



UNIVERSIDAD DE  
COSTA RICA

**iICE**

Instituto de Investigaciones en  
Ciencias Económicas

ISSN 2215-2237

# Serie de Divulgación Económica

---

---

**IDENTIFICACIÓN Y USO DE VARIABLES DE LA  
ENCUESTA DE CONFIANZA DEL  
CONSUMIDOR COMO INDICADORES ADELANTADOS  
DE VARIABLES/AGREGADOS MACROECONÓMICOS**

Alvaro Castro Quesada

330.015.195

C355i Castro Quesada, Álvaro Luis.

Identificación y uso de variables de la encuesta de confianza del consumidor como indicadores adelantados de variables/agregados macroeconómicos / Álvaro Castro Quesada. – San José, C.R. : Universidad de Costa Rica, IICE, 2014.

43 p. : il.-- (Serie divulgación económica; ISSN 2215-2237, IICE-26)

ISBN 978-9968-824-30-9

1. INDICADORES ECONOMICOS – EVALUACION. 2. INDICADORES ECONOMICOS – METODOS ESTADISTICOS. 3. COMPORTAMIENTO DEL CONSUMIDOR. 4. VARIABLES (ESTADISTICA). I. Título. II. Serie.

CIP/2721

CC/SIBDI, UCR

Universidad de Costa Rica

© Instituto de Investigaciones en Ciencias Económicas (IICE)

Ciudad Univeritaria “Rodrigo Facio”, San José Costa Rica.

Prohibida la reproducción total o parcial. Todos los derechos reservados. Hecho el depósito de ley.

La diagramación de este documento estuvo a cargo del IICE.

**IDENTIFICACIÓN Y USO DE VARIABLES DE LA  
ENCUESTA DE CONFIANZA DEL  
CONSUMIDOR COMO INDICADORES ADELANTADOS  
DE VARIABLES/AGREGADOS MACROECONÓMICOS**

Alvaro Castro Quesada



# ÍNDICE

---

---

RESUMEN	1
1. INTRODUCCIÓN	3
<b>2. Observaciones al concepto del icc como indicador adelantado del IMAE</b>	<b>5</b>
Indicador Adelantado Compuesto	5
Indicador del Sentimiento del Consumidor de la Universidad de Michigan	6
<b>3. Metodología de Stock y Watson</b>	<b>7</b>
<b>4. Resultados</b>	<b>9</b>
<b>5. Análisis de Factores</b>	<b>23</b>
Número de Factores	23
Resultados	25
<b>6. Sugerencias</b>	<b>29</b>
Sugerencia Principal : Indicador Adelantado Compuesto de OECD	29
Otras Sugerencias	31
NOTAS	31
REFERENCIAS LITERARIAS	33
APENDICES METODOLOGICOS	35
Modelos Arima	35
Análisis factorial	41



# ÍNDICE DE CUADROS, GRÁFICOS Y RECUADROS

---

---

<b>Cuadro 1:</b> Variación en AIC y BIC al introducir componentes AR de icc, ica e iee	11
<b>Cuadro 2:</b> AR de la serie objeto	13
<b>Cuadro 3:</b> AR de la serie + AR del icc	14
<b>Cuadro 4:</b> AR de la serie + AR del ica	16
<b>Cuadro 5:</b> AR de la serie + AR del iee	18
<b>Cuadro 6:</b> Gráficos de diagnóstico AR (Transporte) +AR (iee)	21
<b>Cuadro 7:</b> Valores Propios de la matriz de correlación de las primeras diferencias	24
<b>Cuadro 8:</b> Cargas Factoriales Estandarizadas	25
<b>Cuadro 9:</b> Matriz de correlación Phi de factores	26
<b>Cuadro 10:</b> Comunalidades de cada variable	27
<b>Gráficos de Diagnóstico</b>	20
<b>Recuadro No.1</b> Prueba de Cambio Estructural para el modelo del sector Transporte <i>Prueba supF de Andrews para AR (transporte, 3) + AR(iee,3)</i>	21
<b>Recuadro No.2</b> Prueba de Cambio Estructural para el modelo del sector Comercio <i>Prueba supF de Andrews para AR (comercio, 2) + AR (iee,3)</i>	22
Valores Propios de la matriz de correlación de las primeras diferencias de las series	24



## RESUMEN

---

---

---

Se identifican el sector de Transporte, Almacenamiento y Comunicaciones, y el sector de Comercio del IMAE, como los más promisorios para emplear el *iee*, índice de expectativas económicas, como variable componente de un futuro Indicador Adelantado Compuesto de varias variables de dichos sectores. En ese sentido, el *iee* exhibe mejor desempeño que el *icc*, índice de confianza del consumidor.

Ninguno de los índices (*icc*, *icea*, *iee*) mostró un buen ajuste de los puntos de giro presentes en las series del IMAE o sus sectores.

La aplicación de un análisis de factores -en el contexto de series de tiempo- permitió entrever aspectos de la dinámica subyacente a las variables de la encuesta. El *icc* aparece relacionado más directamente con las expectativas de evolución en el corto plazo de variables socio-económicas y solo en forma indirecta -a través de la correlación entre el primer y segundo factor- se relaciona con aspectos económicos del ámbito familiar.



# 1. INTRODUCCIÓN

---

---

---

El interés principal es evaluar el *icc* como indicador adelantado de la actividad económica.

En un principio se consideró utilizar el PIB Trimestral como variable para operacionalizar el concepto de actividad económica, sin embargo, el número de observaciones trimestrales de la encuesta no resultó suficiente. Se optó entonces por mensualizar la serie del *icc*.

La mensualización se hizo mediante el paquete Temporal Disaggregation (tempdisagg) del software R. Específicamente, se utilizó el método Denton-Cholette.

Con datos mensuales el concepto de actividad económica se operacionalizó con la variable Índice Mensual de Actividad Económica (IMAE). Se utilizó la Tasa de Variación Media del IMAE. (La Oficina del Censo de Estados Unidos, en la documentación de sus programas X12-Arima y X13-Arima-Seats, no recomienda utilizar tasas de Variación Interanual, ver <http://www.census.gov/srd/www/sapaper/sapaperbytopic.html>)

Para evitar que factores estacionales y/o irregulares influyan en la relación entre el *icc* y el IMAE, se trabajó con el componente de Tendencia-Ciclo de cada serie.

En el caso de la Tasa de Variación Media del IMAE se empleó el componente Tendencia-Ciclo calculado por el Banco Central de Costa Rica.

En el caso del *icc* se obtuvo su Descomposición en Componentes (componente Tendencia-Ciclo, componente Estacional y componente Irregular) mediante el paquete Seasonal-Trend Decomposition (stl) del software R. Este paquete implementa el método no paramétrico LOESS (Cleveland W. S., Devlin S. J., 1988)

Así entonces, se pretende evaluar el componente de Tendencia-Ciclo del *icc* como indicador adelantado del componente Tendencia-Ciclo del IMAE.



## **2. OBSERVACIONES AL CONCEPTO DEL CC COMO INDICADOR ADELANTADO DEL IMAE**

---

---

### **Indicador Adelantado Compuesto**

El empleo de una sola variable como indicador adelantado, especialmente como indicador adelantado de agregados macroeconómicos, ha sido sustituido por el empleo de varias variables. Por lo general se busca incluir variables que reaccionen a distintos sectores de la economía. Es común incluir variables del sector real, del sector monetario, del sector externo, etc.

Un promedio ponderado, en sentido amplio, de éstas variables se utiliza como índice adelantado. A este índice se le ha llamado Indicador Adelantado Compuesto.

El uso de un Indicador Adelantado Compuesto está fundamentado en el hecho observado de que cada recesión es distinta, en el sentido de que es causada o iniciada por diferentes acontecimientos y/o eventos; por ejemplo, una recesión puede tener orígenes externos mientras que la siguiente recesión puede tener su origen en el mercado local. Por consiguiente, los sectores que reaccionen en forma temprana serán diferentes en cada recesión. Siendo así, es muy improbable que una sola variable sea útil como indicador adelantado.

Con la popularización de las Encuestas de Confianza (tanto dirigidas al consumidor como al empresario), se ha vuelto común incluir también en el Indicador Adelantado Compuesto variables provenientes de dichas encuestas. De esta forma se busca incluir en el Indicador Adelantado Compuesto el cambio en las expectativas, tanto del consumidor como empresariales.

La OECD es uno de los organismos con más experiencia en la construcción de Indicadores Adelantados Compuestos y su metodología (OECD, Abril 2012) ha sido aplicada en varios países de Latinoamérica (Yabuta 2010, Pedersen 2009)

En los casos latinoamericanos se ha encontrado la dificultad de que las series de datos constituidas por las variables provenientes de las Encuestas de Confianza son mucho más cortas que las series de datos de agregados macroeconómicos. En algunos casos ésta situación ha impedido su uso (Yabuta 2009)

Otro ejemplo de utilización de Indicador Adelantado Compuesto lo constituye la Oficina de Análisis Económico del Departamento Federal de Comercio de Estados Unidos (Moylan 2010)

## Indicador del Sentimiento del Consumidor de la Universidad de Michigan

La metodología del *icc* sigue muy de cerca la metodología del Indicador del Sentimiento del Consumidor de la Universidad de Michigan o ICS (Index of Consumer Sentiment). Por ejemplo, utiliza exactamente las mismas cinco preguntas<sup>1</sup>, desagrega el índice general en los mismos dos subíndices: *icea* e *iee*, y utiliza tamaños de muestra telefónicas muy similares (<http://www.sca.isr.umich.edu/main.php>)

Para el caso de la Universidad de Michigan, el Index of Consumer Expectation o ICE (el equivalente al *iee* de la Escuela de Estadística de la Universidad de Costa Rica) es utilizado por la Oficina de Análisis Económico del Departamento Federal de Comercio como una de las variables de su Indicador Adelantado Compuesto (Moynan 2010)

Obsérvese que el indicador de Michigan no es utilizado como indicador adelantado ‘univariado’ sino como insumo de un Indicador Adelantado Compuesto, y además, se utiliza el sub-indicador de expectativas, ICE.

