Estimación semiparamétrica de redes de alta dimensión para el análisis de Índices de Precio

Julián López Hernández

Resumen: Se considera el problema de estimación de la estructura de precios en una economía a través de una red no dirigida de alta dimensionalidad. Para modelar las relaciones de dependencia entre los precios, se emplean modelos gráficos gaussianos (GGMs), que definen un grafo $G = (V,E)$, donde el conjunto de aristas E describe relaciones de *dependencia condicional* entre pares de vértices $(i, j) \in V^2$. Dado que los GGMs dependen del supuesto de normalidad conjunta, no satisfecho por las series de precios, se recurre a una extensión semiparamétrica basada en el uso de cópulas gaussianas, conocido como *nonparanormal distribution* (Liu et al., 2009). Para tratar la alta dimensionalidad y obtener una inferencia *sparse*, se utiliza el algoritmo de estimación gráfica Lasso (*GLasso*). La metodología se aplica a un conjunto de datos del índice de precios al consumidor (IPC) de Estados Unidos, que cubre el período 2012-2023. La implementación se lleva a cabo en el software R, utilizando los paquetes *igraph* y *huge*.

Palabras claves: Redes complejas no dirigidas, Modelo Gráfico Gaussiano, Cópula Gaussiana, Lasso gráfico, IPC.

Problema de investigación: redes de correlación con umbral

Los modelos gráficos son una herramienta útil cuando se desea modelar la estructura de un conjunto de datos de alta dimensionalidad, es decir, donde el conjunto de variables d es igual o mayor al tamaño de la muestra n, $p \geq n$. Para su especificación se considera un vector aleatorio de dimensión d, $X = (X_1, X_2, \ldots, X_d)^T$, con n observaciones, sobre el cual se estima un grafo $G = (V,E)$, donde $V = \{1,\ldots,d\}$ representa el conjunto de vértices asociados a las d variables aleatorias del vector X, mientras que el conjunto de aristas E describe la existencia de una relación, *adyacencia*, entre un par de vértices, lo que se denota como $X_i \sim X_j$.

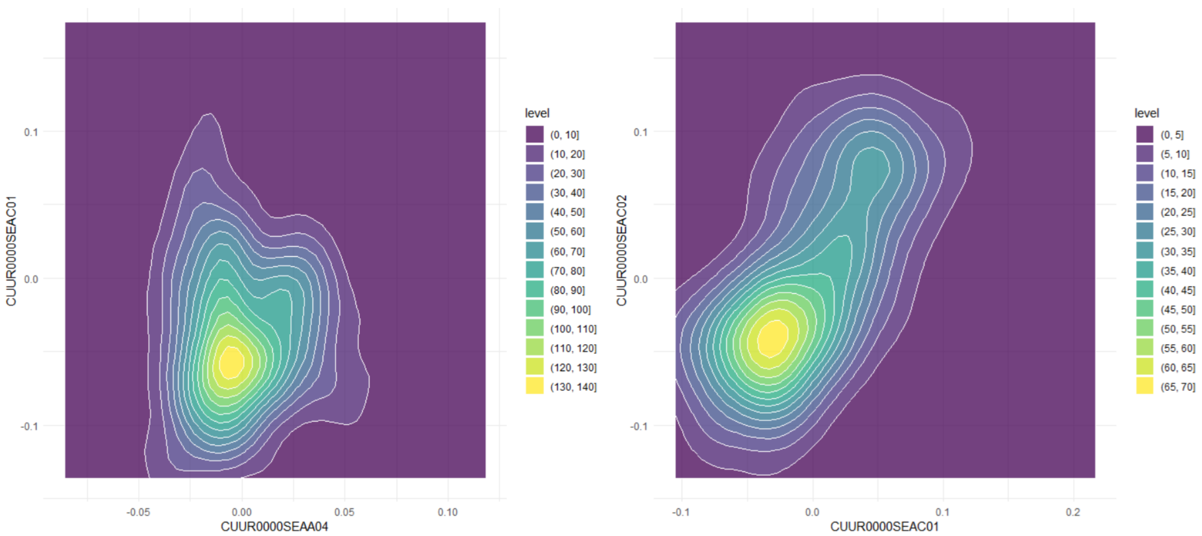
La implementación de modelos gráficos en el análisis de la dinámica de precios en una economía encuentra antecedentes importantes en los trabajos de Gao et al. (2013) y Sarantitis et al. (2018), que establecen las bases metodológicas para inferir las propiedades topológicas de las redes mediante *grafos de correlación* con umbrales. En este enfoque, el vector X representa el conjunto de bienes o clases de bienes dentro de una canasta de consumo, representados por el conjunto de nodos V, con n observaciones que corresponden a registros temporales de los índices de precios. El criterio de adyacencia E se determina utilizando el *coeficiente de correlación de Pearson* ($\hat{\rho}_{i,j}$), y la densidad de la red se controla estableciendo un umbral mínimo u, de manera que las conexiones E se definen como: $E = \{(i, j) \in V^2 : |\hat{\rho}_{i,j}| > u\}$

