INVESTIGACIÓN & DESARROLLO, No. **7**: 5 – 24 (2007)
ISSN 1814-6333

DESCOMPOSICIÓN EN REGRESIÓN LINEAL: UN NUEVO MÉTODO PARA ANÁLISIS DE DETERMINANTES Y TOMA DE DECISIONES

Ernesto Cupé C.
Centro de Investigaciones Económicas y Empresariales (CIEE)
Universidad Privada Boliviana (UPB)
ecupe@upb.edu

RESUMEN

En el contexto de los modelos de regresión lineal, se presenta una novedosa metodología para la descomposición del indicador R^2 en términos de las variables independientes. La metodología se basa en la existencia de una base ortogonal de vectores singulares para el espacio generado por las variables independientes expresadas en términos de desviaciones respecto a su media. Esta metodología permite cuantificar el poder explicativo o la participación de cada una de las variables independientes en la explicación de la variación de la variable dependiente, por lo que en la práctica resulta extremadamente útil en el análisis de los determinantes de la variable dependiente y la toma de decisiones en cualquier campo donde se use un modelo lineal. Se presentan dos aplicaciones, una relativa a los determinantes de la decisión de distribuir dividendos en un conjunto de empresas, y otra relacionada con los determinantes del ingreso laboral en las zonas urbanas de Bolivia. Se pone especial énfasis en evaluar la nueva metodología respecto a la descomposición propuesta por G. Fields, como representante de metodologías de descomposición que admiten participaciones negativas por variable independiente, la cual se ha popularizado en los últimos años rápidamente aplicándose en diversos estudios a lo largo de todo el mundo. La nueva metodología de descomposición garantiza la no negatividad del poder explicativo directo de una variable, por una parte, y reconoce y cuantifica el efecto que genera la correlación entre variables independientes, por otra.

Palabras Clave: Regresión lineal, base ortogonal, vectores singulares, metodologías de descomposición.

1. INTRODUCCIÓN

Si bien existe una amplia literatura sobre métodos de descomposición en diversos contextos de modelización, curiosamente aún es tema de investigación el método de descomposición para el indicador más importante, R^2, del modelo más utilizado, el modelo de regresión lineal. Aparentemente, ya se ha estudiado todo respecto a este ya tradicional modelo y su aplicación ahora se facilita totalmente con el apoyo de una computadora y el software apropiado. La descomposición del poder explicativo del modelo, R^2, en términos de aportes por variable independiente es, sin embargo, un tema pendiente, teóricamente no resuelto y en la práctica muy requerido.

Como es sabido, en el contexto de los modelos lineales y bajo presencia de la constante entre los regresores, el poder explicativo del modelo se mide a través del indicador R^2. En un extremo, si este indicador es igual a la unidad, el conjunto de las variables independientes explica completamente a la variable dependiente; en el otro extremo, si el indicador es cero, el conjunto de variables independientes no explica nada de la variable dependiente. En la práctica de los modelos lineales, el primer indicador que se evalúa luego de una estimación del modelo es precisamente R^2; una vez que se ha obtenido un R^2 satisfactorio, la historia del R^2 habitualmente termina ahí.

La utilidad práctica de R^2, se incrementa notoriamente si puede descomponerse en términos de todas las variables independientes, cada una con su propia participación en la descomposición. Así, no solamente se sabría qué tan bien explica el modelo a la variable dependiente, sino qué tanto de dicha explicación se debe a cada una de las variables independientes, generando un ordenamiento de las variables independientes según su poder explicativo individual, ordenamiento puede ser aprovechado ya no para evaluar impactos ni para pronósticos, sino para la toma decisiones y la definición de políticas. Por ejemplo, a la hora de tomar decisiones, para un inversionista podría ser un factor decisivo saber que entre los determinantes de la distribución de dividendos en las empresas de su medio, la liquidez es el más importante.

Actualmente, se ha popularizado el uso de la metodología de descomposición propuesta por G. Fields, de la Cornell University, al punto que se la ha aplicado en diversos estudios a lo largo de todo el mundo. La descomposición de Fields es simple y directa; sin embargo, tiene algunos problemas metodológicos, como todas las propuestas previas de soluciones extremadamente simples a un problema no tan simple. En el presente estudio se señalan algunas limitaciones de metodologías que, como la de Fields, admiten la posibilidad que una variable independiente tenga poder explicativo negativo; luego, se propone una nueva metodología libre de dichas limitaciones. Las características de la nueva metodología se muestran empíricamente a través de dos aplicaciones, una a la distribución de dividendos empresariales y otra a los determinantes del ingreso laboral en las zonas urbanas de Bolivia.

El documento está organizado de la siguiente manera. En la siguiente sección se hace una revisión del indicador R^2, en la Sección 3 se presenta la metodología de descomposición propuesta por G. Fields; en la Sección 4 se presenta preliminarmente la idea subyacente a la nueva metodología; en la Sección 5 se desarrolla la nueva metodología. En la Sección 6 se presentan dos aplicaciones con datos reales y, finalmente, en la Sección 7 se presentan las conclusiones.

2. R^2 EN LA REGRESIÓN LINEAL

La linealidad del modelo

$$Y = X\ \beta + e \tag{2.1}$$

donde Y es un vector de observaciones de la variable dependiente y, X la matriz de observaciones de las variables independientes, e el vector aleatorio no observado, permite cuantificar el efecto sobre la variable dependiente de cambios en cualquiera de las variables independientes,

$$\frac{\partial\ y}{\partial\ x_i} = \beta_i \tag{2.2}$$

Si los cambios en algunas o todas las variables independientes x_i, componentes del vector x, se dan simultáneamente, el efecto sobre la variable independiente es una transformación lineal definida por

$$\frac{\partial\ y}{\partial\ \mathrm{x}} = \beta' \tag{2.3}$$

donde β' es la transpuesta del vector de coeficientes β. Así, si h_i es el cambio en la i-ésima variable independiente, el efecto simultáneo sobre la variable dependiente está dado por

$$\frac{\partial\ y}{\partial\ \mathrm{x}}(h) = \beta' h = \sum_{i=0}^{k} \beta_i\ h_i \tag{2.4}$$

Entonces, la linealidad del modelo permite descomponer el efecto de cambios simultáneos en dichas variables como suma de efectos por cambios aislados en cada variable independiente. Es importante notar que no existen efectos combinados generados por la interacción entre las variables independientes.

Genéricamente, la regresión lineal es un tema de la estadística aplicada debido a la presencia de un término estocástico no observado en el modelo lineal de regresión, lo que deriva en el interés por las propiedades estadísticas de los parámetros estimados y el cumplimiento de los supuestos estocásticos del modelo. Sin embargo, una de las propiedades principales de la regresión lineal es no estocástica: proporciona la mejor aproximación lineal de la variable dependiente en función de las variables independientes del modelo.

Dadas las variables dependiente e independientes y una vez establecida dicha aproximación lineal óptima, el interés se orienta al grado de aproximación o la calidad del ajuste; es decir, en qué tan bien la variable dependiente estimada, $\hat{Y}$, se aproxima a la variable independiente observada, Y. La forma tradicional de hacerlo es a través del indicador R^2 que, en presencia de una constante entre los regresores, toma valores en el intervalo [0, 1]. Cuanto más próximo a la unidad es el valor de R^2 mejor es el ajuste del modelo y cuanto más próximo a cero la calidad del ajuste empeora. Siguiendo la idea expresada en [2.4], resulta natural requerir por una descomposición similar para R^2.

Se asume que en el modelo [2.1] tiene T observaciones, la constante está entre los regresores y su parámetro es β_0; además, se tienen k variables independientes no constantes con vector de parámetros $\beta=[\beta_1,\ldots,\beta_k]'$ y cuyas observaciones por variable se registran como columnas de la Txk matriz X.

