# Redundancia, entropía e importancia de los factores comunes en las series económicas peruanas: exploración preliminar

**Roddy Rivas-Llosa M.**[§]
Profesor de Economía, Finanzas e Informática Avanzada
Universidad del Pacífico – Lima, Perú.
Noviembre 2000

## Abstract

¿Qué tan coordinadas entre sí se encuentran las series económicas en el Perú? ¿Cuántos componentes independientes se pueden extraer al efectuar un análisis de factores comunes entre las principales series estadísticas disponibles? Si bien la correlación existente entre las series de datos más utilizadas se evidencia en prácticamente cualquier modelo econométrico general, es necesario desarrollar un análisis formal específico para responder convenientemente a estas preguntas. Mediante un enfoque basado en el método de componentes principales aplicado a información anual referida a la centuria comprendida entre los años 1896 y 1995, se concluye que es posible sintetizar 4 variables económicas ortogonales entre sí, que recogen más del 95% de la varianza observada en la totalidad de las 63 series de datos examinadas. Este hecho, que demuestra ser estable en el tiempo, ofrece una base de análisis más eficiente para estudiar la fenomenología económica por medios estadísticos.

## 1. Presentación

La noción de *entropía* destaca entre los elementos teóricos más distintivos de la Física moderna, pues su esencia se encuentra arraigada en la naturaleza a un nivel tan fundamental como las leyes del movimiento de Isaac Newton (1687) o la teoría de la relatividad de Albert Einstein (1916). La idea de entropía, que fuera vislumbrada por primera vez a finales del siglo XIX y afinada durante el siglo XX, se ha filtrado desde sus orígenes físicos hasta alcanzar a la teoría económica y convertirse en otra de sus concepciones inspiradoras.

Entropía significa "desorganización, aleatoriedad; un proceso de degradación hacia un estado último de uniformidad". En estos términos la *Encyclopædia Britannica* define la esencia del concepto, a primera vista bivalente, utilizado en el campo de las ciencias físicas para referirse a la tendencia natural del universo hacia el desorden, el caos. Una tendencia que se muestra sin ambigüedad en la experiencia cotidiana. Basta comprobar, por ejemplo, que el calor tiende a disiparse, fluyendo siempre de los cuerpos más calientes hacia los más fríos[1]. Así, al disponer dos vasijas con agua, una caliente y la otra fría, se tendrá un sistema organizado en el que todas las partículas de agua de 'alta energía' se hallan reunidas en la primera vasija. La entropía (desorganización) de este sistema es, por lo tanto, reducida. Si se conectara ambas vasijas mediante una termocupla[2], bastaría para producir una corriente eléctrica proporcional a la diferencia de temperaturas entre sus contenidos. Esta energía eléctrica podría ser entonces aprovechada para producir trabajo mecánico. Sin embargo, a medida que el calor fluye del primer contenedor al segundo, el sistema se 'desorganiza'

[§] RivasLlosa_RN@UP.edu.pe

[1] Esta regularidad es conocida como la *segunda ley de la termodinámica* y fue enunciada por primera vez en 1850 por el físico alemán Rudolf Clausius. Años más tarde, mientras investigaba la teoría del motor a vapor, este mismo científico propuso el término *entropía* para referirse a la disipación de la energía disponible.

[2] Dispositivo en forma de anillo compuesto por dos cables de metales distintos unidos en sus extremos. Si las uniones se disponen a temperaturas distintas, se producirá un corriente eléctrica en el circuito. Este efecto fue descubierto por el científico alemán Thomas Seebeck en 1821 y lleva su nombre.

(las partículas con alta y baja energía se desagrupan, se mezclan entre sí en ambas vasijas). Resulta evidente que este proceso no podrá mantenerse indefinidamente, pues pasado cierto tiempo se alcanzará el equilibrio energético entre ambos contenedores. La entropía del sistema habrá aumentado al acentuarse la correlación entre las propiedades físicas de sus componentes.

