

# Comparación de los métodos de estimación de máxima verosimilitud y mínimos cuadrados no ponderados en el análisis factorial confirmatorio mediante simulación Monte Carlo

M. Carmen Ximénez y Ana G. García  
Universidad Autónoma de Madrid

En este artículo se investiga la recuperación de factores débiles en el contexto del análisis factorial confirmatorio. La investigación previa se limita al caso del análisis factorial exploratorio. El estudio se realiza mediante simulación Monte Carlo con las siguientes condiciones: comparación entre los métodos de estimación de máxima verosimilitud (ML) y mínimos cuadrados no ponderados (ULS), tamaño muestral (100, 300, 500, 1.000 y 2.000) y nivel de debilidad del factor (saturaciones de 0.25, 0.40 y 0.50). Los resultados indican que con tamaños muestrales pequeños el método ULS recupera el factor débil en muchos de los casos en que ML falla. Esta ventaja está relacionada con la ocurrencia de casos Heywood. Asimismo, la recuperación del factor débil mejora a medida que el nivel de debilidad es menor y que el modelo incluye mayor número de factores.

*Comparison of maximum likelihood and unweighted least squares estimation methods in confirmatory factor analysis by Monte Carlo simulation.* This article examines the recovery of weak factors in the context of confirmatory factor analysis. Previous research only refers to exploratory factor analysis. The study is done by Monte Carlo simulation with the following conditions: comparison of maximum likelihood (ML) and unweighted least squares (ULS) estimation methods, sample size (100, 300, 500, 1.000 and 2.000) and level of factor weakness (loadings of 0.25, 0.40 and 0.50). Results show that with small sample sizes ML failed to recover the weak factor while ULS succeed in many cases. This advantage is related to the occurrence of Heywood cases. Also the weak factor recovery improves as the level of factor weakness decreases and the number of factors in the model increases.

El análisis factorial confirmatorio (AFC) ha proporcionado los medios para construir y evaluar modelos teóricos (Browne, 1984). El modelo de AFC se expresa mediante:

$$\mathbf{x} = \Lambda \boldsymbol{\xi} + \boldsymbol{\delta} \quad (1)$$

Donde  $\mathbf{x}$  es un vector de  $p$  variables observadas,  $\boldsymbol{\xi}$  es un vector de  $q$  factores (donde  $q < p$ ),  $\Lambda$  es la matriz  $p \times q$  de saturaciones factoriales y  $\boldsymbol{\delta}$  el vector de  $p$  errores de medida. Se asume que  $E(\mathbf{x}) = E(\boldsymbol{\xi}) = E(\boldsymbol{\delta}) = \mathbf{0}$  y que  $E(\boldsymbol{\xi}\boldsymbol{\delta}') = \mathbf{0}$ . La matriz de covarianzas de  $\mathbf{x}$  es:

$$\Sigma = \Lambda \Phi \Lambda' + \Theta_{\delta} \quad (2)$$

Donde  $\Phi$  es la matriz  $q \times q$  de covarianzas de los  $\boldsymbol{\xi}$  y  $\Theta_{\delta}$  la matriz  $p \times p$  de covarianzas de los  $\boldsymbol{\delta}$ . Aquí se asume que  $\Phi = \mathbf{I}$  y que  $\Theta_{\delta}$  es diagonal.

Cuando se lleva a cabo un AFC el investigador debe decidir sobre la especificación del modelo, la identificación, el método de

estimación de parámetros y el procedimiento para evaluar el ajuste del modelo (Bollen, 1989). El presente trabajo se centra en el método de estimación de parámetros. Si las observaciones de  $\mathbf{x}$  son independientes y siguen la distribución normal multivariante, suele utilizarse el método de máxima verosimilitud (ML) por sus propiedades asintóticas (proporciona estimadores insesgados, consistentes, eficientes y normalmente distribuidos). Cuando se viola el supuesto de normalidad, uno de los métodos más empleados es el de mínimos cuadrados no ponderados (ULS). Los métodos ML y ULS difieren en sus supuestos y en su función de ajuste. A no ser que el modelo sea saturado (0 grados de libertad) o que su ajuste sea perfecto (es decir, que la función de ajuste sea nula en ambos casos), proporcionan estimaciones ligeramente diferentes (para una descripción detallada de ambos métodos véase Bollen, 1989, pp. 104-113).

En este trabajo se comparan los métodos ML y ULS en relación a la recuperación de factores que presentan saturaciones bajas. Estos factores, denominados *débiles*, se definen aquí como aquellos con saturaciones entre 0.25 y 0.50. Las diferencias entre los métodos ML y ULS en la recuperación de factores han sido ya estudiadas en la investigación previa, pero solamente en el contexto del análisis factorial exploratorio (AFE). Los resultados sugieren que el tipo de error en los datos puede afectar a la recuperación de los factores (Briggs y MacCallum, 2003; MacCallum y Tucker, 1991).