El modelo estimado a partir de [2.1] es

$$Y = \iota\,\hat{\beta}_0 + X\,\hat{\beta} + \hat{e} \tag{2.6}$$

donde ι es un vector Tx1 de unos, $\hat{\beta}_0$, β y $\hat{e}$ son estimaciones de los respectivos parámetros y vectores aleatorio. Con la matriz simétrica e idempotente

$$M_0 = I - \frac{1}{T}\iota\iota' \tag{2.7}$$

donde I es la matriz identidad, [2.6] se puede escribir como,

$$\tilde{Y} = \tilde{X}\,\hat{\beta} + \hat{e} \tag{2.8}$$

donde $\tilde{Y} = M_0 Y$ y $\tilde{X} = M_0 X$. Salvo una constante, la varianza de la variable dependiente puede expresarse como,

$$\tilde{Y}'\tilde{Y} = \hat{\beta}'\tilde{X}'\tilde{Y} + \hat{e}'\hat{e} \tag{2.9}$$

Expresando [2.9] en términos relativos a la varianza de la variable dependiente se tiene,

$$1 = \frac{\hat{\beta}'\tilde{X}'\tilde{Y}}{\tilde{Y}'\tilde{Y}} + \frac{\hat{e}'\hat{e}}{\tilde{Y}'\tilde{Y}} \tag{2.10}$$

de donde se sigue que el indicador R^2, que expresa la varianza explicada por el modelo en términos de la varianza de la variable dependiente, está dado por

$$R^2 = \frac{\hat{\beta}'\tilde{X}'\tilde{Y}}{\tilde{Y}'\tilde{Y}} \tag{2.11}$$

Además de la interpretación estadística en términos de varianza que tiene, el indicador R^2 admite también otras interpretaciones. Por ejemplo, en sentido geométrico es una medida del ángulo entre el vector de observaciones de la variable dependiente y el hiperplano generado por los regresores, todos medidos en términos de desviaciones respecto a su correspondiente media; de hecho, es igual al cuadrado del coseno de dicho ángulo. En sentido gráfico, R^2 es directamente proporcional al área entre la línea asociada a las observaciones de la variable dependiente y la línea de la variable dependiente estimada por el modelo.

Bajo cualquier interpretación, R^2 mide la bondad de ajuste entre la variable dependiente observada y su mejor aproximación lineal en términos de las variables independientes o regresores; la optimalidad de la aproximación está garantizada por el conocido Teorema de la Proyección, válido en espacios tan generales como los Espacios de Hilbert; por eso, ahora la pregunta es cuánto contribuye cada regresor en el logro de dicha aproximación.

3. LA DESCOMPOSICIÓN DE FIELDS

Gary Fields de la Cornell University, propone una sencilla y directa descomposición de R^2, características que han popularizado al método[1] y extendido su uso en diversas áreas y aplicaciones[2], particularmente en estudios sobre determinantes de la desigualdad en el ingreso[3].

[1] Una exposición de su propuesta se encuentra en Fields [4] y un programa para aplicarlo en Fiorio y Jenkins [6], también existe una rutina implementada en Stata.
[2] En [4], G. Fields hace referencia a una amplia serie de países para los que se han realizado estudios aplicando su metodología de descomposición.
[3] Ver por ejemplo Fields and Yoo [5].

La descomposición de Fields expresa R^2 como

$$R^2 = \sum_{i=1}^{k} \hat{\beta}_i \frac{\tilde{X}_i'\tilde{Y}}{\tilde{Y}'\tilde{Y}} \tag{3.1}$$

Así,

$$s(X_i) = \hat{\beta}_i \frac{\tilde{X}_i'\tilde{Y}}{\tilde{Y}'\tilde{Y}} \tag{3.2}$$

es la participación[4] de la i-ésima variable independiente en R^2 y su participación porcentual está dada por

$$p(X_i) = \frac{s(X_i)}{R^2} \tag{3.3}$$

Esta descomposición equivale a expresar la variación de la variable dependiente en términos no de las variaciones de las variables independientes, sino de las componentes de la variación de la misma variable independiente respecto a cada una de las variables independientes expresadas en términos de variaciones.

Si bien el método de descomposición propuesto por G. Fields es simple y directo, presenta dos serias limitaciones de alta importancia práctica. Una de ellas, cuando se manifiesta, dificulta la interpretación y la otra, curiosa y aparentemente, la facilita.

La primera limitación se refiere a que nada garantiza la no negatividad de las participaciones $s(X_i)$ y, dado que se trata de explicar una varianza, una participación negativa no tiene sentido. La participación $s(X_i)$ toma un valor negativo si el parámetro $\hat{\beta}_i$ y la correlación entre la variable independiente X_i y la variable dependiente Y tienen signos diferentes. Por otro lado, esta limitación del método da lugar a la posibilidad de la existencia de variables que individualmente expliquen más del 100% del R^2.

La segunda limitación se refiere a que la descomposición de Fields ignora el efecto combinado que tienen los regresores en la explicación de la varianza de la variable dependiente. El pasar por alto dicho efecto combinado se traduce en un sesgo de sobrestimación, generalmente en presencia de sólo participaciones no negativas, o subestimación, generalmente bajo presencia de alguna participación negativa, del poder explicativo de varianza atribuido a una o varias variables.

4. UNA PROPUESTA PRELIMINAR DE DESCOMPOSICIÓN ORTOGONAL

Es claro que la idea de una descomposición de R^2 es expresar la variación de Y solamente en términos de participaciones individuales de las X_i. Desafortunadamente, ello no es posible en general; con datos reales, dicha imposibilidad se origina en la presencia de algún grado de correlación entre las X_i.

Ante la naturaleza de la causa que genera esta limitación, de manera natural surge la idea de descomponer la variación de Y en términos de la variación de las X_i respecto a un conjunto de variables auxiliares, digamos Z_i, no correlacionadas u ortogonales entre sí. A fin de capturar las direcciones de mayor variación de las X_i, cada Z_j se puede determinar de modo que capture la dirección de mayor variación en el subespacio ortogonal al generado por los vectores Z_i previamente seleccionados. Esto nos lleva a elegir

$$Z_1 = \tilde{X}\alpha_1 \tag{4.1}$$

donde α_1 es un eigenvector unitario de la matriz $\tilde{X}'\tilde{X}$ asociado al mayor eigenvalor, luego se elige

---

[4] En [4], G. Fields expresa [3.2] en su forma equivalente: $s(X_i) = \dfrac{cov\,[X_i\hat{\beta}_i, Y]}{\mathrm{var}[Y]}$.

$$Z_2 = \tilde{X}\alpha_2 \qquad (4.2)$$

donde α_2 es un eigenvector unitario de la matriz $\tilde{X}'\tilde{X}$ asociado a su segundo mayor eigenvalor. Procediendo de esta manera se obtienen k variables ortogonales,

$$Z_i = \tilde{X}\alpha_i, \qquad i = 1, \ldots, k \qquad (4.3)$$

donde los eigenvectores α_i son ortogonales entre sí y están asociados a eigenvalores λ_i ordenados en forma descendente. La existencia de las α_i unitarias y ortogonales está garantizada por la simetría de la matriz $\tilde{X}'\tilde{X}$. Es claro que las Z_i están no correlacionadas, pues

$$\begin{aligned} Z_i' Z_j &= \alpha_i' \tilde{X}' \tilde{X} \alpha_j \\ &= \alpha_i' \lambda_j \alpha_j \\ &= \lambda_j \alpha_i' \alpha_j \\ &= 0 \qquad \text{para } i \neq j \end{aligned} \qquad (4.4)$$

Con Z la matriz cuyas columnas son los vectores Z_i y α la matriz conformada por los vectores α_t como columnas, a continuación se muestra que las variables auxiliares Z_i capturan toda la variación de las X_i ,

$$\begin{aligned} \sum^k Z_i' Z_i &= \text{traza}(Z'Z) \\ &= \text{traza}(\alpha' \tilde{X}'\tilde{X} \alpha) \\ &= \text{traza}(\tilde{X}'\tilde{X} \alpha\alpha') \\ &= \text{traza}(\tilde{X}'\tilde{X}\, I) \\ &= \text{traza}(\tilde{X}'\tilde{X}) \\ &= \sum_{i=1}^k \tilde{X}_i' \tilde{X}_i \end{aligned} \qquad (4.5)$$

donde $(\alpha'\hat{\beta})_i^2$ es el cuadrado del i-ésimo elemento del vector $\alpha'\hat{\beta}$. Así, tomando en cuenta que la media de cada Z_i es cero, se sigue que la variación total de ambos conjuntos de vectores es la misma.