Dados los problemas de consistencia, reproducibilidad y comparabilidad asociados con este enfoque, en esta investigación se propone la implementación de *Modelos Gráficos Gaussianos*, que infieren relaciones mediante la correlación condicional entre pares de variables, controlando por las restantes. Formalmente, las conexiones se definen como: $E = \{(i, j) \in V^2 : \hat{\rho}_{i,j|V\setminus\{i,j\}} \neq 0\}$. Para controlar la densidad de la red, se emplea el algoritmo de penalización *Lasso*, que impone un criterio *sparsity* en la matriz de precisión, Ω, que nos informa de la dependencia condicional entre variables.

Implementación y extensión no paramétrica

Dado que los modelos GGMs exigen la normalidad conjunta de las variables X, se procede a evaluar el supuesto de normalidad multivariada con la *prueba de Mardia*, disponible en el paquete *MVN* de *R*, cuyos resultados nos sugieren rechazar la hipótesis de distribución normal. La figura 1 ilustra la no normalidad conjunta entre pares de variables X.

Figura 1. Gráficos de contorno para evaluación de normalidad conjunta



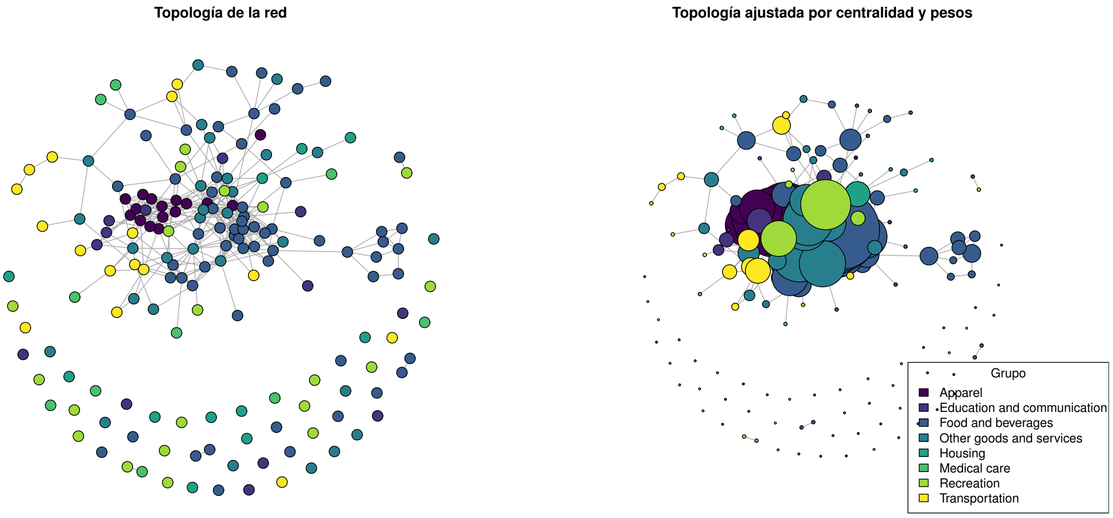
En respuesta a esta limitación, se explora una extensión semiparamétrica utilizando una cópula gaussiana con marginales no paramétricas, lo que se conoce como *nonparanormal distribution* (NPN) (Liu et al., 2009). El método de estimación transforma el vector aleatorio $X = (X_1,\ldots,X_d)$ en $f(X) = (f_1(X_1),\ldots,f_d(X_d))$, donde cada función f_i ajusta la marginal X_i. Tras esta transformación, el nuevo vector $f(X)$ sigue una distribución normal multivariada, es decir, $f(X) \sim NPN(\mu,\Sigma,f)$, aunque las marginales de X no lo sean. La existencia de dicha transformación está garantizada por el teorema de Sklar.

