جزوه آموزشی اقتصادسنجی استاد سیدامیرحسین حسینی

فصل اول: مقدمه و کلیات اقتصادسنجی

۱. اقتصادسنجی چیست؟ (ترکیب تئوری اقتصادی، ریاضیات و آمار).

۲. اهداف اقتصادسنجی (آزمون نظریه‌ها، برآورد ضرایب و پیش‌بینی).

۳. انواع داده‌ها در پژوهش‌های تجربی:

- داده‌های مقطعی (Cross-Sectional)
- داده‌های سری زمانی (Time Series)
- داده‌های ترکیبی و پانل (Panel Data)

٤. مراحل انجام یک پژوهش اقتصادسنجی (مشخص‌سازی مدل، جمع‌آوری داده، برآورد، آزمون فرضیه‌ها و اعتبارسنجی).

فصل دوم: مدل رگرسیون خطی ساده (دو متغیره)

۱. مفهوم رگرسیون و تفاوت آن با همبستگی (Causation vs. Correlation).

۲. تابع رگرسیون جامعه (PRF) و تابع رگرسیون نمونه (SRF).

۳. ماهیت جزء اخلال تصادفی (u_i).

٤. روش حداقل مربعات معمولی (OLS): استخراج جبری و هندسی ضرایب.

٥. فروض اساسی کلاسیک رگرسیون خطی (فروض گاوس–مارکوف).

٦. قضیه گاوس–مارکوف و ویژگی‌های بهترین برآوردکننده خطی نااریب (BLUE).

۷. ضریب تعیین (R^2) و معیار نیکویی برازش.

فصل سوم: استنباط آماری در رگرسیون خطی ساده

۱. فرض نرمال بودن پسماندها و ویژگی‌های توزیعی برآوردکننده‌ها.

۲. فاصله اطمینان برای پارامترهای مدل.

۳. آزمون فرضیه‌ها برای ضرایب (آزمون tاستیودنت).

٤. جدول تحلیل واریانس (ANOVA) و آزمون کلیت رگرسیون (F).

٥. خطاهای رایج در تفسیر ضرایب رگرسیونی.

فصل چهارم: مدل رگرسیون خطی چندمتغیره

۱. تصریح مدل رگرسیون چندمتغیره و نمادگذاری ماتریسی (اختیاری متناسب با سطح کلاس).

۲. تفسیر ضرایب رگرسیون جزئی (مفهوم *ceteris paribus* یا با فرض ثبات سایر شرایط).

۳. ضریب تعیین تعدیل‌شده ($\bar{R}^2$) و تفاوت آن با R^2.

٤. آزمون فرضیه‌های مرکب و مقید (آزمون Fبرای چند متغیر هم‌زمان).

٥. متغیرهای مجازی (Dummy Variables):

- متغیرهای کیفی در عرض از مبدأ و شیب.
- تله متغیر مجازی (Dummy Variable Trap).
- آزمون شکست ساختاری (آزمون چاو – Chow Test).

فصل پنجم: نقض فروض کلاسیک (تشخیص و درمان)

۱. **هم‌خطی چندگانه (Multicollinearity):**

- ماهیت و دلایل بروز.
- پیامدها برای برآوردکننده‌های OLS.
- روش‌های شناسایی (ماتریس همبستگی، شاخص VIF، تحمل).
- راهکارهای رفع هم‌خطی.

۲. **ناهمسانی واریانس (Heteroscedasticity):**

- ماهیت و دلایل بروز (به‌ویژه در داده‌های مقطعی).
- پیامدها برای استنباط آماری.

- آزمون‌های تشخیص (پارک، گلجسر، وایت، بروش-پاگان).
- روش‌های درمان (حداقل مربعات تعمیم‌یافته GLS، خطاهای استاندارد مقاوم وایت).

۳. **خودهمبستگی (Autocorrelation):**

- ماهیت و دلایل (به‌ویژه در سری‌های زمانی).
- پیامدها برای مدل.
- روش‌های تشخیص (دوربین-واتسون، بروش-گادفری/LM).
- روش‌های رفع (تعدیل مدل، تبدیل کوکران-اورکات، روش نیوی-وست).

فصل ششم: خطاهای تصریح مدل و انتخاب متغیرها

۱. حذف متغیر مرتبط (Omitted Variable Bias).

۲. گنجاندن متغیر نامرتبط (Irrelevant Variable).

۳. فرم تبعی نامناسب (خطی در مقابل لگاریتمی، درجه دو، تعاملی).

٤. آزمون‌های خطای تصریح (مانند آزمون RESET رمزی).

فصل هفتم: آشنایی مقدماتی با مباحث پیشرفته (اختیاری / سطح تکمیلی)

۱. **مقدمه‌ای بر سری‌های زمانی مدرن:**

- پایایی (Stationarity) و ریشه واحد (آزمون دیکی-فولر تعمیم‌یافته ADF).
- مفهوم رگرسیون کاذب (Spurious Regression) و هم‌انباشتگی (Cointegration).

۲. **مقدمه‌ای بر داده‌های تابلویی (Panel Data):**

- مزایای داده‌های پانل.
- مدل اثرات ثابت (Fixed Effects) در برابر اثرات تصادفی (Random Effects) و آزمون هاسمن.

۳. **متغیرهای ابزاری (IV) و درون‌زایی (Endogeneity):**

- مفهوم متغیرهای درون‌زا و برون‌زا.

- روش حداقل مربعات دومرحله‌ای (SLS۲).

موضوع ۱: اقتصادسنجی چیست؟ (ترکیب تئوری اقتصادی، ریاضیات و آمار)

۱.۱. ریشه‌شناسی، تاریخچه و پیدایش دیسیپلین

اصطلاح اقتصادسنجی (Econometrics) مرکب از دو واژه یونانی *oikonomia* (مدیریت خانوار/اقتصاد) و *metron* (سنجش و اندازه‌گیری) است. نخستین بار پاول سارانتس در سال ۱۹۱۰ این واژه را به کار برد، اما تثبیت مفهومی و معرفت‌شناختی آن توسط راگنار فریش (Ragnar Frisch) در سال ۱۹۲۶ صورت گرفت. فریش به‌همراه یان تینبرگن (Jan Tinbergen) در سال ۱۹۳۰ «انجمن بین‌المللی اقتصادسنجی» (The Econometric Society) و مجله مشهور *Econometrica* (۱۹۳۳) را پایه‌گذاری کردند؛ این دو در سال ۱۹۶۹ نخستین جایزه نوبل علوم اقتصادی را به دلیل تکوین مدل‌های پویای اقتصادسنجی دریافت نمودند.

۱.۲. ماهیت هم‌افزایی سه‌گانه

اقتصادسنجی جمع جبری ساده یا پیوند صوری سه رشته نیست، بلکه فصل مشترک و ترکیب هم‌افزا از سه ضلع است:

[نظریه اقتصادی]

\ /

**/ اقتصادسنجی **

**/ (هسته مشترک) **

۱. نظریه اقتصادی (Economic Theory): فرضیه‌هایی کیفی بر پایه رفتار بهینه‌یاب عوامل اقتصادی (بنگاه‌ها، مصرف‌کنندگان و دولت) صورت‌بندی می‌کند. برای مثال، نظریه مصرف کینز بیان می‌دارد که میل نهایی به مصرف (MPC) مثبت و کمتر از یک است. نظریه اقتصادی جهت رابطه را مشخص می‌کند، اما اندازه عددی ($Magnitude$) و عدم‌قطعیت تصادفی را به دست نمی‌دهد.

۲. اقتصاد ریاضی (Mathematical Economics): گزاره‌های کیفی تئوری را به معادلات جبری قطعی تبدیل می‌کند. مثلاً رابطه کینز به شکل خطی زیر نوشته می‌شود:

$$C = a + bY\ ,\ 0 < b < ۱$$

اشکال اقتصاد ریاضی این است که جهانی کاملاً قطعی (Deterministic) فرض می‌کند؛ به این معنا که اگر درآمد فردی Yباشد، مصرف او الزاماً و بدون کمترین انحراف برابر با $a + bY$خواهد بود.

۳. آمار اقتصادی و استنباطی (Mathematical Statistics & Economic): آمار اقتصادی متکفل جمع‌آوری داده‌ها، شاخص‌سازی، جدول‌بندی و ترسیم نمودارهاست (مانند حساب‌های ملی و شاخص قیمت مصرف‌کننده). آمار ریاضی نیز تئوری‌های احتمال، آزمون فرضیه و فواصل اطمینان را تدوین می‌کند. با این حال، روش‌های آمار کلاسیک عمدتاً بر داده‌های تجربیِ آزمایشگاهی (Experimental) که در کشاورزی یا پزشکی تولید می‌شوند، تمرکز دارند.

٤. جایگاه یگانه اقتصادسنجی: در اقتصاد، آزمایش‌های کنترل‌شده آزمایشگاهی به دلایل اخلاقی، فنی و مالی عموماً غیرممکن است؛ اقتصاددان نمی‌تواند برای بررسی اثر نقدینگی بر تورم، اقتصاد کشور را به دو نیمه تقسیم کرده و در یک نیمه حجم پول را دو برابر کند! بنابراین داده‌های اقتصادی از نوع مشاهده‌ای و غیرآزمایشگاهی (Observational / Non-experimental Data) هستند. اقتصادسنجی روش‌های ویژه آماری را برای انطباق تئوری بر چنین داده‌هایی تکوین می‌بخشد.

۱.۳. فرمول‌بندی و ساختار تصادفی

گذار از اقتصاد ریاضی به اقتصادسنجی با تبدیل مدل دترمینستیک به یک مدل تصادفی (Stochastic Model) رخ می‌دهد:

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

در اینجا u_iمتغیر تصادفی جزء اخلال است که نماینده تمام نیروهای پنهان، خطاهای اندازه‌گیری و تصادفی بودن رفتار بشری است. بدون جزء اخلال، روش‌های آماری فاقد موضوعیت خواهند بود.

۱.۴. مثال عددی و مفهومی

فرض کنید تابع تقاضای کالایی را مدل می‌کنیم. اقتصاد ریاضی می‌گوید:

$$Q_d = ۱۰۰ - ۲P$$

طبق این مدل، اگر قیمت ۱۰ هزار تومان باشد، تقاضا دقیقاً ۸۰ واحد است. اما اقتصادسنجی مدل را به این شکل می‌بیند:

$$Q_{di} = \beta_0 + \beta_1 P_i + u_i$$

در یک نمونه ۱۰۰ نفری، حتی اگر قیمت برای همه ۱۰ هزار تومان باشد، تقاضای افراد به دلیل درآمد، سن، سلیقه و ترجیحات متفاوت، اعدادی نظیر ۷۲، ۸۵، ۷۸ و ۸۱ خواهد بود. اقتصادسنجی وظیفه دارد پارامترهای β_0 و β_1 را به نحوی برآورد کند که رفتار میانگین جامعه به درستی بازنمایی شود.

۱.۵. خطاهای رایج در درک ماهیت اقتصادسنجی

- خطای اول: تصور اینکه اقتصادسنجی صرفاً کار با نرم‌افزارهای آماری است (تکنیک‌زدگی بدون تسلط بر تئوری اقتصادی).
- خطای دوم: این‌همان‌پنداری علم داده (Data Science) یا یادگیری ماشین با اقتصادسنجی؛ یادگیری ماشین بر بیشینه‌سازی قدرت پیش‌بینی درون نمونه یا بیرون نمونه ($\hat{Y}$) تمرکز دارد، در حالی که هسته اقتصادسنجی بر تفسیر علی پارامترها ($\hat{\beta}$) و پاسخ به پرسش‌های سیاستی ضدفکتیو (Counterfactual Policy Analysis) متمرکز است.

منابع

- ابریشمی، حمید (۱۴۰۱). *مبانی اقتصادسنجی (ترجمه دامودار گجراتی)*، جلد اول، چاپ سی‌ام، تهران: انتشارات دانشگاه تهران. (فصل اول: ماهیت اقتصادسنجی).
- سوری، علی (۱۴۰۲). *اقتصادسنجی (همراه با کاربرد EViews و Stata)*، جلد اول، تهران: انتشارات فرهنگ‌شناسی. (فصل اول: کلیات).
- نوفرستی، محمد (۱۴۰۰). *مقدمه‌ای بر اقتصادسنجی*، تهران: انتشارات خدمات فرهنگی رسا.
- Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach*, 7th Edition, Boston: Cengage Learning. (Chapter 1: “The Nature of Econometrics and Economic Data”, pp. 1-19).
- Stock, J. H., & Watson, M. W. (2019). *Introduction to Econometrics*, 4th Edition, New York: Pearson. (Chapter 1: “Economic Questions and Data”, pp. 31-45).
- Greene, W. H. (2018). *Econometric Analysis*, 8th Edition, New York: Pearson. (Chapter 1: “Econometrics”, pp. 1-12).

موضوع ۲: اهداف اقتصادسنجی (آزمون نظریه‌ها، برآورد ضرایب و پیش‌بینی)

۲.۱. ساختار تحلیلی اهداف سه‌گانه

اقتصادسنجی سه هدف متوالی و تفکیک‌ناپذیر را دنبال می‌کند:

[۱. برآورد پارامترها (Estimation)]

↓

[۲. آزمون نظریه‌ها (Hypothesis Testing)]

↓

[۳. پیش‌بینی و ارزیابی سیاست (Forecasting)]

. هدف اول: برآورد ضرایب و پارامترهای ساختاری (Parameter Estimation)

نظریه‌های اقتصادی از متغیرها و کشش‌ها به صورت مفهومی سخن می‌گویند؛ سیاست‌گذار برای تصمیم‌گیری نیازمند مقادیر عددی دقیق است.

- **صورت‌بندی ریاضی:** اگر مدل تابع تولید کاب–داگلاس به صورت لگاریتمی باشد:

$$\ln(Y) = \beta_0 + \beta_1 \ln(L) + \beta_2 \ln(K) + u$$

برآورد اقتصادسنجی مقادیر عددی عددی $\hat{\beta}_1$(کشش تولیدی نیروی کار) و $\hat{\beta}_2$(کشش تولیدی سرمایه) را استخراج می‌کند.

- **اهمیت اقتصادی:** آیا بازدهی نسبت به مقیاس صعودی است ($\beta_1 + \beta_2 > 1$)، ثابت است ($\beta_1 + \beta_2 = 1$) یا نزولی؟ پاسخ تنها از مسیر برآورد عددی پارامترها ممکن است.

۲.۳. هدف دوم: آزمون نظریه‌ها و فرضیات رفتاری (Hypothesis Testing)

بسیاری از گزاره‌های اقتصادی در قالب فرضیه‌های دوگانه مطرح می‌شوند:

- **نظریه هم‌ارزی ریکاردویی:** آیا کسری بودجه دولت اثری بر مصرف بخش خصوصی ندارد؟ ($H_0: \beta_{Deficit} = 0$).
- **نظریه بازار کار:** آیا شکاف دستمزد جنسیتی پس از کنترل سابقه و تحصیلات معنادار است؟ ($H_0: \beta_{Gender} = 0$ در برابر $H_1: \beta_{Gender} \neq 0$).

- **روش‌شناسی:** در این گام با استفاده از آماره‌های آزمون (t، F، χ^2) و تعیین سطح معناداری ($\alpha = ۰.۰۵$ یا ۰.۰۱)، رد یا عدم رد فرضیه صفر به آزمون گذاشته می‌شود.

۲.۴. هدف سوم: پیش‌بینی مقادیر آتی و ارزیابی سیاست‌ها (Policy Evaluation & Forecasting)

پیش‌بینی در اقتصادسنجی به دو شکل صورت می‌پذیرد:

۱. **پیش‌بینی نقطه‌ای و فاصله‌ای (Interval Forecast & Point):** پیش‌بینی نرخ تورم سال آینده به شرط تحقق نرخ رشد پایه پولی معین.

$$\hat{Y}_{T+۱} = \hat{\beta}_۰ + \hat{\beta}_۱ X_{T+۱}$$

همراه با خطای استاندارد پیش‌بینی ($SE(\hat{Y}_{T+۱})$) و ایجاد فاصله اطمینان ۹۵٪:

$$\hat{Y}_{T+۱} \pm t_{\alpha/۲} \times SE(\text{Forecast})$$

۲. **ارزیابی سناریوهای سیاستی (Policy Evaluation):** بررسی این مسئله که اگر بانک مرکزی نرخ بهره سیاستی را ۲ درصد افزایش دهد، بیکاری طی ۶ ماه آینده چه مسیری را خواهد پیمود.

۳. **نقد لوکاس (Lucas Critique - ۱۹۷۶):** یکی از ژرف‌ترین تذکرات در پیش‌بینی اقتصادسنجی که رابرت لوکاس مطرح کرد: «پارامترهای معادلات اقتصادسنجی کلان ثابت نیستند؛ زمانی که سیاست دولت تغییر می‌کند، انتظارات عقلایی مردم تغییر یافته و در نتیجه ضرایب رفتاری (β) دگرگون می‌شوند.» از این رو، مدل‌های اقتصادسنجی باید پایه‌های خرد (Microfoundations) داشته باشند.

۲.۵. جدول مقایسه‌ای اهداف اقتصادسنجی

هدف	تمرکز اصلی	ابزار ریاضی/آماری	مثال کاربردی در مدیریت و اقتصاد
برآورد ضرایب	محاسبه اندازه عددی اثرات	حداقل مربعات (OLS)، درست‌نمایی بیشینه (MLE)	کشش قیمتی تقاضای بنزین چقدر است؟ ($\hat{\epsilon}_p = -0.35$)
آزمون فرضیه	داوری درباره درستی تئوری	آماره t ، آزمون والد (Wald)، آزمون نسبت درست‌نمایی (LR)	آیا آموزش عالی نرخ بازدهی مثبت دارد؟ ($H_0 : \beta = 0$)
پیش‌بینی	تصویر آینده و شبیه‌سازی سیاست	امید ریاضی شرطی، مدل‌های پویای زمانی و VAR	اگر نرخ ارز ۳۰% جهش کند، تورم پایان سال چند درصد خواهد بود؟

منابع

- دادگر، یداله و صیامی عراقی، ابراهیم (۱۳۹۹). اقتصادسنجی مقدماتی، تهران: انتشارات دانشگاه تربیت مدرس.