Como se puede deducir del ejemplo anterior, la capacidad de un sistema físico para generar trabajo mecánico depende esencialmente de la gradiente de energía que contiene. Esto es, qué tan efectivamente separados y organizados se hallan los elementos con distintos niveles de energía[3]. Entropía, en tanto sinónimo de desorganización informativa, también lo es de redundancia o coordinación en cuanto a las características observables de las partes de un todo. Cuanto mayor sea la correlación entre las características energéticas de los componentes de un sistema, tanto menor será su potencial para generar un efecto de utilidad. Mantengamos este razonamiento en mente por un momento como la **'proposición entrópica'** y regresemos ahora a temas más terrenales…

## 2. Motivación del estudio

La disponibilidad de medios electrónicos de registro y sistemas cada vez más efectivos para medir las variables relevantes de la economía ha favorecido que, en tiempos modernos, se tenga acceso a un volumen creciente de series estadísticas sobre el comportamiento agregado de los agentes. Tal abundancia de datos ha convertido a la econometría en un terreno fértil para el surgimiento de innumerables modelos paramétricos aplicados al mundo real. A través de estos modelos, se persigue hallar regularidades matemáticas sensatas al combinar las magnitudes observadas de las variables relevantes en cada caso, integradas en la forma de 'funciones de comportamiento'. Con excepción posiblemente de los modelos estrictamente autoregresivos, este análisis parte del supuesto de que la varianza observada en una determinada variable puede explicarse en cierta medida por la varianza registrada por otra[4]. El componente no explicado por la segunda podrá ser entonces explicado por una tercera, y de este modo sucesivamente hasta encontrar un modelo con las características buscadas, si es que existe[5].

No obstante, cabe detenerse en un hecho que por su evidencia podría pasarse por alto en este contexto: *las series de datos utilizadas para explicar una variable económica generalmente no son estadísticamente independientes entre sí*. Específicamente, no tienen por qué serlo desde el punto de vista de la correlación estadística. Así, al incluir en un modelo dado una serie adicional para explicar la variable analizada, tan sólo aportará aquella fracción de su varianza que no se encuentre correlacionada con el resto de regresores. Este hecho se encuentra asociado con el problema econométrico denominado **'multicolinealidad'**, fenómeno que reduce, en principio, la eficiencia de las estimaciones obtenidas mediante un modelo. La intensidad de su efecto pernicioso sobre los resultados depende del grado de correlación entre las series de datos utilizadas como regresores. Llevado al caso más extremo —aquél en el que algunos regresores fueran plenamente codependientes entre sí—, la estimación del modelo se haría numéricamente imposible, salvo que se eliminara del análisis el efecto de alguna de las variables colineadas.

[3] Nótese que este mismo principio puede aplicarse en escenarios muy variados. La capacidad de producir energía mediante el agua en una represa, por ejemplo, no se debe solamente a que es posible reunir organizadamente un volumen sustancial de agua en un lugar elevado, sino también a la disponibilidad de un terreno menos elevado donde dejarla caer!

[4] Esto implica, por ejemplo, que los movimientos ascendentes de una van consistentemente acompañados de movimientos similares en otra, ya sea en términos contemporáneos o con un desfase temporal sostenido.

[5] En particular, puede no existir un modelo que disponga del balance necesario entre forma funcional, simplicidad, poder explicativo, poder predictivo y robustez estadística.

En sendos trabajos pioneros, Pearson[6] y Spearman[7] abordaron el problema de la multicolinealidad, subrayando el hecho de que al incluir una variable adicional en un modelo no necesariamente se está incluyendo un factor explicativo adicional, sino un factor *marginal*. En este sentido, cabe concluir que, *ceteris paribus,* una variable aportará tanto mayor información al modelo cuanto menor sea la correlación estadística que guarda con el resto de regresores.

Retomemos ahora la *proposición entrópica* e interpretemos el problema que enfrenta la Economía bajo esta óptica. Cabe afirmar que las economías —léase los sistemas económicos— también exhiben una tendencia natural hacia el desorden (tal vez puede argumentarse a favor de esta afirmación en muchos otros planos, pero nos restringiremos aquí al plano del desorden *informativo*). Dicho comportamiento se manifiesta a través de la existencia de correlación estadística entre sus series representativas ya que, en este marco, desorden se ha de entender como uniformidad o coordinación. Más aún, la existencia de entropía es la causa última de multicolinealidad. De llegar a cancelarse la entropía subyacente, toda la información en una economía estaría organizada hasta el punto de hacer de la econometría una práctica menos accidentada de lo que es, ya que bastaría un puñado de variables completamente independientes entre sí para construir prácticamente cualquier modelo concebible.

