UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Modelo de Lealtad a partir de un análisis de Ecuaciones Estructurales

Robert Romero

Trabajo de grado presentado como requisito para optar al título de:
Profesional en Estadística

Director:
M.Sc Giovanny Babativa Márquez

Línea de Investigación:
Modelos de Ecuaciones Estruturales
Universidad Santo Tomas
Facultad de Estadística, Departamento de Ciencias Económicas
Bogotá, Colombia
2015

Índice general

Índice de tablas

Índice de figuras

Capítulo 1

Introducción

Para una compañía siempre ha sido importante la relación con sus clientes, por ello en diferentes trabajos se han planteado patrones para describir dicha asociación, en esta búsqueda ha cobrado relevancia el uso de modelos estadísticos que permiten establecer la forma en que interactúan las distintas variables que determinan el comportamiento de un cliente.

Adicionalmente, también se ha analizado la causalidad y se ha encontrado un ciclo en el comportamiento de compra, así, cuando un cliente es más leal a una marca, mayor es su grado de recomendación y recompra hacia ésta. Así mismo, se ha buscado establecer la relación entre satisfacción y lealtad ya que no necesariamente un alto grado de satisfacción causa lealtad ni tampoco un alto grado de lealtad causa satisfacción.

Entre los modelos utilizados para describir la relación entre las variables componentes de la satisfacción y la lealtad se encuentran los modelos de regresión, análisis factoriales y últimamente los Modelos de Ecuaciones Estructurales (SEM). Estos últimos, tienden a ser los más adecuados para realizar la estimación debido a que algunas de estas variables son exógenas y endógenas al mismo tiempo y no es fácil establecer una relación directa entre ellas, en este caso el modelo de ecuaciones estructurales permite establecer los efectos directos e indirectos entre las variables, dichos valores son utilizados para construir un indicador que permita establecer estrategias para elevar el grado de lealtad de sus clientes.

En este trabajo se explora la utilización de los Modelos de Ecuaciones Estructurales SEM para construir un modelo de lealtad a partir los componentes del servicio ofrecido al cliente. En una primera parte se describe la teoría general de los modelos SEM, en la segunda parte se realiza el análisis descriptivo de los datos y se examinan las posibles relaciones entre las variables utilizadas, por último se presenta un ejemplo práctico de la estimación con modelos SEM y sus respectivos resultados y por último las conclusiones y futuras líneas de investigación.

Capítulo 2

Objetivos

Objetivo General

Usando la técnica de Ecuaciones Estructurales construir un indicador de lealtad a partir de las dimensiones componentes del servicio

Objetivos Específicos

- Elegir el método de estimación de parámetros más adecuado.
- Medir el efecto (directo, indirecto y total) de las dimensiones componentes del servicio sobre la satisfacción del cliente.
- Determinar la matriz de acción de lealtad a partir del impacto de los componentes del servicio.
- Determinar la expresión matemática para el cálculo del indicador.

Capítulo 3

Marco Teórico

3.1. Conceptos Fundamentales

En la actualidad el análisis de Modelos de Ecuaciones Estructurales (SEM) es una importante técnica de análisis multivariado que se aplica en diferentes campos profesionales, no obstante, las técnicas estadísticas que la soportan no son de reciente aparición y tuvo que pasar gran tiempo antes que se pudiera explotar su potencial, esto gracias a los avances computacionales, que como en muchos otros campos han sido vía arteria del desarrollo de conocimiento, y que permitieron una investigación más profunda de la técnica, la cual se ha apalancado principalmente en estudios de simulación.

A grandes rasgos, los modelos SEM permiten la representación de una serie de hipótesis de relaciones (principalmente lineales) entre una serie de variables medidas y causadas a su vez por una diversidad de fenómenos subyacentes los cuales no son directamente observables, que en este caso se llaman *latentes* (en otros campos de investigación se conocen como factores o constructos). Las variables latentes son de gran importancia en muchas disciplinas pero carecen de un modo preciso de medición en lo relativo a su existencia o influencia en otros fenómenos o variables.

Posibles ejemplos de variables latentes pueden ser la calidad del aire, la felicidad o la inteligencia, fenómenos los cuales dada su inobservabilidad no se pueden medir directamente, razón por la cual, los investigadores definen una serie de herramientas operacionales a través de las cuales se puedan construir de manera indirecta. A estas herramientas se le llama variables manifiestas (observadas) y en la metodología los Modelos SEM sirven como indicadoras del fenómeno subyacente que representan.

Adicionalmente es importante tener en cuenta la diferencia entre variables *exógenas* y variable *endógenas*; las variables exógenas son variables independientes que "*causan* a otras variables en el modelo y cuyos cambios no pueden se explicados por este aunque son afectadas por factores como la edad, el sexo, la religión, filiación política, etc., por su parte las variables endógenas son dependientes en el modelo y por lo tanto son influenciadas

por las variables exógenas ya sea directa o indirectamente. Los cambios en las variables endógenas son explicados por el modelo ya que todas las variables influenciadoras están contempladas en el mismo.

El término SEM es una generalización para varios tipos de modelos y estadísticamente representan una extensión de procedimientos de Modelos Lineales Generales (**MLG**), tales como **ANOVA**, Análisis de Regresión Múltiple y Análisis Factorial, que tienen las siguientes características diferenciadoras de las técnicas de modelización clásica:

- Generalmente se conciben como construcciones teóricas de fenómenos que no son medibles directamente.
- Toman en cuenta los posibles errores de medición de las variables con las cuales se construyen los factores latentes. Las varianzas de los términos de error son los parámetros a estimar al momento de ajustar el modelo, por lo que en este caso es correcto llamarlo Análisis de Estructura de Covarianzas.
- Los modelos son ajustados a partir de matrices de índices de interrelación (matriz de correlaciones o covarianzas), aunque algunas veces también se realizará el análisis sobre las medias de las variables.

Otro punto en común con los modelos clásicos, es el uso de la prueba **F** para la comparación de modelos más restringidos contra modelos saturados.

Entre los tipos de modelos SEM se puede mencionar:

- *Path Analysis:* Este tipo de análisis busca mediante el apoyo de un *path diagram* descomponer la covarianza entre las variables del modelo con el fin de establecer la medida de relación entre los efectos causales y las medidas de covariación. Estas relaciones pueden ser directas, espúreas, indirectas o conjuntas. Adicionalmente permite medir el efecto causal (directo, indirecto y total) que una variable tiene sobre otra en el modelo. Según Long (1983) también son llamados *Modelos de Estructura de Covarianzas* y se descomponen en dos: el modelo de *componente estructural* y el modelo de *componente de medida*.
- *Análisis Factorial Confirmatorio:* Permite analizar los patrones de relación o causación entre variables latentes (constructos) del modelo estructural, con el fin de verificar si éste es válido o si sus interrelaciones generan alguna interpretación plausible.
- *Modelos de Curva de Crecimiento:* Los anteriores modelos citados se basan en datos de corte transversal obtenidos a través de una muestra de individuos en un punto de tiempo t. Los modelos de curva de crecimiento permiten analizar para datos de corte longitudinal, la dinámica de los cambios y evoluciones del comportamiento de los procesos bajo estudio.

Los modelos SEM requieren una serie de hipótesis de relaciones estructurales a-priori las cuales confirmar , por lo que su objetivo primordial es determinar si para un modelo teórico la hipótesis es consistente con los datos, dicha consistencia se refleja en las medidas de ajuste del modelo, las cuales indican que tan bien se reproduce la matriz de covarianzas.

La expresión matemática de estos modelos es generalmente más compleja que la utilizada para otros análisis multivariantes, por ello no fue hasta que apareció el programa *LISREL* (Linear Structural Relations, Jöreskog.1973) que se extendió su uso. Otros programas usados para la estimación de modelos SEM son *EQS* (Equations, Bentler, 1985), *AMOS* (Analysis of Moments Structures,Arbuckle,1997), *MPLUS* (Muthén,1998). En este trabajo utilizo el programa de código abierto R y específicamente el paquete *Lavaan* (Latent Variable Analysis,Yves Rosseel, 2010), el cual incorpora soporte para el análisis factorial confirmatorio, modelos estructurales y de curvas de crecimiento, adicionalmente permite proponer el tipo y dirección de las relaciones que se espera encontrar entre las variables del modelo (al igual que los programas comerciales) , para posteriormente entrar a realizar la estimación de los parámetros.

3.2. Path Diagram (Diagrama de Ruta)

A veces los sistemas de ecuaciones son muy complejos y requieren introducir muchas relaciones entre las variables, en este caso generalmente se prefiere la representación gráfica del modelo en consideración mediante un diagrama causal o *Path Diagram* como el de la figura [3.1], este tipo de representación equivale a un conjunto de ecuaciones que configuran el modelo. En esta representación gráfica se usa una notación especial como se muestra en la tabla [3.1]:

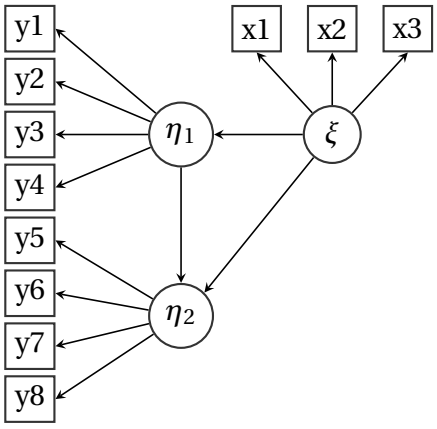


Figura 3.1: Modelo de Ecuación Estructural representado por un Path Diagram

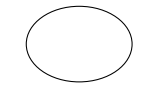
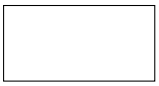
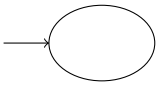
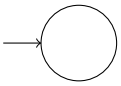
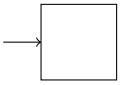
	o		Variable Latente
	o		Variable Observada
	o		Relación entre dos variables
	o		Covarianza entre dos variables
	o		Medida de Error de la variable Latente
	o		Medida de Error de la variable Observada

Tabla 3.1: Símbolos usados comúnmente en análisis SEM

Cuando el modelo incluye tanto variables observadas como latentes, las primeras se representan con cuadrados o rectángulos y las segundas con círculos o elipses, las flechas que salen de las variables latentes a las observadas se llaman *relaciones de medición*. Las variables observadas están afectadas por un término de error aleatorio, el cual es representado en el diagrama por una flecha unidireccional que apunta a la variable observable. La covariación entre dos errores de medición es representada por una flecha bidireccional que los une. La relación entre variables se índica por una flecha unidireccional desde la variable *exógena* a la variable *endógena*.

Según Batista & Coenders (2000) [5] para que el *Path Diagram* represente las teorías causales y de medición de forma equivalente a la que lo hacen los sistemas de ecuaciones, se debe cumplir:

- Todas las relaciones causales deben estar representadas en el diagrama.
- Todas las variables que son causas de las variables *endógenas* deben estar incluidas en el diagrama.
- El diagrama debe ser lo más parsimonioso posible y debe contener aquellas relaciones que pueden justificarse teóricamente.

3.3. Path Analysis

Ya en la sección [3.1] se determinó el principal objetivo del *Path Analysis*, para desarrollar por completo el concepto se detallarán los tipos de relación que puede llevar a que dos variables x y y covarien.

1. x y y pueden covariar si y tiene algún efecto sobre x (o al contrario), como la relación representada por un modelo de regresión simple por ejemplo:

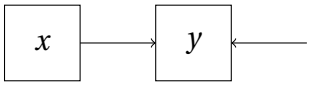


Figura 3.2: Path Diagram Regresión Simple

Estas relaciones se llaman *directas*, aunque también pueden ser *recíprocas*:

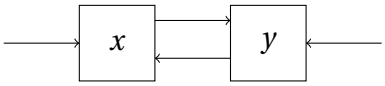


Figura 3.3: Path Diagram relación *recíproca*

2. x y y covarian si tienen una causa común z, este tipo de relación se denomina *espúrea*:

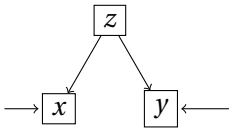


Figura 3.4: Path Diagram relación *espúrea*

3. x y y también covarian si están relacionadas a través de una tercera variable z, este tipo de relación se denomina *indirecta*:

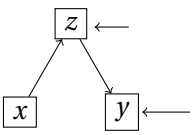


Figura 3.5: Path Diagram relación *indirecta*

Una vez señalados los posibles tipos de covariación entre las variables del modelo, se procede ahora a implantar una serie de reglas de descomposición que permitan establecer las relaciones entre las covariaciones y los parámetros del mismo, una vez fijadas estas relaciones se podrá realizar a partir de ellas el cálculo de estimación de parámetros. Es importante no perder de vista que el ajuste del modelo se realiza sobre las matrices de covarianzas de las variables observadas, las cuales son previamente centradas para dejar de un lado el efecto que pueda tener la media de cada variable observada.

Las varianzas y covarianzas de las variables observadas son en si mismas medidas iniciales del modelo. Según Batista & Coenders (2000) [5] para derivar los demás parámetros se sigue que:

1. La covarianza entre dos variables se calcula como la suma entre los efectos directos, indirectos, espúreos y conjuntos. Cada uno de estos representa en el *Path Diagram* una posible manera de unir las variables. El efecto se calcula como el producto de la varianza de la variable de partida por todos los parámetros asociados a las flechas recorridas hasta llegar a unir las dos variables de interés.

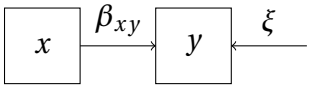
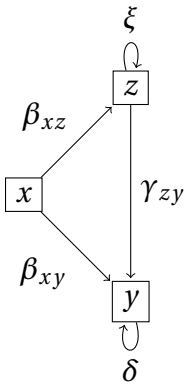


Figura 3.6: Efecto Total de x sobre y



Efecto Directo:	β_{xy}
Efecto Indirecto:	$\beta_{xz} * \gamma_{zy}$
Efecto Total:	$\mathbf{B}_{xy} = \beta_{xy} + \beta_{xz} * \gamma_{zy}$

Figura 3.7: Descomposición del efecto de x sobre y

2. la varianza de una variable exógena se calcula como la varianza del término de error más la varianza explicada por otras variables en el modelo.

Así mismo, la varianza explicada puede expresarse como una función de todas las exógenas con efecto directo sobre la variable endógena, como la suba de la totalidad de productos entre los efectos directos y las covarianzas entre las variables endógena y la exógenas relacionadas por los efectos.

Aplicando las anteriores reglas, se obtiene un sistema de *ecuaciones estructurales* el cual expresa la matriz de covarianzas en función de los parámetros del modelo,

$$\Sigma = \Sigma(\theta)$$

donde θ es un vector que contiene los parámetros del modelo (efectos directos, indirectos, varianzas, covarianzas de errores, de perturbaciones y de variables exógenas).

Ejemplo

Supongamos un modelo de regresión simple sin intercepto en el cual las variables se encuentran centradas respecto de su media, su expresión estaría dada por:

$$y = \beta X + e \tag{3.1}$$

y que se puede representar con el diagrama de la figura [3.8]:

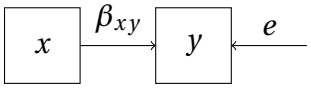


Figura 3.8: Ejemplo: Regresión Simple

Como sabemos este modelo tiene una serie de supuestos que se deben cumplir para su aplicación, estos son:

$$x \sim N(0, \delta)$$
$$e \sim N(0, \xi)$$
$$cov(x, e) = 0$$

Que se puede expresar de manera más general como:

$$\begin{pmatrix} x \\ e \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \delta_{11} & 0 \\ 0 & \xi_{22} \end{pmatrix}\right) \tag{3.2}$$

Considerando [3.1] y [3.2] se tiene una distribución conjunta de 3 parámetros que derivan a

$$\Sigma = \Sigma(\theta)$$

donde Σ representa la matriz de covarianzas poblacionales de las variables manifiestas en el modelo:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

y el vector de parámetros está dado por:

$$\theta = (\delta_{11}, \xi_{22}, \beta_{xy})$$

Generalizando, en un modelo con p variables observables se tiene un número de $p(p+1)/2$ ecuaciones estructurales en el modelo (tantos como elementos distintos en Σ). Para derivar las ecuaciones se hace uso de las reglas de descomposición citadas en la sección [3.1], de manera que la varianza de x es un parámetro directo del modelo, la covarianza entre x y y se calcula multiplicando el coeficiente β que relaciona a x con y por la varianza de x, en tanto que la varianza de y se calcula como la varianza de la variable de error más el producto entre la $cov(x, y)$ por el coeficiente β del modelo:

$$\begin{cases} \sigma_{11} = \delta_{11} \\ \sigma_{21} = \delta_{11}\beta_{xy} \\ \sigma_{22} = \xi_{22} + \sigma 21\beta_{xy} = \xi_{22} + \delta_{11}\beta_{xy}^2 \end{cases} \tag{3.3}$$

Este sistema tiene igual número de ecuaciones que de parámetros a estimar, por lo que se le conoce como *exactamente identificado*, despejando los parámetros se tiene:

$$\begin{cases} \delta_{11} = \sigma_{11} \\ \beta_{xy} = \sigma_{21}/\sigma_{11} \\ \xi_{22} = \sigma_{22} - \sigma_{21}^2/\sigma_{11} \end{cases} \tag{3.4}$$

Ya que no se conoce Σ se utiliza su estimación, entonces el vector de parámetros estimados $\hat{\theta} = (\hat{\delta_{11}}, \hat{\xi_{22}}, \hat{\beta_{xy}})$.

$$\begin{cases} \hat{\delta_{11}} = s_{11} \\ \hat{\beta_{xy}} = s_{21}/s_{11} \\ \hat{\xi_{22}} = s_{22} - s_{21}^2/s_{11} \end{cases} \tag{3.5}$$

Si se tiene;

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} = \begin{pmatrix} 2,6 & 1,8 \\ 1,8 & 3,4 \end{pmatrix}$$

La estimación de los parámetros sería;

$$\begin{cases} \hat{\delta_{11}} = 2,6 \\ \hat{\beta_{xy}} = 1,8/2,6 = 0,69 \\ \hat{\xi_{22}} = 3,4 - 1,8^2/2,6 = 2,15 \end{cases}$$

3.4. Etapas de Construcción de los modelos SEM

Como se ha visto, el análisis de modelos SEM requiere de un conocimiento previo de las posibles interacciones o relaciones entre las variables del modelo tanto exógenas como endógenas, esto quiere decir que en el *Path Diagram* se grafican las hipotéticas relaciones de los fenómenos bajo estudio, a partir de este punto existe una serie de pasos que debe seguir el investigador antes de aplicar el modelo a la vida real.

Las etapas básicas se enumeran a continuación y luego se presentan en un diagrama de flujo en la figura [3.9], estos pasos son iterativos, ya que dependiendo de lo que pase en alguno de ellos se debe retornar al primero:

1. Especificación
2. Identificación
3. Recogida de Datos
4. Estimación
5. Diagnóstico
6. Evaluar si el modelo es adecuado (en caso de no serlo volver al paso 1)
7. Uso

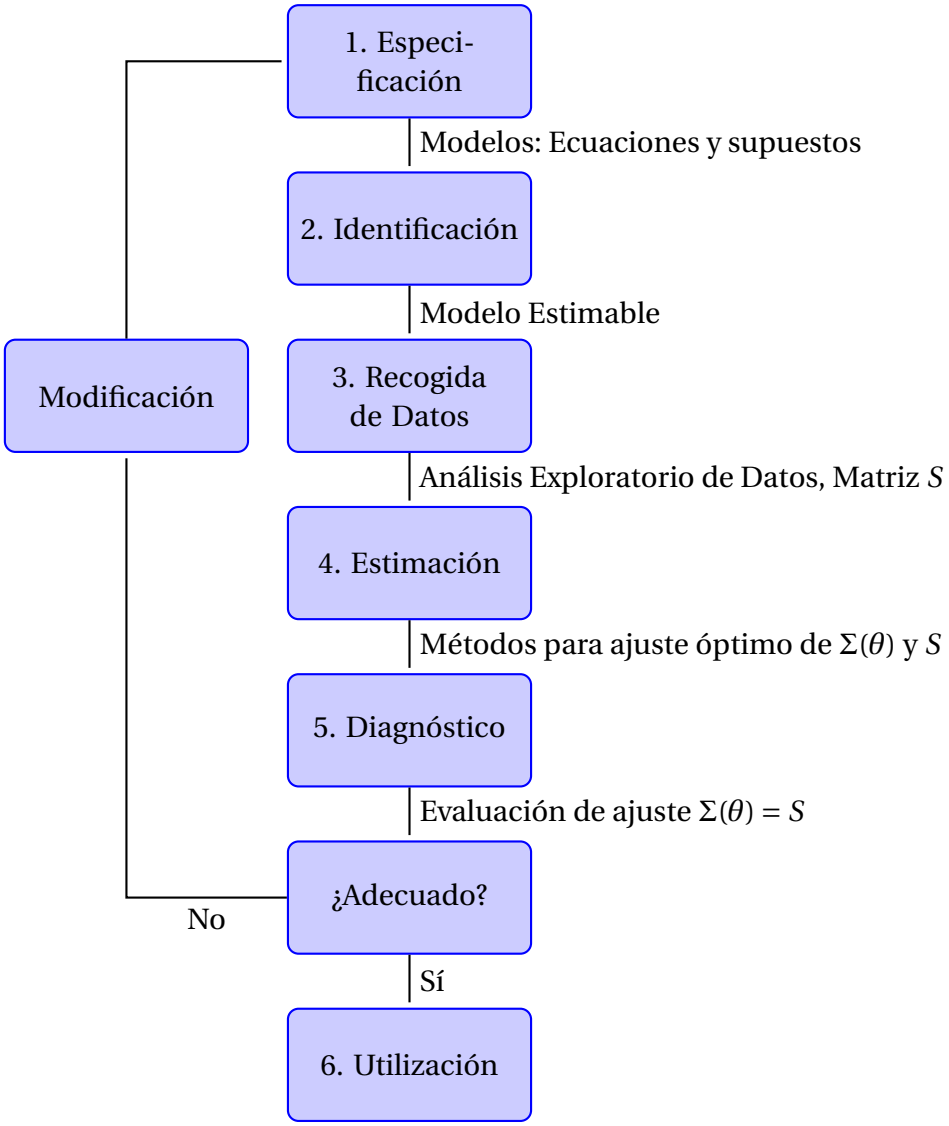


Figura 3.9: Etapas del modelado SEM - Fuente: Batista & Coenders [5]

3.4.1. Etapa de Especificación

Es la representación formal de las hipótesis del modelo SEM, dichas hipótesis tienen que ver con la relación o efecto que puedan tener los factores latentes subyacentes sobre las variables observadas (Modelo de Medida), así como el efecto que tienen algunos factores latentes sobre otros de la misma clase (Modelo Estructural). Esta representación se basa principalmente en el conocimiento teórico que sobre el fenómeno estudiado tenga el investigador, el cual si es detallado, facilitará la identificación del modelo y su posterior estimación (Análisis Confirmatorio), sin embargo, en caso de carecer de información fiable, debe primero explorar las posibles relaciones entre variables con el fin de establecer

un modelo inicial para posteriormente «afinarlo» en la etapa de estimación (Análisis Exploratorio), esta segunda, es la estrategia más usada.