- سوری، علی (۱۴۰۲). اقتصادسنجی پیشرفته، جلد اول، چاپ هفتم، تهران: انتشارات نورعلم. (فصل دوم: ارزیابی فرضیه‌ها).
- برادران شرکاء، حمیدرضا و پیش‌بهار، اسماعیل (۱۳۹۵). اقتصادسنجی کاربردی، تهران: انتشارات دانشگاه علامه طباطبایی.
- Lucas, R. E. (1976). “Econometric Policy Evaluation: A Critique”, *Carnegie-Rochester Conference Series on Public Policy*, 1, pp. 19-46.
- Johnston, J., & DiNardo, J. (1997). *Econometric Methods*, 4th Edition, New York: McGraw-Hill. (Chapter 1).
- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*, 5th Edition, New York: McGraw-Hill. (Chapter 1: “The Nature of Regression Analysis”).

موضوع ۳: انواع داده‌ها در پژوهش‌های تجربی (مقطعی، سری زمانی، ترکیبی و پانل)

۳.۱. داده‌های مقطعی (Cross-Sectional Data)

داده‌های مقطعی شامل مشاهدات گردآوری‌شده از واحدهای گوناگون در **یک مقطع زمانی معین** هستند.

- **ساختار ریاضی:** مجموعه مشاهدات به صورت زیر نمایش داده می‌شود:

$$\{(Y_i, X_{1i}, X_{2i}, \dots, X_{ki})\}_{i=1}^{N}$$

که در آن زیرنویس iمعرف واحد مقطعی (فرد، خانوار، شرکت، شهر یا کشور) است و Nاندازه نمونه را نشان می‌دهد.

- **فرض نمونه‌گیری:** فرض اساسی بر **نمونه‌گیری تصادفی و متغیرهای تصادفی با توزیع یکسان و مستقل (i.i.d.)** استوار است:

$$(Y_i, X_i) \sim \text{i.i.d.}$$

- **چالش ساختاری عمده:** ناهمگنی ذاتی میان واحدهای مختلف. برای مثال اگر درآمد و مصرف ۵۰۰ خانوار را بسنجیم، واریانس خطای مربوط به خانوارهای پردرآمد بسیار بزرگتر از خانوارهای کم‌درآمد است؛ این وضعیت منجر به پدیده **ناهمسانی واریانس (Heteroskedasticity)** می‌شود.

۳.۲. داده‌های سری زمانی (Time Series Data)

داده‌های سری زمانی مشاهدات مربوط به **یک واحد معین** را در **طول زمان** با فواصل منظم (روزانه، ماهانه، فصلی، سالانه) ثبت می‌کنند.

- **ساختار ریاضی:**

$$\{(Y_t, X_{1t}, X_{2t}, \dots, X_{kt})\}_{t=1}^{T}$$

که در آن زیرنویس tنشان‌دهنده زمان است ($t = 1,2, \dots, T$).

- **ویژگی‌های ذاتی:**

۱. ترتیب زمانی غیرقابل تغییر است (نقض فرض جابجایی داده‌ها).

۲. مشاهدات متوالی به یکدیگر وابسته‌اند؛ یعنی $Cov(Y_t, Y_{t-1}) \neq 0$(پدیده وابستگی متوالی یا Serial Correlation).

۳. داده‌ها اغلب دارای روند (Trend) و ناپایایی (Non-stationarity) هستند. رگرسیون بر روی دو متغیر نامرتبط با روند صوتی یکسان به **رگرسیون کاذب (Spurious Regression)** با R^2بسیار بالا و آماره tبه ظاهر معنادار منتهی می‌شود.

۳.۳. مقاطع تلفیقی مکرر (Pooled Cross-Sections)

در این الگو، نمونه‌های مقطعی تصادفی در چند دوره زمانی متفاوت برداشت می‌شوند، اما **افراد یا شرکت‌های درون نمونه در دوره‌های مختلف یکسان نیستند**.

- **مثال:** طرح آمارگیری هزینه و درآمد خانوار مرکز آمار ایران؛ خانوارهای پرسش‌شونده در سال ۱۴۰۰ با خانوارهای سال ۱۴۰۱ متفاوتند.
- **کاربرد اصلی:** ارزیابی تأثیر تغییر سیاست‌ها، اصلاح قوانین، یا تکانه‌های برون‌زا با استفاده از روش **تفاضل در تفاضل (Difference-in-Differences - DiD)**. فرم معادله:

$$Y_{it} = \beta_0 + \beta_1 \text{Treated}_i + \beta_2 \text{Post}_t + \beta_3 (\text{Treated}_i \times \text{Post}_t) + u_{it}$$

که در آن ضریب تعاملی β_3اثر واقعی سیاست را تفکیک می‌کند.

۳.۴. داده‌های تابلویی / ترکیبی (Panel or Longitudinal Data)

داده‌های پانل ساختاری دوبعدی دارند که در آن **واحدهای مقطعی یکسان** $(i = 1, \dots, N)$ **در طول دوره‌های زمانی متوالی** $(t = 1, \dots, T)$ تعقیب و مشاهده می‌شوند.

- **ساختار ریاضی:**

$$\{Y_{it}, X_{1it}, \dots, X_{kit}\} \text{ for } i = 1, \dots, N \text{ and } t = 1, \dots, T$$

کل حجم مشاهدات برابر با $N \times T$است.

- **پانل متوازن و نامتوازن:** اگر برای تمام افراد اطلاعات تمام سال‌ها موجود باشد، پانل متوازن (Balanced) است؛ در غیر این صورت، نامتوازن (Unbalanced) خواهد بود.
- **مزیت تئوریک بی‌نظیر پانل:** تفکیک اثر ناهمگنی‌های فردی مشاهده‌نشده و غیرقابل اندازه‌گیری (Unobserved Individual Heterogeneity) مانند هوش، فرهنگ سازمانی یا استعداد ذاتی:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it}$$

که در آن α_iاثر اختصاصی واحد iاست که در طول زمان ثابت می‌ماند و با روش‌های اثرات ثابت (Fixed Effects) یا اثرات تصادفی (Random Effects) کنترل می‌شود.

۳.۵. جدول تطبیقی ماتریس داده‌ها

نوع داده	بعد مقطع (N)	بعد زمان (T)	تغییر هویت واحدها	چالش آماری غالب
مقطعی (Cross-Section)	بزرگ ($N > 100$)	یک مقطع ($T = 1$)	بدون تغییر	ناهمسانی واریانس، درون‌زایی
سری زمانی (Time Series)	یک واحد ($N = 1$)	بزرگ ($T > 30$)	بدون تغییر	خودهمبستگی، ریشه واحد، ناایستایی
تلفیقی (Pooled Cross-Section)	در هر دوره متفاوت	کوچک ($T = 2, 3, \dots$)	واحدهای جدید در هر سال	ناهمگنی ساختاری در طول زمان
پانل (Panel Data)	بزرگ یا کوچک	بزرگ یا کوچک	همان واحدهای ثابت	همبستگی متقاطع، اثرات ثابت انفرادی

منابع

- افلاطونی، عباس (۱۴۰۱). تجزیه و تحلیل آماری با EViews در تحقیقات حسابداری و مالی، چاپ ششم، تهران: انتشارات ترمه. (فصل پنجم: داده‌های ترکیبی).
- سوری، علی (۱۴۰۲). اقتصادسنجی پیشرفته (جلد دوم: پانل دیتا و سری‌های زمانی غیرخطی)، تهران: نشر نورعلم.
- فلاحی، محمدعلی (۱۳۹۸). اقتصادسنجی داده‌های پانل، مشهد: انتشارات دانشگاه فردوسی مشهد.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd Edition, Cambridge, MA: MIT Press. (Chapter 1 & 10).
- Baltagi, B. H. (2021). *Econometric Analysis of Panel Data*, 6th Edition, Cham: Springer. (Chapter 1: “Introduction”, pp. 1-14).
- Hamilton, J. D. (1994). *Time Series Analysis*, Princeton: Princeton University Press. (Chapter 1).

موضوع ۴: مراحل انجام یک پژوهش اقتصادسنجی (مشخص‌سازی، گردآوری، برآورد، آزمون و اعتبارسنجی)

۴.۱. دیاگرام کلان فرایند پژوهش

یک مطالعه تجربی اقتصادسنجی زنجیره‌ای از مراحل منطقی است که نقص در هر مرحله، اعتبار یافته‌های مراحل بعدی را باطل می‌سازد:

[۱. تدوین نظریه و تبیین مسئله]

↓

[۲. تصریح مدل اقتصادسنجی (Specification)]

↓

[۳. گردآوری و پاکسازی داده‌ها (Data Preparation)]

↓

[۴. برآورد ضرایب و ماتریس کواریانس (Estimation)]

↓

[۵. استنباط آماری و آزمون فرضیه‌ها (Inference)]

↓

[۶. آزمون‌های تشخیصی و اعتبارسنجی (Diagnostics & Validation)]

↓

آیا مدل معتبر است؟

/ \

(بله) (خیر) ——→ بازگشت به گام ۲ (اصلاح تصریح)

↓

[۷. تحلیل اثرات سیاستی و پیش‌بینی (Policy Application)]

گام‌های هفت‌گانه تشریحی

گام ۱: تدوین تئوری و تعیین روابط رفتاری

هیچ مدلی بدون نظریه آغاز نمی‌شود. پژوهشگر متغیر وابسته (Y) و متغیرهای توضیحی کلیدی (X) را مشخص می‌کند. مثلاً در تابع درآمد مینسر (Mincerian Wage Equation): دستمزد تابعی است از سال‌های تحصیل، سابقه کار و توانایی.

گام ۲: مشخص‌سازی مدل اقتصادسنجی (Model Specification)

تعیین فرم ریاضی معادله و ویژگی‌های متغیرها:

- آیا فرم خطی است ($Y = \beta_0 + \beta_1 X + u$)؟
- آیا فرم نیمه‌لگاریتمی است ($\ln(Y) = \beta_0 + \beta_1 X + \beta_2 X^2 + u$)؟
- متغیرهای کنترلی ضروری کدامند تا از تورش متغیر حذف‌شده اجتناب شود؟

گام ۳: جمع‌آوری، پالایش و غربالگری داده‌ها

استخراج داده از منابع موثق، بازرسی مقادیر مفقوده (Missing Values)، شناسایی داده‌های پرت (Outliers) و آزمون ایستایی قبل از برآورد.

گام ۴: برآورد مدل (Estimation)

انتخاب روش تخمین متناسب با ساختار داده و نقض فروض:

- اگر فروض کلاسیک برقرار باشد: روش حداقل مربعات معمولی (OLS).
- اگر درون‌زایی وجود داشته باشد: متغیرهای ابزاری و دومرحله‌ای (IV۲ / SLS).
- اگر متغیر وابسته محدود یا دوحالتی باشد: مدل‌های پروبیت و لاجیت با برآورد درست‌نمایی بیشینه (MLE).

گام ۵: آزمون فرضیه‌ها (Statistical Inference)

محاسبه خطاهای استاندارد، ارزیابی مقادیر t-statistic و مقادیر p-value و فاصله اطمینان ضرایب جهت سنجش معناداری آماری اثرات.

گام ۶: ارزیابی تشخیصی مدل (Misspecification Tests & Diagnostic Checking)

این مرحله سنگ بنای اعتبارسنجی علمی است. مدل باید در بوته این آزمون‌ها سنجیده شود:

- آزمون نرمال بودن پسماندها (Jarque-Bera).
- آزمون ناهمسانی واریانس (Breusch-Pagan / White).

- آزمون خودهمبستگی (Breusch-Godfrey LM test).
- آزمون صحت فرم تبعی (Ramsey RESET).
- بررسی هم‌خطی شدید (VIF).

گام ۷: استفاده از مدل برای سناریوسازی و پیش‌بینی

در صورتی که مدل از تمام فیلترهای تشخیصی عبور کند، جهت راهنمایی مدیران ارشد، برآورد منافع سرمایه‌گذاری، یا تدوین خط‌مشی‌های مالیاتی و پولی به کار گرفته می‌شود.

منابع

- فرخ‌نژاد، رضا و رزم‌جو، حمیدرضا (۱۴۰۰). *روش‌شناسی پژوهش در اقتصاد کاربردی*، تهران: نشر نی.
- سوری، علی (۱۴۰۲). *اقتصادسنجی (جلد اول)*، فصل سوم: مراحل پژوهش تجربی.
- خلیلی عراقی، منصور و سوری، علی (۱۳۹۴). *مقدمه‌ای بر مدلسازی اقتصادسنجی*، تهران: انتشارات دانشگاه تهران.
- Hendry, D. F. (۱۹۹۵). *Dynamic Econometrics*, Oxford: Oxford University Press. (Chapter ۱: "The Methodology of Econometric Modeling").
- Stock, J. H., & Watson, M. W. (۲۰۱۹). *Introduction to Econometrics*, ٤th Edition, Chapter ۹: "Assessing Studies Based on Multiple Regression", pp. ۳۱۳-۳٤٥.
- Kennedy, P. (۲۰۰۸). *A Guide to Econometrics*, ٦th Edition, Malden, MA: Blackwell Publishing. (Chapter ٥: "Specification").

موضوع ۵: مفهوم رگرسیون و تفاوت بنیادین آن با همبستگی (Causation vs. Correlation)

۵.۱. ریشه‌شناسی و قانون بازگشت به میانگین گالتون

واژه رگرسیون (Regression) به معنای «بازگشت یا پس‌روی» نخستین بار توسط **فرانسیس گالتون (Francis Galton)** در سال ۱۸۸۶ در مقاله‌ای با عنوان «Regression towards Mediocrity in Hereditary Stature» مطرح شد. گالتون دریافت که اگرچه پدران قدبلند، پسرانی قدبلند دارند، اما میانگین قد پسران تمایل دارد به سمت میانگین قد کل جامعه «پس‌روی یا بازگشت» کند. این مفهوم بعدها در آمار به تحلیل نحوه تغییرات متغیر وابسته در پاسخ به تغییرات متغیرهای مستقل تعمیم داده شد.

۵.۲. تفاوت ساختاری رگرسیون و همبستگی

بسیاری از پژوهشگران مبتدی همبستگی را با رگرسیون اشتباه می‌گیرند. وجوه تمایز این دو به شرح زیر است:

[ضریب همبستگی: ρ(X,Y)]

[+۱ ,-۱] تقارن محض | فاقد جهت علی | بازه

≠

[ضریب رگرسیون: β]

) | مقیاس متغیرهاY→Xعدم تقارن | جهت‌دار (

۱. **تقارن در برابر عدم تقارن:**

- ضریب همبستگی پیرسون کاملاً متقارن است:

$$r_{xy} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}} = r_{yx}$$

همبستگی Xبا Yدقیقاً برابر با همبستگی Yبا Xاست. در همبستگی هیچ متغیری علت و دیگری معلول نیست.

- تحلیل رگرسیون اساساً نامتقارن است؛ معادله $Y = \beta_0 + \beta_1 X + u$به معنای آن است که متغیر Xاثرگذار بر میانگین شرطی Yاست. رگرسیون معکوس ($X = \alpha_0 + \alpha_1 Y + v$) ضریب متفاوتی تولید می‌کند ($\hat{\alpha}_1 \neq 1/\hat{\beta}_1$).

۲. **واحدهای سنجش:**

- ضریب همبستگی یک عدد بدون بعد (Dimensionless) در بازه $[-۱, +۱]$است.
- ضریب رگرسیون $\beta_۱$دارای بعد و واحد سنجش است $(\frac{\text{واحد } Y}{\text{واحد } X})$ و بیانگر تغییر در واحد Yبه ازای یک واحد تغییر در Xاست.

۵.۳. تفاوت همبستگی آماری و علیت (Correlation is not Causation)

وجود همبستگی قوی بین دو متغیر الزاماً دال بر وجود رابطه علّی نیست. سه عامل عمده در ایجاد همبستگی بدون علیت نقش دارند:

۱. همبستگی کاذب و متغیر حذف‌شده مخدوش‌کننده (Confounding / Omitted Variable)

مثال کلاسیک: همبستگی مثبت بسیار شدید بین فروش بستنی (X) و تعداد غرق‌شدگان در سواحل دریا (Y). آیا خوردن بستنی موجب غرق شدن است؟ خیر! یک متغیر پنهان سوم به نام **دمای هوا** (Z) علت هر دو است:

دمای بالا (Z)

↖ ↗

(افزایش فروش بستنی X) (افزایش شنا در دریا Y)

اگر Zرا در رگرسیون کنترل نکنیم، رابطه بین Xو Yکاذب خواهد بود.

۲. علیت معکوس (Reverse Causality)

ما فرض می‌کنیم $X \to Y$، در حالی که جهت رابطه از $Y \to X$است. برای نمونه، همبستگی مثبت بین تعداد مأموران پلیس در محله‌ها (X) و نرخ جرم و جنایت (Y). رگرسیون ساده ممکن است نشان دهد حضور پلیس جرم را زیاد می‌کند! اما حقیقت این است که شهرداری به محله‌های ناامن‌تر مأمور بیشتری اعزام کرده است.

۳. هم‌زمانی (Simultaneity)

هر دو متغیر به طور هم‌زمان یکدیگر را تعیین می‌کنند $(X \leftrightarrow Y)$؛ مانند قیمت و مقدار در تعادل بازار، یا تورم و نرخ ارز.

۵.۴. چارچوب مدرن بررسی علیت: پیامدهای پتانسیل روبین (Rubin Causal Model)

در اقتصادسنجی مدرن، اثر علی متغیر درمان $(D_i \in \{۰, ۱\})$ بر متغیر دستمزد با تفاوت میانگین دو حالت پتانسیل تعریف می‌شود:

$$\text{اثر درمان برای فرد } i = Y_i(۱) - Y_i(۰)$$

مشکل اساسی استنباط علّی این است که ما در جهان واقعی برای هر فرد تنها یکی از این دو وضعیت را مشاهده می‌کنیم: فرد یا آموزش دانشگاهی دیده ($Y_i(1)$) یا ندیده ($Y_i(0)$)؛ حالت ندیده یک **وضعیت خلاف‌واقع (Counterfactual)** است. اقتصادسنجی تجربی از طریق متغیرهای ابزاری، داده‌های پانل و آزمون‌های ناپیوستگی به دنبال بازسازی شرایط علّی است.

منابع

- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی گجراتی، جلد اول، فصل ۱: «تحلیل رگرسیون و مفاهیم بنیادی».
- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۲: «تحلیل همبستگی و رگرسیون».
- رفیعی، علی و پناهی، حسین (۱۳۹۷). مقدمه‌ای بر استنباط علی در اقتصادسنجی، تبریز: انتشارات دانشگاه تبریز.
- Angrist, J. D., & Pischke, J.-S. (۲۰۱۵). Mastering 'Metrics: The Path from Cause to Effect, Princeton: Princeton University Press. (Chapter ۱: “Randomized Trials”, pp. ۱-٤٦).
- Pearl, J. (۲۰۰۹). Causality: Models, Reasoning, and Inference, ۲nd Edition, Cambridge: Cambridge University Press.
- Stock, J. H., & Watson, M. W. (۲۰۱۹). Introduction to Econometrics, ٤th Edition, Chapter ۱: “Causality and Ideal Conditions”, pp. ۳۸-٤۲.