Al aceptar la existencia de cierto grado de entropía en la economía, surge una interrogante natural. ¿Cómo medir o utilizar en la práctica la redundancia en los datos, con el propósito de sintetizar la información que contienen o entender con mayor claridad las implicancias de esta coordinación? En el pasado, Hotelling[8], Thurstone[9] y Kaiser[10] se hicieron esta misma pregunta y, en el intento de darle respuesta, perfeccionaron los fundamentos del análisis de descomposición de factores, enfoque que fuera sugerido años antes por K. Pearson. Estas investigaciones subrayaron el hecho de que es posible 'concentrar' o 'comprimir' la información registrada en un conjunto de series parcialmente redundantes, mediante el cálculo de un número reducido de *factores* que expliquen la varianza común (*análisis factorial*) o la varianza total (*análisis de componentes principales*) de los datos originales. Este último procedimiento, en particular, surgió como un intento reducir el volumen de los datos numéricos disponibles, de modo tal que se pudiera estudiar con mayor facilidad un fenómeno complejo a través de un número manejable de factores 'abstractos', informativamente equivalentes o estadísticamente equivalentes al conjunto de datos originales.

## 3. El análisis de componentes principales (ACP)

El método de extracción de componentes principales que utilizaremos para analizar las estadísticas peruanas, también conocido como *transformación de Hotelling* o *transformación discreta de Karhunen-Loève*, involucra la manipulación de la base de coordenadas en la que se expresa un conjunto de series de datos. La selección de un nuevo arreglo de ejes ortogonales permite reexpresar los datos originales de tal forma que se cumplan ciertas propiedades útiles para el análisis, sin que se pierda información en el proceso[11].

---

[6] Pearson, K. "On lines and planes of closest fit to systems of points in space." *Philosophical Magazine*, ser. 6, 2, 1901, pp. 559-572.

[7] Spearman, C.H. "General intelligence objectively determined and measured". *American Journal of Psychology*, 15, 1904, pp. 201-293.

[8] Hotelling, H. "Analysis of a complex of statistical variables into principal components". *Journal of Educational Psychology*, 24, 1933, pp. 417-441 y 498-520.

[9] Thurstone, L.L. *Multiple Factor Analysis*. Chicago: University of Chicago Press, 1947.

[10] Kaiser, H.F. "The Varimax criterion for analytic rotation in factor analysis." *Psychometrika*, 23, 1958, pp. 187-200.

[11] En un contexto distinto, González y Woods (González, Rafael C. y Richard E. Woods. *Tratamiento digital de imágenes*, Wilmington: Addison-Wesley Iberoamericana S.A. y Ediciones Díaz de Santos S.A., 1996) presentan una interesante aplicación del ACP para el reconocimiento de formas y la compresión de imágenes digitalizadas. En efecto, algunas técnicas de análisis son compartidas por disciplinas realmente muy distintas!

Partiendo de una matriz formada por un conjunto de 'n' vectores de datos (asociados presumiblemente a sendas series estadísticas relativamente correlacionadas entre sí), el ACP permite construir una base de 'n' vectores ortogonales de la misma dimensión que los primeros. Mediante ponderaciones apropiadas, estos vectores base (llamados precisamente *componentes principales* del sistema) son capaces de dar origen a toda la información originalmente registrada en la matriz de datos observados, y viceversa. No obstante, la proporción de la varianza total original recogida por cada uno de los nuevos ejes rotados no es la misma. El primer eje recoge la mayor proporción de varianza posible, el segundo la mayor proporción posible sujeta a su ortogonalidad con el primero y de este modo sucesivamente. Así, mientras que los primeros componentes recogen una porción importante de la variabilidad original, los últimos prácticamente carecen de varianza. Este hecho da origen a la propiedad más importante de la transformación de Hotelling, a saber: que es posible reconstruir los datos originales con un error estadísticamente no significativo *sin necesidad de hacer uso de los 'n' vectores base en su totalidad*, utilizando tan sólo un subconjunto más pequeño, conformado por los primeros 'p' vectores base. En esencia, si el número 'p' resulta ser relativamente pequeño en relación al número 'n', se habrá conseguido una compresión efectiva de los datos originales desde el punto de vista del espacio dimensional en el que se expresa y analiza la información.