Estudio de simulación

El presente trabajo es continuación del estudio de Briggs y MacCallum (2003) que examina la efectividad de los métodos ML y ULS en el contexto del AFE mediante simulación Monte Carlo. Definen el factor débil con saturaciones entre 0.40 y 0.50 y utilizan tres modelos: el de MacCallum y Tucker (1991), que incluye 12 variables y 3 factores, siendo uno de ellos débil; y dos con 16 variables y 4 factores, uno con un factor débil y otro con dos. Consideran dos tipos de error: error en el modelo (refiriéndose a la falta de ajuste del modelo) y error muestral (definido por el tamaño de la muestra). Encontraron que al introducir error en el modelo, los métodos ML y ULS eran igualmente eficaces si el ajuste era bueno. Si el ajuste era moderado o bajo, ULS recuperaba mejor el factor débil que ML y en las ocasiones en que ML fallaba, ULS no lo hacía. Además, ambos métodos eran igualmente eficaces con muestras grandes pero ULS recuperaba mejor el factor débil con muestras pequeñas ( $n= 100$ ). Concluyeron que esta mejora está asociada a la presencia de casos Heywood. Como los propios autores reconocen (ver p. 53), su estudio tiene dos limitaciones: los factores débiles se definen con saturaciones más altas que las encontradas en la práctica (entre 0.20 y 0.30), el estudio se refiere solamente al AFE y los resultados no pueden generalizarse al caso confirmatorio. El estudio de Olsson, Troye y Howell (1999) sí se refiere al AFC pero compara el método ML con el de mínimos cuadrados generalizados (GLS). Encontraron que ML ofrece mejores estimaciones cuando el ajuste no es perfecto.

En este trabajo se abordan algunas de las limitaciones señaladas por Briggs y MacCallum (2003). Se compara la efectividad de los métodos de estimación ML y ULS en la recuperación de factores débiles en el contexto del AFC. Asimismo, los factores débiles se definen con saturaciones entre 0.25 y 0.50, por lo que se tienen en cuenta distintos niveles de debilidad del factor. Se lleva a cabo un estudio de simulación Monte Carlo para examinar si el tipo de método, el nivel de debilidad del factor y el tamaño de la muestra afectan a la recuperación de factores débiles. El estudio de estos efectos en el contexto del AFC es relevante debido a la gran cantidad de estudios prácticos que utilizan esta técnica (valga como ejemplo los trabajos de Hernández y González-Romá, 2003; o el de Olea, Abad, Ponsoda y Ximénez, 2004).

Para exponer las características del estudio de simulación se sigue el esquema recomendado por Skrondal (2000, pp. 140-141) que propone los seis siguientes pasos para el diseño de un estudio de simulación Monte Carlo en modelos de ecuaciones estructurales:

- (1) *Formulación de problema de investigación.* En este trabajo se estudia el efecto del tipo de método, el nivel de debilidad del factor y el tamaño de la muestra sobre la recuperación de factores débiles en el contexto del AFC.
- (2) *Plan experimental.* El diseño consta de tres variables independientes: tipo de método (ML vs. ULS), nivel de debilidad del factor (saturaciones en el factor débil de 0.25, 0.40 y 0.50) y tamaño muestral (se utilizan muestras de 100, 300, 500, 1.000 y 2.000 observaciones). La variable dependiente es la recuperación del factor débil (se define en el paso 6). Se incluyen todas las combinaciones posibles de tratamientos. Se trata por tanto de un diseño factorial  $2 \times 3 \times 5$  con 30 combinaciones. Se definen tres estructuras factoriales teóricas basadas en el modelo de McCallum y Tucker (1991), que emplea 12 variables observadas normales y 3 factores, siendo el último el factor débil. Las otras dos estructuras teóricas se definen a partir de las mismas variables pero con 1 y 2 factores. Los valores teóricos de cada estructura se muestran en la tabla 1.
- (3) *Simulación.* Las matrices de saturaciones de la tabla 1 se toman como base para simular las matrices de correlaciones en cada tamaño muestral. Para cada condición se simulan 1,000 matrices utilizando el programa PRELIS 2 de Jöreskog y Sörbom (1996a).
- (4) *Estimación.* Para cada matriz de correlaciones simulada se lleva a cabo el AFC con los métodos ML y ULS. Se utiliza la técnica de reducción de la varianza de los números aleatorios comunes, consistente en utilizar los mismos valores aleatorios para todas las condiciones de la variable método (Reuelta y Ponsoda, 2003). Para obtener las estimaciones de parámetros se utiliza el programa LISREL 8.53 de Jöreskog y Sörbom (1996b).
- (5) *Réplica.* Los pasos 3 y 4 se repiten para cada una de las réplicas ( $r = 1, \dots, 30,000$ ).

