

Dehejia&Wahba 2002

I Introduccion

En este paper se discutio el uso de métodos de emparejamiento del propensity score para corregir el problema del sesgo de las muestras seleccionadas dadas por las diferencias observables entre los tratados y los grupos de comparación.

El emparejamiento consiste en tratar a un individuo de control y a uno en tratamiento, pero que tengan características observables similares.

El paper se centro en los métodos de emparejamiento del propensity score ya que el problema de la dimensionalidad suele ser alta.

Con una cantidad pequeña de características el emparejamiento es mas directo, pero cuando hay muchas variables, es difícil de determinar las dimensiones a emparejar o que esquema de peso que adoptar.

El acercamiento tuvo 2 motivaciones.

1.-En algunos temas de interés, información en los resultados de la variable para los grupos de control pueden ser demasiado costosos.

2.-Incluso si la información del resultado esta disponible para todas las unidades de control, el proceso de búsqueda de un mejor subgrupo del grupo de control revela la extensión del overlap entre el grupo tratado el de control en términos de características pre tratamiento.

II Metodos de emparejamiento

a) El rol de la randomizacion

El problema básico en identificar el efecto causal es observaba ya sea bajo tratamiento o bajo control, pero nunca ambos. El efecto básico del tratamiento de interés bajo ajustes no experimentales es el efecto esperado del tratamiento para la población tratada.

Agregar Formula de Propuesta 1

b) Emparejamiento exacto en covariables

Cuando tenemos una falta de unidades de control para experimentar, asumimos que la información puede ser obtenida para un grupo unidades potenciales de comparación, que no son necesariamente sacadas de la misma población que los tratados , donde observamos el set de covariables X.

Propuesta 1: Comparar 2 individuos con las mismas características observables, uno que fue tratado y uno que no, la proposición 1 los compara en un experimento aleatorio.

En principio podemos categorizar la información en subgrupos, cada uno definido para un valor de X. Las limitaciones del método es que se basa en un grupo de comparacion suficiente grande para que ninguna unidad de tratado este sin una unidad de comparación

c) Propensity Score y reducción de la dimensionalidad

Rubin propone usar el propensity score para reducir la dimensionalidad del problema del emparejamiento.

añadir propuesta 3

III Algoritmos de emparejamiento según propensity score

En la discusión asumimos que la propensity score es conocida, pero en realidad no lo es.

Emparejar con propensity score es esencialmente un esquema de peso, lo que determina el peso en las unidades de comparación al estimar los efectos del tratamiento.

Este estimador sigue la propuesta 3. Las expectativas son reemplazadas por medias muestrales y condicionamos $p(X_i)$ al emparejar cada unidad del tratamiento i a un grupo de unidades de comparación j , con un propensity score similar. El objetivo se convierte en emparejar tratados con unidades de comparación cuyos propensity score son lo suficientemente cercanos para condicionar $p(x)$ a la propuesta 3 para ser válida.

Ocurren 3 problemas al emparejar:

1.- Si emparejar con un reemplazo: hacer esto reduce la distancia del propensity score entre tratados y el grupo de comparación, pudiendo emparejarse con el más cercano, incluso si una unidad de comparación ya fue emparejada. En contraste al emparejar sin reemplazo, cuando hay pocas unidades de comparación similares a las de los tratados podemos vernos obligados a emparejar unidades que son muy distintas entre ellas, lo que incrementa el sesgo, pero puede mejorar la precisión de los estimadores.

2.- Cuántas unidades de comparación deben ser emparejadas a las unidades de tratados: Al usar solo una unidad de comparación para cada unidad de tratamiento, se asegura la menor distancia del propensity score entre tratados y de control. Al usar más unidades de control, aumentando la precisión de los estimadores, pero al costo de aumentar el sesgo. Un método de seleccionar un grupo de unidades de control es el método del vecino más cercano, que selecciona a las unidades de control con una propensity score más cercana al de los tratados. Otro método es el emparejamiento según caliper, que usa todos los propensity score con un radio del propensity score.

Para emparejar sin reemplazo, consideramos low-high, high-low y son de emparejamiento aleatorio. En este método, los tratados están en un ranking.

Los individuos de ranking más alto se empareja primero, y el individuo de control emparejado se quita para emparejar más adelante.

Para emparejar con reemplazo se considera al único vecino más cercano y emparejamiento de caliper para un rango de caliper.

Además de usar el peso de en la diferencia de medias para estimar el efecto del tratamiento, también consideramos el peso de la regresión usando tratados y unidades de control emparejada. Una regresión puede mejorar la precisión de los estimadores.

Pero la pregunta es, cual método debemos seguir. Lo cual depende de la información que tengamos y en particular del grado del overlap entre los grupos de control y tratados en términos de propensity score.

Cuando hay un sustancial overlap en la distribución del propensity score entre control y tratados, la mayoría de los algoritmos de emparejamiento llevarán a resultados similares. Cuando los tratados y controlados son muy diferentes, encontrando un emparejamiento al emparejar sin reemplazos puede ser difícil. En particular, si solo hay un grupo pequeño de las unidades de control comparables a los tratados.

