

۱) انتظار داریم عدالت رابطه منفی باشد. چون با تحصیل بیشتر، هزینه فرصت بچه دار شدن زیاد می شود و هر چه سال های تحصیل بیشتر شود، برای مادر کمتر صرفه جویی می شود. پس عدالت رابطه ای تدریجی سال های تحصیل با فرزندان منفی است.

۲) اگر $educ$ با متغیرهای دیگری که می تواند بررسی $children$ تاثیر داشته باشد همبستگی داشته باشد، ضریب $educ$ با تورش همراه است. برای مثال، هر چه $educ$ برای یک فرد زیاد شود، age هم زیاد می شود. و هر چه age زیاد شود

احتمال فرزند زایی بیشتر دارد. این تاثیر، باعث تورش مثبت می $educ$ می شود به طور مشابه در مورد $agebirth$ می توان گفت. هر چه $educ$ بیشتر باشد، $agebirth$

بیشتر است چون مادر در حال تحصیل بوده است. اما باز در $agebirth$ ، تعداد فرزندان ~~تورش~~ کمتر می شود. پس این تاثیر که تورش منفی می $educ$ دارد شاید مادران $urban$ ، تحصیل کمتری داشته باشند و از طرفی سن ازدواج کمتر و تعداد فرزندان بیشتری داشته باشند این هم تاثیر تورش منفی می $educ$ دارد.

۳) علیت معکوس: ما به دنبال تاثیر متغیر x روی y هستیم اما از طرفی y هم روی x تاثیر دارد. برای مثال شرایط بهداشتی و حقوق فرد را در نظر بگیرید. اگر حقوق فرد زیاد باشد، دسترسی بهتری به امکانات بهداشتی دارد و بهداشت بهتری دارد. از طرفی با داشتن شرایط بهداشتی بهتر، سلامتی جسمانی شخص برای کم کردن بیشتری شود یا مثال دیگر، رابطه ای قیمت و تعداد در رابطه عرضه و تقاضا است.

این باعث می شود ضریب x در رابطه ای y از طریق رگرسیون دچار تورش شود و مقدار اشتباهی معادله شود.

باید دنبال راهی باشیم که مادران با فرزند بیشتر به دلایلی، y به x بیشتری داشته باشند مثلاً در بیمارستان قبل از زایمان، ۱ سال تحصیل کنند (۱) اما همین حالتی نداریم پس این مسئله علیت معکوس ندارد.

د) اگر $electric$ ~~درآمد~~ درآمد رابطه ای مثبت داشته باشد که دارد ✓، $electric$ زمانی که درآمد کنترل شود، در تعداد فرزندان تأثیر نداشته باشد و بقیه x ها رابطه ای با y نداشته باشند (اثر در زمان داشته باشد) و بقیه متغیرهای پیشگو، در رابطه ای بین درآمد و $electric$ تأثیری نداشته باشند (یعنی فردی که y به x بیشتری دارد، احتمال ~~درآمد~~ داشتن بیشتری نداشته باشد.)

~~که شرایط برای برقرار داشتن رگرسیون است~~
که شرط آخر برقرار نیست یعنی هرکس که درآمد بیشتری دارد احتمال y به x بیشتری دارد پس $electric$ نمی تواند به عنوان $proxy$ استفاده شود.

ح) در این صورت باید اندازه ی ~~ضریب~~ ضریب به صفر نزدیک تر باشد. یا توجه به این که ضریب منفی است، ضریب باید نزدیک تر از واقعیت بدست آید و تورش ضریب مثبت است.

پ) برای مثال فرض کنید در یکی از استان ها شایع شود که عوامل دریافت آمار، مربوط به اداره ی مالیات هستند و اگر تعداد فرزندان را بیشتر اعلام کنند به ضررشان است.

و در آستانه دیگر شایع شود این مأموران از سازمان مربوط به یارانه هستند و اگر فرزندان خود را بیشتر گزارش کنند، یارانه بیشتری می گیرند. در این صورت گزارش مثبت یا منفی در هنگام گزارش children محدود دارد که در نقشه ~~children~~ $E(U, U_Z)$ در آستانه.

از طرفی دو گزارش هم به دلیل رسوم یا دین، می تواند گزارش مفروش شود.

راه حل استفاده از رگرسیون cluster شده ری متغیر آستانه است. در این صورت

فرض می شود ماتریس زیر، بلوکی قطری است:

$$E(U, U_Z) = \begin{pmatrix} (& 0 & 0 \\ 0 & (& 0 \\ 0 & 0 & () \end{pmatrix}$$

که هر بلوک مربوط به یک آستانه است.

(9)

متغیر ابزاری باید 3 شرط داشته باشد: - هم بستگی بالایی به educ داشته باشد.

- از طریق دیگری به جز educ نتواند روی children تاثیر داشته باشد.

شرط اول: به نظر که در دوس اشاره شد، احتمالاً دانش آموزان تا سن 16 سالگی مجبور به تحصیل

هستند و در این صورت دانش آموزان با $firsthalf = 1$ ، educ بیشتری دارند.

پس educ و firsthalf همبستگی بالایی دارند. (این درست است و معنی دایضرب هم دیده می شود.)

شرط دوم: $firsthalf$ نمی تواند از طریق دیگری روی تعداد بچه ها اثر گذار باشد.

