Estimadores aditivos de efectos

Víctor de Buen

27 de enero de 2007

Resumen

Se obtienen métodos estadísticos que miden los efectos en términos originales de las variables independientes en modelos estadísticos univariantes con filtro lineal, cuando la variable dependiente se introduce bajo transformación no lineal. Se hará especial hincapié en la transformación logarítmica bajo supuesto de normalidad y en el caso en el que se busca una descomposición aditiva de grupos de efectos. Se utilizará la nomenclatura típica del análisis de series temporales pero los resultados son válidos para cualquier tipo de magnitud de ordenamiento de los vectores. Se suponen conocimientos de regresión lineal y de optimización no lineal con restricciones.

Introducción La métrica en la que se introduce en un modelo la variable aleatoria dependiente, se debe escoger teniendo en cuenta la naturaleza de la magnitud observada, la forma en que las variables independendientes interactúan con la misma, y en virtud de criterios estadísticos como la homocedasticidad o el análisis de frecuencias de los residuos.

Sin embargo, a veces es necesario conocer en términos originales cuál es la influencia de cada input en particular, o de ciertos grupos de inputs que den una descomposición aditiva de la variable estudiada, para alcanzar así una visión más gráfica e intuitiva del comportamiento del modelo. De esto precisamente nos encargaremos a continuación.

Antecedentes En el caso general de modelos univariantes con filtro lineal se tendrá que

$$T\left(z_t\right) - \sum_{k=1}^{K} \omega_k x_{kt} = e_t \forall t = 1 \ldots N$$

donde

1. el output z_t o variable dependiente del modelo es un vector aleatorio; y

2. el ruido e_t, son realizaciones de una variable aleatoria de la que no se tiene control ni observación directa pero que se postula con cierta distribución de probabilidad que puede ser fija o variable en el espacio de índices y de la que se pueden deducir diferentes propiedades. Para su cálculo puede utilizarse un conjunto de parámetros adicionales que no vienen al caso en este estudio, como por ejemplo los coeficientes polinomiales si se tratara de un modelo ARIMA;

3. los inputs x_{kt} o variables independientes pueden ser tanto aleatorias como deterministas;

4. los parámetros del filtro lineal ω_k se estiman junto aquellos posibles parámetros que intervengan en el cálculo del ruido;

5. $T : D \subset \Re \rightarrow I \subset \Re$ es una función de transformación biyectiva en su dominio con función inversa conocida $T^{-1} : I \subset \Re \rightarrow D \subset \Re$;

Definiciones Supongamos que deseamos saber qué parte de z_t se debe exclusivamente al input x_{jt}. En primer lugar definiremos la variable aleatoria output filtrado del input j-ésimo z_{jt} correspondiente al valor del output si no existiera el efecto del input j-ésimo

$$T(z_{jt}) - \sum_{k \neq j} \omega_k x_{kt} = T(z_t) - \sum_{k=1}^{K} \omega_k x_{kt} = T(z_t) - \omega_j x_{jt} - \sum_{k \neq j} \omega_k x_{kt}$$

$$T(z_{jt}) = T(z_t) - \omega_j x_{jt}$$

$$z_{jt} = T^{-1}(T(z_t) - \omega_j x_{jt})$$

La diferencia con el valor real es la variable aleatoria buscada, o sea, el efecto o filtro en términos originales del input j-ésimo

$$f_{jt} = z_t - z_{jt} = z_t - T^{-1}(T(z_t) - \omega_j x_{jt})$$

Se define también el efecto o filtro conjunto en términos originales de un subconjunto de inputs de al menos dos elementos $J \subset K$ como

$$F_{Jt} = z_t - T^{-1}\left(T(z_t) - \sum_{i \in J} \omega_i x_{it}\right)$$

Por último, se define el efecto o filtro global de todos los inputs o simplemente filtro como

$$F_t = z_t - T^{-1}\left(T(z_t) - \sum_{i=1}^{K} \omega_i x_{it}\right)$$

Caso de la transformación logarítmica Por ejemplo, en el caso de la distribución logarítmica se tiene que el efecto en términos originales es

$$f_{jt} = z_t - z_{jt} = z_t - \exp(\ln(z_t) - \omega_j x_{jt}) = z_t(1 - \exp(-\omega_j x_{jt}))$$

$$F_{Jt} = z_t\left(1 - \exp\left(-\sum_{i \in J} \omega_i x_{it}\right)\right)$$

Cuando el valor del filtro transformado es uniformemente bastante más pequeño que 1 en valor absoluto, entonces su exponencial se puede aproximar como

$$\exp(-x) = 1 - x + O\left(x^2\right)$$

$$f_{jt} = z_t\left(\omega_j x_{jt} + O\left((\omega_j x_{jt})^2\right)\right)$$

Por este motivo a veces se interpreta el efecto en términos de porcentaje sobre la serie original, aunque se trata simplemente de una aproximación que no siempre es válida.