Como comentario adicional se apunta que es importante distinguir el uso estadístico del indicador de la Universidad de Michigan como insumo de un Indicador Adelantado Compuesto, del uso periodístico y de las interpretaciones/pronósticos por parte de analistas y tomadores de decisiones, cuando los resultados son divulgados por los medios de prensa.

Finalmente se recuerda que el consumo privado interno es un componente con más peso relativo dentro de la economía estadounidense que dentro de la economía costarricense.

Los anteriores comentarios plantean algunas limitantes a-priori al uso del *icc* como indicador adelantado ‘univariado’ del IMAE. Por otra parte y en sentido positivo, dichos comentarios apuntan a la conveniencia de explorar la relación del *icc* con sectores del IMAE.

Stock y Watson (Stock J., Watson M. 1999) propusieron una metodología para pronosticar la inflación con base en una gran cantidad de indicadores económicos. En su estudio evaluaron individualmente 168 indicadores según su capacidad predictiva con respecto a la inflación.

---

1 Falta

### 3. METODOLOGÍA DE STOCK Y WATSON

---

---

---

La metodología de evaluación de indicadores es la siguiente:

- a- Estimar un modelo autoregresivo,  $AR(p)$ , de la variable a predecir, en nuestro caso, el IMAE.
- b- El orden del modelo, o número de rezagos,  $p$ , viene dado por el valor que maximice algún criterio de información. Por lo general se utilizan el criterio de Akaike (AIC) o el criterio de Schwarz (BIC)
- c- La capacidad predictiva de dicho modelo autoregresivo se evalúa mediante la raíz del error cuadrático medio. Para esto se divide la muestra total y se proyectan los dos últimos años de la muestra.
- d- Al modelo autoregresivo se le agrega la variable que se quiere probar, por ejemplo se agrega el *icc*. Dicha variable entra también en forma rezagada, es decir como componente autoregresivo.
- e- La capacidad predictiva de dicho modelo autoregresivo ampliado se evalúa también mediante la raíz del error cuadrático medio.
- f- Se observa si la raíz del error cuadrático medio del modelo ampliado es menor que la del modelo original, en cuyo caso la variable agregada es un predictor adecuado.

La longitud -aún mensualizada- de la serie del *icc*, no resultó suficiente para poder dividir la muestra en una submuestra de entrenamiento y una submuestra de predicción. Adicionalmente, la presencia de un solo punto de giro<sup>2</sup> en el periodo que abarca la serie, impide también realizar dicha partición. Lo anterior no permite emplear el criterio del error cuadrático medio.

Se optó, entonces, por utilizar la muestra completa -sin particiones- y emplear el criterio de disminución en el AIC y/o BIC (se es consciente que conceptualmente el AIC y/o BIC no son sustitutos del MSE, por el hecho, sobradamente conocido, que un buen ajuste no implica necesariamente una buena predicción)

La metodología, arriba reseñada, se aplicó también a la relación del *icc* con cada sector del IMAE.

Adicionalmente, todo el conjunto de estimaciones se replicó empleando el *icea* y empleando el *iee*.

Para las estimaciones y la selección automática del número de rezagos se usó la función `ar.mle()` del paquete `stats` y también se utilizó el paquete `dynamic linear models (dynlm)`, ambos del software R.

## 4. RESULTADOS

---

---

Los resultados se observan en los siguientes cuadros.

El cuadro No. 1 contiene las variaciones relativas en el AIC y en el BIC cuando se pasa de un modelo con solo un componente AR de la serie objetivo (IMAE o algún sector) a un modelo ampliado con un componente AR del *icc* (alternativamente con un componente AR del *icea* y con un componente AR del *iee*).

Los cuadros del No. 2 al No. 5 contienen el detalle de estadísticos de ajuste y de diagnóstico. Incluyen los valores de los coeficientes, el  $r^2$ , el  $r^2$  ajustado, los p-value de las pruebas de Box-Ljung y Box-Pierce, y una indicación de presencia de valor extremo y/o influyente.

El Cuadro No. 2 presenta dichos resultados para el modelo con solo el componente AR de la serie objetivo, mientras que los cuadros No. 3 al No. 5 corresponden a los resultados de los modelos con la adición del componente AR del *icc*, *icea* e *iee* respectivamente.

El cuadro No. 1 permite identificar, con base en los mayores porcentajes de disminución del AIC y/o BIC, el sector de Transporte, Almacenamiento y Comunicaciones y el sector de Comercio como los más indicados para emplear el *iee* como una variable a incluir en un Indicador Adelantado Compuesto de dichos sectores.

En efecto, para el sector Transporte, Almacenamiento y Comunicaciones la adición del componente AR de *iee* reduce AIC y BIC en 23% y 13% respectivamente. En el sector Comercio las correspondientes reducciones son de 12% y 9%.

El mejor desempeño relativo del *iee* sobre el *icc* recuerda el caso del Indicador de Michigan donde la Oficina de Análisis Económico del Departamento Federal de Comercio de Estados Unidos incluye el Index of Consumer Expectation en su Indicador Adelantado Compuesto.

Siendo difícil de establecer en forma rigurosa, se puede argumentar que resulta intuitivo que el sector de Transporte, Almacenamiento y Comunicaciones y el sector de Comercio, sectores directamente relacionados con el consumo privado interno, presenten los mayores porcentajes de disminución del AIC y/o BIC cuando se adiciona un componente AR de variables relacionadas con las expectativas del consumidor.

Pasando a los cuadros No. 2 al No. 5, todos los modelos estimados resultan adecuados en el sentido que las pruebas de Box-Ljung y Box-Pierce no rechazan la hipótesis nula de independencia de los errores. Sin embargo, todos los modelos estimados presentan presencia de valor extremo/influencial, precisamente en el punto de giro. Por tanto, la adición de un componente AR del índice (*icc,icea,iee*) al modelo AR de la serie objetivo (IMAE o algún sub-sector) no es suficiente para lograr un buen ajuste del punto de giro. A manera de ejemplo, se presentan los gráficos de diagnóstico (ver más adelante: Gráficos de Diagnóstico) para el sector Transporte,Almacenamiento y Comunicaciones, donde se evidencia la presencia de valor extremo/influencial en el punto de giro.

En conclusión, se identifican el sector de Transporte,Almacenamiento y Comunicaciones y el sector de Comercio como los más indicados para emplear el *iee* como una variable a incluir en un Indicador Adelantado Compuesto y, como era de esperar, el empleo de un indicador adelantado ‘univariado’ no es suficiente para lograr un buen ajuste del punto de giro.

Es pertinente mencionar que el buen ajuste de los puntos de giro sigue siendo un reto incluso para las oficinas gubernamentales que cuentan con sistemas bien establecidos de Indicadores Adelantados Compuestos. Este tema tomó especial relevancia con la última recesión de los países industrializados declarada en el tercer cuatrimestre del 2008 (las críticas a los Indicadores Adelantados arrancan con Koopmans (Koopmans 1947), una referencia más reciente es el estudio de Koenig y Emery (Koenig E., Emery K. 1993).

Dada la presencia de valor extremo/influencial en el punto de giro, para el sector de Transporte,Almacenamiento y Comunicaciones y para el sector de Comercio se aplicó una prueba de estabilidad estructural de los parámetros: la prueba supF o generalización de Andrews de la F de Chow (Andrews 1993, Andrews DWK, Ploberger W. 1994). Para ello se utilizó el paquete Structural Change (strucchange) del software R.

Los resultados para el sector Transporte,Almacenamiento y Comunicaciones se presentan en el recuadro No. 1 y los resultados para el sector Comercio se presentan en el recuadro No. 2. En ambos casos se rechaza la hipótesis nula de estabilidad estructural de los parámetros sobre el periodo completo de la serie.

Estos resultados abren un espacio para en un futuro, con series de datos más extensas, aplicar algún modelo del tipo Markov Switching iniciados por Hamilton (Hamilton 1989).