Tabla 1  
Parámetros teóricos para la simulación de matrices de correlaciones en cada modelo

	Modelo de un factor		Modelo de dos factores		Modelo de tres factores		
	Factor 1		Factor 1	Factor 2	Factor 1	Factor 2	Factor 3
X <sub>1</sub>	0.25		0.80	0	0.95	0	0
X <sub>2</sub>	0.25		0.80	0	0.95	0	0
X <sub>3</sub>	0.25		0.80	0	0.95	0	0
X <sub>4</sub>	0.25		0.80	0	0.95	0	0
X <sub>5</sub>	0.25		0.80	0	0.95	0	0
X <sub>6</sub>	0.25		0.80	0	0	0.70	0
X <sub>7</sub>	0.25		0.80	0	0	0.70	0
X <sub>8</sub>	0.25		0	0.40	0	0.70	0
X <sub>9</sub>	0.25		0	0.40	0	0.70	0
X <sub>10</sub>	0.25		0	0.40	0	0	0.50
X <sub>11</sub>	0.25		0	0.40	0	0	0.50
X <sub>12</sub>	0.25		0	0.40	0	0	0.50

Nota: las saturaciones en los factores débiles se muestran en negrita.

- (6) *Análisis de resultados.* En primer lugar, se eliminan las soluciones no-convergentes (según el criterio de LISREL: tras 250 iteraciones) y se contabilizan los casos Heywood.

Para estudiar la recuperación de factores débiles se emplean dos coeficientes de comparación de factores. El primero, el *coeficiente de congruencia* ( $\phi$ ) atribuido a Burt (1949) y definido por Tucker (1951) y Tucker y Lewis (1973). Se calcula mediante la siguiente expresión:

$$\phi_k = \frac{\sum_{i=1}^p \lambda_{ik(t)} \lambda_{ik(e)}}{\sqrt{\left( \sum_{i=1}^p \lambda_{ik(t)}^2 \right) \left( \sum_{i=1}^p \lambda_{ik(e)}^2 \right)}} \quad (3)$$

Donde  $p$  es el número de variables que definen el factor  $k$ ,  $\lambda_{ik(t)}$  es la saturación teórica para la variable observada  $i$  en el factor  $k$  y  $\lambda_{ik(e)}$  la correspondiente saturación estimada. Este coeficiente evalúa la similaridad en la forma y la magnitud de la saturación teórica y la estimada. Oscila entre  $-1$  y  $+1$  y se interpreta como el coeficiente de correlación de Pearson.

El segundo, el índice *RMSD* (*Root Mean Square Deviation*) de Levine (1977). Se calcula mediante la siguiente expresión:

$$\delta_k = \sqrt{\frac{\sum_{i=1}^p (\lambda_{ik(t)} - \lambda_{ik(e)})^2}{p}} \quad (4)$$

El índice  $\delta$  captura las diferencias en magnitud en todas las direcciones. Su valor oscila entre 0 y 2, siendo 0 cuando todas las saturaciones son iguales y 2 si valen 1 pero tienen signo opuesto. Cuanto menor sea el valor de  $\delta$ , mayor es la congruencia entre la saturación teórica y la estimada. Para ambos índices se obtienen los estadísticos descriptivos.

Siguiendo la recomendación de Skrondal (2000), se utiliza un modelo sencillo para analizar los resultados, que sólo incluye el efecto principal de cada factor y la doble interacción sobre la variable dependiente. El modelo es el siguiente:

$$\text{RFD} = \mu + \text{Método} + \text{Factor} + n + \text{Método} \times \text{factor} + \text{Método} \times n + \text{Factor} \times n \quad (5)$$

Donde: RFD: recuperación del factor débil: (coeficiente  $\phi$  e índice  $\delta$ )

Método: ML vs. ULS

Factor: nivel de debilidad del factor (0.25, 0.40 y 0.50)

$n$ : tamaño de la muestra (100, 300, 500, 1000 y 2000)

Para contrastar los efectos incluidos en este modelo se lleva a cabo un análisis de varianza (ANOVA). Dado el gran número de observaciones ( $N=30,000$ ) los resultados incluyen las pruebas de significación estadística para cada uno de los efectos y también las de significación práctica, medidas mediante el estadístico  $\eta^2$  (eta cuadrado) que se interpreta como la proporción de varianza explicada. También se realizan las comparaciones múltiples para los efectos que resultan significativos. Los análisis mencionados en este apartado se llevan a cabo con el programa SPSS 12.

## Resultados

Sólo se encontraron soluciones que no convergen en las muestras con  $n=100$ . En el modelo de un factor, que incluye el factor más débil, 60 muestras no convergen con el método ML y 11 con el método ULS. Asimismo, en 53 muestras (el 88%) en las que ML no converge, ULS sí lo hace. En el modelo de dos factores, 31 muestras no convergen con ML y 13 con ULS, ocurriendo que en 19 muestras (el 61%) cuando ML no converge, ULS sí lo hace. En el modelo de tres factores, que incluye el factor menos débil, 12 muestras no convergen con ningún método. Como se observa, las soluciones no-convergentes son más frecuentes con el método de estimación ML y con saturaciones del factor débil entre 0.25 y 0.40. Este resultado es congruente con lo encontrado por Anderson y Gerbing (1984). Todas las soluciones no-convergentes fueron eliminadas. En cuanto a los casos Heywood, también se encuentra que sólo aparecen en las muestras con  $n=100$ . En el modelo de un factor se producen 84 casos con el método ML y 16 con el método ULS, en el de dos factores 42 con ML y 18 con ULS y en el de tres factores 65 con ambos métodos. En el modelo de un factor, en el 88% de las muestras cuando el método ML produce casos Heywood, el método ULS no lo hace, y en el de dos factores dicho porcentaje es del 60%.