Descripción del problema general Obsérvese que si, como en el ejemplo anterior, la transformación es no lineal entonces las esperanzas de los efectos en términos originales no son aditivas, es decir, la esperanza del efecto unión de dos o más inputs no es igual a la esperanza de la suma de los efectos por separado

$$E[F_{Jt}] \neq \sum_{j \in J} E[f_{jt}]$$

y lo mismo le puede pasar a otros estimadores como el máximo verosímil. De forma general se puede exigir que se cumpla la propiedad aditiva con respecto a varios subconjuntos de al menos $J_s \forall s = 1 \ldots S$

Sin embargo se pueden construir otros estimadores de los efectos basados en estimadores de las realizaciones no observadas $\widetilde{\omega}_{jt}$ en cada posición t de los parámetros ω_j, de los cuales sí que tenemos al menos alguna aproximación de su distribución conjunta

$$\widetilde{f}_{jt} = z_t - T^{-1}\left(T\left(z_t\right) - \widetilde{\omega}_{jt} x_{jt}\right)$$

$$\widetilde{F}_{st} = \widetilde{F}_{J_s t} = z_t - T^{-1}\left(T\left(z_t\right) - \sum_{i \in J_s} \widetilde{\omega}_{it} x_{it}\right)$$

Buscaremos estos estimadores de forma que sean congruentes con la propiedad aditiva respecto a la familia de subconjuntos $\{J_s\}_{s=1\ldots S}$ o sea, que cumplan las propiedades

$$\widetilde{F}_{J_s t} - \sum_{j \in J_s} \widetilde{f}_{jt} = 0 \forall s = 1 \ldots S$$

al mismo tiempo que alcanzan alguna propiedad típica de los estimadores como que los $\widetilde{\omega}_j$ tengan máxima verosimilitud. Es decir se trata de añadir una lista de restricciones al método estadístico de estimación seleccionado.

Nótese que el número de restricciones S no puede superar al de inputs implicados $C = card\left[\bigcup_{s=1}^{S} J_s\right]$. Si se pretendiera conseguir un conjunto de estimadores completamente aditivos habría que incluir todos los subconjuntos de dos o más inputs, es decir, tendríamos $S = 2^K - K$ restricciones lo cual sería inviable en la práctica con apenas unos pocos inputs. Por lo tanto hay que ser cautos al elegir los subconjuntos de inputs sobre los que se pretende que se cumpla la aditividad.

Si los subconjuntos de inputs tienen una jerarquía o estructura arbórea, entonces se pueden añadir las condiciones de aditividad correspondientes a cada nodo de forma trivial sin aumentar demasiado la complejidad.

Una solución máximo-verosímil en caso de normalidad En este caso se tiene al menos una estimación de la media y de la matriz de covarianzas de los parámetros y en el caso de residuos normales los parámetros siguen una t de student multivariante que se puede aproximar por una multinormal:

$$\omega = (\omega_j) \ \ N\left(\widehat{\omega} = (\widehat{\omega}_j), \Sigma = (\sigma_{ij})\right) \wedge i, j \in \bigcup_{s=1}^{S} J_s$$

donde supondremos independencia entre los inputs, es decir: Σ es definida positiva y existe una transformación del tipo $\Sigma = L \cdot L^T$.

Nótese que se refiere a los índices implicados en la familia de subconjuntos de inputs que no tienen porqué ser todos los del filtro lineal. Resolveremos el problema para cada $t = 1 \ldots N$ por separado, para disminuir la complejidad aunque las variables no tienen porqué ser independientes, por lo que podemos prescindir de dicho índice. Tomando la transformación $\zeta = L^{-1}\left(\widetilde{\omega} - \widehat{\omega}\right)$ el problema es entonces el de maximizar la verosimilitud bajo el supuesto de normalidad es equivalente al de minimizar

$$\min \zeta^T \zeta$$

sujeto a las restricciones

$$\widetilde{F}_{J_s} - \sum_{j \in J_s} \widetilde{f}_j = 0 \forall s = 1 \ldots S$$

donde recordamos que

$$\widetilde{\omega} = \widehat{\omega} + L\zeta$$

$$\widetilde{f}_j = z - T^{-1}\left(T\left(z\right) - \widetilde{\omega}_j x_j\right)$$

$$\widetilde{F}_s = z - T^{-1}\left(T\left(z\right) - \sum_{i\in J}\widetilde{\omega}_i x_i\right)$$