موضوع ۶: تابع رگرسیون جامعه (PRF) و تابع رگرسیون نمونه (SRF)

۶.۱. مفهوم جامعه و تابع رگرسیون جامعه (Population Regression Function - PRF)

در نظریه اقتصادسنجی، جامعه آماری (Population) به کل مجموعه مشاهداتی اطلاق می‌شود که موضوع مطالعه هستند. فرض کنید متغیر تصادفی وابسته Y(مثلاً مخارج مصرفی خانوار) و متغیر توضیحی X(مثلاً درآمد هفتگی خانوار) باشند. برای هر مقدار معین از X، یک توزیع احتمال شرطی برای Yوجود دارد.

امید ریاضی شرطی (Conditional Expectation) Yبه شرط مقدار معین X_iرا با $E(Y \mid X_i)$نمایش می‌دهیم. مکان هندسی نقاط امید ریاضی شرطی، تشکیل‌دهنده **تابع رگرسیون جامعه (PRF)** است:

$$E(Y \mid X_i) = f(X_i)$$

اگر شکل تابعی خطی در نظر گرفته شود:

$$E(Y \mid X_i) = \beta_0 + \beta_1 X_i$$

در این معادله:

- β_0: عرض از مبدأ جامعه (تراز پایه مصرف در درآمد صفر).
- β_1: شیب جامعه (میل نهایی به مصرف، یعنی تغییر در امید ریاضی Yبه ازای یک واحد تغییر در X).

از آنجا که یک مشاهده انفرادی Y_iالزاماً روی خط امید ریاضی شرطی منطبق نیست، انحراف هر فرد از میانگین جامعه توسط جزء اخلال جامعه (u_i) تعریف می‌شود:

$$u_i = Y_i - E(Y \mid X_i) \implies Y_i = E(Y \mid X_i) + u_i$$

بنابراین شکل تصادفی (Stochastic) تابع رگرسیون جامعه عبارت است از:

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

۶.۲. تابع رگرسیون نمونه (Sample Regression Function - SRF)

در واقعیت به دلیل محدودیت منابع، دسترسی به کل اعضای جامعه ناممکن است؛ لذا یک نمونه تصادفی به حجم n از جامعه استخراج می‌شود: $\{(X_۱, Y_۱), (X_۲, Y_۲), \dots, (X_n, Y_n)\}$.

برآورد حاصل از این نمونه را **تابع رگرسیون نمونه (SRF)** می‌نامیم:

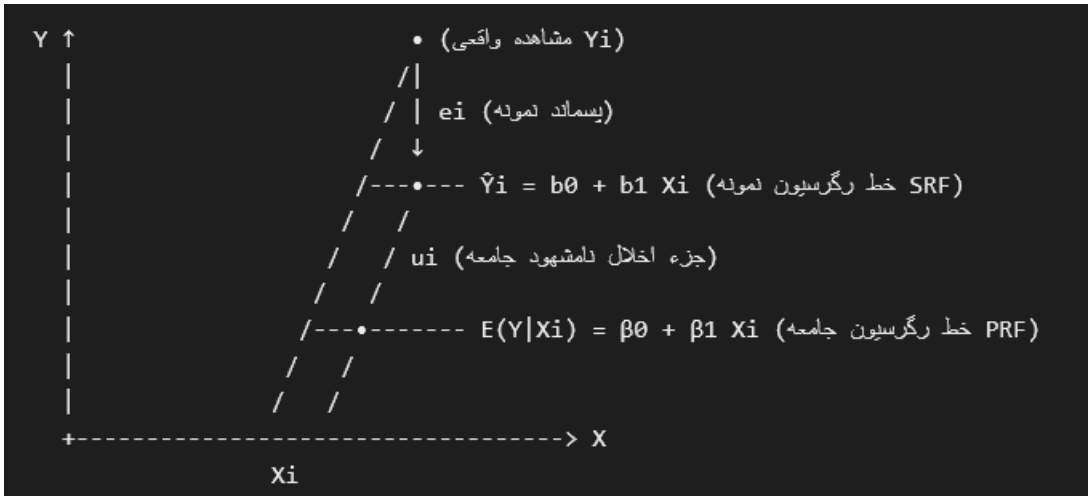
$$\hat{Y}_i = \hat{\beta}_۰ + \hat{\beta}_۱ X_i$$

در این رابطه:

- $\hat{Y}_i$: مقدار برازش‌شده (Fitted / Predicted Value) یا برآوردگر $E(Y \mid X_i)$.
- $\hat{\beta}_۰$: برآوردگر پارامتر ناپیدای جامعه $\beta_۰$.
- $\hat{\beta}_۱$: برآوردگر پارامتر ناپیدای جامعه $\beta_۱$.

انحراف هر مقدار مشاهده‌شده در نمونه (Y_i) از مقدار برازش‌شده توسط خط رگرسیون نمونه ($\hat{Y}_i$)، **پسماند نمونه (Residual)** نامیده شده و با e_i یا $\hat{u}_i$ نشان داده می‌شود:

$$e_i = Y_i - \hat{Y}_i \implies Y_i = \hat{\beta}_۰ + \hat{\beta}_۱ X_i + e_i$$



۶.۳. تمایز پارادایمی بین PRF و SRF

تفاوت میان PRF و SRF تفاوت میان **حقیقت ناپیدای جهان واقع** و **تقریب ذهنی ما بر پایه شواهد تجربی محدود** است:

بعد مقایسه	تابع رگرسیون جامعه (PRF)	تابع رگرسیون نمونه (SRF)
پارامترها	β_0, β_1 (ثابت ولی نامعلوم)	$\hat{\beta}_0, \hat{\beta}_1$ (متغیرهای تصادفی، تابعی از نمونه)
خطای انحراف	u_i (جزء اخلال تصادفی جامعه، غیرقابل مشاهده)	e_i (پسماند نمونه، دقیقاً قابل محاسبه و مشاهده)
رابطه مقادیر	$Y_i = \beta_0 + \beta_1 X_i + u_i$	$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$
تغییرپذیری	در کل جامعه واحد و یکتاست	از یک نمونه به نمونه دیگر تغییر می‌کند (Sampling Variability)

۶.۴. خطای رایج در ادبیات پژوهش

خلط نمادگذاری u_i(جامعه) با e_i(نمونه). بسیاری از دانشجویان پسماندهای محاسبه‌شده نرم‌افزار را جزء اخلال می‌نامند، حال آنکه پسماندها صرفاً سایه‌ای از جزء اخلال واقعی هستند که خود تابع برآوردگرهای $\hat{\beta}$هستند.

منابع

- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی (دامودار گجراتی)، تهران: انتشارات دانشگاه تهران، فصل ۲: «تحلیل رگرسیون دومتغیره: ایده‌های اساسی».
- سوری، علی (۱۴۰۲). اقتصادسنجی (همراه با کاربرد نرم‌افزارهای EViews و Stata)، جلد اول، فصل ۲: «رگرسیون دومتغیره»، صص ۳۵–۶۰.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics: A Modern Approach, ۷th ed., Chapter ۲: “The Simple Regression Model”, pp. ۲۰-۳۵.
- Gujarati, D. N., & Porter, D. C. (۲۰۰۹). Basic Econometrics, ۵th ed., McGraw-Hill, Chapter ۲: “Two-Variable Regression Analysis: Some Basic Ideas”.

موضوع ۷: ماهیت جزء اخلال تصادفی (u_i)

۷.۱. ضرورت وجودی جزء اخلال در مدل‌سازی اقتصادی

اگر جهان روابط اقتصادی قطعی (Deterministic) بود، اقتصادسنجی کاربردی نداشت. اما چرا رفتار انسان‌ها و متغیرهای اقتصادی دارای یک مؤلفه آشوبناک و تصادفی است؟ جزء اخلال (u_i) کانون پذیرش جهل سیستماتیک ما در اقتصادسنجی است و به دلایل بنیادین زیر ظاهر می‌شود:

۱. **حذف متغیرهای تأثیرگذار جزئی (Omission of Minor Variables):** در مدل بررسی مخارج مصرفی بر حسب درآمد، ده‌ها عامل دیگر نظیر اندازه خانواده، سن سرپرست، تمایلات مذهبی و انتظارات تورمی دخیل هستند. گنجاندن همه آن‌ها ناممکن یا داده‌هایشان در دسترس نیست. بنابراین، اثر تجمیعی این متغیرهای خرد در u_iادغام می‌گردد.

۲. **غیرقابل پیش‌بینی بودن رفتار انسانی (Randomness in Human Behavior):** حتی اگر شرایط محیطی و اقتصادی دو فرد دوقلو کاملاً یکسان باشد، ترجیحات و تصمیمات لحظه‌ای آن‌ها انحرافات تصادفی نشان می‌دهد.

۳. **خطای اندازه‌گیری در داده‌ها (Measurement Errors):** ارقام گزارش‌شده در پرسشنامه‌های خانوار یا صورت‌های مالی حسابداری اغلب با بیش‌اظهاری، کم‌اظهاری یا خطای گردکردن همراه‌اند. این خطاها درون u_iجذب می‌شوند.

٤. **شکل تبعی تقریبی (Proxy Functional Form):** اگر رابطه واقعی کمی منحنی یا لگاریتمی باشد ولی پژوهشگر یک تقریب خطی ساده را استفاده کند، تفاوت این دو فرم به جزء اخلال رانده می‌شود.

۷.۲. تفکیک مفهومی رفتار سیستماتیک از تصادفی

در معادله:

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{بخش سیستماتیک (معین)}} + \underbrace{u_i}_{\text{بخش تصادفی (نوفه)}}$$

بخش سیستماتیک نشان‌دهنده میانگین رفتار توده‌ای افراد جامعه است؛ در حالی که جزء اخلال، پراکندگی حول این خط میانگین را تبیین می‌نماید.

۷.۳. قضیه حد مرکزی و رفتار تصادفی u_i

بر اساس قضیه حد مرکزی (Central Limit Theorem - CLT)، اگر یک متغیر از جمع تعداد بسیار زیادی متغیر تصادفی مستقل با توزیع‌های گوناگون که هیچ‌یک اثر غالب بر دیگری ندارد تشکیل شده باشد، توزیع مجموع آن‌ها به سمت **توزیع نرمال** میل خواهد کرد:

$$u_i = \sum_{k=1}^{M} v_{k,i} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

این برهان، بنیان فرضیه کلاسیک «نرمال بودن جزء اخلال» را در نمونه‌های بزرگ شکل می‌دهد.

منابع

- نوفرستی، محمد (۱۴۰۰). مقدمه‌ای بر اقتصادسنجی، تهران: انتشارات خدمات فرهنگی رسا، فصل دوم: «مدل رگرسیون خطی ساده».
- دادگر، یداله و صیامی عراقی، ابراهیم (۱۳۹۹). اقتصادسنجی مقدماتی، تهران: دانشگاه تربیت مدرس، فصل ۳: «توزیع و مشخصات جزء خطای تصادفی».
- Goldberger, A. S. (۱۹۹۱). A Course in Econometrics, Harvard University Press, Chapter ۱۳: “Regression and Probability Models”.
- Kennedy, P. (۲۰۰۸). A Guide to Econometrics, ٦th ed., Blackwell, Chapter ۳: “The Classical Linear Regression Model”.

موضوع ۸: روش حداقل مربعات معمولی (OLS): استخراج جبری و هندسی ضرایب

۸.۱. منطق حداقل مربعات کارل فردریش گاوس

روش OLS (Ordinary Least Squares) که توسط **گاوس (Gauss)** در سال ۱۷۹۵ ابداع شد، به دنبال یافتن خطی است که مجموع مجذورات فواصل عمودی مشاهدات از خط برازش‌شده را کمینه‌سازی کند.

چرا مجذور انحرافات؟

- جمع ساده پسماندها ($\sum e_i$) انحرافات مثبت و منفی را خنثی می‌کند و به صفر کاذب می‌انجامد.
- استفاده از قدر مطلق ($\sum |e_i|$) به دلیل مشتق‌ناپذیری در نقطه صفر، استخراج تحلیلی و جبر ماتریسی را پیچیده می‌سازد.
- مجذور کردن جریمه تصاعدی برای انحرافات بزرگتر ایجاد می‌کند و محاسبات دیفرانسیلی ساده دارد.

۸.۲. استخراج جبری پارامترهای OLS (اثبات با مشتق جزئی)

تابع هدف عبارت است از کمینه‌سازی مجموع مربعات پسماندها (Sum of Squared Residuals - SSR):

$$S(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

شرط لازم برای کمینه‌سازی (First Order Conditions - FOC)، برابر صفر قرار دادن مشتقات جزئی نسبت به $\hat{\beta}_0$ و $\hat{\beta}_1$ است:

گام ۱: مشتق جزئی نسبت به $\hat{\beta}_0$:

$$\frac{\partial S}{\partial \hat{\beta}_0} = 2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-1) = 0$$

تقسیم بر ۲−:

$$\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \implies \sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i = 0 \quad \text{(معادله نرمال اول)}$$

تقسیم طرفین بر n:

$$\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X} = 0 \implies \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

این رابطه اثبات می‌کند که خط رگرسیون OLS همواره از نقطه میانگین داده‌ها $(\bar{X}, \bar{Y})$عبور می‌کند.

گام ۲: مشتق جزئی نسبت به $\widehat{\beta}_1$:

$$\frac{\partial S}{\partial \hat{\beta}_1} = 2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-X_i) = 0$$

$$\sum_{i=1}^{n} X_i \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0 \implies \sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0$$ (معادله نرمال دوم)

گام ۳: جایگذاری $\widehat{\beta}_0$در معادله نرمال دوم:

$$\sum_{i=1}^{n} X_i Y_i - \left(\bar{Y} - \hat{\beta}_1 \bar{X}\right) \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0$$

$$\sum_{i=1}^{n} X_i Y_i - \bar{Y} \sum X_i + \hat{\beta}_1 \bar{X} \sum X_i - \hat{\beta}_1 \sum X_i^2 = 0$$

می‌دانیم $\sum X_i = n\bar{X}$؛ بنابراین:

$$\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y} = \hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right)$$

سرانجام:

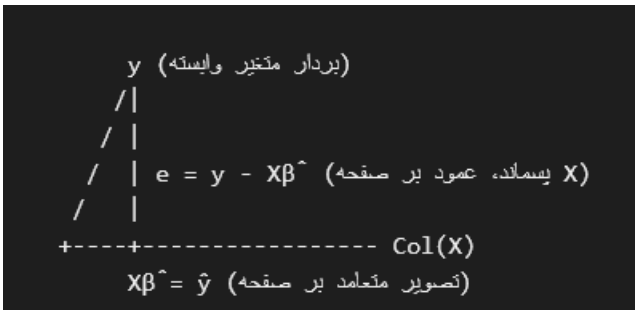
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\widehat{Cov}(X,Y)}{\widehat{Var}(X)}$$

که در آن $x_i = X_i - \bar{X}$ و $y_i = Y_i - \bar{Y}$ متغیرهای انحراف از میانگین هستند.

۸.۳. تفسیر هندسی OLS (تصویر متعامد – Orthogonal Projection)

در جبر خطی، اگر بردار متغیر وابسته $y \in \mathbb{R}^n$ و فضای ستونی ماتریس داده‌ها $Col(X)$ باشد، خطای برآورد بردار پسماند $e = y - X\hat{\beta}$ است. کمینه‌سازی طول بردار پسماند $\| e \|^2$ زمانی محقق می‌شود که بردار پسماند بر فضای ستونی X **عمود** (**Orthogonal**) باشد:

$$X'e = 0 \implies X'(y - X\hat{\beta}) = 0 \implies X'X\hat{\beta} = X'y \implies \hat{\beta} = (X'X)^{-1}X'y$$



منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۲: «برآورد مدل رگرسیون با روش OLS»، صص ۶۵–۸۵.
- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی، جلد اول، فصل ۳: «مدل رگرسیون خطی دو متغیره: مسئله تخمین».
- Greene, W. H. (۲۰۱۸). Econometric Analysis, ۸th ed., Pearson, Chapter ۳: “The Classical Linear Regression Model: Least Squares”, pp. ۵۱-۷۸.
- Strang, G. (۲۰۱۶). Introduction to Linear Algebra, ۵th ed., Wellesley-Cambridge Press, Section ٤,۳: “Projections and Least Squares”.

موضوع ۹: فروض اساسی کلاسیک رگرسیون خطی (فروض گاوس–مارکوف)

برای آنکه برآوردگرهای OLS ویژگی‌های آماری بهینه داشته باشند، مجموعه‌ای از فروض پایه‌ای در مورد ساختار مدل و توزیع احتمال جزء اخلال ضروری است:

فرض ۱: خطی بودن در پارامترها (Linearity in Parameters)

مدل در **پارامترهای** β خطی است، حتی اگر در متغیرهای X غیرخطی باشد:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \text{ یا } Y_i = \beta_0 + \beta_1 X_i^2 + u_i \text{ (معتبر)}$$

اما مدلی مانند $Y_i = \beta_0 + X_i^{\beta_1} + u_i$ در پارامتر غیرخطی است و با OLS استاندارد قابل برآورد نیست.

فرض ۲: میانگین شرطی صفر جزء اخلال (Zero Conditional Mean)

امید ریاضی خطای تصادفی به ازای هر مقدار معلوم متغیر توضیحی برابر با صفر است:

$$E(u_i \mid X_i) = 0$$

این فرض متضمن این است که هیچ متغیر مؤثری از قلم نیفتاده و متغیر توضیحی با خطای تصادفی همبستگی ندارد ($\text{Cov}(X_i, u_i) = 0$، یعنی X کاملاً **برون‌زا** (**Exogenous**) است).

فرض ۳: همسانی واریانس (Homoscedasticity)

واریانس جزء اخلال برای تمام مشاهدات ثابت و مستقل از مقادیر X_i است:

$$\text{Var}(u_i \mid X_i) = E(u_i^2 \mid X_i) = \sigma^2 \; \forall i = 1, \dots, n$$

اگر واریانس خطا برای داده‌های مختلف تغییر کند، مدل دچار **ناهمسانی واریانس** (**Heteroscedasticity**) است.

فرض ۴: عدم وجود خودهمبستگی (No Autocorrelation / Serial Independence)

کوواریانس بین هر دو جزء اخلال در مشاهدات متمایز برابر با صفر است:

$$\text{Cov}(u_i, u_j, \mid X_i X_j) = E(u_i u_j) = 0 \; (i \neq j)$$

این فرض در سری‌های زمانی حیاتی است؛ نقض آن به معنی همبستگی خطای یک دوره با دوره‌های قبل است.

