios con más indicadores por FD.

Bibliografía

Abad, F. J., Olea, J., Ponsoda, V. & García, C. (2011). Medición en ciencias sociales y de la salud. Madrid: Síntesis.

American Psychological Association, American Educational Research Association, y National Council on Measurement in Education (1999). Standards for educational and psychological test y manuals. Washington, DC: American Psychological Association.

Anderson, J. C., y Gerbing, D. W. (1988). Structural Equation Modeling in Practice: A review and recommended two-step approach. Psychological Bulletin, 103 (3), 411-423.

Batista-Foguet, J. M. & Coenders, G. (2000): Modelos de Ecuaciones Estructurales. La Muralla, Madrid.

Bollen, K. A. (1989). Structural equations with latent variables. New York, NY: John Wiley & Sons.

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. Multivariate Behavioral Research, 38, 25-56.

Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York: Guilford.

Browne, M. W. (1984) Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62–83.

Cudeck, R., & MacCallum, R. C. (Eds.). (2007). Factor Analysis at 100: Historical Developments and Future Directions. Mahwah, NJ: Lawrence Erlbaum SAGE Publications, Inc.

Elosua, P. (2003). Sobre la validez de los tests. Psicothema 15(2), 315-321.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4(3), 272-299.

Ferrando, P.J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en Psicología. Papeles del Psicólogo, 31, 18-33.

Ferrando, P.J.; Lorenzo, U. (1993). Algunas relaciones entre el modelo de un factor común y el modelo logístico de dos parámetros. Psicothema, 5(2), 403-412.

Flora, D.B., & Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychological Methods, 9, 466-491.

Forero, C.G., Maydeu-Olivares, A. & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. Structural Equation Modeling, 16, 625–641.

Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. Quality and Quantity, 24, 387- 404.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. Psychometrika 59, 381-389.

Jöreskog, K.G. & Sörbom, D. (2006). LISREL 8.8 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.

MacCallum, R. C., & Tucker, L. R (1991). Representing sources of error in the common factor model: Implications for theory and practice. Psychological Bulletin, 109, 502-511.

MacCallum, R. C., Tucker, L. R, & Briggs, N. E. (2001). An alternative perspective on parameter estimation in factor analysis and related methods. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future*. (pp. 39-57). Lincolnwood, IL: SSI.

Magnus, J.R., & Neudecker, H. (1999). Matrix Differential Calculus with Applications in Statistics and Econometrics (Revised Edition). Chichester: John Wiley & Sons, Ltd.

McDonald, R. P. (1982). Linear vs. nonlinear models in latent trait theory. Applied Psychological Measurement, 6, 379–396.

Messick, S. (1989). Validity. En R. L. Linn (Ed.), Educational measurement (pp. 13-103). New York: American Council on Educational/Macmillan.

Messick, S. (1993). Foundations of validity: Meaning and consequences in psychological assessment. Educational Testing Service. Princeton, New Jersey.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50, 741-749.

Mulaik, S. A. (2009) Linear Causal Modeling with Structural Equations. Boca Raton FL, CRC Press, Taylor and Francis Group.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. Psychometrika, 43, 551-560.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49, 115-132.

Muthén, B. (1993). Goodness of fit with categorical and other non-normal variables. In K. A. Bollen, & J. S. Long (Eds.), Testing Structural Equation Models (pp. 205-243). Newbury Park, CA: Sage.

Muthén, B., & Kaplan D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. British Journal of Mathematical and Statistical Psychology, 38, 171-189.

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. British Journal of Mathematical and Statistical Psychology, 45, 19-30.

Muthén, B., du Toit, S.H.C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. **http://pages.gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf**.

Olsson, U.H., Foss, T., Troye, S. V., & Roy D. Howell (2000). The Performance of ML, GLS and WLS Estimation in Structural Equation Modeling Under Conditions of Misspecification and Nonnormality. Structural Equation Modeling, 7 (4), 557-595.

Revuelta, J. y Ponsoda, V. (2003). Simulación de modelos estadísticos en ciencias sociales. Madrid. La Muralla.

Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: attacking the conventional wisdom. Multivariate Behavioral Research, 35, 137-167.

Skrondal, A. and Rabe-Hesketh, S. (2004). Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models. Boca Raton, FL: Chapman & Hall/CRC.

Tucker, L.R. (1951). A method for synthesis of factor analysis studies. Personnel Research Section Report, 984. Washington, D. C.: Department of the Army.

Tucker, L.R. y Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Ximénez, C. y García, A.G. (2005). Comparación de los métodos de estimación de máxima verosimilitud y mínimos cuadrados no ponderados en el análisis factorial confirmatorio mediante simulación Monte Carlo. Psicothema, Vol. 17, nº 3, pp. 528-535.

Ximénez, C. (2006). A Monte Carlo Study of Recovery of weak factor loadings in Confirmatory Factor Analysis. Structural Equation Modeling, 13, 587-614.

Ximénez, C. (2007). Effect of variable and subject sampling on recovery of weak factors in CFA. Methodology, 3, 67-80.

Ximénez, C. (2009). Recovery of weak factor loadings in confirmatory factor analysis under conditions of model misspecification. Behavior Research Methods, 41, 1038-1052.

ANEXOS

ESTUDIO 1: Modelo de MacCallum y Tucker (1991)

Figura 7: Promedios y dispersión del índice C_k en función de los patrones de estimación definidos

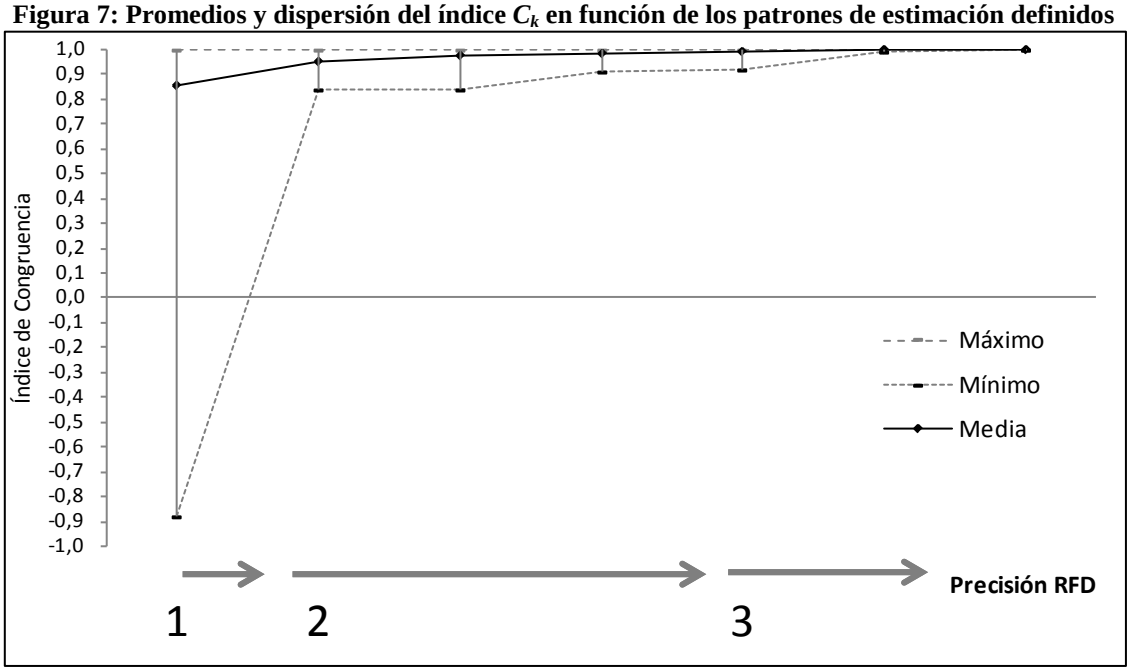
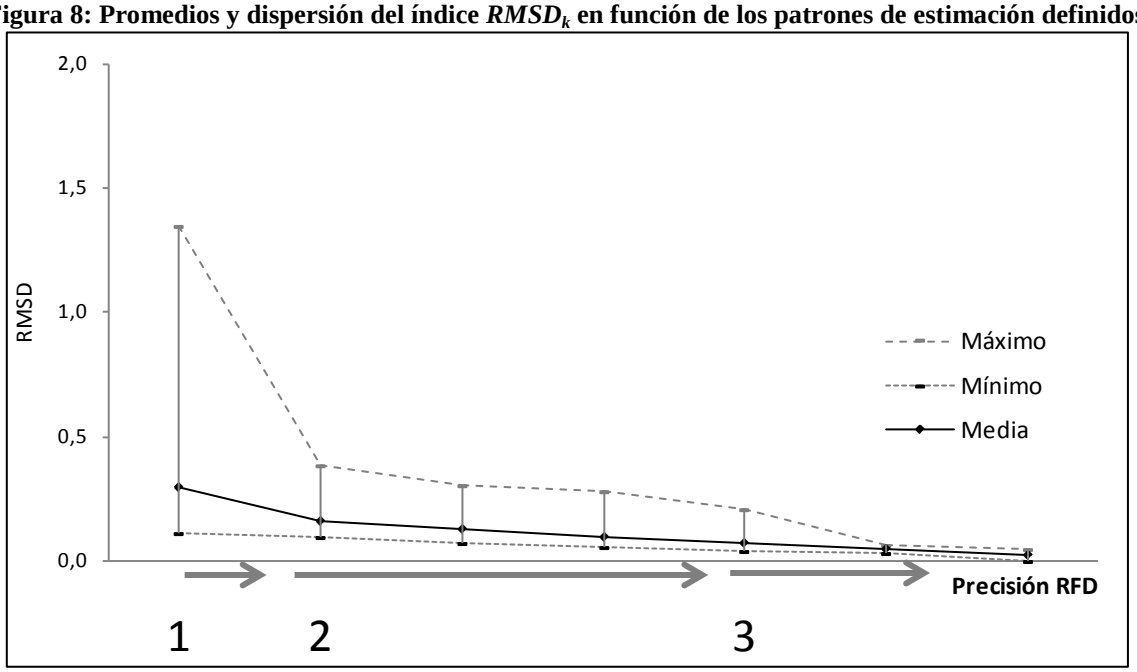


