دانشگاه الزهرا – دانشکده علوم اجتماعی و اقتصاد

اقتصادسنجی کارشناسی ارشد

(آزمون‌های مربوط به خطای تصریح)

استاد: دکتر صفرزاده

فروردین ۹۹

طرح مسئله

با فرض برقراری فروض کلاسیک، برآوردکننده‌های حداقل مربعات معمولی ویژگی‌های برآورد کننده‌های مطلوب را دارا هستند و می‌توان رویه‌های استنتاج آماری را روی آنها پیاده کرد. اما در این زمینه برخی سؤالات اساسی و قابل تامل وجود دارد؛ که مهمترین آنها به شرح زیر است:

- از کجا بدانیم که فروض کلاسیک در مورد مجموعه داده‌های ما صادق است؟
- ویژگی‌های جزء اخلال غیر قابل مشاهده را از کجا بدانیم؟
- از کجا بدانیم که چه متغیرهایی در ماتریس مشاهدات X لحظ شوند و با چه فرم تبعی‌ای؟
- اگر هر یک از فرض‌های پایه‌ای کلاسیک نقض شوند، چه مشکلی برای برآوردکننده‌های حداقل مربعات ایجاد می‌شود؟
- آیا این برآوردکننده‌ها همچنان قابل استفاده هستند، یا دچار نقصان اساسی شده و گمراه کننده می‌شوند؟
- آیا تحت مفروضات غیر کلاسیکی برآورد کننده‌های رقیب مناسب و رویه‌های استنتاج متناسب با آنها وجود دارد؟

هر گاه هر یک از فروض کلاسیک برقرار نباشد، خطای تصریح پیش می‌آید. هر چند برخی از خطاهای تصریح خیلی جزئی هستند، ولی برخی دیگر ممکن است خیلی اساسی و مشکل ساز باشند. بنابراین توجه به امکان وجود خطای تصریح و آزمون وجود آنها خیلی مهم است.

خطای تصریح (Specification Error)

تصریح مدل‌های خطی معمولا بر اساس ویژگی‌های ماتریس‌های X و ε صورت می‌گیرد. برخی از مهمترین مفروضات در مورد این ماتریس‌ها به صورت زیر است.

۱- ارتباط بین متغیرها خطی است.

$$Y = X\beta + \varepsilon \qquad (۱)$$

۲-جملات اخلال به صورت یکسان و مستقل از هم و با واریانس مساوی توزیع شده‌اند.

$$\varepsilon_i \sim iid(0,\sigma^2);\ i = 1,2,\ldots.n \qquad (\text{۲ الف})$$

همچنین فرض می‌شود این جملات اخلال از نوع گوسی هستند.

$$\varepsilon_i \sim iidN(0,\sigma^2);\ i = 1,2,\ldots.n \qquad (\text{۲ ب})$$

۳-جملات اخلال و متغیرهای توضیحی مستقل از هم هستند.

$$E(X_{it},\varepsilon_s) = 0\ \ i = 1,2,\ldots k\ \&\ t,s = 1,2,\ldots..n \qquad (۳)$$

۴- ماتریس X غیر تصادفی بوده و دارای رتبه کامل ستونی است.

$$Rank[X] = k \qquad (۴)$$

مسائل مربوط به جمله اخلال (ε)

۱- فرض (۲ الف) برقرار باشد ولی (۲ب) برقرار نباشد. در این صورت ویژگی‌های BLUE بودن برآوردکننده‌های حداقل مربعات حفظ می‌شود ولی رویه‌های استنتاج صرفا به صورت مجانبی خواهد بود؛ به عبارت دیگر در حالت حدی معتبر خواهند بود.

۲-ماتریس واریانس-کوواریانس به صورت قطری بوده ولی عناصر روی قطر اصلی ثابت نباشند. به عبارت دیگر فرض واریانس همسانی (Homoscedasticity) نقض شود.

$$E(\varepsilon\varepsilon') = diag[\sigma_1^2 \ \ \sigma_2^2 \ \ ...\sigma_n^2].$$

این مسئله ساده‌ترین شکل واریانس ناهمسانی (Heteroscedasticity) است که اغلب در داده‌های مقطعی مشاهده می‌شود؛ هر چند این مسئله و اشکال پیچیده‌تر آن در داده‌های سری زمانی هم پیش می‌آید.