La etapa de especificación es flexible, razón por la cual una gran cantidad de modelos pueden ser generados, sin embargo no todos los modelos pueden ser estimados ya que dependen del número de parámetros desconocidos a identificar. Los parámetros son de tres tipos:

- *Libres*: Desconocidos y no restringidos, son los parámetros a estimar.
- *Restringidos*: Corresponden a dos o más parámetros que a pesar de ser desconocidos, deben tomar el mismo valor al estimarse.
- *Fijos*: Son parámetros conocidos, a los que se les asigna un valor previo a la estimación, que para fijar la escala de la variable latente suele ser uno (1), o en el caso de fijar las covarianzas cero (0).

Adicionalmente, los modelos SEM según su complejidad pueden expresar relaciones entre factores (Modelos Estructurales) o relaciones entre factores y variables manifiestas (Modelos de Medida), independientemente que se usen juntos o no, siguen una serie de supuestos que se deben cumplir necesariamente para la correcta especificación del modelo.

3.4.1.1. Modelo Estructural

El componente estructural en los modelos SEM explica la relación causal entre las variables latentes, la figura [3.7] representa el efecto de x sobre y a través de una variable mediadora z, y como antes se mencionó si los efectos entre $x - z$ y entre $z - y$ son diferentes de cero, se dice que se tiene un efecto indirecto de x sobre y a través de z, igualmente si el efecto entre x y y que no pasa por z es diferente de cero, se tiene entonces un efecto directo de x sobre y. Esta estructura puede ser representada de manera más general como:

$$\eta = B\eta + \Gamma\xi + \zeta \tag{3.6}$$

Donde η es un vector de variables latentes endógenas de dimensión $px1$, B es una matriz de coeficientes que relacionan a estas variables latentes y que tiene dimensión pxp y ξ es un vector que contiene las variable latentes exógenas del modelo y tiene dimensión $nx1$, Γ es una matriz de pxn que contiene los coeficientes que relacionan a ξ con η y finalmente ζ es un vector de $px1$ que contiene los el término de error de la ecuación (o disturbancia). Los supuestos básicos del modelo estructural son:

$$
\begin{aligned}
\mathrm{E}(\eta) &= 0 \\
\mathrm{E}(\xi) &= 0 \\
\mathrm{E}(\zeta) &= 0 \\
\mathrm{E}(\xi\zeta^{'}) = E(\zeta\xi^{'}) &= 0 \\
(I-B) &: \text{Tiene inversa}
\end{aligned}
$$

3.4.1.2. Modelo de Medida

En algunas ocasiones, x y y pueden ser medidas *sin error*, por lo que en este caso formulamos de manera simple las ecuaciones para un Modelo de Medida

$$x = \Lambda_x \xi + \delta_k \qquad\qquad y = \Lambda_y \eta + \epsilon_k \qquad (3.7)$$

y es un vector de resultados con dimensión $px1$, x vector de variables independientes de tamaño $rx1$, las matrices que contienen las cargas factoriales o coeficientes de regresión son Λ_x para los factores independientes y Λ_y para los dependientes y sus dimensiones respectivamente son rxs y pxt donde s y t corresponden al número de factores de cada tipo (Nótese que y puede ser un factor latente endógeno o una variable observada endógena).

Los supuestos para el Modelo de Medida son:

$$E(x) = E(\delta) = E(y) = E(\epsilon) = 0$$
$$E(\xi\delta^{'}) = E(\delta\xi^{'}) = E(\eta\epsilon^{'}) = E(\epsilon\eta^{'}) = 0$$
$$E(\delta\epsilon^{'}) = E(\epsilon\delta^{'}) = 0$$

3.4.1.3. Supuestos sobre la distribución de los datos

Como se vio en las secciones [3.4.1.1] y [3.4.1.2], se asume que la distribución de los factores exógenos, perturbaciones y errores de medida es multivariante normal en cada caso y como se sabe, la alteración en los supuestos de distribución conduce a sesgos en la estimación y afecta el contraste de las hipótesis especificadas, aunque en este caso se encuentran estrategias de estimación para la no normalidad que se expondrán más adelante.

Adicionalmente se asume que el modelo se ajuste para todas las observaciones, por lo que al momento de hacer el levantamiento de la información se debe asegurar que cada una de ellas sean independientes e idénticamente distribuidas.

3.4.2. Matriz de Covariazas Implícita

La generalización de un modelo en el cual aparecen factores latentes relacionados con variables observadas es:

$$y = By + \Gamma x + \zeta \tag{3.8}$$

donde

B = pxp matriz de coeficientes
Γ = pxn matriz de coeficientes
y= $nx1$ vector de variable respuesta
x= $nx1$ vector de variable independientes
ζ = $nx1$ vector de error en la ecuación

Como se vió en [3.3], la principal hipótesis del sistema de los modelos SEM es

$$\Sigma = \Sigma(\theta)$$

La relación entre Σ y $\Sigma(\theta)$ es línea base para entender la identificación, la estimación y el ajuste de los modelos SEM.

Bollen (1989)[7] estableció que $\Sigma(\theta)$ es la reunión de 3 componentes:

1. $\Sigma_{yy}(\theta)$
2. $\Sigma_{xy}(\theta)$
3. $\Sigma_{xx}(\theta)$

Considerando primero $\Sigma_{yy}(\theta)$, la matriz implícita de covarianzas para y se calcula como:

$$\begin{aligned}
\Sigma_{yy}(\theta) = &\quad E(yy') \\
= &\quad \text{E}[(\text{I - B})^{-1}(\Gamma x + \zeta)((I-B)^{-1}(\Gamma x + \zeta))'] \\
= &\quad \text{E}[(\text{I - B})^{-1}(\Gamma x + \zeta)(x'\Gamma' + \zeta')((I-B)^{-1})'] \\
= &\quad (\text{I - B})^{-1}(E(\Gamma x x'\Gamma^{-1}) + E(\Gamma x\zeta') + E(\zeta\zeta'))(I-B)^{-1'}
\end{aligned}$$

Por lo que

$$\Sigma_{yy}(\theta) = (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I-B)^{-1'} \tag{3.9}$$

donde

Φ = matriz de covarianzas de x
Ψ = matriz de covarianzas de ζ

Entonces la matriz implícita de covarianzas de x:

$$\Sigma_{xx}(\theta) = E(xx')$$

$$\Sigma_{xx}(\theta) = \Phi \tag{3.10}$$

y finalmente:

$$\begin{aligned} \Sigma_{xy}(\theta) = &\quad E(xy') \\ = &\quad \mathrm{E[x((I - B)^{-1}(\Gamma x + \zeta))']} \end{aligned}$$

$$\Sigma_{xy}(\theta) = \Phi\Gamma'(I-B)^{-1'} \tag{3.11}$$

Ya que:

$$\Sigma(\theta) = \begin{bmatrix} \Sigma_{yy}(\theta) & \Sigma_{xy}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{xx}(\theta) \end{bmatrix}$$

Se tiene que la expresión final de la matriz de covarianzas implícita es:

$$\Sigma(\theta) = \begin{bmatrix} (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I-B)^{-1'} & \Phi\Gamma'(I-B)^{-1'} \\ \Phi\Gamma'(I-B)^{-1'} & \Phi \end{bmatrix} \tag{3.12}$$

3.4.3. Etapa de Identificación

La identificación es un tema que a menudo se presenta confuso y su preocupación radica en el hecho de si se pueden o no estimar los parámetros del modelo SEM, cuándo el modelo se puede estimar es porque el modelo está identificado, en caso contrario se dirá que el modelo no es identificable, sin embargo cuando se trata de ecuaciones múltiples, no se puede decir fácilmente si todos los parámetros se identifican o no. Por ejemplo considere las siguientes ecuaciones:

$$x + y = 10$$

$$3x + y = 7$$

$$x + 5y = 12$$

Con la primera ecuación $x + y = 10$ existe un infinito número de soluciones ya que x puede tomar cualquier valor por lo que se dice que no es identificable (*nunca identificado*). Con las dos primeras ecuaciones el sistema tendría una única solución para x y y por

lo que el sistema es *identificado*, pero si optamos por escoger las 3 ecuaciones para resolver el sistema tendremos una *sobre identificación* ya que podremos encontrar múltiples valores para x y y que pueden satisfacer el sistema, en este caso habría que imponer restricciones sobre algunos parámetros. Los usuarios de modelos SEM prefieren trabajar con los modelos *sobre identificados* ya que aquellos justamente identificados son triviales y no permiten contrastar las teorías sobre las relaciones causales tratadas en el modelo.

El mismo principio se puede aplicar a [3.12], los parámetros a identificar que necesitan ser identificados están en θ cuyo vector contiene t parámetros libres de **B**, Γ, Φ y Ψ. Si uno de los parámetros desconocidos de θ puede ser escrito como una función de uno o más elementos de Σ, este parámetro es identificado y si todos los parámetros de θ son identificados, el modelo es identificable. Bollen expone que la sobre identificación para Σ sucede cuando se tiene $\phi_{11} = VAR(x_1) = COV(x_1, y_1)$ y ambos, $COV(x_1, y_1)$ y $VAR(x_1)$ son identificados.

3.4.3.1. *t-Rule*

Consiste en calcular el número de elementos no redundantes en la matriz de covarianzas de las variables observadas el cual debe ser mayor que el número de parámetros desconocidos en θ:

$$t \leq \left(\frac{1}{2}\right)\Big(p+q)(p+q+1) \tag{3.13}$$

Donde $p+q$ son el número de variables observadas y t es el número de parámetros libres y distintos en θ, entonces, el lado derecho de la desigualdad representa el número de elementos no redundantes en Σ.

Esta regla permite conocer cuándo un modelo no es identificado, aunque no permite discernir si es sobre identificado o no, así que adicionalmente al cálculo de *t-Rule* se debe tener en cuenta que el valor de los grados de libertad (g) puede ayudar a discernir el tipo de modelo de la siguiente manera:

- *g<0*: Modelos nunca identificados, los parámetros podrían tomar infinitos valores, razón por la cual son indeterminados.
- *g=0*: Posiblemente identificados, modelos en los que puede existir una única solución para los parámetros, que ajuste la matriz Σ.
- *g>0*: Posiblemente sobre identificados, Modelos que incluyen menos parámetros que varianzas y covarianzas tiene Σ.

3.4.4. Etapa de Estimación

Esta etapa se basa en las relaciones entre las varianzas y covarianzas de los parámetros y las variables originales, por lo que los diferentes métodos de estimación buscan que la discrepancia sea mínima

$$S - \hat{\Sigma} \approx 0$$

Se habia visto anteriormente que

$$\Sigma(\theta) = \begin{bmatrix} (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I-B)^{-1'} & \Phi\Gamma'(I-B)^{-1'} \\ \Phi\Gamma'(I-B)^{-1'} & \Phi \end{bmatrix}$$

Si los parámetros de la población fueran conocidos y el modelos estructural estuviera bien identificado se tendría que $\Sigma = \Sigma(\theta)$. Por ejemplo considérese la ecuación:

$$y_1 = \gamma_{11} x_1 + \zeta \tag{3.14}$$

En esta ecuación puede observarse que $B = 0$ y se tendría la siguiente matriz de covarianzas muestral para x_1 y y_1:

$$\mathbf{S} = \begin{bmatrix} var(y_1) & cov(y_1, x_1) \\ cov(y_1, x_1) & var(y_1) \end{bmatrix}$$

En [3.12] se estiman los valores de la matriz de covarianzas implícita que según la ecuación [3.16] tomaría la siguiente forma:

$$\Sigma(\hat{\theta}) = \begin{bmatrix} \hat{\phi_{11}}\hat{\gamma_{11}}\hat{\gamma_{11}} + \hat{\psi_{11}} & \hat{\phi_{11}}\hat{\gamma_{11}} \\ \hat{\phi_{11}}\hat{\gamma_{11}} & \hat{\phi_{11}} \end{bmatrix}$$

Para ilustrar este procedimiento, supongamos que

$$\mathbf{S} = \begin{bmatrix} 2,6 & 2,3 \\ 2,3 & 3,4 \end{bmatrix}$$

Ahora digamos que $\hat{\gamma_{11}} = 1$, $\hat{\phi_{11}} = 2,8$ y $\hat{\psi_{11}} = -0,4$

$$\Sigma(\hat{\theta}) = \begin{bmatrix} 2,4 & 2,8 \\ 2,8 & 2,8 \end{bmatrix}$$

Y la matriz residual $S - \Sigma(\hat{\theta})$ indicaría que tanto ajusta el modelo

$$S - \hat{\Sigma} = \begin{bmatrix} 0,2 & -0,5 \\ -0,5 & 0,6 \end{bmatrix}$$

Las funciones de discrepancia más utilizadas para realizar la estimación de los parámetros son las siguientes:

3.4.4.1. Máxima Verosimilitud

ML *(Maximum Likelihood)* es la más extendida función de estimación de parámetros y busca minimizar la siguiente función de ajuste:

$$F_{ML} = log \mid \Sigma(\theta) \mid + tr(S\Sigma^{-1}(\theta)) - log \mid S \mid -(p+q) \tag{3.15}$$

donde $\Sigma(\theta)$ y **S** son matrices simétricas y semidefinidas positivas de manera que todos sus menores principales son positivos, por lo tanto son matrices invertibles. Adicionalmente x y y son variables $iid.$ tales que $y \sim N(\mu_y, \sigma_y)$ y $x \sim N(\mu_x, \sigma_x)$ (Los programas computacionales de estimación centran las variables previamente de manera que $\mu = 0$).

Substituyendo $\hat{\Sigma}$ por $\Sigma(\theta)$ y teniendo en cuenta que $\hat{\Sigma} = S$ se tiene

$$F_{ML} = log \mid S \mid + tr(I) - log \mid S \mid -(p+q)$$

$$F_{ML} = tr(I) - (p+q)$$

Dado que $SS^{-1} = I$, a la vez, $tr(I) = (p+q)$ por lo que

$$F_{ML} = 0$$

Ejemplo (Tomado de Bollen (1989)[7])

Para comprobar la forma en que opera esta función de ajuste, vamos a tomar la ecuación

$$y_1 = x_1 + \zeta \tag{3.16}$$

teniendo en cuenta que

$$\hat{\Sigma} = \begin{bmatrix} \hat{\phi_{11}} + \hat{\psi_{11}} & \hat{\phi_{11}} \\ \hat{\phi_{11}} & \hat{\phi_{11}} \end{bmatrix}$$

$$\mid \hat{\Sigma} \mid = \quad \hat{\phi_{11}}(\hat{\phi_{11}} + \hat{\psi_{11}}) - \hat{\phi_{11}}^2$$

$$\mid \hat{\Sigma} \mid = \quad \hat{\phi_{11}}\hat{\psi_{11}}$$

se sustituye $\hat{\Sigma}$ por $\Sigma(\theta)$ en la función de ajuste del método **ML** obteniendo

$$F_{ML} = log(\hat{\phi_{11}}\hat{\psi_{11}}) + \hat{\psi_{11}}^{-1}(var(y_1) - 2cov(y_1, x_1) + var(x_1)) + \hat{\phi_{11}} var(x_1)$$
$$- log[var(y_1)var(x_1) - (cov(y_1, x_1))^2] - (1+1) \tag{3.17}$$

Dado que nos encontramos ante un problema de minimización de F_{ML} se calculan sus derivadas parciales con respecto a $\hat{\phi_{11}}$ y $\hat{\psi_{11}}$ y se igualan a cero

$$\frac{\partial F_{ML}}{\partial \hat{\phi_{11}}} = \hat{\phi_{11}}^{-1} - \hat{\phi_{11}}^{-2} var(x_1) \tag{3.18}$$

$$\frac{\partial F_{ML}}{\partial \hat{\psi_{11}}} = \hat{\psi_{11}}^{-1} - \hat{\psi_{11}}^{-2} (var(y_1) - 2cov(y_1, x_1) + var(x_1)) \tag{3.19}$$

Igualando a cero y resolviendo se llega a

$$\hat{\phi_{11}} = var(x_1) \tag{3.20}$$

$$\hat{\psi_{11}} = var(y_1) - 2cov(y_1, x_1) + var(x_1) \tag{3.21}$$

La condición suficiente para lo valores que minimizan a F_{ML} es que la matriz formada por las segundas derivadas de la función de ajuste, sea definida positiva

$$\begin{bmatrix} -\hat{\phi_{11}}^{-2} + 2\hat{\phi_{11}}^{-3} var(x_1) & 0 \\ 0 & -\hat{\psi_{11}}^{-2} + 2\hat{\psi_{11}}^{-3} (var(y_1) - 2cov(y_1, x_1) + var(x_1)) \end{bmatrix} \tag{3.22}$$

3.4.4.2. Mínimos Cuadrados Generalizados

GLS *(Generalized Least Squares)* minimiza el cuadrado de las desviaciones entre S y $\Sigma(\theta)$ de la siguiente función de ajuste

$$F_{GLS} = (\frac{1}{2}) tr[([S - \Sigma(\theta)]\mathbf{W}^{-1})^2] \tag{3.23}$$

en la que substituyendo $\hat{\Sigma}$ por $\Sigma(\theta)$ y teniendo en cuenta que $\hat{\Sigma} = S$ fácilmente se observa que

$$F_{GLS} = 0$$

En este caso $\mathbf{W}^{-1}$ corresponde a una matriz de pesos para los residuos y bien puede ser una matriz aleatoria que converge en probabilidad a una matriz definida positiva cuando $N \to \infty$ o también una matriz positiva de constantes en otros casos.

Cuando $\mathbf{W}^{-1} = I$, la función de ajuste toma el nombre de **ULS** *(Unweighted Least Squares)* y el proceso de minimización de la función es igual al realizado para **ML**.

La escogencia de $\mathbf{W}^{-1}$ se debe hacer de tal manera que satisfaga los supuestos que sobre **S** se tienen, principalmente la condición de normalidad de sus elementos.

3.4.4.3. Mínimos Cuadrados Ponderados

WLS *(Weigthed Least Squares)*, también llamado **ADF** *(Asymptotic Distribution-Free)*, es un método libre de distribución, es decir, no requiere una distribución específica de las variables observadas y produce resultados asintóticamente válidos (para muestras grandes). Minimiza la siguiente función de ajuste:

$$F_{WLS} = [s - \sigma(\theta)]' \mathbf{W}^{-1} [s - \sigma(\theta)] \tag{3.24}$$

Donde s es un vector de $(\frac{1}{2})(p+q)(p+q+1)$ elementos no duplicados vectorizados de **S**, igualmente $\sigma(\theta)$ es un vector que contiene los elementos no duplicados de $\Sigma(\theta)$ y θ es un vector de $px1$ parámetros. La matriz **W** es una matriz cuadrada definida positiva de pesos con rango $(\frac{1}{2})(p+q)(p+q+1)$ y la cual pondera las diferencias cuadráticas entre las matrices de covarianzas muestrales y estimadas.