Figura 8: Promedios y dispersión del índice $RMSD_k$ en función de los patrones de estimación definidos



NOTA: El eje horizontal recoge la suma del tipo de estimación del parámetro *lambda* de los 3 indicadores que representan al FD. Codificación: 1 (estimación imprecisa – I), 2 (estimación aceptable – A), y 3 (estimación óptima – O). El rango de esta suma oscila entre 3 y 9. El valor 1 de las figuras refleja aquellas soluciones en las que los 3 indicadores tienen estimaciones de tipo I, el valor 2 refleja el punto a partir del cual al menos una estimación es de tipo A, y el valor 3 refleja el punto a partir del cual al menos una estimación es del tipo O (y el resto del tipo A).

Tabla 13: Significación práctica de la estimación de los indicadores FD por condición

Precisión de las estimaciones (% vertical)		**ML**			**ULS**		
		x_{10}	x_{11}	x_{12}	x_{10}	x_{11}	x_{12}
N = 100	[1] I-*subestimada*	24,4%	23,8%	25,9%	24,8%	24,0%	26,1%
	[2] A	10,8%	12,2%	13,3%	10,8%	12,2%	13,2%
	[3] O	27,0%	25,0%	24,3%	26,9%	25,0%	24,3%
	[4] A	13,5%	12,0%	10,3%	13,4%	12,0%	10,3%
	[5] I-*sobreestimada*	24,3%	27,0%	26,1%	24,2%	26,9%	26,2%
N = 300	[1] I-*subestimada*	12,5%	11,9%	12,6%	12,5%	11,9%	12,6%
	[2] A	14,1%	15,5%	17,1%	14,1%	15,5%	17,1%
	[3] O	42,1%	41,0%	46,7%	42,1%	41,0%	46,7%
	[4] A	17,8%	16,6%	14,1%	17,8%	16,6%	14,1%
	[5] I-*sobreestimada*	13,5%	15,0%	9,5%	13,5%	15,0%	9,5%
N = 500	[1] I-*subestimada*	6,1%	6,6%	7,0%	6,1%	6,6%	7,0%
	[2] A	14,6%	17,6%	15,7%	14,6%	17,6%	15,7%
	[3] O	54,9%	52,1%	55,5%	54,9%	52,1%	55,5%
	[4] A	17,4%	14,8%	15,3%	17,4%	14,8%	15,3%
	[5] I-*sobreestimada*	7,0%	8,9%	6,5%	7,0%	8,9%	6,5%
N = 1.000	[1] I-*subestimada*	2,0%	2,2%	2,0%	2,0%	2,2%	2,0%
	[2] A	12,1%	13,2%	12,1%	12,1%	13,2%	12,1%
	[3] O	69,1%	69,2%	71,0%	69,1%	69,2%	71,0%
	[4] A	15,1%	12,5%	13,4%	15,1%	12,5%	13,4%
	[5] I-*sobreestimada*	1,7%	2,9%	1,5%	1,7%	2,9%	1,5%
N = 2.000	[1] I-*subestimada*	0,0%	0,0%	0,1%	0,0%	0,0%	0,1%
	[2] A	5,7%	7,9%	7,5%	5,7%	7,9%	7,5%
	[3] O	86,5%	85,1%	86,0%	86,5%	85,1%	86,0%
	[4] A	7,7%	6,9%	6,3%	7,7%	6,9%	6,3%
	[5] I-*sobreestimada*	0,1%	0,1%	0,1%	0,1%	0,1%	0,1%

[1] Estimación imprecisa (I-*subestimada*), [2] Estimación aceptable (A), [3] Estimación óptima (O), [4] Estimación aceptable (A), [5] Estimación imprecisa (I-*sobreestimada*)

Tabla 14: Tipo de RFD en función del tamaño muestral y del método de estimación (con y sin casos Heywood)

Patrones (% vertical)	**ML** (*con Heywood*)					**ULS** (*con Heywood*)				
	N=100	N=300	N=500	N=1.000	N=2.000	N=100	N=300	N=500	N=1.000	N=2.000
[4]	47,8%	16,8%	7,1%	1,1%	0,0%	47,9%	16,8%	7,1%	1,1%	0,0%
[3]	37,8%	37,8%	27,3%	9,9%	0,4%	37,8%	37,8%	27,3%	9,9%	0,4%
[2]	12,9%	35,0%	47,5%	53,6%	35,4%	12,9%	35,0%	47,5%	53,6%	35,4%
[1]	1,4%	10,4%	18,1%	35,4%	64,2%	1,4%	10,4%	18,1%	35,4%	64,2%
.Patrones (% vertical)	**ML** (*sin Heywood*)					**ULS** (*sin Heywood*)				
	N=100	N=300	N=500	N=1.000	N=2.000	N=100	N=300	N=500	N=1.000	N=2.000
[4]	45,8%	15,4%	6,7%	0,8%	0,0%	45,9%	15,4%	6,7%	0,8%	0,0%
[3]	37,7%	35,5%	22,8%	7,9%	0,4%	37,7%	35,5%	22,8%	7,9%	0,4%
[2]	15,1%	39,1%	53,4%	57,4%	37,3%	15,0%	39,1%	53,4%	57,4%	37,3%
[1]	1,4%	10,0%	17,1%	33,9%	62,3%	1,4%	10,0%	17,1%	33,9%	62,3%

[1] RFD óptima: estimación conjunta del FD entre ± 0,05 respecto al valor teórico. INTETRPRETACIÓN DIRECTA.
[2] RFD precisa: sin estimaciones I, soluciones mixtas con estimaciones O y A. INTETRPRETACIÓN DIRECTA.
[3] RFD imprecisa: soluciones con alguna estimación I. EVALUACIÓN DIMENSIONALIDAD.
[4] RFD inaceptable: soluciones con 2 o 3 estimaciones I. EVALUACIÓN DIMENSIONALIDAD.

Tabla 15: Tipo de RFD y promedio de C_k y $RMSD_k$ por condición

		ML					ULS				
		N=100	N=300	N=500	N=1.000	N=2.000	N=100	N=300	N=500	N=1.000	N=2.000
C_k	**[4]**	,921	,959	,971	,976	---	,905	,959	,971	,976	---
	[3]	,981	,986	,988	,990	,989	,981	,986	,988	,990	,989
	[2]	,996	,995	,995	,996	,997	,996	,995	,995	,996	,997
	[1]	,999	,999	,999	,999	,999	,999	,999	,999	,999	,999
$RMSD_k$	**[4]**	,219	,146	,123	,109	---	,227	,146	,123	,109	---
	[3]	,116	,093	,084	,077	,079	,116	,093	,084	,077	,079
	[2]	,057	,055	,054	,050	,044	,057	,055	,054	,050	,044
	[1]	,027	,026	,028	,025	,023	,027	,026	,028	,025	,023

[1] RFD óptima: estimación conjunta del FD entre ± 0,05 respecto al valor teórico. INTETRPRETACIÓN DIRECTA.
[2] RFD precisa: sin estimaciones I, soluciones mixtas con estimaciones O y A. INTETRPRETACIÓN DIRECTA.
[3] RFD imprecisa: soluciones con alguna estimación I. EVALUACIÓN DIMENSIONALIDAD.
[4] RFD inaceptable: soluciones con 2 o 3 estimaciones I. EVALUACIÓN DIMENSIONALIDAD.