فرض ۵: تغییرپذیری و عدم تصادفی بودن متغیر توضیحی

مقادیر Xدر طول نمونه ثابت و دارای واریانس مثبت هستند:

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 > 0$$

(متغیر Xنباید تک‌ارزش یا ثابت باشد، در غیر این صورت مخرج کسر شیب صفر و نامعین می‌شود).

جدول ساختار نقض فروض کلاسیک گاوس–مارکوف

عنوان فرض	نماد ریاضی	در صورت نقض چه نامیده می‌شود؟	پیامد اصلی OLS
برون‌زایی متغیرها	$E(u \mid X) = 0$	درون‌زایی (Endogeneity)	برآوردگرها اریب و ناسازگار می‌شوند
همسانی واریانس	$Var(u_i) = \sigma^2$	ناهمسانی واریانس (Heteroscedasticity)	برآوردگرها کارا نیستند؛ آزمون t و F نامعتبر است
استقلال خطاها	$Cov(u_i, u_j) = 0$	خودهمبستگی (Autocorrelation)	خطای استاندارد کمتر از واقع برآورد می‌شود
پراکندگی متغیر مستقل	$\sum x_i^2 > 0$	فقدان اطلاعات (No variation)	پارامترها غیرقابل محاسبه می‌شوند

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۴: «فروض کلاسیک مدل رگرسیون خطی».
- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی گجراتی، جلد اول، فصل ۳.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ۲ & ۳: “The Mechanics and Assumptions of OLS”.
- Stock, J. H., & Watson, M. W. (۲۰۱۹). Introduction to Econometrics, Chapter ٤: “Linear Regression with One Regressor”, pp. ۱۰۵-۱۳۵.

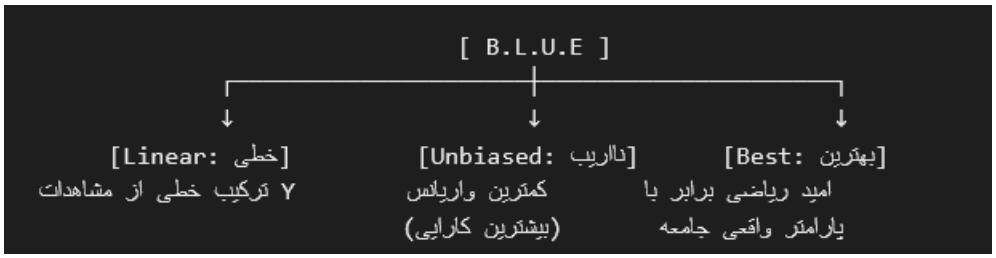
موضوع ۱۰: قضیه گاوس–مارکوف و ویژگی‌های بهترین برآوردکننده خطی نااریب (BLUE)

۱۰.۱. متن قضیه گاوس–مارکوف (The Gauss-Markov Theorem)

قضیه گاوس–مارکوف اعلام می‌دارد:

«تحت برقراری فروض کلاسیک رگرسیون (فروض ۱ تا ۵)، برآوردکننده‌های OLS یعنی $\hat{\beta}_0$ و $\hat{\beta}_1$، در میان تمامی برآوردکننده‌های خطی و نااریب، دارای کمترین واریانس هستند؛ یعنی **BLUE (Best Linear Unbiased Estimator)** می‌باشند.»

۱۰.۲. تشریح مؤلفه‌های چهارگانه ویژگی BLUE



۱. خطی بودن (Linearity)

برآوردکننده $\hat{\beta}_1$ یک تابع خطی از متغیر وابسته نمونه Y_i است.

اثبات:

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum_{i=1}^{n} \left(\frac{x_i}{\sum x_k^2} \right) Y_i = \sum_{i=1}^{n} k_i Y_i$$

که در آن وزن‌های $k_i = \frac{x_i}{\sum x_k^2}$ فقط تابعی از مقادیر معلوم X هستند و خواص جبری زیر را دارند:

$$\sum k_i = 0, \quad \sum k_i X_i = 1, \quad \sum k_i^2 = \frac{1}{\sum x_i^2}$$

۲. ناارییی (Unbiasedness)

امید ریاضی برآوردکننده باید دقیقاً برابر پارامتر واقعی جامعه باشد: $E(\hat{\beta}_1) = \beta_1$.

اثبات:

جایگذاری $Y_i = \beta_0 + \beta_1 X_i + u_i$ در فرمول خطی:

$$\hat{\beta}_1 = \sum k_i (\beta_0 + \beta_1 X_i + u_i) = \beta_0 \sum k_i + \beta_1 \sum k_i X_i + \sum k_i u_i$$

با اعمال خواص وزنی ($\sum k_i = 0$ و $\sum k_i X_i = 1$):

$$\hat{\beta}_1 = \beta_1 + \sum k_i u_i$$

اکنون امید ریاضی شرطی می‌گیریم:

$$E(\hat{\beta}_1 \mid X) = E(\beta_1 \mid X) + \sum k_i E(u_i \mid X)$$

به موجب فرض صفر بودن میانگین شرطی خطا ($E(u_i \mid X) = 0$):

$$E(\hat{\beta}_1) = \beta_1 + 0 = \beta_1 \blacksquare$$

۳. بهترین / کمترین واریانس (Best / Minimum Variance / Efficiency)

ابتدا واریانس برآوردگر OLS را محاسبه می‌کنیم:

$$Var(\hat{\beta}_1) = E\left[(\hat{\beta}_1 - \beta_1)^2\right] = E\left[\left(\sum k_i u_i\right)^2\right] = \sum k_i^2 Var(u_i) = \sigma^2 \sum k_i^2 = \frac{\sigma^2}{\sum x_i^2}$$

حال برآوردکننده خطی و نااریب دلخواه دیگری مانند $\tilde{\beta}_1 = \sum w_i Y_i$ را فرض می‌کنیم. وزن‌ها را به صورت $w_i = k_i + d_i$ تعریف می‌کنیم.

برای آنکه $\tilde{\beta}_1$ نااریب باشد، باید:

$$\sum w_i = 0 \implies \sum d_i = 0 \text{ و } \sum w_i X_i = 1 \implies \sum d_i X_i = 0$$

محاسبه واریانس برآوردکننده جدید:

$$\mathrm{Var}(\tilde{\beta}_1) = \sigma^2 \sum w_i^2 = \sigma^2 \sum (k_i + d_i)^2 = \sigma^2 \left(\sum k_i^2 + \sum d_i^2 + 2 \sum k_i d_i \right)$$

اما:

$$\sum k_i d_i = \sum \left(\frac{x_i}{\sum x_j^2} \right) d_i = \frac{1}{\sum x_j^2} \sum x_i d_i = 0$$

بنابراین:

$$\mathrm{Var}(\tilde{\beta}_1) = \sigma^2 \sum k_i^2 + \sigma^2 \sum d_i^2 = \mathrm{Var}(\hat{\beta}_1) + \sigma^2 \sum d_i^2$$

از آنجا که $\sigma^2 \sum d_i^2 \geq 0$ است، همواره داریم:

$$\mathrm{Var}(\tilde{\beta}_1) \geq \mathrm{Var}(\hat{\beta}_1)$$

تساوی تنها زمانی برقرار می‌شود که $d_i = 0$ باشد، یعنی $\tilde{\beta}_1 = \hat{\beta}_1$. این قضیه اثبات نهایی کمترین واریانس بودن برآوردگر OLS در کلاس برآوردگرهای خطی نااریب است.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۴: «قضیه گاوس–مارکوف و ویژگی‌های ضرایب»، صص ۹۲–۱۱۵.
- نوفرستی، محمد (۱۴۰۰). مقدمه‌ای بر اقتصادسنجی، تهران: رسا، صص ۷۵–۹۰.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ۳: “The Gauss-Markov Theorem”, pp. ۹۲-۹۸.
- Greene, W. H. (۲۰۱۸). Econometric Analysis, Chapter ٤: “Statistical Properties of the Least Squares Estimator”, pp. ۹٥-۱۲۰.
- Hayashi, F. (۲۰۰۰). Econometrics, Princeton University Press, Chapter ۱: “Finite-Sample Properties of OLS”.

موضوع ۱۱: ضریب تعیین (R^2) و معیار نیکویی برازش (Goodness of Fit)

۱۱.۱. فلسفه و شهود اقتصادی نیکویی برازش

هدف از رگرسیون تبیین نوسانات و تغییرپذیری متغیر وابسته (Y) بر پایه تغییرات متغیر(های) توضیحی (X) است. اما پرسش بنیادی این است: «چه سهمی از کل تغییرات متغیر وابسته واقعاً توسط مدل ما توضیح داده شده و چه مقدار از آن به دست تصادف و جزء اخلال سپرده شده است؟» معیار سنجش این قدرت تبیین، **ضریب تعیین (Coefficient of Determination)** است که با $R^۲$نشان داده می‌شود.

۱۱.۲. اثبات ریاضی و تجزیه تغییرپذیری کل (ANOVA Identity)

مقدار مشاهده‌شده متغیر وابسته (Y_i) را می‌توان نسبت به میانگین کل نمونه ($\bar{Y}$) به‌صورت زیر تفکیک کرد:

$$\mathrm{Y_i} - \overline{\mathrm{Y}} = \left(\mathrm{Y_i} - \widehat{\mathrm{Y}}_\mathrm{i}\right) + \left(\widehat{\mathrm{Y}}_\mathrm{i} - \overline{\mathrm{Y}}\right) = \mathrm{e_i} + \left(\widehat{\mathrm{Y}}_\mathrm{i} - \overline{\mathrm{Y}}\right)$$

طرفین این رابطه را به توان ۲ می‌رسانیم و روی کل نمونه جمع می‌بندیم:

$$\sum_{i=۱}^{n} (Y_i - \bar{Y})^۲ = \sum_{i=۱}^{n} \left[\left(\hat{Y}_i - \bar{Y}\right) + e_i\right]^۲$$

$$\sum_{i=۱}^{n} (Y_i - \bar{Y})^۲ = \sum_{i=۱}^{n} (\hat{Y}_i - \bar{Y})^۲ + \sum_{i=۱}^{n} e_i^۲ + ۲\sum_{i=۱}^{n} (\hat{Y}_i - \bar{Y}) e_i$$

بر اساس ویژگی‌های جبری روش حداقل مربعات معمولی (OLS)، پسماندها بر مقادیر برازش‌شده عمود هستند ($\sum e_i \hat{Y}_i = ۰$) و جمع جبری پسماندها صفر است ($\sum e_i = ۰$)، بنابراین جمله حاصل‌ضرب متقاطع دقیقاً صفر می‌شود:

$$\sum_{i=۱}^{n} (\hat{Y}_i - \bar{Y}) e_i = \sum \hat{Y}_i \, e_i - \bar{Y} \sum e_i = ۰ - ۰ = ۰$$

در نتیجه، معادله تجزیه مجموع مربعات (Sum of Squares Decomposition) حاصل می‌شود:

$$TSS = ESS + SSR$$

که در آن:

- **مجموع مربعات کل (Total Sum of Squares - TSS):** کل واریانس نمونه‌ای متغیر وابسته:

$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

- **مجموع مربعات رگرسیون یا توضیح‌داده‌شده (Explained Sum of Squares - ESS):** تغییراتی که توسط خط رگرسیون تبیین شده‌اند:

$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

- **مجموع مربعات پسماندها (Sum of Squared Residuals - SSR):** تغییرات تبیین‌نشده ناشی از خطای تصادفی:

$$SSR = \sum_{i=1}^{n} e_i^2$$

۱۱.۳. تعریف و دامنه مقادیر R^2

با تقسیم طرفین معادله بر TSS:

$$\frac{TSS}{TSS} = \frac{ESS}{TSS} + \frac{SSR}{TSS} \implies 1 = \frac{ESS}{TSS} + \frac{SSR}{TSS}$$

ضریب تعیین به‌صورت نسبت تغییرات توضیح‌داده‌شده به کل تغییرات تعریف می‌شود:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$$

ویژگی‌های کلیدی:

۱. **دامنه تغییرات:** در مدل‌هایی که دارای عرض از مبدأ (β_0) هستند، همواره داریم:

$$0 \leq R^2 \leq 1$$

- اگر $R^2 = ۱$: برازش کامل رخ داده و همه نقاط دقیقاً روی خط رگرسیون قرار دارند ($SSR = ۰$).
- اگر $R^2 = ۰$: متغیر توضیحی هیچ توانایی در پیش‌بینی متغیر وابسته ندارد و $\hat{Y}_i = \overline{Y}$.

۲. **ارتباط با ضریب همبستگی پیرسون:** در رگرسیون دومتغیره (ساده)، ضریب تعیین دقیقاً برابر با مجذور ضریب همبستگی ساده میان Xو Yاست:

$$R^2 = r_{XY}^2$$

۱۱.۴. خطاهای رایج در تفسیر R^2

- **سوءتفسیر در داده‌های مقطعی در برابر سری‌های زمانی:** در داده‌های مقطعی (نظیر بودجه خانوار)، ارقام R^2در حدود ۰/۲ تا ۰/۴ کاملاً طبیعی و قابل قبول هستند، در حالی که در سری‌های زمانی ممکن است به دلیل روندهای کاذب (Spurious Trend) مقدار R^2بالای ۰/۹ ظاهر شود که لزوماً نشان‌دهنده کیفیت مدل نیست.
- **رگرسیون بدون عرض از مبدأ:** در مدل‌های رگرسیون که از مبدأ عبور می‌کنند (بدون β_0)، الزاماً $\sum e_i = ۰$برقرار نیست و رابطه $TSS = ESS + SSR$نقض می‌شود؛ در نتیجه نرم‌افزارها ممکن است برای R^2مقادیر منفی یا بزرگتر از ۱ گزارش کنند.

۱۱.۵. مثال عددی کاربردی

فرض کنید در بررسی مصرف (Y) بر اساس درآمد (X) روی یک نمونه ۱۰ نفره مقادیر زیر حاصل شده است:

$$TSS = ۱۲۰۰ \, , \; SSR = ۱۸۰$$

محاسبه ضریب تعیین:

$$ESS = TSS - SSR = ۱۲۰۰ - ۱۸۰ = ۱۰۲۰$$

$$R^2 = \frac{۱۰۲۰}{۱۲۰۰} = ۰.۸۵$$

تفسیر: ۸۵ درصد از تغییرات مصرف در این نمونه توسط تغییرات درآمد تبیین می‌شود و تنها ۱۵ درصد آن ناشی از عوامل تصادفی و متغیرهای خارج از مدل است.

منابع

- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی (دامودار گجراتی)، انتشارات دانشگاه تهران، جلد اول، فصل ۳، صص ۸۵–۱۰۲.
- سوری، علی (۱۴۰۲). اقتصادسنجی (همراه با کاربرد EViews و Stata)، جلد اول، فصل ۳: «معیار نیکویی برازش»، صص ۹۵–۱۱۵.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics: A Modern Approach, ۷th ed., Chapter ۲: “Mechanics of OLS and Goodness-of-Fit”, pp. ۳۸–٤٤.
- Gujarati, D. N., & Porter, D. C. (۲۰۰۹). Basic Econometrics, ٥th ed., McGraw-Hill, Chapter ۳, pp. ۷۳–۸٦.

موضوع ۱۲: فرض نرمال بودن پسماندها و ویژگی‌های توزیعی برآوردکننده‌ها

۱۲.۱. ضرورت ورود به مدل رگرسیون خطی نرمال کلاسیک (CNLRM)

تا این مرحله، فروض گاوس–مارکوف فقط نشان دادند که برآوردکننده‌های OLS بهترین برآوردکننده خطی نااریب (BLUE) هستند؛ اما هیچ توزیع احتمالی مشخصی برای آن‌ها به دست ندادند. بدون اطلاع از توزیع احتمال $\hat{\beta}$، استخراج **فاصله‌های اطمینان** و اجرای **آزمون‌های فرضیه** (t و F) غیرممکن است. برای حل این مشکل، مدل رگرسیون خطی نرمال کلاسیک (CNLRM) فرض نرمال بودن را به جزء اخلال اضافه می‌کند:

$$u_i \sim \text{i.i.d. } \mathcal{N}(۰, \sigma^۲)$$

۱۲.۲. ویژگی‌های توزیعی برآوردکننده‌های OLS

از آنجا که برآوردکننده شیب رگرسیون یک ترکیب خطی از مشاهدات متغیر وابسته است:

$$\hat{\beta}_۱ = \sum_{i=۱}^{n} k_i Y_i \left(k_i = \frac{x_i}{\sum x_i^۲} \right)$$

و $Y_i = \beta_۰ + \beta_۱ X_i + u_i$خود تابعی خطی از متغیر تصادفی نرمال u_iاست، بنابراین $\hat{\beta}_۱$نیز دقیقاً دارای **توزیع نرمال** خواهد بود:

$$\hat{\beta}_۱ \sim \mathcal{N}\left(\beta_۱, Var(\hat{\beta}_۱)\right) \implies \hat{\beta}_۱ \sim \mathcal{N}\left(\beta_۱, \frac{\sigma^۲}{\sum x_i^۲}\right)$$

به طور مشابه برای عرض از مبدأ:

$$\hat{\beta}_۰ \sim \mathcal{N}\left(\beta_۰, \sigma^۲ \left[\frac{۱}{n} + \frac{\bar{X}^۲}{\sum x_i^۲}\right]\right)$$

۱۲.۳. توزیع واریانس پسماندها و استقلال آماری

تحت فرض نرمال بودن:

۱. برآوردکننده واریانس جامعه یعنی $s^۲ = \frac{\sum e_i^۲}{n-۲} = \frac{SSR}{n-۲}$پس از ضرب و تقسیم در درجات آزادی، دارای توزیع کای‌دو ($\chi^۲$) است:

$$\frac{(n-۲)s^۲}{\sigma^۲} = \frac{\sum e_i^۲}{\sigma^۲} \sim \chi^۲_{(n-۲)}$$

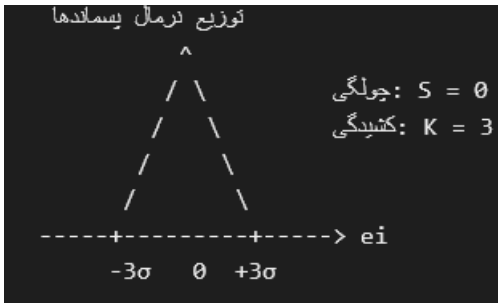
۲. بر اساس **قضیه باسو (Basu's Theorem)**، برآوردکننده‌های شیب و عرض از مبدأ ($\hat{\beta}_۰, \hat{\beta}_۱$) از نظر آماری کاملاً **مستقل** از واریانس پسماند ($s^۲$) هستند. این استقلال سنگ بنای استخراج آزمون tاست.

