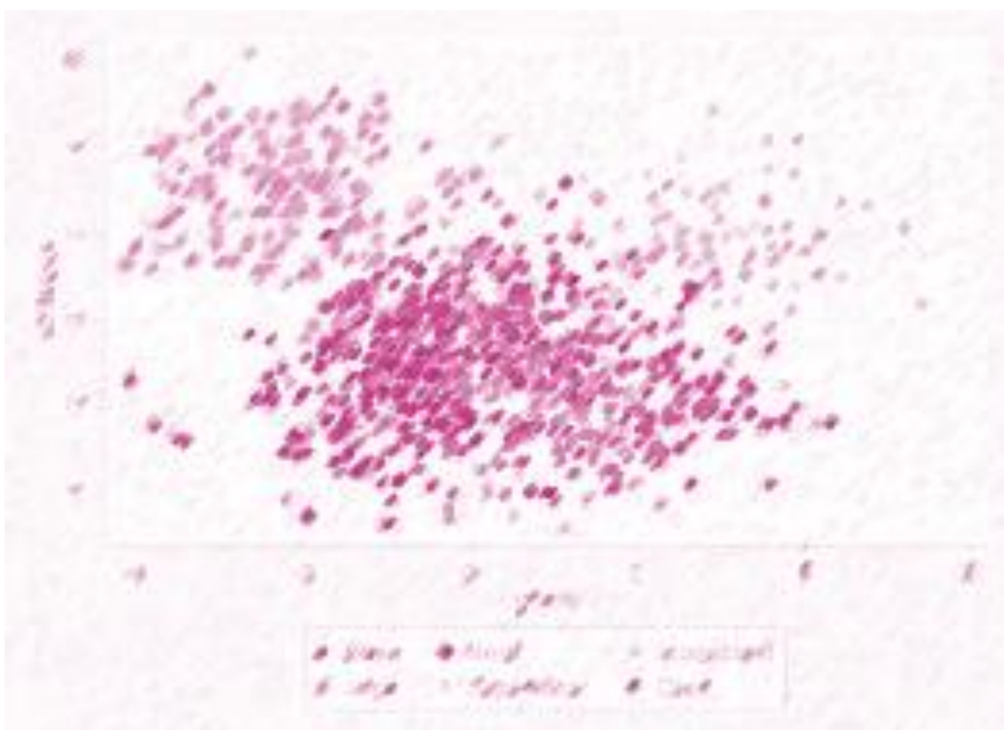
پاسخ‌نامه امتحان پایان‌ترم اقتصادسنجی دوره فرعی

محمد مهدی بناساز



تابستان ۱۳۹۳

دانشکده مدیریت و اقتصاد

دانشگاه صنعتی شریف

سوال اول

الف) متغیرهای زیر را تعریف می‌کنیم:

M = Mean Contingency Fee, LR = Ln Recovery

معادله رگرسیون زیر را درنظر بگیرید؛ نقطه شکستگی را LR_0 می‌نامیم.

$$M = a + b.LR + c.D + d.D \times LR + u \quad (1)$$

$$D = \begin{cases} 0, & if\ LR < LR_0 \\ 1, & if\ LR \geq LR_0 \end{cases}$$

معادله رگرسیون قبل از شکستگی:

$$M = a + b.LR + u$$

معادله رگرسیون بعد از شکستگی:

$$M = a + b.LR + c.D + d.D \times LR + u$$

همانطور که در نمودار مشخص است، در نقطه شکستگی مقادیر M برابر است. پس داریم:

$$a + b.LR_0 = (a + c) + (b + d).LR_0 \ \rightarrow c + d.LR_0 = 0$$

عبارت D ($c + d.LR_0 = 0$) را از سمت راست معادله رگرسیون ۱ کم می‌کنیم تا به معادله زیر برسیم.

$$\boldsymbol{M = a + b.LR + d.D(LR - LR_0) + u}$$

ضریب a مربوط به شیب منحنی قبل از شکستگی و ضریب d شیب منحنی بعد از شکستگی را نمایندگی می‌کنند.

ب) از آزمون‌های cusum, cusumq استفاده می‌کنیم.

ج) بعد از مشخص شدن نقطه شکستگی کافی است در معادله رگرسیونی که برای الف بدست آوردیم معنی‌داری تخمین ضریب d آزمون شود.