۳- جملات اخلال ممکن است به صورت دو به دو همبستگی داشته باشند.

$$E(\varepsilon_t\varepsilon_{t-s}) \neq 0 \ ,(s \neq 0)$$

در داده‌های سری زمانی احتمال وجود همبستگی در جملات اخلال مجاور شدیدتر است و ممکن است که این مسئله در فواصل زمانی بیشتر کاهش پیدا کند.

مسائل مربوط به ماتریس X

۱- از قلم افتادن متغیرهای مناسب

۲- لحاظ کردن متغیرهای نامرتبط (زاید)

۳- شکل تبعی نادرست

۴- ماتریس X رتبه ستونی کامل نداشته باشد. در این شرایط امکان بدست آورد ضرایب یگانه وجود نخواهد داشت. در ادبیات اقتصادسنجی این مسئله زمانی پیش می‌آید که متغیرهای لحاظ شده در ماتریس X همخطی (Collinearity) داشته باشند.

همخطی در میان متغیرها دو نوع می‌تواند وجود داشته باشد:

- ✓ همخطی قوی: $X_2 = aX_3$. در این شرایط امکان برآورد ضرایب وجود نخواهد داشت.
- ✓ همخطی ضعیف: $X_2 = aX_3 + b$. در این صورت امکان برآورد ضرایب وجود دارد ولی ضرایب کارایی خود را از دست می‌دهند.

مسائل مربوط به ماتریس X (ادامه)

۵- بین متغیرهای توضیحی و جملات اخلال همبستگی وجود داشته باشد. این مسئله در شرایط مختلف می‌تواند ایجاد شود که متعارف‌ترین آن زمانی است که وقف متغیر وابسته به صورت متغیر توضیحی در سمت راست معادله رگرسیون قرار گیرد. به عنوان مثال:

$$Y_t = \gamma + \sum_{i=۱}^{s} \alpha_i Y_{t-i} + \sum_{i=۱}^{k} \beta_i X_i + \varepsilon_i \quad s,t = ۱,۲,...t$$

در این شرایط متغیر توضیحی با جملات اخلال جاری و آتی همبستگی ندارد ولی با جملات اخلال گذشته همبسته خواهد بود. در این صورت برآورد کننده‌های حداقل مربعات معمولی در نمونه‌های کوچک تورش دار خواهد بود ولی در نمونه‌های بزرگ دارای ویژگی سازگاری بوده و به طور مجانبی توزیع نرمال خواهند داشت. مشکل زمانی حادتر خواهد شد که متغیرهای توضیحی با جملات اخلال جاری همبستگی داشته باشند. در این صورت برآورد کننده‌های حداقل مربعات تورش‌دار و ناسازگار خواهند بود. این مشکل زمانی ایجاد می‌شود که در اندازه‌گیری متغیرهای توضیحی خطا صورت گیرد و یا اینکه با معادلات همزمان سر و کار داشته باشیم.

۶- متغیرهای لحاظ شده در ماتریس X نامانا باشند. (Nonstationary Variables)

مسائل مربوط به پارامترها

فرض ضمنی در رابطه (۱) این است که بردار β در طول تمام مشاهدات نمونه‌ای واقعی یا ممکن، ثابت باشد. اما ممکن است این ضرایب در اثر شکست‌های ساختاری ناگهانی در آنها و یا در واکنش به تحولات تدریجی اجتماعی و محیطی تغییر کنند.

ارزیابی مدل و آزمون‌های تشخیص

مدل‌سازی استاندارد اقتصادسنجی در عمل موارد زیر را در بر می‌گرفت:

- مدلسازی بر اساس مبانی نظری یا یافته‌های پیشین اقتصادسنجی
- برآورد ضرایب مدل تصریح شده با استفاده از داده‌های در دسترس
- وارسی نتایج برآورد شده وآماره‌های توام با آن جهت قضاوت در مورد دقت مدل تصریح شده

این واررسی معمولا بر موارد زیر تمرکز داشت:

- ✓ برازش کلی
- ✓ تطابق علامت ضرایب برآورد شده با انتظارات پیشینی
- ✓ معنی‌داری آماری ضرایب برآورد شده
- ✓ وجود خودهمبستگی بین جملات اخلال

اگر مدلی بر اساس معیارهای فوق رضایت بخش بود، معادله مورد برآورد به ادبیات موضوع در این زمینه اضافه شده و از آن برای پیش بینی دوره‌های آتی استفاده می‌شد. ولی اگر بر این مبنا رضایت بخش نبود، پژوهشگر باید به مدل‌سازی مجدد می‌پرداخت تا یک معادله رضایت‌بخش‌تری بدست آورد.