پس متغیر ابزاری خوبی است.

h) آموزش ناشی از ~~تفاوت~~ واریانس اندازه گیری را در حالت بررسی می کنیم:

$$Plm \beta_{IV} = \beta + \frac{Cov(Z, u)}{Cov(x, u)} \frac{\sigma_u}{\sigma_x}, \quad Plm \beta_{OLS} = \beta + Cov(x, u) \frac{\sigma_u}{\sigma_x^2}$$

$$\Rightarrow \frac{Cov(Z, u)}{Cov(x, u)} < Cov(x, u) \Rightarrow \boxed{Cov(Z, u) < Cov(x, u)^2}$$

Z: ابزاری
x: اصلی
u: باقی مانده ها

ضریب u در ستون 4، مقدار بزرگتری دارد. این بدین دلیل اثر متغیرهای معذبانی

مثل درآمد در ستون 4 به خوبی از بین برده شده و نتیجه قابل اعتماد تر است.

در ستون 4: از نظر آماری معنی دار $\Rightarrow 3.67 > 2$

$$t\text{-score} = \frac{-0.237}{0.0646}$$

لپس وایت نشان داده می شود. ✓

23-34 مورد مناسبی برای یافتن اثر آلودگی نیست. چون شاید افراد مراجعه کننده در مناطق آلوده تر،

بیماران آسیب پذیرتر یا سالم تری داشته باشند. برای مثال شاید مسافران و افراد بارنقشه تر جمع دهند

که در مناطق با آلودگی کمتر سکونت کنند و مثلاً به الواسان می روند. در این صورت کو بایاس

در مراجعات منطبق کمتر آلوده وجود دارد که ناشی از تفاوت نوع بیماران در دو منطقه است.

فقط برای زمان ۰:

$$Y_{im0} = \alpha + \gamma \delta_m + e_{im}, \quad \delta_m = \begin{cases} 0 & m \in \text{Control} \\ 1 & m \in \text{treatment} \end{cases}$$

Control: کم آلوده ها
Treatment: پر آلودگی ها

اگر: حذف باشد. اعظم T است.

$$E[Y_{ic0} | \delta_m = 1] - E[Y_{ic0} | \delta_m = 0] = 0$$

افراد: در کم آلوده زندگی کنند
افراد: در کم آلوده باشد

~~فرض می شود اگر افراد گروه T در منطقه C (کم آلوده) قرار بگیرند، مراجعات تنفسی مشابهی خواهند داشت.~~

فرض می شود اگر افراد گروه T در منطقه C (کم آلوده) قرار بگیرند، مراجعات تنفسی مشابهی خواهند داشت. اما این فرض به دلیل انتخاب افراد بر اساس زندگی خودشان غلط است و افراد ضعیف تر تمایل بیشتری دارند در C زندگی کنند.

b

49-34

در این مورد فقط برای $\delta_m = 1$

$$y_{it} = \alpha + \gamma Post_t, \quad Post_t = \begin{cases} 1 & t=1 \\ 0 & t=0 \end{cases}$$

شش ماه اول را $t=0$ نشان می دهیم.

فرض

$$E[y_{it+1} | t=2] - E[y_{it+1} | t=1] = 0$$

فرض شده تنها عاملی که می تواند زمان عوض می شود و مراجعات تلفنی به بیمارستان ها را در گروه T عوض می کند. آلودگی بوده که صحیح نیست. نباید برای مثل آب شهر مسوم شود یا زلزله باید و ...
 ناهای در دوره دوم، کل مراجعات زیاد شود در نتیجه درصد مراجعات تلفنی کم شود.
 در این صورت فرض گفته شده غلط است. یعنی الزاماً فقط شرایط آلودگی عوض نشده!

$$y_{it} = \alpha + \beta_1 \delta_m + \beta_2 Post_t + \gamma \times \delta_m \times Post_t + u_{it}$$

(c)

$$\gamma = (49-34) - (28-23) = 15-5 = 10$$

تفسیر: اثر آلودگی بیشتر. و درصد مراجعات تلفنی بیشتر است.

چون آلودگی در $t=1$ در $\delta_m = 1$ ، بیشتر افزایش یافته، و درصد مراجعات تلفنی

ناشی از این هم کاهش داده و افزایش آلودگی نسبت به ناهایی $\delta_m = 0$ است.

d) در این مدل اثرات fixed effect در نظر گرفته شده. برای مثال شاید یک بیمارستان

به طور اختصاصی برای قلب باشد یا اختصاصی ریه باشد. در این صورت عرض از
مبدأ متفاوتی دارند. نکته در مورد P_{mt} شاید مشکل کمی بودن آب به طور دائمی
در ~~بعضی~~ بیمارستانها یا آتش سوزی بزرگی اتفاق می افتد که ~~مصرفان~~ ^{بیماران}
را بسیار زیاده می کند. بهتر است از این مدل استفاده کرد تا خطای کمتری داشته باشد.
اما ~~مجبور~~ ^{مجبور} است باید جبهی ضریب ترکیب شود تا تفاضل در تفاضل را در برگیرد.

$$Y_{imt} = \alpha_i + \beta_t + \beta_1 P_{mt} + \beta_2 P_{mt} \times \delta_m + e_{it}$$

در مدل گفته شده در سوال، اثر تفاضل در تفاضل در نظر گرفته نمی شود.

e) نیز چون به برای هر بیمارستان امفانه شده است و یک بیمارستان همیشه m بیمار
دارد. پس امفانه کردن β_m بی مورد است.

f) فرض تغییرات موازی. یعنی اگر عدد سرد نمی شود و در تقسیم آلودگی در مناطق پرالودگی
بیشتر از مناطق کم آلوده، افزایش نمی یابد، در هر مراجعین در هر بیمارستان متناسب
عوض می شود و همه در حد یک عدد ثابت تعبیری کردند.

g) اگر بیش از ۳ دوره داشته باشیم، می توانیم تغییرات موازی را تحقیق کنیم. یعنی بفهمیم
آیا Y_{imt} ها به ازای m های مختلف به هم زیاد می کمی شوند یا خیر.