**Cuadro I**  
**Variación en AIC y BIC al introducir componentes AR de ica, icea e ice**

AR de la serie	IMAE	Manufactura	Minas	Agricultura	Transporte	Otr Ser Empr	Elect y Agua	Construcción	Comercio	Finan/Seguros	Hoteles	Inter finan Ind
AIC	273.1029	417.5064	453.2659	249.3607	80.61116	169.852	190.9765	324.7275	251.4093	174.1329	201.2426	173.4605
BIC	280.3231	424.7266	460.4861	258.9385	90.18896	188.6024	198.1966	334.3053	258.6295	181.3531	219.9931	180.6806
AR de la serie + AR de ICC												
IMAE		Manufactura	Minas	Agricultura	Transporte	Otr Ser Empr	Elect y Agua	Construcción	Comercio	Finan/Seguros	Hoteles	Inter finan Ind
AIC	261.1161	408.034	453.2589	246.7392	65.30869	166.5105	188.4347	325.1054	226.5814	170.0101	196.6701	173.4151
BIC	275.4828	422.4007	467.6255	263.5003	82.06983	192.2923	202.8014	341.8665	240.9481	184.3768	222.4519	1887.7818
% variación												
AIC	-4.39	-2.27	0.00	-1.05	-18.98	-1.97	-1.33	0.12	-9.88	-2.37	-2.27	-0.03
BIC	-1.73	-0.55	1.55	1.76	-9.00	1.96	2.32	2.26	-6.84	1.67	1.12	3.93
AR de la serie + AR de ICEA												
IMAE		Manufactura	Minas	Agricultura	Transporte	Otr Ser Empr	Elect y Agua	Construcción	Comercio	Finan/Seguros	Hoteles	Inter finan Ind
AIC	266.6623	412.648	452.9807	244.7602	71.87214	170.4966	190.276	327.5007	237.2201	172.1944	196.0461	172.6748
BIC	281.029	427.0147	467.3474	261.5213	88.63328	196.2785	204.6427	344.2618	251.5868	186.5611	221.8279	187.0415
% variación												
AIC	-2.36	-1.16	-0.06	-1.84	-10.84	0.38	-0.37	0.85	-5.64	-1.11	-2.58	-0.45
BIC	0.25	0.54	1.49	1.00	-1.72	4.07	3.25	2.98	-2.72	2.87	0.83	3.52

Continúa...

**Cuadro I**  
**Variación en AIC y BIC al introducir componentes AR de icc, icea e iee**      Continuación...

AR de la serie + AR de IEE												
	IMAE	Manufactura	Minas	Agricultura	Transporte	Otr Ser Empr	Elect y Agua	Construcción	Comercio	Finan/Seguros	Hoteles	Inter finan Ind
AIC	258.5939	405.8754	452.9747	247.0785	61.83503	165.3002	187.3952	324.6652	221.7391	170.212	194.5045	173.9306
BIC	272.9606	420.2421	467.3414	263.8396	78.59617	191.082	201.7619	341.4263	236.1058	184.5787	220.2863	188.2973
% variación												
AIC	-5.31	-2.79	-0.06	-0.92	-23.29	-2.68	-1.88	-0.02	-11.80	-2.25	-3.35	0.27
BIC	-2.63	-1.06	1.49	1.89	-12.85	1.31	1.80	2.13	-8.71	1.78	0.13	4.22



**Cuadro 3**  
**AR de la serie + AR del icc**

AR de la serie + AR de ICC												
	IMAE	Manufactura	Minas	Agricultura	Transporte	Otr Ser Empr	Elect y Agua	Construcción	Comercio	Finan/Seguros	Hoteles	Inter finan Ind
AR de la serie												
orden	2	2	2	3	3	7	2	3	2	2	7	2
coef 1	1.0265	1.0201	1.1235	1.03028	1.10395	1.001731	1.0633	1.18609	1.03608	1.1391	1.03107	1.1897
coef 2	-0.1271	-0.1412	-0.17	-0.01542	-0.16134	0.001916	-0.1406	-0.08979	-0.11721	-0.1534	-0.02858	-0.2161
coef 3				-0.07878	0.02001	-0.051185		-0.11661			0.04474	
coef 4						-0.008911					-0.06245	
coef 5						-0.012264					0.04311	
coef 6						0.00121					-0.05099	
coef 7						-0.109281					-0.01708	
+												
AR de ICC												
orden	3	3	3	3	3	3	3	3	3	3	3	3
coef 1	-1.3602	-2.6868	1.4575	0.19981	-0.61103	0.398758	-0.3658	-1.38977	-2.24443	-0.8892	-0.80224	-0.4036
coef 2	3.0928	6.2795	-2.4698	0.03243	1.33168	-0.5306	0.9268	3.04205	4.66056	1.7289	1.83562	0.6822
coef 3	-1.7225	-3.5829	1.0118	-0.22891	-0.71511	0.185997	-0.5565	-1.64964	-2.411063	-0.8378	-1.03281	-0.2754
coef 4												
coef 5												
R2	0.9665	0.9147	0.9374	0.968	0.9981	0.9981	0.9659	0.9828	0.9635	0.9941	0.9682	0.9933
R2 adj	0.9643	0.9091	0.9333	0.9655	0.998	0.9978	0.9636	0.9815	0.9611	0.9937	0.9635	0.9929
p-values												
Box-Pierce	0.8254	0.8244	0.8252	0.9455	0.6438	0.996	0.8635	0.8516	0.5724	0.9581	0.9851	0.9353
Box-Ljung	0.8222	0.8212	0.822	0.9445	0.6377	0.996	0.861	0.8488	0.5653	0.9573	0.9848	0.9341

Continúa...



**Cuadro 4**  
**AR de la serie + AR del ica**

AR de la serie + AR de ICC												
	IMAE	Manuf�ctura	Minas	Agricultura	Transporte	Otr Ser Empr	Elect y Agua	Construcci�n	Comercio	Finan/Seguros	Hoteles	Inter finan Ind
AR de la serie												
orden	2	2	2	3	3	7	2	3	2	2	7	2
coef 1	1.00804	1.0178	1.1346	1.04452	1.03711	1.1044116	1.0664	1.17834	0.95907	1.1381	0.948262	1.2017
coef 2	-0.11762	-0.1325	-0.1756	-0.04482	-0.04284	-0.024216	-0.1432	-0.07103	-0.03887	-0.1645	0.01889	-0.242
coef 3				-0.07747	-0.04202	-0.0464838		-0.1299			0.001807	
coef 4						-0.0280647					-0.003857	
coef 5						0.0008334					0.006304	
coef 6						-0.0210547					0.014313	
coef 7						-0.0559126					-0.037897	
+												
AR de ICC												
orden	3	3	3	3	3	3	3	3	3	3	3	3
coef 1	0.09354	0.8078	2.8157	1.52316	-0.10298	0.8852883	0.3476	-0.56949	-1.0589	-0.8399	0.195919	-0.604
coef 2	0.38284	-0.4323	-5.6372	-2.53442	0.42547	-1.6301795	-0.4635	1.5064	2.68581	1.6968	0.106425	1.1171
coef 3	-0.46284	-0.3625	2.8224	1.01687	-0.31401	0.7683273	0.1215	-0.93225	-1.61984	-0.8519	-0.301434	-0.5068
coef 4												
coef 5												
R2	0.9641	0.9097	0.9377	0.9688	0.998	0.998	0.9651	0.9823	0.9584	0.994	0.9685	0.9934
R2 adj	0.9618	0.9038	0.9336	0.9663	0.9978	0.9977	0.9628	0.9809	0.9556	0.9936	0.9638	0.9929
p-values												
Box-Pierce	0.8959	0.8748	0.8248	0.9397	0.9397	0.8472	0.8959	0.8764	0.6616	0.9726	0.9443	0.9056
Box-Ljung	0.894	0.8725	0.8216	0.9386	0.8979	0.8443	0.894	0.8742	0.6557	0.9721	0.9432	0.9039

Contin a...

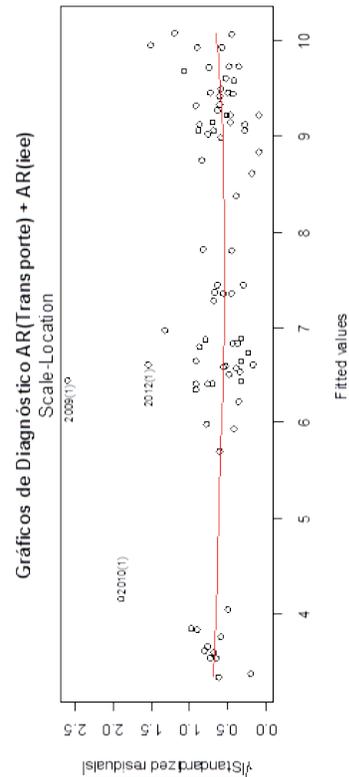
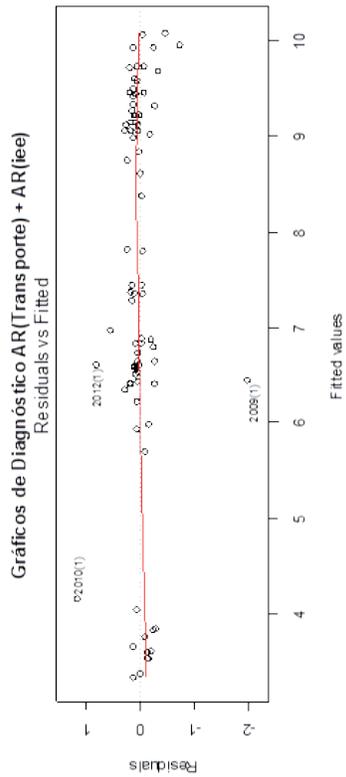
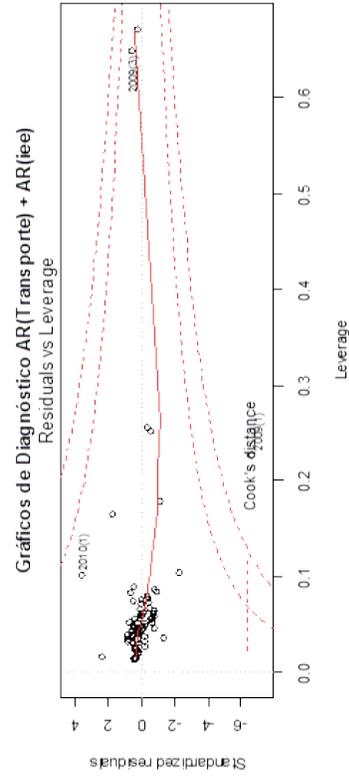
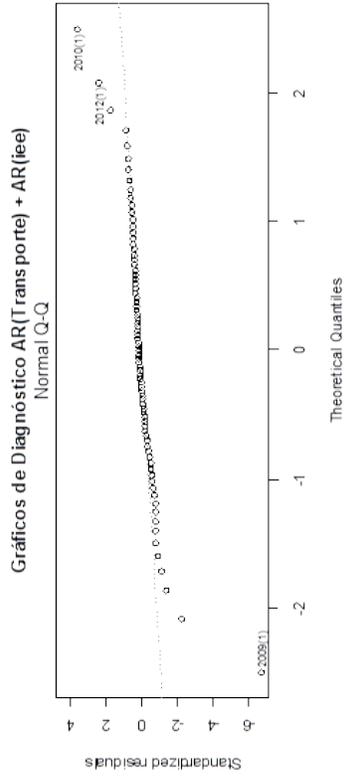