رویکرد نوین در ارزیابی مدل و آزمون‌های تشخیص

در سال ۱۹۷۵ رویکرد سنتی در مورد مدلسازی اقتصادسنجی مورد انتقاد شدید دنیس سارگان (Denis Sargan) در مدرسه اقتصادی لندن قرار گرفت. ایشان بیان داشتند که: «*مدل‌های تصریح شده ضمن اینکه باید با فرایندهای داده‌کاوی مورد ارزیابی و اعتبارسنجی قرار گیرند؛ بلکه باید از تمام راه‌های ممکن نیز مورد آزمون قرار گیرند و فقط مدل‌های تصریح شده‌ای ماندگار خواهند بود که از بوته تمام این آزمون‌ها سرافراز بیرون بیایند.*»

این رویکرد بعد از آن توسط دیوید هندری (David Hendry) و همکاران او توسعه یافت و بر این اساس علاوه بر آزمون‌های متعارف پیش‌گفته، انواع آزمون‌های دیگر هم مورد توجه قرار گرفت.

آزمون‌های مربوط به ثبات پارامترها

یکی از مهمترین معیارها برای معادله برآورد شده این است که این معادله برای داده‌های خارج از نمونه استفاده شده برای برآورد هم اعتبار داشته باشد؛ به عبارت دیگر پارامترهای مدل برآورد شده باید ثبات (Parameter Constancy) داشته باشند.

ثبات ضرایب را از راه‌های مختلف می‌توان آزمون کرد که یکی از مفیدترین این آزمون‌ها، دقت پیش‌بینی است.

آزمون پیش‌بینی چو (The Chow Forecast test)

اگر پارامترهای برآورد شده ثابت باشند، پیش بینی‌های خارج نمونه‌ای حاصل از این برآورد با احتمالات مشخص در محدوده‌های قابل محاسبه از داده‌های نمونه‌ای قرار خواهند گرفت. خطاهای پیش‌بینی بزرگ‌تر، در رابطه با ثبات پارامترها تردید ایجاد می‌کند، و برعکس هرچه خطاهای پیش بینی کوچکتر باشد، اعتماد به ثبات ضرایب بیشتر می‌شود. بر این اساس یکی از راه‌های آزمون ثبات ضرایب افراز مشاهدات به دو گروه n_1 و n_2 است؛ که از n_1 مشاهده برای برآورد ضرایب استفاده می‌شود و با استفاده از ضرایب بدست آمده n_2 مشاهده باقی‌مانده پیش بینی می‌شود. در کارهای تجربی معمولا ۵، ۱۰ یا ۱۵ درصد مشاهدات برای پیش بینی انتخاب می‌شوند.

آزمون دقت پیش بینی معمولا به اعتبار مقاله ارزشمند چو (۱۹۶۰) به ایشان نسبت داده می‌شود. فرآیند این آزمون به صورت زیر است:

- ابتدا بردار ضرایب با استفاده از n_1 مشاهده به روش حداقل مربعات برآورد می‌شود.

$$b_1 = (X_1'X_1)^{-1}X_1'Y \qquad (۵)$$

که در آن $X_i, Y_i (i = 1,2)$ به افراز مشاهدات به دو گروه n_1 و n_2 مربوط می‌شود.

- با استفاده از بردار برآورد شده (b_1) بردار Y_2 پیش بینی می‌شود.

$$\hat{Y}_2 = X_2 b_1 \qquad (۶)$$

آزمون پیش‌بینی چو (ادامه)

- بردار خطای پیش بینی بدست آمده و توزیع نمونه‌ای آن تحت فرضیه صفر مبنی بر ثبات ضرایب، تحلیل می‌شود.

