

Linearitás és feloldása, nemlineáris modellek

Ferenci Tamás
tamas.ferenci@medstat.hu

Utoljára frissítve: 2023. május 12.

Tartalom

- 1 Előljáróban: a marginális hatás általánosabb értelmezése
- 2 A linearitás feloldása
 - Emlékeztetőül
 - Az additivitás feloldása: az interakció
 - A változónkénti linearitás feloldása
- 3 Néhány nevezetes, paraméterében nemlineáris modell
- 4 Specifikációs tesztek

A marginális hatás fogalma

- Marginális hatás: a magyarázó változó kis növelésének hatására mekkora az eredményváltozó *egységnyi magyarázóváltozó-növelésre jutó* változása
- Tipikus egyszerűsítés: a magyarázó változó egységnyi növelésének hatására mennyit változik az eredményváltozó
- (Hiszen a kettő ugyanaz, ha a változó hatása lineáris)
- Idáig az i -edik magyarázó változó ilyen módon értelmezett marginális hatása és a β_i számértéke gyakorlatilag szinonima volt

A marginális hatás precízebben

- Definíció alapján a marginális hatás: $\frac{\Delta Y}{\Delta X_j}$, ha ΔX_j kicsiny
- Ugye egyetemen vagyunk \rightarrow a marginális hatás $\frac{\partial Y}{\partial X_j}$
- A többváltozós lineáris regresszió eddigi (sokasági) modelljében
 $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$, ezért

$$\begin{aligned}\frac{\partial Y}{\partial X_j} &= \frac{\partial}{\partial X_j} [\beta_0 + \beta_1 X_1 + \dots + \\ &\quad + \dots + \beta_{j-1} X_{j-1} + \beta_j X_j + \beta_{j+1} X_{j+1} + \dots + \beta_k X_k + \varepsilon] = \\ &= \beta_j\end{aligned}$$

- ...hát ezért tekinthettük eddig a marginális hatást és a becsült regressziós koefficiens szinonimának!

A linearitás következményei

- A linearitás két dolgot vont maga után:
 - Mindegy, hogy honnan indulva növelem a változót egy egységgel (a változó hatása lineáris)
 - Mindegy, hogy a többi változó milyen szinten van rögzítve (additivitás)
- E kettőt fogjuk most feloldani

Interakció mint a linearitás egyféle feloldása

- Eddigi modellünkben a marginális hatások a többi változó szintjétől függetlenül állandóak voltak
- Például: 1 Ft pluszjövedelem taglétszámtól függetlenül azonos többletkiadást jelent...?
- Ha nem, akkor azt mondjuk, hogy a két változó között *interakció* van: az egyik marginális hatásának *nagyságát* befolyásolja a másik *szintje*
- A kapcsolat tehát marginális hatás és szint között van (nem marginális hatás és marginális hatás vagy szint és szint között!)
- Kézenfekvő indulás: az egyik változó szintje *lineárisan* hasson a másik marginális hatására; sokaságban felírva:

$$(\beta_J + \beta_{JT}\text{Tag}) J_{0v},$$

ahol β_{JT} az interakció hatását kifejező (lineáris) együttható

Interakció

- Helyezzük ezt be a (sokasági) regresszióba:

$$Y = \beta_0 + (\beta_J + \beta_{JT}\text{Tag}) \text{Jov} + \beta_T \text{Tag} + \varepsilon,$$

azonban felbontva a zárójelet:

$$\begin{aligned} Y &= \beta_0 + \beta_J \text{Jov} + \beta_{JT} \text{Tag} \cdot \text{Jov} + \beta_T \text{Tag} + \varepsilon = \\ &= \beta_0 + \beta_J \text{Jov} + (\beta_T + \beta_{JT} \text{Jov}) \text{Tag} + \varepsilon \end{aligned}$$

- Tehát az interakció *szükségképp*, automatikusan „szimmetrikus”: ha az egyik változó szintje hat a másik marginális hatására akkor szükségképp fordítva is: a másik szintje is hatni fog az előbbi marginális hatására

Interakció

- Azaz „egyszerre” lesz igaz, hogy $(\beta_J + \beta_{JT}\text{Tag})$ Jov és $(\beta_T + \beta_{JT}\text{Jov})$ Tag: attól függően, hogy milyen szempontból nézzük (melyik marginális hatását vizsgáljuk, ezt még ld. később is)
- A regresszióban így elég egyszerűen ennyit írni:

$$\beta_T\text{Tag} + \beta_J\text{Jov} + \beta_{JT}(\text{Jov} \cdot \text{Tag}).$$

