بسم الله الرحمن الرحيم

ENDOGENEITY-MEASUREMENT ERROR

سمینار درس مدل های خطی۱

ارائه دهنده: فرزانه بیگی

دی ۹۷

روش حداقل مربعات معمولی

درآمارحداقل مربعات معمولی روشی است برای برآورد پارامترهای مجهول درمدل رگرسیون خطی ازطریق کمینه کردن اختلاف بین مقدار واقعی ومقدارمشاهده شده است .

مدل زیر را در نظر بگیرید:

$$y_i = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + u_i$$

فرضیاتی که برای مدل رگرسیون خطی وجود دارند به قرار زیر است .

فرض ۱. میانگین u_i ها صفراست .

$$E(u_i) = 0$$

فرض ۲ . عدم وجود خود همبستگی بین u ها

$$\text{cov}(u_i, u_j) = E(u_i u_j) = 0$$

فرض ۳ . یکسانی واریانس $\mathrm{var}(u_i) = \delta^2$

فرض ۴ . کوواریانس صفر بین u_i , x_i $\mathrm{cov}(u_i, x_i) = E(u_i x_i) = 0$

برای کاربرد روش حداقل مربعات معمولی در مدل رگرسیون خطی هیچ فرضی در مورد توزیع احتمالی جزء خطا ارائه نکردیم . تنها فرضی که در مورد u_i مد نظر قرارگرفت عبارت بودند از اینکه اجزاء مذکور دارای امید صفر وعدم همبستگی و واریانس ثابت است.

درون زایی در مدل رگرسیون چند گانه رخ می دهد.اگر حداقل یکی ازمتغیرهای

مستقل مدل رگرسیونی درون زا باشند .

اگر متغییردرون زا در مدل وجود داشته باشد مشکل درون زایی رخ میدهد

فرض ۴ نقض می شود .

$$E\left(u_i x_i\right) \neq 0$$

یعنی این جمله با خطا وابستگی خواهد داشت.

متغیربرون زا (مستقل):متغیری است که هیچگونه تاثیری از

سایرمتغیرهای موجود در مدل دریافت نمی کند.

متغیر درونزا (وابسته):متغیری است که از جانب سایرمتغیرهای

موجود درمدل تاثیرمی پذیرد.

تا به حال تمام متغیرهای مدل های رگرسیونی برون زا بودند.

اگر x درون زا باشد $\operatorname{cov}(x_i, e_i) \neq 0$

اگر x برون زا باشد $\operatorname{cov}(x_i, e_i) = 0$

عوامل درون زایی (باعث می شود X با e_i همبسته شود)

1.OMITTED VARIABLES

2.MEASUREMENT ERROR

3.SIMULTANEITY IN SIMULTANEOUS EQUATIONS MODELS

حضور متغیرهای درون زا در مدل های آماری ناسازگاری واریبی براوردگرهای معمول پارامترهای مدل را به دنبال دارد. روش های متعددی دراین حالت ارائه شد که مشکل ناسازگاری و اریبی را تنها در حالت بزرگ نمونه ای حل کرده اند .یکی از این روش ها مبتنی بر استفاده متغیرهای ابزاری و روش حداقل مربعات دو مرحله ای است که باعث حذف درون زایی متغیرمورد نظر است.[1]

وقتی متغیردرون زا درمدل داشته باشیم دراین صورت برآوردگرهای OLS اریب خواهند شد مثلا

$$y = \beta_0 + \beta_1 x + \beta_2 T + \tau$$

$$X = \gamma_0 + \gamma_1 T + \varepsilon$$

پس با خطا دررابطه است پس با خطا همبستگی دارد پس

$$T = \frac{Y - \beta_0 - \beta_1 X - \varepsilon}{\beta_2}$$

$$X = \gamma_0 + \gamma_1 \frac{Y - \beta_0 - \beta_1 X - \tau}{\beta_2} + \varepsilon$$