En estas condiciones, considerando que,

$$\begin{aligned} \alpha' \tilde{X}'\tilde{X} \alpha &= D \\ Z'Z &= D \end{aligned} \qquad (4.6)$$

donde D es la matriz diagonal con las λ_i en la diagonal principal, es posible descomponer la variación de Y en términos de las Z_i :

$$\begin{aligned} \tilde{Y}'\tilde{Y} &= \hat{\beta}' \tilde{X}'\tilde{X} \hat{\beta} + \hat{e}'\hat{e} \\ &= \hat{\beta}' \alpha D \alpha' \hat{\beta} + \hat{e}'\hat{e} \\ &= \hat{\beta}' \alpha Z'Z \alpha' \hat{\beta} + \hat{e}'\hat{e} \\ &= \sum_{i=1}^k (\alpha'\hat{\beta})_i^2 Z_i' Z_i + \hat{e}'\hat{e} \end{aligned} \qquad (4.7)$$

donde $(\alpha'\hat{\beta})_i^2$ es el cuadrado del i-ésimo elemento del vector $\alpha'\hat{\beta}$.
La contribución de la i-ésima variable auxiliar en la determinación de R^2 es

$$S_{CP}(i) = (\alpha'\hat{\beta})_i^2 \frac{Z_i'Z_i}{\tilde{Y}'\tilde{Y}} = (\alpha'\hat{\beta})_i^2 \frac{\lambda_i}{\tilde{Y}'\tilde{Y}} \tag{4.8}$$

Cada uno de los términos de la descomposición [4.8] es no negativo y su suma reproduce R^2. Lamentablemente, esta descomposición sólo es posible en términos de las variables auxiliares y no de las variables originales; más aún, los vectores α_i, de dimensión k y que desempeñan un rol central en la descomposición, no tienen relación directa con las variables originales X_i. La propuesta que se presenta en la siguiente sección salva esta limitación.

5. UN NUEVO MÉTODO DE DESCOMPOSICIÓN

El aspecto novedoso en la propuesta de descomposición ortogonal de la sección anterior empieza cuando se trabaja con la matriz $\tilde{X}'\tilde{X}$ en lugar de la matriz $X'X$. Si bien el recurrir a variables ortogonales asociadas a la matriz $\tilde{X}'\tilde{X}$ resuelve el problema de descomposición de R^2 como suma de términos no negativos asociados a variables individuales, sin embargo no permite expresar la descomposición en términos de las variables originales o relacionarlas directamente con las variables originales. A fin de salvar esta limitación, es necesario trabajar con las matrices $\tilde{X}$ y $\tilde{X}'$ por separado.

La matriz $\tilde{X}$ puede interpretarse como la representación matricial de una transformación lineal de $\mathbb{R}^k$ en $\mathbb{R}^T$; a su vez, la matriz $\tilde{X}'$ es la representación matricial de una transformación lineal del espacio dual de $\mathbb{R}^T$, $\mathbb{R}^{*T}$, en el espacio dual de $\mathbb{R}^k$, $\mathbb{R}^{*k}$. Entre $\mathbb{R}^T$ y $\mathbb{R}^{*T}$ existe un isomorfismo natural con representación matricial la matriz identidad I_T, lo mismo que entre los espacios duales $\mathbb{R}^k$ y $\mathbb{R}^{*k}$ con representación matricial la matriz identidad I_k. En todos los casos se han considerado las bases canónicas de los espacios respectivos para la representación matricial.

Escribiendo,

$$\tilde{X}'\tilde{X} = I_k \tilde{X}' I_T \tilde{X} \tag{5.1}$$

se tiene que la matriz $\tilde{X}'\tilde{X}$ es la representación matricial de una transformación lineal de $\mathbb{R}^k$ en $\mathbb{R}^k$, resultado de la composición de cuatro transformaciones lineales que van de $\mathbb{R}^k$ a $\mathbb{R}^T$, de $\mathbb{R}^T$ a $\mathbb{R}^{*T}$, de $\mathbb{R}^{*T}$ a $\mathbb{R}^{*k}$ y de $\mathbb{R}^{*k}$ a $\mathbb{R}^k$. En ese contexto, se puede mostrar[5] que existen matrices U_{TxT}, V_{kxk} y D_{Txk} tales que

i) U y V, son matrices ortonormales. Las columnas de U y V forman bases ortonormales de $\mathbb{R}^T$ y $\mathbb{R}^k$, respectivamente.

ii) D es una matriz cuyos únicos elementos no nulos son los elementos de la fila y columna i, para i = 1,…,k (se assume que no existe multicolinealidad perfecta entre los regresores, de lo contrario se debe sustituir k por $r = rango(\tilde{X})$); dichos elementos tienen los valores d_{ii} positivos y se consideran ordenados en forma descendente.

iii) Con U_i y V_i las columnas i-ésimas de las matrices U y V, respectivamente, se tiene

$$\begin{cases} \tilde{X} V_i = d_{ii} U_i \\ \tilde{X}' U_i = d_{ii} V_i \end{cases} \quad \text{para } i = 1, \ldots, k \tag{5.2a}$$

$$\tilde{X}' U_i = 0 \qquad \text{para } i = k+1, \ldots, T \tag{5.2b}$$

iv) Las columnas de U capturan ortogonalmente las direcciones de mayor variación de las variables originales y columnas de $\tilde{X}$. Las columnas de V capturan ortogonalmente las direcciones de mayor variación de las filas de $\tilde{X}$.

v) $$\tilde{X} = UDV' \tag{5.2c}$$

---

[5] Este resultado es conocido como el *Teorema de Descomposición de Valor Singular de Matrices* y es válida para cualquier matriz. Ver Datta [3].

La expresión [5.2c] es la *Descomposición de Valor Singular* de la matriz $\tilde{X}$, los vectores U_i y V_i son los *vectores singulares* de $\tilde{X}$, asociados a los *valores singulares* d_{ii}.

5.1 DESCOMPOSICIÓN POR VARIABLES ORTOGONALES

A partir de la descomposición de valor singular de la matriz $\tilde{X}$ y [2.11], se tiene la siguiente descomposición ortogonal de R^2,

$$R^2 = \sum_{i=1}^{k} \frac{(\hat{\beta}'\tilde{X}'U_i)^2}{\tilde{Y}'\tilde{Y}} \qquad [5.3]$$

Entonces, las participaciones de las variables ortogonales U_i en R^2 están dadas por

$$s_i^{\perp} = \frac{(\hat{\beta}'\tilde{X}'U_i)^2}{\tilde{Y}'\tilde{Y}} \qquad \text{para} \quad i = 1, \ldots, k \qquad [5.4]$$

La expresión [5.3] descompone R^2 como suma de cantidades no negativas

$$R^2 = \sum_{i=1}^{k} s_i^{\perp} \qquad [5.5]$$

Cada participación $s_i^{\perp}$ también pueden interpretarse como la participación de las variables independientes en la explicación de la varianza de la variable dependiente a través de la variable auxiliar U_i.

En la práctica, el vector U_i eventualmente puede ser interpretado en términos de las variables originales, a través de su representación lineal en términos de las variables originales o de un análisis de correlación y, a partir de ello, la descomposición ortogonal [5.5] puede ser interpretada en relación a las variables independientes originales.

5.2 DESCOMPOSICIÓN POR VARIABLES INDEPENDIENTES

En un contexto de proyección ortogonal, el mismo que se aplica en una regresión lineal, en la anterior sección se ha reconocido la existencia de participaciones combinadas en R^2. Cada participación combinada se da entre variables independientes y no entre variables ortogonales, una variable ortogonal captura la participación de las variables independientes en R^2 y, por la naturaleza de la varianza, esa combinación se puede descomponer de a pares; más aún, las variables ortogonales también capturan la participación directa de cada variable independiente. Resulta natural, entonces, agrupar las participaciones de las variables independientes originales capturadas ortogonalmente y determinar las participaciones directas y combinadas de dichas variables.

