

El «Indicador Sintético de Posición Sectorial» (ISPS), el «Indicador de Clima Industrial» (ICI), el «Indicador Factorial [Dinámico] de la Industria» (IFI) o el «Indicador de Diversificación Energética» (IDE) son algunos ejemplos de indicadores sintéticos, utilizados en los ámbitos de las políticas industrial y energética, que difieren no solo en cuanto a su finalidad -análisis estructural vs coyuntural-, sino que también, relevantemente, en cuanto a su metodología de cálculo.

ASPECTOS METODOLÓGICOS DE LOS INDICADORES SINTÉTICOS DE INDUSTRIA Y ENERGÍA

Un «indicador sintético» o agregado, como su propio nombre indica, sintetiza o agrega varios indicadores parciales -o subcomponentes- con el objeto de facilitar el estudio simplificado de realidades complejas por medio de la comparación entre entidades o el análisis de su evolución temporal -histórica o adelantada-. Es únicamente en el contexto de un ejercicio particular de *benchmarking* de este tipo donde cabe interpretar los valores concretos que para un indicador agregado se calculen.

Según la guía metodológica OCDE/CCI [0], en la construcción de un indicador sintético se han de seguir ciertas etapas (1) entre las cuales la de **ponderación y agregación**, que puede llevarse a cabo en diferentes dimensiones o niveles jerárquicos, es especialmente relevante. Cuestiones a considerar en este respecto son la mecánica de síntesis -aritmética/aditiva vs geométrica/multiplicativa- y la magnitud y el sentido (2) de las contribuciones de los indicadores parciales, en cuanto que pueden resultar en una compensación (3), una contabilización múltiple -en casos de alta correlación entre subcomponentes- o una consideración desequilibrada (4) de los mismos.

La forma más elemental de realizar la agregación de indicadores parciales es calculando su media aritmética. Una alternativa algo más elaborada sería calculando una media ponderada, en la que las ponderaciones/pesos pueden escogerse con criterios o bien **subjetivos** (5) o bien **objetivos**, destacando entre estos aquellos basados en métodos estadísticos de **análisis multivariante** como los dos que sucintamente se describen a continuación.

El «Análisis de Componentes Principales» (ACP -o PCA, en su acrónimo en inglés-) es una técnica descriptiva de reducción de la dimensionalidad que, partiendo de los indicadores parciales o variables originales, construye un número manejable de combinaciones lineales, denominadas «componentes principales», de forma que sean de varianza máxima y ortogonales entre sí -incorreladas-, recogiendo por orden de importancia una parte sustancial de la variabilidad de aquellas:

$$Y = A \cdot X$$

siendo

X el vector ($px1$) de variables originales u observadas

A la matriz (qxp) de cargas factoriales, con $q < p$

Y el vector ($qx1$) de componentes principales

Se trata de un ejercicio de optimización lagrangiana iterada que resulta en la **diagonalización** de la matriz de varianzas-covarianzas de las observaciones de las variables originales (6), siendo cada autovalor la varianza total explicada por el respectivo componente y cada autovector los coeficientes en su combinación lineal -filas de la matriz A -. Estos coeficientes se denominan «cargas factoriales» o «factores de carga» (7) (*factor loadings*) y sus cuadrados representan, a modo de sensibilidad, la parte de la varianza de cada variable original explicada por el respectivo componente principal. Los autovalores se ordenan decrecientemente, de manera que para la elección de q se tomarán en cuenta tantos de estos hasta que se considere adecuado el nivel acumulado de varianza total explicada. Una explicación intuitiva de esta técnica es su interpretación geométrica como búsqueda de las direcciones principales de un hiperelipsoide por rotación a una base ortonormal, que en este caso constituiría unos ejes "naturales" de datos en el sentido de que sobre los mismos se maximiza la variabilidad de las observaciones.