اریبی و سازگاری OLS

$$y_i = b_0 + b_1 x_i + e_i$$

$$\hat{b}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{N}(\mathrm{x}_i - \bar{\mathrm{x}})(\mathrm{y}_i - \bar{y})}{\sum_{i=1}^{N}(\mathrm{x}_i - \bar{\mathrm{x}})^2} = \frac{\mathrm{cov}(\mathrm{x}_i, \mathrm{y}_i)}{\mathrm{var}(\mathrm{x}_i)}$$

$$= \frac{\mathrm{cov}(\mathrm{x}_i, \mathrm{b}_0 + \mathrm{b}_1\mathrm{x}_i + \mathrm{e}_i)}{\mathrm{var}(\mathrm{x}_i)} \Rightarrow \hat{b_1} = b_1 + \frac{\mathrm{cov}(\mathrm{x}_i, \mathrm{e}_i)}{\mathrm{var}(\mathrm{x}_i)}$$

$$\frac{\mathrm{cov}(\mathrm{x}_i, [\mathrm{b}_0 + \mathrm{b}_1\mathrm{x}_i + \mathrm{e}_i])}{\mathrm{var}(\mathrm{x}_i)} = \frac{\mathrm{cov}(\mathrm{x}_i, \mathrm{b}_0) + \mathrm{cov}(\mathrm{x}_i, \mathrm{b}_1\mathrm{x}_i) + \mathrm{cov}(\mathrm{x}_i, \mathrm{e}_i)}{\mathrm{var}(\mathrm{x}_i)}$$

$$\frac{0 + b_1\,\mathrm{var}(\mathrm{x}_i) + \mathrm{cov}(\mathrm{x}_i, \mathrm{e}_i)}{\mathrm{var}(\mathrm{x}_i)} = \frac{0}{\mathrm{var}(\mathrm{x}_i)} + \frac{b_1\,\mathrm{var}(\mathrm{x}_i)}{\mathrm{var}(\mathrm{x}_i)} + \frac{\mathrm{cov}(\mathrm{x}_i, \mathrm{e}_i)}{\mathrm{var}(\mathrm{x}_i)}$$

$$= b_1 + \frac{\operatorname{cov}(x_i, e_i)}{\operatorname{var}(x_i)}$$

به دنبال این معادله

$$\hat{b}_1 = b_1 + \frac{\operatorname{cov}(x_i, e_i)}{\operatorname{var}(x_i)}$$

است که برای اینکه OLS سازگار باشد باید $\operatorname{cov}(x_i, e_i) = 0$

بر قرارباشداما

$$\operatorname{cov}(x_i, e_i) \neq 0 \quad \Rightarrow \quad \hat{b}_1 = b_1 + \frac{\operatorname{cov}(x_i, e_i)}{\operatorname{var}(x_i)}$$

1. MEASUREMENT ERROR AN IN INDEPENDENT

مدل رگرسیون ساده را در نظر بگیرید

$$y_i = \beta_0 + \beta_1 X_i + u_i$$

x_i باخطا ها اندازه گیری می شود. در این صورت به جای X_i $\tilde{X}_i = X_i + e_i$

را داریم. فرض می کنیم X_i و e_i ناهمبسته است . یعنی $E(X_i e_i) = 0$

بنابراین معادلات رگرسیون به صورت زیر است .

$$Y_i = \beta_0 + \beta_1 \tilde{X}_i + u_i - \beta_1 e_i$$

در این صورت درون زایی رخ می دهد

$$= \beta_0 + \beta_1 \hat{X}_i + v_i$$

درون زا بودن متغیرها باعث اریبی و

$$E(\hat{X}_i V_i) = E((X_i + e_i)(u_i - \beta_1 e_i))$$

سازگاری براورد گر حاصل از روش براورد کمترین توانهای دوم معمولی OLS می شود

$$= \beta_1 \operatorname{var}(e_i) \neq 0$$

برآورد گر OLS از β_1

$$\beta_1 = \beta_1 + \sum_{i=1}^{N}(X_i - \bar{X})(\mathrm{u}_i - \beta_1 e_i) \Big/ \sum_{i=1}^{N}(X_i - \bar{X})$$

$$\xrightarrow{\mathrm{P}} \beta_1 - \beta_1 \frac{\mathrm{var(e)}}{\mathrm{var(X)} + \mathrm{var(e)}}$$

$$p\lim(\hat{\beta}_1) = \beta_1 \frac{\mathrm{var(x)}}{\mathrm{var(X)} + \mathrm{var(e)}} \le \beta_1$$