ESTUDIO 2: Variabilidad en el factor débil y correlación entre factores

Tabla 16: Calidad en la RFD ante variaciones del factor débil, correlación entre factores y variaciones en FF_1 y FF_2 (*excluidos los casos Heywood*)

λ_{FD} x ρ x λ_{FF1} xλ_{FF2}			C_k				$RMSD_k$			
			ML		ULS		ML		ULS	
			$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x
λ_{FD} = 0,2	ρ = 0	λ_1=0,7 y λ_2=0,5	**,662**	,426	**,617**	,489	**,240**	,118	**,247**	,129
		λ_1=0,7 y λ_2=0,7	**,665**	,423	**,618**	,490	**,239**	,118	**,247**	,126
		λ_1=0,9 y λ_2=0,5	**,660**	,429	**,575**	,539	**,241**	,120	**,255**	,135
		λ_1=0,9 y λ_2=0,7	**,662**	,426	**,612**	,498	**,240**	,119	**,249**	,129
		Total	**,662**	,426	**,606**	,504	**,240**	,118	**,249**	,130
	ρ = 0,3	λ_1=0,7 y λ_2=0,5	**,681**	,401	**,688**	,387	,192	,091	,179	,078
		λ_1=0,7 y λ_2=0,7	**,701**	,387	**,717**	,351	,181	,086	,169	,072
		λ_1=0,9 y λ_2=0,5	**,712**	,364	**,693**	,379	,181	,082	,175	,074
		λ_1=0,9 y λ_2=0,7	**,731**	,342	**,719**	,349	,172	,077	,167	,070
		Total	**,706**	,374	**,704**	,367	,181	,085	,172	,074
λ_{FD} = 0,3	ρ = 0	λ_1=0,7 y λ_2=0,5	**,873**	,246	**,862**	,282	,172	,102	,175	,106
		λ_1=0,7 y λ_2=0,7	**,873**	,246	**,850**	,314	,172	,100	,178	,113
		λ_1=0,9 y λ_2=0,5	**,875**	,235	**,834**	,357	,172	,099	,182	,119
		λ_1=0,9 y λ_2=0,7	**,878**	,227	**,839**	,344	,171	,097	,181	,119
		Total	**,875**	,239	**,846**	,325	,172	,099	,179	,115
	ρ = 0,3	λ_1=0,7 y λ_2=0,5	**,863**	,268	**,868**	,202	,158	,092	,158	,074
		λ_1=0,7 y λ_2=0,7	**,879**	,233	**,881**	,173	,149	,083	,152	,068
		λ_1=0,9 y λ_2=0,5	**,896**	,148	**,868**	,192	,147	,069	,159	,071
		λ_1=0,9 y λ_2=0,7	**,903**	,130	**,882**	,164	,142	,065	,152	,067
		Total	**,885**	,203	**,875**	,183	,149	,078	,155	,070
λ_{FD} = 0,4	ρ = 0	λ_1=0,7 y λ_2=0,5	,939	,152	,948	,080	,138	,087	,134	,072
		λ_1=0,7 y λ_2=0,7	,948	,083	,939	,156	,134	,073	,137	,088
		λ_1=0,9 y λ_2=0,5	,932	,194	,940	,147	,140	,099	,137	,084
		λ_1=0,9 y λ_2=0,7	,944	,122	,936	,169	,136	,078	,138	,090
		Total	,941	,144	,941	,143	,137	,085	,137	,084
	ρ = 0,3	λ_1=0,7 y λ_2=0,5	,932	,225	,947	,099	,130	,098	,133	,064
		λ_1=0,7 y λ_2=0,7	,941	,195	,951	,075	,124	,089	,130	,060
		λ_1=0,9 y λ_2=0,5	,959	,047	,945	,084	,118	,057	,138	,063
		λ_1=0,9 y λ_2=0,7	,959	,074	,949	,074	,116	,056	,134	,060
		Total	,947	,156	,948	,084	,122	,077	,134	,062

Se han resaltado en negrita aquellos promedios RFD que se encuentran dentro de los valores considerados como inadecuados.

Figura 9: Promedio de C_k por condición (*excluidos los casos Heywood*)

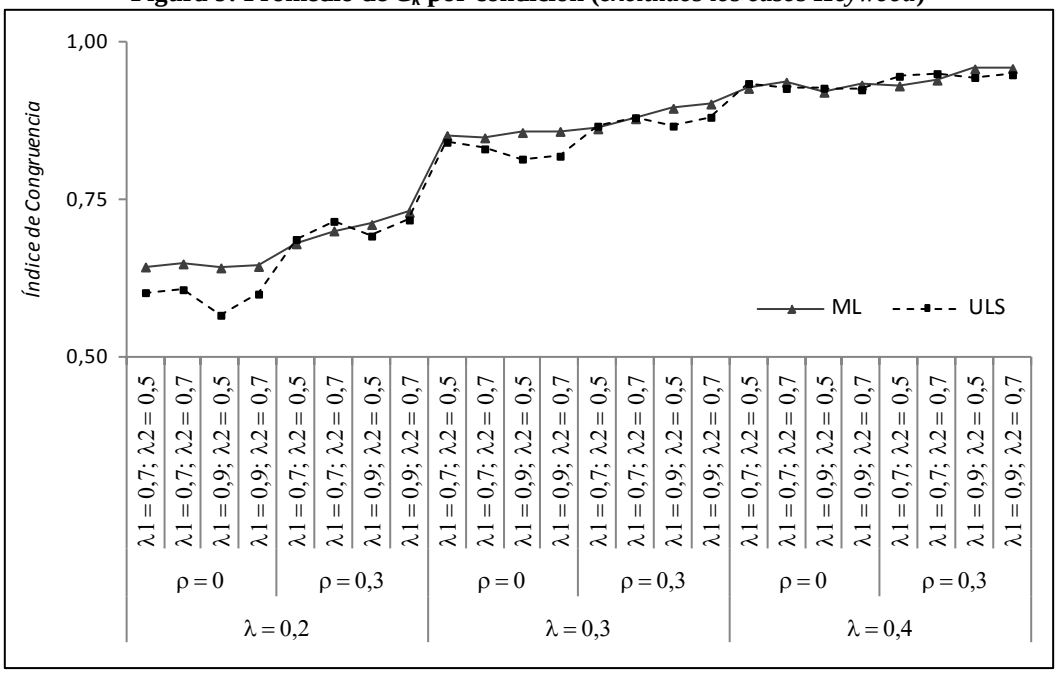


Figura 10: Promedio de $RMSD_k$ por condición (*excluidos los casos Heywood*)