۱۲.۴. آزمون نرمال بودن پسماندها (آزمون ژاک–برا - Jarque-Bera)

برای آزمون این فرض در کارهای تجربی، رفتار پسماندهای حاصل از مدل (e_i) بررسی می‌شود. آزمون ژاک–برا بر پایه دو گشتاور توزیع یعنی **چولگی (Skewness - S)** و **کشیدگی (Kurtosis - K)** بنا شده است:

$$JB = n\left[\frac{S^۲}{۶} + \frac{(K-۳)^۲}{۲۴}\right] \sim \chi^۲_{(۲)}$$

- برای یک توزیع نرمال استاندارد: $S = ۰$و $K = ۳$.
- فرض صفر: پسماندها نرمال هستند: $H_۰$.
- اگر مقدار احتمال (p–value) گزارش‌شده در EViews برای آماره JBکمتر از ۰۵/۰ باشد، فرض نرمال بودن رد می‌شود.



منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۴: «توزیع احتمال برآوردکننده‌ها و مدل CNLRM»، صص ۱۱۶–۱۳۵.
- نوفرستی، محمد (۱۴۰۰). مقدمه‌ای بر اقتصادسنجی، تهران: رسا، فصل ۳، صص ۶۰–۷۸.
- Greene, W. H. (۲۰۱۸). Econometric Analysis, ۸th ed., Pearson, Chapter ٤: "Statistical Properties of the Classical Regression Model", pp. ۹٥–۱۱٥.
- Hayashi, F. (۲۰۰۰). Econometrics, Princeton University Press, Chapter ۱, pp. ۳۲–٤٥.

موضوع ۱۳: فاصله اطمینان برای پارامترهای مدل (Confidence Intervals)

۱۳.۱. منطق استنتاج بازه‌ای در برابر برآورد نقطه‌ای

برآورد نقطه‌ای یک عدد واحد برای پارامتر ارائه می‌دهد (مثلاً $\hat{\beta}_۱ = ۱/۴$). اما از آنجا که این تخمین بر پایه یک نمونه تصادفی است، اگر نمونه دیگری گرفته شود، عدد متفاوتی حاصل می‌شود. **فاصله اطمینان (Confidence Interval)** محدوده‌ای حول تخمین نقطه‌ای ارائه می‌دهد که با درجه اطمینان مشخص (مثلاً ۹۵٪) پارامتر حقیقی جامعه ($\beta_۱$) را در بر می‌گیرد.

۱۳.۲. استخراج ریاضی متغیر محوری و توزیع t

از آنجا که انحراف معیار واقعی جامعه (σ) نامعلوم است، انحراف معیار تخمینی پسماندها (s) را جایگزین می‌کنیم:

$$SE(\hat{\beta}_۱) = \frac{s}{\sqrt{\sum x_i^۲}} = \frac{\sqrt{\frac{SSR}{n-۲}}}{\sqrt{\sum x_i^۲}}$$

با تقسیم متغیر نرمال بر ریشه متغیر کای‌دو تقسیم بر درجه آزادی‌اش، کمیت محوری زیر حاصل می‌شود که دارای **توزیع tاستیودنت با $n-۲$درجه آزادی** است:

$$t = \frac{\hat{\beta}_۱ - \beta_۱}{SE(\hat{\beta}_۱)} \sim t_{(n-۲)}$$

برای سطح اطمینان $۱-\alpha$ (مثلاً با $\alpha = ۰/۰۵$برای اطمینان ۹۵٪):

$$P\left(-t_{\alpha/۲,\, n-۲} \leq \frac{\hat{\beta}_۱ - \beta_۱}{SE(\hat{\beta}_۱)} \leq t_{\alpha/۲,\, n-۲}\right) = ۱ - \alpha$$

با بازآرایی رابطه، فرمول فاصله اطمینان استخراج می‌شود:

$$\hat{\beta}_۱ - t_{\alpha/۲,\, n-۲} \cdot SE(\hat{\beta}_۱) \leq \beta_۱ \leq \hat{\beta}_۱ + t_{\alpha/۲,\, n-۲} \cdot SE(\hat{\beta}_۱)$$

یا به صورت فشرده:

$$\beta_۱ \in \left[\hat{\beta}_۱ \pm t_{\text{critical}} \cdot SE(\hat{\beta}_۱)\right]$$

۱۳.۳. مثال عددی گام‌به‌گام

فرض کنید در یک رگرسیون با ۲۰ مشاهده ($n = ۲۰$):

$$\hat{\beta}_۱ = ۰.۷۲\,,\; SE(\hat{\beta}_۱) = ۰.۱۵$$

برای محاسبه فاصله اطمینان ۹۵٪:

- درجات آزادی: $df = n - ۲ = ۲۰ - ۲ = ۱۸$
- مقدار بحرانی توزیع t در سطح ۵٪ دوطرفه از جدول: $t_{۰.۰۲۵,\,18} = ۲.۱۰۱$
- حاشیه خطا (Margin of Error): $ME = ۲.۱۰۱ \times ۰.۱۵ = ۰.۳۱۵$
- کران پایین: $۰.۷۲ - ۰.۳۱۵ = ۰.۴۰۵$
- کران بالا: $۰.۷۲ + ۰.۳۱۵ = ۱.۰۳۵$

تفسیر دقیق آماری: با ۹۵٪ اطمینان، پارامتر واقعی شیب جامعه ($\beta_۱$) در بازه $[۰,/۴۰۵۱/۰۳۵]$ قرار دارد. از آنجا که صفر درون این بازه قرار نگرفته است، شیب از لحاظ آماری در سطح ۵ درصد تفاوت معناداری با صفر دارد.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۴: «فاصله اطمینان برای ضرایب»، صص ۱۳۰–۱۴۵.
- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی، فصل ۵: «آزمون فرضیه و فاصله اطمینان».
- Stock, J. H., & Watson, M. W. (۲۰۱۹). Introduction to Econometrics, ٤th ed., Pearson, Chapter ٥: “Hypothesis Tests and Confidence Intervals in Single-Regressor Models”, pp. ١٤٥–١٦٥.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ٤, pp. ١٢٠–١٣٥.

موضوع ۱۴: آزمون فرضیه‌ها برای ضرایب (آزمون tاستیودنت)

۱۴.۱. ساختار آزمون فرضیه کلاسیک

آزمون فرضیه چارچوبی است برای تصمیم‌گیری در خصوص اینکه آیا شواهد تجربی ارائه‌شده توسط نمونه، یک فرضیه اقتصادی معین را تأیید می‌کنند یا رد.

تعریف فرض صفر و فرض مقابل:

در اغلب مطالعات، فرضیه عدم وجود رابطه (بی‌اثر بودن متغیر توضیحی) به عنوان فرض صفر در نظر گرفته می‌شود:

$$\begin{cases} H_0: \beta_1 = 0 & \text{(متغیر توضیحی هیچ اثری بر متغیر وابسته ندارد)} \\ H_1: \beta_1 \neq 0 & \text{(متغیر توضیحی اثر معنادار دارد - آزمون دوطرفه)} \end{cases}$$

۱۴.۲. آماره آزمون t (**t-statistic Calculated**)

تحت فرض صفر ($H_0: \beta_1 = 0$):

$$t_{calc} = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$$

۱۴.۳. قواعد تصمیم‌گیری

دو رویکرد معادل برای قضاوت وجود دارد:

رویکرد اول: مقایسه با مقدار بحرانی (Critical Value Approach)

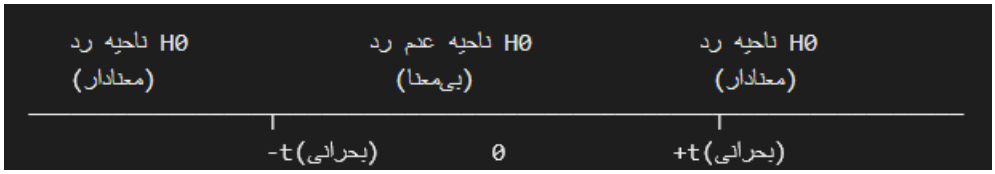
اگر سطح خطا $\alpha = 0/05$باشد:

- اگر $|t_{calc}| > t_{\alpha/2, n-2}$باشد: فرض صفر H_0**رد می‌شود** (ضریب از نظر آماری معنادار است).
- **قانون سرانگشتی ۲ (Rule of Thumb):** برای نمونه‌های با درجات آزادی بیش از ۲۰، مقدار بحرانی tدر سطح ۵٪ تقریباً برابر با ۲ است. بنابراین، اگر قدر مطلق آماره tبزرگتر از ۲ باشد، ضریب در سطح ۹۵٪ معنادار تلقی می‌شود.

رویکرد دوم: مقدار احتمال (p-value Approach)

مقدار p-value کمترین سطح خطایی است که در آن می‌توان فرض صفر را رد کرد.

- اگر $p\text{-value} < \alpha$(مثلاً کمتر از ۰۵/۰): H_0رد می‌شود.
- در نرم‌افزارهای مدرن (EViews, Stata, R)، تصمیم‌گیری مستقیم بر اساس ستون Prob.یا p-value انجام می‌گیرد.



۱۴.۴. آزمون‌های یکطرفه (One-Tailed Tests)

اگر نظریه اقتصادی جهت خاصی را دیکته کند (مثلاً درآمد اثر **مثبت** بر مصرف دارد یا قیمت اثر **منفی** بر تقاضا دارد):

$$H_0: \beta_1 \leq 0 \text{ در برابر } H_1: \beta_1 > 0$$

در این حالت کل سطح خطای αدر یک دنباله توزیع قرار می‌گیرد و مقدار بحرانی جدول در ستون t_αخوانده می‌شود.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۴: «آزمون معناداری ضرایب رگرسیون»، صص ۱۴۵–۱۶۵.
- نوفرستی، محمد (۱۴۰۰). مقدمه‌ای بر اقتصادسنجی، فصل ۳.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ٤: “Multiple Regression Analysis: Inference”, pp. ۱۱۸–۱۳۸.
- Gujarati, D. N., & Porter, D. C. (۲۰۰۹). Basic Econometrics, Chapter ٥: “Two-Variable Regression: Interval Estimation and Hypothesis Testing”.

موضوع ۱۵: جدول تحلیل واریانس (ANOVA) و آزمون کلیت رگرسیون (F)

۱۵.۱. مفهوم و ساختار تحلیل واریانس (Analysis of Variance)

جدول ANOVA ابزاری ساختاریافته برای تفکیک اجزای واریانس متغیر وابسته به بخش توضیح‌داده‌شده توسط مدل و بخش خطای تبیین‌نشده است. این جدول مبنای محاسبه آماره آزمون کلیت رگرسیون (F) به شمار می‌رود.

۱۵.۲. اجزای جدول استاندارد ANOVA

در یک مدل رگرسیون با n مشاهده و یک متغیر توضیحی:

منبع تغییرات (Source)	مجموع مربعات (SS)	درجات آزادی (df)	میانگین مربعات ($MS = SS/df$)	آماره F
رگرسیون (مدل)	$ESS = \sum(\hat{Y}_i - \bar{Y})^2$	1	$MS_{reg} = \frac{ESS}{1}$	$F = \frac{MS_{reg}}{MS_{res}} = \frac{ESS/1}{SSR/(n-2)}$
پسماند (خطا)	$SSR = \sum e_i^2$	$n-2$	$MS_{res} = \frac{SSR}{n-2} = s^2$	—
کل (Total)	$TSS = \sum(Y_i - \bar{Y})^2$	$n-1$	$MS_{tot} = \frac{TSS}{n-1} = s_Y^2$	—

۱۵.۳. فرضیه آزمون کلیت رگرسیون و ارتباط با R^2

آزمون F اعتبار کل مدل را می‌سنجد:

$$\begin{cases} H_0: \beta_1 = 0 & \text{(کل مدل فاقد اعتبار است و متغیر توضیحی نقشی ندارد)} \\ H_1: \beta_1 \neq 0 & \text{(مدل در مجموع معنادار است)} \end{cases}$$

با تقسیم صورت و مخرج آماره F بر TSS، رابطه مستقیم آن با ضریب تعیین استخراج می‌شود:

$$F = \frac{ESS/1}{SSR/(n-2)} = \frac{\frac{ESS}{TSS}}{\frac{SSR}{TSS} \cdot \frac{1}{n-2}} = \frac{R^2}{\frac{1-R^2}{n-2}} \implies F = \frac{R^2}{1-R^2} \cdot (n-2)$$

این رابطه نشان می‌دهد که آزمون کلیت رگرسیون (F) آزمونی هم‌ارز برای معناداری آماری ضریب تعیین (R^2) است: اگر $R^2 = 0$ باشد، $F = 0$ خواهد بود.

۱۵.۴. رابطه یکتای آماره t و آماره F در رگرسیون دومتغیره

در مدل رگرسیون ساده دومتغیره، آزمون معناداری شیب (t) دقیقاً معادل آزمون معناداری کل مدل (F) است و اتحاد ریاضی زیر همواره برقرار است:

$$F = t^2$$

این تساوی نشان می‌دهد در مدل ساده، آزمون F هیچ اطلاعات اضافه‌ای نسبت به آزمون t به دست نمی‌دهد؛ اما اهمیت واقعی F در مدل‌های چندمتغیره نمایان می‌شود که در آن‌ها معناداری هم‌زمان چند متغیر سنجیده می‌شود.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۳ و ۴: «جدول تحلیل واریانس و آزمون F»، صص ۱۶۵–۱۸۰.
- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی، فصل ۵.
- Greene, W. H. (۲۰۱۸). Econometric Analysis, Chapter ۵: “Hypothesis Tests and Model Selection”, pp. ۱٤۰–۱٦۰.
- Stock, J. H., & Watson, M. W. (۲۰۱۹). Introduction to Econometrics, Chapter ۵ & ۷.

موضوع ۱۶: خطاهای رایج در تفسیر ضرایب رگرسیونی

۱۶.۱. ریشه‌یابی خطاهای رایج تحلیلی در تفسیر خروجی‌ها

در پژوهش‌های تجربی، بارها دیده می‌شود که ضرایب رگرسیونی به اشتباه تفسیر می‌گردند. درک این خطاها برای هر پژوهشگر و تحلیل‌گر ضروری است:

۱. خلط همبستگی ساده با علیت اقتصادی (Correlation is not Causation)

وجود ضریب رگرسیونی معنادار ($\hat{\beta}_1 > 0$) صرفاً نشان‌دهنده یک الگوی هم‌حرکتی در داده‌های نمونه است. علیت نیازمند تأیید تئوری، توالی زمانی و کنترل اثر سایر عوامل است. برای نمونه، همبستگی مثبت میان تعداد بستنی فروخته‌شده و نرخ غرق‌شدگی در دریا به معنی علت بودن بستنی نیست؛ بلکه عامل سوم یعنی «دمای هوا در فصل تابستان» متغیر مخدوش‌کننده است.

۲. اریبی ناشی از متغیرهای حذف‌شده (Omitted Variable Bias - OVB)

اگر متغیری با متغیر وابسته مرتبط باشد و با متغیر توضیحی موجود در مدل نیز همبستگی داشته باشد ولی از مدل حذف گردد، ضریب تخمین‌زده‌شده برای متغیر حاضر هم اثر خود و هم اثر متغیر پنهان را جذب کرده و دچار اریبی شدید می‌شود:

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \cdot \frac{Cov(X_1, X_2)}{Var(X_1)}$$

تنها در صورتی ضریب دچار اریبی نمی‌شود که متغیر حذف‌شده اصلاً با متغیر حاضر همبستگی نداشته باشد ($Cov(X_1, X_2) = 0$).

۳. نادیده گرفتن مقیاس سنجش و واحد متغیرها (Scale Fallacy)

اگر درآمد را از «میلیون تومان» به «ریال» تبدیل کنیم، ضریب رگرسیونی متناظر یک میلیون برابر کوچک‌تر می‌شود، در حالی که معناداری آماری (t-statistic) و ضریب تعیین (R^2) دست‌نخورده باقی می‌ماند. بنابراین اندازه عددی ضریب به خودی خود ملاک اهمیت اقتصادی متغیر نیست.

۴. خطای برون‌یابی (Extrapolation Fallacy)

روابط تخمین‌زده‌شده در OLS فقط در محدوده مقادیر متغیرهای درون نمونه معتبر است. به عنوان مثال، اگر مدلی برای رابطه سن بین ۲۰ تا ۵۰ سال و بهره‌وری نیروی کار برآورد شده باشد، نمی‌توان با آن بهره‌وری یک فرد ۸۰ ساله را پیش‌بینی کرد؛ زیرا ممکن است در سنین بالا ماهیت رابطه از خطی به مقعر تغییر یابد.

چک‌لیست تحلیل پنج‌گانه این خروجی:

۱. **ضرایب (Coefficient):** ضریب متغیر درآمد ۰/۷۲است؛ یعنی به ازای هر واحد افزایش درآمد، مصرف به طور متوسط ۰/۷۲واحد افزایش می‌یابد.

۲. **آماره‌های tو مقدار Prob:** برای متغیر درآمد، آماره $t = ۴/۸۰$و مقدار $Prob = ۰/۰۰۰۱$است که چون کمتر از ۵ درصد است، رابطه کاملاً معنادار است.

ضریب تعیین ($R^۲$): برابر با ۰/۵۶۱۱است؛ یعنی ۱/۵۶ درصد تغییرات مصرف با درآمد توضیح داده می‌شود.

۳. **آماره Fو سطح احتمال آن:** مقدار $F = ۲۳/۰۴$با $Prob = ۰/۰۰۰۱$نشان‌دهنده معناداری کلی مدل رگرسیون در سطح بالای ۹۹ درصد است.

٤. **ارتباط Fو t:** دقت شود که $۲۳/۰۴ \approx (۴/۸)^۲$که تطابق کامل با رابطه تئوریک $F = t^۲$دارد.

منابع

- سوری، علی (۱۴۰۲). *اقتصادسنجی (جلد اول)*، فصل ۵: «تفسیر نتایج رگرسیون و خطاهای رایج»، صص ۱۸۰–۲۰۰.
- ابریشمی، حمید (۱۴۰۱). *مبانی اقتصادسنجی (گجراتی)*، فصل ۵ و ۶.
- Wooldridge, J. M. (۲۰۲۰). *Introductory Econometrics*, Chapter ۳: “Multiple Regression Analysis: Estimation - Omitted Variable Bias”, pp. ۷۸–۹۲.
- Stock, J. H., & Watson, M. W. (۲۰۱۹). *Introduction to Econometrics*, Chapter ٦: “Omitted Variable Bias”, pp. ۱۸۵–۲۰۵.