بنابراین براورد گر OLS ناسازگار است

1. MEASUREMENT ERROR IN DEPENDENT VARIABLE

فرض کنید که

$$y = b_0 + b_1 x + u \qquad true$$

$$y = y^{true} + e \qquad observe$$

یعنی متغییر وابسته با خطا اندازه گیری می شود

با جایگذاری معادله(۲)در(۱) داریم

$$y - e = b_0 + b_1 x + u$$

$$y = b_0 + b_1 x + u + e$$

$$y = b_0 + b_1 x + v \qquad st \quad v = u + e$$

$$E(\mathrm{e}) = E(\mathrm{u}) = 0$$

$$\mathrm{cov}(\mathrm{x},\mathrm{u}) = \mathrm{cov}(\mathrm{x},\mathrm{e}) = 0$$

$$E(y_i e_i) \neq 0$$

بنابراین تخمین OLSنا اریب هستند وواریانس

$$\mathrm{var}(\hat{\beta}) = \frac{\sigma_u{}^2}{N\,\mathrm{var}(x)} \qquad \textit{true}$$

$$\mathrm{var}(\hat{\beta}) = \frac{\sigma_u{}^2 + \sigma^2{}_e}{N\,\mathrm{var}(\mathrm{x})} \qquad \textit{estimate}$$

2.simultaity

یعنی علاوه بر اینکه متغیر پاسخ درون زا است حداقل یکی از متغیر های مستقل مدل هم درون زا است . برای حل این مشکل باید متغیر درون زا با خطا همبستگی نداشته باشد پس از روش متغیر ابزاری برای براورد پارامتر ها استفاده می کنیم.

فرض کنید x,y دو متغیر با هم تعریف می شوند که هر کدام از یکدیگر تاثیر

می گیرند. فرض کنید که دو معادله ساختاری وجود دارد.

$$\begin{aligned} y_i &= \beta_1 X_i + \gamma_1 Z_i + u_i \\ Z_i &= \beta_2 X_i + \gamma_2 y_i + v_i \end{aligned}$$

برآورد هردو معادله منجر به درون زایی می شود .در اولین معادله ساختاری $E(Z_i u_i) \neq 0$

را داریم . برای به دست آوردن Z_i فرض می کنیم که

$$1 - \gamma_1 \gamma_2 \neq 0$$

$$z_i = \frac{\beta_2 + \gamma_2 \beta_1}{1 - \gamma_1 \gamma_2} X_i + \frac{1}{1 - \gamma_1 \gamma_2} v_i + \frac{\gamma_2}{1 - \gamma_1 \gamma_2} u_i$$

فرض کنید X_i, v_i با u_i ناهمبسته هستند بنابراین داریم:

$$E(Z_i u_i) = \frac{\gamma_2}{1 - \gamma_1 \gamma_2} E(u_i u_i) \neq 0$$

3.OMITTED VARIABLE

$$Y_1 = b_1 x_t + b_2 w_t + e_t$$

True

فرض کنیم داده x_2 از مدل حذف شده در این صورت مدل زیر را تخمین بزنید

$$Y_t = b_1 x_t + v_t$$

we estimate

که $v_t = b_1 w_t + v_t$ است. در این صورت

اگر $\operatorname{cov}(x, w) \neq 0$

بنابراین $\operatorname{cov}(x, v) \neq 0$

راهکار برای حل این مشکل استفاده از متغیر ابزاری و روش حداقل مربعات

دو مرحله ای است.(می توانند براورد گر های سازگار تولید کنند)

متغیرابزاری برای تخمین اثرتصادفی تعدادی از متغیرهای x روی متغیر

دیگری مثل y یک ابزار سومی است مانند z است

دو شرط برای متغیر ابزاری

$$instrument \quad relevance : \mathrm{cov}(Z_i , X_i) \neq 0$$

$$instrument \quad exogeneity : \mathrm{cov}(z_i , e_i) = 0$$

برای مدل زیر

$$y_i = \beta_0 + \beta_1 X_i + u_i$$

متغییر ابزاری z_i را در نظر بگیرید.به طوری که

$$\mathrm{cov}(z, x) \neq 0$$
$$\mathrm{cov}(z, u) = 0$$

براوردگر iv

$$0 = \mathrm{cov}(z, u) = \mathrm{cov}(z, y - \beta_0 - \beta_1 X_i)$$
$$= \mathrm{cov}(z, y) - \beta_1 \mathrm{cov}(z, x)$$