A partir de la descomposición de R^2 en términos de las variables independientes dada por

$$R^2 = \sum_{i=1}^{k}\left(\sum_{j=1}^{k}\frac{(\hat{\beta}_i\tilde{X}_i'U_j)^2}{\tilde{Y}'\tilde{Y}}\right) + \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\left(2\sum_{l=1}^{k}\frac{(\hat{\beta}_i\tilde{X}_i'U_l)(\hat{\beta}_j\tilde{X}_j'U_l)}{\tilde{Y}'\tilde{Y}}\right) \qquad (5.6)$$

se tiene que la participación directa de la variable X_i en la explicación de la varianza de la variable dependiente está dada por

$$q_i = \hat{\beta}_i^2 \sum_{j=1}^{k}\frac{(\tilde{X}_i'U_j)^2}{\tilde{Y}'\tilde{Y}} \qquad (5.7)$$

el término $\hat{\beta}_i^2 \dfrac{(\tilde{X}_i'U_j)^2}{\tilde{Y}'\tilde{Y}}$ representa la parte de la participación directa de la i-ésima variable independiente capturada por la j-ésima variable ortogonal.

La participación combinada de las variables X_i y X_j es,

$$q_{i\,j} = 2\hat{\beta}_i\hat{\beta}_j \sum_{l=1}^{k} \frac{(\tilde{X}_i'U_l)(\tilde{X}_j'U_l)}{\tilde{Y}'\tilde{Y}}, \qquad i < j, \quad i = 1,\ldots,k-1, \quad j = 2,\ldots,k \tag{5.8}$$

donde $2\hat{\beta}_i\hat{\beta}_j \dfrac{(\tilde{X}_i'U_l)(\tilde{X}_j'U_l)}{\tilde{Y}'\tilde{Y}}$ representa la parte de la participación combinada de las variables independientes i-ésima y j-ésima capturada por la l-ésima variable ortogonal.

Las participaciones directas q_i son no negativas y junto a las participaciones combinadas q_{ij} descomponen R^2,

$$R^2 = \sum_{i=1}^{k} q_i + \sum_{i<j}^{i=k-1,\, j=k} q_{i\,j} \tag{5.9}$$

Las participaciones combinadas no necesariamente son no negativas. La presencia de participaciones combinadas negativas refleja la presencia en R^2 de efecto correlación de variables independientes y es un reconocimiento de la existencia de efecto cruzado que tienen las variables independientes en la descomposición de la varianza de la variable dependiente. No reconocer esta propiedad de los datos, como en la descomposición propuesta por Fields u otros métodos de descomposición que admiten participaciones directas negativas[6], implica que los efectos combinados son atribuidos y distribuidos como efectos directos.

5.3 DESCOMPOSICIÓN DIRECTA

Si bien la descomposición ortogonal, la participación de las variables independientes a través de las ortogonales y la interpretación de las variables ortogonales en términos de las variables independientes, proporcionan una radiografía clara de la composición de R^2 en términos de las variables independientes originales, eventualmente podría ser suficiente con la descomposición por variables independientes directamente. A continuación se presentan resultados para dicha descomposición directa que no hacen referencia directa a las variables ortogonales.

Tomando en cuenta que la matriz U es ortonormal, se tiene

$$\begin{aligned} \sum_{j=1}^{k} (\tilde{X}_i'U_j)^2 &= \sum_{j=1}^{k} (\tilde{X}_i'U_j)(U_j'\tilde{X}_i) \\ &= \tilde{X}_i'(\sum_{j=1}^{k} U_j\, U_j')\tilde{X}_i \\ &= \tilde{X}_i'(UU')\tilde{X}_i \\ &= \tilde{X}_i'\tilde{X}_i \end{aligned} \tag{5.10}$$

por lo que la participación directa de la variable X_i se puede expresar simplemente como

$$q_i = \hat{\beta}_i^2 \frac{\tilde{X}_i'\tilde{X}_i}{\tilde{Y}'\tilde{Y}} \tag{5.11}$$

Similarmente, se obtiene que la participación combinada de las variables X_i y X_j se expresa también como,

$$q_{i\,j} = 2\hat{\beta}_i\hat{\beta}_j \frac{\tilde{X}_i'\tilde{X}_j}{\tilde{Y}'\tilde{Y}}, \qquad i < j, \quad i = 1,\ldots,k-1, \quad j = 2,\ldots,k \tag{5.12}$$

Así, (5.11) y (5.12) permiten descomponer R^2 de una manera simple a partir directamente de los parámetros estimados y las variables independientes.

[6] Ver, por ejemplo, Morduch y Sicular [8].

6. APLICACIONES

En esta sección se presentan dos aplicaciones de la nueva metodología de descomposición y se comparan los resultados con los obtenidos aplicando la descomposición de Fields. Las diferencias y consideraciones respectivas valen para otras metodologías en las que se admiten participaciones negativas y se ignoran las participaciones combinadas.

6.1 DETERMINANTES DE DIVIDENDOS EMPRESARIALES[7]

Con el objeto de identificar los determinantes de la decisión de repartir dividendos, a partir de los balances anuales de un conjunto de 56 empresas se han generado 19 indicadores sobre rentabilidad y eficiencia, expansión y necesidad de fondos para expansión, estructura financiera, generación de fondos y nivel de endeudamiento, y liquidez.

TABLA 6.1 - DETERMINANTES DE DIVIDENDOS EMPRESARIALES
Variables Relacionadas

Clasificación	Variable	Descripción
Rentabilidad y Eficiencia	RE1	Valor agregado/Ventas
	RE2	Valor agregado/Inmovilizado Neto
	RE3	Utilidad antes de intereses e impuestos/Activo Total
	RE4	Utilidad antes de intereses e impuestos/Ventas
	RE5	Ventas/Activo Total
	RE6	Utilidad después de intereses e impuestos/Fondos Propios
Expansión y necesidades de fondos para expansión	ENFP1	Tasa de variación del activo total
	ENFP2	Tasa de variación del inmovilizado neto
	ENFP3	Tasa de variación de los fondos propios
	ENFP4	Tasa de variación de las ventas
Estructura financiera, generación de fondos y nivel de endeudamiento	EFGFD1	Deudas totales/Recursos propios
	EFGFD2	Deudas a corto plazo/Deudas a largo plazo
	EFGFD3	Utilidad antes de intereses e impuestos/Gastos financieros
	EFGFD4	Gastos financieros/Capacidad autofinanciación
	EFGFD5	Activo total/capacidad autofinanciación
Liquidez	L1	Activo circulante/Deudas a corto plazo
	L2	(Activo circulante-Existencias)/Deudas a corto plazo
	L3	Capacidad autofinanciación/Deudas a corto plazo
	L4	Deudas a corto plazo/Ventas
Ratio de Dividendos	DIV	Dividendos/Utilidad después de intereses e impuestos

Luego del análisis econométrico usual, se concluye que la distribución de dividendos, DIV, se explica por la tasa de variación de fondos propios, ENFP3, endeudamiento en términos de recursos propios, EFGFD1, capacidad de cobertura de gastos financieros, EFGFD3, y liquidez en términos de obligaciones por deudas a corto plazo, L2.

Los parámetros estimados tienen los signos esperados. Un incremento en la tasa de variación de los fondos propios se refleja en menor disponibilidad para dividendos, coeficiente negativo de ENFP3; una disminución de solvencia frente a deudas (incremento del ratio de deudas sobre recursos propios) se refleja en menores dividendos, coeficiente negativo de EFGFD1; aumentos de la capacidad de cobertura de gastos financieros se traducen en mayores dividendos, coeficiente positivo de EFGFD3; e incremento de la liquidez en términos de obligaciones de corto plazo por deudas, coeficiente L2 positivo.

La variable de mayor impacto individual sobre la distribución de dividendos es la liquidez, medida por L2, seguida por endeudamiento, medido por EFGFD1, capacidad de cobertura de gastos financieros, medido por EFGFD3, y variaciones en fondos propios, medido por ENFP3, en ese orden. Salvo en el caso del coeficiente de endeudamiento, la hipótesis nula de coeficiente nulo se rechaza a un nivel de significancia menor al 5%, incluyendo la constante.

[7] La base de datos y el enfoque en la elaboración de indicadores de esta aplicación se han tomado de Gonzáles [9] y Carrascal [2], respectivamente.