Por su parte, el «Análisis Factorial» (AF -o FA, en su acrónimo en inglés-) supone que existe un número reducido de variables latentes hipotéticas o «factores comunes» subyacentes -artificiales y no observables- que explican una gran proporción de la varianza compartida por las variables originales -varianza común-, capturando así sus interrelaciones. El modelo contempla parsimoniosamente que, además de estas «comunalidades» (8), existen «factores únicos» o «especificidades» idiosincráticas de perturbación/ruido para cada una de las variables:

$$X = A \cdot F + U$$

siendo

X el vector ($px1$) de variables originales u observadas

A la matriz (pxq) de cargas factoriales, con $q < p$

F el vector ($qx1$) de factores comunes

U el vector ($px1$) de factores únicos

Se tiene entonces que para la entidad k la observación t es

$$X_{it}^k = \sum_{j=1}^q a_{ij}^k \cdot F_{jt}^k + U_{it}^k \quad (\text{con } i = 1, 2, \dots, p)$$

donde a las F_{jt}^k (con $j = 1, 2, \dots, q$) se las denomina «puntuaciones factoriales» (*factor scores*). Es decir, si n es el número de observaciones de cada entidad (9), el modelo asocia la matriz (pxn) de observaciones de las variables originales (10) con una matriz ($qx n$) de puntuaciones factoriales, que son por tanto las coordenadas de aquellas en el hiperespacio de dimensión reducida $q < p$ en el que puede considerarse que se generan.

El AF es una técnica estadística que se utiliza generalmente con carácter exploratorio, si bien cabe también hacerlo con carácter confirmatorio –test de hipótesis-. Su uso inferencial requiere de los datos variabilidad y la interdependencia que subyace en el modelo, que habrá de ser oportunamente contrastada. A diferencia del ACP, en el AF existen diferentes métodos para la extracción de los factores comunes y estos –y sus cargas factoriales asociadas- no constituyen una solución única (11) a no ser que se imponga algún criterio adicional. Así por ejemplo, y con el denominado «Método del Factor Principal» –en el que, al igual que en el ACP en el que se basa, se busca que los factores comunes sean incorrelados (12) y expliquen la máxima varianza compartida-, la formalización del problema resulta en una **descomposición en valores singulares** (DVS –o SVD en su acrónimo en inglés-) iterada (13) que parte de una estimación inicial de la matriz de varianzas-covarianzas de las perturbaciones (14). Para dotarlos de una mayor interpretabilidad, los factores comunes se someten a una rotación que los asocie a “bloques” razonablemente disjuntos de variables originales (15), para lo cual existen también diferentes métodos.

Con este enfoque AF, el peso de cada variable original en un indicador sintético ponderado se toma proporcional a su comunalidad. Por tanto, una vez extraídos los factores comunes y obtenidas sus cargas factoriales en cada variable original, el peso correspondiente se calcularía como la suma de los cuadrados de aquellas, dividida por la totalidad de la varianza común explicada por los factores retenidos –de manera que todos los pesos sumen uno-. Esto es, el valor del indicador sintético para la entidad k correspondiente a la observación t sería:

$$I_t^k = w_1^k \cdot X_{1t}^k + w_2^k \cdot X_{2t}^k + \dots + w_p^k \cdot X_{pt}^k = \sum_{i=1}^p w_i^k \cdot X_{it}^k$$

$$\text{donde } w_i^k = \frac{\sum_{j=1}^q (a_{ij}^k)^2}{\sum_{i=1}^p \sum_{j=1}^q (a_{ij}^k)^2} \text{ y por tanto } \sum_{i=1}^p w_i^k = 1$$

Resaltar que el método tiene el inconveniente de que la llegada de nuevos datos exige la reevaluación del modelo. Además, y por sus características ya apuntadas, requiere un especial énfasis en el análisis tanto de su robustez como de la coherencia de sus resultados con el marco teórico-conceptual en el que se concibe el indicador sintético.

