# Covariables dependientes del tiempo: conceptos

Las covariables que varían aleatoriamente a lo largo del tiempo a menudo se denominan estocásticas, es decir, los valores de la covariable en cualquier ocasión no pueden predecirse con precisión, ya que están gobernados por un mecanismo aleatorio. Cuando una covariable varía en el tiempo y es estocástica, surgen nuevos problemas con respecto a la interpretación y estimación de los parámetros de regresión en modelos para datos longitudinales.

Para covariables estacionarias (no dependientes del tiempo):

Cuando una covariable es variable en el tiempo y estocástico puede que no necesariamente se mantenga. Por ejemplo, la suposición será violada cuando el valor actual de , dado , predice el valor posterior de . En ese caso

y se dice que confunde la relación entre y

En general, cuando esto no se cumple, los valores precedentes y / o posteriores de la covariable variable en el tiempo confunden la relación entre y ; esto puede llevar a estimaciones sesgadas de .

Se dice que una covariable variable en el tiempo es externa cuando los valores actuales y anteriores de la respuesta en la ocasión j (, ..., ), dados los valores actuales y precedentes de la covariable que varía con el tiempo (, ... ., ), no predice el valor posterior de . Más formalmente, una covariable que varía en el tiempo es externa (o a veces se denomina exogena) cuando

, ... ., , ..., , ... ., (1)

de lo contrario, se dice que la covariable es interna (o endogena).

Tenga en cuenta que cuando una covariable es externa,

,

En principio, es posible examinar la suposición de que una covariable variable en el tiempo es externa al considerar modelos de regresión para la dependencia de en , ..., (o alguna función (es) conocida (s) de , ..., ) y , ..., (o alguna función conocida de , ..., ). La ausencia de cualquier relación entre y , ..., , dado el perfil de covariable anterior, , ..., , proporciona soporte para la validez de la suposición de que el proceso de covariable es externo.

A los parámetros de regresión se les puede dar una interpretación causal solo cuando se puede asumir además que las covariables variables en el tiempo son externas con respecto a la variable de respuesta.

Si es distinto a el estimador de beta puede ser inconsistente del verdadero valor Beta.

**(Pepe and Anderson)**

Cuando hay covariables dependientes del tiempo, sugieren que los modelos marginales se estimen por ecuaciones de estimación generalizadas (GEE) con matriz de correlación independiente.

Aunque el estimador GEE no es necesariamente consistente con una estructura de correlación de trabajo arbitraria cuando hay covariables dependientes del tiempo, el estimador GEE es consistente con la correlación de trabajo independiente (R = I)

**(Tze Leung Lai y Dylan Small)**

Un componente clave en nuestro enfoque es una clasificación de covariables dependientes del tiempo en tres tipos.

**Tipo I**: clasificamos una covariable xj dependiente del tiempo como de tipo I si satisface

(1)

Una condición suficiente para que todas las covariables sean de Tipo I es

(2)

Para un modelo lineal, una condición suficiente para que una covariable xj sea de Tipo I es que

(3)

Las variables que satisfacen esto plausiblemente incluyen variables de edad, tiempo y la asignación de tratamiento para el sujeto i en el momento t en un ensayo aleatorizado cruzado.

**Tipo II**: clasificamos una covariable xj dependiente del tiempo como de tipo II si satisface

(4)

Tenga en cuenta que la clase de las covariables de tipo I es un subconjunto de la clase de las covariables de tipo II. Una condición suficiente para que todas las covariables sean del tipo II es que

(5)

Esta condición dice que el proceso covariable dependiente del tiempo no se ve afectado por la respuesta yit en el tiempo t condicional en xit.