En la práctica econométrica usual, este procedimiento puede ser utilizado para minimizar el efecto de la multicolinealidad implícita en los datos, pues permite realizar regresiones equivalentes con regresores ortogonales entre sí[12]. Ya que se trata de una técnica reversible, resulta factible recuperar el valor de los parámetros estimados asociados al modelo teórico originalmente planteado a partir de los parámetros estimados del modelo ortogonal equivalente[13].

A continuación se presenta una breve derivación del procedimiento en términos formales. Tomemos como punto de partida una matriz de datos estadísticos observados, tal que contenga 'n' columnas y 't' filas. Comúnmente cada columna estará asociada a una variable, mientras que cada fila lo estará a una observación de la misma. En primer lugar, esta información será normalizada (sustrayendo a cada elemento la media de su correspondiente columna y luego dividiéndolo entre la desviación estándar de la columna original).[14]

Sea $X^*$ la matriz *(t x n)* de valores normalizados para las 'n' variables y A una matriz de dimensión *(n x n)* de coeficientes de ponderación. El producto matricial de ambos objetos dará lugar a una nueva matriz $Z^*$ de la misma dimensión que $X^*$, en la que cada columna corresponderá a una suma ponderada de las columnas de la matriz original[15]. Nótese que si la matriz A es invertible, la transformación será reversible *postmultiplicando* cada término de (1) por su inversa.

$$Z^* = X^* A \tag{1}$$

[12] Para un estudio más detallados de la eficacia del método de componentes principales con este propósito, puede revisarse, por ejemplo, Gurmu. S., p. Rilstone, y S. Stern. "Semiparametric Estimation of Count Regression Models." *Journal of Econometrics*, 88, 1, 1999, pp. 123-150.

[13] Otros métodos comúnmente utilizados para aminorar los efectos del problema de multicolinealidad son el análisis del coeficiente de inflación de varianza sugerido por Klein, (Klein, L. *An Introduction to Econometrics*. Englewood Cliffs: Prentice-Hall, 1962) o el estimador sesgado de regresión de cresta analizado por Judge, et al. (Judge, G., C. Hill, W. Griffiths, y T. Lee. *The Theory and Practice of Econometrics*. New York: John-Wiley and Sons, 1985).

[14] Nótese que esta transformación será reversible en tanto se conserven en una matriz auxiliar de *(n x 2)* elementos los valores calculados de la media y desviación estándar originales.

[15] El asterisco utilizado en la notación de la matriz Z se usa para subrayar el hecho de que sus columnas también se encontrarán normalizadas (es decir, tendrán media cero y varianza unitaria), al ser combinaciones lineales de columnas previamente normalizadas.

Por otra parte, si se fijan los ponderadores contenidos en A de tal modo que las columnas de la matriz Z sean ortogonales entre sí, se dirá que la matriz $Z^*$ contiene los *componentes principales* de la matriz A. Esta condición puede alcanzarse si, al calcular la matriz de varianza-covarianza de $Z^*$, se garantiza la obtención una matriz diagonal. Los elementos de la diagonal principal de esta matriz cuadrada mostrarían la varianza de cada componente principal, mientras que el resto de elementos (siendo nulos) revelaría la ortogonalidad entre cada par de componentes distintos.

Al calcular la matriz de varianza-covarianza de $Z^*$ puede verse que ésta equivale a:

$$\begin{aligned} Z^{*}{'}Z^{*} &= (X^{*}A)'(X^{*}A) \\ Z^{*}{'}Z^{*} &= A'X^{*}{'}X^{*}A \\ Z^{*}{'}Z^{*} &= A'V\ A \end{aligned} \qquad (2a - 2c)$$

Donde V representa la matriz de varianza-covarianza de $X^*$ o, equivalentemente, la matriz de correlaciones de X.

En este punto, puede verse que la condición de extracción de componentes principales de $X^*$ es congruente con el problema de diagonalizar la matriz simétrica V, puesto que al reexpresar (2c) se tiene:

$$D = A'V\ A \qquad (3)$$

Donde la matriz diagonal D representa la matriz de varianza-covarianza de $Z^*$. Se desprende de la teoría de diagonalización de matrices reales simétricas[16] que la matriz D contendrá en su diagonal principal los valores propios de la matriz V[17], mientras que la matriz A contendrá, en cada una de sus columnas, un vector propio de V asociado al valor propio contenido en la columna correspondiente de la matriz D[18].