موضوع ۱۷: تصریح مدل چندمتغیره و نمادگذاری ماتریسی (Matrix Formulation of the Multiple Regression Model)

۱۷.۱. ضرورت و مزایای فرمول‌بندی ماتریسی

در مدل رگرسیون ساده دومتغیره، روابط جبری از طریق بسط جمع‌ها ($\sum$) به سادگی قابل پیگیری بود. اما هنگامی که تعداد متغیرهای توضیحی افزایش می‌یابد (k متغیر)، استفاده از علامت‌های مجموع جبری به روابط بسیار طویل، مستعد خطا و از نظر ریاضی ناکارآمد تبدیل می‌شود. **جبر ماتریسی** چارچوب فشرده، ظریف و ساختاریافته‌ای فراهم می‌سازد که تحلیل مدل با هر تعداد متغیر توضیحی و هر حجم نمونه‌ای را در قالب چند معادله ماتریسی یکدست خلاصه می‌کند.

۱۷.۲. تصریح دستگاه معادلات رگرسیون

فرض کنید رابطه بین یک متغیر وابسته (Y) و $k-۱$ متغیر توضیحی را برای یک نمونه nتایی بررسی می‌کنیم:

$$Y_i = \beta_۱ + \beta_۲X_{i2} + \beta_۳X_{i3} + \cdots + \beta_kX_{ik} + u_i \ , \ i = ۱,۲, \dots, n$$

این دستگاه شامل nمعادله مجزا است:

$$\begin{cases} Y_۱ = \beta_۱ + \beta_۲X_{۱۲} + \beta_۳X_{۱۳} + \cdots + \beta_kX_{۱k} + u_۱ \\ Y_۲ = \beta_۱ + \beta_۲X_{۲۲} + \beta_۳X_{۲۳} + \cdots + \beta_kX_{۲k} + u_۲ \\ \vdots \\ Y_n = \beta_۱ + \beta_۲X_{n2} + \beta_۳X_{n3} + \cdots + \beta_kX_{nk} + u_n \end{cases}$$

۱۷.۳. ساختار بردارها و ماتریس‌های بنیادین

این دستگاه به شکل ماتریسی فشرده زیر نوشته می‌شود:

$$y = X\beta + u$$

تعریف اجزای معادله:

۱. **بردار متغیر وابسته (y):** برداری ستونی با ابعاد $n \times ۱$:

$$y = \begin{bmatrix} Y_۱ \\ Y_۲ \\ \vdots \\ Y_n \end{bmatrix}$$

۲. **ماتریس داده‌های متغیرهای توضیحی یا ماتریس طرح (X):** ماتریسی با ابعاد $n \times k$. ستون نخست آن تماماً عدد ۱است تا ضریب عرض از مبدأ ($\beta_۱$) را در خود جای دهد:

$$X = \begin{bmatrix} ۱ & X_{۱۲} & X_{۱۳} & \dots & X_{۱k} \\ ۱ & X_{۲۲} & X_{۲۳} & \dots & X_{۲k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ۱ & X_{n2} & X_{n3} & \dots & X_{nk} \end{bmatrix}_{n \times k}$$

۳. **بردار پارامترهای مجهول جامعه (β):** برداری ستونی با ابعاد $k \times ۱$:

$$\beta = \begin{bmatrix} \beta_۱ \\ \beta_۲ \\ \vdots \\ \beta_k \end{bmatrix}_{k \times ۱}$$

٤. **بردار خطای تصادفی (u):** برداری ستونی با ابعاد $n \times ۱$:

$$u = \begin{bmatrix} u_۱ \\ u_۲ \\ \vdots \\ u_n \end{bmatrix}_{n \times ۱}$$

۱۷.۴. بازنویسی فروض گاوس–مارکوف در قالب جبر ماتریسی

در قالب ماتریسی، فروض کلاسیک به زیبایی و با نهایت ایجاز تعریف می‌شوند:

۱. **میانگین شرطی پسماندها صفر است:**

$$E(u \mid X) = ۰ \text{ یا } \begin{bmatrix} E(u_۱) \\ E(u_۲) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} ۰ \\ ۰ \\ \vdots \\ ۰ \end{bmatrix}$$

۲. **ماتریس واریانس–کوواریانس پسماندها (همسانی واریانس و عدم خودهمبستگی):**

$$Var(u \mid X) = E(uu' \mid X) = \begin{bmatrix} Var(u_1) & Cov(u_1,u_2) & \dots & Cov(u_1,u_n) \\ Cov(u_2,u_1) & Var(u_2) & \dots & Cov(u_2,u_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(u_n,u_1) & Cov(u_n,u_2) & \dots & Var(u_n) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 I_n$$

که در آن I_nماتریس همانی $n \times n$است.

۳. **رتبه کامل ستونی ماتریس X(عدم همخطی کامل):**

$$\text{rank}(X) = k < n$$

این فرض تضمین می‌کند که هیچ ستونی ترکیب خطی دقیقی از سایر ستون‌ها نیست؛ در نتیجه ماتریس مربعی $X'X$با مرتبه $k \times k$**وارون‌پذیر (Nonsingular)** خواهد بود و معکوس آن یعنی $(X'X)^{-1}$وجود دارد.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (همراه با کاربرد EViews و Stata)، جلد اول، فصل ۶: «تحلیل ماتریسی رگرسیون چندمتغیره»، صص ۲۱۰–۲۲۸.
- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی (دامودار گجراتی)، دانشگاه تهران، جلد اول، پیوست الف و ب: «عملیات ماتریسی در اقتصادسنجی».
- Greene, W. H. (۲۰۱۸). Econometric Analysis, ۸th ed., Pearson, Chapter ۲ & ۳: “The Classical Multiple Linear Regression Model”, pp. ۲۰–۴۸.
- Johnston, J., & DiNardo, J. (۱۹۹۷). Econometric Methods, ٤th ed., McGraw-Hill, Chapter ۳: “The Vector and Matrix Basis of Econometrics”, pp. ٦۰–۸۵.

موضوع ۱۸: استخراج جبری برآوردکننده OLS ماتریسی و ماتریس واریانس–کوواریانس

۱۸.۱. استخراج معادله نرمال و برآوردکننده $\widehat{\beta}$

معادله نمونه‌ای رگرسیون بر حسب پسماندها (e) عبارت است از:

$$y = X\hat{\beta} + e \implies e = y - X\hat{\beta}$$

معیار حداقل مربعات معمولی (OLS)، کمینه‌سازی مجموع مربعات پسماندها (SSR) است:

$$SSR = \sum_{i=1}^{n} e_i^2 = e'e = \left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right)$$

با بسط ترانهاده:

$$SSR = \left(y' - \hat{\beta}'X'\right)\left(y - X\hat{\beta}\right) = y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

از آنجا که عبارت $y'X\hat{\beta}$یک کمیت اسکالر (1×1) است، ترانهاده آن با خودش برابر است:

$$\left(y'X\hat{\beta}\right)' = \hat{\beta}'X'y$$

بنابراین:

$$SSR = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

برای حداقل کردن SSR، نسبت به بردار $\hat{\beta}$مشتق ماتریسی می‌گیریم و برابر صفر قرار می‌دهیم:

$$\frac{\partial(SSR)}{\partial\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0$$

با ساده‌سازی، **معادلات نرمال OLS** حاصل می‌شود:

$$X'X\hat{\beta} = X'y$$

با توجه به فرض رتبه کامل، وارون ماتریس $X'X$وجود دارد. با ضرب طرفین از چپ در $(X'X)^{-1}$:

$$\hat{\beta} = (X'X)^{-1}X'y$$

این فرمول بنیادی‌ترین فرمول در علم اقتصادسنجی پارامتریک است.

۱۸.۲. اثبات ناااریبی برآوردکننده OLS ماتریسی

با جایگذاری مقدار واقعی جامعه ($y = X\beta + u$) در فرمول برآوردکننده:

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + u)$$

$$\hat{\beta} = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u$$

چون $(X'X)^{-1}X'X = I$، نتیجه می‌شود:

$$\hat{\beta} = \beta + (X'X)^{-1}X'u$$

امید ریاضی شرطی بردار ضرایب را به دست می‌آوریم:

$$E(\hat{\beta} \mid X) = E(\beta \mid X) + E[(X'X)^{-1}X'u \mid X]$$

چون βبردار ثابتی است و بر اساس فرض $E(u \mid X) = 0$:

$$E(\hat{\beta} \mid X) = \beta + (X'X)^{-1}X'E(u \mid X) = \beta + 0 = \beta$$

این اثبات نشان می‌دهد که $\hat{\beta}$یک **برآوردکننده ناااریب (Unbiased)** است.

۱۸.۳. استخراج ماتریس واریانس–کوواریانس $Var(\hat{\beta})$

بر اساس تعریف ماتریس پراکندگی:

$$Var(\hat{\beta} \mid X) = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \mid X\right]$$

با جایگذاری $\hat{\beta} - \beta = (X'X)^{-1}X'u$:

$$Var(\hat{\beta} \mid X) = E[((X'X)^{-1}X'u)((X'X)^{-1}X'u)' \mid X]$$

$$Var(\hat{\beta} \mid X) = E[(X'X)^{-1}X'uu'X(X'X)^{-1} \mid X]$$

$$Var(\hat{\beta} \mid X) = (X'X)^{-1}X'E(uu' \mid X)X(X'X)^{-1}$$

با توجه به فرض همسانی واریانس و عدم خودهمبستگی ($E(uu' \mid X) = \sigma^2 I$):

$$Var(\hat{\beta} \mid X) = \sigma^2 (X'X)^{-1} X'X (X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

تشریح ساختار این ماتریس:

ماتریس $Var(\hat{\beta})$دارای ابعاد $k \times k$است:

$$Var(\hat{\beta}) = \begin{bmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & \dots & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & Var(\hat{\beta}_2) & \dots & Cov(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_k, \hat{\beta}_1) & Cov(\hat{\beta}_k, \hat{\beta}_2) & \dots & Var(\hat{\beta}_k) \end{bmatrix}$$

- **عناصر روی قطر اصلی:** واریانس نمونه‌ای هر یک از برآوردکننده‌ها را به دست می‌دهند که جذر آن‌ها خطای استاندارد ($SE(\hat{\beta}_j)$) است.
- **عناصر خارج از قطر:** کوواریانس بین برآوردهای ضرایب مختلف را نشان می‌دهند.

منابع

- سوری، علی (۱۴۰۲). *اقتصادسنجی (جلد اول)*، فصل ۶: «استخراج ماتریسی برآوردکننده OLS»، صص ۲۲۵–۲۴۵.
- نوفرستی، محمد (۱۴۰۰). *مقدمه‌ای بر اقتصادسنجی*، فصل ۴: «رگرسیون چندمتغیره در جبر ماتریسی»، صص ۸۵–۱۱۰.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics: A Modern Approach, Appendix E: “The Linear Regression Model in Matrix Form”, pp. ۷۸۵–۸۰۲.
- Greene, W. H. (۲۰۱۸). Econometric Analysis, Chapter ٤, pp. ٦٦–٧٩.

موضوع ۱۹: ماتریس تصویر (Hat Matrix)، ماتریس پسماندساز و قضیه فریش–وو–لوول (FWL)

۱۹.۱. هندسه ماتریس تصویر (The Hat Matrix)

مقادیر برازش‌شده توسط مدل ($\hat{y}$) را می‌توان بر حسب ماتریس داده‌ها بازنویسی کرد:

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y$$

اگر ماتریس H(ماتریس تصویر یا Hat Matrix) را به صورت زیر تعریف کنیم:

$$H = X(X'X)^{-1}X'$$

آن‌گاه:

$$\hat{y} = Hy$$

این ماتریس از آن رو Hat خوانده می‌شود که بر سر متغیر وابسته کلاهک برازش ($\hat{}$) می‌گذارد.

۱۹.۲. ماتریس پسماندساز (M)

به طور مشابه، بردار پسماندها (e) را می‌توان مستقیماً از yاستخراج نمود:

$$e = y - \hat{y} = y - Hy = (I_n - H)y$$

ماتریس پسماندساز به صورت $M = I_n - H$تعریف می‌شود:

$$e = My$$

خواص ریاضی شگفت‌انگیز ماتریس‌های Hو M:

۱. **تقارن (Symmetric):**

$$H' = H\ ,\ M' = M$$

۲. **خودتوانی یا یکسان‌توانی (Idempotent):** ضرب ماتریس در خودش برابر با خودش است:

$$HH = H\ ,\ MM = M$$

۳. **تعامد متقابل:** حاصل‌ضرب آن‌ها ماتریس صفر است ($HM = ۰$) و علاوه بر این $MX = ۰$.

٤. **رد (Trace) و درجات آزادی:**

$$\text{tr}(H) = \text{tr}(X(X'X)^{-۱}X') = \text{tr}((X'X)^{-۱}X'X) = \text{tr}(I_k) = k$$
$$\text{tr}(M) = \text{tr}(I_n) - \text{tr}(H) = n - k$$

این ویژگی اثبات می‌کند که درجات آزادی پسماندها در رگرسیون چندمتغیره دقیقاً $n - k$است.

۱۹.۳. قضیه فریش–وو–لوول (Frisch-Waugh-Lovell - FWL)

یکی از زیباترین قضایای اقتصادسنجی کاربردی که شهود مفهوم «سایر شرایط ثابت» (*ceteris paribus*) را عیان می‌سازد، قضیه FWL است.

صورت مسأله:

فرض کنید ماتریس متغیرهای توضیحی را به دو بلوک $X_۱$و $X_۲$تفکیک کنیم:

$$y = X_۱\beta_۱ + X_۲\beta_۲ + u$$

می‌خواهیم ضریب $\beta_۲$را بدون تخمین کل مدل یا با خالص‌سازی اثرات محاسبه کنیم.

بیان قضیه FWL:

ضریب برآوردشده $\hat{\beta}_۲$در رگرسیون فوق دقیقاً برابر است با ضریب حاصل از یک رگرسیون دومتغیره میان پسماندهای خالص‌شده دو معادله کمکی زیر:

۱. **گام اول:** رگرسیون yبر روی $X_۱$برآورد شده و پسماندهای آن ذخیره می‌شود:

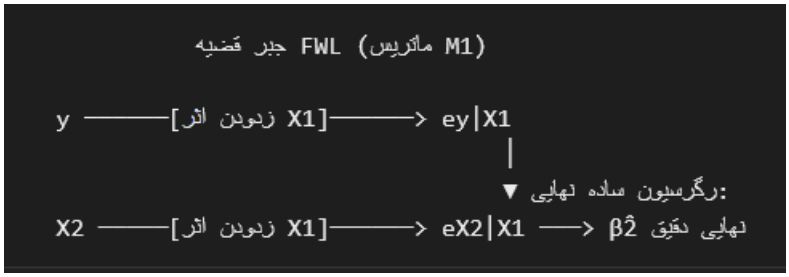
$$e_{y|X_۱} = M_۱ y$$

۲. **گام دوم:** تک‌تک متغیرهای موجود در $X_۲$بر روی $X_۱$رگرسیون شده و پسماندهای خالص آن‌ها استخراج می‌شود:

$$e_{X_۲|X_۱} = M_۱ X_۲$$

۳. **گام سوم:** رگرسیون پسماند به پسماند اجرا می‌شود:

$$\hat{\beta}_2 = \left(e'_{X_2|X_1} e_{X_2|X_1}\right)^{-1} e'_{X_2|X_1} e_{y|X_1} = (X_2' M_1 X_2)^{-1} X_2' M_1 y$$



اهمیت مفهومی: قضیه FWL ثابت می کند ضریب هر متغیر در رگرسیون چندمتغیره، اثر خالص آن متغیر پس از کسر و تخلیه کامل اثرات خطی متغیرهای دیگر مدل بر آن متغیر و بر متغیر وابسته است.

منابع

- سوری، علی (۱۴۰۲). *اقتصادسنجی (جلد اول)*، فصل ۶: «ماتریس‌های متعامد و قضیه FWL»، صص ۲۵۰–۲۶۵.
- Davidson, R., & MacKinnon, J. G. (۲۰۰٤). Econometric Theory and Methods, Oxford University Press, Chapter ۲: “The Geometry of Linear Regression and The FWL Theorem”, pp. ٦۲–۷٥.
- Hayashi, F. (۲۰۰۰). Econometrics, Princeton University Press, Chapter ۱, pp. ۱۸–۲۷.

موضوع ۲۰: ضریب تعیین تعدیل‌شده ($\bar{R}^2$) و معیارهای انتخاب مدل (AIC, SC)

۲۰.۱. نقیصه ریاضی ضریب تعیین معمولی (R^2) در رگرسیون چندمتغیره

یک ویژگی نامطلوب در ضریب تعیین معمولی ($R^۲ = ۱ - \frac{SSR}{TSS}$) این است که تابعی غیرکاهشی از تعداد متغیرهای توضیحی است. یعنی اگر متغیر کاملاً بی‌ارتباط یا نامربوطی (مانند پیش‌بینی قیمت مسکن با تعداد تصادفات رانندگی در قطب جنوب!) به مدل افزوده شود، مقدار SSR هیچ‌گاه افزایش نیافته و $R^۲$ همواره افزایش می‌یابد (یا دست‌کم ثابت می‌ماند). این موضوع پژوهشگران را با خطر بیش‌برازش (Overfitting) مواجه می‌سازد.

۲۰.۲. استخراج ضریب تعیین تعدیل‌شده ($\bar{R}^2$)

هنری تیل (Henri Theil) معیاری پیشنهاد داد که به جای استفاده از مجموع مربعات خام، از میانگین مربعات تعدیل‌شده با درجات آزادی بهره می‌برد:

$$\bar{R}^۲ = ۱ - \frac{SSR/(n-k)}{TSS/(n-۱)} = ۱ - \frac{s^۲}{s_Y^۲}$$

رابطه ریاضی میان $\bar{R}^2$ و R^2:

با قرار دادن $SSR = TSS(۱ - R^۲)$:

$$\bar{R}^۲ = ۱ - \left(\frac{n-۱}{n-k}\right)(۱ - R^۲)$$

ویژگی‌های حیاتی $\bar{R}^2$:

۱. **جریمه برای متغیرهای اضافی:** کسر $\frac{n-۱}{n-k}$ با اضافه شدن متغیر توضیحی جدید بزرگتر از ۱ می‌شود. متغیر جدید تنها در صورتی $\bar{R}^۲$ را بالا می‌برد که آماره t متناظر با آن از نظر قدر مطلق بزرگتر از ۱ باشد ($|t| > ۱$).