La tabla 2 muestra la media y la desviación típica para el coeficiente de congruencia ( $\phi$ ) y el índice *RMSD* ( $\delta$ ) para el factor débil en todas las condiciones del estudio.

Globalmente los resultados indican que en todas las condiciones se obtienen promedios de los índices de recuperación del factor débil aceptables, excepto en el modelo de un factor con  $n=100$ , donde la media del coeficiente  $\phi$  es menor de 0.80 y la del índice  $\delta$  mayor de 0.20. Se observa que en el modelo de 3 factores los resultados son idénticos con ambos métodos. Sin embargo, en los de 1 y 2 factores, se obtienen mejores resultados con el método ULS y el error típico es menor. Asimismo, a medida que aumenta el tamaño muestral disminuye esta diferencia y no se produce con muestras grandes. Siguiendo lo encontrado por Briggs y MacCallum (2003), se analiza si la ventaja del método ULS en las soluciones con  $n=100$  está asociada con la ocurrencia de casos Heywood. La parte inferior de la tabla 2 muestra que en todos los modelos se produce un aumento en la media de  $\phi$  y una disminución en la de  $\delta$  al eliminar los casos Heywood, y que en los modelos de 1 y 2 factores el método ULS proporciona mejores resultados.

Un análisis más detallado de los casos Heywood en las muestras con  $n=100$  revela un patrón interesante. Al seleccionar las soluciones en las que tanto el método ML como ULS producen casos Heywood, se observa que la recuperación del factor débil es pobre, independientemente del método de estimación. Al seleccionar aquellas donde sólo el método ML produce casos Heywood, se observa que ML falla en recuperar el factor débil en todos los casos mientras que ULS lo recupera adecuadamente en el 60% de los casos. Por último, al seleccionar las soluciones donde sólo el método ULS produce casos Heywood, no hay diferencias entre los métodos y ambos recuperan adecuadamente el factor débil.

La figura 1 muestra las diferencias entre los métodos en más detalle, pues contiene los gráficos de dispersión para el coeficiente  $\phi$  en cada uno de los modelos con  $n=100$  en todas las réplicas (panel izquierdo) y eliminando las que presentan casos Heywood (panel derecho). Cada gráfico muestra un patrón distintivo de puntos. La mayoría están concentrados en la parte superior derecha, indicando las réplicas en que ML y ULS obtienen coeficientes  $\phi$

mayores de 0.80 y por tanto recuperan el factor débil adecuadamente. Sin embargo, en muchos otros casos (puntos por debajo de 0.80 en el eje de ordenadas y por encima de 0.80 en el de abscisas) se observa como ULS recupera el factor débil adecuadamente y ML no. Por último, en algunos casos los coeficientes  $\phi$  son bajos pues ninguno de los métodos recupera el factor débil adecuadamente. Al eliminar los casos Heywood se observa cómo desaparecen algunos de los dos últimos casos. Este patrón es especialmente pronunciado en el modelo de un factor débil, aunque también se observa en el modelo de dos factores, donde los casos en que el método ULS recupera el factor débil adecuadamente y el método ML no, desaparecen al eliminar los casos Heywood.

La tabla 3 muestra los resultados del ANOVA que pone a prueba el modelo planteado en la ecuación (5) en todas las réplicas y eliminando las que presentan casos Heywood.

Como se observa, en relación al coeficiente  $\phi$  en todas las réplicas el modelo explica un 35.10% de la varianza. El efecto del tipo de método no es estadísticamente significativo ( $p= 0.273$ ) cuando todas las variables independientes se introducen simultáneamente y tampoco el de las dobles interacciones. El mayor efecto encontrado es el del tamaño muestral ( $\eta^2= 0.158$ ), seguido del de la interacción tamaño muestral  $\times$  debilidad del factor ( $\eta^2= 0.121$ ) y del de la debilidad del factor ( $\eta^2= 0.072$ ). Las comparaciones múltiples para analizar el efecto del tamaño muestral indican que existen diferencias estadísticamente significativas entre todas las medias excepto las de  $n= 1.000$  y  $2.000$ , por lo que el factor débil se recupera mejor a medida que aumenta el tamaño de la muestra. Este resultado es congruente con lo encontrado por MacCallum, Widaman, Zhang y Hong (1999). En cuanto al efecto del nivel de debilidad del factor, todas las diferencias resultan significativas, siendo:

Tabla 2  
Medias y desviaciones típicas del coeficiente de congruencia ( $\phi$ ) y el índice RMSD ( $\delta$ ) de los factores débiles en el estudio Monte Carlo