Previamente a la construcción de un indicador agregado, sus indicadores parciales se han de adimensionalizar en aras a su comparabilidad por medio de una **normalización/tipificación** (16), para la cual un método sencillo es el denominado «MÍN-MÁX» que reescala las observaciones (V) originales transformándolas linealmente en unas observaciones normalizadas (W) acotadas entre dos valores mínimo y máximo –de ahí el nombre- escogidos a priori:

$$W(V) = \frac{V - V_{\text{mín}}}{V_{\text{máx}} - V_{\text{mín}}} \cdot (W_{\text{máx}} - W_{\text{mín}}) + W_{\text{mín}}$$

Una **representación gráfica** muy visual para mostrar la información contenida en un indicador sintético es el «diagrama de radar» o “tela de araña”, en el que los radios se corresponden con cada uno de los indicadores parciales. Otros gráficos propios del análisis multivariante ACP/AF son el «diagrama de sedimentación» – o *Scree Plot*, que ilustra cómo se acumula la varianza al retener el modelo un mayor número de componentes/factores-, el «círculo de correlaciones» –que representa las cargas factoriales de las distintas variables originales mediante puntos en un plano cuyas dimensiones se corresponden con dos componentes seleccionadas al efecto- y el más informativo *Biplot* –que muestra tanto los factores de carga como las puntuaciones factoriales estimadas (17)-.

Como es natural, los paquetes de *software* estadístico como *Matlab*®, *Minitab*®, *R*®, *SAS*®, *SPSS*®, *Stata*®, *Statgraphics*® o *XLSTAT*® incluyen ACP y AF entre sus funcionalidades de análisis multivariante.

Una tercera metodología de ponderación, también con estimación endógena u objetiva de pesos, es la del «**Beneficio de la Duda**» (*BOD*, en su acrónimo en inglés) (18). Se trata de un caso particular del «**Análisis de Envoltante de Datos**» (*AED* –o *DEA* en su acrónimo en inglés-), en el que se construye una frontera eficiente en base a un modelo no paramétrico de **programación lineal**, calculándose por separado los pesos que maximizan el valor del indicador compuesto para cada entidad. Se consigue así una mayor aceptabilidad del mismo –las ponderaciones son las que más favorecen a cada entidad, lo que explica la denominación de la metodología y

TABLA
INDICADORES SINTÉTICOS DE INDUSTRIA Y ENERGÍA

	Media Aritmética	Análisis Factorial
Estructural/Histórico	Indicador Sintético de Posición Sectorial-ISPS [6]	Indicador de Diversificación Energética-IDE [8]
Coyuntural/Adelantado	Indicador de Clima Industrial-ICI [7]	Indicador Factorial [Dinámico] de la Industria-IFI [9]

Fuente: Elaboración propia

el consenso que propicia-, incorporando en su propio diseño la idea de *benchmarking* -como distancia de cada entidad a la frontera eficiente- y con la ventaja adicional de su invarianza ante el método de normalización de indicadores parciales escogido. Por contra, este enfoque presenta inconvenientes en cuestión de comparabilidad entre entidades, multiplicidad de soluciones o la falta de incentivos dinámicos que genera.

El uso de indicadores sintéticos para el análisis estructural histórico no demanda mucho más aparato metodológico que el descrito. Por contra, en la perspectiva dinámica propia del **análisis de coyuntura**, en la que se pretende anticipar el futuro próximo por medio de indicadores -simples o agregados- de carácter adelantado, se han de someter estos a un procesado de **series temporales** para deslindar sus diferentes componentes de ciclo-tendencia, efectos de estacionalidad y calendario, y errático/accidental/residual, suavizándolas de aquellos y, eventualmente, proyectando el último dentro de un intervalo de confianza con la ayuda de un modelo predictivo.

La elaboración de indicadores "soft" cualitativos a partir de encuestas opináticas de confianza ha sido una solución tradicionalmente adoptada para superar el inconveniente que para el seguimiento estrecho en el tiempo de una actividad económica representa el carácter retrasado de los indicadores "hard" cuantitativos (19). Una confrontación gráfica de niveles frente a variaciones temporales de indicadores de confianza es una sencilla pero potente herramienta visual de análisis cíclico -identificación de fases y puntos de giro del ciclo económico- conocida como «**trazador de clima económico**» (20).