سوال دوم

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

$$\hat{y} = 4 + 0.4x_1 + 0.9x_2 \,, R^2 = \frac{8}{60} \,, \sum error^2 = 520, \;\; n = 29$$

$$X'X = \begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum {x_1}^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_2 x_1 & \sum {x_2}^2 \end{bmatrix} = \begin{bmatrix} 29 & 0 & 0 \\ 0 & 50 & 10 \\ 0 & 10 & 80 \end{bmatrix}$$

معادله‌های بسیار مهم که در این قسمت از آن‌ها استفاده می‌کنیم. این فرمول‌های در ضمیمه A کتاب درسی آمده است!

$$\sum_{i=1}^{n} (x_{1i} - \overline{x_1})^2 = \sum_{i=1}^{n} x_{1i}^2 - n(\overline{x_1})^2$$

$$\sum_{i=1}^{n} (x_{1i} - \overline{x_1})(x_{2i} - \overline{x_2}) = \sum_{i=1}^{n} x_{1i} x_{2i} - n(\overline{x_1}.\overline{x_2})$$

آزمون زیر را انجام می‌دهیم:

$$\begin{cases} H_0: \beta_1 + \beta_2 = 1 \\ H_1: \beta_1 + \beta_2 \neq 1 \end{cases}$$

آماره آزمون:

$$\boldsymbol{t = \frac{(\widehat{\beta_1} + \widehat{\beta_2}) - 1}{\sqrt{Var(\widehat{\beta_1}) + Var(\widehat{\beta_2}) + 2cov(\widehat{\beta_1}, \widehat{\beta_2})}}}$$

$$Var\left(\widehat{\beta_1}\right) = \frac{\widehat{\sigma^2}}{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2 \times (1-R_1^2)} = \frac{\frac{\sum error^2}{n-3}}{\sum_{i=1}^{n} x_{1i}^2 - n(\overline{x_1})^2 \times \left(1-\frac{Cov(x_1,x_2)^2}{var(x_1)var(x_2)}\right)}$$

$$= \frac{\frac{\sum error^2}{n-3}}{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2 \times \left(1-\frac{\left(\sum_{i=1}^{n}(x_{1i}-\overline{x_1})(x_{2i}-\overline{x_2})\right)^2}{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2\sum_{i=1}^{n}(x_{2i}-\overline{x_2})^2}\right)}$$

$$= \frac{\frac{\sum error^2}{n-3}}{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2 \times \left(1-\frac{\left(\sum_{i=1}^{n}(x_{1i}-\overline{x_1})(x_{2i}-\overline{x_2})\right)^2}{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2\sum_{i=1}^{n}(x_{2i}-\overline{x_2})^2}\right)}$$

$$= \frac{\frac{\sum error^2}{n-3}}{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2 \times \left(\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2\sum_{i=1}^{n}(x_{2i}-\overline{x_2})^2-\left(\sum_{i=1}^{n}(x_{1i}-\overline{x_1})(x_{2i}-\overline{x_2})\right)^2}{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2\sum_{i=1}^{n}(x_{2i}-\overline{x_2})^2}\right)}$$

$$= \frac{\frac{\sum error^2}{n-3}}{\left(\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2\sum_{i=1}^{n}(x_{2i}-\overline{x_2})^2-\left(\sum_{i=1}^{n}(x_{1i}-\overline{x_1})(x_{2i}-\overline{x_2})\right)^2}{\sum_{i=1}^{n}(x_{2i}-\overline{x_2})^2}\right)}$$

$$= \frac{\frac{\sum error^2}{n-3}}{\left(\frac{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2\sum_{i=1}^{n}(x_{2i}-\overline{x_2})^2-\left(\sum_{i=1}^{n}(x_{1i}-\overline{x_1})(x_{2i}-\overline{x_2})\right)^2}{\sum_{i=1}^{n}(x_{2i}-\overline{x_2})^2}\right)} = \cdots$$

برای کوتاه شدن محاسبات، با استفاده از فرم ماتریسی داریم (حروف **پررنگ** نشانه بردار است):

$$Var\left(\widehat{\boldsymbol{\beta}}\middle|\boldsymbol{X}\right) = \sigma^2 \times (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} 0.0345 & 0 & 0 \\ 0 & 0.0205 & -0.0026 \\ 0 & -0.0026 & 0.0128 \end{bmatrix}$$

$$Var\widehat{(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})} = \widehat{\sigma^2} \times (\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{\sum error^2}{n-3} \times \begin{bmatrix} 0.0345 & 0 & 0 \\ 0 & 0.0205 & -0.0026 \\ 0 & -0.0026 & 0.0128 \end{bmatrix}$$

$$= \frac{520}{26} \times \begin{bmatrix} 0.0345 & 0 & 0 \\ 0 & 0.0205 & -0.0026 \\ 0 & -0.0026 & 0.0128 \end{bmatrix}$$

$$= \begin{bmatrix} 0.6897 & 0 & 0 \\ 0 & 0.4103 & -0.0513 \\ 0 & -0.0513 & 0.2564 \end{bmatrix}$$