	$\phi$				$\delta$			
	ML		ULS		ML		ULS	
	M	D. T.	M	D. T.	M	D. T.	M	D. T.
<b>En todas las réplicas:</b>								
Modelo de un factor								
$n= 100$	0.699	0.406	0.707	0.424	0.197	0.103	0.187	0.098
$n= 300$	0.935	0.104	0.940	0.085	0.090	0.029	0.088	0.026
$n= 500$	0.966	0.017	0.967	0.016	0.067	0.014	0.066	0.014
$n= 1.000$	0.983	0.007	0.984	0.007	0.047	0.009	0.046	0.009
$n= 2.000$	0.992	0.004	0.992	0.004	0.033	0.007	0.033	0.007
Total	0.917	0.213	0.918	0.205	0.085	0.075	0.084	0.071
Modelo de dos factores								
$n= 100$	0.933	0.100	0.939	0.094	0.148	0.075	0.141	0.063
$n= 300$	0.982	0.014	0.982	0.013	0.080	0.026	0.078	0.025
$n= 500$	0.989	0.008	0.989	0.008	0.061	0.020	0.060	0.020
$n= 1000$	0.995	0.004	0.995	0.004	0.043	0.014	0.043	0.014
$n= 2000$	0.997	0.002	0.997	0.002	0.030	0.010	0.030	0.010
Total	0.979	0.051	0.981	0.047	0.072	0.055	0.070	0.050
Modelo de tres factores								
$n= 100$	0.955	0.058	0.955	0.058	0.156	0.111	0.156	0.111
$n= 300$	0.987	0.016	0.987	0.016	0.081	0.040	0.081	0.040
$n= 500$	0.992	0.009	0.992	0.009	0.062	0.029	0.062	0.029
$n= 1000$	0.996	0.004	0.996	0.004	0.044	0.019	0.044	0.019
$n= 2000$	0.998	0.002	0.998	0.002	0.030	0.013	0.030	0.013
Total	0.986	0.032	0.986	0.032	0.075	0.071	0.075	0.071
<b>Sin casos Heywood:</b>								
Modelo de un factor								
$n= 100$	0.708	0.405	0.709	0.423	0.190	0.094	0.186	0.096
Total	0.921	0.208	0.919	0.219	0.084	0.070	0.083	0.070
Modelo de dos factores								
$n= 100$	0.936	0.095	0.940	0.092	0.144	0.061	0.139	0.058
Total	0.980	0.048	0.981	0.046	0.071	0.052	0.070	0.049
Modelo de tres factores								
$n= 100$	0.965	0.039	0.965	0.039	0.137	0.066	0.137	0.066
Total	0.988	0.022	0.988	0.022	0.070	0.053	0.070	0.053
Nota: M es la media y D. T. la desviación típica.								

$\phi_{\lambda_{ij}=0.50} > \phi_{\lambda_{ij}=0.40} > \phi_{\lambda_{ij}=0.25}$ . Es decir, el factor débil se recupera mejor a medida que disminuye su debilidad. Por último, las comparaciones para la interacción factor  $\times n$  indican que para el mayor nivel de debilidad ( $\lambda_{ij}=0.25$ ) existen diferencias entre todas las medias salvo las de  $n=1.000$  y  $2.000$  ( $\eta^2=0.233$ ), para el nivel medio ( $\lambda_{ij}=0.40$ ) sólo hasta  $n=500$  ( $\eta^2=0.013$ ) y para el menor nivel ( $\lambda_{ij}=0.50$ ) entre  $n=100$  y los restantes tamaños muestrales ( $\eta^2=0.007$ ). Este hallazgo es congruente con lo encontrado por MacCallum y Tucker (1991) en el contexto del AFE que indica que con communalidades altas, se obtienen soluciones estables con  $n$  pequeños.

Con el índice  $\delta$ , los resultados son similares a los ya comentados para el coeficiente  $\phi$ , con la excepción de que el efecto del tipo de método sí resulta estadísticamente significativo ( $p=0.005$ ) y también el de la interacción método  $\times n$  ( $p=0.004$ ). Una posible razón por la que se producen resultados diferentes con  $\phi$  y  $\delta$  puede tener que ver con los casos en que uno de los métodos converge y el otro no. La figura 2 incluye los gráficos de dispersión de  $\phi$  y  $\delta$  según la convergencia de cada método en los modelos de 1 y 2 factores con  $n=100$ . Las líneas discontinuas limitan las muestras donde el factor débil se recupera adecuadamente y donde no. En general, existe una relación fuerte entre  $\phi$  y  $\delta$  (valores altos de  $\phi$  les corresponde valores bajos en  $\delta$  y viceversa). En cuanto al método ULS, se obtienen valores muy altos en  $\delta$  en aquellas muestras donde ML no converge. De hecho, ULS recupera adecuadamente el factor débil sólo en el 10% de esas muestras. Esto indica que, aunque ULS converge, lo hace en una solución muy alejada del valor teórico. Este resultado es más acusado en el modelo de dos factores. Con respecto al método ML, se aprecia que  $\delta$  toma valores extremos con mucha más frecuencia que al utilizar ULS. Hay pocas muestras donde ULS no converge y ML sí, pero aun así, ML proporciona valores muy ex-

tremos de  $\phi$  y  $\delta$ . Además, ML produce soluciones extremas en muchas muestras en las que ULS converge. Los resultados del ANOVA se ven afectados por estos casos en que se produce convergencia a valores extremos. Por ello se repitieron los análisis utilizando sólo los casos donde no hay problemas de convergencia. Los resultados aparecen en la parte inferior de la tabla 3 e indican que con ambos índices el efecto del tipo de método y de la interacción método  $\times n$  son significativos. Las comparaciones múltiples revelan que las medias de  $\delta$  obtenidas con ULS son menores pero sólo para  $n=100$  ( $\eta^2=0.001$ ). No obstante, este efecto no es suficientemente grande para ser útil en la práctica y desaparece al eliminar los casos Heywood.