Diferentes matrices **W** llevarán a diferentes ajustes y por lo tanto diferentes estimaciones, por lo cual es importante la correcta escogencia de ésta, la matriz **W** será a menudo una una función de n, bajo el supuesto de $lim_{n\to\infty} V_n = V$, en este caso corresponde a la matriz asintótica de covarianzas, la cual es calculada a partir de la matriz de varianza-covarianza muestral junto con los momentos de cuarto orden (ver [8]), de la siguiente manera:

$$W_{ij,kl} = s_{ijkl} - s_{ij} s_{kl} \tag{3.25}$$

en la que;

$$s_{ij} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{N-1} \tag{3.26}$$

$$s_{kl} = \frac{\sum_{i=1}^{N} (x_k - \bar{x}_k)(x_l - \bar{x}_l)}{N-1} \tag{3.27}$$

Por su parte s_{ijkl} se relaciona con la curtosis multivariada y se calcula como:

$$s_{ijkl} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}_i)(x_j - \bar{x}_j)(x_k - \bar{x}_k)(x_l - \bar{x}_l)}{N-1} \tag{3.28}$$

La muestra mínima requerida para una estimación de parámetros por el método *WLS* es de $\frac{1}{2}\left(p+q\right)\left(p+q+1\right)$.

3.4.4.4. Mínimos Cuadrados Ponderados Diagonalizados

DWLS *(Diagonally Weigthed Least Squares)* es un método robusto derivado del método *WLS* y está basado en la matriz de correlaciones policóricas de las variables incluidas en el análisis. Este método usa la diagonal de la matriz $\mathbf{W}\rho\rho$ de pesos invertida calculada para el método *WLS* en la siguiente función de ajuste:

$$F_{DWLS} = [\hat{\rho} - \rho(\theta)]^{'} diag(W_{\rho\rho})^{-1} [\hat{\rho} - \rho(\theta)] \tag{3.29}$$

A diferencia de *WLS* este método puede ser usado con pequeñas muestras y especialmente con variables observadas de tipo ordinal.

3.4.4.5. Estimación con variables ordinales

En algunas ocasiones las variables observadas son de tipo ordinal, como por ejemplo los casos en que se usa una escala de Likert para medir la satisfacción de los usuarios de un producto o servicio (1='Nada Satisfecho', 2='Poco Satisfecho', ..., 5='Completamente Satisfecho') y no pueden ser tratadas como variables continuas, en este caso la matriz de covarianzas no se puede calcular como medida de asociación ya que no estimaría bien las relaciones entre las variables, por lo que para usar variables de tipo ordinal en modelos SEM se debe recurrir a técnicas diferentes a las tradicionales pero que sean a la vez robustas.

A propósito, las llamadas *correlaciones policóricas* son apropiadas en este caso como punto de partida para las estimaciones, sin embargo no debe usarse con la estrategia de estimación **ML** ya que no obstante ser consistente en sus estimaciones, produciría errores estándar asintóticamente incorrectos (Jöreskog and Sörbom 1989) [20]. Estudios realizados por Múthen (1984) [26] y Aish & Jöreskog (1990) [1] han mostrado que la estimación por Mínimos Cuadrados Ponderados genera una mejor matriz asintótica de correlaciones estimadas.

El supuesto de las *correlaciones policóricas* es que para cada variable ordinal y hay una variable continua subyacente $y^* \sim N(\mu_y, \sigma_y^2)$, este supuesto no es comprobable por si solo, sin embargo (Jöreskog and Sörbom 1989) [20] proporcionan una prueba de normalidad bivariante para las variables subyacentes en el software PRELIS2.

La conexión entre y y y^* se da según la siguiente función escalonada:

$$y = i \text{ si } \quad \tau_{i-1} < y^* \leq \tau_i \quad \text{ para } i = 1, 2, ..., m \tag{3.30}$$

donde

$$\tau_0 = -\infty, \quad \tau_1 < \tau_2 < \cdots < \tau_{m-1}, \quad \tau_m = +\infty$$

son los llamados *umbrales*; con m categorías, hay *m-1* umbrales.

Sea $\phi(y_l^*, y_m^*, \rho)$ una función de distribución de probabilidad tal que

$$\phi(y_l^*, y_m^*, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} exp\{-\frac{y_l^* - 2_l^* y_m^* + y_l^{*2}}{2(1-\rho^2}\} \tag{3.31}$$

Dado que el interés es una correlación, sin pérdida de generalidad se puede asumir que $\mu_{y_l^*} = \mu_{y_m^*} = 0$ y $\sigma_{y_l^*} = \sigma_{y_m^*} = 1$. Sea $\Phi_2(\xi_1, \eta_1, \rho)$ la distribución bivariada normal acumulativa de y_l^* y y_m^* a ξ_1 y η_1 respectivamente, de tal manera que:

$$\Phi_2(\xi_1, \eta_1, \rho) = \int_{-\infty}^{\xi_1} \int_{-\infty}^{\eta_1} \phi(y_l^*, y_m^*, \rho) d_{y_l^*} d_{y_m^*} \tag{3.32}$$

Resolviendo esta integral se tiene,

$$\pi_{11} = \Phi_2(\xi_1, \eta_1, \rho) \tag{3.33}$$

Entonces π es una función de ρ y los umbrales ξ y η para una observación con $y_l^* = 1$ y $y_m^* = 1$, y de manera general

$$_{kl}(\theta) = \Phi_2(\xi_k, \eta_l) - \Phi_2(\xi_{k-1}, \eta_l) - \Phi_2(\xi_k, \eta_{l-1}) + \Phi_2(\xi_{k-1}, \eta_{l-1}) \tag{3.34}$$

en el que θ es un vector de multiparámetros tal que $\theta = [\rho, \theta] = [\rho, \xi_1, ..., \xi_{k-1}, \eta_1, ..., \eta_{l-1}]$, definiendo así que

$$\pi_{kl} = h_{kl}(\theta) \tag{3.35}$$

Entonces [24] si $\pi = [\pi_{11}, \pi_{12}, ..., \pi_{kl}]'$ y $h(\theta) = [h_{11}(\theta), h_{12}(\theta), ..., h_{kl}(\theta)]$, las correlaciones policóricas estarían dadas por:

$$\pi = h(\theta) \tag{3.36}$$

3.4.5. Diagnóstico del Modelo

Para conocer si las relaciones de covariación identificadas en el modelo SEM son aceptables o no, es necesario realizar una medida del ajuste del mismo, esto es, realizar una estimación de la bondad de ajuste de $\Sigma = \Sigma(\theta)$. Muchos de ellos se derivan del valor de χ^2 valor el cual representa la diferencia entre S y $\hat{\Sigma}$ (recordar que esta última se deriva a partir del modelo planteado). Los índices de ajuste se pueden clasificar en:

- Funciones de discrepancia: Como la medida de χ^2, RMR, RMSEA.
- Índices comparativos de ajuste entre el modelo nulo y el modelo objetivo (nested models): Como NFI, RFI y CFI.
- Índices de proporción de varianza: Como GFI y PGFI

3.4.5.1. Chi-Square χ^2

Es la medida global de ajuste para el estimador de ML, la cual sigue una distribución asintótica χ^2_g donde n corresponde al tamaño de la muestra, tiene $g = (p+q)(p+q+1-t)$ grados de libertad y t parámetros libres. Específicamente se escribe como:

$$\chi^2 = N[tr(S\Sigma^{-1}) + log \mid \Sigma \mid - log \mid S \mid -(p+q)] \tag{3.37}$$

Con lo que se observa que χ^2 es sensible a los valores de N, lo que lleva a que sea extremadamente propenso a rechazar cuando las muestras son grandes (N ≥ 1000), $porloquenoseusamuchoc$

3.4.5.2. Root Mean Square Residual (RMR)

Calcula la diferencia entre S y Σ, es decir, el promedio de estos residuos se computa para medir hasta qué punto el modelo estima bien la matriz de covarianzas,

$$RMR = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}(s_{ij} - \hat{\sigma_{ij}})}{k(k+1)/2}} \tag{3.38}$$

Con $k = p+q$. Al igual que χ^2 grandes valores en el índice indican un mal ajuste, siendo cero el valor óptimo.

3.4.5.3. Root Mean Square Error of Approximation (RMSEA)

Se basa en el parámetro de no centralidad y puede interpretarse como el error de aproximación por medio de los grados de libertad, se calcula como:

$$RMSEA = \sqrt{\frac{\chi^2 - g}{g(N-1)}} \tag{3.39}$$

Donde N indica el tamaño de la muestra y g los grados de libertad del modelo. Valores de $RMSEA \leq 0.1$ indican un ajuste aceptable.

3.4.5.4. Normed Fit Index (NFI)

Mide la variación porcentual de falta de ajuste del estadístico χ^2 brindado por el modelo anidado respecto del modelo base, expresada como:

$$NFI = \frac{\chi^2_o - \chi^2_1}{\chi^2_o} \tag{3.40}$$

Por lo que $0 \leq NFI \leq 1$. Este índice no es muy usado ya que no tiene en cuenta los grados de libertad, por lo que favorece la adopción de modelos poco parsimoniosos.

3.4.5.5. Comparative Fit Index (CFI)

Evalua la eficiencia relativa entre el modelo propuesto y el modelo nulo, el cual se asume independiente y con covarianzas poblacionales iguales a cero

$$CFI = 1 - \frac{max((\chi_1^2 - dof_1), 0)}{max((\chi_o^2 - dof_o), 0)} \tag{3.41}$$

Donde χ_o^2 representa el valor de $chi - cuadrado$ para el modelo nulo y χ_1^2 el valor de $chi - cuadrado$ para el modelo propuesto. Valores de $CFI \geq 0.9$ indican un buen ajuste del modelo.

3.4.5.6. Relative Fit Index (RFI)

Se creó con el propósito de tener una medida de ajuste alternativa en ausencia de normalidad que no dependa del tamaño de la muestra:

$$RFI = \frac{\frac{\chi_o^2}{dof_o} - \frac{\chi_1^2}{dof_1}}{\frac{\chi_o^2}{dof_o}} \tag{3.42}$$

Donde $\chi_o^2 y \text{dof}_o$ corresponden al valor del estadístico $chi - cuadrado$ y los grados de libertad del modelo nulo respectivamente, mientras que $\chi_1^2 y \text{dof}_1$ son los valores de $chi - cuadrado$ y grados de libertad del modelo propuesto.

3.4.5.7. Goodness of Fit Index (GFI)

Este índice mide la variabilidad total explicada por el modelo respecto de la matriz de covarianzas muestral, está dado por la expresión:

$$GFI = 1 - \frac{tr[(s - \hat{\sigma})' W (s - \hat{\sigma})]}{tr[s' W]} \tag{3.43}$$

En la cual s y $\hat{\sigma}$ corresponde a los elementos vectorizados de S y $\hat{\Sigma}$ respectivamente, W por su parte es la matriz de pesos escogida según el método de estimación utilizado. Está definida de tal manera que $GFI \in (0, 1)$, valores altos de GFI indican un mejor ajuste.

3.4.5.8. Parsimony Goodness of Fit Index (PGFI)

Este estadístico intenta realizar un ajuste de GFI por falta de parsimonia, incluyendo en el cálculo la relación entre los grados de libertad del modelo calculado y el modelo base.

$$PGFI = \frac{dof_1}{dof_o} * GFI \tag{3.44}$$

En la tabla [3.2] se detallan los valores óptimos de ajuste de cada test mencionado en este trabajo:

Test	Sigla	Valor óptimo de Ajuste
chi-square	χ^2	≈ 0
Root Mean Square Residual	RMR	≤ 0.1
Root Mean Square Error of Approximation	RMSEA	≤ 0.1
Normed Fit Index	NFI	≥ 0.9
Comparative Fit Index	CFI	≥ 0.9
Relative Fit Index	RFI	≥ 0.9
Goodness of Fit Index	GFI	≥ 0.9
Parsimony Goodness of Fit Index	PGFI	≥ 0.9

Tabla 3.2: Medidas de Ajuste del Modelo

3.4.6. Modificación del Modelo

Generalmente el modelo inicial no se ajusta plenamente en la etapa de diagnóstico, esto puede obedecer a problemas de especificación, identificación o estimación, aunque generalmente es debido a problemas en la restricción de los parámetros. El objetivo de la etapa de modificación es realizar cambios que lleven a obtener un mejor diagnóstico y ajuste de $\Sigma(\theta)$, sin embargo, dada la característica exploratoria de esta etapa, los cambios deben introducirse uno a uno, con el fin de identificar plenamente las fuentes y la magnitud de la variación de un modelo a otro. Cada cambio genera un nuevo 'modelo' del cual se debe estudiar sus medidas de ajuste y diagnóstico, con el objetivo de saber si el ajuste realizado lleva a una mejor aproximación de la matriz de covarianzas.

Uno de los métodos más utilizados en esta etapa es la evaluación del comportamiento en la estadística χ^2 al momento de liberar parámetros restringidos inicialmente, ya que este estadístico permite comparar dos modelos que tienen las mismas variables pero se diferencian en sus parámetros.

3.4.6.1. Índices de Modificación

Los Índices de Modificación, conocida también como *test de multiplicadores de Lagrange* es una prueba que permite por medio del análisis de los residuos conocer el nivel de cambio que se produce en χ^2 al momento de la liberación de los parámetros con el fin de optimizar el ajuste entre S y $\Sigma(\theta)$, por lo que es una herramienta frecuentemente utilizada den la etapa de modificación.

Siendo $g = \partial f / \partial \theta$ un vector gradiente de la función de ajuste f, el Índice de Modificación está dado por la expresión [43]:

$$MI = \frac{\frac{1}{2}\hat{g}_1}{\hat{k} - \hat{d}' \hat{E^{-1}} \hat{d}} \tag{3.45}$$

donde $\hat{g_1}$ representa el vector gradiente para el parámetro liberado $\hat{g_1} = \partial f/\partial\theta_1$, $\hat{d}$ es un vector derivado de segundo orden de los valores esperados $E[\partial^2 f/(\partial\theta\partial\theta_1)]$, denota a $E[\partial^2 f/(\partial\theta_1\partial\theta_1)]$ evaluadas en $\hat{\theta}$ y $\hat{\theta_1}$.

Dichas derivadas se obtienen a partir del Polinomio de Taylor para la función de ajuste usada, que genéricamente se escribe:

$$f \approx \hat{f} + \begin{bmatrix} \theta & - & \hat{\theta} \\ \theta_1 & - & \hat{\theta_1} \end{bmatrix}' \begin{bmatrix} \hat{g} \\ \hat{g_1} \end{bmatrix}' + \frac{1}{2} \begin{bmatrix} \theta & - & \hat{\theta} \\ \theta_1 & - & \hat{\theta_1} \end{bmatrix}' \begin{bmatrix} \hat{E} & \hat{d} \\ \hat{d}' & \hat{k} \end{bmatrix} \begin{bmatrix} \theta & - & \hat{\theta} \\ \theta_1 & - & \hat{\theta_1} \end{bmatrix} \tag{3.46}$$

¿Cuándo parar de hacer modificaciones?

Varios modelos anidados pueden generar un buena estimación de $\Sigma(\theta)$ por lo que se deben tener criterios suplementarios, según Batista & Coenders [5] pueden ser los siguientes:

- Analizar la cantidad de cambio en χ^2.
- Analizar los índices de bondad de ajuste que tienen en cuenta la parsimonia del modelo (CFI, RMSEA).

3.5. Análisis Factorial

El propósito esencial del análisis factorial es describir las relaciones de covarianza entre muchas variables, en términos de unos pocos factores subyacentes, los cuales están conformados por aquellos grupos de variables que presentan una alta correlación intragrupal y a la vez una casi nula correlación intergrupal, entonces, es creíble que cada grupo de variables represente una única construcción o factor latente, el cual es el responsable de las correlaciones encontradas.

Para desarrollar el modelo consideraré un vector centrado de observaciones (respecto de su media) compuesto por una parte sistemática y un error no observable. La parte sistemática está compuesta por una serie de variables no observables directamente que se llaman factores latentes, la parte no sistemática corresponde a los errores los cuales están incorrelacionados.

El modelo descrito entonces es:

$$X = \Lambda F + \xi \tag{3.47}$$

es decir

$$\begin{array}{rcl} \mathrm{x}_1 & = & \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1m} f_m + e_1 \\ \mathrm{x}_2 & = & \lambda_{21} f_1 + \lambda_{22} f_2 + \cdots + \lambda_{2m} f_m + e_2 \\ \vdots & = & \vdots \\ \mathrm{x}_p & = & \lambda_{p1} f_1 + \lambda_{p2} f_2 + \cdots + \lambda_{pm} f_m + e_p \end{array}$$

El coeficiente λ_{ij} se conoce como *loading* (carga) de la $i-$ésima variable en el $j-$ésimo factor, por lo que la matriz Λ se denomina *matriz de cargas factoriales*. Se debe tener en cuenta que el $i-$ésimo factor e_i es asociado únicamente al al $p-$ésimo individuo, las $p+m$ variables aleatorias f y e no son observables, diferenciádose en este ítem caso del modelo de regresión múltiple.

Pueden distinguirse dos tipos de Análisis Factorial: Análisis Factorial Exploratorio (EFA) y en Análisis Factorial Confirmatorio (CFA). El EFA tiene como objetivo determinar el número de factores subyacentes que agrupan las variables observadas; por otro lado, en el CFA es el investigador quién con un conocimiento a-priori basado en la teoría y conocimientos previos de la situación, fija la relación que subyace entre las agrupaciones de variables y es en este caso que debe formularse las respectivas hipótesis de causación o covariación a contrastar.

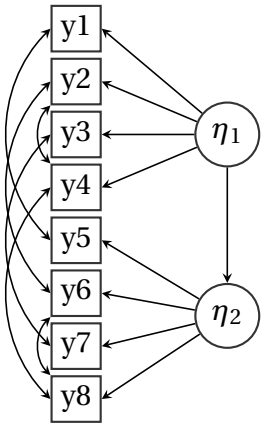


Figura 3.10: Ejemplo Modelo Factorial Confirmatorio

En la figura [3.10] se tiene un ejemplo de CFA, en el cual se plantea la hipótesis que las variables observadas y_1, y_2, y_3, y_4 son causadas por el factor η_1, mientras que las variables y_5, y_6, y_7, y_8 son causadas por η_2, factor el cual a su vez es afectado de manera directa por η_1, además se supone a-priori una covariación existente entre las variables observadas, la cual se cree de pequeña magnitud para las variables asociadas a factores diferentes, y de alta magnitud para las pertenecientes a un mismo factor.

Capítulo 4

Modelo de Lealtad

4.1. Introducción

Tal vez una de las inquietudes más frecuente de las empresas sin importar el país de origen, es conocer qué tan leales son sus clientes, para este fin se han construido múltiples tipos de modelos que abordan temáticas que van desde el capital de marca hasta la evaluación de las dimensiones del servicio final ofrecido.

El presente trabajo toma como línea de base la simulación de una investigación sobre la calidad en la atención recibida por el usuario en una empresa prestadora de servicios. Es importante tener en cuenta que a diferencia de la evaluación de productos físicos, la evaluación del servicio se basa en componentes intangibles los cuales están sujetos a la subjetividad del usuario según la experiencia vivida [40].

Para ello se tomó una medida de satisfacción con el servicio recibido, mediante las siguientes variables:

- *intenc*: La intención de tomar nuevamente los servicios de la empresa
- *recom*: La disposición a recomendar la empresa a otras personas
- *horario*: Los horarios de atención
- *ubicacion*: La ubicación de las sedes en el lugar donde se necesita
- *comodidad*: La comodidad de las sedes
- *senaliza*: La Señalización al interior de las sedes
- *seguridad*: La seguridad que siente en las sedes
- *personal*: Cantidad de personas atendiendo
- *tiempo*: Tiempo de espera en área de recaudo
- *amabi*: Amabilidad y respeto recibido en recaudo

- *agili*: Agilidad de la persona que atiende en recaudo
- *manejo*: Disposición de las filas en recaudo
- *espera*: Tiempo de espera al asesor
- *respet*: Amabilidad y respeto del asesor
- *interes*: Interés demostrado por el asesor
- *claro*: Claridad de la información que brindó el asesor
- *soluc*: Solución efectiva a las necesidades
- *ag*: Agilidad en la respuesta recibida
- *cupos*: Los cupos otorgados
- *requisit*: Los requisitos exigidos para adquirir el servicio
- *trami*: La agilidad en los trámites
- *antigue*: El reconocimiento de la antigüedad como cliente
- *comporta*: El reconocimiento por el buen uso del servicio
- *acompa*: El servicio post-venta
- *nuevo*: El ofrecimiento de nuevos productos

Dichas mediciones se realizaron en una escala ordinal de uno (1) a diez (10), siendo 1 la menos calificación y 10 la mayor calificación. El enfoque del trabajo será realizar el cálculo de un indicador de lealtad del cliente a partir del método de Ecuaciones Estructurales, el cual permitirá descomponer los efectos tanto directos como indirectos que sobre la lealtad tienen las diferentes dimensiones latentes del servicio. Previo a realizar el estudio estructural, se hace un análisis factorial exploratorio con el fin de encontrar las relaciones subyacentes entre las variables observadas.