و بنابراین داریم

$$\hat{\beta}_{1,\mathrm{IV}} = \frac{\mathrm{cov}(z, y)}{\mathrm{cov}(z, x)}$$

$$\hat{\beta}_{1,IV} = \beta_1 + \frac{\sum_{i=1}^{N}(Z_i - \bar{z})u_i}{\sum_{i=1}^{N}(Z_i - \bar{z})(x_i - \bar{x})}$$

بنابراین برآوردگر IV سازگاراست.

$$\frac{\sum_{i=1}^{N}(z_i - \bar{z})u_i}{N} \xrightarrow{P} 0 \, as \to \infty$$

$$P\lim(\hat{\beta}_{1,iv}) = \beta_1$$

واریانس مجانبی برآورد کننده IV به صورت زیر است.

$$\operatorname{var}(\hat{\beta}_{1,iv}) \approx \sigma^2_u \frac{\operatorname{var}(z)}{N \, . \operatorname{cov}(z, x)^2}$$

با مقایسه با $\beta_1^{ols} = \frac{\mathrm{cov}(x, y)}{\mathrm{var}(x)}$ براورد کننده iv سازگار است در

حالی که براورد کننده OLS ناسازگار است.

روش د یگر برای برآورد پارامتر های مدل های درون زا روش کمترین توان

های دوم دو مرحله ای است که دقت بهتری نسبت به روش کمترین توان های

دوم معمولی دارد اما برآورد گرهای حاصل از این روش نیز تنها در حالت

بزرگ نمونه ای ناریب و سازگار است

یکی از روش هایی که می توانیم به کار ببریم روش حداقل مربعات دو مرحله

ای $2sls$ است. در مرحله اول هر یک از متغیر های درون زا در معادله هدف

روی همه متغیر های برون زا درمدل رگرسیون می شود و مقدارپیش بینی شده

از این رگرسیون به دست می آید. بخشی از x که با e ناهمبسته است را

جدا می کنیم

$$x_i = b_0 + b_1 z_i + v_i$$

مرحله ۱.رگرسیون کردن x روی z

و سپس ارزش پیش بینی شده را ذخیره می کنیم. $\tilde{x} = p_z X$

درمرحله دوم رگرسیون هدف به صورت معمول تخمین زده

می شود و هر یک از متغیرهای برون زا با ارزش پیش بینی شده از اولین

مرحله مدل جایگزین می شود

روش های آزمون

ENDOGENEITY

Endogeneity روش های آزمون

	oLs	iv
X not endogenous	Consistent&Efficint	**Consistent &inefficint**
X endogenous	Inconsistent	**consistent**

HAUSMAN TEST آزمون

$H_0 : \operatorname{cov}(x, u) = 0$

β_{IV} *and* β_{ols} *are similar*

$H_1 : \operatorname{cov}(x, u) \neq 0$

β_{IV} *and* β_{ols} *are different*

آزمون STATISTIC

$$H = (\hat{\beta}^{ols} - \hat{\beta}^{Iv})'(\mathrm{v}(\hat{\beta}^{ols} - \hat{\beta}^{Iv})^{-1}(\hat{\beta}^{ols} - \hat{\beta}^{Iv}) \sim \chi_k{}^2$$

K که تعداد متغیرهای رگرسیونی است.

منابع:

1.[BLACK,BERGER,ANDSCOTT(1998)].BOUNDINGPARAMETERESTIMATESWITHNONCLASSICALMEASUREMENTERROR,.JOURNALOFTHE AMERICANSTATISTICALASSOCIATION95,739-748.
2.[HYSLOPANDIMBENS(2001)].BIASFROMCLASSICALANDOTHERFORMSOF MEASUREMENTERROR,.JOURNALOFBUSINESSAND ECONOMICSTATISTICS19,475-481
3.[AIGNER(1973)].REGRESSIONWITHABINARYINDEPENDENTVARIABLESUBJECTTO ERRORSOFOBSERVATION,.JOURNALOFECONOMETRICS1,49-60

با تشکر از توجه شما