TABLA 6.2 - ESTIMACIÓN DEL MODELO DE DETERMINANTES DE DIVIDENDOS EMPRESARIALES

Dependent Variable: DIV
Method: Least Squares
Sample: 1 56
Included observations: 56

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.238354	0.111044	2.146477	0.0366
ENFP3	-5.97E-05	2.38E-05	-2.503850	0.0155
EFGFD1	-0.073515	0.041894	-1.754794	0.0853
EFGFD3	0.006368	0.003074	2.071525	0.0434
L2	0.259618	0.034094	7.614771	0.0000

R-squared	0.589202	Mean dependent var	0.461071
Adjusted R-squared	0.556982	S.D. dependent var	0.661584
S.E. of regression	0.440347	Akaike info criterion	1.282540
Sum squared resid	9.889202	Schwarz criterion	1.463375
Log likelihood	-30.91113	F-statistic	18.28713
Durbin-Watson stat	2.110723	Prob(F-statistic)	0.000000

En conjunto, las variables independientes explican el R^2 = 58.9% de la variación de los dividendos. A fin de realizar una evaluación comparativa de la descomposición propuesta por G. Fields y la nueva metodología en la determinación del poder explicativo de cada variable, se presentan a continuación los resultados de ambas descomposiciones.

Por simplicidad y en correspondencia con la notación empleada en el desarrollo general de este artículo, se adopta la siguiente notación:

X1 = ENFP3, X2 = EFGFD1, X3 = EFGFD3, X4 = L2

De acuerdo a la descomposición de Fields, Tabla 6.3, la liquidez es con mucho la variable con mayor poder explicativo relativo, 82%. Cada una de las demás variables tiene un poder explicativo relativo menor al 10% y en orden de importancia son endeudamiento, 8%, capacidad de cobertura de gastos financieros, 6%, y variación de fondos propios, 4%.

TABLA 6.3 - DETERMINANTES DE DIVIDENDOS EMPRESARIALES
Descomposición de Fields

Participación	Regresores				Total
	X1	X2	X3	X4	
s(Xi)	0.023	0.049	0.035	0.482	0.589
p(Xi)	4%	8%	6%	82%	100%

Por otra parte, la descomposición ortogonal identifica cuatro variables ortogonales ordenadas según las direcciones de mayor variación de las variables independientes. De acuerdo al Tabla 6.4, en este caso cada variable ortogonal está asociada a una variable independiente; U1 con la variación de fondos propios, U2 con la capacidad de cobertura de gastos financieros, U3 con liquidez y U4 con endeudamiento.

TABLA 6.4 - DETERMINANTES DE DIVIDENDOS EMPRESARIALES
Matriz de Correlación Variables Independientes vs. Variables Ortogonales

	X1	X2	X3	X4
U1	-1.000	0.116	0.062	-0.173
U2	0.000	-0.196	0.998	-0.062
U3	0.000	-0.424	0.000	0.947
U4	0.000	0.876	0.001	0.265

La descomposición ortogonal, Tabla 6.5, muestra que la variable U3 (altamente correlacionada con la variable de liquidez) concentra el 93% del poder explicativo relativo de la variación de la variable dividendos, seguida por las variables ortogonales U2 (altamente correlacionada con la capacidad de cobertura de gastos financieros), U1 (altamente correlacionada con la variación de fondos propios), y U4 (altamente correlacionada con endeudamiento).

TABLA 6.5 - DETERMINANTES DE DIVIDENDOS EMPRESARIALES Y DESCOMPOSICIÓN ORTOGONAL

Participación	Regresores ortogonales				Total
	U1	U2	U3	U4	
s'(Ui)	0.010	0.032	0.545	0.002	0.589
p(Ui)	2%	5%	93%	0%	100%

Considerando la correspondencia uno a uno que se da en este caso entre variables independientes y ortogonales, se observa que ambas descomposiciones describen una concentración del poder explicativo relativo en una sola variable, la liquidez. Sin embargo, parecen existir diferencias en el orden por poder explicativo relativo en las siguientes variables independientes. A diferencia de lo que establece la descomposición de Fields, donde endeudamiento tiene el segundo mayor poder explicativo, la descomposición ortogonal señala que la variable endeudamiento podría ser menos importante de lo que parece, de hecho podría ser la menos importante. Para ver esto con mayor detalle, pasamos a considerar la descomposición ortogonal en términos de las variables independientes.

En los Tablas 6.6 y 6.7 se presentan las participaciones directas y combinadas, respectivamente, en términos de las variables independientes. El total de las participaciones combinadas tiene signo negativo y junto al total de las participaciones directas reproduce R^2, las magnitudes de las participaciones combinadas son relativamente pequeñas, reflejo de baja colinealidad entre las variables independientes; más aún, prácticamente la totalidad de las participaciones directas de cada variable independiente son capturadas por sólo una variable ortogonal.

TABLA 6.6 - DESCOMPOSICIÓN ORTOGONAL Y DIVIDENDOS EMPRESARIALES
Participaciones Directas de Variables Independientes

	Participaciones Directas				Total
	X1	X2	X3	X4	
U1	0.053	0.000	0.000	0.015	0.068
U2	0.000	0.001	0.036	0.002	0.039
U3	0.000	0.005	0.000	0.447	0.452
U4	0.000	0.021	0.000	0.035	0.056
Total	0.053	0.027	0.036	0.499	0.615

La variable de liquidez es la que mayor efecto combinado muestra, particularmente con la variable de fondos propios, produciendo un efecto negativo de 0.056 puntos. Esta variable, fondos propios, también tiene efecto combinado con cada una de las demás variables, aunque el más significativo se da con liquidez.

TABLA 6.7 - DESCOMPOSICIÓN ORTOGONAL Y DIVIDENDOS EMPRESARIALES
Participaciones Combinadas de Variables Independientes

	Participaciones Combinadas						
	X1 X2	X1 X3	X1 X4	X2 X3	X2 X4	X3 X4	Total
U1	-0.009	0.005	-0.056	0.000	0.005	-0.003	-0.058
U2	0.000	0.000	0.000	0.012	-0.003	-0.017	-0.007
U3	0.000	0.000	0.000	0.000	0.093	0.000	0.093
U4	0.000	0.000	0.000	0.000	-0.054	0.000	-0.054
Total	-0.009	0.005	-0.056	0.012	0.041	-0.020	-0.023

En una presentación de resultados que omite la referencia a las variables ortogonales, Tabla 6.8 y Tabla 6.9, a los que se pueden llegar también por descomposición directa, se confirma la predominancia de la variable de liquidez en la explicación de la variación de la distribución de dividendos empresariales; sin embargo, también se confirma la importancia de la variable de fondos propios que resulta ser la segunda en importancia de poder explicativo relativo en términos de su participación directa, 9%. Las variables de capacidad de cobertura de gastos financieros y endeudamiento tienen similar poder explicativo de manera directa, aunque la de endeudamiento es levemente inferior.

TABLA 6.8 - DESCOMPOSICIÓN DIRECTA Y DIVIDENDOS EMPRESARIALES
Participaciones Directas de Variables Independientes

	Participaciones Directas				
Participación	X1	X2	X3	X4	Total
Q(Xi)	0.053	0.027	0.036	0.499	0.615
P(Xi)	9%	5%	6%	85%	104%

Las participaciones combinadas más importantes, se dan entre la variable de liquidez y las variables de fondos propios (signo negativo) y endeudamiento (signo positivo), reflejando la forma en que afectan conjuntamente, una vez descontada la participación directa, dichas variables en la explicación de la variación de los dividendos.

TABLA 6.9 - DESCOMPOSICIÓN DIRECTA Y DIVIDENDOS EMPRESARIALES
Participaciones Combinadas de Variables Independientes

	Participaciones Combinadas						
Participación	X1 X2	X1 X3	X1 X4	X2 X3	X2 X4	X3 X4	Total
Q(Xi)	-0.009	0.005	-0.056	0.012	0.041	-0.020	-0.026
P(Xi)	-1%	1%	-10%	2%	7%	-3%	-4%
R2	Participaciones Directas + Participaciones Combinadas						0.589
	Participaciones Directas + Participaciones Combinadas, en %.						100%

Los resultados del análisis por descomposición apoyan la idea de que las empresas quiebran por caja, no por utilidad; ello se refleja en la importancia predominante de la variable de liquidez entre los determinantes de la distribución de dividendos.