- ...mindkét – másik szintjétől függő – marginális hatás ebből kiadódik, függően attól, hogy hogyan bontjuk fel a zárójelet (melyik változót vizsgáljuk)
- Ez a marginális hatás pontosabb értelmezése mellett még szebben látható

A marginális hatás interakciók esetén

- Ha interakció van, például a l -edik és az m -edik tag között, akkor az l -edik marginális hatása:

$$\begin{aligned}\frac{\partial Y}{\partial X_l} &= \frac{\partial}{\partial X_l} [\beta_0 + \beta_1 X_1 + \dots + \\ &+ \dots + \beta_l X_l + \dots + \beta_m X_m + \dots + \beta_k X_k + \beta_{lm} X_l X_m + \varepsilon] = \\ &= \beta_l + \beta_{lm} X_m\end{aligned}$$

- Így precíz az előbbi állításunk arról, hogy ha az egyik szerint vizsgáljuk a marginális hatást, akkor az a másik szintjétől fog függeni (gondoljuk hozzá a másik szerinti deriválást is!)

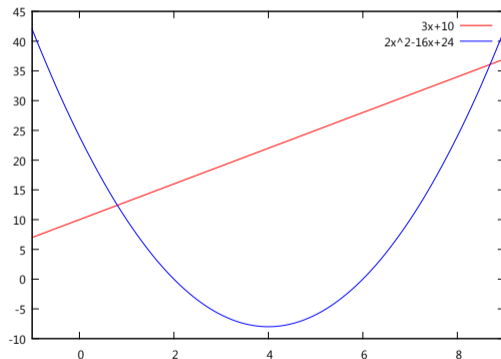
Motiváló példa: kvadratikus hatás

- Már volt: mit jelent az, ha megsértjük a „marginális hatás nem függ attól, hogy a többi magyarázó változót milyen szinten rögzítjük” következményét a linearitásnak
- És ha a „marginális hatás nem függ attól, hogy milyen szintről indulva növeljük a változót” következményt szeretnénk feloldani?
- Például: 1 évvel idősebb életkor kiinduló életkortól függetlenül azonos kiadásváltozást jelent...?
- Használjunk a lineáris függvényforma helyett más, például négyzeteset (parabolát):

$$\frac{\partial}{\partial X_j} [\dots + \beta_j X_j + \beta_{jj} X_j^2 + \dots] = \beta_j + 2\beta_{jj} X_j$$

Motiváló példa: kvadratikus hatás

Szemléletesen az egy magyarázó változós esetben:

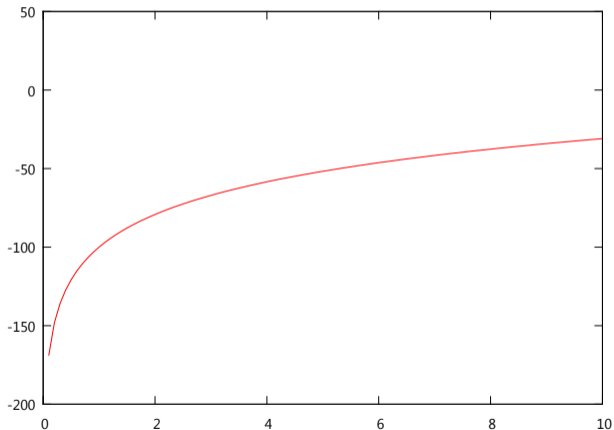


Szükségképpen nyilvánvaló (első derivált előjelet vált): $\beta_j + 2\beta_{jj}X_j = 0 \Rightarrow X_j = -\frac{\beta_j}{2\beta_{jj}}$