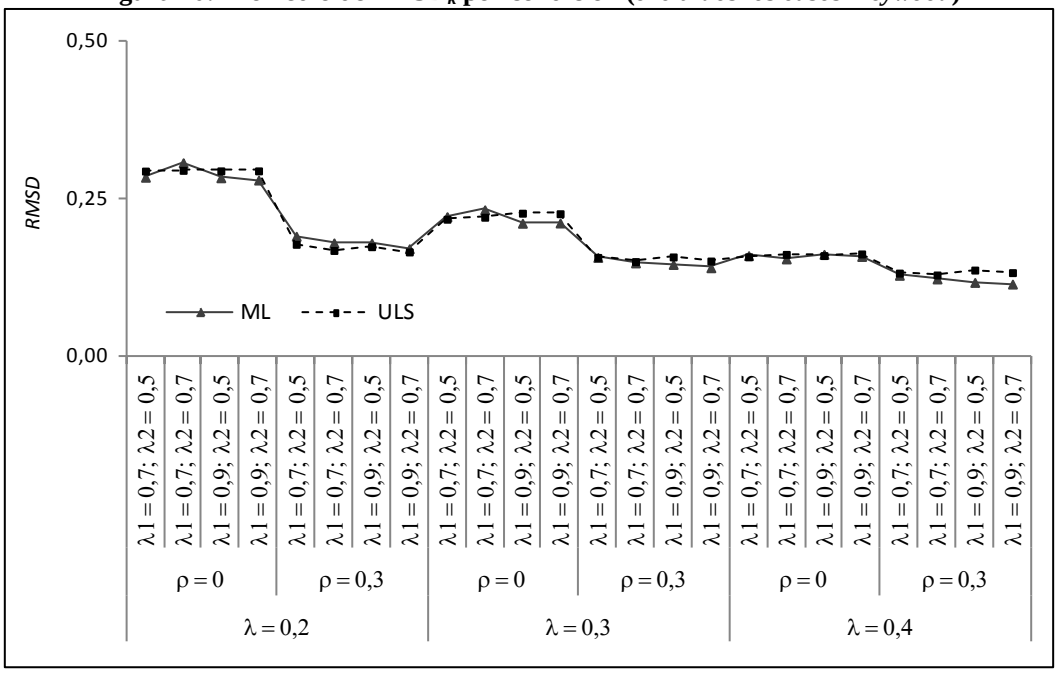


Figura 11: Promedios y dispersión del índice C_k en función de los patrones de estimación definidos ($\lambda_{FD\,=\,0,20}$)

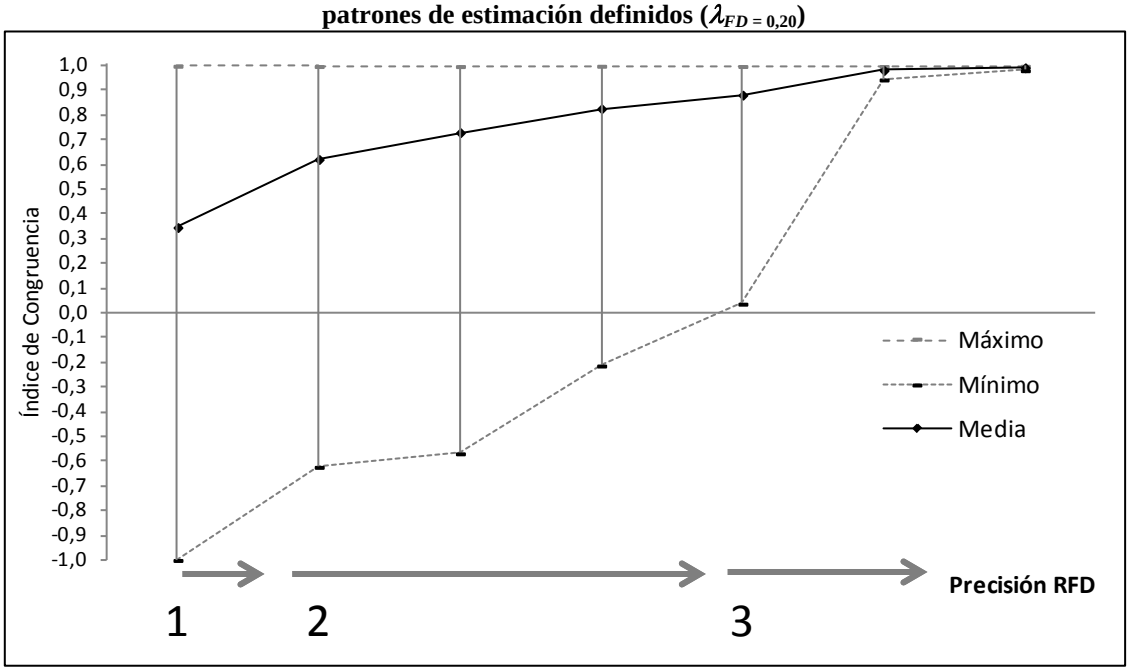
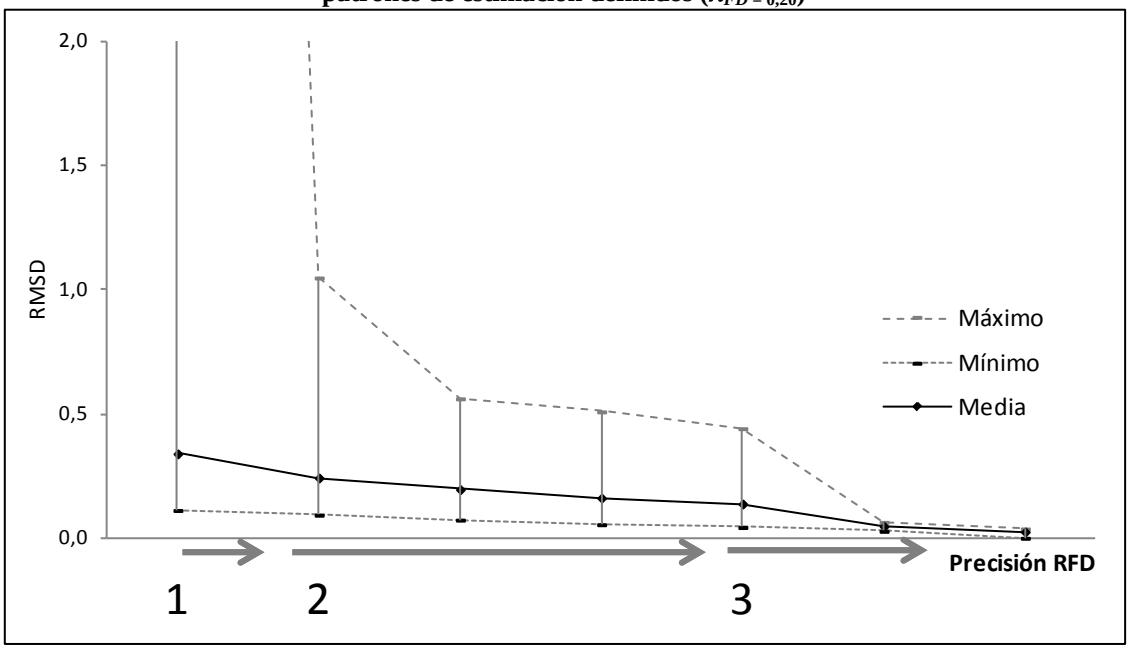


Figura 12: Promedios y dispersión del índice $RMSD_k$ en función de los patrones de estimación definidos ($\lambda_{FD\,=\,0,20}$)



NOTA: El eje horizontal recoge la suma del tipo de estimación del parámetro *lambda* de los 3 indicadores que representan al FD. Codificación: 1 (estimación imprecisa – I), 2 (estimación aceptable – A), y 3 (estimación óptima – O). El rango de esta suma oscila entre 3 y 9. El valor 1 de las figuras refleja aquellas soluciones en las que los 3 indicadores tienen estimaciones de tipo I, el valor 2 refleja el punto a partir del cual al menos una estimación es de tipo A, y el valor 3 refleja el punto a partir del cual al menos una estimación es del tipo O (y el resto del tipo A).

Figura 13: Promedios y dispersión del índice C_k en función de los patrones de estimación definidos ($\lambda_{FD\ =\ 0,30}$)