**Cuadro 5**  
**AR de la serie + AR del ico**

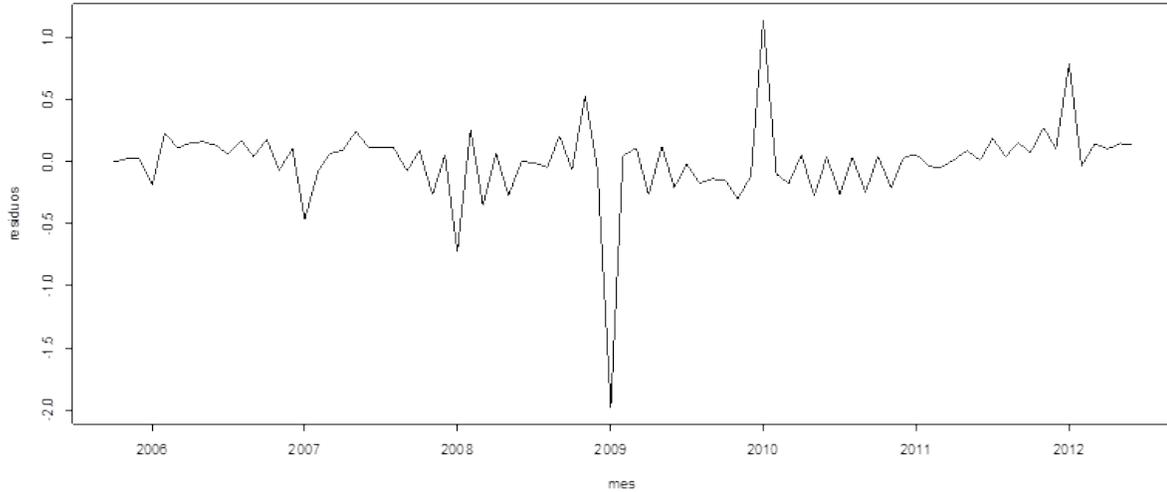
AR de la serie + AR de IEE												
	IMAE	Manufactura	Minas	Agricultura	Transporte	Otr Ser Empr	Elect y Agua	Construccion	Comercio	Finan/Seguros	Hoteles	Inter finan Ind
AR de la serie												
orden	2	2	2	3	3	7	2	3	2	2	7	2
coef 1	1.0428	1.0327	1.1148	1.04398	1.12741	0.992085	1.0719	1.19294	1.04774	1.1333	1.04986	1.1868
coef 2	-0.1343	-0.1494	-0.1623	-0.01783	-0.20571	0.001837	-0.1435	-0.09636	-0.1315	-0.142	-0.05441	-0.2069
coef 3				-0.08087	0.04618	-0.052539		-0.11531			0.06491	
coef 4						-0.008282					-0.06635	
coef 5						-0.012802					0.03848	
coef 6						0.002081					-0.04146	
coef 7						-0.107723					-0.02974	
+												
AR de IEE												
orden	3	3	3	3	3	3	3	3	3	3	3	3
coef 1	-1.3761	-2.8841	0.7514	-0.16137	-0.59765	0.232	-0.4479	-1.27196	-2.01039	-0.7033	-0.86512	-0.26
coef 2	2.9832	6.3534	-1.0503	0.62714	1.24911	-0.267416	1.03	2.70785	4.04465	1.3448	1.85945	0.4093
coef 3	-1.599	-3.4613	0.2976	-0.46345	-0.64516	0.088447	-0.5784	-1.4343	-2.02908	-0.6406	-0.99382	-0.1474
coef 4												
coef 5												
R2	0.9675	0.917	0.9377	0.9679	0.9982	0.9981	0.9663	0.9829	0.9656	0.9941	0.9691	0.9933
R2 adj	0.9654	0.9115	0.9336	0.9653	0.9981	0.9979	0.9641	0.9816	0.9633	0.9937	0.9645	0.9928
p-values												
Box-Pierce	0.8672	0.8338	0.8287	0.8969	0.644	0.9534	0.8577	0.8497	0.7211	0.973	0.9281	0.9411
Box-Ljung	0.8648	0.8307	0.8255	0.895	0.6381	0.9525	0.8551	0.8469	0.7162	0.9725	0.9267	0.94

Continúa...

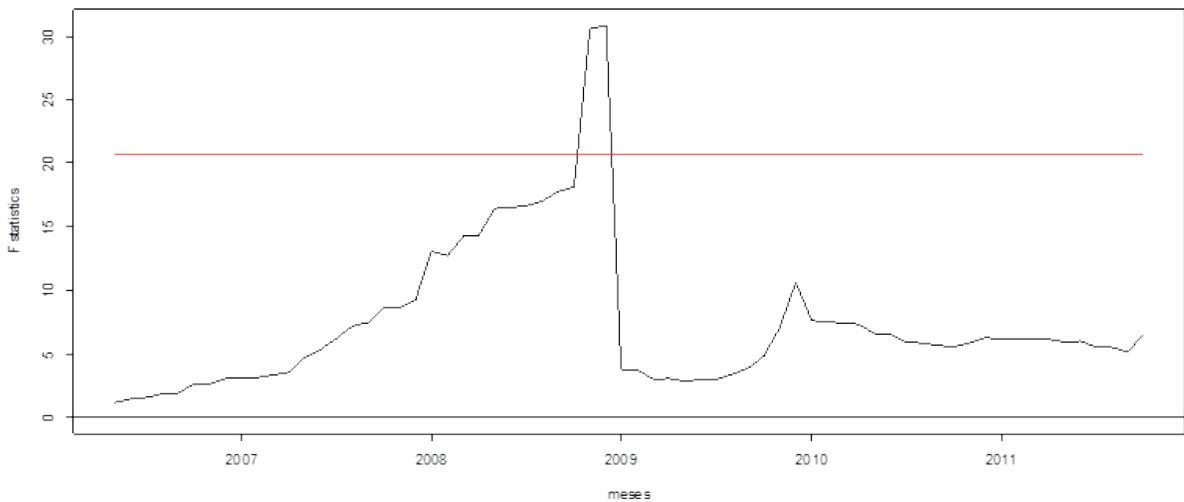




**Cuadro No.6**  
**Gráficos de diagnostico AR (Transporte) +AR (iee)**



**Recuadro No.1**  
**Prueba de Cambio Estructural para el modelo del sector Transporte**  
**Prueba supF de Andrews para AR (transporte, 3) + AR(iee,3)**



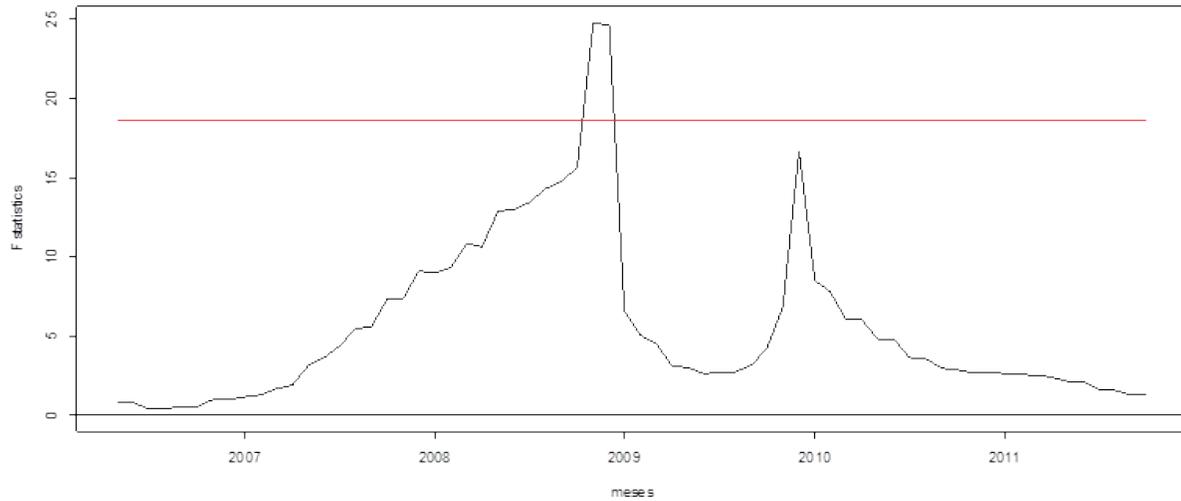
supF test

data: model\_transporte\_strchange2

sup.F = 30.8352, p-value = 0.0009316

---

**Recuadro No.2**  
**Prueba de Cambio Estructural para el modelo del sector Comercio**  
**Prueba supF de Andrews para AR (comercio, 2) + AR (iee,3)**



supF test

data: model\_comercio\_strchange2

sup.F = 24.7874, p-value = 0.004503

## 5. ANÁLISIS DE FACTORES

---

---

Se ha mencionado, por parte de analistas y tomadores de decisión, que la evolución del *icc* puede estar influenciada por variables extra-económicas; en particular, se han señalado las noticias del entorno político como una fuente plausible de variación.