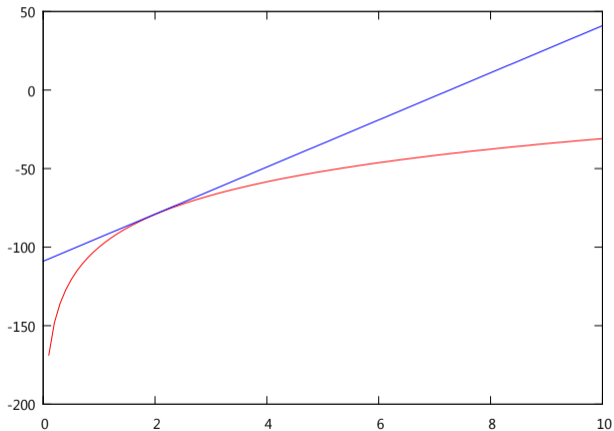
Linearitás, mint közelítés

- Az élet általában nemlineáris
- Miért használunk mégis lineáris modelleket: mert sokszor nem térnek el (nagyon) a valóságtól, de mégis sokkal könnyebben kezelhetőek matematikailag
- (Taylor-soros érvelés!)
- Ez tehát az esetek többségében egy közelítés
- Mint ilyen: vizsgálni kell az érvényességi határokat
- „Munkaponti linearizálás”

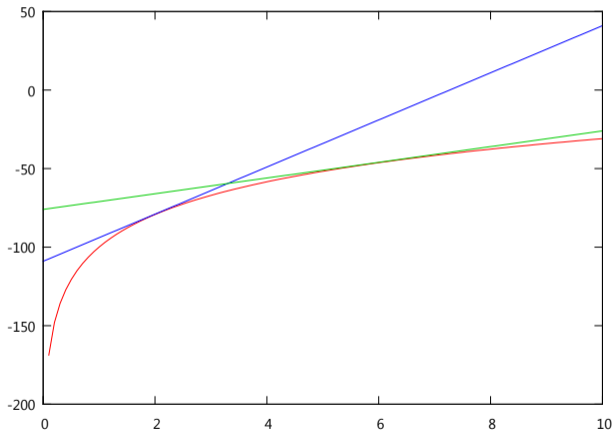
Érvényességi határok



Érvényességi határok



Érvényességi határok



Érvényességi határok

- Az érvényességi határokat az eddig látott modellekben is érdemes végiggondolni
- Azonnal kézenfekvő példa: a konstans (nagyon sok esetben)
- De sok meredekségnél is megragadható ez (fogyasztási függvény példája)
- Ez is egyfajta munkaponti linearizálás

Nemlinearitás fajtái

- Az $\beta_1 + \beta_2 X + \beta_3 X^2$ egy nemlineáris kifejezés (matematikailag)
- De figyelme: ennek ellenére minden további nélkül, tökéletesen kezelhető pusztán az eddig látott (lineáris!) eszköztárral, hiszen az OLS-nek mindegy, hogy a második magyarázó változó értékei történetesen épp az első négyzetei
- (Egészen addig nincs baj, amíg a kapcsolat nem lineáris)
- Nem úgy mint az $\beta_1 X^{\beta_2} \rightarrow$ ez nem becsülhető OLS-sel
- A megkülönböztetés végett az első esetet változójában, a másodikat paraméterében nemlineáris modellnek nevezzük
- Mi van „nemlinearitást okozó pozícióban”

Változójában nemlineáris modell

- Jellemző: továbbra is fennáll a „változók konstansokkal szorozva majd összeadva” (tehát: lineáris kombinációs) struktúra
- De elképzelhető, hogy egy változó egy „eredeti” változó transzformáltja
- Itt szükségképp nemlineáris transzformációról beszélünk!
- Vegyük észre, hogy az „eredeti” és „transzformált” közti megkülönböztetés teljesen mesterséges (csak mi tudjuk, hogy mi volt az adatbázisban bemenő adatként), az OLS-nek mindegy
- Ide tartozik a kvadratikus hatás, általában az X^a magyarázó változók, a $\log_a X$, az a^X stb., ahol a konstans
- Az előzőek miatt a becslés ugyanaz, egyedül az interpretálás igényel további tárgyalást

Paramétereiben nemlineáris modell

- Megsérti a lineáris kombináció struktúráját: paraméter nem csak szorzóként szerepel a regresszióban
- Például X^β , $\log_\beta X$ stb.
- Ez már nem becsülhető OLS-sel: az eredményváltozó nem állítható elő mátrixműveletekkel
- Más módszert fogunk használni

Interakció és kvadratikus hatás revisited

- Az előzőek fényében nyilvánvaló: a kvadratikus hatás egyfajta (igen egyszerű) változójában nemlineáris modell
- Az interakció szintén változóbeli nemlinearitás, de nem annyira kézenfekvő módon (mindenképp indokolt a külön tárgyalása)