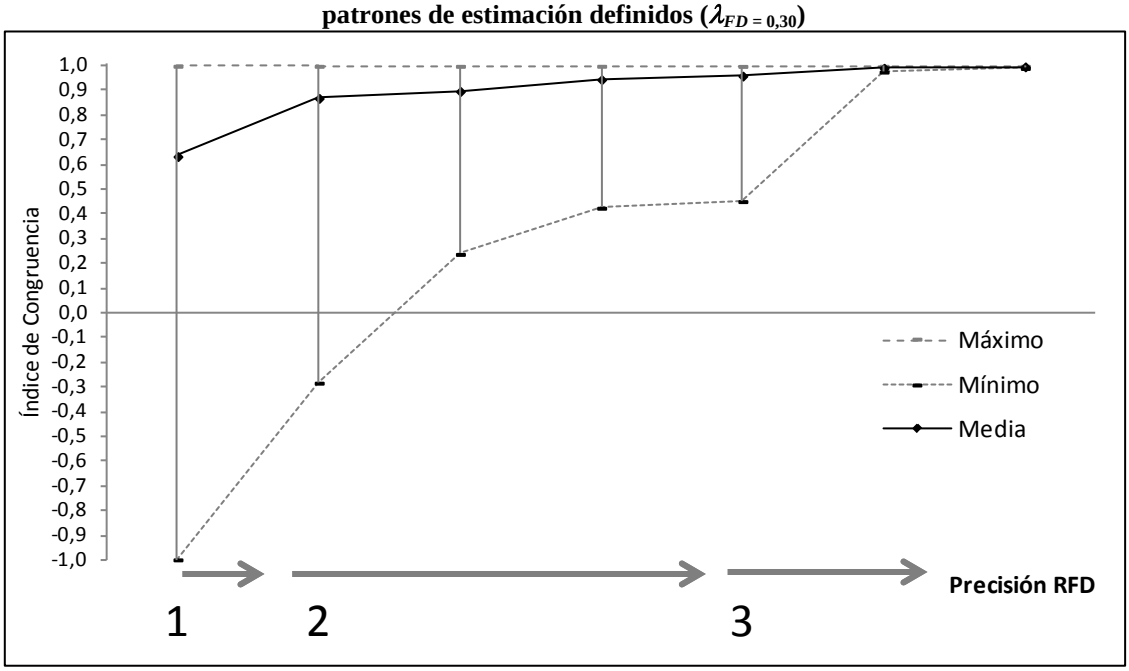
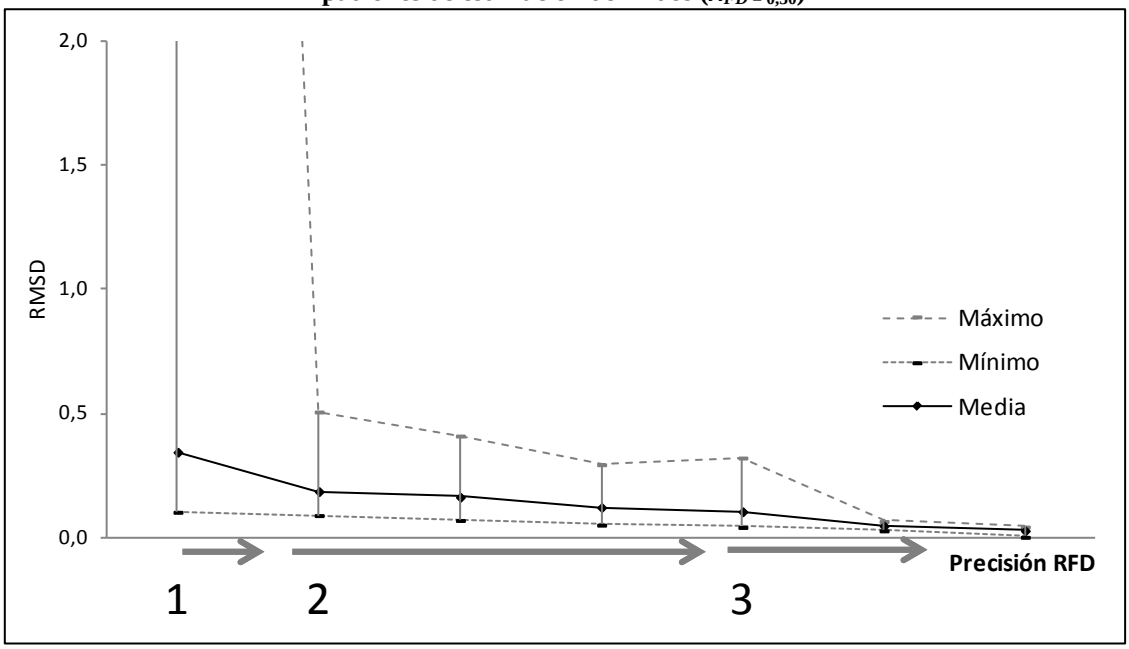


Figura 14: Promedios y dispersión del índice $RMSD_k$ en función de los patrones de estimación definidos ($\lambda_{FD\ =\ 0,30}$)



NOTA: El eje horizontal recoge la suma del tipo de estimación del parámetro *lambda* de los 3 indicadores que representan al FD. Codificación: 1 (estimación imprecisa – I), 2 (estimación aceptable – A), y 3 (estimación óptima – O). El rango de esta suma oscila entre 3 y 9. El valor 1 de las figuras refleja aquellas soluciones en las que los 3 indicadores tienen estimaciones de tipo I, el valor 2 refleja el punto a partir del cual al menos una estimación es de tipo A, y el valor 3 refleja el punto a partir del cual al menos una estimación es del tipo O (y el resto del tipo A).

Figura 15: Promedios y dispersión del índice C_k en función de los patrones de estimación definidos ($\lambda_{FD\ =\ 0,40}$)

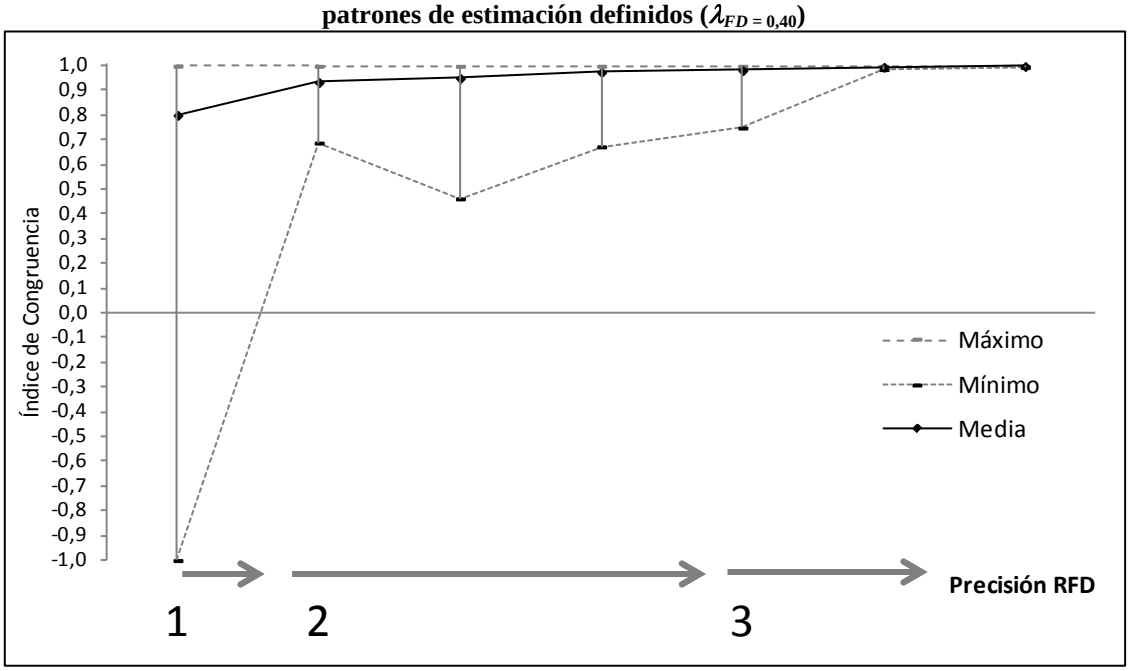
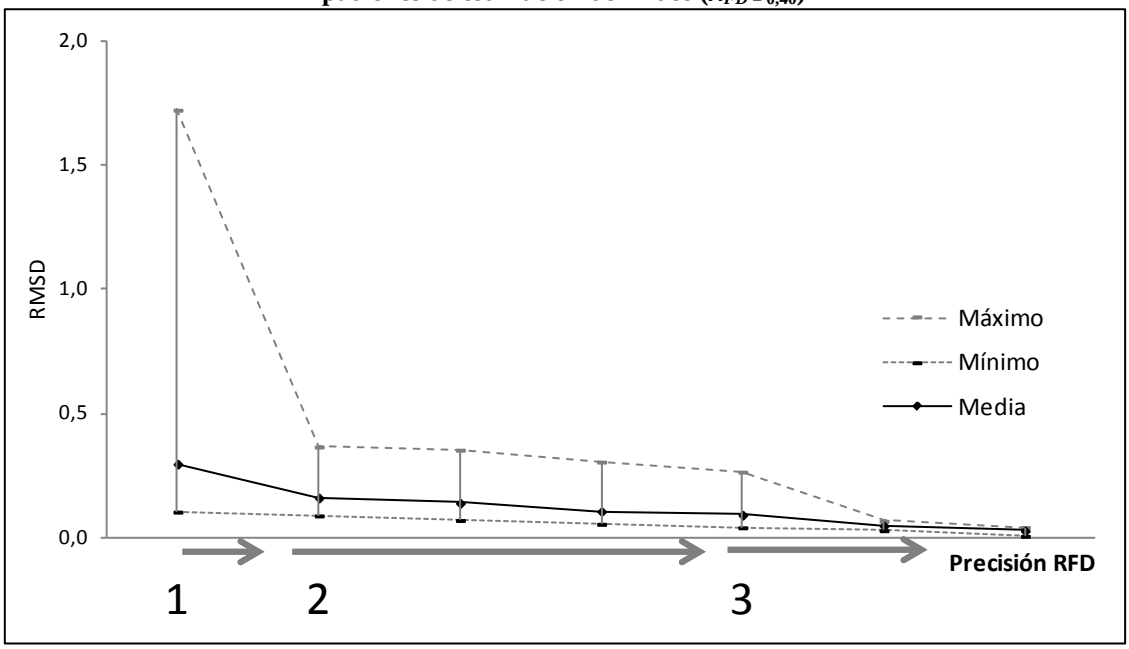