6.2 DETERMINANTES DEL INGRESO LABORAL (I)

El análisis por descomposición es particularmente útil en el estudio de los determinantes de desigualdad en el ingreso[8]. En esta sección se aplica el análisis por descomposición al estudio de los determinantes del ingreso laboral en el área

[8] Ver por ejemplo Arayama *et Al*. [1], Fields y Yoo [5], Morduch y Sicular [8], Salardi [10], Taiwo [11], Wan y Zhou, Zhangyue [13].

urbana de Bolivia, en el que solamente se considera a la población en edad de trabajar y cuya condición de actividad es de ocupados. Los datos provienen de la Encuesta de Hogares del año 2005[9], a partir de esta base se han elaborado varios indicadores como potenciales variables explicativas del ingreso laboral (w) [10].

Luego del análisis econométrico previo y usual se ha llegado a explicar el (logaritmo del) ingreso laboral en términos de las variables experiencia (Expercia), años de escolaridad (Aescola), condición de jefe de hogar (Jefe) y condición de asalariado (Asalrdo). Siguiendo la práctica generalizada en este tipo de modelos se ha incluido la variable experiencia al cuadrado (Expercia2)[11], cuyo coeficiente resulta ser altamente significativo al igual que los coeficientes de las demás variables independientes. El reporte de la estimación del modelo se presenta en el Tabla 6.10[12].

Los parámetros estimados tienen los signos esperados. Existe una relación directa entre el (logaritmo del) ingreso laboral y la experiencia, años de escolaridad, condición de jefe de hogar y de asalariado. En el caso del cuadrado de la experiencia, el signo es consistente con resultados de otros estudios con especificaciones similares[13].

TABLA 6.10 - ESTIMACIÓN DEL MODELO DE DETERMINANTES DE INGRESO LABORAL

Dependent Variable: LNW
Method: Least Squares
Sample: 1 3822
Included observations: 3822

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.113064	0.106135	19.90920	0.0000
EXPERCIA	0.109993	0.005573	19.73524	0.0000
EXPERCIA2	-0.001323	9.20E-05	-14.38130	0.0000
AESCOLA	0.098683	0.006838	14.43180	0.0000
JEFE	1.520367	0.059048	25.74791	0.0000
ASALRDO	1.480297	0.059168	25.01868	0.0000

R-squared	0.403215	Mean dependent var	6.039002
Adjusted R-squared	0.402433	S.D. dependent var	2.235389
S.E. of regression	1.728011	Akaike info criterion	3.933388
Sum squared resid	11394.66	Schwarz criterion	3.943197
Log likelihood	-7510.704	F-statistic	515.6519
Durbin-Watson stat	1.735586	Prob(F-statistic)	0.000000

En conjunto, las variables independientes explican el R2 = 40.3% de la variación del (logaritmo del) ingreso laboral. A fin de realizar una evaluación comparativa con la descomposición propuesta por G. Fields, se presentan a continuación los resultados de ambas descomposiciones. En correspondencia con la notación empleada en el desarrollo general de este artículo, se adopta la siguiente notación:

X1 = EXPERCIA,   X2 = EXPERCIA2,   X3 = AESCOLA,   X4 = JEFE,   X5 = ASALRDO

De acuerdo a la descomposición de Fields, Tabla 6.11, la condición de jefe de hogar es la variable con mayor poder explicativo relativo, 38%, seguida por la condición de asalariado, 29%, y la experiencia, 23%. Los años de escolaridad tienen un poder explicativo menor, 12%, y la experiencia al cuadrado tiene una participación negativa, -2%, en la descomposición de R^2.

[9] Encuesta de Hogares realizado por el Instituto Nacional de Estadística bajo el programa de Mejoramiento de Condiciones de Vida, MECOVI, correspondiente al año 2005.
[10] Los indicadores han sido elaborados por Carlos Foronda R., investigador del Centro de Investigaciones Económicas y Empresariales (CIEE) de la Universidad Privada Boliviana (UPB).
[11] Por ejemplo, en Wan y Zhangyue [13] se incluyen las variables Education y Education Squared, Age y Age Squared, en el marco de un modelo Mincer estándar.
[12] Debido al propósito ilustrativo de esta aplicación, no se discute el conocido problema de sesgo de selección en la estimación de este tipo de modelos.
[13] Ver, por ejemplo, Wan y Zhou [13].

TABLA 6.11 - DETERMINANTES DEL INGRESO LABORAL (I)
Descomposición de Fields

Participación	Regresores					Total
	X1	X2	X3	X4	X5	
s(Xi)	0.091	-0.009	0.048	0.154	0.119	0.403
p(Xi)	23%	-2%	12%	38%	29%	100%

Por otra parte, Tabla 6.12, la descomposición ortogonal muestra que solamente tres variables ortogonales de las cinco, son las que explican la variación del ingreso laboral. En particular, la variable ortogonal U4 tiene un poder explicativo de 55%.

TABLA 6.12 - DETERMINANTES DEL INGRESO LABORAL (I) Y DESCOMPOSICIÓN ORTOGONAL

Participación	Regresores ortogonales					Total
	U1	U2	U3	U4	U5	
s'(Ui)	0.000	0.075	0.105	0.222	0.0000	0.403
p(Ui)	0%	19%	26%	55%	0%	100%

La matriz de correlación entre las variables independientes y las variables ortogonales, Tabla 6.13, muestra que la variable ortogonal U1, que captura la dirección de mayor variación de las variables independientes, está altamente correlacionada con las variables independientes experiencia y experiencia al cuadrado, igual a la unidad con signo negativo con esta última; así, la variable ortogonal U1 es una variable de experiencia que prácticamente captura la participación de dos variables independientes en la descomposición de R2 y señala la posible redundancia en este sentido de una de las dos variables independientes. A su vez, la variable U3, aunque en menor grado, está altamente correlacionada con la variable de años de escolaridad. La correlación de las demás variables ortogonales se da con varias variables independientes.

TABLA 6.13 - DETERMINANTES DEL INGRESO LABORAL (I)
Matriz de Correlación Variables Independientes vs. Variables Ortogonales

	X1	X2	X3	X4	X5
U1	-0.947	-1.000	0.522	-0.103	0.256
U2	0.320	0.000	-0.148	0.223	-0.071
U3	-0.036	0.000	-0.840	-0.172	-0.166
U4	0.000	0.000	0.003	-0.713	-0.698
U5	0.001	0.000	0.000	-0.634	0.644

La posibilidad de redundancia entre las variables experiencia y experiencia al cuadrado se incrementa si se toma en cuenta que la correlación entre ellas es 0.95. En este punto, ya es necesario expresar la descomposición ortogonal en términos de las variables independientes, cuyos resultados se muestran en el Cuadro 6.14.

A diferencia de lo que ocurre en la descomposición propuesta por G. Fields, en la que la descomposición de R^2 no genera señales de alarma sobre la existencia de alta multicolinealidad (salvo por la existencia de una sin sentido participación negativa de la variable experiencia al cuadrado) que afecta a la descomposición, la descomposición ortogonal expresada en términos de las variables independientes muestra claramente la gravedad de la situación[14].

[14] Que no es detectada por el estadístico t de los respectivos coeficientes, pues aún con una correlación de 0.95 los coeficientes de ambas variables son altamente significativos.

TABLA 6.14 - DESCOMPOSICIÓN DIRECTA Y DETERMINANTES DEL INGRESO LABORAL (I)
Participaciones Directas de Variables Independientes

	Participaciones Directas					
Participación	X1	X2	X3	X4	X5	Total
Q(Xi)	0.632	0.328	0.047	0.115	0.110	1.232
P(Xi)	157%	81%	12%	28%	27%	305%

Una primera revisión de los resultados muestra que el poder explicativo directo de la variable de experiencia supera ampliamente el 100% y, a su vez, el poder explicativo directo de la variable experiencia al cuadrado es superior al 80%. Las participaciones directas de las otras variables en la descomposición se mantienen en rangos razonables, confirmando que la multicolinealidad se focaliza en las variables experiencia y experiencia al cuadrado.