El modelo teórico que se va a exponer, contempla que las variables arriba citadas conforman dimensiones de servicio las cuales tienen un efecto indirecto en la lealtad y directo en la satisfacción del cliente, la cual a su vez afecta directamente la lealtad del cliente.

4.2. Análisis Exploratorio y consistencia de los datos

Después de haber enumerado las variables con las cuales se va a calcular el Indicador de Lealtad mediante el modelo global de Ecuaciones Estructurales y teniendo en cuenta el carácter ordinal de estas, en esta sección se muestran los estadísticos descriptivos básicos de las variables en estudio, adicionalmente se reportan las pruebas de normalidad.

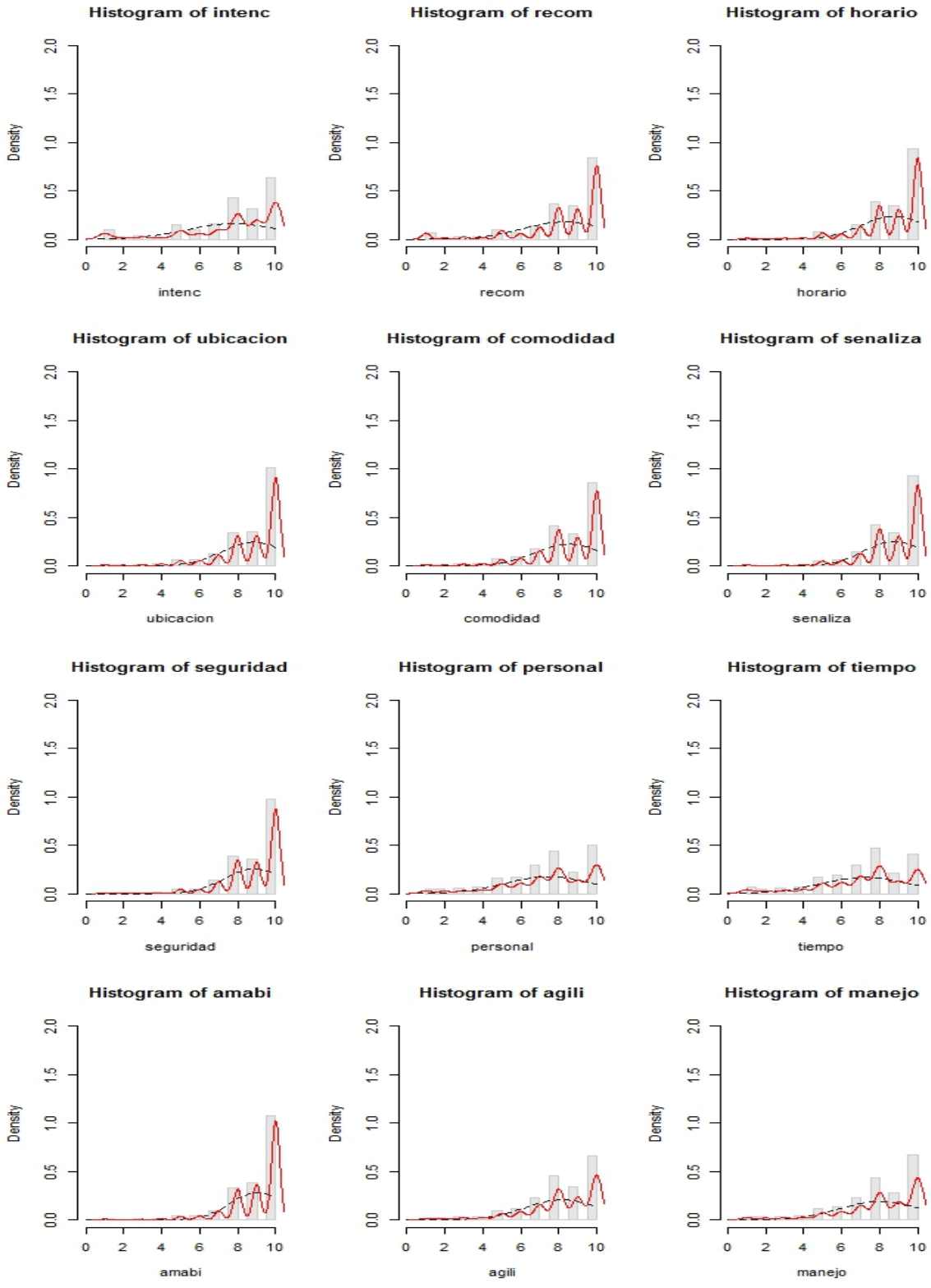
En la tabla [4.1] se observa que el tamaño de la muestra es de 8262 casos y no tiene valores perdidos, el promedio de calificación es en general superior a puntos, adicionalmente es importante resaltar que en la etapa exploratoria no se descartan de antemano variables para el análisis.

var	muestra	mean	sd	min	max	range
horario	8262	8,72	1,68	1	10	9
ubicacion	8262	8,84	1,63	1	10	9
comodidad	8262	8,55	1,78	1	10	9
senaliza	8262	8,77	1,59	1	10	9
seguridad	8262	8,86	1,55	1	10	9
personal	8262	7,48	2,28	1	10	9
tiempo	8262	7,24	2,36	1	10	9
amabi	8262	9,04	1,42	1	10	9
agili	8262	8,18	1,94	1	10	9
manejo	8262	8,04	2,1	1	10	9
espera	8262	7,79	2,2	1	10	9
respet	8262	9,12	1,39	1	10	9
interes	8262	8,57	1,89	1	10	9
claro	8262	8,76	1,71	1	10	9
soluc	8262	8,45	1,98	1	10	9
ag	8262	8,54	1,87	1	10	9
cupos	8262	7,99	2,34	1	10	9
requisit	8262	8,16	2,09	1	10	9
trami	8262	8,2	2,05	1	10	9
antigue	8262	7,62	2,66	1	10	9
comporta	8262	7,8	2,53	1	10	9
acompa	8262	7,67	2,53	1	10	9
nuevo	8262	7,63	2,49	1	10	9

Tabla 4.1: Estadísticos Descriptivos

La escala para cada una de las variables se encuentra entre 1 y 10, la cual es suficientemente amplia para que ser tratadas como variables numéricas (aunque no continuas), razón por la cual es válido realizar análisis y pruebas estadísticas con el fin de verificar los supuestos de los modelos.

En la figura [4.1] se presentan los histogramas de cada una de las variables con el fin de realizar una inspección visual de su comportamiento:


Histogram of intenc
Density
intenc
Histogram of recom
recom
Histogram of horario
horario
Histogram of ubicacion
ubicacion
Histogram of comodidad
comodidad
Histogram of senaliza
senaliza
Histogram of seguridad
seguridad
Histogram of personal
personal
Histogram of tiempo
tiempo
Histogram of amabi
amabi
Histogram of agili
agili
Histogram of manejo
manejo

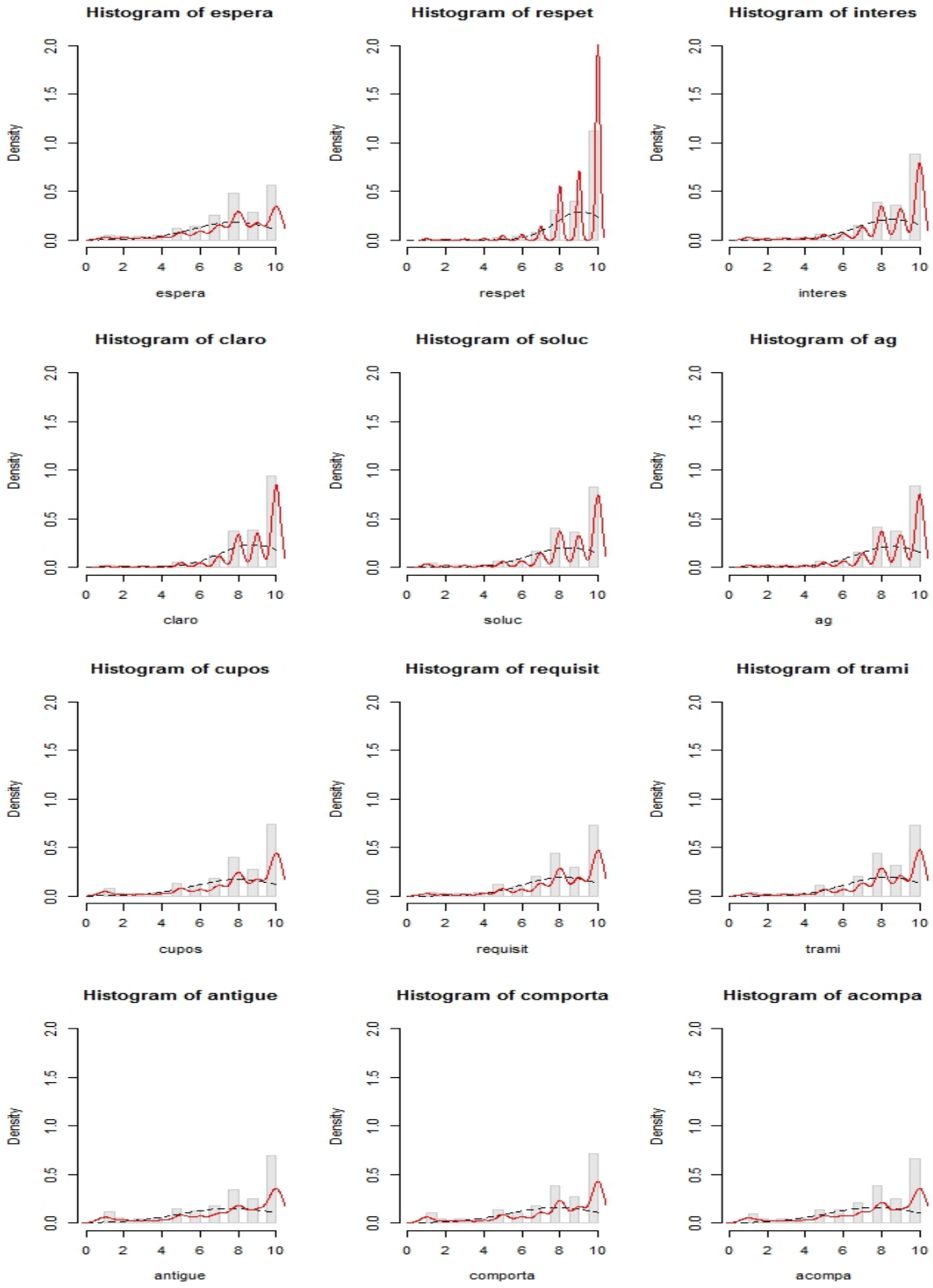


Figura 4.1: Histogramas de las variables estudiadas

A simple vista se observa que las variables bajo estudio no siguen una distribución normal, sin embargo se reporta el estadístico de Anderson Darling y su p-valor asociado en la tabla [4.2], teniendo en cuenta que su hipótesis nula es:

$$H_0 : X \sim N(\mu, \sigma)$$

var	**Estadística**	**p-value**
horario	631,5722	0,00
ubicacion	733,9707	0,00
comodidad	545,6445	0,00
senaliza	609,0171	0,00
seguridad	668,7004	0,00
personal	250,9835	0,00
tiempo	236,6273	0,00
amabi	802,5184	0,00
agili	391,6763	0,00
manejo	391,1865	0,00
espera	338,9154	0,00
respet	876,6862	0,00
interes	630,6738	0,00
claro	684,7826	0,00
soluc	603,0671	0,00
ag	585,1464	0,00
cupos	504,0901	0,00
requisit	453,0338	0,00
trami	459,0459	0,00
antigue	474,4710	0,00
comporta	509,7517	0,00
acompa	432,8783	0,00
nuevo	404,6410	0,00

Tabla 4.2: Test de Anderson Darling para contraste de normalidad

g1p:	103,7106
chi.skew:	142809,5
p.value.skew:	0,00
g2p:	1320,195
z.kurtosis:	998,6971
p.value.kurt:	0,00
chi.small.skew:	142865,7
p.value.small:	0,00

Tabla 4.3: Test de Normalidad Multivariada de Mardia

En este caso el test de Anderson Darling rechaza la hipótesis de normalidad para cada variable observada, igualmente los resultadios de la Prueba de Mardia presentados en la tabla [4.3] rechaza la hipótesis de normalidad multivariada, sin embargo los modelos de ecuaciones estructurales manejan métodos robustos para variables no normales, por lo cual es viable su aplicación.

Para analizar la consistencia de los datos, se utilizó el coeficiente Alpha de Cronbach, el cual estima la fiabilidad de la información a través del conjunto de ítems que se midieron. El procedimiento general para el cálculo del Alpha de Cronbach parte de la matriz de correlaciones de Pearson, ya que como demostraron Gelin, Beasley y Zumbo (2003), la utilización de una escala Likert con una cantidad mayor a 6 categorías estabiliza el coeficiente [12].

alpha	**std.alpha**	**Guttman's Lambda 6**
0,96	0,96	0,97

Tabla 4.4: Alpha de Cronbach

El coeficiente Alpha de Cronbach es de 0.96, por lo que se concluye que las variables recogen con alta fiabilidad la información requerida [44].

4.3. Análisis Factorial Exploratorio

Antes de entrar en la etapa de construcción del modelo, es necesario dar coherencia a la diversidad de información que se está midiendo mediante alguna técnica que permita encontrar, a partir de su estructura de correlación, relaciones subyacentes entre los vectores de análisis con el fin de definir grupos de variables que estén altamente correlacionadas entre sí a k factores latentes que expliquen la mayor cantidad de varianza de la matriz $\mathbf{X}$ original. Existen diversas técnicas estadísticas de interdependencia, en este caso se utiliza el Análisis Factorial Exploratorio **(EFA)**.

No obstante el **(EFA)** se base en el supuesto de normalidad, se considera su uso en este trabajo ya que el objetivo previo a la especificación del modelo estructural es contextualizar la situación y tener un modelo de medición que proporcione una base para un análisis causal de relaciones entre variables latentes [23].

4.3.1. Tratamiento de Variables Ordinales

El paso previo para la realización de un **(EFA)** es hacer una evaluación de la matriz de correlaciones con el fin de establecer si se justifica su aplicación; sin embargo se debe tener en cuenta que en este caso se está tratando con variables discretas ordinales, por lo que se debe estudiar la conveniencia de usar la matriz de Correlaciones de Pearson, ya que en algunas ocasiones no es apropiada para el análisis; en estos casos son las llamadas *correlaciones policóricas* las que deben emplearse como punto de partida [27]. Este tipo de correlaciones se usa para relacionar características que aunque en principio son continuas se utilizó una escala ordinal para medirlas, un claro ejemplo son las características medidas mediante las escalas de Likert utilizadas en este trabajo.

> Un enfoque típico para modelar variables ordinales es asumir que para cada variable ordinal y_i hay una variable subyacente y_i^*, y que cada y_i se relaciona a y_i^* a través de la función de paso:
>
> $$y_i = k \text{ cuando } \tau_{ik-1} < y_i^* \leq \tau_{ik}$$
>
> Para $k = 1, ..., m_i$, donde $\tau_i = -\infty$, $\tau_{ik} < tau_{ik+1}$, $tau_{imi} = \infty$. Los parámetros τ_i con $i = 1, ..., m_{i-1}$, son llamados umbrales de la $i-ésima$ variable[13].

Sin embargo, dependiendo del método de estimación, la matriz de correlaciones policóricas puede ser no definida positiva, con lo que el análisis factorial no sería posible, así que tomando como apoyo computacional el paquete *psych* del entorno *R*, se analiza por pruebas de bootstrap a las matrices de correlación de pearson y policóricas para determinar con cuál de ellas se realiza el **(EFA)**. En esta simulación los casos se crean mediante MASR y las correlaciones se calculan tantas veces como iteraciones hay. La media de correlación y su respectiva desviación estándar son calculadas en base a la transformación Z de Fisher de las correlaciones.

Si se denota por $\hat{\rho}$ a la correlación policórica de interés, el intervalo de la transformación de Fisher estará definido por:

$$z(\hat{\rho}) \pm z_{\delta/2} * SE(\hat{\rho})/(1-\hat{\rho}^2)$$

donde

$$z(\hat{\rho}) = 0.5 * ln[(1+\rho)/(1-\rho)]$$

y en el que $SE(\hat{\rho})$ es el error estándar de la correlación policórica [17].

Los resultados se muestran a continuación:

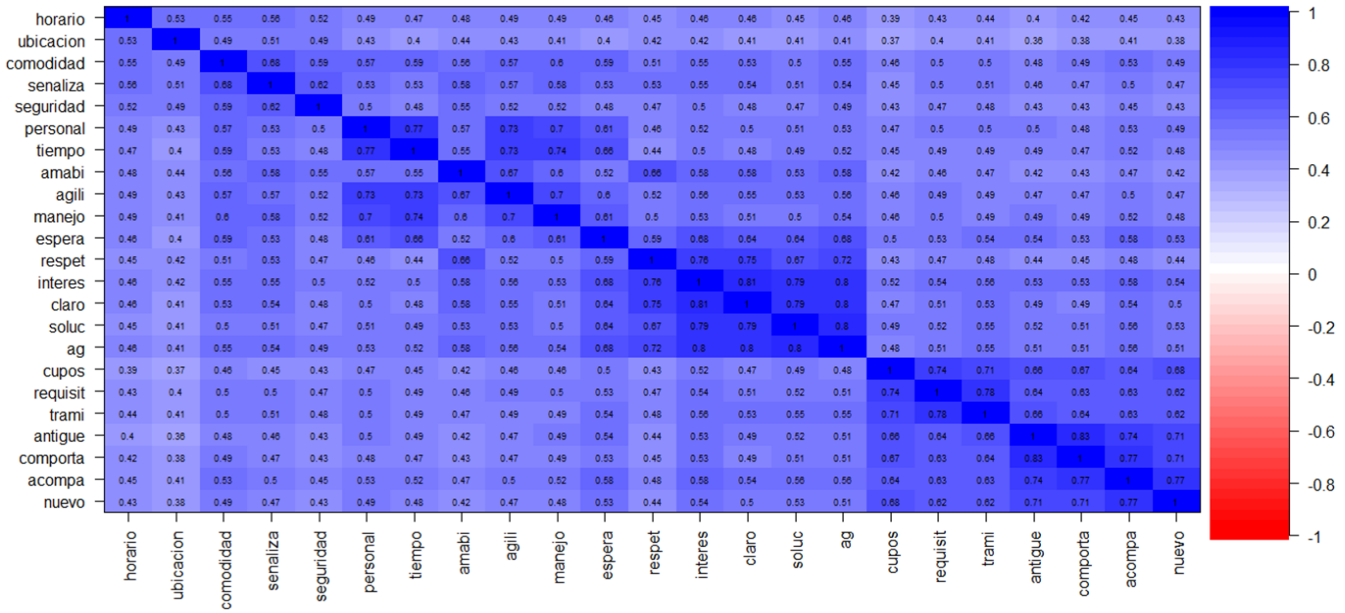


Figura 4.2: Matriz de Correlación de Pearson

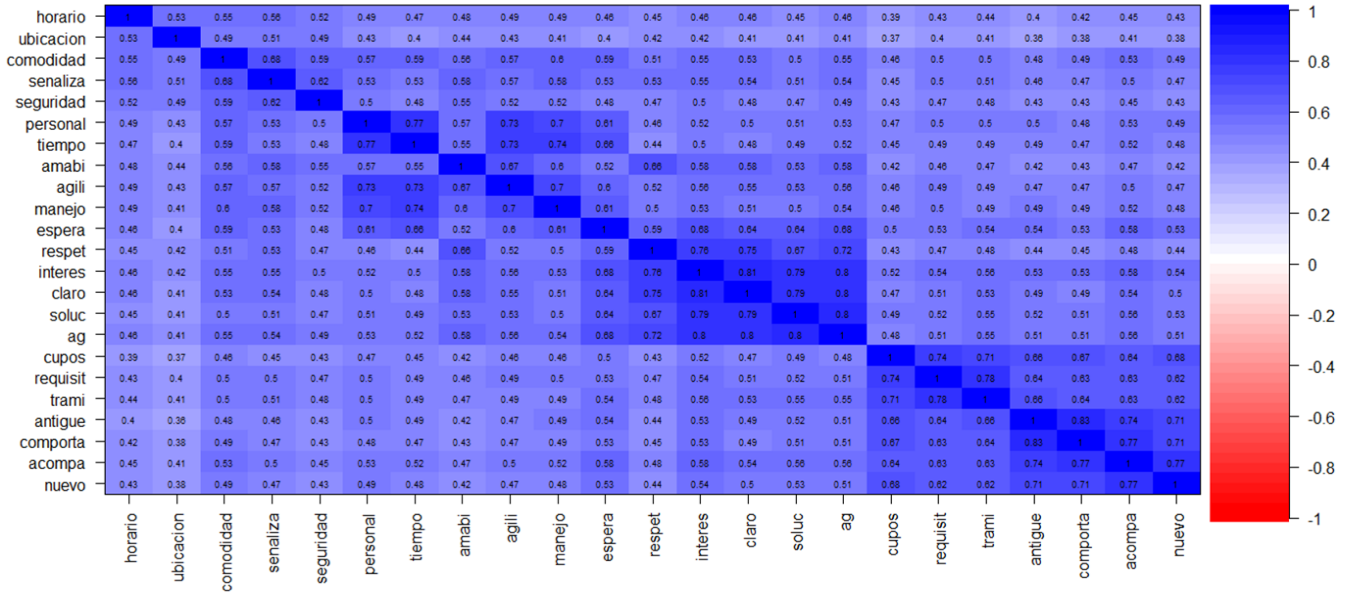


Figura 4.3: Matriz de Correlación Policórica

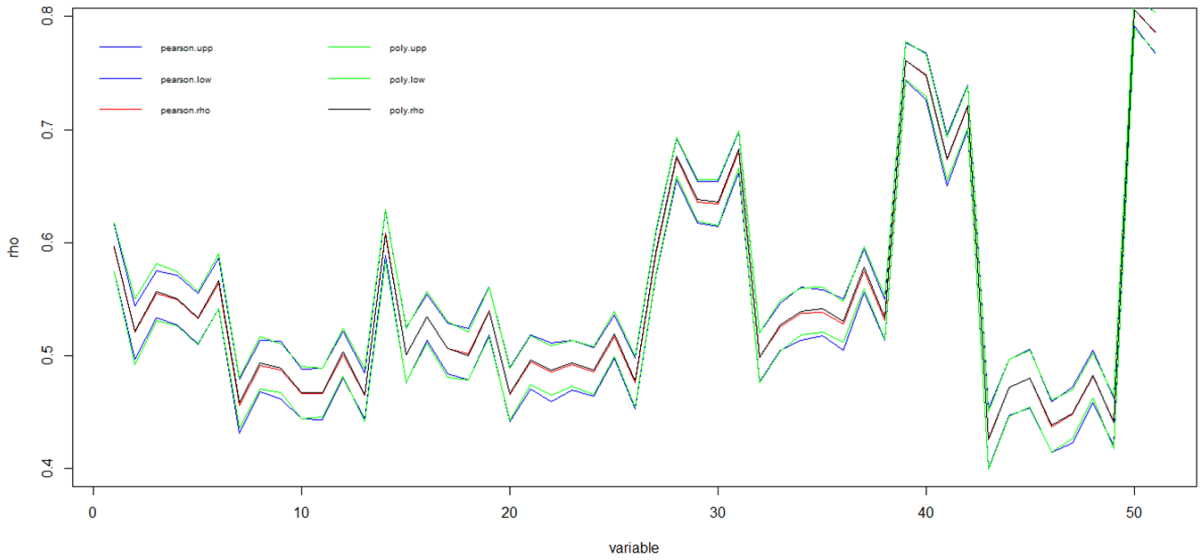


Figura 4.4: CI (95 %) generados por Bootstrap

En la gráficas [4.2] [4.3] y [4.4] se observa que los resultados generados por el método de correlaciones policóricas y la matriz de Correlación de Pearson para datos continuos son muy similares entre sí, y a la vez, difieren en su comportamiento de los resultados obtenidos a través del método de Spearman, esto sumado a la posibilidad de obtener una matriz de correlación policórica no definida positiva define que el presente trabajo se decante por la utilización de la Correlación Clásica de Pearson para la realización del Análisis Factorial Exploratorio.

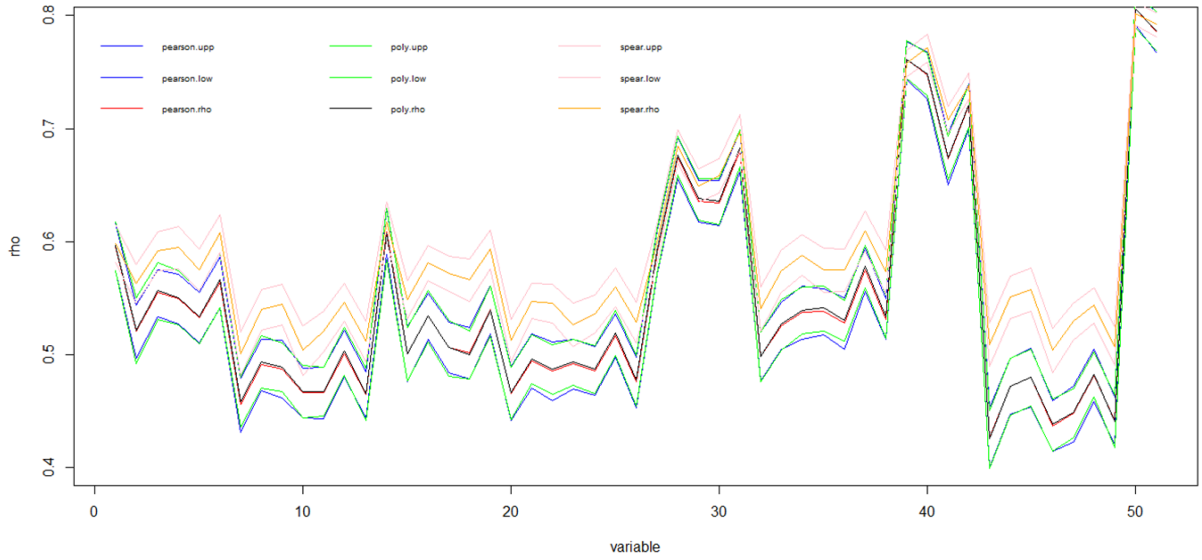


Figura 4.5: CI (95 %) generados por Bootstrap incluyendo Correlación de Spearman

4.3.2. Optimalidad de la estrategia y número de factores a retener

Antes de conducir cualquier tipo de análisis factorial, es necesario estudiar de antemano si es posible o recomendable aplicar la técnica a los datos que se tienen; hay varias formas de hacer este estudio y todas se basan en el estudio de la estructura de correlaciones de las variables. Ya se observó en la sub-sección anterior la matriz de correlación y visualmente parece lógico aplicar este análisis para encontrar la estructura subyacente de relación entre las variables, sin embargo, para confirmar esta presunción se aplicarán el Test de Esfericidad de Bartlett y el Índice de KMO.

4.3.2.1. Test de Esfericidad de Bartlett

El Test de Esfericidad de Bartlett compara si la matriz de correlación $R = (r_{ij})_{(pxp)}$ difiere significativamente de la matriz identidad, es decir:

$$H_0 : r_{ij(pxp)} = I_{(pxp)}$$

vs.

$$H_1 : r_{ij(pxp)} \neq I_{(pxp)}$$

Con el propósito de medir la relación global entre las variables, se calcula $\mid R \mid$. Bajo: $H_0 \mid R \mid = 1$; pero si las variables están altamente correlacionadas entonces $\mid R \mid \approx 0$.

Es estadístico de Bartlett está definido como

$$\mathbf{B} = -\left(n - 1 - \frac{2 * p + 5}{6}\right) * ln \mid R \mid$$

Bajo, H_0: **B** $\sim X^2_v$ con $v = p * (p-1)/2$, obteniendo los resultados expuestos en el cuadro [4.5]:

B.statistic:	164.219
degrees of freedom:	253
p.value:	0,00

Tabla 4.5: Test de Esfericidad de Bartlett

La prueba de esfericidad de Bartlett rechaza a todos los niveles de significancia H_0, por lo cual se concluye que es adecuado aplicar un análisis factorial a las variables bajo estudio.

4.3.2.2. Índice Kaiser-Meyer-Olkin (KMO)

El índice KMO busca el mismo objetivo que el test de Bartlett, pero esta técnica busca aislar la influencia que puedan tener el conjunto de variables sobre la relación de un par

de ellas; para lograr aislar este efecto hace uso de las correlaciones parciales entre pares de variables. La matriz de correlación parcial $\mathbf{P} = (p_{ij})$ se obtiene a partir de la inversa de la matriz de correlación inicial $\mathbf{R}^{-1} = (v_{ij})$ de la siguiente manera:

$$p_{ij} = -\frac{v_{ij}}{\sqrt{v_{ii}v_{jj}}}$$

El Índice KMO tiene rango acotado entre 0 y 1 y se cálcula:

$$KMO = \frac{\sum_i \sum_{j \neq i} r_{ij}^2}{\sum_i \sum_{j \neq i} r_{ij}^2 + \sum_i \sum_{j \neq i} p_{ij}^2}$$

En este caso $KMO = 0,9198213 > 0,5$, por lo que según [18] se concluye que es adecuado realizar el análisis factorial.

4.3.2.3. Número de factores a retener

Como regla inicial, el número de factores a retener debe ser cercano al número de auto-valores positivos de la matriz de correlación [18], sin embargo en este caso es posible obtener una gran cantidad de valores propios positivos pero muy cercanos a cero, lo que en algunas ocasiones hace que se maneje gran cantidad de factores que aportan muy poca información de análisis, en este caso se utilizan dos técnicas para escoger el número de factores a retener.

4.3.2.3.1. Regla de Kaiser-Guttman: Es la más fácil de aplicar, la más utilizada y la que paquetes estadísticos populares como SPSS tienen incorporada por defecto, los pasos para obtener el número de factores a retener son los siguientes:

1. Obtener los valores propios a_i de la matriz de correlaciones **R** inicial.
2. Determinar el número de valores propios tales que $a_i > 1.0$.
3. Este es el número de factores que se retendrá en el análisis factorial.

Analizando el screeplot generado por la regla de Kaiser-Guttman que se encuentra en la gráfica [[**?**]] se observa que los factores a retener son 3, sin embargo un problema de esta técnica es que descarta los valores propios que son muy cercanos a 1 sin evaluar si son estadísticamente significativos.

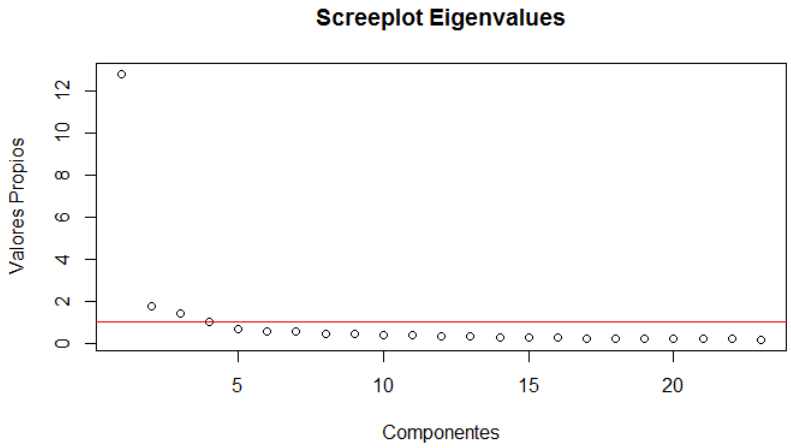


Figura 4.6: Screeplot Kaiser-Guttman

4.3.2.3.2. Regla de Very Simple Structure (VSS): VSS compara la matriz de correlación original por una versión simplificada de la matriz de factores original siguiendo los siguientes pasos [34]:

1. Encontrar una solución inicial con k factores, con el método de extracción de preferencia.
2. Rotar la solución para maximizar el criterio de rotación preferido, llame a este la matriz de factores patrón F_{k^*}.
3. Para una solución de estructura simple para el factor v, reemplace los $k-v$ elementos más pequeños en cada fila de F_{k^*} con ceros. Esta será la matriz simplificada S_{vk}, esto es lo que se hace en la práctica cuando se interpreta el factor por sus cargas más altas.
4. Para evaluar la eficacia de un factor de solución rotada particular F_k, ajuste un modelo de estructura simple de factor complejo v considerando que tan bien la matriz

$$R_v^* = S_{vk}\Phi S_{vk}^T$$

(donde Φ es la matriz de factores inter-correlación) reproduce la matriz de correlaciones iniciales R, es decir, encontrar la matriz residual:

$$\overline{R_v} = R - R_v^* = R - S_{vk}\Phi S_{vk}^T$$

5. Como un índice de ajuste de $\overline{R_v}$ a R, encontrar uno menos el cociente del cuadrado medio de la correlación residual sobre el cuadrado medio de la correlación original:

$$VSS_{vk} = 1 - \frac{MS_{\bar{r}}}{MS_r}$$

donde los grados de libertad con los que se calculan los cuadrados medios son el número de correlaciones estimado menos el número de parámetros libres en S_{vk}. Los cuadrados medios son encontrados para los elementos de la matriz diagonal inferior en R y $\bar{R}$.

6. Entonces para determinar el número adecuado de factores a extraer, se debe encontrar el valor del criterio de VSS para todos los valores desde k hasta el rango de la matriz. El número óptimo de factores interpretables es el número de k factores que maximizan a VSS_{vk}.

Los resultados obtenidos por el método VSS con asistencia del paquete $psych$ del entorno R muestran que el número de óptimo a retener es de 4 factores ya que en este punto se encuentra el mayor número de factores con los que $VSS_{vk} \approx 1$.

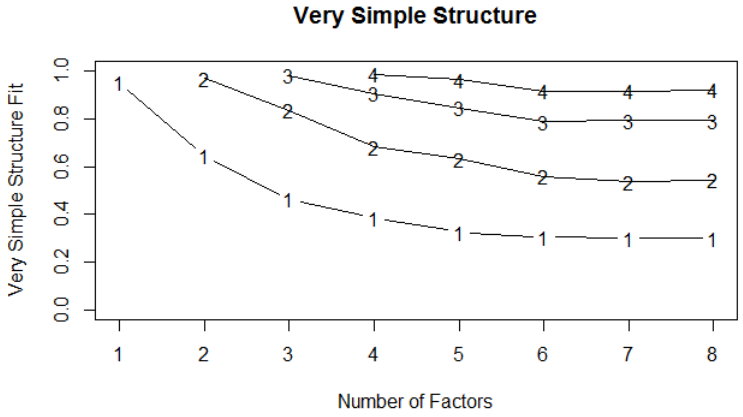


Figura 4.7: Factores a retener por la regla VSS

4.3.2.4. Método de Extracción

El método usado para la extracción de los factores es el de Factor Principal, en el cual extrae la matriz factorial con la propiedad de que los factores explican la máxima varianza y adicionalmente son incorrelacionados, en este se supone que existe un factor común subyacente a las variables por lo que busca extraer la máxima varianza en cada factor para que al final los k factores resultantes expliquen la mayor cantidad de varianza común, aunque las cargas de los factores extraídos no difieren sustancialmente del método de la Componente Principal [35].

El método se aplica a la matriz de correlación R en la cual antes de ser descompuesta espectralmente se reemplazan los unos en la diagonal por las comunalidades para que vía multiplicadores de Lagrange se pueda maximizar la varianza explicada por el primer factor, se itera este proceso con los $p-1$ factores restantes hasta que los cálculos converjan en la matriz factorial.

A este método se le aplicó la rotación Varimax, con el fin de maximizar los pesos de cada factor esperando así que cada variable sea bien representada en solo uno de ellos y a la vez se minimice el máximo número de variables correlacionadas a cada factor.

El análisis resultante extrayendo los 4 factores mediante el método citado se expone a en la tabla [4.6] y en la tabla [4.7]:

item	**FAC1**	**FAC2**	**FAC3**	**FAC4**	**comunalidad**	**unicicidad**
antigue		0,78			0,75	0,25
comporta		0,78			0,75	0,25
nuevo		0,73			0,68	0,32
cupos		0,73			0,66	0,34
acompa		0,72			0,72	0,28
requisit		0,67			0,65	0,35
trami		0,67			0,65	0,35
claro			0,79		0,81	0,19
interes			0,76		0,82	0,18
ag			0,76		0,8	0,2
soluc			0,74		0,76	0,24
respet			0,7		0,69	0,31
espera	0,46		0,49		0,64	0,36
tiempo	0,79				0,81	0,19
personal	0,71				0,73	0,27
agili	0,66				0,71	0,29
manejo	0,65				0,7	0,3
senaliza				0,65	0,66	0,34
seguridad				0,6	0,56	0,44
horario				0,57	0,51	0,49
comodidad				0,57	0,63	0,37
ubicacion				0,55	0,44	0,56
amabi				0,45	0,58	0,42

Tabla 4.6: Resultados del Análisis Factorial Exploratorio

item	FAC1	FAC2	FAC3	FAC4
SS loadings	4,99	4,26	3,32	3,14
Proportion Var	0,22	0,19	0,14	0,14
Cumulative Var	0,22	0,4	0,55	0,68
Proportion Explained	0,32	0,27	0,21	0,2
Cumulative Proportion	0,32	0,59	0,8	1,00
Correlation of scores with factors	0,94	0,94	0,9	0,84
Multiple R square of scores with factors	0,88	0,88	0,82	0,7
Minimum correlation of possible factor scores	0,76	0,75	0,64	0,41

Tabla 4.7: Estadísticos de Resumen

DIMENSION	VARIABLE
TRATO	antigue: El reconocimiento de la antigüedad como cliente comporta: El reconocimiento por el buen uso del servicio nuevo: El ofrecimiento de nuevos productos cupos: Los cupos otorgados acompa: El servicio post-venta requisit: Los requisitos exigidos para adquirir el servicio trami: La agilidad en los trámites
ASESORES	claro: La claridad de la información que le brindó el asesor interes: El interés que demostró por sus necesidades ag: La agilidad en la respuesta soluc: La Solución efectiva a sus requerimientos respet: La amabilidad y respeto en el trato del asesor espera: El tiempo de espera previo a la atención del asesor
CAJAS	tiempo: El tiempo de espera en las filas personal: La cantidad de personas que atienden las cajas agili: La agilidad de las personas que atienden las cajas manejo: El manejo adecuado de las filas para clientes
OFICINAS	senaliza: La Señalización al interior de las oficinas seguridad: La seguridad que siente en la oficina horario: Los horarios de atención de las oficinas comodidad: La comodidad de las Oficinas ubicacion: La ubicación de las oficinas en el lugar donde se necesita amabi: La amabilidad y respeto de las personas que atienden

En el análisis de la etapa exploratoria se encuentra un total de 4 factores que recogen un 68 % de la varianza; estos factores generan las dimensiones subyacentes del servicio según las evaluaciones realizadas a los usuarios. Estas dimensiones diferencian claramente cada uno de los momentos de atención que se presentan en las empresas de servicios.

1. Primer Factor: El primer factor es conformado por un total de 7 variables con ponderaciones mayores a 0.65, a su vez estas variables no alcanzan en otros factores importancias mayores a 0.25, todas estas variables se refieren a la satisfacción con el trato de servicios ofrecido por la empresa.
2. Segundo Factor: Está conformado por 6 variables asociadas a la atención personal recibida por los asesores de servicio cuyas ponderaciones son mayores a 0.49.
3. Tercer Factor: Compuesto por cuatro variables con ponderaciones mayores a 0.65 y está conformada por la variables que evalúan la atención recibida en cajas.
4. Cuarto Factor: Lo integran 6 variables las cuales tienen ponderaciones mayores a 0,45 y conforman la dimensión subyacente de atención en oficinas.

4.4. Fase de Aplicación

4.4.1. Etapa de Especificación

En esta sección se busca establecer formalmente el modelo; anteriormente se establecieron 4 configuraciones de interdependencia, en las que los factores son medidas latentes de la satisfacción con cada uno de los momentos del servicio y en las que se asume la existencia de una relación lineal entre los factores y variables observadas.

Es así que el primer factor toma la forma presentada en la figura [4.8]:

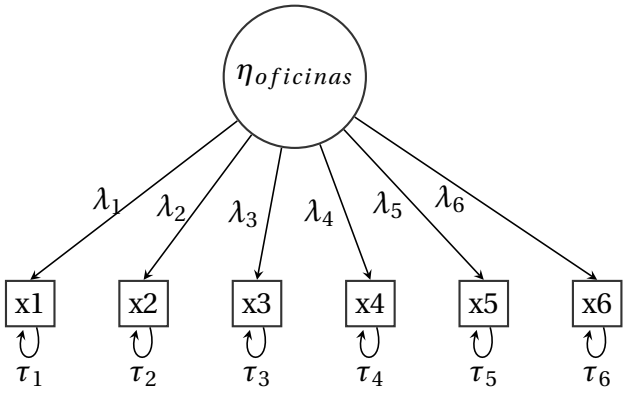


Figura 4.8: Path Diagram Factor 1 - OFICINAS

El *path diagram* expresa el efecto directo que tiene la variable subyacente de satisfacción con las oficinas en las variables observadas, que se pueden convertir al modelo de

medida

$$\mathbf{X} = \Lambda\eta + \tau$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \end{bmatrix} [\eta] + \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \\ \tau_6 \end{bmatrix}$$

donde η es la variable latente OFICINAS, Λ es la matriz de efectos desconocidos a estimar τ es la matriz de error de medida y finalmente **X** es la matriz de variables observadas donde x_1=senaliza, x_2=seguridad, x_3=horario, x_4=comodidad, x_5=ubicacion y x_6=amabi.