Nemlinearitás kezelése: NLS

- Vegyük észre, hogy a \min_{β} ESS célfüggvény akkor is tartható, ha nemlineáris modellt specifikálunk!
- (Csak az ESS számításához szükséges \hat{Y} -ok másképp jönnek ki, de ez a fenti optimalizáció szempontjából *teljesen mindegy*)
- Oldjuk meg ezt az optimalizációs feladatot!
- Ez a nem-lineáris legkisebb négyzetek (NLS, non-linear least squares) módszere

Nemlinearitás kezelése: NLS

- Sajnos mondani könnyebb, mint a gyakorlatban kivitelezni; szemben a lineáris specifikációval, a kritériumfelület nem kvadratikus, emiatt nincs egyetlen művelettel megtalálható optimum
- Van-e egyáltalán egyértelmű (globális) optimum? Mi van, ha több lokális optimum létezik?

Nemlinearitás kezelése: NLS

- Ettől el is tekintve, a konkrét optimalizáció számos gyakorlati problémát vethet fel, mivel valamilyen iteratív algoritmus kell
- Több lehetőség van, különféle előnyökkel és hátrányokkal (Gauss–Newton keresés, Levenberg–Marquardt algoritmus, konjugált gradiens keresés stb.), de mind rengeteg numerikus kérdést vet fel:
 - Meg tudjuk találni az optimumot? Biztosan? (Lehet-e baj a konvergenciával? Mi legyen a konvergencia-kritérium?)
 - Mennyi idő alatt találjuk meg?
 - Milyen kezdőértékből induljunk? (Milyen a módszer numerikus stabilitása?)
 - stb. stb. stb.

Nemlinearitás kezelése: algebrai linearizáció

- Mi a fenti (mindig alkalmazható) módszerrel szemben egy másik (könnyebb, de nem mindig alkalmazható) módszert fogunk vizsgálni: algebrai linearizálás
- Alkalmas transzformációval a nemlineáris problémát lineárisá alakítjuk, azt OLS-sel megoldjuk, majd a kapott eredményeket visszatranszformáljuk az eredeti transzformáció inverzével
- Például: $Y = \beta_1 X^{\beta_2} \varepsilon$ paramétereiben nemlineáris ...
- ... de mindkét oldal logaritmusát véve $\log Y = \log \beta_1 + \beta_2 \log X + \varepsilon'$ már az!
- Adatbázis logaritmálása, eredmények visszahatványozása
- (Amint mondtuk, nem mindig alkalmazható, de azért nagyon sok, gyakorlatilag fontos esetben igen)
- Természetesen itt is eltérő, specifikus értelmezések jelenthetnek meg

Log-log modell

- Például a Cobb-Douglas termelési modell:

$$Y = \beta_1 L^{\beta_L} K^{\beta_K} \varepsilon,$$

ahol Y a kibocsátás, L a munka, K a tőke (ill. általában a termelési tényezők) felhasználása

- Elaszticitása:

$$\text{El}_L(L, K) = \frac{\frac{dY}{Y}}{\frac{dL}{L}} = \frac{dY}{dL} \frac{L}{Y} = \beta_1 \beta_L L^{\beta_L-1} K^{\beta_K} \frac{L}{\beta_1 L^{\beta_L} K^{\beta_K}} = \beta_L$$

- Ezért nevezik *konstans elaszticitású* modellnek is
- Kezelése linearizálással: mindkét oldalt logaritmáljuk

$$\log Y = \log \beta_1 + \beta_L \log L + \beta_K \log K + \varepsilon'$$

Log-log modell

- Minden változót (eredmény és összes magyarázó is) logaritmálni kell
- Innen a modell neve
- Csak a konstans lesz logaritmálva, a többi koefficiens a transzformáció ellenére (ill. épp azért...) közvetlenül kapjuk
- Volumenhozadék (skálahozadék): $\beta_K + \beta_L$ viszonya 1-hez

Log-lin modell

- Például a jövedelem alakulása:

$$Y = e^{\beta_1 + \beta_2 X + \varepsilon}$$

- Linearizálás ismét mindkét oldal logaritmálásával:

$$\log Y = \beta_1 + \beta_2 X + \varepsilon$$

- Elnevezés logikája így már látható: az eredményváltozó logaritmálva, de a magyarázó változók maradnak szintben
- Növekedési ráta: $e^{\beta_1 + \beta_2(X+1) + u} = Ye^{\beta_2}$, pillanatnyi növekedési ütem: $\beta_2 = \frac{d \log Y}{dX} = \frac{1}{Y} \frac{dY}{dX}$
- Elaszticitás: $El_X(Y) = \frac{dY}{dX} \frac{X}{Y} = \beta_2 X$, tehát csak X -től függ

Lin-log modell – kakukktojás!