Más sofisticada metodológica y computacionalmente es la **predicción en tiempo real** (*nowcasting*), cada vez más popular por la creciente disponibilidad de datos de alta frecuencia, que trata de aprovechar la información contenida en indicadores tanto *hard* como *soft*, incorporándola a medida que fluyen sus datos (*real time data flow*), para construir así un **termómetro** de la realidad dinámica a monitorizar.

La técnica utilizada en este caso es la del «**Análisis Factorial Dinámico**», que en esencia consiste en la formulación dinámica de un modelo AF utilizando la representación de **espacio de estados** -en la que se deslindan en dos ecuaciones los aspectos de medición y evolución temporal de estos- y el **filtro/observador de Kalman** -algoritmo recursivo de predicción, medición y corrección por retroalimentación sucesivas- para la estimación de los factores latentes de estado y la predicción de los indicadores parciales no disponibles -que resultan en paneles desequilibrados de datos por ser distinta la frecuencia de aparición de los mismos-.

A la vista de todo lo expuesto, y para finalizar, cabría clasificar los indicadores sintéticos citados en el encabezado de esta nota según su uso y metodología de agregación/ponderación de acuerdo a la tabla que se incluye (entre corchetes se referencian las respectivas notas metodológicas).

■ Antonio Moreno-Torres Gálvez

NOTAS

- (1) Definición del marco teórico-conceptual; selección de datos; estimación/imputación de datos faltantes; análisis exploratorio multivariante; normalización; ponderación y agregación; análisis de robustez y sensibilidad; interpretación de los resultados; vínculos con otros indicadores; y presentación de los resultados. Sobre normalización y representación gráfica se comenta algún aspecto más adelante en esta nota.
- (2) En principio los indicadores parciales han de contribuir al indicador sintético en el mismo sentido de la realidad que este trata de modelar, de manera que eventualmente se habrán de invertir, sustituyéndolos por sus respectivos recíprocos, aquellos de los preexistentes que lo hagan en sentido contrario.
- (3) Situación por la que una entidad obtiene puntuación alta en un indicador agregado, pese a puntuar muy bajo en uno de los indicadores parciales que lo componen, por el efecto compensatorio de una puntuación muy alta en otro de estos. Lo que puede mitigarse por medio de una agregación geométrica.
- (4) Como la desigual ponderación implícita resultante de la agregación por medias aritméticas en una jerarquía desbalanceada de niveles de subcomponentes.

- (5) Véase, por ejemplo, las referencias [1], [2] y [3].
- (6) De correlaciones, en versión estandarizada. Se trata de una matriz (pxp) simétrica -ortogonalmente diagonalizable- y semidefinida positiva –autovalores no negativos-.
- (7) Nótese que las cargas factoriales podrían resultar ser negativas.
- (8) En puridad este término se refiere a la parte de la varianza observada explicada por los factores, siendo la variabilidad remanente la «unicidad».
- (9) O el número de individuos, en el caso más general en el que no se dé una interpretación temporal a esta dimensión longitudinal del modelo de datos.
- (10) A partir de la matriz \bar{X} de observaciones centradas en medias, la matriz (pxp) de varianzas-covarianzas puede expresarse como $S = \frac{1}{n} \cdot \bar{X} \cdot \bar{X}^T$.
- (11) Puesto que en el producto $A \cdot F$ no son observables ni A ni F .
- (12) De no serlo, se habla de un modelo factorial "oblicuo".
- (13) En una generalización de la descomposición espectral, los «valores singulares» de una matriz A rectangular (pxq) son las raíces cuadradas de los autovalores –nunca negativos- de la matriz $A \cdot A^T$ cuadrada (pxp) , simétrica y semidefinida positiva, de manera que $A \cdot A^T = P \cdot D \cdot P^T = (P \cdot D^{\frac{1}{2}}) (D^{\frac{1}{2}} \cdot P^T)$ donde P es una matriz ortonormal de autovectores de $A \cdot A^T$ y $D^{\frac{1}{2}}$ es una matriz diagonal con los valores singulares de A , ambas (pxp) . En la iteración para aproximar $A_{(pxq)}$ se usa $P_{(pxq)} \cdot D_{(qxq)}^{\frac{1}{2}}$.
- (14) Por hipótesis esta matriz (pxp) de varianzas-covarianzas de las U_{it}^k es diagonal. Denominándola E , en el modelo de factores ortogonales es $S - E = A \cdot A^T$.
- (15) Las columnas de la matriz A de cargas factoriales rotada contendrán así valores altos para ciertas variables originales y pequeños para otras.
- (16) Un modo de normalización común es la estandarización $N(0,1)$.
- (17) El modelo AF, aunque definido a nivel estructural, está indeterminado a nivel de datos, por lo que hay varios métodos para la medición de las puntuaciones factoriales. Aspecto no relevante en el contexto de esta nota.
- (18) Véase la referencia [4].
- (19) Así por ejemplo, a partir de la «Encuesta de Coyuntura Industrial» (ECI) se elabora un «Indicador de Clima Industrial» (ICI) que anticipa razonablemente bien al «Índice de Producción Industrial» (IPI).
- (20) Véase la referencia [5].