Discusión y conclusiones

En este artículo se examina la efectividad de los métodos de estimación ML y ULS en relación a la recuperación de factores débiles a partir de matrices de correlaciones con una estructura factorial conocida en el contexto del AFC mediante simulación Monte Carlo. El estudio es continuación del de Briggs y MacCallum (2003) que se limita al AFE y define el factor débil con saturaciones entre 0.40 y 0.50. El presente estudio examina también el efecto del tamaño muestral y considera un rango más amplio de nivel de debilidad del factor (de 0.25 a 0.50). Los resultados indican que con tamaños muestrales pequeños (p.e.  $n=100$ ) el método ULS recupera el factor débil en muchos de los casos en que el método ML falla. Esta ventaja del método ULS desaparece a medida que el tamaño de la muestra aumenta y está relacionada con la ocurrencia de casos Heywood. En las soluciones en que solamente el método ML produce casos Heywood, el factor débil no se recupera adecuadamente en ninguno de los casos mientras que el mé-

Tabla 3  
Resultados de los ANOVA para el coeficiente de congruencia ( $\phi$ ) y el índice RMSD ( $\delta$ ) en el estudio Monte Carlo

	En todas las réplicas				Eliminando los casos Heywood			
	$\phi$		$\delta$		$\phi$		$\delta$	
	$p$	$\eta^2$	$p$	$\eta^2$	$p$	$\eta^2$	$p$	$\eta^2$
<b>Soluciones que convergen con algún método:</b>								
Método	0.273	0.000	0.005	0.000	0.529	0.000	0.024	0.000
Factor	0.000	0.072	0.000	0.016	0.000	0.074	0.000	0.028
$n$	0.000	0.158	0.000	0.503	0.000	0.146	0.000	0.563
Método $\times$ Factor	0.683	0.000	0.125	0.000	0.904	0.000	0.294	0.000
Método $\times n$	0.752	0.000	0.004	0.001	0.983	0.000	0.156	0.000
Factor $\times n$	0.000	0.121	0.000	0.023	0.000	0.123	0.000	0.044
Total:		0.351		0.543		0.343		0.635
<b>Soluciones que convergen con ambos métodos:</b>								
Método	.014	.000	.000	.000	.050	.000	.010	.000
Factor	.000	.069	.000	.014	.000	.071	.000	.025
$n$	.000	.153	.000	.494	.000	.142	.000	.553
Método $\times$ Factor	.160	.000	.024	.000	.365	.000	.064	.000
Método $\times n$	.019	.000	.000	.001	.202	.000	.002	.001
Factor $\times n$	.000	.114	.000	.020	.000	.117	.000	.039
Total:		0.336		0.529		0.330		0.618

Nota: la tabla incluye los resultados de las pruebas de significación para el estadístico F del ANOVA y los valores de  $\eta^2$  (proporción de varianza explicada).