La transformación se realizó siguiendo el enfoque Liu et al. (2009), en el cual se aplica una transformación de rangos empíricos a cada marginal X_i, de modo que los valores transformados estén distribuidos uniformemente en el intervalo [0, 1]. Posteriormente, estos rangos se transforman mediante la función inversa de la distribución normal estándar Φ^{-1}, obteniendo así un vector con distribución normal multivariada, aunque las marginales originales no sigan una distribución normal.

Para la construcción del grafo, se utilizó la librería *huge* de *R*, especializada en el análisis de datos de alta dimensionalidad en la inferencia de redes. El algoritmo *graphical lasso (glasso)* se aplicó para estimar la matriz de precisión dispersa, permitiendonos identificar las relaciones condicionales entre las variables. Posteriormente, se utilizó la función *huge.select()* para ajustar el parámetro de regularización óptimo, λ, el cual se definió en función del criterio de estabilidad *StARS*.

Finalmente, el grafo se construyó con la librería *igraph*, para lo cual se utilizó el algoritmo *Fruchterman-Reingold*. La librería *igraph* nos permite representar graficamente el grado de centralidad de los vértices y los pesos de los enlaces, como se ilustra en la figuran 2.

Figura 2. Modelo gráfico de cópula gaussiana semiparamétrica



Referencias

Álvarez, E., Brida, J. G., Martínez, M., and Mones, P. (2022). Análisis de redes complejas: un estudio de la inflación en uruguay. *Revista Finanzas y Política Económica*, 14(1):131–166.

Alvarez, E., Brida, J. G., and Mones, P. (2021). Dinámica de la estructura de precios en uruguay. *Rect@: Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, 22(1):1–19.

Barrat, A., Barthelemy, M., and Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge university press.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Csardi, G. and Nepusz, T. (2006). The igraph software. *Complex syst*, 1695:1–9.

Gao, X., An, H., and Zhong, W. (2013). Features of the correlation structure of price indices. *PLoS one*, 8(4):e61091.

Kapçiu, R. and Kalluçi, E. (2024). Inflation study: Network analysis and complex networks implementation research article. In *IAI ACADEMIC CONFERENCE PROCEEDINGS*, page 9.

Kolaczyk, E. D. and Csárdi, G. (2014). *Statistical analysis of network data with R*, volume 65. Springer.

Lafferty, J., Liu, H., and Wasserman, L. (2012). Sparse nonparametric graphical models.

Liu, H., Han, F., Yuan, M., Lafferty, J., and Wasserman, L. (2012). High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326.

Liu, H., Lafferty, J., and Wasserman, L. (2009). The nonparanormal: semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 10(10).

Nelsen, R. B. (2006). *An introduction to copulas*. Springer.

Papadimitriou, T., Gogas, P., and Sarantitis, G. A. (2014). Convergence of european business cycles: A complex networks approach. *Computational Economics*, 47:97–119.

Sarantitis, G. A., Papadimitriou, T., and Gogas, P. (2018). A network analysis of the united kingdom's consumer price index. *Computational Economics*, 51:173–193.

Sun, Q., Gao, X., Wen, S., Chen, Z., and Hao, X. (2018). The transmission of fluctuation among price indices based on granger causality network. *Physica A: Statistical Mechanics and its Applications*, 506:36–49.

Tu, C. (2014). Cointegration-based financial networks study in chinese stock market. *Physica A: Statistical Mechanics and its Applications*, 402:245–254.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press.

Xiao, C., Ye, J., Esteves, R. M., and Rong, C. (2016). Using spearman's correlation coefficients for exploratory data analysis on big dataset. *Concurrency and Computation: Practice and Experience*, 28(14):3866–3878.