El factor CAJAS se puede representar gráficamente según la gráfica [4.9]:

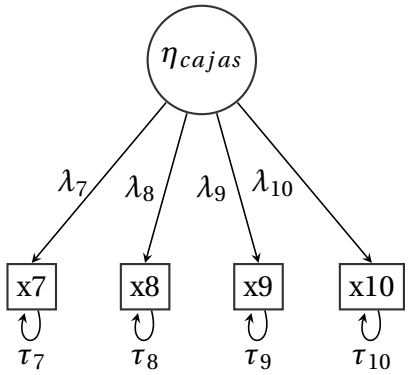


Figura 4.9: Path Diagram Factor 2 - CAJAS

y la ecuación que representa su modelo de medida

$$\begin{bmatrix} x_7 \\ x_8 \\ x_9 \\ x_{10} \end{bmatrix} = \begin{bmatrix} \lambda_7 \\ \lambda_8 \\ \lambda_9 \\ \lambda_{10} \end{bmatrix} [\eta] + \begin{bmatrix} \tau_7 \\ \tau_8 \\ \tau_9 \\ \tau_{10} \end{bmatrix}$$

donde η es la variable latente cajas, Λ es la matriz de efectos desconocidos a estimar τ es la matriz de error de medida y finalmente **X** es la matriz de variables observadas donde x_7=tiempo, x_8=personal, x_9=agili, x_{10}=manejo.

En el *path diagram* [4.10] representa al factor de ASESORES se puede observar la misma estructura del anterior, por lo que se genera una ecuación similar

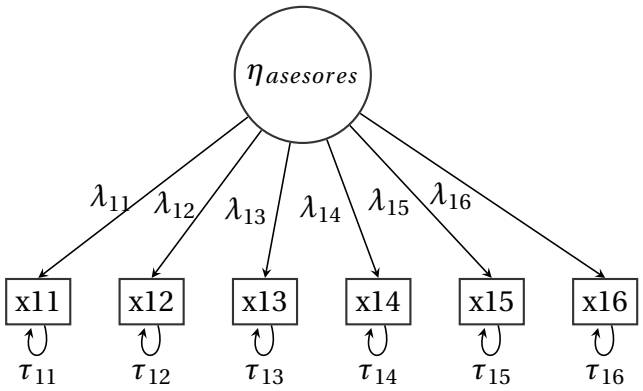


Figura 4.10: Path Diagram Factor 3 - ASESORES

Igualmente este *path diagram* expresa el efecto directo que tiene la variable subyacente de asesores en las variables observadas, entonces su modelo de medida:

$$\begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \\ x_{15} \\ x_{16} \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \lambda_{12} \\ \lambda_{13} \\ \lambda_{14} \\ \lambda_{15} \\ \lambda_{16} \end{bmatrix} [\eta] + \begin{bmatrix} \tau_{11} \\ \tau_{12} \\ \tau_{13} \\ \tau_{14} \\ \tau_{15} \\ \tau_{16} \end{bmatrix}$$

En la que η es la variable latente asesores, Λ es la matriz de efectos desconocidos a estimar τ es la matriz de error de medida y finalmente $\mathbf{X}$ es la matriz de variables observadas donde x_{11}=claro, x_{12}=interes, x_{13}=ag, x_{14}=soluc, x_{15}=respet y x_{16}=espera.

Finalmente el factor TRATO es representado gráficamente como se muestra en [4.11]

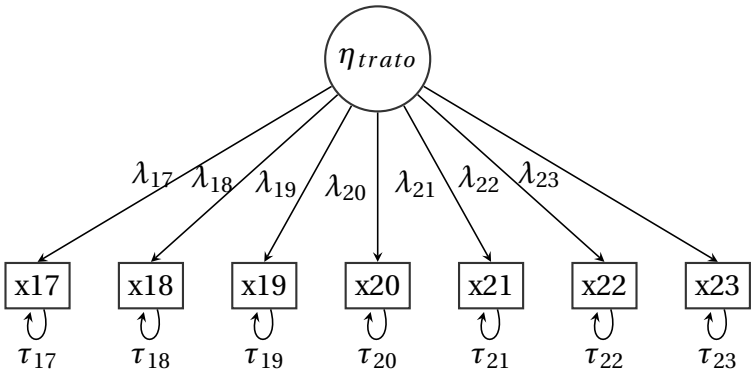


Figura 4.11: Path Diagram Factor 4 - TRATO

y la ecuación que lo representa

$$\begin{bmatrix} x_{17} \\ x_{18} \\ x_{19} \\ x_{20} \\ x_{21} \\ x_{22} \\ x_{23} \end{bmatrix} = \begin{bmatrix} \lambda_{17} \\ \lambda_{18} \\ \lambda_{19} \\ \lambda_{20} \\ \lambda_{21} \\ \lambda_{22} \\ \lambda_{23} \end{bmatrix} [\eta] + \begin{bmatrix} \tau_{17} \\ \tau_{18} \\ \tau_{19} \\ \tau_{20} \\ \tau_{21} \\ \tau_{22} \\ \tau_{23} \end{bmatrix}$$

igualmente η es la variable latente TRATO, Λ es la matriz de efectos desconocidos a estimar τ es la matriz de error de medida y finalmente **X** es la matriz de variables observadas donde x_{17}=antigue, x_{18}=comporta, x_{19}=requisit, x_{20}=cupos, x_{21}=acompa, x_{22}=nuevo y x_{23}=trami.

Para finalizar la etapa de especificación no se debe perder de vista que los anteriores factores latentes determinados de manera exploratoria tienen efecto (directo e indirecto) sobre la satisfacción del servicio, por lo que en esta etapa se definen las siguientes hipótesis estructurales:

- H1: El factor subyacente oficinas tiene un efecto significativo sobre la satisfacción con el servicio, es decir $\eta_{oficinas} \neq 0$.
- H2: El factor subyacente cajas tiene un efecto significativo sobre la satisfacción con el servicio, es decir $\eta_{cajas} \neq 0$.
- H3: El factor subyacente asesores tiene un efecto significativo sobre la satisfacción con el servicio, es decir $\eta_{asesores} \neq 0$.
- H4: El factor subyacente trato tiene un efecto significativo sobre la satisfacción con el servicio, es decir $\eta_{trato} \neq 0$.
- H5: La satisfacción general tiene un efecto significativo sobre la lealtad con la marca, es decir $\beta_{sat_{gen}} \neq 0$.
- H6: La recomendación tiene un efecto significativo sobre la lealtad con la marca, es decir $\beta_{recom} \neq 0$.
- H7: La intención de recompra tiene un efecto significativo sobre la lealtad con la marca, es decir $\beta_{intenc} \neq 0$.

En este caso la satisfacción general está dada por el diagrama [4.12]:

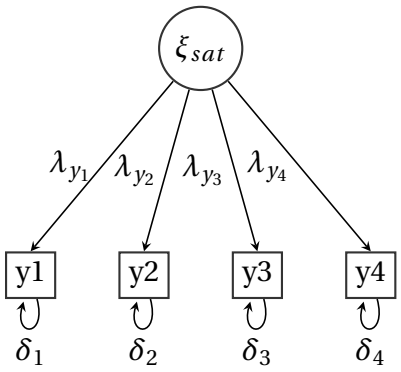


Figura 4.12: Path Diagram Factor SATISFACCIÓN

y la ecuación que representa su modelo de medida

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{y_1} \\ \lambda_{y_2} \\ \lambda_{y_3} \\ \lambda_{y_4} \end{bmatrix} [\xi] + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

donde ξ es la variable latente de satisfacción general, Λ es la matriz de efectos desconocidos a estimar δ es la matriz de error de medida y finalmente **Y** es la matriz de variables observadas donde y_1=planta física, y_2=área financiera, x_9=atención personal, x_{10}=variedad de productos.

Por lo que podemos observar en la gráfica [4.13] que el *path analysis* genera las siguientes ecuaciones para el modelo estructural:

$$\xi = \gamma^T \eta + \varsigma$$

$$[\xi] = [\gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4] \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix} + [\varsigma]$$

Y finalmente:

$$Leal = \beta^T v + \varepsilon$$

donde v es un vector que contiene las variables exógenas que afectan la lealtad de manera directa;

$$Leal = [\beta_1 \quad \beta_2 \quad \beta_3] \begin{bmatrix} \xi \\ recom \\ inten \end{bmatrix} + [\varepsilon]$$

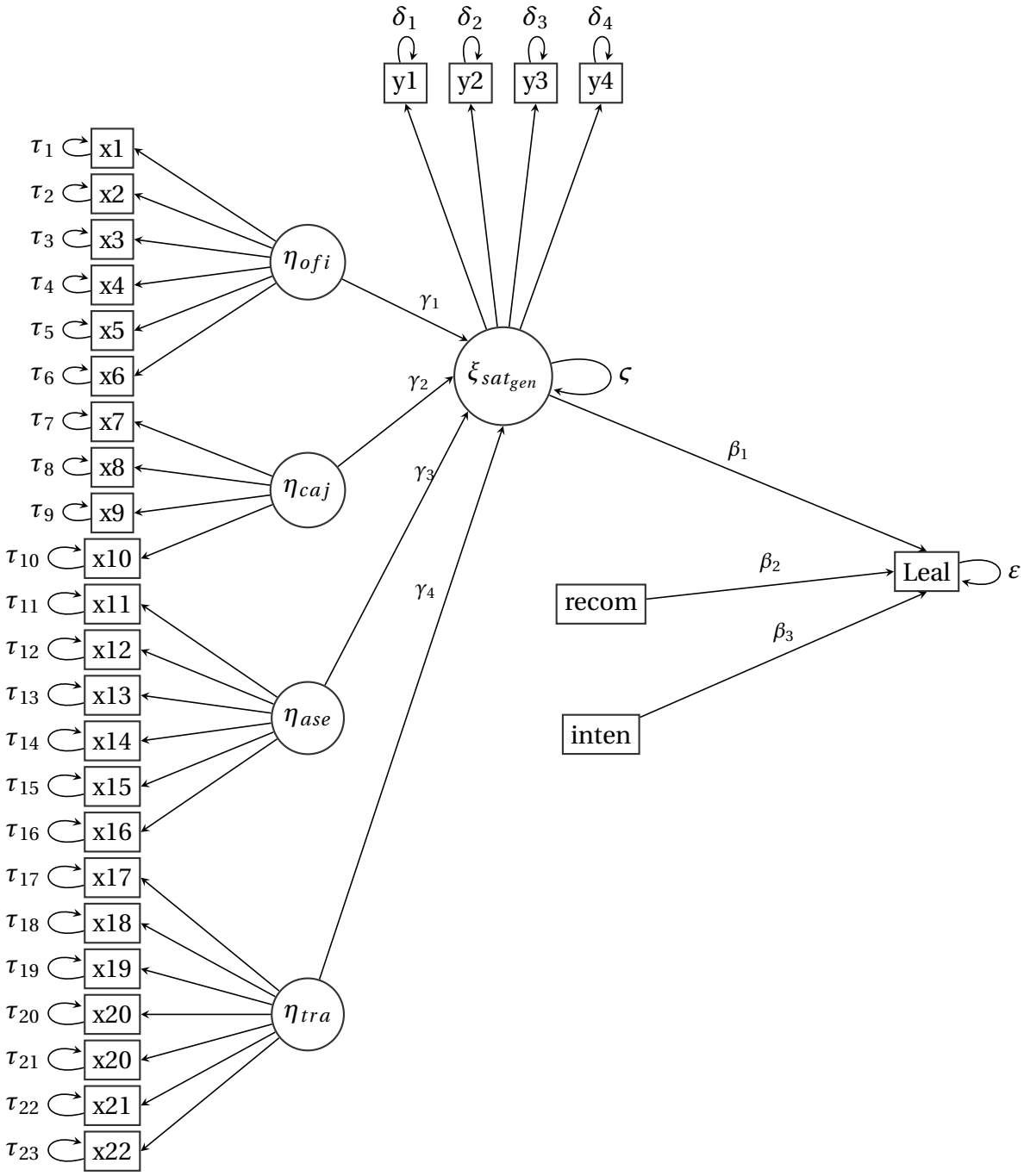


Figura 4.13: Path Diagram Etapa Especificación

4.4.2. Etapa de Identificación

A primera vista y asumiendo que todas las variables observadas son incorrelacionadas, el modelo se encontraría sobreidentificado, ya que analizando los grados de libertad

$$dof = m(m+1)/2 - 2m - \xi(\xi-1)/2 = 400 > 0$$

Sin embargo, no existe una prueba estadística para realizar la identificación del modelo, así como tampoco existe una condición necesaria, suficiente y general para hacerlo, aunque existen ciertas normas según el tipo de modelo estructural que se esté tratando. Dado que este modelo contiene variables latentes el cumplimiento de las siguientes condiciones simultáneas, son suficientes para aceptar el modelo como identificado [19]:

1. Cada variable latente tiene al menos dos variables observadas que se relacionan con una sola variable latente.

2. Cada variable latente tiene al menos una variable observada con un efecto directo distinto de cero que se usa para fijar la escala de la variable latente. el modelo para η tiene una estructura identificada.

4.4.3. Etapa de Estimación

4.4.3.1. Método de Estimación apropiado

Dada la no normalidad de las variables bajo estudio y su carácter ordinal, se escoge para la estimación una técnica que no haga ningún supuesto sobre la distribución de las variables, de manera que la kurtosis que se presenta no afecten los errores estándar de las estimaciones.

Este método minimiza la función:

$$F_{DWLS} = [\hat{\rho} - \rho(\theta)]' \, diag(W_{\rho\rho})^{-1} [\hat{\rho} - \rho(\theta)]$$

Dado que estos métodos buscan reducir la distancia entre las varianzas observadas y el diseño modelado, mediante la matriz **W** calculada por los Mínimos Cuadrados Ponderados (WLS), será este precisamente el método de estimación escogido para el modelo, particularmente el método robusto ($DWLS$) que se basa en las correlaciones policóricas de las variables observadas y que arrojan los resultados que se observan en la gráfica [4.14] y en las tablas [4.8] y [4.9]:

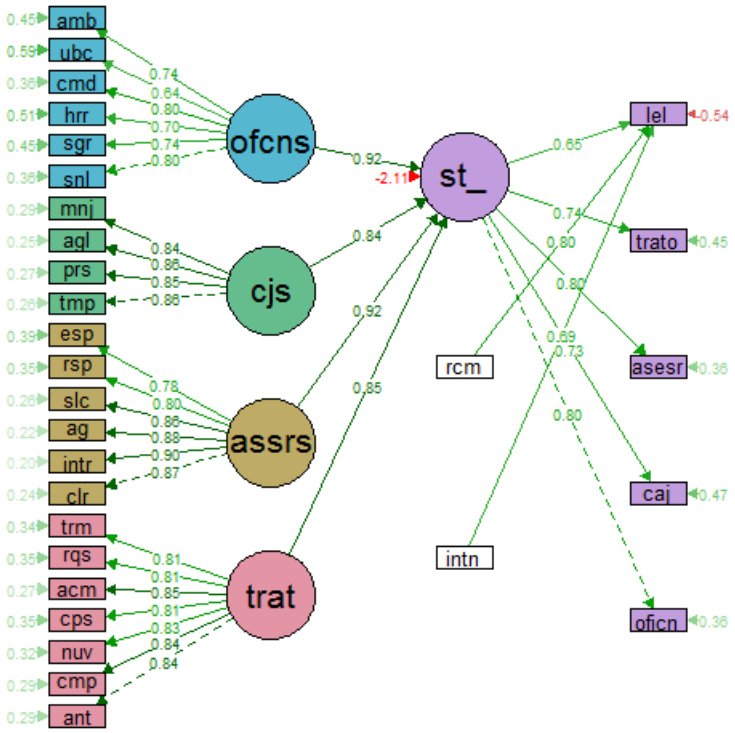


Figura 4.14: Estimación Inicial del Modelo

Variable/Factor	Estimate	Std.err	Z-value	P(>|z|)	Std.lv	Std.all
Regressions:						
sat_gene						
oficinas	1,042	0,017	60,355	0	0,924	0,924
cajas	0,59	0,009	63,886	0	0,836	0,836
asesores	0,885	0,014	65,284	0	0,919	0,919
trat	0,545	0,007	83,501	0	0,848	0,848
lealtad						
intenc	0,585	0,02	29,922	0	0,585	0,688
recom	0,761	0,028	27,443	0	0,761	0,803
sat_gene	0,958	0,009	105,588	0	1,377	0,653

Tabla 4.8: Estimaciones Iniciales

Variable/Factor	Estimate	Std.err	Z-value	P(>|z|)	Std.lv	Std.all
trat =						
antigue	1				2,237	0,842
comporta	0,956	0,011	85,695	0	2,138	0,845
nuevo	0,922	0,011	85,2	0	2,062	0,828
cupos	0,845	0,01	82,581	0	1,89	0,807
acompa	0,968	0,011	86,102	0	2,165	0,855
requisit	0,753	0,009	82,184	0	1,683	0,807
trami	0,744	0,009	81,573	0	1,665	0,812
asesores =						
claro	1				1,492	0,872
interes	1,133	0,017	65,851	0	1,69	0,895
ag	1,104	0,017	65,858	0	1,647	0,883
soluc	1,148	0,017	65,895	0	1,712	0,863
respet	0,747	0,012	61,478	0	1,115	0,804
espera	1,148	0,017	66,513	0	1,713	0,78
cajas =						
tiempo	1				2,036	0,863
personal	0,953	0,014	68,827	0	1,94	0,852
agili	0,824	0,012	66,645	0	1,678	0,864
manejo	0,871	0,013	67,399	0	1,773	0,843
oficinas =						
senaliza	1				1,275	0,802
seguridad	0,899	0,015	59,536	0	1,146	0,741
horario	0,92	0,015	59,539	0	1,172	0,699
comodidad	1,123	0,018	61,802	0	1,431	0,802
ubicacion	0,82	0,014	58,543	0	1,045	0,64
amabi	0,823	0,014	58,096	0	1,049	0,741
sat_gene =						
planta_fisica	1				1,437	0,801
financiera	0,864	0,008	109,268	0	1,243	0,73
atencion	0,968	0,009	112,323	0	1,391	0,802
variedad	1,004	0,009	112,506	0	1,444	0,745

Tabla 4.9: Estimaciones Iniciales (cont.)

Los resultados de la gráfica muestran que la recomendación tiene un mayor efecto sobre la lealtad con la marca que el producido por la recompra o la misma satisfacción general con el servicio, sin embargo se observa un problema de mala especificación del modelo, ya que la varianza del factor Satisfacción General es negativa, por lo que se debe mejorar la estructura de covarianzas del modelo inicial para evitar este problema. Hay dos tipos de efectos sobre la lealtad; los Efectos Indirectos que son aquellos que afectan la variable a través de otras (En este caso tienen efecto indirecto la satisfacción con las oficinas,

las cajas, la asesoría y el trato), y tienen efecto directo sobre la lealtad la satisfacción general y las variables observadas recomendación e intención de recompra. En los cuadros se puede observar que en general las variables ingresadas en el modelo son estadísticamente significativas en el cálculo de cada factor.

El efecto total de cada variable se calcula realizando la multiplicación de sus estimaciones siguiendo en el diagrama la ruta que conecta la variable en cuestión con la lealtad. Los efectos calculados se muestran a en la tabla [4.10].

variable	**estima**	**ef.indirecto**	**ef.directo**	**total**
oficinas	0,924	0,603372	0	0,603372
cajas	0,836	0,545908	0	0,545908
asesores	0,919	0,600107	0	0,600107
trat	0,848	0,553744	0	0,553744
intenc	0,803	0	0,803	0,803
recom	0,653	0	0,653	0,653

Tabla 4.10: Efecto sobre la lealtad

Inicialmente se observa que entre aquellas variables que afectan indirectamente la lealtad, es la relación con los asesores y el factor oficinas las que tienen el mayor efecto, mientras que a nivel de efecto directo es la intención de recompra quién genera mayor afectación a la medida de lealtad.

Una condición de los modelos de ecuaciones estructurales, es que la calidad de su ajuste se mide a través de la capacidad que tienen las estimaciones para reproducir la matriz de covarianzas muestral, o bien la matriz de correlaciones en el caso de las estimaciones estandarizadas, la cual se deriva a través de

$$\Sigma(\theta) = \Sigma(\hat{\theta})$$

donde $\hat{\theta} = (\Lambda, \eta, \xi)$ es el vector de parámetros que relaciona las covarianzas entre las variables y los parámetros, si el modelo está bien ajustado, la matriz residual debe ser o más cercana posible a la matriz nula.