بردار خطای پیش بینی به صورت زیر خواهد بود:

$$d = Y_2 - \hat{Y}_2 = Y_2 - X_2 b_1 \qquad (7)$$

اگر معادله $Y = X\beta + \varepsilon$ با فرض $E(\varepsilon\varepsilon') = \sigma^2 I$ برای دو گروه مشاهدات برقرار باشد؛ بردار خطای پیش بینی را می‌توان به صورت زیر بازنویسی کرد:

$$d = Y_2 - X_2 b_1 = \varepsilon_2 - X_2(b_1 - \beta)$$

بنابراین $E(d) = 0$ بوده و ماتریس واریانس – کوواریانس آن به صورت زیر خواهد بود.

$$\begin{aligned} \mathrm{var}(d) &= E(dd') = \sigma^2 I_{n_2} + X_2 . \mathrm{var}(b_1) . X_2' \\ &= \sigma^2 [I_{n_2} + X_2 (X_1' X_1)^{-1} X_2'] \qquad (8) \end{aligned}$$

اگر جملات اخلال گوسی باشند؛

$$X_2' d \sim N[0, \mathrm{var}(d)]$$

و به این اعتبار خواهیم داشت:

$$d' [\mathrm{var}(d)]^{-1} d \sim \chi^2_{n_2}$$

آزمون پیش‌بینی چو (ادامه)

همچنین می‌دانیم:

$$e_1'e_1 / \sigma^2 \sim \chi^2_{n_1-k}$$

که در آن $e_1'e_1$ مجموع مجذورات باقیمانده‌های رگرسیون برآورد شده است. از آنجا که دو توزیع χ^2 فوق مستقل از هم هستند، پس می‌توان تحت فرضیه ثبات پارامترها آماره زیر را تعریف کرد:

$$F = \frac{d'[I_{n_2} + X_2(X_1'X_1)^{-1}X_2']^{-1}d \big/ n_2}{e_1'e_1 \big/ n_1 - k} \sim F_{n_2, n_1-k} \qquad (9)$$

استخراج $\mathrm{var}(d)$ بر این اساس استوار بود که σ^2 برای دو گروه داده‌ها یکسان است. پس آماره F در رابطه (۹) به شرط برقراری این فرض اعتبار دارد. اگر واریانس‌ها یکسان نباشند آماره F ممکن است سطح معنی‌داری درست را بیش از حد نشان دهد.

آزمون پیش‌بینی چو (ادامه)

یک راه جایگزین روشن‌تر نیز برای استخراج آزمون دقت پیش‌بینی وجود دارد، که پاگان و نیکولز (Pagan & Nicholls) ارائه داده‌اند. فرض بگیرید که پارامترها در دو گروه داده‌ها یکسان نباشند:

$$
\begin{aligned}
&Y_1 = X_1\beta + \varepsilon_1 \\
&Y_2 = X_2\alpha + \varepsilon_2 = X_2\beta + X_2(\alpha-\beta) + \varepsilon_2 = X_2\beta + \gamma + \varepsilon_2 \qquad (۱۰) \\
&;\gamma = X_2(\alpha-\beta);\ if\ \ \gamma = 0 \Rightarrow \alpha = \beta
\end{aligned}
$$

یعنی بردار پارامترها در دو دوره برآورد شده و پیش‌بینی شده یکسان است.

$$
\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ X_2 & I_{n_2} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \qquad (۱۱)
$$

در رابطه فوق خواهیم داشت:

$$
X'X = \begin{bmatrix} X_1'X_1 + X_2'X_2 & X_2' \\ X_2 & I \end{bmatrix}
$$

که جهت سهولت اندیس n_2 از ماتریس I حذف شده است.

$$
(X'X)^{-1} = \begin{bmatrix} (X_1'X_1)^{-1} & -(X_1'X_1)^{-1}X_2' \\ -X_2(X_1'X_1)^{-1} & I + X_2(X_1'X_1)^{-1}X_2' \end{bmatrix}
$$

آزمون پیش‌بینی چو (ادامه)