Figura 16: Promedios y dispersión del índice $RMSD_k$ en función de los patrones de estimación definidos ($\lambda_{FD\ =\ 0,40}$)



NOTA: El eje horizontal recoge la suma del tipo de estimación del parámetro *lambda* de los 3 indicadores que representan al FD. Codificación: 1 (estimación imprecisa – I), 2 (estimación aceptable – A), y 3 (estimación óptima – O). El rango de esta suma oscila entre 3 y 9. El valor 1 de las figuras refleja aquellas soluciones en las que los 3 indicadores tienen estimaciones de tipo I, el valor 2 refleja el punto a partir del cual al menos una estimación es de tipo A, y el valor 3 refleja el punto a partir del cual al menos una estimación es del tipo O (y el resto del tipo A).

Tabla 17: Significación práctica de la estimación de FD por condición (λ_{FD} = 0,2)

Precisión de las estimaciones (% vertical)			ML		ULS			
			x_{10}	x_{11}	x_{10}	x_{11}	x_{10}	x_{11}
FF: λ_1=0,7 y λ_2=0,5	ρ = 0	[1] I-*subestimada*	**17,5%**	30,9%	32,7%	**20,1%**	32,7%	34,2%
		[2] A	10,7%	11,4%	9,0%	10,3%	11,0%	8,7%
		[3] O	22,6%	18,4%	22,6%	21,5%	18,4%	22,6%
		[4] A	8,1%	7,4%	8,9%	8,0%	7,1%	8,7%
		[5] I-*sobreestimada*	**41,0%**	31,8%	26,8%	**40,1%**	30,7%	25,9%
	ρ = 0,3	[1] I-*subestimada*	29,0%	32,8%	30,6%	31,9%	33,3%	29,4%
		[2] A	9,5%	9,2%	9,3%	10,3%	10,0%	12,0%
		[3] O	19,4%	19,2%	20,0%	20,2%	20,5%	21,5%
		[4] A	9,3%	9,9%	9,3%	7,9%	10,0%	9,5%
		[5] I-*sobreestimada*	32,8%	28,9%	30,8%	29,7%	26,2%	27,6%
FF: λ_1=0,7 y λ_2=0,7	ρ = 0	[1] I-*subestimada*	**18,3%**	30,1%	31,6%	**19,9%**	33,0%	33,0%
		[2] A	10,7%	11,4%	9,3%	10,5%	11,6%	9,2%
		[3] O	22,5%	18,5%	22,9%	22,2%	17,7%	22,6%
		[4] A	8,3%	7,4%	9,3%	7,6%	7,0%	8,8%
		[5] I-*sobreestimada*	**40,1%**	32,5%	27,0%	**39,9%**	30,7%	26,4%
	ρ = 0,3	[1] I-*subestimada*	28,9%	31,8%	28,3%	30,1%	32,4%	28,0%
		[2] A	9,7%	9,4%	10,2%	12,1%	9,2%	11,5%
		[3] O	19,9%	19,5%	20,6%	20,8%	21,6%	23,7%
		[4] A	10,3%	10,9%	10,3%	8,7%	10,4%	9,0%
		[5] I-*sobreestimada*	31,2%	28,4%	30,6%	28,3%	26,4%	27,8%
FF: λ_1=0,9 y λ_2=0,5	ρ = 0	[1] I-*subestimada*	**17,7%**	30,6%	33,1%	**20,8%**	34,1%	35,7%
		[2] A	10,8%	11,4%	8,9%	10,6%	11,1%	8,4%
		[3] O	22,1%	18,3%	22,8%	21,7%	17,9%	22,4%
		[4] A	8,1%	7,6%	9,0%	8,0%	7,1%	8,2%
		[5] I-*sobreestimada*	**41,2%**	32,2%	26,2%	**39,0%**	29,9%	25,3%
	ρ = 0,3	[1] I-*subestimada*	27,9%	32,4%	28,7%	31,1%	33,8%	29,7%
		[2] A	10,1%	8,2%	9,3%	10,4%	9,2%	11,7%
		[3] O	19,9%	20,2%	20,6%	21,2%	20,6%	20,5%
		[4] A	9,3%	10,4%	9,8%	8,8%	11,0%	9,1%
		[5] I-*sobreestimada*	32,8%	28,8%	31,6%	28,5%	25,4%	29,0%
FF: λ_1=0,9 y λ_2=0,7	ρ = 0	[1] I-*subestimada*	**18,7%**	29,8%	32,2%	**20,5%**	32,4%	34,4%
		[2] A	10,5%	11,8%	8,9%	10,0%	11,6%	8,4%
		[3] O	22,4%	18,7%	22,9%	21,8%	17,8%	22,2%
		[4] A	8,0%	7,5%	8,9%	8,0%	7,1%	8,5%
		[5] I-*sobreestimada*	**40,4%**	32,2%	27,1%	**39,6%**	31,1%	26,5%
	ρ = 0,3	[1] I-*subestimada*	28,0%	31,1%	27,0%	29,9%	31,9%	27,8%
		[2] A	10,2%	8,2%	10,7%	11,9%	10,0%	11,6%
		[3] O	20,4%	20,1%	21,2%	21,9%	21,5%	23,1%
		[4] A	9,8%	11,9%	10,2%	8,4%	10,9%	8,9%
		[5] I-*sobreestimada*	31,6%	28,7%	30,9%	27,9%	25,7%	28,6%

[1] Estimación imprecisa (I-*subestimada*), [2] Estimación aceptable (A), [3] Estimación óptima (O), [4] Estimación aceptable (A), [5] Estimación imprecisa (I-*sobreestimada*)

Tabla 18: Significación práctica de la estimación de FD por condición (λ_{FD} = 0,3)