Con la intención de hacer un estudio exploratorio de las influencias subyacentes a las variables medidas en la encuesta, se aplicó un Análisis de Factores.

En el contexto de series de tiempo la técnica más conocida es la llamada Análisis de Factores Dinámico. Esta técnica hace uso de Variables de Estado para modelar la dinámica de los factores no observables subyacentes a las variables medidas. Esto requiere conocimiento e información adicional sobre dichos factores, o en su defecto, postular algún tipo de hipótesis acerca de la dinámica de los factores.

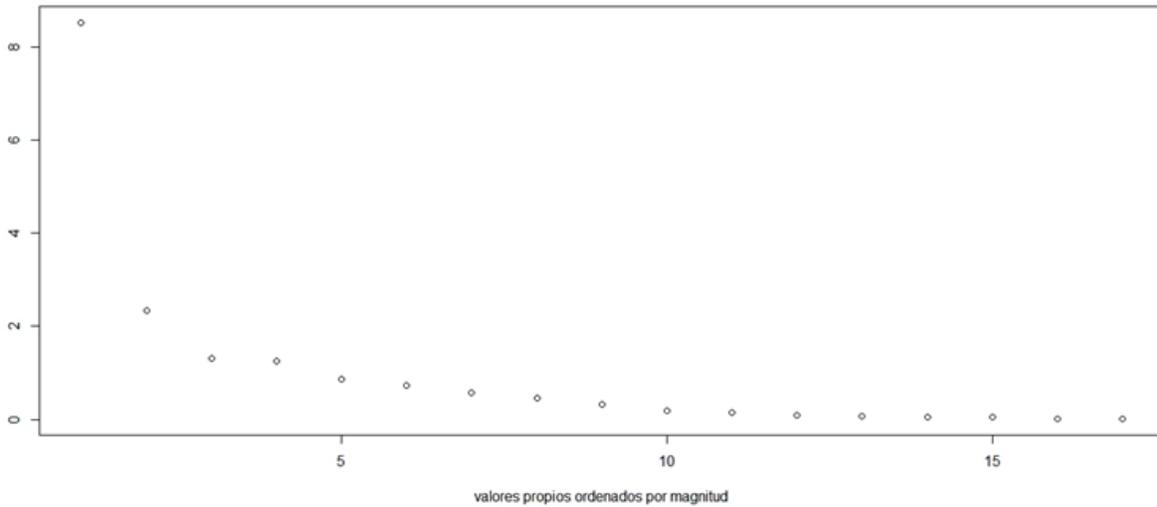
Otra técnica para la aplicación del Análisis de Factores en un contexto de series de tiempo, es la propuesta por Gilbert y Meijer (Gilbert P. D., Meijer E., 2005), la cual no precisa explicitar ningún patrón dinámico de los factores no observables. El paquete Time Series Factor Analysis (tsfa), del software R, implementa dicha técnica.

Mediante dicho paquete se trabajó las siguientes series trimestrales: **icc**, Tasas de interés serán mayores prox 12m (**in1mayores**), Precios serán mayores prox 12m (**in2mayores**), Ingreso familiar crecerá más que los precios (**in3mayor**), Espera que ingreso familiar sea mayor que año pasado (**in4mayor**), Política económica gob actual es buena (**gb1bueno**), Desempleo aumentará en prox 12 m (**gb2mayor**), Pobreza será mayor prox 12 m (**gb3mayor**), Buenos tiempos para comprar casa (**bd1buenos**), Buenos tiempos para comprar carro (**bd5buenos**), Hogar tiene capacidad de gastos discrecionales (**gasdis**), Puede comprar muebles (**gd1r**), Pude comprar vehículo (**gd2r**), Puede realizar viaje de placer (**gd3r**), Puede comprar casa (**gd4r**), Pueden comprar microcomputadora (**gd5r**), Pueden comprar terreno (**gr6r**).

### Número de Factores

Las series se trabajan en sus primeras diferencias, por lo que el análisis inicia con la extracción de los valores propios de la matriz de correlación de las primeras diferencias.

### Valores Propios de la matriz de correlación de las primeras diferencias de las series



**Cuadro No. 7**  
**Valores Propios de la matriz de correlación**  
**de las primeras diferencias**

Eigenvalues	% explicación	acumulado
<b>8.53350776</b>	<b>50.2%</b>	<b>50.2%</b>
<b>2.34751524</b>	<b>13.8%</b>	<b>64.0%</b>
<b>1.31808781</b>	<b>7.8%</b>	<b>71.8%</b>
<b>1.25841037</b>	<b>7.4%</b>	<b>79.2%</b>
0.86979088	5.1%	84.3%
0.73715879	4.3%	88.6%
0.56680775	3.3%	91.9%
0.44893079	2.6%	94.6%
0.32212389	1.9%	96.5%
0.17718398	1.0%	97.5%
0.14270243	0.8%	98.4%
0.08515676	0.5%	98.9%
0.07163671	0.4%	99.3%
0.04973518	0.3%	99.6%
0.04096012	0.2%	99.8%
0.01968903	0.1%	99.9%
0.01060252	0.1%	100.0%

Siguiendo el criterio de incluir en el análisis todos los valores propios mayores a 1, el análisis se corrió con cuatro factores. El cuadro No. 7 permite constatar que con cuatro factores se explica el 79% de la variancia. Se observa también que el primer factor explica el 50%, y los dos primeros factores explican un 64% de la variancia.

## Resultados

El cuadro No. 8 muestra las cargas factoriales estandarizadas luego de aplicar una rotación oblicua.

**Cuadro No. 8**  
**Cargas Factoriales Estandarizadas**

Código	Nombre	Factor 1	Factor 2	Factor 3	Factor 4
<i>icc</i>	Indice Confianza del Consumidor	0.011600594	<b>0.758465537</b>	0.15338814	0.325754461
in1mayores	Tasas de interés serán mayores prox 12m	-0.033345033	-0.170133514	0.01104388	<b>-0.676174320</b>
in2mayores	Precios serán mayores prox 12m	0.223055316	-0.041061587	0.09416049	<b>-0.785886397</b>
in3mayor	Ingreso familiar crecerá más que los precios	<b>0.455007140</b>	0.139100940	-0.01456539	0.004237909
in4mayor	Espera que ingreso familiar sea mayor que año pasado	0.046256505	0.272658536	-0.31103156	0.225360903
gb1bueno	Política económica gob actual es buena	0.383273954	0.038299320	<b>0.43066706</b>	<b>0.605140205</b>
gb2mayor	Desempleo aumentará en prox 12 m	0.007695668	<b>-0.935695111</b>	0.21278412	0.126471613
gb3mayor	Pobreza será mayor prox 12 m	-0.047863135	<b>-0.964907725</b>	-0.13720790	-0.012739630
bd1buenos	Buenos tiempos para comprar casa	0.195724656	0.290032669	0.20277630	<b>0.410700489</b>
bd5buenos	Buenos tiempos para comprar carro	0.269664345	-0.005806707	-0.08444789	<b>0.719897950</b>
gasdis	Hogar tiene capacidad de gastos discrecionales	<b>0.624763978</b>	0.071784227	<b>-0.55356106</b>	0.142736411
gd1r	Puede comprar muebles	<b>0.607084547</b>	0.107690569	<b>-0.52417812</b>	0.030547559
gd2r	Puede comprar vehículo	<b>0.755279105</b>	0.091836682	-0.12625956	0.131580365
gd3r	Puede realizar viaje de placer	<b>0.652331631</b>	-0.112816983	-0.14427200	0.249386027
gd4r	Puede comprar casa	<b>1.050216066</b>	-0.011782457	0.17630397	-0.160284573
gd5r	Pueden comprar microcomputadora	0.297906633	0.086268999	<b>-0.41425643</b>	0.370840313
gd6r	Pueden comprar terreno	<b>0.984480334</b>	0.081912110	0.01941365	-0.069507837

De la distribución de las mayores cargas factoriales en cada factor, tanto de signo positivo como negativo, se constata que:

- el primer factor tiene relación con aspectos económicos del ámbito familiar, pues en dicho factor las mayores cargas -todas positivas- son en las siguientes variables:

Ingreso familiar crecerá más que los precios (**in3mayor**), hogar tiene capacidad de gastos discrecionales (**gasdis**), Puede comprar muebles (**gd1r**), puede comprar vehículo (**gd2r**), puede realizar viaje de placer (**gd3r**), puede comprar casa (**gd4r**), pueden comprar terreno (**gd6r**). Es decir, variables relacionadas con la expectativa de crecimiento real del ingreso familiar y relacionadas con la capacidad familiar de realizar cierto tipo de gastos importantes.

- el segundo factor tiene relación negativa con las variables Desempleo aumentará en prox 12 meses (**gb2mayor**) y Pobreza será mayor prox 12 meses (**gb3mayor**), y relación positiva con el *icc*.