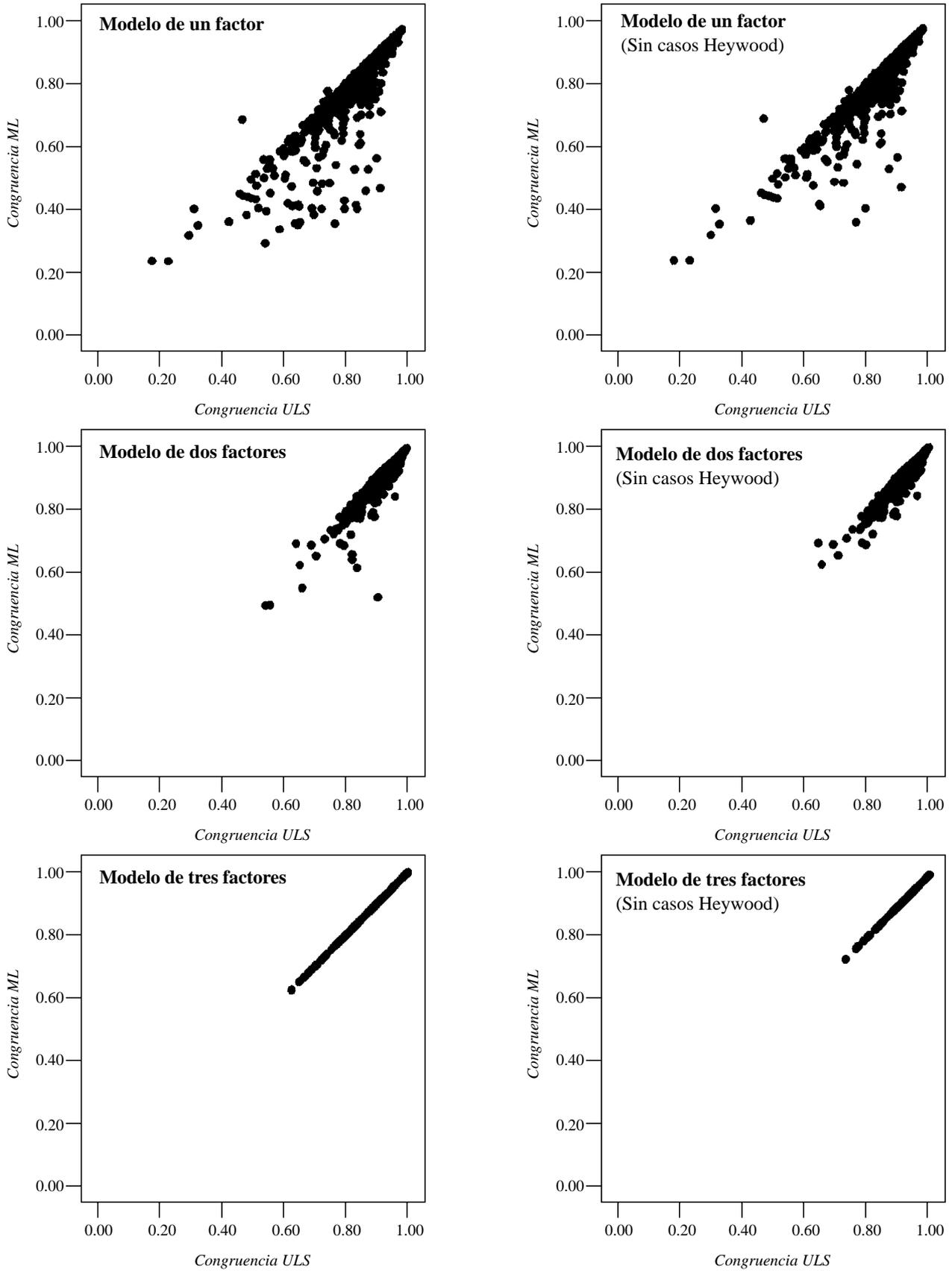


Figura 1. Coeficiente de congruencia para el factor débil en las soluciones ML y ULS en cada modelo con  $n=100$  (con y sin casos Heywood)

todo ULS sí lo recupera satisfactoriamente en la mayoría. Otro resultado interesante es que la recuperación del factor débil mejora a medida que el nivel de debilidad del factor disminuye y que el modelo incluye mayor número de factores. De hecho, con saturaciones a partir de 0.40 el factor débil se recupera adecuadamente incluso con muestras de  $n=100$ .

Los resultados del presente trabajo son congruentes con los encontrados por Briggs y MacCallum (2003) en el contexto del AFE. No obstante, resulta difícil establecer comparaciones entre ambos estudios puesto que este trabajo sólo se refiere a modelos correctamente especificados. Estamos llevando a cabo otro estudio donde se examina si la falta de especificación del modelo afecta a la recuperación del factor débil. Otro aspecto, no incluido en este trabajo y que sería interesante examinar, es si el imponer distintos ti-

pos de restricciones en los parámetros del modelo afecta a la recuperación del factor débil.

En síntesis, los resultados de este trabajo sugieren que pese a que el método ML produzca estimaciones más eficientes que ULS por sus propiedades asintóticas, puede fallar en la recuperación de los factores más débiles, especialmente si sus saturaciones toman valores entre 0.25 y 0.40, si el tamaño de la muestra es pequeño y si se producen casos Heywood. Aumentar el tamaño de la muestra puede evitar este problema. De lo contrario, en estas condiciones, se recomienda utilizar el método ULS.

Agradecimientos

Este trabajo está financiado por los proyectos PBR-541A-2-640 de la DGICYT y 06/HSE/0005/2004 de la Comunidad de Madrid.