Precisión de las estimaciones (% vertical)			ML		ULS			
			x_{10}	x_{11}	x_{10}	x_{11}	x_{10}	x_{11}
FF: λ_1=0,7 y λ_2=0,5	ρ = 0	[1] I-*subestimada*	24,4%	26,4%	26,5%	24,5%	26,6%	26,8%
		[2] A	10,4%	11,2%	10,2%	10,4%	11,2%	10,2%
		[3] O	24,8%	22,0%	26,4%	24,6%	21,8%	26,2%
		[4] A	9,0%	9,1%	7,5%	8,8%	9,2%	7,6%
		[5] I-*sobreestimada*	31,5%	31,3%	29,5%	31,6%	31,1%	29,2%
	ρ = 0,3	[1] I-*subestimada*	25,2%	29,2%	25,1%	29,3%	31,3%	27,2%
		[2] A	10,7%	9,8%	12,2%	11,1%	10,5%	13,1%
		[3] O	25,1%	23,2%	23,0%	21,8%	22,8%	23,1%
		[4] A	10,0%	10,3%	11,5%	10,5%	11,1%	9,7%
		[5] I-*sobreestimada*	29,0%	27,5%	28,2%	27,3%	24,3%	26,9%
FF: λ_1=0,7 y λ_2=0,7	ρ = 0	[1] I-*subestimada*	24,4%	26,4%	26,8%	25,0%	27,1%	27,1%
		[2] A	10,3%	11,3%	10,1%	10,4%	11,4%	10,0%
		[3] O	24,7%	21,9%	26,3%	24,5%	21,4%	26,3%
		[4] A	8,9%	9,2%	7,5%	8,7%	9,4%	7,2%
		[5] I-*sobreestimada*	31,6%	31,1%	29,2%	31,4%	30,7%	29,3%
	ρ = 0,3	[1] I-*subestimada*	24,4%	28,0%	24,7%	29,1%	29,9%	26,0%
		[2] A	10,1%	10,2%	11,0%	11,0%	10,6%	12,6%
		[3] O	26,2%	23,9%	24,6%	22,6%	23,8%	25,4%
		[4] A	11,0%	11,7%	12,0%	10,9%	12,2%	9,7%
		[5] I-*sobreestimada*	28,3%	26,2%	27,7%	26,4%	23,5%	26,3%
FF: λ_1=0,9 y λ_2=0,5	ρ = 0	[1] I-*subestimada*	24,0%	26,0%	26,4%	25,3%	27,6%	28,4%
		[2] A	10,5%	11,5%	10,2%	10,1%	11,2%	9,7%
		[3] O	24,9%	22,0%	26,5%	24,5%	21,3%	26,1%
		[4] A	9,0%	9,1%	7,6%	8,8%	8,9%	7,3%
		[5] I-*sobreestimada*	31,6%	31,4%	29,2%	31,3%	31,0%	28,5%
	ρ = 0,3	[1] I-*subestimada*	23,2%	28,6%	23,5%	29,0%	32,5%	27,2%
		[2] A	11,1%	9,1%	11,7%	12,1%	9,4%	13,5%
		[3] O	26,3%	24,4%	26,0%	22,8%	22,5%	22,2%
		[4] A	10,2%	11,3%	10,4%	8,9%	11,6%	10,1%
		[5] I-*sobreestimada*	29,2%	26,6%	28,4%	27,2%	24,0%	27,0%
FF: λ_1=0,9 y λ_2=0,7	ρ = 0	[1] I-*subestimada*	24,0%	25,8%	26,5%	25,4%	27,5%	28,2%
		[2] A	10,4%	11,5%	10,2%	10,2%	11,3%	9,7%
		[3] O	24,9%	22,0%	26,4%	24,3%	21,4%	26,0%
		[4] A	9,0%	9,2%	7,6%	8,8%	9,0%	7,6%
		[5] I-*sobreestimada*	31,7%	31,5%	29,3%	31,3%	30,8%	28,6%
	ρ = 0,3	[1] I-*subestimada*	23,3%	27,9%	23,0%	27,7%	30,9%	27,3%
		[2] A	10,9%	9,4%	11,7%	13,1%	10,2%	11,4%
		[3] O	27,0%	24,1%	25,9%	22,8%	22,7%	24,6%
		[4] A	11,0%	11,8%	11,3%	10,4%	12,9%	11,3%
		[5] I-*sobreestimada*	27,8%	26,8%	28,1%	26,0%	23,3%	25,4%

[1] Estimación imprecisa (I-*subestimada*), [2] Estimación aceptable (A), [3] Estimación óptima (O), [4] Estimación aceptable (A), [5] Estimación imprecisa (I-*sobreestimada*)

Tabla 19: Significación práctica de la estimación de FD por condición (λ_{FD} = 0,4)

Precisión de las estimaciones (% vertical)			ML		ULS			
			x_{10}	x_{11}	x_{10}	x_{11}	x_{10}	x_{11}
FF: λ_1=0,7 y λ_2=0,5	ρ = 0	[1] I-*subestimada*	24,1%	23,9%	25,9%	23,7%	23,6%	25,4%
		[2] A	11,4%	13,4%	11,8%	11,5%	13,4%	11,8%
		[3] O	27,7%	24,4%	28,8%	28,0%	24,4%	28,9%
		[4] A	11,6%	12,3%	9,7%	11,6%	12,6%	9,8%
		[5] I-*sobreestimada*	25,1%	25,9%	23,9%	25,1%	25,9%	24,2%
	ρ = 0,3	[1] I-*subestimada*	20,0%	24,9%	20,5%	23,8%	27,8%	23,9%
		[2] A	12,8%	11,3%	12,6%	13,8%	11,2%	13,5%
		[3] O	30,8%	28,1%	30,3%	26,0%	26,7%	27,0%
		[4] A	11,7%	13,9%	13,1%	11,8%	12,5%	10,9%
		[5] I-*sobreestimada*	24,7%	21,8%	23,5%	24,6%	21,8%	24,7%
FF: λ_1=0,7 y λ_2=0,7	ρ = 0	[1] I-*subestimada*	23,8%	23,4%	25,3%	24,2%	24,0%	25,7%
		[2] A	11,6%	13,4%	11,8%	11,4%	13,2%	11,7%
		[3] O	27,9%	24,5%	28,9%	28,0%	24,4%	28,7%
		[4] A	11,7%	12,6%	9,8%	11,6%	12,6%	9,7%
		[5] I-*sobreestimada*	25,0%	26,1%	24,3%	24,8%	25,9%	24,2%
	ρ = 0,3	[1] I-*subestimada*	18,8%	23,5%	20,0%	23,6%	27,5%	22,8%
		[2] A	12,6%	13,3%	12,9%	14,0%	10,6%	13,6%
		[3] O	32,4%	27,6%	29,7%	26,6%	27,2%	29,0%
		[4] A	12,6%	14,4%	13,6%	12,6%	13,2%	10,4%
		[5] I-*sobreestimada*	23,6%	21,2%	23,8%	23,2%	21,5%	24,2%
FF: λ_1=0,9 y λ_2=0,5	ρ = 0	[1] I-*subestimada*	23,8%	24,3%	26,1%	23,9%	23,7%	25,7%
		[2] A	11,4%	13,4%	11,8%	11,5%	13,4%	11,7%
		[3] O	28,0%	24,2%	28,9%	28,0%	24,4%	28,7%
		[4] A	11,6%	12,4%	9,4%	11,6%	12,6%	9,6%
		[5] I-*sobreestimada*	25,1%	25,6%	23,8%	24,9%	25,9%	24,2%
	ρ = 0,3	[1] I-*subestimada*	18,3%	24,4%	19,4%	24,9%	29,9%	25,3%
		[2] A	13,4%	10,8%	12,9%	14,1%	10,3%	13,2%
		[3] O	32,0%	28,7%	30,9%	25,0%	24,9%	25,6%
		[4] A	12,4%	14,2%	12,9%	10,8%	13,0%	12,2%
		[5] I-*sobreestimada*	23,9%	21,9%	23,9%	25,2%	21,9%	23,7%
FF: λ_1=0,9 y λ_2=0,7	ρ = 0	[1] I-*subestimada*	23,9%	23,7%	25,6%	24,2%	23,9%	25,9%
		[2] A	11,3%	13,4%	11,8%	11,4%	13,3%	11,6%
		[3] O	28,0%	24,3%	28,8%	27,9%	24,4%	28,7%
		[4] A	11,6%	12,5%	9,7%	11,6%	12,6%	9,7%
		[5] I-*sobreestimada*	25,1%	26,0%	24,2%	24,9%	25,9%	24,1%
	ρ = 0,3	[1] I-*subestimada*	18,4%	23,8%	19,1%	24,1%	28,7%	24,4%
		[2] A	12,4%	10,8%	12,6%	14,5%	11,0%	12,4%
		[3] O	33,6%	29,6%	30,1%	25,8%	25,4%	27,2%
		[4] A	12,5%	14,7%	14,6%	11,6%	14,1%	12,3%
		[5] I-*sobreestimada*	23,1%	21,1%	23,6%	24,0%	20,8%	23,7%

[1] Estimación imprecisa (I-*subestimada*), [2] Estimación aceptable (A), [3] Estimación óptima (O), [4] Estimación aceptable (A), [5] Estimación imprecisa (I-*sobreestimada*)

Tabla 20: Tipo de RFD en función del nivel de debilidad del factor 3 (% marginal-fila), de la correlación entre factores y del método de estimación