Ya que los elementos de la diagonal de D corresponden a la varianza de los componentes principales del sistema, cabe subrayar dos propiedades importantes. La primera es que siendo la suma de los valores propios de una matriz igual a la traza de la matriz, la suma de los valores propios de V será igual a la traza de V, que es igual a la suma de las varianzas de los vectores en $X^*$. En consecuencia, será posible calcular la proporción de varianza explicada por cada componente principal dividiendo el correspondiente valor propio entre la suma de todos los valores propios. La segunda es que, de ordenarse descendentemente los valores propios de la matriz V, se llegará a un punto en el que la proporción de varianza explicada acumulada por los correspondientes componentes principales sea estadísticamente igual a la varianza total de las series observadas (traza de V).

Sobre la base de la última propiedad, puede concluirse que, si bien el proceso de inversión de la ecuación (1) resulta exacto al tomar los ‘n’ componentes principales, se puede conseguir una aproximación conveniente de la matriz $X^*$ mediante el uso de las matrices A y $Z^*$ truncadas ($A_p$ y $Z_p^*$ respectivamente en la ecuación 4b), en las que se incluyan únicamente las ‘p’ primeras columnas

[16] Nótese que si bien el problema de diagonalización en su forma general se enuncia como $D=A^{-1}\ V\ A$, el teorema de Schur permite afirmar que la matriz A es unitaria (ortogonal) y que, por lo tanto, puede diagonalizarse la matriz simétrica real V cumpliéndose que $A^{-1} = A'$.

[17] Asumiremos, por conveniencia, que los valores propios en D están dispuestos en orden descendente según su valor.

[18] Ya que el problema de extracción de vectores propios dispone, al menos, de un grado de libertad para fijar la escala de los componentes de vector (un número mayor de grados de libertad en el caso de valores propios de multiplicidad mayor a uno), se escogerán aquellos que formen una base ortonormal (es decir, vectores con norma igual a la unidad).

(descartando los componentes con aporte de varianza muy bajo o nulo). El producto que se muestra en (4b) tiene dimensiones *(t x p) (p x n) = (t x n)*, coincidentes con la matriz original de datos.

$$Z^{*} A' = X^{*}$$
$$Z_p^{*} A_p' \cong X^{*} \qquad (4a , 4b)$$

El valor del parámetro 'p' está bajo el control del modelador, aunque se dispone de reglas referenciales para su elección. Kaiser[19], por ejemplo, indica que se conserven solamente aquellos factores cuyos valores propios asociados sean mayores a la unidad. Análisis posteriores han encontrado que este criterio tiende a sobreestimar el número de factores[20].

## 4. Análisis de entropía en la información estadística peruana (1896-1995)

Para este análisis se tomó como base de datos referencial el conjunto de estadísticas presentado por Seminario y Beltrán[21]. La información estadística utilizada corresponde a un total de 63 series distintas (con correlación menor a la unidad para cualquier par de series) referidas al período de 100 años comprendido entre 1896 y1995[22].

El ACP muestra la existencia de cuatro componentes principales significativos de acuerdo con el criterio de Kaiser. En conjunto, permiten explicar en 95.6% de la *varianza total* observada en las series de datos. El Cuadro 1 detalla estos resultados.

Cuadro 1
**Descomposición de los datos en componentes principales**
**Información utilizada: años 1896 – 1995 (100 años)**

| Componente principal | Valores propios | | |
|---|---|---|---|
| | Valor propio | % de varianza explicada | % de varianza acumulada |
| 1 | 51.727 | 82.106 | 82.106 |
| 2 | 5.310 | 8.429 | 90.534 |
| 3 | 1.988 | 3.156 | 93.690 |
| 4 | 1.207 | 1.915 | 95.606 |
| 5 - 63 | -- | 4.394 | 100.000 |