Caso de la transformación logarítmica El problema en este caso se plantea del siguiente modo

$$\min_{\delta}\zeta^T\zeta$$

sujeto a las restricciones

$$z\left(1-\exp\left(-\sum_{i\in J}\widetilde{\omega}_i x_i\right)\right) = z\left(\sum_{j\in J_s}\left(1-\exp\left(-\widetilde{\omega}_j x_j\right)\right)\right)\forall s = 1\ldots S$$

que también se pueden escribir así

$$1-\exp\left(-\sum_{i\in J}\widetilde{\omega}_i x_i\right) = C_s - \sum_{j\in J_s}\exp\left(-\widetilde{\omega}_j x_j\right)\forall s = 1\ldots S$$

donde $C_s = card\left(J_s\right)$

Una variante del problema: descomposición aditiva del filtro global El caso más usual en la práctica, al que llamaremos descomposición aditiva del filtro global, es aquel en el que se tienen los inputs agregados parcialmente en efectos del mismo tipo y se desea que la suma de los efectos en términos originales sea igual al efecto conjunto total máximo-verosímil, es decir, se tiene una partición $\{J_u\}_{u=1\ldots U}$ de $\{1\ldots K\}$ y se ha de cumplir, prescindiento del subíndice t

$$\sum_{u=U}^{U}\widetilde{F}_u = \widehat{F}$$

donde

$$\widetilde{F}_u = \widetilde{F}_{J_u} = z - T^{-1}\left(T\left(z\right) - \sum_{i\in J_u}\widetilde{\omega}_i x_i\right)$$

$$\widehat{F} = z - T^{-1}\left(T\left(z\right) - \sum_{i=1}^{K}\widehat{\omega}_i x_i\right)$$

Es decir se da por buena la estimación conjunta para que no interfiera con la transformación máximo-verosímil del ruido del modelo y se modifican sólo las estimaciones de los grupos de inputs para que den una descomposición aditiva del filtro total en términos originales.

La solución máximo-verosímil en caso de normalidad En este caso los parámetros de los inputs pertenecientes al mismo grupo pueden modificarse solidariamente puesto que sólo se tiene en cuenta la suma de sus efectos. Puesto que los parámetros ω_j siguen aproximadamente una multinormal conocida, cualquier vector de combinaciones lineales de ellos es una multinormal conocida. En este caso definiremos

$$g = \ N\left(\widehat{g}, \Upsilon = (\gamma_{uv})_{u,v=1\ldots U}\right)$$

$$g_u = \sum_{i \in J_u} \omega_j x_i$$

$$\widehat{g}_u = \sum_{i \in J_u} \widehat{\omega}_j x_i$$

$$\gamma_{uv} = \sum_{i \in J_u} \sum_{j \in J_v} \sigma_{ij} x_i x_j$$

Ahora habría que minimizar las nuevas sumas de cuadrados, donde, por comodidad notacional se ha usado de nuevo la letra L para descomponer la nueva matriz de covarianzas $\Upsilon = L \cdot L^T$, que evidentemente es también definida positiva. Tomando la transformación $\eta = L^{-1}(g - \widehat{g})$ el problema es entonces el de minimizar

$$\min y(\eta) = \eta^T \eta$$

sujeto a la restricción no lineal

$$h(\eta) = \sum_{u=U}^{U} z - T^{-1}\left(T(z) - \left(\widehat{g}_u + \sum_{v=1}^{U} L_{uv}\eta_v\right)\right) - \widehat{F} = 0$$

Una solución aproximada Si no tenemos suficiente información o recursos de cálculo adecuados una solución aproximada de este problema sería repartir la discordancia, volviendo a poner los índices t para mayor claridad

$$d_t = \widehat{F}_t - \sum_{u=1}^{U} \widehat{F}_{ut}$$

de forma proporcional a una medida de dispersión s_u de las series F_{ut}, como por ejemplo podría ser

$$s_u = T^{-1}\left(\sqrt{\gamma_{uu}}\right)$$

y calculando los estimadores de los filtros como

$$\widetilde{F}_{ut} = \widehat{F}_{ut} + \frac{s_u d}{\sum_{v=1}^{U} s_v}$$

que evidentemente dan una descomposición aditiva

$$\sum_{u=U}^{U} \widetilde{F}_{ut} = \sum_{u=U}^{U}\left(\widehat{F}_{ut} + \frac{s_u d}{\sum_{v=1}^{U} s_v}\right) = \sum_{u=U}^{U} \widehat{F}_{ut} + d = \widetilde{F}_{ut}$$

Es decir modificamos más los filtros más variables y menos los más estables para alcanzar el objetivo de una forma muy rápida, aunque no muy robusta, ya que la medida de dispersión elegida no concuerda exáctamente con la correspondiente a la función de verosimilitud.