REFERENCIAS

- [0] "Handbook on Constructing Composite Indicators. Methodology and User Guide". OECD & JRC (2008). Disponible en <https://www.oecd.org/sdd/42495745.pdf>
- [1] "International Energy Security Risk Index Report". Global Energy Institute-U.S. Chamber of Commerce (2018). Disponible en https://www.globalenergyinstitute.org/sites/default/files/energyrisk_intl_2018.pdf
- [2] "Environmental Performance Index. Technical Appendix". Yale Center for Environmental Law & Policy (2018). Disponible en <https://epi.envirocenter.yale.edu/downloads/epi2018technicalappendixv05.pdf>
- [3] "El riesgo regulatorio en el sector energético. Índice de Evaluación de la Calidad Regulatoria (ICRE)". Serrano Calle, Silvia. EOI (2013). Disponible en <https://www.eoi.es/es/savia/publicaciones/20731/el-riesgo-regulatorio-en-el-sector-energetico-indice-de-evaluacion-de-la-calidad-regulatoria-icre>
- [4] "An Introduction to 'Benefit Of The Doubt' Composite Indicators". Cherchye, L. et al. Social Indicators Research (2007). Disponible en <https://link.springer.com/article/10.1007/s11205-006-9029-7>
- [5] "El Trazador de Clima Económico". Almendros Ulibarri, José Manuel. Economía Industrial, número 395 (2015, 1T). Disponible en <https://www.mincotur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/395/NOTA%201%20ALMENDROS.pdf>
- [6] "Competitividad de los sectores manufactureros en la industria española. Análisis a través de los Indicadores [Sintéticos] de Posición Sectorial" (sobre el ISPS). Guerediaga Alonso, M^a Ángeles. Economía Industrial, número 383 (2012, 1T). Disponible en <https://www.mincotur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/383/NOTAS.pdf>
- [7] "Encuesta de Coyuntura Industrial (ECI): Metodología" (sobre el ICI). MINCOTUR. Disponible en <https://industria.gob.es/es-ES/estadisticas/Documents/Encuesta%20Coyuntura%20Industrial/Metodologia-ECI.pdf>
- [8] "Una aproximación a un Indicador de Diversificación Energética (IDE)". Cuevas Galindo, Ángel. Economía Industrial, número 388 (2013, 2T). Disponible en <https://www.mincotur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/388/%C3%81ngel%20Cuevas%20Galindo.pdf>
- [9] "Medición de la actividad industrial en tiempo real: un Indicador Factorial [Dinámico] de la Industria (IFI)". Cuevas Galindo, Ángel. Economía Industrial, número 382 (2011, 4T). Disponible en <https://www.mincotur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/382/NOTAS.pdf>