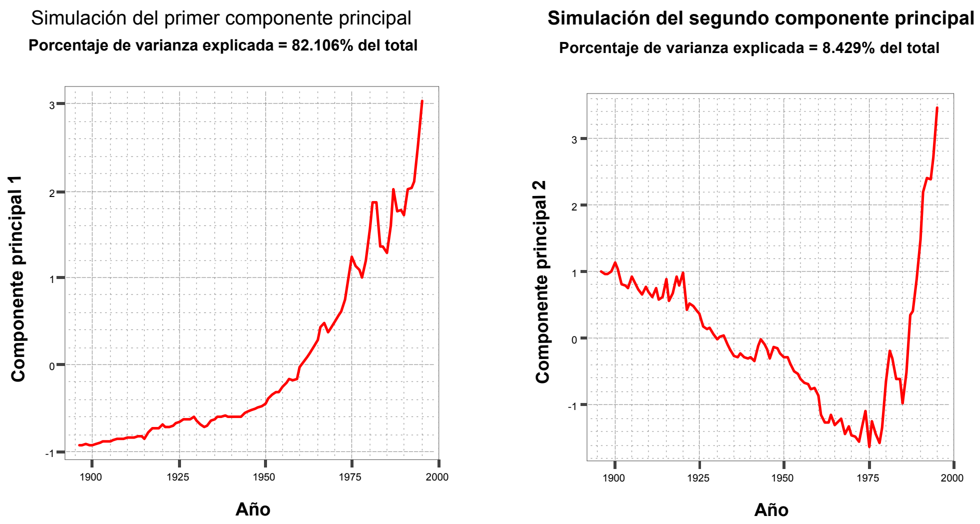
Los resultados deben interpretarse del siguiente modo. Es posible construir 63 series de datos sintéticas utilizando sendos vectores de ponderación (contenidos en la matriz 'A'). No obstante, basta tomar las tres primeras para explicar el 93.6% de la varianza observada en las 63 series de datos originales o las dos primeras para explicar el 90% de su varianza total. Esto implica que, si

[19] Kaiser, H.F. "The application of electronic computers to factor analysis". *Educational and Psychological Measurement*, 20, 1960, pp. 141-151.
[20] Pruebas alternativas han sido propuestas por Cattell (1966), Velicer (1976), Bartlett (1950, 1951), Horn (1965) y más recientemente por Gurmu, Rilstone y Stern (1999). Las referencias relevantes se incluyen en la bibliografía.
[21] Seminario, Bruno y Arlette Beltrán. *Crecimiento económico en el Perú: 1896-1995. Nuevas evidencias estadísticas*. Lima: Centro de Investigación de la Universidad del Pacífico, 1998, Serie Documento de Trabajo N°32.
[22] Cuatro variables del total de 67 series presentadas por Seminario y Beltrán fueron obviadas por no contar con observaciones completas o varianza suficiente para llevar a cabo el procedimiento ACP. Con el propósito de encontrar la mayor cantidad posible de componentes, se evitó hacer cualquier ajuste sobre las estadísticas consignadas, tales como la corrección por precios, naturaleza real o nominal, etc.

bien cada observación trae asociadas 63 dimensiones de análisis distintas (es decir, se representaría como un punto en un espacio de tal número de dimensiones), bastarían tres dimensiones para graficar dicha observación y aproximar las relaciones geométricas con el resto de observaciones con un error esperado de 6.31%, en promedio.

En el Gráfico 1 se pueden ver los dos primeros componentes principales calculados, expresados en la forma de series de tiempo. Se aprecia que el primero de ellos recoge prácticamente 10 veces la varianza que el segundo captura. Podría concluirse que las series de datos que están detrás de estos dos factores representativos tienen, en su mayor parte, una tendencia de crecimiento similar a la mostrada por el primer componente. Cabe destacar, sin embargo, que el segundo componente revela que algunas variables sufren un quiebre sensible en su tendencia entre los años 1975 y 1980. Este cambio puede ser positivo o negativo en función del signo de la ponderación que asuma el segundo componente en la serie específica[23].

Gráfico 1
**Componentes principales extraídos para la muestra completa**




Los hallazgos comentados en los párrafos anteriores demuestran ser estables en el tiempo. Al repetir el proceso de extracción de componentes para las cinco muestras que se exhiben en el Cuadro 2, se concluye que los cuatro primeros componentes principales recogen al menos 94% de la varianza total de las series examinadas.

[23] Conviene recordar que el efecto mostrado en los componentes se encuentra normalizado y que para llegar a la matriz de series X se debe pasar primero por la inversión a través de la matriz A' y luego la inversión de normalización efectuada para pasar de X a $X^*$. Esta última es particularmente importante, pues implica multiplicar cada valor por la desviación estándar de la serie originalmente observada.