Patrones (% vertical)		ML		ULS		TOTAL	
		$\rho = 0$	$\rho = 0{,}3$	$\rho = 0$	$\rho = 0{,}3$	ML	ULS
λ_{FD}=0,2	**[4]**	63,0%	61,7%	64,0%	58,5%	64,1%	62,6%
	[3]	32,0%	31,3%	31,2%	34,1%	30,5%	31,7%
	[2]	4,6%	6,4%	4,5%	6,6%	4,9%	5,0%
	[1]	0,4%	0,7%	0,4%	0,8%	0,6%	0,7%
λ_{FD}=0,3	**[4]**	57,8%	53,4%	58,2%	55,0%	55,3%	56,4%
	[3]	32,4%	36,0%	32,1%	35,7%	34,4%	34,1%
	[2]	8,1%	9,3%	8,0%	8,2%	8,8%	8,1%
	[1]	1,7%	1,4%	1,7%	1,1%	1,5%	1,4%
λ_{FD}=0,4	**[4]**	46,7%	40,3%	46,7%	46,8%	43,4%	46,7%
	[3]	35,8%	41,1%	35,8%	39,3%	38,5%	37,6%
	[2]	14,6%	15,8%	14,6%	12,2%	15,2%	13,4%
	[1]	2,9%	2,9%	2,9%	1,8%	2,9%	2,3%

Tabla 21: Tipo de RFD en función de la carga factorial de los FF y del método de estimación

Patrones (% vertical)		ML				ULS			
		λ_1=0,70 λ_2=0,50	λ_1=0,70 λ_2=0,70	λ_1=0,90 λ_2=0,50	λ_1=0,90 λ_2=0,70	λ_1=0,70 λ_2=0,50	λ_1=0,70 λ_2=0,70	λ_1=0,90 λ_2=0,50	λ_1=0,90 λ_2=0,70
λ_{FD}=0,2	**[4]**	63,2%	61,5%	62,8%	61,0%	61,2%	60,2%	61,0%	59,3%
	[3]	30,9%	32,0%	31,4%	31,9%	32,6%	32,9%	33,1%	33,7%
	[2]	5,3%	5,9%	5,2%	6,5%	5,6%	6,1%	5,4%	6,4%
	[1]	0,6%	0,5%	0,6%	0,5%	0,6%	0,8%	0,5%	0,6%
λ_{FD}=0,3	**[4]**	56,4%	55,2%	54,7%	54,7%	56,4%	55,8%	57,2%	56,2%
	[3]	33,6%	34,1%	35,4%	34,6%	34,5%	34,4%	34,0%	33,5%
	[2]	8,8%	9,1%	8,2%	9,0%	7,8%	8,4%	7,4%	9,0%
	[1]	1,2%	1,5%	1,7%	1,7%	1,3%	1,5%	1,4%	1,3%
λ_{FD}=0,4	**[4]**	44,2%	43,4%	43,2%	42,9%	46,5%	45,8%	47,9%	46,8%
	[3]	38,1%	38,3%	38,8%	38,7%	37,9%	38,3%	36,9%	37,2%
	[2]	14,6%	15,4%	15,2%	15,6%	13,3%	13,6%	13,1%	13,5%
	[1]	3,0%	2,9%	2,8%	2,9%	2,3%	2,4%	2,1%	2,5%

[1] RFD óptima: estimación conjunta del FD entre ± 0,05 respecto al valor teórico. INTETRPRETACIÓN DIRECTA.
[2] RFD precisa: sin estimaciones I, soluciones mixtas con estimaciones O y A. INTETRPRETACIÓN DIRECTA.
[3] RFD imprecisa: soluciones con alguna estimación I. EVALUACIÓN DIMENSIONALIDAD.
[4] RFD inaceptable: soluciones con 2 o 3 estimaciones I. EVALUACIÓN DIMENSIONALIDAD.

Tabla 22: Tipo de RFD y promedio de C_k y $RMSD_k$ en función de la correlación entre factores y del método de estimación

		ML				ULS			
		$\rho = 0$		$\rho = 0{,}3$		$\rho = 0$		$\rho = 0{,}3$	
		C_k	$RMSD_k$	C_k	$RMSD_k$	C_k	$RMSD_k$	C_k	$RMSD_k$
λ_{FD}=0,2	**[4]**	,827	,212	,829	,157	,797	,216	,822	,161
	[3]	,849	,221	,888	,140	,849	,221	,878	,145
	[2]	,977	,072	,972	,073	,978	,069	,973	,087
	[1]	,992	,044	,994	,040	,992	,044	,994	,073
λ_{FD}=0,3	**[4]**	,768	,261	,786	,175	,734	,266	,784	,174
	[3]	,936	,142	,931	,120	,937	,142	,928	,128
	[2]	,988	,075	,987	,073	,988	,075	,987	,080
	[1]	,997	,051	,997	,059	,997	,049	,996	,062
λ_{FD}=0,4	**[4]**	,758	,273	,777	,178	,725	,278	,785	,174
	[3]	,965	,116	,954	,116	,965	,116	,951	,124
	[2]	,993	,070	,993	,069	,993	,070	,992	,073
	[1]	,998	,042	,998	,048	,998	,042	,998	,047

Tabla 23: Tipo de RFD y promedio de C_k y $RMSD_k$ en función de la carga factorial de los FF y ML

Método ML		λ_1=0,70 λ_2=0,50		λ_1=0,70 λ_2=0,70		λ_1=0,90 λ_2=0,50		λ_1=0,90 λ_2=0,70	
		C_k	$RMSD_k$	C_k	$RMSD_k$	C_k	$RMSD_k$	C_k	$RMSD_k$
λ_{FD}=0,2	**[4]**	,812	,186	,825	,185	,833	,177	,842	,173
	[3]	,867	,180	,870	,176	,870	,178	,872	,175
	[2]	,975	,072	,974	,070	,974	,077	,971	,071
	[1]	,993	,041	,993	,052	,993	,029	,995	,042
λ_{FD}=0,3	**[4]**	,761	,219	,774	,219	,784	,209	,795	,204
	[3]	,932	,132	,934	,129	,932	,130	,935	,128
	[2]	,988	,072	,988	,073	,988	,075	,987	,075
	[1]	,997	,060	,997	,055	,997	,051	,997	,056
λ_{FD}=0,4	**[4]**	,751	,226	,765	,225	,774	,216	,785	,211
	[3]	,958	,117	,960	,115	,958	,117	,959	,116
	[2]	,992	,070	,993	,070	,993	,069	,993	,069
	[1]	,998	,044	,998	,045	,998	,044	,998	,046

[1] RFD óptima: estimación conjunta del FD entre ± 0,05 respecto al valor teórico. INTETRPRETACIÓN DIRECTA.
[2] RFD precisa: sin estimaciones I, soluciones mixtas con estimaciones O y A. INTETRPRETACIÓN DIRECTA.
[3] RFD imprecisa: soluciones con alguna estimación I. EVALUACIÓN DIMENSIONALIDAD.
[4] RFD inaceptable: soluciones con 2 o 3 estimaciones I. EVALUACIÓN DIMENSIONALIDAD.

Tabla 24: Tipo de RFD y promedio de C_k y $RMSD_k$ en función de la carga factorial de los FF y ULS

Método ULS		λ_1=0,70 λ_2=0,50		λ_1=0,70 λ_2=0,70		λ_1=0,90 λ_2=0,50		λ_1=0,90 λ_2=0,70	
		C_k	$RMSD_k$	C_k	$RMSD_k$	C_k	$RMSD_k$	C_k	$RMSD_k$
λ_{FD}=0,2	**[4]**	,812	,185	,819	,183	,800	,188	,815	,185
	[3]	,863	,180	,867	,177	,865	,180	,866	,177
	[2]	,974	,082	,974	,080	,976	,083	,974	,084
	[1]	,995	,054	,994	,064	,992	,089	,994	,068
λ_{FD}=0,3	**[4]**	,761	,214	,771	,211	,751	,216	,767	,213
	[3]	,933	,134	,932	,133	,931	,136	,932	,134
	[2]	,987	,079	,988	,076	,988	,077	,987	,078
	[1]	,997	,057	,997	,054	,996	,059	,997	,053
λ_{FD}=0,4	**[4]**	,758	,219	,767	,215	,748	,220	,765	,216
	[3]	,958	,120	,958	,120	,957	,121	,957	,122
	[2]	,992	,071	,993	,070	,992	,072	,992	,073
	[1]	,998	,043	,998	,044	,998	,044	,998	,044

[1] RFD óptima: estimación conjunta del FD entre ± 0,05 respecto al valor teórico. INTETRPRETACIÓN DIRECTA.
[2] RFD precisa: sin estimaciones I, soluciones mixtas con estimaciones O y A. INTETRPRETACIÓN DIRECTA.
[3] RFD imprecisa: soluciones con alguna estimación I. EVALUACIÓN DIMENSIONALIDAD.
[4] RFD inaceptable: soluciones con 2 o 3 estimaciones I. EVALUACIÓN DIMENSIONALIDAD.