۲. **امکان منفی شدن:** در مدل‌های بسیار ضعیف با متغیرهای نامناسب، ممکن است $\bar{R}^۲$ منفی شود.

۳. **همواره کمتر بودن از ضریب تعیین خام:**

$$\bar{R}^۲ \leq R^۲$$

۲۰.۳. معیارهای اطلاعاتی مدل‌سازی (Schwarz Criteria & Akaike)

برای گزینش بهترین تصریح و پیشگیری از خطای تصریح، از توابع درست‌نمایی جریمه‌دار استفاده می‌شود:

- **معیار آکائیکه (Akaike Information Criterion - AIC):**

$$AIC = e^{۲k/n}\frac{\sum e_i^۲}{n} \implies \ln(AIC) = \left(\frac{۲k}{n}\right) + \ln\left(\frac{SSR}{n}\right)$$

- **معیار شوارتز–بیزین (Schwarz Information Criterion - SIC / BIC):**

$$SIC = n^{k/n}\frac{\sum e_i^۲}{n} \implies \ln(SIC) = \left(\frac{k\ln n}{n}\right) + \ln\left(\frac{SSR}{n}\right)$$

قاعده تصمیم‌گیری: در مقایسه میان دو یا چند مدل رقیب، مدلی که مقادیر AIC و SIC کمتری داشته باشد، مدل برتر و بهینه‌تر تلقی می‌شود. جریمه SIC برای متغیرهای اضافه بسیار سنگین‌تر از AIC است.

منابع

- ابریشمی، حمید (۱۴۰۱). *مبانی اقتصادسنجی*، فصل ۷: «مدل‌های رگرسیون چندمتغیره»، صص ۲۳۰–۲۴۵.
- سوری، علی (۱۴۰۲). *اقتصادسنجی (جلد اول)*، صص ۲۷۰–۲۸۲.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ٦: “Multiple Regression Analysis: Further Issues”, pp. ۲۰۰–۲۱٥.
- Stock, J. H., & Watson, M. W. (۲۰۱۹). Introduction to Econometrics, Chapter ٦ & ۷.

موضوع ۲۱: آزمون‌های قیود خطی و آزمون کلی والد (test-F Wald)

۲۱.۱. فلسفه آزمون‌های قیود فرضیه‌ای

در بسیاری از نظریه‌های اقتصادی، ادعاهایی مطرح می‌شود که متضمن روابط هم‌زمان میان چندین پارامتر است. برای مثال:

- **بازده ثابت نسبت به مقیاس در تابع تولید کاب–داگلاس:** $\beta_۲ + \beta_۳ = ۱$
- **عدم توهم پولی در تقاضای پول:** مجموع کشش‌ها برابر صفر باشد.
- **حذف هم‌زمان گروهی از متغیرها:** $\beta_۳ = \beta_۴ = ۰$

آزمون tبه تنهایی قادر به آزمون روابط چندمتغیره یا ترکیبات خطی میان ضرایب نیست؛ زیرا کوواریانس میان برآوردها ($Cov(\hat{\beta}_j, \hat{\beta}_m)$) را در نظر نمی‌گیرد. ابزار جامع برای حل این مسأله، **آزمون قیود خطی F(آزمون والد)** است.

۲۱.۲. فرمول‌بندی ماتریسی قیود فرضیه

هر مجموعه از qقید خطی را می‌توان در قالب ماتریسی زیر فرموله کرد:

$$H_۰: R\beta = r \quad \text{در برابر} \quad H_۱: R\beta \neq r$$

که در آن:

- R: ماتریس ضرایب قیود با ابعاد $q \times k$و با رتبه کامل سطری (q).
- r: بردار ستونی مقادیر ثابت قیود با ابعاد $q \times ۱$.

مثال ملموس:

در مدلی با ۴ متغیر توضیحی ($Y = \beta_۱ + \beta_۲X_۲ + \beta_۳X_۳ + \beta_۴X_۴ + u$)، فرض کنید می‌خواهیم دو قید هم‌زمان را بیازماییم:

۱. $\beta_۲ + \beta_۳ = ۱$

۲. $\beta_۴ = ۰$

ماتریس Rو بردار rبه شکل زیر تنظیم می‌شوند:

$$R = \begin{bmatrix} ۰ & ۱ & ۱ & ۰ \\ ۰ & ۰ & ۰ & ۱ \end{bmatrix}_{۲\times۴} , \beta = \begin{bmatrix} \beta_۱ \\ \beta_۲ \\ \beta_۳ \\ \beta_۴ \end{bmatrix} , r = \begin{bmatrix} ۱ \\ ۰ \end{bmatrix}_{۲\times۱}$$

۲۱.۳. آماره آزمون والد ماتریسی

بردار اختلاف $R\hat{\beta} - r$مقدار انحراف برآوردهای نمونه‌ای از فرضیه صفر را می‌سنجد. آماره آزمون Fوالد به صورت فرم درجه دوم زیر محاسبه می‌شود:

$$F = \frac{\left(R\hat{\beta} - r\right)' [R(X'X)^{-۱}R']^{-۱}\left(R\hat{\beta} - r\right)}{q \cdot s^۲} \sim F_{(q, n-k)}$$

که در آن qتعداد قیود مستقل (درجه آزادی صورت) و $n - k$درجه آزادی مخرج است.

۲۱.۴. رویکرد مدل مقید و نامقید (Restricted vs. Unrestricted Regression)

روش بسیار ساده‌تر و معادل تجربی برای محاسبه آماره آزمون فوق بدون درگیر شدن با جبر ماتریسی، برازش دو مدل رگرسیون است:

۱. **مدل نامقید (Unrestricted Model - UR):** تخمین مدل کامل با تمام متغیرها و بدون هیچ قیدی:

$$R^۲_{UR} \text{ و } SSR_{UR}$$

۲. **مدل مقید (Restricted Model - R):** تخمین مدلی که قیود فرض صفر ($H_۰$) در معادلات آن اعمال و تحمیل شده است:

$$R^۲_R \text{ و } SSR_R$$

فرمول آماره F:

از آنجا که تحمیل قید همواره خطای مدل را افزایش می‌دهد، همواره داریم $SSR_R \geq SSR_{UR}$. آماره آزمون عبارت است از:

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n-k)}$$

یا بر حسب ضریب تعیین:

$$F = \frac{(R_{UR}^{۲} - R_{R}^{۲})/q}{(۱ - R_{UR}^{۲})/(n-k)}$$

قاعده تصمیم‌گیری:

اگر $F_{\text{calc}} > F_{\text{critical}}(q, n-k)$یا p-value $< ۰/۰۵$باشد، فرض صفر رد می‌شود؛ یعنی قیود تحمیل‌شده با داده‌های تجربی سازگاری ندارند و نمی‌توان این متغیرها را حذف یا ضرایب آن‌ها را محدود کرد.

۲۱.۵. مثال عددی تجربی

فرض کنید روی یک نمونه ۶۰ تایی ($n = ۶۰$) با ۴ متغیر توضیحی ($k = ۴$) رگرسیون نامقید برآورد شده و $SSR_{UR} = ۱۵۰$به دست آمده است. دو قید خطی هم‌زمان ($q = ۲$) تحمیل شده و در رگرسیون مقید مقدار $SSR_R = ۱۹۵$گزارش می‌شود:

- درجات آزادی صورت: $q = ۲$
- درجات آزادی مخرج: $n - k = ۶۰ - ۴ = ۵۶$

$$F_{\text{calc}} = \frac{(۱۹۵ - ۱۵۰)/۲}{۱۵۰/۵۶} = \frac{۴۵/۲}{۲.۶۷۸} = \frac{۲۲.۵}{۲.۶۷۸} \approx ۸.۴۰$$

مقدار بحرانی جدول Fدر سطح خطای ۵٪ با درجات آزادی ۲ و ۵۶ تقریباً برابر با $۳/۱۶$است. چون $۸/۴۰ > ۳/۱۶$است، فرض صفر با قاطعیت رد می‌شود و قیود پذیرفته نیستند.

منابع موضوع ۲۱

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۷: «آزمون فرضیه‌های خطی و آزمون والد»، صص ۲۸۵–۳۱۰.
- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی، فصل ۸: «آزمون‌های قیود بر ضرایب».
- Greene, W. H. (۲۰۱۸). Econometric Analysis, Chapter ۵: “Testing Linear Restrictions”, pp. ۱۲۰–۱۳۸.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ٤: “Testing Multiple Linear Restrictions: The F Test”, pp. ۱٤۰–۱۵۸.

موضوع ۲۲: ماهیت متغیرهای مجازی (Dummy Variables) و تله متغیر مجازی (Dummy Variable Trap)

۲۲.۱. فلسفه و لزوم ورود متغیرهای کیفی به رگرسیون

تاکنون متغیرهای مورد استفاده در مدل‌ها همگی از جنس متغیرهای کمی یا مقداری (مانند درآمد، قیمت، سن، موجودی سرمایه) با ماهیت فاصله‌ای یا نسبی بودند. با این حال، در پدیده‌های اقتصادی و اجتماعی، بسیاری از عوامل تعیین‌کننده ماهیت کیفی دارند؛ عواملی نظیر:

- جنسیت (مرد / زن)
- وضعیت تأهل (مجرد / متأهل)
- نژاد، قومیت یا موقعیت جغرافیایی (شمال / جنوب، پایتخت / سایر شهرها)
- نوع مالکیت بنگاه (دولتی / خصوصی)
- دوره‌های بحران اقتصادی، جنگ، تحریم یا تغییرات ناگهانی سیاست‌گذاری

برای لحاظ کردن این ویژگی‌های کیفی مقوله‌ای (Categorical) در یک مدل رگرسیون خطی، از **متغیرهای مجازی یا موهومی (Dummy Variables)** استفاده می‌شود که با مقادیر دوگانه عددی کدگذاری می‌گردند:

$$D_i = \begin{cases} ۱ & \text{اگر ویژگی مورد نظر در مشاهده } i \text{ وجود داشته باشد} \\ ۰ & \text{اگر ویژگی مورد نظر وجود نداشته باشد} \end{cases}$$

۲۲.۲. تله متغیر مجازی (Dummy Variable Trap)

یکی از شایع‌ترین خطاهای تصریح در اقتصادسنجی کاربردی، گرفتاری در **تله متغیر مجازی** است.

بیان مسأله:

فرض کنید متغیری کیفی دارای m طبقه متمایز و مانعةالجمع است (مانند ۴ فصل سال: بهار، تابستان، پاییز، زمستان). اگر پژوهشگر برای هر ۴ طبقه، متغیر مجازی بسازد ($D_۱, D_۲, D_۳, D_۴$) و هم‌زمان عرض از مبدأ ($\beta_۰$) را نیز در مدل نگه دارد:

$$Y_i = \beta_۰ + \beta_۱ X_i + \gamma_۱ D_{۱i} + \gamma_۲ D_{۲i} + \gamma_۳ D_{۳i} + \gamma_۴ D_{۴i} + u_i$$

از آنجا که هر مشاهده الزماً متعلق به یکی از این ۴ طبقه است، در هر سطر داده‌ها همواره جمع متغیرهای مجازی برابر با عدد ۱ است:

$$\sum_{j=۱}^{۴} D_{ji} = D_{۱i} + D_{۲i} + D_{۳i} + D_{۴i} = ۱ \; \forall i$$

ستون متناظر با عرض از مبدأ در ماتریس طرح (X) نیز برداری از اعداد ۱ است:

$$x_{۰} = \begin{bmatrix} ۱ \\ ۱ \\ \vdots \\ ۱ \end{bmatrix} \implies x_{۰} = D_{۱} + D_{۲} + D_{۳} + D_{۴}$$

این رابطه خطی دقیق بدین معناست که ستون عرض از مبدأ ترکیب خطی کامل ۴ ستون متغیرهای مجازی است. در نتیجه:

$$\text{rank}(X) < k \implies \det(X'X) = ۰$$

ماتریس $X'X$غیرقابل وارون شده و برآوردکننده OLS یعنی $(X'X)^{-۱}X'y$**نامعین** می‌گردد. به این وضعیت **همخطی کامل (Perfect Multicollinearity)** یا **تله متغیر مجازی** می‌گویند.

۲۲.۳. راهکارهای گریز از تله متغیر مجازی

برای رهایی از این تله، دو قاعده جبری وجود دارد:

قاعده اول: حذف یک متغیر مجازی و حفظ عرض از مبدأ (رویکرد رایج)

اگر یک پدیده کیفی دارای mحالت باشد، حداکثر $m - ۱$متغیر مجازی وارد مدل می‌شود. طبقه‌ای که متغیر مجازی آن حذف می‌شود، به عنوان **دسته پایه (Base / Benchmark / Reference Category)** تعریف می‌گردد. در این حالت، تمامی ضرایب نسبت به دسته پایه سنجیده و تفسیر می‌شوند.

قاعده دوم: حذف عرض از مبدأ و حفظ تمام mمتغیر مجازی

اگر مدل بدون عرض از مبدأ برآورد شود، می‌توان تمام mمتغیر مجازی را وارد کرد. در این ساختار، ضریب هر متغیر مجازی مستقیماً نشان‌دهنده میانگین همان طبقه است، اما به دلیل حذف عرض از مبدأ، مقدار $R^{۲}$ممکن است اعتبار تحلیلی مرسوم را از دست بدهد. بنابراین رویکرد اول همواره ترجیح داده می‌شود.

منابع موضوع ۲۲

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۸: «متغیرهای مجازی و کاربرد آن در اقتصادسنجی»، صص ۳۲۱–۳۴۰.
- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی (گجراتی)، دانشگاه تهران، جلد دوم، فصل ۹: «رگرسیون بر متغیرهای مجازی»، صص ۴۴۵–۴۷۵.

- Wooldridge, J. M. (۲۰۲۰). *Introductory Econometrics: A Modern Approach*, ۷th ed., Chapter ۷: “Multiple Regression Analysis with Qualitative Information: Binary (or Dummy) Variables”, pp. ۲۲۰–۲۳۵.
- Gujarati, D. N., & Porter, D. C. (۲۰۰۹). *Basic Econometrics*, Chapter ۹: “Dummy Variable Regression Models”, pp. ۲۷۷–۲۹۵.

موضوع ۲۳: اثرات متغیر مجازی بر عرض از مبدأ و تفسیر گروه پایه (Benchmark Category)

۲۳.۱. مدل متغیر مجازی بر عرض از مبدأ (Differential Intercept Coefficient)

ساده‌ترین نوع استفاده از متغیر مجازی، فرض تغییر در سطح مبنای متغیر وابسته بدون تغییر در شیب رابطه است. به مدل دستمزد زیر دقت کنید:

$$W_i = \beta_0 + \beta_1 S_i + \gamma_1 D_i + u_i$$

که در آن:

- W_i: دستمزد ساعتی (هزار تومان)
- S_i: سال‌های تحصیل (متغیر کمی)
- D_i: متغیر مجازی جنسیت ($D_i = 1$ برای مردان و $D_i = 0$ برای زنان)

۲۳.۲. تفکیک مدل بر اساس دسته‌ها و تفسیر ضرایب

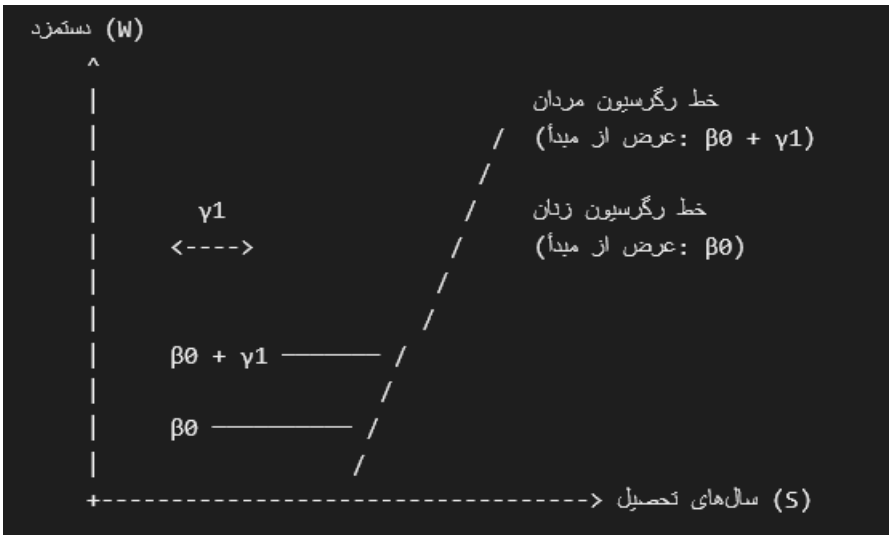
امید ریاضی شرطی دستمزد را برای هر دو گروه استخراج می‌کنیم:

۱. **برای زنان ($D_i = 0$ – گروه مبنا یا مرجع):**

$$E(W_i, | S_i D_i = 0) = \beta_0 + \beta_1 S_i + \gamma_1(0) = \beta_0 + \beta_1 S_i$$

۲. **برای مردان ($D_i = 1$):**

$$E(W_i, | S_i D_i = 1) = \beta_0 + \beta_1 S_i + \gamma_1(1) = (\beta_0 + \gamma_1) + \beta_1 S_i$$



تفسیر دقیق ضرایب:

- β_0: عرض از مبدأ **گروه پایه (زنان)**؛ یعنی دستمزد پایه زنانی که سابقه تحصیلی ندارند ($S = 0$).
- β_1: بازده تحصیلی (شیب خط رگرسیون)؛ به ازای هر یک سال افزایش تحصیل، دستمزد هم برای زنان و هم برای مردان به اندازه یکسان β_1افزایش می‌یابد (فرض شیب یکسان).
- γ_1: **ضریب تفاوت عرض از مبدأ (Differential Intercept)**؛ اختلاف میانگین دستمزد مردان و زنان در هر سطح معین و یکسان از تحصیلات. اگر $\gamma_1 > 0$و از نظر آماری معنادار باشد، نشان‌دهنده شکاف دستمزد مثبت به نفع مردان پس از کنترل اثر تحصیلات است.

۲۳.۳. متغیرهای طبقه‌ای چندگانه (بیش از دو حالت)

اگر بخواهیم منطقه سکونت را با سه موقعیت «تهران»، «سایر مراکز استان‌ها» و «مناطق روستایی» وارد کنیم:

- متغیرهای مجازی: D_1برای تهران و D_2برای مراکز استان.
- گروه پایه: مناطق روستایی ($D_1 = 0, D_2 = 0$).