Específicamente con relación al *icc*, se observa que tiene su carga factorial mayoritariamente en el segundo factor, un factor relacionado con la expectativa de evolución en el corto plazo de variables socio-económicas, y además, no exhibe carga factorial en el primer factor, relacionado con aspectos económicos familiares. Al considerar estos resultados hay que tener presente que los factores finalmente obtenidos se correlacionan entre sí, pues se usó una rotación no ortogonal. La correlación estimada entre estos dos factores es de 0.397 (la matriz completa de correlación entre factores se presenta más adelante):

- en el cuarto factor las variables que tienen mayores cargas factoriales positivas son: Política económica gob actual es buena (**gb1bueno**), Buenos tiempos para comprar casa (**bd1buenos**) y Buenos tiempos para comprar carro (**bd5buenos**). Y las variables con mayores cargas factoriales negativas son: Tasas de interés serán mayores prox 12meses (**in1mayores**) y Precios serán mayores prox 12meses (**in2mayores**). Este factor está mayoritariamente relacionado con las expectativas sobre las condiciones económico/financieras (precios y tasas de interés) que afectan la compra de casa y/o carro.
- el tercer factor presenta una distribución de cargas factoriales menos clara, especialmente en sus signos. Por otra parte, resalta el hecho que los estimados del coeficiente de correlación de este factor con cualquiera de los otros tres factores son negativos. Tal y como se aprecia en la matriz de correlación entre factores que se presenta en el Cuadro No. 9.

**Cuadro No. 9**  
**Matriz de correlación Phi de factores**

	[,1]	[,2]	[,3]	[,4]
[1,]	1.0000000	0.3971906	-0.1788222	0.3895763
[2,]	0.3971906	1.0000000	-0.1558579	0.4883971
[3,]	-0.1788222	-0.1558579	1.0000000	-0.1367091
[4,]	0.3895763	0.4883971	-0.1367091	1.0000000

Finalmente, el cuadro No. 10 presenta las comunalidades para cada serie. Para efectos de este estudio hay que resaltar la comunalidad de aproximadamente 91% de la serie del *icc*.

**Cuadro No. 10**  
**Comunalidades de cada variable**

Código	Nombre	Comunalidad
icc		90.6%
in1mayores		62.5%
in2mayores		57.9%
in3mayor		28.2%
in4mayor		35.3%
gb1bueno		78.0%
gb2mayor		87.0%
gb3mayor		95.7%
bd1buenos		50.1%
bd5buenos		76.9%
gasdis		99.5%
gd1r		86.1%
gd2r		79.9%
gd3r		60.0%
gd4r		96.3%
gd5r		64.0%
gd6r		97.9%

La conclusión a resaltar es que el *icc* aparece relacionado más directamente con las expectativas de evolución en el corto plazo de variables socio-económicas y solo en forma indirecta -a través de la correlación entre el primer y segundo factor- se relaciona con aspectos económicos del ámbito familiar. Se puede argumentar que ésta mayor exposición del *icc* a variables macro en contraposición a variables micro/familiares hace más plausible una influencia de variables extra-económicas.



## 6. SUGERENCIAS

---

---

---

### Sugerencia Principal : Indicador Adelantado Compuesto de OECD

#### Metodología

Una continuación en cierta forma natural de este trabajo sería aplicar la metodología de la OECD de Indicadores Adelantados Compuestos, la cual se describe a continuación. Para mayores detalles referirse a la fuente original (OECD, Abril 2012)

El objetivo general de la metodología es proporcionar señales tempranas de los puntos de giro en la actividad económica.

El proceso metodológico completo consiste en la aplicación de cinco grandes etapas, cada una de ellas conformada por varias sub-etapas. El mapa del proceso es el siguiente:

#### **PRE-SELECCION**

- Serie Objetivo
- Series Componentes

#### **FILTRADO**

- Periodicidad
- Ajuste Estacional
- Detección de Valores Extremos
- Identificación del Ciclo
- Normalización

#### **EVALUACION**

- Número de meses de Adelanto
- Adecuación al Ciclo Completo
- Ciclos omitidos/Ciclos extras
- Desempeño

#### **AGREGACION**

- Ponderación
- Corrimiento de Rezagos e Inversión
- Agregación

**PRE-SELECCION**

- Serie Objetivo  
Debe ser una serie de base mensual y adecuada proxy de la actividad económica.
- Series Componentes  
Se seleccionan con base en los siguientes criterios:
  - Relevancia Económica
  - Cobertura
  - Frecuencia de la serie
  - Sujeta a revisiones periódicas
  - Disponibilidad oportuna
  - Longitud

**FILTRADO**

- Esta etapa busca eliminar los demás elementos de la serie que no permitan la adecuada determinación del ciclo de cada una. Elementos tales como tendencia, estacionalidad y valores extremos son removidos.
- Periodicidad  
Consiste en transformar las series trimestrales a una base mensual.
  - Ajuste Estacional  
Como su nombre lo indica, se desestacionaliza las series. La OECD requiere que se aplique X12ARIMA o TRAMO-SEATS
  - Detección de Valores Extremos  
Se utiliza TRAMO-SEATS para identificar y sustituir varios tipos de valores extremos: aditivos, transitorios, cambios de nivel.
  - Identificación del Ciclo  
Luego de eliminar estacionalidad y valores extremos se procede a remover la tendencia de largo plazo y el componente irregular (variación de alta frecuencia). Esto se realiza mediante la aplicación de un filtro band-pass (Hodrick-Prescott o Christiano-Fidgerald).
  - Normalización  
Las diferentes series componentes están en distintas escalas por lo que se debe realizar un proceso de normalización.

**EVALUACION**

- Número de meses de Adelanto  
El número de meses de adelanto varía con cada punto de giro. Se utilizan, entonces, como resúmenes, el número promedio de adelanto y el número mediano de adelanto. Es deseable obtener entre 6 y 9 meses de adelanto.
- Adecuación al Ciclo Completo  
Es deseable no solo un adelanto mínimo promedio/mediano de 6 meses en los puntos de giro, sino una adecuación del índice adelantado compuesto a las fases ascendentes y descendentes del ciclo.

- Ciclos omitidos/Ciclos extras  
El índice no debe emitir señales falsas (extra ciclos) ni mostrar insensibilidad (omitir ciclos)
- Desempeño  
Las series componentes se pueden agregar en modos alternativos para obtener diferentes índices compuestos. El desempeño de cada uno de estos índices alternativos se compara entre sí con base en los criterios antes mencionados.

### ***AGREGACION***

- Ponderación  
Las series componentes se ponderan en forma uniforme para obtener el índice compuesto. Se reconoce que el proceso de normalización introduce una ponderación implícita.
- Corrimiento de Rezagos e Inversión  
El corrimiento de rezagos es una transformación en la estructura de rezagos de las series que busca maximizar el número de meses de adelanto.  
Por otra parte, algunas de las series componentes suelen exhibir comportamiento contracíclico por lo que requieren su inversión.
- Agregación

Se utilizan procedimientos de agregación robustos a valores faltantes.

### Tiempo

Por lo común se inicia con un conjunto de 40 o más series (representando diferentes sectores de la actividad económica) y cada una de ellas es sometida a varias, o, a todas las etapas del proceso anteriormente descrito.

Se recomienda un mínimo de un año para cumplir con todas las etapas.

### Otras Sugerencias

Se mencionan dos sugerencias adicionales.

1-) Aplicar un Modelo de Ecuaciones Estructurales mediante el módulo AMOS de SPSS que formalice la estructura subyacente encontrada mediante el análisis de factores.

2-) Hacer un Análisis de Sentimiento (Sentiment Analysis) mediante text-mining de las noticias relacionadas de economía nacional, economía internacional, gobierno y administración. Este Análisis de Sentimiento proporcionaría un "índice de noticias".

### NOTAS

1. Q1 "We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?"

Q2 "Now looking ahead--do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?"

Q3 "Now turning to business conditions in the country as a whole--do you think that during the next twelve months we'll have good times financially, or bad times, or what?"

Q4 "Looking ahead, which would you say is more likely--that in the country as a whole we'll have continuous good times during the next five years or so, or that we will have periods of widespread unemployment or depression, or what?"

Q5 "About the big things people buy for their homes--such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or bad time for people to buy major household items?"

2. En este trabajo de carácter exploratorio y con series de datos cortas, no se usó una definición formal de ciclo ni de punto de giro. La metodología de la OECD provee pautas al respecto (OECD, Abril 2012)

## REFERENCIAS

---

---

Andrews DWK, Tests for parameter instability and structural change with unknown change point, *Econometrica*, 61, 821-856, 1993

Andrews DWK, Ploberger W., Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, 60 (4), 953-966

Cleveland, William S.; Devlin, Susan J. Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting, *Journal of the American Statistical Association* 83 (403): 596–610, 1988.

Gilbert P. D., Meijer E., Time Series Factor Analysis with an Application to Measuring Money, University of Groningen, Research School SOM Research Report 05F10, Nov. 2005

J. D. Hamilton, A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, Vol. 57, No. 2., pp. 357-384, 1989.

<http://www.census.gov/srd/www/sapaper/sapaperbytopic.html>

<http://www.sca.isr.umich.edu/main.php>

Koening E., Emery K., Why the composite index of leading indicators doesn't lead, Federal Reserve Bank of Dallas, Research Paper No. 9318, May 1993.

Koopmans T.C., Measurement without theory, *The Review of Economic Statistics*, Vol XXIX (August, 1947)

Moylan C. E., Cyclical Indicators for the United States, Third International Seminar on Early Warning and Business Cycle Indicators, Moscow Russian Federation, November 17-19, 2010

OECD System of Composite Leading Indicators, April, 2012

Pedersen, M. Un indicador líder compuesto para la actividad económica en Chile, *Revista Monetaria*, vol. XXXII, No. 2 abril-junio, CEMLA, 2009.