Para un modelo lineal, una condición suficiente para que xj sea de Tipo II es que

(6)

**Tipo III**: Clasificamos una covariable xj dependiente del tiempo para que sea de Tipo III si no es de Tipo II, es decir,

(7)

Comparando GMM con GEE, cuando ambos métodos combinan las mismas ecuaciones de estimación, GMM tiene la misma eficiencia asintótica que GEE si la estructura de correlación de trabajo en GEE está especificada correctamente, pero GMM es más asintóticamente eficiente cuando la correlación de trabajo en GEE está mal escrito

Aunque GMM proporciona la flexibilidad para combinar las ecuaciones de estimación en (4), si algunas de las ecuaciones de estimación en (4) son inválidas para que la covariable sea de hecho de Tipo III, entonces el estimador GMM que combina las ecuaciones de estimación en (4) será inconsistente. En consecuencia, es importante poder probar si una covariable es de Tipo II vs. Tipo III (o Tipo I frente a Tipo II o III).

Nuestro enfoque general de GMM para el análisis de regresión marginal con covariables dependientes del tiempo es el siguiente. A menos que existan razones de peso para pensar que una covariable dependiente del tiempo es de Tipo I o Tipo II, asumimos que es de Tipo III y usamos las condiciones de momento (14) para ello. Si hay razones sustanciales para pensar que una covariable es de Tipo II (o Tipo I), entonces probamos la hipótesis nula de que es de Tipo II (o Tipo I) versus la alternativa de que es de Tipo III y si el test t no se rechaza, usamos las condiciones de momento (10) (u (8)) en nuestro estimador GMM. Llamamos al estimador GMM que elige qué condiciones de momento usar en base a la prueba (22) el estimador de Selección de Momento GMM. El estimador de Selección de Momento GMM gana eficiencia para covariables de Tipo I y Tipo II en comparación con GEE con la correlación de trabajo independiente cuando nuestra hipótesis de que una covariable es de Tipo I o Tipo II es correcta. El estimador de Selección de Momento GMM permanece consistente cuando nuestra hipótesis de que una covariable es de Tipo I o Tipo II es incorrecta siempre y cuando la prueba basada en (22) sea consistente.

**(Shrinkage)**

Nuestro método separa las funciones de estimación en dos grupos. Un grupo siempre se usa. En el contexto de este documento, este grupo corresponde a todas las funciones de estimación bajo el supuesto de correlación de independencia. En otras situaciones, este grupo también puede incluir la estimación de funciones que se sabe que son insesgadas a priori. El segundo grupo son aquellos que pueden mejorar la eficiencia asintótica de las estimaciones de los parámetros. Este grupo incluye todas las demás funciones de estimación, algunas de las cuales pueden ser informativas, pero algunas pueden ser no informativas o incluso estar sesgadas. Este grupo debe manejarse con delicadeza ya que la inclusión de funciones de estimación no informativas o sesgadas puede perjudicar el rendimiento.

A diferencia de los métodos de selección previos para identificar las funciones de estimación no informativas y sesgadas, creamos parámetros de contracción que reducen adecuadamente las funciones de estimación en este grupo, de acuerdo con la probabilidad de que cada una sea una ecuación de estimación sesgada, desinformativa o informativa. De esta forma, nuestro método resuelve el problema de alta dimensión de encontrar exhaustivamente el mejor subconjunto de funciones de estimación de todas las funciones de estimación de candidatos.

El método de Lai y Small (2007) requiere la prueba de cada covariable dependiente del tiempo para Tipo I, II vs. III. Como en cualquier prueba estadística, existe la posibilidad de falsos positivos, especialmente cuando el número de pruebas (covariables) es moderadamente grande, lo que puede ser el caso en muchas situaciones prácticas. Por lo tanto, su método no garantiza una mejora en la eficiencia sobre un GEE con una suposición de correlación de independencia.

En este trabajo, consideramos un método que tiene los siguientes atributos: (1) mejora de la eficiencia sobre un GEE con una suposición de correlación de independencia para ciertos tipos de covariables dependientes del tiempo, como los definidos por Lai y Small (2007) y (2) robustez contra las funciones de estimación sesgadas que podrían haberse incluido debido a una covariable dependiente del tiempo falsamente clasificada.

Llegamos a las siguientes conclusiones. Nuestro método funciona mejor cuando hay covariables de Tipo III y cuando el poder de la prueba que identifica las covariables de Tipo II versus III es bajo. Cuando los datos constan solo de covariables de Tipo II, su rendimiento es similar a otros estimadores que usan una prueba, todos son mejores que usar un GEE con la suposición de independencia.