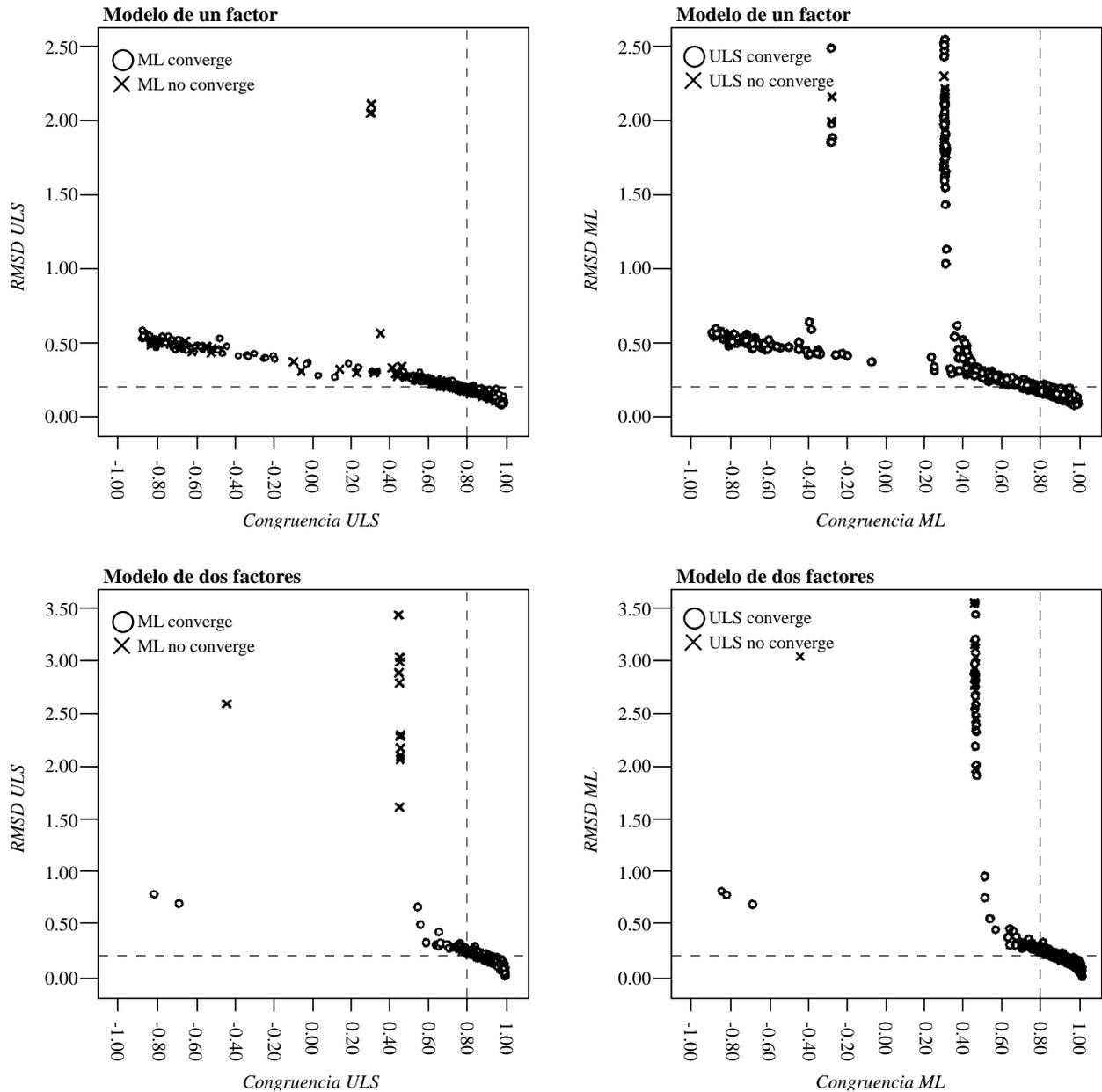


Figura 2. Gráficos de dispersión para el coeficiente  $\phi$  y el índice  $\delta$  según la convergencia de los métodos ML y ULS en los modelos de uno y dos factores con  $n=100$

## Referencias

- Anderson, J.C. y Gerbing, D.W. (1984). The effect of sampling error on convergence, improper solutions and goodness of fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49, 155-173.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Briggs, N.E. y MacCallum, R.C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, 38, 25-56.
- Browne, M.W. (1984). Asymptotically distribution-free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.
- Burt, C. (1949). Factor analysis and canonical correlations. *British Journal of Statistical Psychology*, 1, 95-106.
- Hernández, A. y González-Romá, V. (2003). Evaluating the multiple-group mean and covariance structure analysis model for the detection of differential item functioning in polytomous ordered items. *Psicothema*, 15, 322-327.
- Jöreskog, K.G. y Sörbom, D. (1996a). *PRELIS 2: User's reference guide*. Scientific Software International. 3rd edition. Chicago, IL.
- Jöreskog, K.G. y Sörbom, D. (1996b). *LISREL 8: User's reference guide*. Scientific Software International. 2nd edition. Lincolnwood, IL.
- Levine, M.S. (1977). *Canonical correlation analysis and factor comparison techniques*. Sage, Beverly Hills.
- MacCallum, R.C. y Tucker, L.R. (1991). Representing sources of error in the common factor model: implications for theory and practice. *Psychological Bulletin*, 109, 501-511.
- MacCallum, R.C., Widaman, K.F., Zhang, S. y Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.
- Olea, J., Abad, F.J., Ponsoda, V. y Ximénez, M.C. (2004). Un test adaptativo informatizado para evaluar el conocimiento del inglés escrito: diseño y propiedades psicométricas. *Psicothema*, 16, 519-525.
- Olsson, U.H., Troye, S.V. y Howell, R.D. (1999). Theoretical fit and empirical fit: the performance of maximum likelihood versus generalized least squares estimation in structural equation models. *Multivariate Behavioral Research*, 34, 31-58.
- Revuelta, J. y Ponsoda, V. (2003). *Simulación de modelos estadísticos en las 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.
- 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.