TABLA 6.15 - DESCOMPOSICIÓN DIRECTA Y DETERMINANTES DEL INGRESO LABORAL (I)
Participaciones Combinadas de Variables Independientes

	Participaciones Combinadas										
Participación	X1 X2	X1 X3	X1 X4	X1 X5	X2 X3	X2 X4	X2 X5	X3 X4	X3 X5	X4 X5	Total
Q(Xi)	-0.862	-0.177	0.094	-0.136	0.130	-0.040	0.097	0.008	0.041	0.017	-0.828
P(Xi)	-214%	-44%	23%	-34%	32%	-10%	24%	2%	10%	4%	-205%
R2	Participaciones Directas + Participaciones Combinadas										0.403
	Participaciones Directas + Participaciones Combinadas, en %.										100%

El reporte de las participaciones combinadas, Cuadro 6.15, muestra que la alta multicolinealidad entre las variables de experiencia se manifiesta en extremadamente desproporcionadas participaciones combinadas en las relaciones que intervienen. En particular, la participación combinada entre experiencia y experiencia al cuadrado es negativa y supera el 200%.

Así, aunque en términos de la participación directa las variables de mayor poder explicativo del (logaritmo del) ingreso laboral son experiencia y experiencia al cuadrado, su alta participación combinada negativa relativiza dicho resultado inicial y señala la necesidad de repensar en la especificación del modelo.

6.3 DETERMINANTES DEL INGRESO LABORAL (II)

Debido a las señales de alarma puestas de manifiesto por la nueva metodología de descomposición, se ha excluido la variable experiencia al cuadrado de la especificación del modelo de ingreso laboral en las zonas urbanas de Bolivia. Los resultados de la estimación del modelo re-especificado, Tabla 6.16, muestran que los coeficientes estimados continúan siendo significativos y que R^2 se reduce levemente a 0.371.

En correspondencia con la notación empleada en el desarrollo general de este artículo, esta vez se adopta la siguiente notación:

X1 = EXPERCIA, X2 = AESCOLA, X3 = JEFE, X4 = ASALRDO

De acuerdo a la descomposición de Fields, Tabla 6.17, la condición de asalariado es la variable con mayor poder explicativo, 46%, seguida por la condición de jefe de hogar, 32%, años de escolaridad, 14% y experiencia, 8%. No se tienen participaciones negativas.

TABLA 6.16 - ESTIMACIÓN DEL MODELO DE DETERMINANTES DE INGRESO LABORAL (II)

Dependent Variable: LNW
Method: Least Squares
Sample: 1 3822
Included observations: 3822

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.594182	0.103405	25.08767	0.0000
EXPERCIA	0.035739	0.002154	16.58863	0.0000
AESCOLA	0.109146	0.006980	15.63692	0.0000
JEFE	1.670789	0.059661	28.00486	0.0000
ASALRDO	1.473719	0.060740	24.26269	0.0000

R-squared	0.370870	Mean dependent var	6.039002
Adjusted R-squared	0.370210	S.D. dependent var	2.235389
S.E. of regression	1.773989	Akaike info criterion	3.985645
Sum squared resid	12012.23	Schwarz criterion	3.993820
Log likelihood	-7611.568	F-statistic	562.5266
Durbin-Watson stat	1.704430	Prob(F-statistic)	0.000000

TABLA 6.17- DETERMINANTES DEL INGRESO LABORAL (II)
Descomposición de Fields

Participación	Regresores				Total
	X1	X2	X3	X4	
s(Xi)	0.030	0.053	0.118	0.170	0.371
p(Xi)	8%	14%	32%	46%	100%

Por otra parte, la descomposición ortogonal, Tabla 6.18, muestra que la participación se concentra en dos de las cuatro variables ortogonales; en particular, la variable ortogonal U4 tiene un poder explicativo de 67%.

TABLA 6.18 - DETERMINANTES DEL INGRESO LABORAL (II) Y DESCOMPOSICIÓN ORTOGONAL

Participación	Regresores ortogonales				Total
	U1	U2	U3	U4	
s'(Ui)	0.010	0.111	0.250	0.000	0.371
p(Ui)	3%	30%	67%	0%	100%

La matriz de correlación entre las variables ortogonales y las variables independientes, Tabla 6.19, muestra que cada una de las dos primeras variables está altamente correlacionada a una variable explicativa diferente, la variable ortogonal U1 con la variable experiencia y la variable ortogonal U2 con la variable de años de escolaridad.

TABLA 6.19 - DETERMINANTES DEL INGRESO LABORAL (II)
Matriz de Correlación Variables Independientes vs. Variables Ortogonales

	X1	X2	X3	X4
U1	-0.999	0.547	0.267	-0.167
U2	-0.043	-0.837	-0.165	-0.179
U3	0.000	0.003	-0.644	-0.774
U4	0.000	0.000	-0.698	0.584

Como consecuencia de las relaciones de correlación, la participación de cada una de las dos primeras variables independientes es capturada por la respectiva variable ortogonal correlacionada, Tabla 6.20; las participaciones de las variables dicotómicas de jefe de hogar y asalariado son capturadas por todas las variables ortogonales. El valor del total de participaciones directas es prácticamente igual al valor de R^2.

TABLA 6.20 - DESCOMPOSICIÓN DIRECTA Y DETERMINANTES DEL INGRESO LABORAL (II)
Participaciones Directas de Variables Independientes

Participación	Participaciones Directas				Total
	X1	X2	X3	X4	
U1	0.067	0.017	0.008	0.004	0.095
U2	0.000	0.040	0.003	0.004	0.048
U3	0.000	0.000	0.045	0.083	0.128
U4	0.000	0.000	0.053	0.047	0.100
Total	0.067	0.058	0.109	0.139	0.372

A diferencia de lo que ocurría con las participaciones combinadas en el modelo que incluía la variable experiencia al cuadrado, las participaciones combinadas en el modelo re-especificado no presenta valores desproporcionados; en particular, la participación combinada total no alcanza a 1% de R^2.

TABLA 6.21- DESCOMPOSICIÓN DIRECTA Y DETERMINANTES DEL INGRESO LABORAL (II)
Participaciones Combinadas de Variables Independientes

Participación	Participaciones Combinadas						
	X1 X2	X1 X3	X1 X4	X2 X3	X2 X4	X3 X4	Total
U1	-0.068	-0.045	0.032	0.023	-0.016	-0.011	-0.085
U2	0.004	0.001	0.001	0.022	0.027	0.007	0.063
U3	0.000	0.000	0.000	0.000	0.000	0.122	0.122
U4	0.000	0.000	0.000	0.000	0.000	-0.100	-0.100
Total	-0.063	-0.044	0.034	0.045	0.010	0.018	-0.001

Expresando en forma resumida los resultados de la descomposición ortogonal en términos de las variables independientes, se verifica que la descomposición de R^2 se reduce prácticamente a las participaciones directas; si bien existen participaciones combinadas positivas y negativas de alguna magnitud, éstas se compensan representando en el total menos del 1% de R^2[15].

TABLA 6.22 - DESCOMPOSICIÓN DIRECTA Y DETERMINANTES DEL INGRESO LABORAL (II)
Participaciones Directas de Variables Independientes

Participación	Participaciones Directas				Total
	X1	X2	X3	X4	
Q(Xi)	0.067	0.058	0.109	0.139	0.372
P(Xi)	18%	16%	29%	37%	100%

La participación combinada negativa generada entre las variables de experiencia y jefe de hogar se compensa con la positiva generada por la participación positiva generada entre las variables jefe de hogar y años de escolaridad; la participación combinada negativa generada entre las variables experiencia y años de escolaridad se compensan parcialmente con la participación combinada positiva generada por la interacción de la variable años de escolaridad con las variables jefe de hogar y condición de asalariado.

Por tanto, de acuerdo a la descomposición propuesta en el presente artículo, la variable independiente de mayor poder explicativo es la condición de asalariado, 37%, seguida de la variable de condición de jefe de hogar, 29%, la variable de experiencia, 18%, y la variable de años de escolaridad, 16%.

[15] Las participaciones porcentuales que se reportan en los cuadros 6.22 y 6.23 se han redondeado a unidades enteras.