براورد ضرایب معادله (۱۱) به روش حداقل مربعات معمولی به صورت زیر خواهد بود:

$$\begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1} & -(X_1'X_1)^{-1}X_2' \\ -X_2(X_1'X_1)^{-1} & I + X_2(X_1'X_1)^{-1}X_2' \end{bmatrix} \begin{bmatrix} X_1'Y_1 + X_2'Y_2 \\ Y_2 \end{bmatrix}$$

$$= \begin{bmatrix} (X_1'X_1)^{-1}X_1'Y_1 \\ Y_2 - X_2(X_1'X_1)^{-1}X_1'Y_1 \end{bmatrix} = \begin{bmatrix} b_1 \\ d \end{bmatrix} \qquad (۱۲)$$

بنابراین k ضریب اول برآوردی به روش حداقل مربعات معمولی برای پارامتر β یعنی b_1 از طریق n_1 مشاهده بدست می‌آید؛ و n_2 ضریب باقیمانده که بردار γ برآورد می‌کند، به طور ساده همان خطاهای پیش‌بینی هستند که در رابطه (۷) تعریف شده بود.

بنابراین برآورد معادله (۱۱) به روش حداقل مربعات به صورت زیر خواهد بود:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & o \\ X_2 & I_{n_2} \end{bmatrix} \begin{bmatrix} b_1 \\ d \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \qquad (۱۳)$$

آزمون پیش‌بینی چو (ادامه)

معادله دوم در رابطه (۱۳) به صورت زیر است:

$$Y_2 = X_2 b_1 + d + e_2$$

اگر d را از رابطه (۷) جایگزاری کنیم خواهیم داشت $e_2 = 0$. بنابراین RSS از برازش معادله (۱۱) به طور ساده به صورت $e_1'e_1$ خواهد بود. از رابطه (۱۰) مشخص است که فرضیه ثبات ضرایب β همان معادل فرضیه $H_0 : \gamma = 0$ است. این فرضیه در واقع از طریق بررسی معنی‌داری همزمان n_2 متغیر در رابطه (۱۳) آزمون می‌شود. این امر همان کاربرد مستقیم رویه آزمون قیدهای خطی است که در جلسات قبل بررسی شد. اگر در آماره آزمون قیدهای خطی جایگزاری متناسب انجام شود؛ آن آماره آزمون به صورت زیر در خواهد آمد:

$$F = \frac{d'[\mathrm{var}(d)]^{-1} d \big/ n_2}{e_1'e_1 \big/ n_1 - k} \sim F_{n_2, n_1 - k} \qquad (۱۴)$$

که درجه آزادی مخرج کسر فوق از تفاضل تعداد مشاهدات $(n_1 + n_2)$ و پارامترهای برآورد شده $(k + n_2)$ به دست آمده است.

آزمون پیش‌بینی چو (ادامه)

در نهایت محاسبات ساده‌تری هم در قالب رگرسیون‌های مقید نامقید قابل دستیابی است. در این مورد رگرسیون نامقید به شکل رابطه (۱۳) خواهد بود که مجموع مجذورات خطای آن به صورت $RSS = e_1'e_1$ بود. در این حالت رگرسیون مقید از طریق تحمیل قید $\gamma = 0$ بدست خواهد آمد؛ که به صورت رابطه زیر با $RSS = e_*'e_*$ قابل برآورد است.

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} b_* + e_*$$

آماره آزمون برای $\gamma = 0$ به صورت زیر خواهد بود:

$$F = \frac{(e_*'e_* - e_1'e_1)/n_2}{e_1'e_1/(n_1 - k)} \sim F_{n_2, n_1 - k} \qquad (۱۵)$$

بنابراین در کارهای تجربی آزمون چو را به صورت زیر می‌توان انجام داد:

- در ابتدا با استفاده از n_1 مشاهده اول Y_1 را روی X_1 رگرس کرده و RSS آنرا به صورت $e_1'e_1$ بدست آورید.
- همان رگرسیون را روی تمام مشاهدات $(n_1 + n_2)$ برازش کرده و RSS مقید را به صورت $e_*'e_*$ بدست آورید.
- در رابطه (۱۵) جایگزاری کرده و با محاسبه مقدار آماره و مقایسه آن با مقادیر بحرانی در جدول در مورد رد یا عدم رد فرضیه ثبات پارامترها تصمیم‌گیری کنید.