$$\boldsymbol{t} = \frac{(\widehat{\boldsymbol{\beta}_1} + \widehat{\boldsymbol{\beta}_2}) - \mathbf{1}}{\sqrt{\boldsymbol{Var}(\widehat{\boldsymbol{\beta}_1}) + \boldsymbol{Var}(\widehat{\boldsymbol{\beta}_2}) + \mathbf{2}\boldsymbol{cov}(\widehat{\boldsymbol{\beta}_1}, \widehat{\boldsymbol{\beta}_2})}} = \frac{\mathbf{0.3}}{\sqrt{\mathbf{0.1683} + \mathbf{0.0657} - \mathbf{0.1026}}}$$

$$= \frac{\mathbf{0.3}}{\mathbf{0.3625}} = \mathbf{0.8276}$$

با توجه به اینکه آماره t به دست آمده از ۲ کوچک‌تر است، نمی‌توانیم فرضیه نول را رد کنیم.

سوال سوم

داده‌های ما از ۹۰ شهرستان طی ۷ سال (۸۱ الی ۸۷) تهیه شده است. و در این سوال ما به دنبال تأثیر تعداد پلیس سرانه بر جرم و جنایت هستیم.

الف) اگر به تمام قسمت‌های سوال، مخصوصا قسمت ۲ نگاهی بیاندازیم می‌فهمیم؛ در سال‌های بعد از ۸۶ (که سال ۸۶ و ۸۷ در دامنه مطالعاتی ما قرار دارد)، در نیمی از شهرستان‌ها به صورت تصادفی سیاست افزایش پلیس اجرا شده است. در واقع برای نیمی از استان‌ها در سال ۸۶ و ۸۷ ما شاهد تغییر سیاست بوده‌ایم که بهتر است اثرات این تغییر کیفی را با متغیرهای مجازی شکار کنیم.

با توجه به توضیحات بالا ما باید برای نیمی از شهرستان‌ها در سال‌های ۸۶ و ۸۷ متغیر مجازی تعریف کنیم و از آن‌جایی که شهرستان‌ها به طور تصادفی انتخاب شده اند، این شهرستان‌ها در سال‌های ۸۶ و ۸۷ دچار تغییر کیفی شده‌اند و لازم نیست شهرستان‌ها در همه زمان‌ها متغیر مجازی داشته باشند.

رگرسیون‌های ۳، ۸، ۹، ۱۰ و ۱۱ در نظر اول برای این قسمت مناسب به‌نظر می‌رسند. اما ما باید یکی از آن‌ها را با توجیحات آماری به عنوان بهترین مدل برای پاسخ‌گویی به سوال انتخاب کنیم. در رگرسیون ۳ تنها برای شیب سال‌های ۸۶ و ۸۷ به طور مجزا متغیر مجازی در نظرگرفته شده است. ایراد اول این است که تنها ما برای شیب متغیر مجازی درنظر گرفته‌ایم و مدل خود را به این محدود کرده‌ایم که عرض از مبدأ ثابت می‌ماند و مشکل

بعدی این است که طبق استدلال ما سال‌های ۸۶ و ۸۷ کیفیتی متفاوت از هم نیستند و می‌توانستیم متغیر مجازی را برای هر دو مشترکا تعریف کنیم و به جای ۳ قسمت شدن نمونه، نمونه را به دو قسمت تقسیم کنیم.

رگرس ۸ تنها متغیر مجازی برای شیب تعداد پلیس سرانه را مربوط به نیمی از شهرستان‌ها در همه زمان‌ها در نظر گرفته است. نکته اول این است که نیمی از شهرستان‌ها به دلیل تغییر سیاست فقط در سال‌های ۸۶ به بعد با بقیه شهرستان‌ها متفاوت است و این شهرستان‌ها در سال‌های قبل از ۸۶ از لحاظ کیفی متفاوت نیستند. پس بهتر است متغیر مجازی فقط برای نیمی از شهرستان‌ها در سال‌های ۸۶ و ۸۷ تعریف شود، نه همه زمان‌ها. و ایراد دوم این است که این مدل عرض از مبدا را ثابت فرض کرده است و محدودیت بر روی مدل قرار داده است که مطلوب نیست.

یکی از ایرادات رگرس ۱۰ نیز این است که با تعریف متغیر مجازی فقط برای شیب تعداد پلیس سرانه قید اضافه بر مدل تحمیل کرده است و عرض از مبدأ را ثابت گرفته است. پس در آخر به دو رگرسیون ۹ و ۱۱ می‌رسیم. رگرسیون ۱۱ همه نوع ترکیب متغیر مجازی را هم برای نیمی از شهرستان‌ها در سال‌های ۸۶ و ۸۷ ، و هم برای نیمی از شهرستان‌ها در همه زمان‌ها دارد. رپرس ۹ همان رگرس ۱۱ است ولی متغیر مجازی برای عرض از مبدأ برای نیمی از شهرستان‌ها در همه زمان‌ها را ندارد.