Descomposición aditiva del filtro global bajo transformación logarítmica En este caso los filtros en términos originales se definen como

$$\widetilde{F}_u = z\left(1 - \exp\left(-g_u\right)\right)$$

$$\widehat{F} = z\left(1 - \exp\left(-\sum_{u=1}^{U} g_u\right)\right)$$

La solución máximo-verosímil en caso de normalidad El problema es el de minimizar

$$\min y\left(\eta\right) = \eta^T \eta$$

sujeto a la restricción no lineal

$$h\left(\eta\right) = \sum_{u=1}^{U}\left(1 - \exp\left(-\widehat{g}_u - \sum_{v=1}^{U} L_{uv}\eta_v\right)\right) - \frac{\widehat{F}}{z} = 0$$

reescribible como

$$h\left(\eta\right) = U - \frac{\widehat{F}}{z} - \sum_{u=1}^{U}\exp\left(-\widehat{g}_u - \sum_{v=1}^{U} L_{uv}\eta_v\right) = 0$$

o en forma matricial

$$h\left(\eta\right) = \kappa - \mathbf{1}_U^T \chi\left(\eta\right) = 0$$

donde

$$\kappa = U - \frac{\widehat{F}}{z}$$

$$\mathbf{1}_U^T = \left(1 \ldots 1\right) \in \Re^{1\times U}$$

$$\chi\left(\eta\right) = \left(\chi_u\left(\eta\right)\right)_{u=1\ldots U} \in \Re^{U\times 1}$$

$$\chi_u\left(\eta\right) = \exp\left(-\widehat{g}_u - \sum_{v=1}^{U} L_{uv}\eta_v\right)$$

Utilizaremos también la notación matricial

$$\chi\left(\eta\right) = \exp\left(-\widehat{g} - L\eta\right)$$

en la que exponencial se aplica elemento a elemento del vector.

Al haber sólo restricciones de igualdad es lo mismo que minimizar sin restricciones la función lagrangiana

$$\min \mathcal{L}\left(\eta, \lambda\right) = y\left(\eta\right) + \lambda h\left(\eta\right)$$

y lo mismo que resolver el sistema de $U + 1$ ecuaciones no lineales

$$\triangledown_\eta \mathcal{L}\left(\eta, \lambda\right) = \triangledown_\eta y\left(\eta\right) + \lambda \triangledown_\eta h\left(\eta\right) = 0$$

$$h\left(\eta\right) = 0$$

Puesto que

$$\triangledown_\eta y\left(\eta\right) = 2\eta$$

$$\nabla_\eta h(\eta) = L^T \chi(\eta)$$

se tiene el sistema no lineal

$$2\eta + \lambda L^T \chi(\eta) = 0$$

$$\kappa - \mathbf{1}_U^T \chi(\eta) = 0$$

Obsérvese que la primera ecuación es en realidad una identidad vectorial de U ecuaciones mientras que la primera es una identidad escalar. Multiplicamos ambas ecuaciones por los premultiplicadores del término no lineal $\chi(\eta)$ en la otra ecuación para obtener el par de ecuaciones escalares

$$2\mathbf{1}_U^T L^{-T} \eta + \lambda \mathbf{1}_U^T \chi(\eta) = 0$$

$$\lambda\kappa - \lambda \mathbf{1}_U^T \chi(\eta) = 0$$

De la suma de ambas se deduce

$$2\mathbf{1}_U^T L^{-T} \eta + \lambda\kappa = 0$$

lo cual permite despejar

$$\lambda = \frac{-2\mathbf{1}_U^T L^{-T} \eta}{\kappa}$$

y sustituir en las ecuaciones originales para obtener el siguiente sistema no lineal

$$\kappa\eta - \left(\mathbf{1}_U^T L^{-T} \eta\right) L^T \chi(\eta) = 0$$

$$\kappa - \mathbf{1}_U^T \chi(\eta) = 0$$

Algoritmo A continuación se ofrece un método iterativo no diferencial diseñado *ad-hoc* para este problema. Obsérvese que si $\kappa = 0$ entonces la solución trivial es válida, es decir, $\eta = 0$. En otro caso se parte de una solución trivial de la segunda ecuación

$$\chi_u^{(1)} = \frac{\kappa}{U} \forall u = 1 \ldots U$$

y se ejecuta un ciclo cuya n-esima iteración sería en notación matricial

$$\eta^{(n)} = L^{-1} \left(-\widehat{g} - \ln\left(\chi^{(n)}\right)\right)$$

$$\eta^{(n+1)} = \frac{1}{k} \left(\mathbf{1}_U^T L^{-T} \eta^{(n)}\right) L^T \chi^{(n)}$$

$$x = \exp\left(\widehat{g} + L\eta^{(n+1)}\right)$$

$$\chi^{(n+1)} = \frac{\kappa}{U \mathbf{1}_U^T x} x$$

Obsérvese que las funciones logaritmo y exponencial se aplican elemento a elemento del vector.