Stock, James and Mark W. Watson. "Forecasting Inflation," *Journal of Monetary Economics*, 1999, v44 (2,Oct), 293-335.

Yabuta Yoriko, New tools for tracking the Mexican Business Cycle, Third International Seminar on

Early Warning and Business Cycle Indicators, Moscow Russian Federation, November 17-19, 2010

## APENDICES METODOLOGICOS

---

---

---

### MODELOS ARIMA

En 1970, G. E. P. Box y G. M. Jenkins desarrollaron un cuerpo metodológico destinado a **Identificar, Estimar y Diagnosticar** modelos de series de tiempo, por tal razón también se conocen como modelos Box-Jenkins.

En el contexto de modelos ARIMA una serie de tiempo se conceptúa como una muestra de un proceso estocástico. Es decir, se considera que un proceso estocástico es el generador de la serie de tiempo observada.

La metodología ARIMA asume, y este es un supuesto de gran importancia, que la gran mayoría de series de tiempo que se observan son generadas -generadas de forma exacta, o de forma aproximada- por un proceso débilmente estacionario MA(q), o por un proceso débilmente estacionario AR(p), o un proceso débilmente estacionario ARMA(p,q), que es una combinación de ambos. Se comprenderá entonces que dada la serie de tiempo que se desea proyectar, es de suma importancia poder identificar si proviene de un proceso MA(q) o, de un proceso AR(p) o, de un proceso ARMA(p,q) (suponiendo que sea débilmente estacionario). Esto se hace mediante la inspección visual de las llamadas Funciones de Autocorrelación y Autocorrelación Parcial.

### Función de Autocorrelación (ACF)

Es intuitivo esperar que, por lo general, dos observaciones consecutivas de una serie de tiempo estén relacionadas. Una posible forma de medir dicha relación es utilizar el coeficiente de correlación, aplicando dicho coeficiente a los pares de observaciones consecutivas. Como los pares con los que se hace el cálculo corresponden a la misma variable, al resultado se le denomina auto-correlación.

De igual forma se puede calcular el coeficiente de correlación con los pares de observaciones separadas por 2 unidades de tiempo, por 3 unidades de tiempo y así sucesivamente.

Se puede hacer un gráfico con dichos coeficientes: utilizando el eje horizontal para las unidades de separación, y el eje vertical para los coeficientes. Este es el gráfico de la función de autocorrelación (ACF).

### Función de Autocorrelación Parcial (PACF)

Si se sigue el mismo procedimiento, pero en lugar del coeficiente de correlación simple de Pearson se utiliza el coeficiente de correlación parcial, se obtiene el gráfico de la función de autocorrelación parcial (PACF).

En forma intuitiva se puede argumentar que la forma de la ACF y de la PACF de una serie de tiempo (en otras palabras: los valores particulares de los coeficientes y en

ese orden específico) debe estar determinada, en un alto grado, por el proceso estocástico que generó dicha serie.

Es decir: si la serie de tiempo que se desea proyectar proviene de un proceso MA(q), sus ACF y PACF deben tener un patrón particular conocido (un proceso MA(1) tiene una ACF y una PACF con una forma específica y exclusiva, igualmente para un MA(2), para un MA(3) y así sucesivamente)

De la misma manera, si la serie de tiempo que se desea proyectar proviene de un proceso AR(p), sus ACF y PACF tienen un patrón particular conocido.

Llegamos así a la llamada etapa de Identificación dentro de los modelos ARIMA.

### **Etapa de Identificación**

La etapa de Identificación consiste en calcular la función de autocorrelación, ACF, y la función de autocorrelación parcial, PACF, de la serie que se desea proyectar; luego compararlas visualmente con las correspondientes funciones ACF y PACF de los modelos teóricos AR(p), MA(q) o ARMA(p,q). Finalmente se escoge el modelo (es una escogencia preliminar) ARIMA cuyas ACF y PACF teóricas sean lo más parecidas a las recién calculadas ACF y PACF.

Una vez identificado el proceso, se pasa a la etapa de estimación.

### **Etapa de Estimación**

Si el proceso de Identificación tuvo éxito, se procede a la estimación de los parámetros del modelo.

Por ejemplo, suponga que se identificó un AR(2) como el proceso generador de la serie que se desea proyectar. La fórmula de un AR(2) es (obviando el término de error) :

$$X_t = a_1 X_{t-1} + a_2 X_{t-2}$$

En la etapa de Estimación se aplican procedimientos de Máxima Verosimilitud para obtener estimados numéricos de  $a_1$  y  $a_2$

### **Etapa de Diagnóstico**

y la ACF de los residuos obtenidos en la estimación deben ser "ruido blanco" . Si la ACF de los residuos obtenidos en la estimación no son "ruido blanco", habrá que realizar una nueva estimación incorporando la estructura más parecida al modelo teórico que se pueda intuir con la comparación visual con los modelos teóricos.

Se desprende de esta última etapa que la metodología ARIMA es de carácter iterativo.

### **Transformaciones**

En los casos prácticos donde la serie de tiempo que se desea proyectar No se puede considerar como proveniente de un proceso estocástico débilmente estacionario, la metodología ARIMA recomienda aplicar transformaciones a los datos de tal forma que se genere una nueva serie de datos que Sí se pueda considerar como proveniente de un proceso estocástico débilmente estacionario.

Las transformaciones más comunes son: Diferencias Sucesivas, Rezagos y Logaritmo Natural.

Rezagos: consiste en desfazar cada observación de la serie. Si cada observación se desfaza un período se habla de rezago de un período, si cada observación se desfaza dos períodos se habla de rezago de dos períodos, y así sucesivamente.

Box y Jenkins hacen uso intensivo del Operador de Rezagos  $Ly_t = y_{t-1}$ , y de los llamados Polinomios de Rezagos. Estas herramientas pertenecen a una tradición – se podría decir que perdida o al menos en desuso- de Cálculo de Diferencias Finitas y Ecuaciones en Diferencias, herramientas muy usadas en cálculo actuarial, y más recientemente, en Matemáticas Discretas y Control de Sistemas Lineales Discretos

Logaritmo Natural: consiste en tomar el logaritmo natural de los datos. Es una transformación que busca lograr que las varianzas de las variables aleatorias no dependan del tiempo y sean finitas. Dicha transformación es un caso particular de una clase más general de transformaciones para estabilizar la varianza, conocidas como Transformaciones Box-Cox, a las cuales se recurre si la transformación logaritmo no es exitosa.

Diferencias Sucesivas: consiste en tomar diferencias de observaciones sucesivas. Es una transformación que busca lograr que las esperanzas matemáticas de las variables aleatorias sean constantes. De ser necesario se aplica más de una vez. Box y Jenkins escriben que, en su experiencia, la mayoría de las veces es suficiente con tomar 2 diferencias sucesivas (es decir un  $d$  máximo de 2 en la notación de Box-Jenkins). Es posible incurrir también en lo que se conoce como sobre-diferenciación. ¿Cuántas veces tomar diferencias? formalmente este es un tema de pruebas de raíz unitaria.

### **Criterios de Información**

Hasta el momento se ha descrito la metodología original sugerida por los mismos Box y Jenkins:

- Se asume que la gran mayoría de series de tiempo que se observan son generadas -generadas de forma exacta, o de forma aproximada- por un proceso débilmente estacionario de tipo AR, MA o ARMA.
- Si el proceso no fuera débilmente estacionario, se aplican diferencias sucesivas a la serie (y otras transformaciones si vienen al caso)
- Se calculan las ACF y PACF
- Se inicia la Etapa de Identificación
- Etapa de Estimación
- Etapa de Diagnóstico

Sin embargo se ha vuelto muy popular, debido a las facilidades de cómputo de la actualidad, escoger el modelo ARIMA (escoger los parámetros  $p, d, q$  del proceso  $ARIMA(p,d,q)$ ) haciendo uso de los llamados Criterios de Información. Es decir, la Etapa de Identificación es sustituida por el uso de Criterios de Información.

Con tales criterios se reduce el grado de subjetividad implícito en la comparación entre, por

un lado, la ACF y PACF teóricas, y por otro, la ACF y la PACF empíricas. Los criterios de información más comúnmente empleados son: el Criterio de Información de Akaike, el Criterio de Información Bayesiano y el Criterio de Información de Akaike corregido.

#### Criterio de Información de Akaike AIC

Fue sugerido por H. Akaike en 1974. Se puede interpretar que establece un balance entre el sesgo y la variancia en un ajuste particular, también se puede interpretar que establece un balance entre la exactitud y la complejidad (siendo la complejidad medida en forma aproximada mediante el número de parámetros incorporados en el modelo). En este último sentido se suele afirmar que el AIC es un criterio penalizado: penaliza si se agrega un parámetro adicional al modelo.

$$AIC = n \log \left( \frac{SSE}{n} \right) + 2(k + 2),$$

donde k es el número de parámetros incorporados en el modelo.

#### Criterio de Información Bayesiano BIC

La derivación teórica del BIC se fundamenta en la teoría bayesiana y se suele atribuir su creación a Schwarz en 1978, por lo que también se conoce como criterio de información de Schwarz.