به نظر می‌رسد هم مدل ۹ و هم مدل ۱۱ برای پاسخ‌گویی به این قسمت مناسب است ولی دست آخر اگر بخواهیم یک مدل را انتخاب کنیم و تحلیل کمی مربوطه را انجام دهیم، مدل ۹ را انتخاب می‌کنیم، زیرا R^2 تعدیل شده برای آن بزرگ‌تر است.

```
. reg  lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc   lpolpc_p  lpolpc_86_87 lpolpc_p_86_87 d_p_86_87

      Source |       SS       df       MS              Number of obs =     630
-------------+------------------------------           F(  9,   620) =   91.68
       Model |  117.835258     9  13.0928065           Prob > F      =  0.0000
    Residual |  88.5450841   620  .142814652           R-squared     =  0.5710
-------------+------------------------------           Adj R-squared =  0.5647
       Total |  206.380342   629  .328108652           Root MSE      =  .37791

------------------------------------------------------------------------------
     lcrmrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lprbarr |  -.7268143   .0368283   -19.74   0.000    -.7991376   -.6544911
    lprbconv |  -.5536826   .0268676   -20.61   0.000    -.6064451   -.5009201
    lprbpris |    .233072   .0669378     3.48   0.001     .1016197    .3645243
     lavgsen |  -.0621449   .0553586    -1.12   0.262    -.1708579    .0465681
      lpolpc |   .3604659   .0317962    11.34   0.000     .2980247    .4229072
    lpolpc_p |   .0070712   .0055207     1.28   0.201    -.0037704    .0179127
lpolpc_86_87 |  -.0064424   .0073937    -0.87   0.384    -.0209621    .0080772
lpolpc_p_~87 |  -.1437422    .135233    -1.06   0.288    -.4093124     .121828
   d_p_86_87 |  -1.015014   .8849846    -1.15   0.252    -2.752945    .7229168
       _cons |  -2.218925   .2437125    -9.10   0.000    -2.697527   -1.740324
------------------------------------------------------------------------------
```

همانطور که از رگرس ۹ در بالا معلوم است، ضرایب مربوط به متغیرهای مجازی نیمی از شهرستان‌ها و سال‌های بعد از ۸۶ معنی‌دار نیست و از لحاظ آماری صفر است. در واقع ۱ درصد افزایش تعداد پلیس به صورت سرانه باعث افزایش ۰٫۳۶ درصدی جرم وجنایت می‌شود. خطای استاندارد این تخمین ۰٫۰۳۲ است.

ب) برای بررسی خطای تصریح از آزمون RESET Ramsey استفاده می‌کنیم، که اطلاعات مربوط به آن در رگرس (۷) آمده است. روند آزمون به این صورت است که برای ضرایب تخمین زده شده برای متغیرهای lcrmrte_hat_2 و lcrmrte_hat_3 آزمون F انجام می‌دهیم و آماره آزمون را حساب می‌کنیم. حال اگر P-value این عدد را حساب کنیم به خطای آزمون می‌رسیم و می‌توانیم آزمون را انجام دهیم.

ج) برای شهرستان‌هایی که به عنوان کاندیدای متغیر پرت در نظر داریم از متغیرهای مجازی استفاده می کنیم و معنی داری ضریب متغیرهای مجازی را بررسی می کنیم. کاندیدا ها را می‌توان از روی رسم توزیع متغیرهای موردنظر برای شهرها و بررسی نواحی کم احتمال این توزیع‌ها انتخاب کرد.

د) رگرس (۱۱) – با توجه به آماره t این اثرات چه در عرض از مبدا و چه در شیب بی تأثیر بوده اند.

ه) رگرس (۱۴) – ۰٫۴۱۳۷۷۱

ز) وجود درون زایی

ح) مالیات سرانه از طرفی در رگرس نبوده‌است. و در رگرسیون ۱۵ با تعداد پلیس سرانه هم‌بستگی دارد و طبق آزمون F رگرس ۱۵ این همبستگی معنی‌دار است. از طرفی از نظر شهودی به نظر نمی‌آید مالیات سرانه با جرم و جنایت هم‌بسته باشد. از این جهات مالیات سرانه IV مناسبی است.

ط) رگرسیون ۱۸ - خیر

سوال چهارم

الف) معادلات را به ازای متغیرهای درون‌زا، در هم جایگذاری می کنیم.

ب)

ردیف	معادله اول	معادله دوم
۱	Just Identified/ OLS	unidentified
۲	Just Identified/ OLS	Over identified /2SLS
۳	Just identified/ OLS	Just identified/ OLS
۴	Over identified /2SLS	unidentified
۵	Just identified/ 2SLS	Just identified/2SLS