4.4.3.2. Diagnóstico de ajuste del modelo

Para realizar el diagnóstico de bondad de ajuste del modelo, se han diseñado varias pruebas, las cuales obedecen a motivaciones diferentes, algunas contrastan el ajuste de $\Sigma(\theta)$, algunas otras verifican la parsimonia del modelo y otras comparan el modelo ajustado contra el modelo de base saturado. Sin embargo existen diferentes situaciones que pueden alterar el resultado de las pruebas, asi que antes de escoger alguna con la que cual realizar el diagnóstico de ajuste del modelo, se realizan 1000 simulaciones por Bootstrap a

diferentes niveles de muestra (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000). En la gráfica [4.15] se observa que en particular la prueba χ^2 aumenta su estadístico a medida que crece el tamaño de la muestra, llevando a rechazar siempre para n grandes, este resultado se encuentra en línea con las investigaciones que han concluido que para muestras grandes pequeñas variaciones entre la matrices de covarianzas muestral y estimada son detectadas como significativas [6].

Con la prueba **RMR** sucede lo contrario ya que a medida que aumenta la muestra sus estimaciones decaen, encontrando su estabilidad para tamaños de muestra superiores a 2.000 casos; por su parte las pruebas **TLI**, **RFI** y **GFI**, alcanzan estabilidad en tamaños de muestra de 500 o 1000 elementos. Se muestran como las más estables las pruebas **CFI**, **RMSEA** y **PGFI** , que en su orden contrastan el ajuste de $\Sigma(\theta)$, residuos del modelo y contraste contra el modelo saturado.

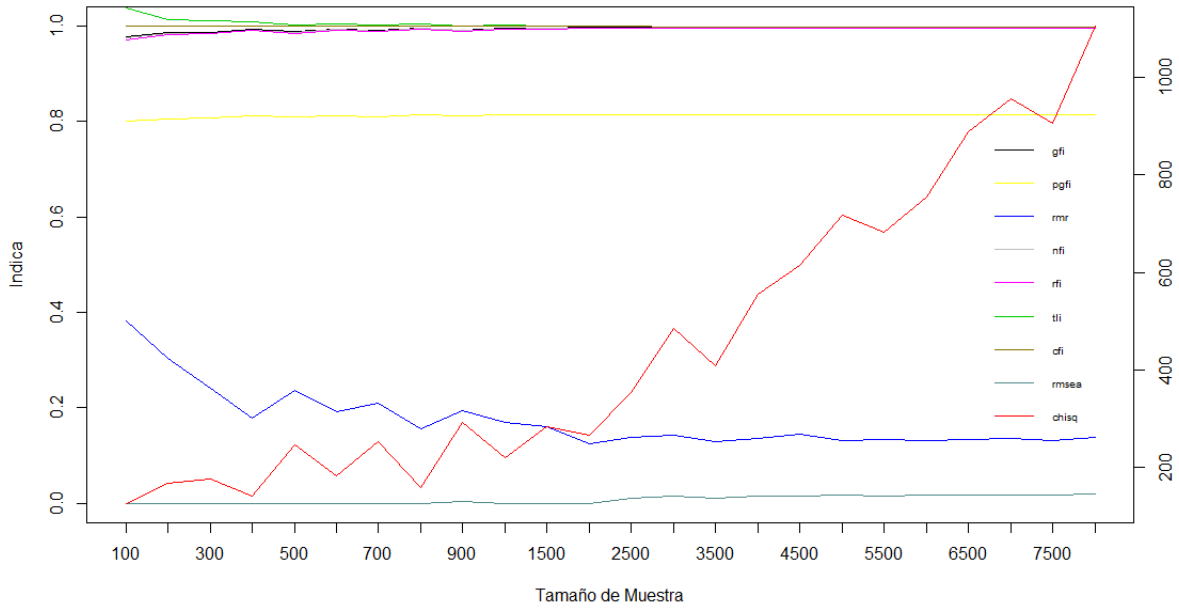


Figura 4.15: Estabilidad de ajuste a diferentes niveles de muestra

test	n=100	n=200	n=300	n=400	n=500	n=600	n=700	n=800
chisq	40,924	172,587	134,674	140,172	174,318	238,995	200,552	234,652
gfi	0,989	0,984	0,989	0,994	0,993	0,992	0,994	0,994
pgfi	0,809	0,805	0,808	0,812	0,812	0,811	0,813	0,813
rmr	0,245	0,354	0,264	0,202	0,216	0,203	0,170	0,184
nfi	0,987	0,981	0,987	0,993	0,992	0,991	0,993	0,993
rfi	0,986	0,979	0,985	0,992	0,991	0,990	0,992	0,993
tli	1,096	1,015	1,018	1,009	1,006	1,002	1,004	1,002
cfi	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
rmsea	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Tabla 4.11: Estabilidad de ajuste a diferentes niveles de muestra-1

test	n=900	n=1000	n=1500	n=2000	n=2500	n=3000	n=3500	n=4000
chisq	156,400	281,855	235,966	290,892	379,124	467,831	505,438	614,377
gfi	0,996	0,995	0,997	0,997	0,997	0,997	0,997	0,997
pgfi	0,814	0,813	0,815	0,815	0,815	0,815	0,815	0,815
rmr	0,152	0,171	0,146	0,143	0,148	0,135	0,136	0,142
nfi	0,995	0,994	0,996	0,996	0,996	0,996	0,996	0,996
rfi	0,995	0,993	0,995	0,996	0,995	0,995	0,996	0,996
tli	1,004	1,000	1,001	1,000	0,999	0,998	0,998	0,998
cfi	1,000	1,000	1,000	1,000	0,999	0,998	0,998	0,998
rmsea	0,000	0,000	0,000	0,003	0,011	0,014	0,015	0,017

Tabla 4.12: Estabilidad de ajuste a diferentes niveles de muestra-2

test	n=4500	n=5000	n=5500	n=6000	n=6500	n=7000	n=7500	n=8000
chisq	589,637	644,513	698,778	847,680	836,607	990,642	985,634	1151,096
gfi	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997
pgfi	0,815	0,815	0,816	0,815	0,815	0,815	0,815	0,815
rmr	0,133	0,136	0,125	0,140	0,133	0,139	0,135	0,138
nfi	0,997	0,997	0,997	0,996	0,997	0,996	0,997	0,996
rfi	0,996	0,996	0,997	0,996	0,996	0,996	0,996	0,996
tli	0,998	0,998	0,998	0,997	0,998	0,997	0,997	0,997
cfi	0,998	0,998	0,998	0,998	0,998	0,997	0,998	0,997
rmsea	0,015	0,016	0,016	0,018	0,017	0,019	0,018	0,019

Tabla 4.13: Estabilidad de ajuste a diferentes niveles de muestra-3

test	Valor	Valor Óptimo de Ajuste	Resultado
χ^2	228551,8671	$\approx 0,0$	Mal ajuste
cfi	0,4568	≈ 1	Mal ajuste
rmsea	0,2624	$\leq 0,1$	Mal ajuste
pgfi	0,4522	≈ 1	Mal ajuste

Tabla 4.14: Índices de Bondad de Ajuste

Haciendo un análisis de las pruebas de ajuste del modelo más estables y presentadas en el cuadro [4.14], se encuentra que el modelo inicial no presenta un ajuste adecuado de la matriz de covarianzas muestral.

4.4.4. Modificación del Modelo

4.4.4.1. Índices de Modificación

Los resultados de este modelo muestran un mal ajuste, por lo que procede será realizar las modificaciones en busca de uno mejor, estas modificaciones buscan liberar parámetros que se fijaron en cero (por ejemplo la covarianza entre dos variables). Para este fin se utiliza la prueba de multiplicadores de Lagrange (Índices de Modificación) para maximizar el ajuste del modelo; esta prueba permiten establecer las modificaciones que aportaran un mayor cambio en el estadístico de χ^2 resultantes de la liberación de estos parámetros restringidos a cero. Sin embargo, establecer estas modificaciones depende básicamente de las implicaciones que sobre la teoría del modelo conlleve dicho cambio, por lo que no todos los cambios son viables, o incluso, algún cambio puede reportar variaciones en la teoría del modelo.

Estas modificaciones se incluyen una a una y tras cada nuevo cambio se revisa de nuevo el ajuste del modelo, hasta encontrar el adecuado. El contraste entre cada modelo se realizará por medio del estadístico LR Test que contrasta la diferencia entre las log-verosimilitudes de los modelos (inicial y modificado) mediante la estadística:

$$LR = -2(l(\pi^0|y) - l(\pi^1|y)) \sim \chi^2_1$$

Las modificaciones se detienen en el momento que cada liberación de parámetros realizada no genera grandes cambios de χ^2 entre los modelos anidados (séptimo modelo en este caso), en las gráficas [4.16] y [4.17] se observa que a partir del modelo número siete, los nuevos modelo anidados no son significativamente diferentes al modelo anterior, por lo que se introduciran al modelo inicial estos 7 cambios, los cuales consisten en la liberación de los parámetros de covarianza entre los factores latentes que afectan de manera directa a la satisfacción general.

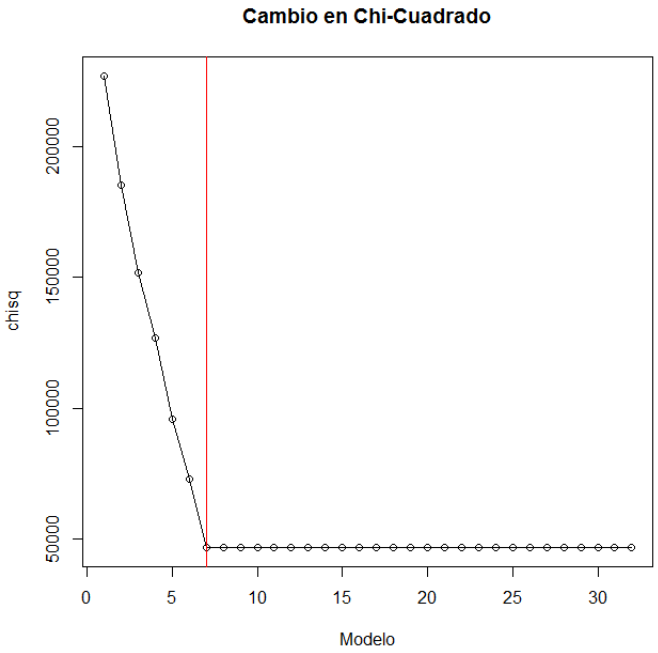


Figura 4.16: Índices de Modificación - Cambios en Chi-Cuadrado

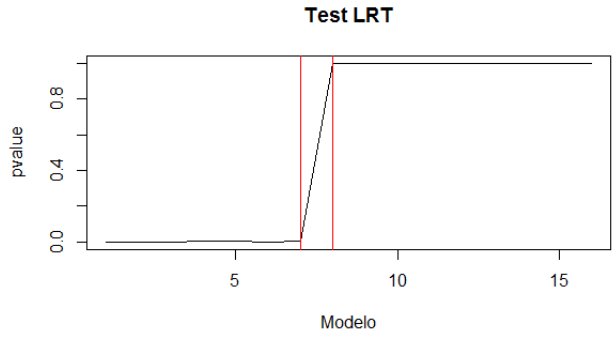


Figura 4.17: Índices de Modificación - Test LRT

4.4.5. Modelo Definitivo

Verificando los test de bondad de ajuste, se observa que finalmente se ha llegado a un buen modelo, tanto en variabilidad total explicada, como a nivel de parsimonia y residuos mínimos.

test	Valor	Valor Óptimo de Ajuste	Resultado
chisq	46935,5		
cfi	0,9	≈ 1	Buen Ajuste
rmsea	0,1	$\leq 0,1$	Buen Ajuste
pgfi	0,8	≈ 1	Buen Ajuste

Tabla 4.15: Índices de Bondad de Ajuste finales

Un vistazo a la matriz de covarianzas residual [4.16], permite observar que el modelo replica de una manera óptima las covarianzas muestrales, de manera que de esta manera se puede entender la mejora de las estadísticas de bondad de ajuste de **CFI** y **RMSEA**.

Las estimaciones de los parámetros permiten observar que las dimensiones de trato y asesores son las que más efecto tienen sobre la satisfacción general de los clientes, a la vez dicha satisfacción afecta en mayor medida la recomendación lo que se ve reflejado en el coeficiente que tienen esta variable sobre la respuesta de Lealtad. Así mismo que observa que la mayor covarianza se encuentra entre los factores oficinas y cajas, por lo que un posible ajuste del modelo que contemple estos dos factores como uno solo podría también tomarse en cuenta.

Las estimaciones de los parámetros se muestran en la gráfica [4.18]:

Tabla 4.16: Tabla de Covarianzas Residuales

antigu	comprt	nuevo	cupos	acompa	requst	trami	claro	inters	ag	soluc	respet	espera	tiempo	persnl	agili	manejo	senalz	segrdd	horari	comddd	ubiccn	amabi	plnt_f	finncr	atencn	varidd	leal	intenc	recom	
antigue	0,0																													
comporta	1,0	0,0																												
nuevo	0,2	0,2	0,0																											
cupos	0,1	0,1	0,2	0,0																										
acompa	0,1	0,3	0,4	-0,3	0,0																									
requisit	-0,2	-0,3	-0,3	0,4	-0,5	0,0																								
trami	-0,2	-0,3	-0,3	0,2	-0,5	0,4	0,0																							
claro	-0,1	-0,1	0,0	-0,1	0,0	0,0	0,1	0,0																						
interes	0,0	0,0	0,1	0,0	0,1	0,0	0,1	0,2	0,0																					
ag	-0,1	-0,1	-0,1	-0,1	0,1	0,0	0,1	0,2	0,1	0,0																				
soluc	0,0	0,0	0,1	0,0	0,1	0,0	0,1	0,3	0,2	0,3	0,0																			
respet	-0,1	-0,1	-0,1	-0,1	-0,1	0,0	0,0	0,2	0,2	0,1	0,0	0,0																		
espera	0,1	0,1	0,1	0,0	0,2	0,0	0,1	-0,3	-0,3	-0,2	-0,3	-0,2	0,0																	
tiempo	0,1	0,0	0,0	0,0	0,1	0,1	0,0	-0,2	-0,2	-0,1	-0,1	-0,2	0,7	0,0																
personal	0,1	0,0	0,1	0,0	0,1	0,1	0,0	-0,1	-0,1	-0,1	-0,1	-0,1	0,4	0,3	0,0															
agili	-0,1	-0,1	-0,1	-0,1	-0,1	0,0	0,0	0,0	0,0	0,1	0,0	0,1	0,2	0,0	0,0	0,0														
manejo	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-0,1	-0,1	0,0	-0,1	0,0	0,3	0,1	-0,1	-0,2	0,0													
senaliza	-0,1	0,0	0,0	0,0	0,0	0,1	0,1	0,0	0,0	0,0	-0,1	0,1	0,0	-0,1	-0,1	0,0	0,1	0,0												
seguridad	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	0,0	0,0	-0,1	0,0	0,0	-0,1	-0,1	0,0	0,0	0,1	0,0											
horario	0,0	0,0	0,1	0,0	0,0	0,1	0,1	0,0	-0,1	0,0	-0,1	0,0	0,0	-0,1	0,0	0,0	0,0	0,0	0,0	0,0										
comodidad	0,0	0,0	0,0	0,0	0,1	0,0	0,0	0,0	-0,1	0,0	-0,1	0,0	0,1	0,1	0,0	0,0	0,1	0,1	0,0	0,0	0,0									
ubicacion	-0,1	0,0	0,0	0,0	0,1	0,1	0,1	0,0	0,0	0,0	0,0	0,1	-0,1	-0,1	-0,1	-0,1	-0,1	0,0	0,1	0,3	0,0	0,0								
amabi	-0,2	-0,1	-0,1	-0,1	-0,1	0,0	0,0	0,2	0,1	0,1	0,0	0,3	0,0	0,0	0,1	0,3	0,2	-0,1	0,0	-0,1	-0,2	-0,1	0,0							
planta_fisica	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	0,0	-0,1	-0,1	-0,1	0,0	0,0	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,2	0,0	0,0	0,0						
financiera	-0,3	-0,3	-0,3	-0,2	-0,3	-0,1	-0,2	-0,1	-0,1	-0,1	-0,2	0,0	-0,1	0,5	0,5	0,7	0,4	0,1	0,1	0,1	0,1	0,0	0,4	0,1	0,0					
atencion	-0,2	-0,2	-0,2	-0,2	-0,1	-0,1	-0,1	0,5	0,5	0,5	0,5	0,4	0,2	-0,2	-0,1	0,0	0,0	-0,1	0,0	-0,1	-0,1	-0,1	0,1	0,1	0,0	0,0				
variedad	0,5	0,5	0,8	0,7	0,5	0,6	0,5	-0,2	-0,1	-0,2	-0,1	-0,1	-0,3	-0,2	-0,2	-0,2	-0,2	-0,1	-0,1	-0,1	-0,1	0,0	-0,2	-0,1	-0,2	-0,2	0,0			
leal	0,2	0,2	0,2	0,1	0,2	0,1	0,1	0,1	0,0	0,0	0,2	-0,1	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	-0,1	0,0	-0,1	0,1	-0,2	0,3	-0,1	0,0	0,3	0,0		
intenc	3,0	2,9	2,8	2,6	3,1	2,4	2,4	1,9	2,2	2,1	2,4	1,4	2,3	2,3	2,2	1,8	2,0	1,5	1,4	1,5	1,8	1,4	1,2	2,3	1,6	2,0	2,4	0,0	0,0	
recom	3,1	2,9	2,8	2,6	3,1	2,4	2,4	2,1	2,3	2,2	2,5	1,4	2,4	2,3	2,3	1,9	2,0	1,7	1,5	1,6	1,9	1,5	1,4	2,5	1,7	2,1	2,5	0,0	0,0	0,0

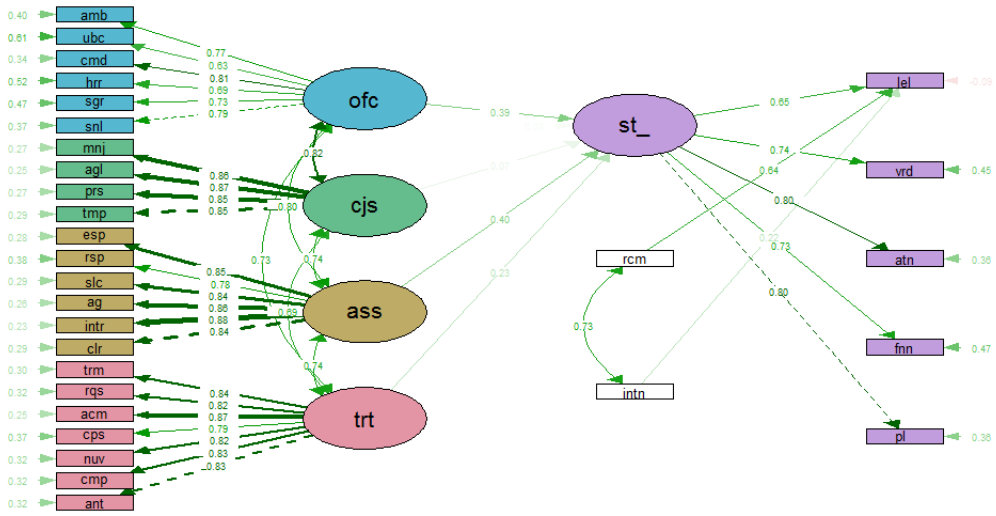


Figura 4.18: Path Diagram Modelo Ajustado

Variable/Factor	Estimate	Std.err	Z-value	P(>|z|)	Std.lv	Std.all
Regressions:						
sat_gene ~						
oficinas	0,442	0,041	10,908	0	0,388	0,388
cajas	0,05	0,019	2,586	0,01	0,07	0,07
asesores	0,395	0,025	15,796	0	0,395	0,395
trat	0,149	0,01	14,335	0	0,227	0,227
leal ~						
intenc	0,185	0,053	3,516	0	0,185	0,217
recom	0,611	0,062	9,822	0	0,611	0,645
sat_gene	0,957	0,009	105,589	0	1,376	0,652

Tabla 4.17: Estimaciones Modelo Ajustado

Variable/Factor	Estimate	Std.err	Z-value	P(>|z|)	Std.lv	Std.all
Latent variables:						
trat =~						
antigue	1				2,195	0,826
comporta	0,953	0,008	122,828	0	2,092	0,827
nuevo	0,933	0,008	123,158	0	2,048	0,822
cupos	0,845	0,007	118,678	0	1,855	0,792
acompa	1,002	0,008	125,176	0	2,199	0,868
requisit	0,783	0,007	119,869	0	1,72	0,825
trami	0,78	0,007	119,038	0	1,713	0,835
asesores =~						
claro	1				1,438	0,841
interes	1,151	0,011	108,845	0	1,655	0,877
ag	1,118	0,01	109,03	0	1,607	0,862
soluc	1,161	0,011	108,702	0	1,67	0,841
respet	0,757	0,007	101,08	0	1,088	0,785
espera	1,3	0,012	112,387	0	1,869	0,85
cajas =~						
tiempo	1				1,995	0,845
personal	0,975	0,008	123,324	0	1,944	0,854
agili	0,843	0,007	118,633	0	1,682	0,866
manejo	0,904	0,008	120,148	0	1,803	0,857
oficinas =~						
senaliza	1				1,26	0,793
seguridad	0,895	0,009	102,826	0	1,127	0,729
horario	0,921	0,009	102,431	0	1,16	0,692
comodidad	1,15	0,011	107,947	0	1,449	0,812
ubicacion	0,81	0,008	99,617	0	1,021	0,625
amabi	0,869	0,009	101,832	0	1,095	0,773
sat_gene =~						
planta_fisica	1				1,438	0,801
financiera	0,865	0,008	109,323	0	1,244	0,731
atencion	0,967	0,009	112,364	0	1,39	0,801
variedad	1,004	0,009	112,53	0	1,444	0,745

Tabla 4.18: Estimaciones Modelo Ajustado (cont.)