IV The data

A) El programa National Supported work

Era un programa federal que proveía experiencia laboral que habían enfrentado problemas económicas y sociales antes de enrolarse en el programa. Los candidatos para el experimento fueron seleccionados por un criterio elegible, y después se asignó (o no) al tratamiento.

B) Distribution of the treatment and comparison samples.

PSID -> population survey of income dynamics

CPS -> Current population survey

PSID y CPS Las unidades muestrales son 8 o 9 años más viejas que las del grupo NSW, su composición étnica y si tienen educación secundaria completa, mientras que el NSW eran dropouts de secundaria, y más, las ganancias pretratamiento son mucho más altas para los de control que de tratamiento.

V Resultados del emparejamiento

Fue aplicado a la muestra NSW-CPS. Las figuras 3 a 5 tienen en común que alrededor de las primeras 100 unidades tratadas están bien emparejadas a su contraparte de comparación.

En la figura 3, las unidades que fueron elegidas de manera aleatoria primero, obtuvieron mejores parejas, pero aquellos que lo hicieron después no están bien emparejados porque hay pocas unidades de comparación porque los individuos tratados ya fueron usados

De la misma forma en la figura 4 donde las unidades fueron emparejadas de menor a mayor, las unidades de comparación con propensity score altos están agotadas y las parejas son encontradas entre las unidades de comparación con estimadores mucho más bajos.

De manera parecida, al emparejar de mayor a menor, la calidad empieza a bajar después de las primeras unidades tratadas.

En la figura 6 es obtenida por el método de emparejamiento más cercano. Se nota que al emparejar con un reemplazo podemos evitar el deterioro de la calidad de las parejas. Podemos

notar que las líneas coinciden. Cuando es plana un individuo de control esta emparejada con varios tratados.

La muestras emparejadas están mas cerca de la NSW que todo el grupo CPS (grupo de comparación).

A diferencia del CPS, las muestras de emparejados sin reemplazos no son muy comparables con el NSW. Es razonable comparar en términos de edad, escolaridad y etnia, pero en términos de ganancias pre tratamiento, observamos grandes diferencias.

Como resultado, no es sorprendente que los estimadores del impacto del tratamiento, tanto en la diferencia de medias como en regresiones están lejos del punto de referencia experimental.

En conclusión, los métodos de emparejamiento según propensity score llevan a un estimador razonablemente certero del efecto del tratamiento.

Al seleccionar un subgrupo apropiado del grupo de comparación, una simple diferencia en medias lleva a un estimado del efecto del tratamiento cercano al punto de referencia experimental.

Se vuelve en una decisión importante entre métodos de emparejamiento cuando existe un overlap muy pequeño entre grupo tratado y de comparación.

Cuando hay un overlap minimo, el emparejar con reemplazo parece ser la mejor opción.

Cuando muchas unidades de comparación tienen overlap con el grupo tratado, emparejar sin reemplazo es una mejor opción

VI Testeo

a) Supuesto de la prueba del emparejamiento

La estructura especial de la información que usamos nos permite probar el supuesto que esta bajo el propensity score matching. Porque tenemos un grupo experimental de control y 2 grupos no experimentales de comparación, podemos probar el supuesto que, al condicionar el propensity score, las ganancias en un estado no tratado son independientes de la asignación del tratamiento.

Los estimadores sesgados serian las ganancias para el grupo de control experimental menos las ganancias para el grupo de comparación no experimental condicionado al propensity score estimado.

Tanto para el CPS y el PSID, vemos un rango de estimadores sesgados que son grandes para valores bajos del propensity score estimado. Ellos representan al grupo que eran menos probables de estar en el grupo de tratamiento.

Los estimadores sesgados no son estadísticamente significativos.

b) Probando la sensibilidad de la especificidad del propensity score

Una limitación potencial de los métodos del propensity score es tener que estimar el propensity score.

Incluso, un investigador que no tenga el beneficio “experimental benchmark estimate” preferirían estimadores que especifiquen completamente y de manera exitosa todas las covariables observadas.

VII Conclusiones

La calidad del estimador surge del resultado de comparar esta limitado por la calidad en general del grupo de comparación usado.

La conclusión del análisis es que es extremadamente valioso revisar si es comparable el tratamiento y unidades de comparación en términos de las características pre tratamiento.

El método del propensity score realmente remarca el hecho que la mayoría de las unidades de comparación son muy diferentes de las unidades tratadas.

VIII reflexiones

Fortalezas

Existe 2 grupos de comparación para la NSW, lo que puede ayudar a la investigación.

Posee una gran cantidad de formas de estimar los resultados.

Los métodos pueden estimar de manera acertada los efectos del tratamiento en un contexto no experimental

Posibles criticas

Un problema importante es si el supuesto de variables observables son validas o si el proceso depende de las variables que no son observables.

No todos los estimadores son realmente útiles.

En la práctica no se pueden hacer ciertas pruebas.