- Az előzőek alapján már világos a jelentése (pl. terület és kínálati ár összefüggése):

$$Y = \beta_1 + \beta_2 \log X + \varepsilon$$

- Miért kakukktojás?
- β_2 értelmezése:

$$\frac{dY}{dX} = \frac{\beta_2}{X} \Rightarrow \beta_2 = \frac{dY}{dX/X}$$

- Elaszticitás:

$$\text{El}_X(X) = \frac{\beta_2}{X} \frac{X}{Y} = \frac{\beta_2}{Y},$$

tehát csak Y -től függ (közvetlenül)

Reciprok modell – kakukktojás

- Például keresleti modell:

$$Y = \beta_1 + \frac{\beta_2}{X} + \varepsilon$$

- Miért kakukktojás?
- Aszimptotikusan: $\lim_{X \rightarrow \infty} \mathbb{E}(Y | X) = \beta_1$
- Határkiadás: $\frac{dY}{dX} = -\frac{\beta_2}{X^2}$

- Elaszticitás:

$$\text{El}_X(X) = -\frac{\beta_2}{X^2} \frac{X}{Y} = \frac{\beta_2}{XY}$$

- Paraméterek értelmezése, β_2 előjelének jelentősége az „aszimptotikus” viselkedés szempontjából: az élvezeti cikkek példája

A specifikációs tesztek

- Itt már nagyon erősen felmerül a kérdés: hogyan dönthetek a különféle függvényformák között?
- Ld. a termelési függvény példáját → megadható lineárisan és Cobb-Douglas jelleggel (eredmény nagyon nem mindegy)
- Hogyan lehet *analitikusan* dönteni?
- Az előző példára: BM-teszt, PE-teszt stb.
- Általánosságban (nem csak log/lin kérdésekre, mint az előzőek): ún. specifikációs tesztek

Egy egyszerű specifikációs teszt

- Egy egyszerű ötlet: adjuk hozzá a magyarázó változókhoz a magyarázó változók valamilyen nemlineáris transzformáltját (tipikusan négyzeteiket vagy logaritmusait)...
- ...és nézzük meg, hogy *együttesen* szignifikánsak-e!
- Ha igen, az specifikációs hibára utal
- (Tehát figyelem, ezt alapvetően nem arra használjuk, hogy arra következtessünk, hogy egy adott konkrét változó négyzetét vagy logaritmusát hozzá kell-e adni a függvényformához, hanem összességében nézzük őket, specifikációs tesztként)
- Hátrány: sok szabadságfokot használ el, és csak elég speciális alakú nemlinearitásokkal tesztel

Ramsey RESET-je

- A modellspecifikáció *általános* tesztje; emiatt előnye: nem egy adott specifikációs kérdésre keres választ, hanem általában vizsgálja, hogy a specifikáció jó-e; hátránya, hogy ha nemleges választ ad, nem derül ki, hogy pontosan mi a specifikáció baja
- Trükk: új regressziót becsül, melynek eredményváltozója ugyanaz, de a magyarázó változókhoz hozzáadja az eredeti regresszió becsült eredményváltozójának magasabb hatványait (\hat{Y}^3 -ig néha \hat{Y}^4 -ig is):

$$Y = \beta'_0 + \beta'_1 X_1 + \beta'_2 X_2 + \dots + \beta'_k X_k + \gamma_1 \hat{Y}^2 + \gamma_2 \hat{Y}^3 + \varepsilon'$$

- Amit tesztelni kell az előzőhöz hasonlóan: $H_0 : \gamma_1 = \gamma_2 = 0$ (F -próba és LM-próba is van, heteroszkedaszticitásra is robusztussá tehető)
- Ilyen módon takarékos a szabadsági fokokkal, csak 2-3-at használ el (a fenti trükkel „összesűriti” a magyarázó változókat), ráadásul általánosabb alakú nemlinearitások is beleférnek ebbe, mint a csak négyzetekkel/logaritmusokkal tesztelés
- Specifikációs teszt, tehát kihagyott változó detektálására általában nem alkalmas