Variable/Factor	Estimate	Std.err	Z-value	P(>|z|)	Std.lv	Std.all
Covariances:						
trat ~~						
asesores	2,329	0,022	106,305	0	0,738	0,738
oficinas	2,012	0,019	103,253	0	0,728	0,728
cajas	3,008	0,028	108,676	0	0,687	0,687
asesores ~~						
oficinas	1,442	0,015	94,379	0	0,796	0,796
cajas	2,109	0,021	99,596	0	0,736	0,736
cajas ~~						
oficinas	2,07	0,021	98,007	0	0,824	0,824
intenc ~~						
recom	4,03	0,103	39,127	0	4,03	0,729

Tabla 4.19: Estimaciones Modelo Ajustado (cont.)

Los resultados presentados en las tablas [4.17], [4.18] y [4.19], permiten comprobar el sistema de hipótesis, concluyendo que los factores subyacentes especificados en el modelo tienen un efecto significativo sobre la satisfacción, así como también son significativas cada una de las variables que afectan la lealtad de un cliente con la marca. En el modelo final también se observa la relación de covariación entre los factores componentes del servicio, lo que quiere decir que su estudio no se debe hacer de manera aislada como se haría en cualquier modelo de regresión multivariada, sino que se deben tener en cuenta las correlaciones existentes entre las diferentes dimensiones del servicio ya que unas afectan a las otras.

La expresión matemática del modelo inicial de medida es la siguiente:

$$\mathbf{X} = \Lambda\eta + \tau$$

$$\begin{bmatrix} antigue \\ comporta \\ nuevo \\ cupos \\ acompa \\ requisit \\ trami \\ claro \\ interes \\ ag \\ soluc \\ respet \\ espera \\ tiempo \\ personal \\ agili \\ manejo \\ senaliza \\ seguridad \\ horario \\ comodidad \\ ubicacion \\ amabi \end{bmatrix} = \begin{bmatrix} 0,826 & 0 & 0 & 0 \\ 0,827 & 0 & 0 & 0 \\ 0,822 & 0 & 0 & 0 \\ 0,792 & 0 & 0 & 0 \\ 0,868 & 0 & 0 & 0 \\ 0,825 & 0 & 0 & 0 \\ 0,835 & 0 & 0 & 0 \\ 0 & 0,841 & 0 & 0 \\ 0 & 0,877 & 0 & 0 \\ 0 & 0,862 & 0 & 0 \\ 0 & 0,841 & 0 & 0 \\ 0 & 0,785 & 0 & 0 \\ 0 & 0,85 & 0 & 0 \\ 0 & 0 & 0,845 & 0 \\ 0 & 0 & 0,854 & 0 \\ 0 & 0 & 0,866 & 0 \\ 0 & 0 & 0,857 & 0 \\ 0 & 0 & 0 & 0,793 \\ 0 & 0 & 0 & 0,729 \\ 0 & 0 & 0 & 0,692 \\ 0 & 0 & 0 & 0,812 \\ 0 & 0 & 0 & 0,625 \\ 0 & 0 & 0 & 0,773 \end{bmatrix} \begin{bmatrix} \eta_{trato} \\ \eta_{asesores} \\ \eta_{cajas} \\ \eta_{oficinas} \end{bmatrix} + \begin{bmatrix} 0,317 \\ 0,317 \\ 0,325 \\ 0,373 \\ 0,247 \\ 0,32 \\ 0,302 \\ 0,294 \\ 0,232 \\ 0,258 \\ 0,292 \\ 0,384 \\ 0,277 \\ 0,285 \\ 0,27 \\ 0,25 \\ 0,265 \\ 0,372 \\ 0,469 \\ 0,521 \\ 0,341 \\ 0,609 \\ 0,402 \end{bmatrix}$$

En tanto que el modelo estructural:

$$[\xi_{satisfacción}] = [0,369 \quad 0,292 \quad -0,098 \quad 0,231] \begin{bmatrix} \eta_{trato} \\ \eta_{asesores} \\ \eta_{cajas} \\ \eta_{oficinas} \end{bmatrix} + [0,025]$$

$$Leal = 0,65 * satisfacción + 0,64 * recomendación + 0,22 * intención - 0.093$$

4.4.6. Impacto de las componentes del servicio sobre la lealtad

La estimación del impacto que tiene cada una de las dimensiones del servicio sobre la lealtad del cliente, se hace uso del llamado efecto de la variable (Directo, Indirecto y Total), el cual se calcula como el producto entre la varianza de la variable de partida y los parámetros asociados a las flechas recorridas hasta unir las dos variables de interés.

variable	estima	ef.indirecto	ef.directo	total
oficinas	0,388	0,252976	0	0,252976
cajas	0,07	0,04564	0	0,04564
asesores	0,395	0,25754	0	0,25754
trat	0,227	0,148004	0	0,148004
sat_{gen}	0,652	0	0,652	0,652
intenc	0,217	0	0,217	0,217
recom	0,645	0	0,645	0,645

Tabla 4.20: Efecto Final sobre la Lealtad del Cliente

En la tabla [4.20] este se puede observar que dentro de las dimensiones de satisfacción las de mayor efecto son los asesores y las oficinas, en tanto que las cajas tienen el menor impacto en la satisfacción. A nivel global el mayor efecto sobre la lealtad lo tiene la recomendación.

Comparando a través del gráfico [4.19] el efecto de las variables en el modelo de medida contra la satisfacción declarada inicialmente, se obtiene una mejor lectura de la situación, se puede apreciar cuáles son aquellas variables que tienen un alto impacto en la lealtad y a la vez tienen baja satisfacción declarada, por ejemplo, el tiempo de espera de una persona en asesoría tiene un alto impacto en la lealtad del cliente y sin embargo presenta una valoración inferior al promedio, por lo que en este caso la empresa debería poner especial esfuerzo en mejorar este indicador.

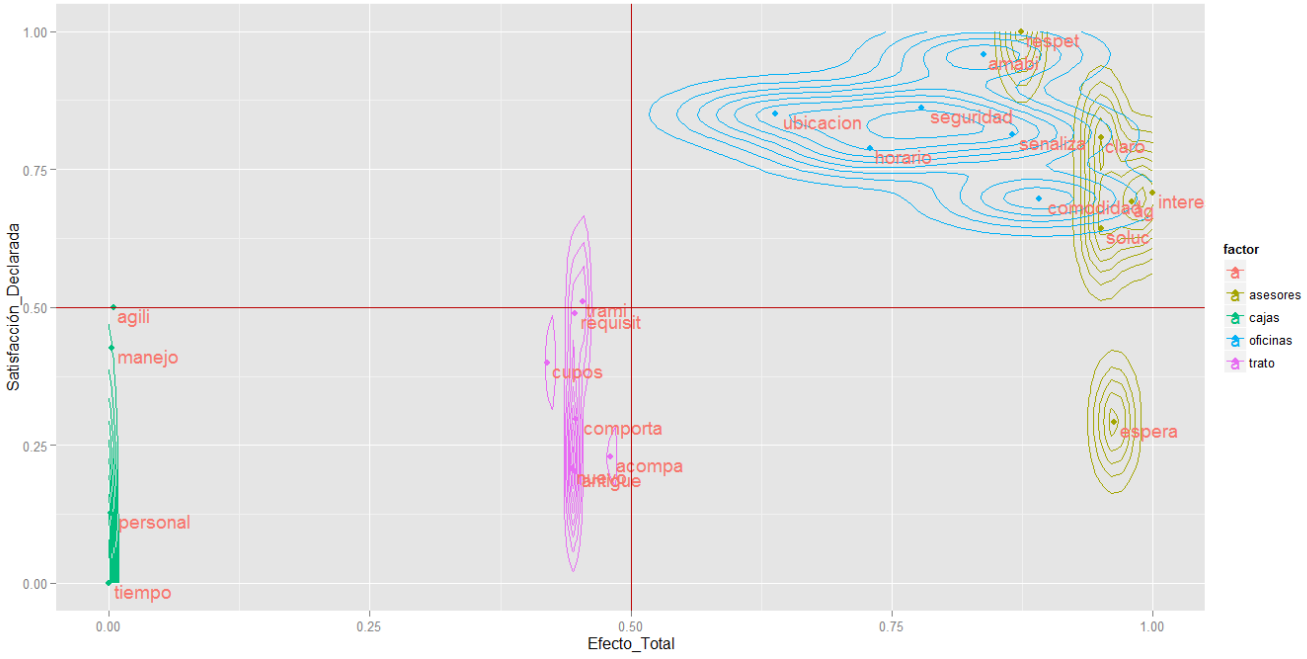


Figura 4.19: Mapa de Acción

4.4.7. Cálculo del Indicador de Lealtad

Diferentes enfoques se pueden utilizar para el cálculo del indicador, sin embargo, el propósito de este trabajo es extraer toda la información derivada como base para el cálculo, se propone realizar el cálculo del indicador de lealtad usando el efecto total de cada una de las componentes del servicio como pesos que ponderan el indicador de tal manera que

$$Lealtad = \sum_{i=1}^{3} w_i v \tag{4.1}$$

donde w_i es un vector de ponderaciones construidos con los efectos totales de las variables $recom$, $intenc$ y el factor sat_{gen}, tal que

$$w_i = \frac{\beta_i}{\sum_{i=1}^{3} \beta_i} \tag{4.2}$$

$$\sum_{i=1}^{3} w_i = 1 \tag{4.3}$$

y v es una matriz conformada por las variables de satisfacción general (sat_{gen}), recomendación ($recom$) e intención de recompra ($intenc$).

Entonces la expresión matemática del Indicador de Lealtad sería:

$$Lealtad = w_1 * sat_{gen} + w_2 * recom + w_3 * intenc \tag{4.4}$$

$$Lealtad = 0,431 * sat_{gen} + 0,143 * recom + 0,426 * intenc \tag{4.5}$$

Capítulo 5

Conclusiones

Las conclusiones son un compendio de lo que ya se mencionó a lo largo de este trabajo, sin embargo sintetizando y de acuerdo a los objetivos del trabajo se tiene que:

- El mejor método de estimación se encuentra que la función de ajuste por mínimos cuadrados ponderados diagonalizados (DWLS) es la que mejor se ajusta a los datos, ya que es una función libre de distribución y adicionalmente utiliza la matriz de correlaciones policóricas para variables ordinales en la etapa de estimación.

- En su orden la satisfacción general con el servicio y la recomendación del mismo, son las variables que mayor efecto tienen sobre la lealtad del cliente. Se puede observar igualmente que son los constructos del servicio relacionados a los asesores y las oficinas los que mayor efecto tienen sobre la satisfacción general con el servicio.

- La Matriz de Acción permite observar los atributos que componen los factores de asesores y oficinas son los mejor valorados y a la vez son los que mayor efecto tienen sobre la lealtad, sin embargo el atributo *espera* tiene una calificación inferior al promedio por lo que con miras a mejorar la lealtad se debe apuntar a subir su calificación promedio. Los atributos componentes del factor cajas presentan en general una satisfacción inferior al promedio y son los que menos efecto tienen sobre la lealtad.

  Adicionalmente durante el desarrollo del presente trabajo surgieron otras inquietudes como por ejemplo el comportamiento de los indicadores de ajuste del modelo y como mejorar el ajuste del mismo, al respecto se puede concluir que:

- En los estudios exploratorios generalmente se analiza la manera en que se correlacionan las variables ordinales para formar factores latentes, en estas situaciones se debe utilizar la matriz de correlaciones policóricas, sin embargo la mayoria de paquetes comerciales aún no ha difundido el uso de éstas por lo que la realización de un Análisis Factorial Exploratorio podría ser cuestionado, en este trabajo se concluye que para escalas tipo Likert de 10 puntos la matriz de correlaciones policóricas es muy cercana a la matriz de correlación de pearson, tanto en el valor de la correlación

como en su comportamiento, por lo que en estos casos es válido utilizar la matriz de pearson para realizar el Análisis Factorial en variables ordinales.

- Como estadística de diagnóstico para el ajuste del modelo se encontró que no obstante ser la más popular, la medida χ^2 no es robusta ya que se ve afectada de manera directa por el tamaño de la muestra aumentando su valor a medida que aumenta n. En lo relacionado con las medidas de RMR, GFI, NFI, RFI y TLI, encuentran estabilidad en muestras superiores a 1.000 casos, por lo que en muchos casos no va a ser permisivo usarlas dados los costos de levantar muestras grandes. Cabe aclarar que este análisis se realizó sobre el método de estimación DWLS, por lo que en otros métodos de estimación este comportamiento se desconoce.

- Sobre la parte práctica se debe ser riguroso con el análisis descriptivo y validación de supuestos de los datos, ya que aunque no se demuestra directamente en este trabajo, analizando las funciones de ajuste se puede entender fácilmente que los valores de los parámetros cambian dependiendo de la función escogida, no obstante también se debe tener en cuenta el tamaño de la muestra, ya que los métodos de distribución libre requieren gran cantidad de muestra para poder realizar las estimaciones.

Capítulo 6

Futúras líneas de investigación

Como futuras líneas de investigación se puede abordar el estudio de los valores iniciales para los parámetros restringidos que sirven de pivote para la estimación, por ejemplo, se podría utilizar la comunalidad de la variable que mayor contribución tiene a la creación del factor como valor inicial en la etapa de estimación.

Otra línea de investigación podría ser el estudio de la matriz de pesos **W**, controlando por ejemplo los grados de *miss-especification* de esta matriz y analizando su repercusión sobre la estimación de los parámetros.

Una posible limitación de este trabajo es la característica líneal del modelo, así que se podría estudiar de manera comparativa los resultados que pueden ofrecer técnicas como los modelos de Ecuaciones Estructurales Difusos (FSM) y llegar a conocer su aporte a la especificación del modelo.

Bibliografía

[1] Aish, A.M., Joreskog, K.G., *Apanel model for political efficacy and responsiveness: an application of LISREL 7 with weighted least squares* Qual Quant. 24, 1990.

[2] Anderson, James C. Gerbing, David W., *Structural equation modeling in practice* Psychological Bulletin, 1988.

[3] Anderson, James C. Gerbing, David W., *Monte Carlo evaluations of goodness of fit indices for structural equation models* Sage Periodical Press, 1992.

[4] Anderson, James C. Gerbing, David W., *The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis* Psychometrica, 1984.

[5] Batista F. Joan M., Coenders G. Germá, *Modelos de Ecuaciones Estructurales.* Editorial La Muralla, 2000.

[6] Bentler, P. M. Bonett, D. G., *Significance test and goodness of fit in the analysis of covariance structures* Psychological Bulletin, 1980.

[7] Bollen, Kenneth A., *Structural equations with latent variables* Psychological Bulletin, 1980.

[8] Browne, M.W., *Asymptotically distribution-free methods for the analysis of covariance structures* British Journal of Mathematical and Statistical Psychology, 1984.

[9] Byrne Barbara, *Structural equation modeling with AMOS* Routledge, 2009.

[10] Chan Wai, *Comparing indirect effects in SEM- A sequential model fitting method using covariance equivalent specifications* Structural Equation Modeling: A Multidisciplinary Journal, 2007.

[11] Fox J., *Effect analysis in structural equation models II- Calculation of specific indirect effects* Sociological Methods & Research, 1985.

[12] Gelin, M.N., Beasley, T.M., y Zumbo, B.D., *What is the impact on scale reliability and exploratory factor analysis of a Pearson correlation matrix when some respondents are not able to follow the rating scale?*. Annual meeting of the American Educational Research Association (AERA), 2003.

[13] Germà Coenders , Albert Satorra & Willem E. Saris, *Alternative approaches to structural modeling of ordinal data: A Monte Carlo study.* Structural Equation Modeling: A Multidisciplinary Journal, 4:4, 1997.

[14] Hancock Gregopry & Mueller Ralph, *Structural Equation Modeling a second course* IAP, 2006.

[15] Heuchenne, Christian, *A sufficient rule for identification in structural equation modeling icluding the null B and recursive rules as extreme cases* Structural Equation Modeling: A Multidisciplinary Journal, 2009.

[16] Holger Brandt, Augustin Kelava & Andreas Klein, *A Simulation Study Comparing Recent Approaches for the Estimation of Nonlinear Effects in SEM Under the Condition of Nonnormality* Structural Equation Modeling: A Multidisciplinary Journal, 2014.

[17] Hoyle, Rick H., *Handbook of Structural Equation Modeling.* The Guilford Press, 2012.

[18] Field, A., *Discovering Statistics using SPSS for Windows.* Sage publications, 2000.

[19] Kenneth A. Bollen & Walter R. Davis, *Two Rules of Identification for Structural Equation Models.* Structural Equation Modeling: A Multidisciplinary Journal, 2009.

[20] Jöreskog, K.G., Sörbom, D., *LISREL 7: A Guide to the Program and Applications* SPSS Publications, 1989.

[21] Kenneth A. Bollen & Walter R. Davis, *Two Rules of Identification for Structural Equation Model* Structural Equation Modeling: A Multidisciplinary Journal, 2009.

[22] Kline Rex B., *Principles and practices of Structural Equation Modeling* The Guilford Press, 2004.

[23] Loehlin, John C., *Latent Variable Models* Laurence Erlbaum Associates Publishers, 2004.

[24] Lyhagen,J., Ornstein, P., *A Rank Based estimator of the Polychoric Correlation* Uppsala University, 2011.

[25] Marcoulides George A., Schumacker Randall E., *New developments and techniques in structural equation modeling* Lawrence Erlbaum Associates, 2001.

[26] Muthén, B., *A general structural equationmodel with dichotomous, ordered categorical, and continuous latent variable indicators* Psychometrika 49, 1984.

[27] Olsson, Ulf Henning, *Maximum likelihood estimation of the polychoric correlation coefficient.* Psychometrika, Volume 44, 1979.

[28] Olsson, Ulf Henning, Foss Tron, Troye Sigurd V. & Howell Roy D., *The Performance of ML, GLS, and WLS Estimation in Structural Equation Modeling Under Conditions of Misspecification and Nonnormality* Structural Equation Modeling: A Multidisciplinary Journal, 2009.

[29] Ory David, Mokhtarian Patricia, *The impact of non-normality,sample size and estimation technique on goodness-of-t measures in structural equation modeling: evidence from ten empirical models of travel behavior* Springer Science+Business Media, 2009.

[30] Pui-Wa Lei, *Evaluating estimation methods for ordinal data in structural equation modeling* Springer Science + Business Media, 2007.

[31] Pui-Wa Lei, Qiong Wu, *Introduction to structural equation modeling- Issues and practical considerations* Pennsylvania State University, 2007.

[32] Raudenbush S. W., Sampson R., *Assessing direct and indirect effects in multilevel designs with latent variables* Sociological Methods & Research, 1999.

[33] Raykov Tenko & Marcoulides George, *A first course in Structural Equation Modeling* Lawrence Erlbaum Associates, 2004.

[34] Revelle, W. & Rocklin, T., *Very Simple Structure: an Alternative Procedure for Estimating the Optimal Number of Interpretable Factors.* Multivariate Behavioral Research, 1979.

[35] Rietveld, T. & Van Hout, R., *Statistical Techniques for the Study of Language and Language Behaviour.* De Gruiter Mouton, 1993.

[36] Rigdon Edward E., *Calculating degrees of freedom for a structural equation modeling* Lawrence Erlbaum Associates, 1994.

[37] Rosseel Yves, *lavaan: an R package for structural equation modeling and more Version 0.3-1 (BETA)* Ghent University, 2010.

[38] Rosseel Yves, *The lavaan tutorial* Ghent University, 2013.

[39] Satorra Albert, *Robustness issues in structural equation modeling* Kluwer Academic Publishers, 1990.

[40] Saurina C., *Evaluación de un modelo de medida de la calidad en el sector servicios.* Estadística Española, 1997.

[41] Şimşek Gülhayat G. & Noyan Fatma, *Structural equation modeling with ordinal variables: a large sample case study* Yildiz Technical University, 2012.

[42] Skrondal Anders, Rabe-Hesketh Sophia, *Generalized latent variable modeling* Chapman & Hall/CRC, 2004.

[43] Sörbom,Dag, *Model Modification* Uppsala University, 1989.

[44] Streiner D., *Starting at the beginning: an introduction to coefficient alpha and internal consistency.* Journal of personality assessment, 2003.

[45] Xitao Fan , Bruce Thompson & Lin Wang, *Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes* Structural Equation Modeling: A Multidisciplinary Journal, 1999.