Cuadro 2
**Descomposición de los datos en componentes principales**
**Comparación de resultados por subperíodos**

| Muestra tomada | % de varianza explicada acumulada por los primeros K componentes principales | | |
|---|---|---|---|
| | **K = 3** | **K = 4** | **K = 5** |
| 1896 – 1915 | 90.222 | 94.000 | 97.257 |
| 1896 – 1935 | 93.552 | 96.330 | 97.943 |
| 1896 – 1955 | 92.625 | 95.461 | 97.318 |
| 1896 – 1975 | 93.811 | 95.786 | 97.406 |
| 1896 – 1995 | 93.690 | 95.606 | 96.994 |
| **Mínimo** | **90.222** | **94.000** | **96.994** |

## 5. Comentarios finales

El análisis llevado a cabo demuestra la existencia de un número reducido y estable de componentes principales relevantes en las series económicas compendiadas por Seminario y Beltrán para la economía peruana durante un lapso de 100 años. No obstante, las series sintéticas generadas como resultado intermedio del procedimiento ACP ofrecen tanto respuestas como preguntas pendientes para futuros estudios. Constituyen respuestas en tanto abren la posibilidad de estimar modelos econométricos breves, libres del problema de multicolinealidad y ricos en información estadística. Pero presentan una interrogante abierta, en la medida en que inducen a un examen detenido para intentar explicar su comportamiento en el tiempo y obtener conclusiones sobre las implicancias de sus ciclos, tendencias y quiebres.

Finalmente, cabe señalar que a la luz de las conclusiones de este análisis preliminar, la adición de una mayor cantidad y variedad de series estadísticas como parte de la información de trabajo se presenta como una extensión natural de esta investigación. Los resultados de un estudio más completo prometen ser alentadores.

## 6. Referencias bibliográficas

Bartlett, M.S. "Tests of significance in factor analysis." *British Journal of Psychology*, 3, 1950, pp. 77-85.

__________ "A further note on tests of significance in factor analysis." *British Journal of Psychology*, 4, 1951, pp. 1-2.

Cattell, R.B. "The Scree test for the number of factors". *Multivariate Behavioral Research*, 1, 1966, pp. 245-276.

González, Rafael C. y Richard E. Woods. *Tratamiento digital de imágenes*, Wilmington: Addison-Wesley Iberoamericana S.A. y Ediciones Díaz de Santos S.A., 1996.

Gurmu. S., p. Rilstone, y S. Stern. "Semiparametric Estimation of Count Regression Models." *Journal of Econometrics*, 88, 1, 1999, pp. 123-150.

Horn, J.L. "A rationale and technique for estimating the number of factors in factor analysis." *Psychometrika*, 30, 1965, pp. 179-185.

Hotelling, H. "Analysis of a complex of statistical variables into principal components". *Journal of Educational Psychology*, 24, 1933, pp. 417-441 y 498-520.

Judge, G., C. Hill, W. Griffiths, y T. Lee. *The Theory and Practice of Econometrics*. New York: John-Wiley and Sons, 1985.

Kaiser, H.F. "The Varimax criterion for analytic rotation in factor analysis." *Psychometrika*, 23, 1958, pp. 187-200.

__________ "The application of electronic computers to factor analysis". *Educational and Psychological Measurement*, 20, 1960, pp. 141-151.

Klein, L. *An Introduction to Econometrics*. Englewood Cliffs: Prentice-Hall, 1962

Pearson, K. "On lines and planes of closest fit to systems of points in space." *Philosophical Magazine*, ser. 6, 2, 1901, pp. 559-572.

Seminario, Bruno y Arlette Beltrán. *Crecimiento económico en el Perú: 1896-1995. Nuevas evidencias estadísticas*. Lima: Centro de Investigación de la Universidad del Pacífico, 1998, Serie Documento de Trabajo N°32.

Spearman, C.H. "General intelligence objectively determined and measured". *American Journal of Psychology*, 15, 1904, pp. 201-293.

Thurstone, L.L. *Multiple Factor Analysis*. Chicago: University of Chicago Press, 1947.

Velicer, W.F. "Determining the number of components from the matrix of partial correlations." *Psychometrika*, 41, 1976, pp. 321- 327.