TABLA 6.23 - DESCOMPOSICIÓN DIRECTA Y DETERMINANTES DEL INGRESO LABORAL (II)
Participaciones Combinadas de Variables Independientes

Participación	Participaciones Combinadas						Total
	X1 X2	X1 X3	X1 X4	X2 X3	X2 X4	X3 X4	
Q(Xi)	-0.063	-0.044	0.034	0.045	0.010	0.018	-0.001
P(Xi)	-17%	-12%	9%	12%	3%	5%	0%
R2	Participaciones Directas + Participaciones Combinadas						0.371
	Participaciones Directas + Participaciones Combinadas, en %.						100%

Finalmente, aún en una situación como la presente, con participaciones combinadas no significativas, las participaciones directas que se obtienen son diferentes según se aplique la metodología propuesta por G. Fields o la nueva metodología. En ambos casos, la condición de asalariado es la variable con mayor poder explicativo, pero en diferentes magnitudes; la descomposición de Fields le asigna una participación de 46% y la propuesta nueva le asigna una participación bastante menor, 37%.

En las demás variables, la diferencia no solamente implica cambios en magnitudes sino también cambios en el orden de importancia, Tabla 6.24[16]. En particular, según la descomposición de Fields existe una diferencia notoria entre el poder explicativo de la variable experiencia y el poder explicativo de la variable años de escolaridad, y se da a favor de años de escolaridad; según la nueva metodología, la diferencia es leve y a favor de la variable experiencia. Según la metodología de G. Fields, la condición de asalariado tiene un poder explicativo casi seis veces más que la variable de experiencia; en cambio, dicha relación es de dos veces según la nueva metodología.

TABLA 6.24 - DESCOMPOSICIÓN DIRECTA Y DETERMINANTES DEL INGRESO LABORAL (II)
Participaciones Combinadas de Variables Independientes

Método de Descomposición	Experiencia	Años de Escolaridad	Jefe de Hogar	Asalariado	Total
Métodología de G. Fields	8%	14%	32%	46%	100%
Metodología Nueva	18%	16%	29%	37%	100%

7. CONCLUSIONES

El análisis por descomposición de R^2 es extremadamente útil en la práctica, no solamente porque amplía considerablemente las posibilidades de análisis en el contexto de los modelos lineales, sino también porque posibilita la toma de mejores decisiones de acción al respaldarlas técnicamente.

Actualmente, se ha popularizado el uso de la metodología de descomposición propuesta por G. Fields y se la ha aplicado en diversos estudios a lo largo de todo el mundo. Un atractivo de la descomposición de Fields es que se obtiene de una manera simple y directa; sin embargo, este método de descomposición no garantiza la no negatividad de los componentes de la descomposición, por una parte, y no reconoce la existencia del efecto generado en la interacción entre regresores, por otra. La presencia de participaciones negativas de una variable en la descomposición de R^2, que mide varianza en términos relativos, carece de sentido; el efecto combinado que la descomposición de Fields ignora puede ser significativo, particularmente bajo presencia de alta multicolinealidad entre las variables independientes del

[16] Debido a que el total de las participaciones combinadas es prácticamente cero en este caso, el total de las participaciones directas es prácticamente 100%.

modelo, ignorar dicho efecto se traduce en considerable sub o sobre-estimación del poder explicativo de una o varias variables.

La nueva metodología de descomposición se basa en la existencia de una base ortogonal de vectores singulares para el espacio columna de la matriz cuyas columnas son precisamente las variables independientes expresadas en términos de desviaciones respecto a su media, la existencia de dicha base está garantizada por el Teorema de Descomposición de Valor Singular de Matrices. Dicha base define un conjunto de regresores ortogonales en las direcciones de mayor variación de las variables independientes.

Las variables ortogonales descomponen R^2 como suma de participaciones no negativas de cada una de ellas. Frecuentemente estas variables admiten una interpretación en términos de las variables independientes y la descomposición puede ser interpretada en esos términos; sin embargo, en la práctica, la principal utilidad de los regresores ortogonales es que muestran la estructura ortogonal de la participación directa y combinada de una variable independiente, la cual es una verdadera radiografía de la forma en que cada regresor participa en la conformación de R^2; metodológicamente, muestran que las participaciones negativas están asociadas a participaciones combinadas y no ha participaciones directas.

La nueva metodología, además de la participación directa de cada variable en la explicación de R^2, reconoce la existencia de participaciones combinadas que se generan en la correlación entre regresores. Esta metodología descompone R^2 como suma de participaciones directas de cada variable independiente y participaciones combinadas de los regresores. Las participaciones directas, en cada caso igual al poder explicativo directo de la respectiva variable, son siempre no negativas y dependen tanto del respectivo coeficiente estimado por la regresión como de la misma variable independiente; las participaciones combinadas se dan siempre entre dos variables independientes y pueden ser positivas o negativas en función de los respectivos coeficientes estimados por regresión y la correlación entre las variables.

La aplicación de la nueva metodología permite detectar la presencia de alta multicolinealidad, a veces no detectada por los estadísticos t de una regresión, en sentido que los coeficientes estimados pueden ser estadísticamente significativos aún bajo condiciones de alta multicolinealidad. La aplicación de análisis por descomposición a la explicación del ingreso laboral en las zonas urbanas de Bolivia es un ejemplo de ello. Como era de esperarse, la magnitud de las participaciones cruzadas está en función del grado de multicolinealidad entre las variables independientes. Multicolinealidad siempre existe en la práctica, pero en la medida que sea de menor magnitud, la descomposición se concentra en las participaciones directas de cada una de las variables independientes; una alta multicolinealidad se refleja en participaciones combinadas considerables y eventualmente desproporcionadas.

La descomposición en términos de las variables ortogonales puede expresarse en términos de las variables independientes de dos maneras. Una, agregando las participaciones directa y combinadas de las variables independientes capturadas por las variable ortogonales; otra, por descomposición directa. Este último procedimiento proporciona una forma simple y directa de cálculo, aunque sin la estructura explícita de la descomposición ortogonal de las participaciones directas y combinadas.

8. BIBLIOGRAFÍA

[1] Arayama, Yuko; Moo Kim, Jong; Kimhi, Ayal. *Determinants of Income Inequality among Korean Farm Households*. Economic Research Center. Discussion Paper No.161 November 2006.
[2] Carrascal, U.; Gonzáles, Y.; Rodríguez, B. *Análisis Econométrico con EViews*. Alfaomega-RaMa. México 2001.
[3] Datta, Biswa Nath. *Numerical Linear Algebra and Applications*. International Thomson Publishing Company. 1994.
[4] Fields, Gary S., *Regression-Based Decompositios: A New Tool for Managerial Decision-Making*, Departamente of Labor Economics, Cornell University. March 2004.
[5] Fields, Gary; Yoo, Gyeongjoon. *Falling Labor Income Inequality in Korea's Economic Growth: Patterns and Underlying Causes*. Review of Income and Wealth. Series 46, Number 2, June 2000
[6] Fiorio, Carlo V.; Jenkins, Stephen P. *ineqrbd: Regression-based inequality decomposition, following Fields (2003)*. UKSUG. September 2007
[7] Lebart, Ludovic; Morineau, Alain; Piron, Marie. *Statistique Exploratoire Multidimensionalle*. Dunod. Paris, 1995.
[8] Morduch, J.; Sicular, T. *Rethinking Inequality Decomposition, with Evidence from Rural China*. The Economic Journal 112:93-106. 2002.

[9] Pedraz Gonzáles, R. *Determinantes de la decisión de Repartir Dividendos*. Revista CEFGESTION. N° 14, España 1999.

[10] Salardi, Paola, *How much of Brazilian Inequality can we explain? An attempt of income differentials decomposition using the PNAD 2002*. Quaderni del Dipartamento di Economia Pubblica e Territoriale n. 1/2005.

[11] Taiwo, Awoyemi. *Explaining Income Inequality in Nigeria: A Regressio-Based Decomposition Using Household Data*. Department of Agricultural Economics. University of Ibadan, Nigeria.

[10] The World Bank. *Spatial Inequality in Vietnam: A Regression-based Decomposition*. 2003.

[12] Wan, Guang Hua. *Regression-based Inequality Decomposition: Pitfalls and a Solution Procedure.* World Institute for Development Economic Research. Discussion Paper No. 2002/101. 2002.

[13] Wan, Guanghua; Zhou, Zhangyue. *Income Inequality in Rural China Regression-Based Decomposition Using Household Data*. Review of Development Economics, 9(1), 107–120, 2005.