El BIC se diferencia del AIC en la forma en que penaliza la inclusión de un parámetro adicional al modelo.

$$BIC = n \log \left( \frac{SSE}{n} \right) + (k + 2) \log(n)$$

Existe cierta controversia sobre cuál criterio es más apropiado, pero se ha observado que por lo general el modelo que minimiza el AIC suele también minimizar el BIC -aunque esto no es un resultado formal- por lo que el tema no es de particular importancia en las aplicaciones. Formalmente lo que se ha establecido es que ambos criterios son asintóticamente equivalentes, lo cual, con todo y ser un resultado teórico interesante, no es de relevancia en contextos aplicados con muestras pequeñas.

#### Criterio de Información de Akaike corregido AICc

Precisamente se ha modificado el AIC para que incorpore una corrección para muestras finitas, a tal coeficiente se le denota AICc.

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

donde k es el número de parámetros incorporados en el modelo, n es el tamaño de muestra. El AICc penaliza en mayor medida que el AIC la inclusión de un parámetro adicional al modelo.

Entonces, utilizando los criterios de información con los modelos ARIMA, la metodología a

seguir queda de la siguiente manera:

- Se asume que la gran mayoría de series de tiempo que se observan son generadas -generadas de forma exacta, o de forma aproximada- por un proceso débilmente estacionario de tipo AR, MA o ARMA.
- Si el proceso no fuera débilmente estacionario, se aplican diferencias sucesivas a la serie (y otras transformaciones si vienen al caso)
- Se calculan AICc, AIC, BIC para los diferentes modelos ARIMA
- Se escoge el modelo ARIMA con menor AICc, AIC, BIC
- Etapa de Estimación

### ***Elección de un Modelo de Series de Tiempo***

Dados dos o más modelos de series de tiempo alternativos, se suelen citar razones a priori y razones a posteriori para escoger entre ellos.

#### Razones A Priori

Debido a que no existe una teoría general de series de tiempo, a priori no es posible afirmar categóricamente que un modelo sea mejor que otro.

Se suelen citar razones a priori con el fin de limitar el número de posibles modelos a evaluar/validar. Sin embargo, en el tanto el énfasis sea proyectar o extrapolar, el veredicto final debe ser a posteriori.

Algunas de las razones a priori que se citan son:

- El Horizonte de Pronóstico  
Que tan lejos se quiera pronosticar. Por ejemplo, los modelos de suavizamiento exponencial son muy útiles en el corto y muy corto plazo.
- Rapidez y Facilidad de Aplicación y de Actualización  
La rapidez y facilidad de aplicación en el día a día es un factor generalmente apreciado por la gerencia y altos mandos de la organización. La actualización ante cambios importantes del contexto evita incurrir en grandes errores individuales de predicción.
- Una teoría estadísticamente más rigurosa  
Por ejemplo en este aspecto los modelos ARIMA aventajan a los tradicionales modelos de Suavizamiento Exponencial: los modelos de Suavizamiento Exponencial no cuentan con el equivalente a la Etapa de Identificación de la metodología ARIMA siendo así de naturaleza mucho más ad hoc, los procedimientos de estimación de los modelos ARIMA son estadísticamente más rigurosos, los modelos ARIMA cuentan con intervalos de confianza  
En general las ventajas teórico-estadísticas son razones a priori que apuntan a un ajuste estadístico correcto. Sin embargo, un ajuste correcto no garantiza una buena proyección. En términos de los Criterios de Información recién mencionados: el modelo con el menor AICc es el que mejor se ajusta a los datos observados, NO NECESARIAMENTE es el que mejor predice los valores futuros.  
Justamente por ello se hace necesario las Medidas de Error/Exactitud que en breve se expondrán.

- La fuerza relativa de los diferentes componentes de la serie.

Por ejemplo, si una serie tiene una fuerte tendencia que prevalece sobre los demás componentes, es plausible el llamado método de regresión con errores AR(1)

O si la serie es dominada por un fuerte componente estacional, es plausible el llamado método de regresión de componentes armónicos.

- Si los patrones son estáticos (constantes en el tiempo) o dinámicos (cambian en el tiempo)

En esta línea, y a riesgo de sobre-simplificar, se puede anotar que los métodos de regresión asumen patrones estáticos, los modelos ARIMA asumen que los patrones varían según un proceso estocástico moderado (y que no obstante permite la suficiente variación como para describir una amplia gama de fenómenos), finalmente los modelos de Variable de Estado especifican ecuaciones para los patrones en sí.

### Razones A Posteriori

Las evaluaciones/validaciones de los modelos alternativos se hacen en forma empírica, haciendo uso de las llamadas Medidas de Error/Exactitud y haciendo uso del Procedimiento de Validación Cruzada.

En cada periodo futuro del horizonte de pronóstico se comete un error de predicción, debido a que el valor predicho no es igual al valor que efectivamente se realiza/observa.

La forma en que se mide esa discrepancia en cada periodo, por ejemplo en forma de una diferencia, o en forma de razón, o mediante una diferencia porcentual, o con una diferencia cuadrática, produce varias medidas alternativas de error/exactitud.

También la forma en que dichas discrepancias individuales se agregan en un solo número representativo de todo el horizonte de pronóstico, produce varias medidas alternativas, aunque por lo general se agregan mediante una suma simple.

### **Medidas de Error/Exactitud**

Entre las medidas de error/exactitud más utilizadas se encuentran el MAPE, el MAD y el MSD.

#### MAPE (Porcentaje promedio absoluto de error)

Mide la exactitud de los valores proyectados de la serie de tiempo. La exactitud se expresa como un porcentaje.  $y_t$  es igual al valor observado,  $\hat{y}_t$  es el valor estimado y  $n$  el número de observaciones.

$$\text{MAPE} = \frac{\sum |y_t - \hat{y}_t|}{n} \times 100 \quad (y_t \neq 0)$$

### MAD (Desviación media absoluta)

Mide la exactitud de los valores proyectados de la serie de tiempo. Expresa la exactitud en las mismas unidades de los datos.

$$MAD = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n}$$

### MSD (Desviación cuadrática media)

Es el promedio de las diferencias cuadráticas.

$$MAD = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|^2}{n}$$

El MSD es la medida más utilizada, aunque es más sensible a errores anormales de pronóstico que el MAD.

## ANÁLISIS FACTORIAL

El Análisis Factorial es un método de estadística multivariable cuyo objetivo es la reducción de datos. Tuvo su origen en la psicología cuantitativa y en la actualidad es aplicado en gran variedad de campos.

Se acostumbra clasificar el análisis factorial en dos tipos: exploratorio y confirmatorio.

El análisis factorial exploratorio se usa para tratar de descubrir la estructura interna de un número relativamente grande de variables. La hipótesis a priori del investigador es que pueden existir una serie de factores asociados a grupos de variables. Las cargas de los distintos factores se utilizan para intuir la relación de éstos con las distintas variables. Es el tipo de análisis factorial más común.

El análisis factorial confirmatorio trata de determinar si el número de factores obtenidos y sus cargas se corresponden con los que cabría esperar a la luz de una teoría previa acerca de los datos. La hipótesis a priori es que existen unos determinados factores preestablecidos y que cada uno de ellos está asociado con un determinado subconjunto de las variables. El análisis factorial confirmatorio se puede considerar un modelo particular de la clase de Modelos de Ecuaciones Estructurales (SEM).

A continuación se introduce el análisis factorial exploratorio en términos intuitivos. Es intuitivo pensar que: "si dos variables tienen un factor en común, la correlación entre ellas se deba -en parte- a dicho factor"

También es intuitivo esperar que: "entre más las variables dependan del factor, más dicho factor explicará la correlación entre ellas"

Esto motiva expresar cada variable como una suma ponderada de dos términos: el factor y un término de error.

Siguiendo con el ejemplo, supóngase ahora que ambas variables comparten no uno sino dos factores en común. Ahora cada variable se expresaría como la suma ponderada de tres términos: los dos factores y el término de error.

Es inmediato llegar al caso de  $N$  variables compartiendo  $K$  factores en común. Cada variable expresada como la combinación lineal de los  $K$  factores y el término de error. En el tanto  $K$  sea considerablemente menor que  $N$ , se logra el objetivo de reducción de datos.

A este sistema  $N \times K$  no se le puede aplicar ningún método de regresión pues los factores son No observables.

Las  $N$  variables Sí son observables y se puede calcular el coeficiente de correlación entre cada par. Así pues el análisis factorial inicia con el cálculo de la matriz de correlación.

A continuación se describen –muy someramente- los pasos o tareas específicas que se siguen para completar un análisis de factores exploratorio.

En cada uno de los pasos el analista dispone de varias alternativas para realizar la misma tarea y la metodología no proporciona reglas claras –ni se ha logrado un consenso- para escoger entre una u otra alternativa. Por ejemplo existen varios métodos para determinar el número de factores.

Esta característica del análisis de factores exploratorio ha sido criticada debido a la discrecionalidad en que, necesariamente, incurre el analista. Por otra parte, también se argumenta que el éxito y popularidad del análisis se debe precisamente a esta flexibilidad.

1. Generar la matriz de correlación entre todas las variables involucradas
2. Determinar el número de factores a extraer

El método más común es obtener los valores propios de la matriz de correlación.

Se decide extraer tantos factores como valores propios mayores a uno se obtengan.

3. Extraer los factores

El método más común es realizar un análisis de Componentes Principales de la matriz de correlación.

4. Se rotan los factores para mejorar su interpretabilidad

Los factores se interpretan según su mayor o menor asociación a variables específicas. Si un factor no está claramente asociado a unas variables, es de difícil interpretación. Con las rotaciones se busca que el patrón de asociación del factor sea más definido, es decir: claramente asociado a unas variables.

Existen rotaciones ortogonales y rotaciones oblicuas. Por lo general se aplican rotaciones ortogonales.