معادله:

$$W_i = \beta_0 + \beta_1 S_i + \gamma_1 D_{1i} + \gamma_2 D_{2i} + u_i$$

- عرض از مبدأ روستاییان: β_0
- عرض از مبدأ ساکنان تهران: $\beta_0 + \gamma_1$(تفاوت دستمزد تهران نسبت به روستا: γ_1)
- عرض از مبدأ سایر مراکز استان‌ها: $\beta_0 + \gamma_2$(تفاوت دستمزد سایر استان‌ها نسبت به روستا: γ_2)
- تفاوت دستمزد تهران در مقایسه با سایر مراکز استان‌ها: $\gamma_1 - \gamma_2$

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، صص ۳۲۵–۳۴۵.
- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی، جلد دوم، فصل ۹.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ۷, pp. ۲۲۲–۲۲۹.
- Stock, J. H., & Watson, M. W. (۲۰۱۹). Introduction to Econometrics, ٤th ed., Chapter ۸: “Nonlinear Regression Functions”, pp. ۲٦۰–۲۷۵.

موضوع ۲۴: اثرات متقابل و متغیرهای مجازی شیب (Slope Dummies and Interaction Terms)

۲۴.۱. ضرورت مدل‌سازی تغییرات در شیب

در موضوع قبلی فرض شد که متغیر مجازی فقط سطح مبدأ را جابه‌جا می‌کند و شیب رابطه (نرخ بازده تحصیلات) برای هر دو گروه یکسان است. اما در دنیای واقعی ممکن است بازدهی هر سال تحصیل برای گروه‌های مختلف یکسان نباشد (مثلاً بازده هر سال تحصیل در بازار کار برای مردان بالاتر از زنان باشد، یا بهره‌وری سرمایه در شرکت‌های بورسی بیشتر از شرکت‌های غیربورسی باشد). برای آزمون این مسأله از **جملات متقابل یا تعاملی (Interaction Terms)** بهره می‌گیریم.

۲۴.۲. تصریح مدل تعاملی کامل

با ضرب متغیر مجازی در متغیر پیوسته، یک **متغیر مجازی شیب (Slope Dummy)** ایجاد می‌شود:

$$W_i = \beta_0 + \beta_1 S_i + \gamma_0 D_i + \gamma_1 (D_i \times S_i) + u_i$$

تفکیک معادلات برای دو گروه:

۱. **گروه پایه (زنان: $D_i = 0$):**

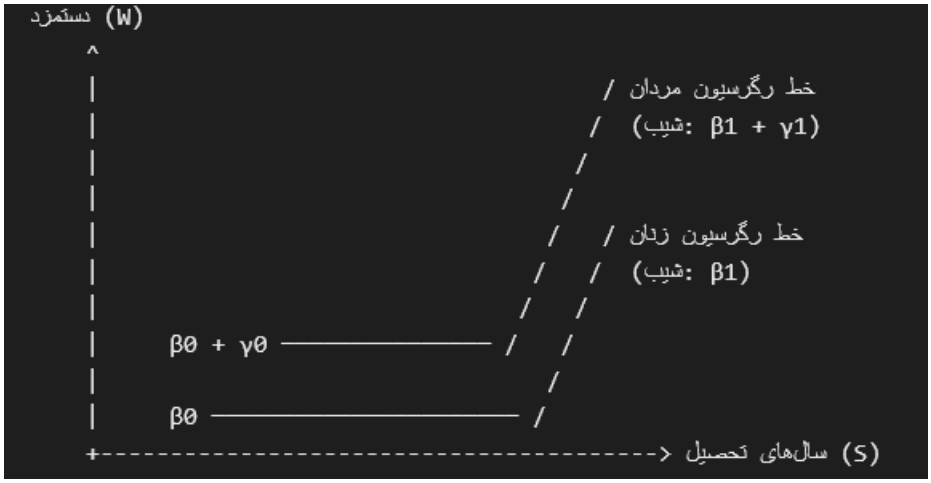
$$E(W_i, | S_i D_i = 0) = \beta_0 + \beta_1 S_i$$

- عرض از مبدأ: β_0
- شیب (بازده تحصیل): β_1

۲. **گروه هدف (مردان: $D_i = 1$):**

$$E(W_i, | S_i D_i = 1) = (\beta_0 + \gamma_0) + (\beta_1 + \gamma_1) S_i$$

- عرض از مبدأ: $\beta_0 + \gamma_0$
- شیب (بازده تحصیل): $\beta_1 + \gamma_1$



آزمون‌های آماری مرتبط با این مدل:

۱. **آزمون یکسانی عرض از مبدأ:** $H_0: \gamma_0 = 0$ (با آزمون t).

۲. **آزمون یکسانی شیب (بازدهی برابر تحصیل):** $H_0: \gamma_1 = 0$ (با آزمون t). اگر $\gamma_1 > 0$ و معنادار باشد، نشان می‌دهد شیب خط مردان تندتر است و بازدهی هر سال آموزش برای آن‌ها بالاتر است.

۳. **آزمون هم‌زمانی برابری دو رابطه:** $H_0: \gamma_0 = 0$ و $\gamma_1 = 0$ (با آزمون مقید F).

۲۴.۳. اثر متقابل دو متغیر مجازی با یکدیگر (Dummy-on-Dummy Interaction)

می‌توان اثر متقابل دو ویژگی کیفی مستقل را نیز بررسی کرد؛ مثلاً جنسیت ($D_M = 1$ اگر مرد) و تحصیلات دانشگاهی ($D_U = 1$ اگر مدرک دانشگاهی دارد):

$$W_i = \alpha_0 + \alpha_1 D_{Mi} + \alpha_2 D_{Ui} + \alpha_3 (D_{Mi} \times D_{Ui}) + u_i$$

- زن غیردانشگاهی ($D_M = 0, D_U = 0$): دستمزد پایه $= \alpha_0$
- مرد غیردانشگاهی ($D_M = 1, D_U = 0$): دستمزد پایه $= \alpha_0 + \alpha_1$
- زن دانشگاهی ($D_M = 0, D_U = 1$): دستمزد پایه $= \alpha_0 + \alpha_2$
- مرد دانشگاهی ($D_M = 1, D_U = 1$): دستمزد پایه $= \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$

ضریب α_3 اصطلاحاً **پاداش تعاملی (Interaction Premium)** نام دارد و نشان می‌دهد که آیا داشتن هم‌زمان هر دو ویژگی اثری فراتر از جمع ساده اثرات جداگانه دارد یا خیر.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، صص ۳۴۵–۳۶۰.
- نوفرستی، محمد (۱۴۰۰). مقدمه‌ای بر اقتصادسنجی، فصل ۵: «متغیرهای موهومی شیب و تعاملی»، صص ۱۲۰–۱۴۲.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ۷: “Interactions Involving Dummy Variables”, pp. ۲۳۵–۲۴۸.
- Greene, W. H. (۲۰۱۸). Econometric Analysis, Chapter ٦, pp. ۱۸۰–۱۹۵.

موضوع ۲۵: مدل‌سازی فصلی‌بودن با متغیرهای مجازی در سری‌های زمانی (Seasonal Adjustment)

۲۵.۱. ماهیت نوسانات فصلی در داده‌های کلان اقتصادی

بسیاری از سری‌های زمانی اقتصادی مانند مصرف برق، تولید محصولات کشاورزی، حجم سفرها و گردشگری، نقدینگی، خرید شب عید و نرخ بیکاری دارای نوسانات چرخه‌ای درون‌سالی معین و منظمی هستند که به آن **فصلی‌بودن (Seasonality)** می‌گویند. برای جلوگیری از تداخل نوسانات تقویمی با اثرات واقعی متغیرها، مدل‌سازی یا زدودن اثرات فصلی الزامی است.

۲۵.۲. تصریح رگرسیون با متغیرهای مجازی فصلی

فرض کنید داده‌های مورد بررسی سه‌ماهه (Quarterly) هستند. برای جلوگیری از تله متغیر مجازی، ۳ متغیر مجازی تعریف کرده و یکی از فصول (مثلاً فصل چهارم/زمستان) را به عنوان گروه پایه لحاظ می‌کنیم:

$$Y_t = \alpha_0 + \delta_1 Q_{1t} + \delta_2 Q_{2t} + \delta_3 Q_{3t} + \beta X_t + u_t$$

که در آن:

- $Q_{1t} = 1$ اگر مشاهده t در فصل بهار باشد و در غیر این صورت ۰.
- $Q_{2t} = 1$ اگر در فصل تابستان باشد و در غیر این صورت ۰.
- $Q_{3t} = 1$ اگر در فصل پاییز باشد و در غیر این صورت ۰.
- فصل زمستان (Q_{4t}): فصل مبنا.

تفسیر مقادیر فصلی:

- میانگین شرطی در زمستان: $\alpha_0 + \beta X_t$
- میانگین شرطی در بهار: $(\alpha_0 + \delta_1) + \beta X_t$
- میانگین شرطی در تابستان: $(\alpha_0 + \delta_2) + \beta X_t$
- میانگین شرطی در پاییز: $(\alpha_0 + \delta_3) + \beta X_t$

ضرایب $\delta_1, \delta_2, \delta_3$ تفاوت سطح فعالیت اقتصادی در هر فصل را نسبت به فصل زمستان می‌سنجند.

۲۵.۳. آزمون وجود فصلی‌بودن (Seasonal F-test)

برای پاسخ به این پرسش که «آیا متغیر وابسته در طول فصول نوسان معناداری را تجربه می‌کند؟»، آزمون مقید Fوالد تحت فرضیه صفر زیر اجرا می‌شود:

$$H_۰: \delta_۱ = \delta_۲ = \delta_۳ = ۰$$

- اگر فرض صفر رد شود، وجود الگوی فصلی معنادار از نظر آماری تأیید می‌گردد و متغیرهای فصلی باید حتماً در تصریح نهایی حفظ شوند تا مانع اریبی متغیر حذف‌شده گردند.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۸: «تعدیل فصلی با متغیرهای مجازی»، صص ۳۶۰–۳۷۲.
- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی، جلد دوم، فصل ۹.
- Enders, W. (2014). Applied Econometric Time Series, 4th ed., Wiley, Chapter 1: “Difference Equations and Seasonality”, pp. 22–35.
- Johnston, J., & DiNardo, J. (1997). Econometric Methods, Chapter 4, pp. 110–125.

موضوع ۲۶: شکست ساختاری و آزمون ثبات پارامترها (Chow Test for Structural Break)

۲۶.۱. مفهوم پدیده شکست ساختاری (Structural Break)

یکی از فروض بنیادین OLS در داده‌های سری زمانی، **ثبات پارامترها (Parameter Stability)** در طول کل دوره نمونه‌گیری است؛ یعنی بردار ضرایب βدر طول زمان ثابت بماند. اما رویدادهای بنیادین تاریخی و تکانه‌های برون‌زا نظیر:

- وقوع انقلاب، جنگ تحمیلی، یا بحران‌های حاد ژئوپلیتیک
- شوک‌های نفتی شدید (نظیر شوک‌های ۱۹۷۳ و ۱۹۷۹ یا سقوط قیمت نفت)
- آزادسازی‌های ارزی، تحریم‌های بانکی یا تغییر رژیم‌های پولی

می‌توانند شیب یا عرض از مبدأ روابط اقتصادی را به طور دائمی تغییر دهند. اگر شکست ساختاری نادیده گرفته شود، مدلی که برای کل دوره تخمین زده می‌شود دچار **خطای فاحش تصریح (Specification Error)** شده و برآوردها کاملاً گمراه‌کننده خواهند بود.

۲۶.۲. آزمون شکست ساختاری چاو (Gregory Chow Test)

گرگوری چاو روشی سیستماتیک برای آزمون این مسأله ارائه داد که آیا رابطه اقتصادی قبل و بعد از یک نقطه عطف تاریخی (T_0) دچار دگرگونی ساختاری شده است یا خیر.

منطق آزمون:

فرض کنید حجم کل مشاهدات nباشد. نمونه را در زمان وقوع شوک (T_0) به دو زیرنمونه تفکیک می‌کنیم:

- زیرنمونه اول: مشاهدات ۱تا n_1(دوره قبل از شکست ساختاری)
- زیرنمونه دوم: مشاهدات $n_1 + 1$تا n(دوره پس از شکست با تعداد $n_2 = n - n_1$)

سه مدل رگرسیون مجزا با روش OLS برآورد می‌گردند:

۱. **مدل تجمیعی برای کل دوره (Pooled Model):**

$$Y_t = \beta_0 + \beta_1 X_t + u_t \implies SSR_P \ (df = n - k)$$

۲. **مدل زیرنمونه اول (دوره قبل از شوک):**

$$Y_t = \beta_0^{(1)} + \beta_1^{(1)} X_t + u_t \implies SSR_1 \ (df = n_1 - k)$$

۳. **مدل زیرنمونه دوم (دوره بعد از شوک):**

$$Y_t = \beta_0^{(2)} + \beta_1^{(2)} X_t + u_t \implies SSR_2 \ (df = n_2 - k)$$

۲۶.۳. استخراج آماره آزمون چاو

مجموع مربعات خطای نامقید برابر است با جمع پسماندهای دو مدل تفکیک‌شده:

$$SSR_{UR} = SSR_1 + SSR_2$$

در حالی که مدل تجمیعی کل دوره به منزله مدل مقید است ($SSR_R = SSR_P$).

فرضیه صفر و فرضیه مقابل:

$$\begin{cases} H_0: \beta^{(1)} = \beta^{(2)} & \text{(عدم شکست ساختاری؛ پارامترها در هر دو دوره یکسان‌اند)} \\ H_1: \beta^{(1)} \neq \beta^{(2)} & \text{(T_0 وقوع شکست ساختاری در زمان)} \end{cases}$$

آماره آزمون Fچاو به صورت زیر استخراج می‌شود:

$$F_{Chow} = \frac{\big(SSR_P - (SSR_1 + SSR_2)\big)/k}{(SSR_1 + SSR_2)/(n - 2k)} \sim F_{(k,\, n-2k)}$$

که در آن kتعداد پارامترهای تخمین‌زده‌شده در هر رگرسیون است.

قاعده تصمیم‌گیری:

اگر $F_{Chow} > F_{critical}(k, n-2k)$یا $p\text{-value} < 0.05$باشد، فرض صفر رد می‌شود و وقوع شکست ساختاری تأیید می‌گردد. در این صورت، تخمین یک معادله واحد برای کل دوره خطاست و باید برای هر دوره مدل جداگانه‌ای تخمین زده شود یا اثر شوک با متغیرهای مجازی کنترل گردد.

۲۶.۴. هم‌ارزی آزمون چاو با متغیرهای مجازی تعاملی

آزمون چاو را می‌توان عیناً از طریق تخمین یک مدل واحد با متغیر مجازی گسست زمانی ($D_t = 1$ برای زمان‌های پس از شوک) و ضرب آن در تمامی متغیرهای مستقل اجرا کرد:

$$Y_t = \beta_0 + \beta_1 X_t + \gamma_0 D_t + \gamma_1 (D_t \times X_t) + u_t$$

آزمون هم‌زمان $H_0: \gamma_0 = \gamma_1 = 0$از طریق آزمون والد F، دقیقاً آماره‌ای برابر با آماره آزمون چاو تولید می‌کند، با این مزیت که خطاهای استاندارد هر شکست را نیز گزارش می‌دهد.

۲۶.۵. مثال عددی آزمون چاو

فرض کنید برای بررسی تابع مصرف ایران طی سال‌های ۱۳۶۰ تا ۱۴۰۰ ($n = ۴۰$ سال):

- پارامترها: عرض از مبدأ و درآمد ($k = ۲$).
- زمان فرضی شکست ساختاری: سال ۱۳۷۲ (آغاز برنامه پنج‌ساله دوم و تعدیل اقتصادی).
- طول زیرنمونه‌ها: $n_۱ = ۱۳$ سال و $n_۲ = ۲۷$ سال.

نتایج برآورد SSR:

$$SSR_P = ۱۲۵۰\ ,\ SSR_۱ = ۲۸۰\ ,\ SSR_۲ = ۴۴۰$$

محاسبه آماره آزمون:

$$SSR_{UR} = SSR_۱ + SSR_۲ = ۲۸۰ + ۴۴۰ = ۷۲۰$$

درجات آزادی صورت: $k = ۲$

درجات آزادی مخرج: $n - ۲k = ۴۰ - ۲(۲) = ۳۶$

$$F_{Chow} = \frac{(۱۲۵۰ - ۷۲۰)/۲}{۷۲۰/۳۶} = \frac{۵۳۰/۲}{۲۰} = \frac{۲۶۵}{۲۰} = ۱۳.۲۵$$

مقدار بحرانی توزیع Fدر سطح ۵٪ با درجات آزادی ۲ و ۳۶ برابر با ۳/۲۶است. چون ۳/۲۶ < ۱۳/۲۵، فرض صفر ثبات پارامترها با قاطعیت رد می‌شود و وقوع تغییر ساختاری در تابع مصرف محرز است.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۸: «آزمون شکست ساختاری چاو»، صص ۳۷۲–۳۹۰.
- نوفرستی، محمد (۱۴۰۰). مقدمه‌ای بر اقتصادسنجی، فصل ۵، صص ۱۴۰–۱۵۸.
- Chow, G. C. (۱۹۶۰). “Tests of Equality Between Sets of Coefficients in Two Linear Regressions”, Econometrica, ۲۸(۳), pp. ۵۹۱–۶۰۵.
- Greene, W. H. (۲۰۱۸). Econometric Analysis, Chapter ۶: “Structural Change and Parameter Constancy”, pp. ۱۹۵–۲۱۰.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ۷, pp. ۲۴۸–۲۵۵.

موضوع ۲۷: ماهیت و پیامدهای ناهمسانی واریانس (Heteroscedasticity)

۲۷.۱. بازتعریف فرض همسانی واریانس (Homoscedasticity) در برابر ناهمسانی

یکی از فروض پنج‌گانه بنیادی گاوس–مارکوف این بود که واریانس جزء اخلال برای تمامی مشاهدات مقدار ثابت و یکسانی دارد:

$$Var(u_i \mid X) = \sigma^2 \; \forall i = 1,2, \dots, n$$

اما در بسیاری از مطالعات تجربی (به‌ویژه در داده‌های مقطعی نظیر خانوار، بنگاه‌ها یا شهرها)، با افزایش مقادیر متغیر توضیحی (مثلاً درآمد یا اندازه بنگاه)، میزان پراکندگی یا نوسان پسماندها نیز افزایش یا کاهش می‌یابد. به این پدیده ناهمسانی واریانس (Heteroscedasticity) می‌گویند؛ یعنی:

$$Var(u_i \mid X) = \sigma_i^2$$

مثال شهودی:

رابطه میان «درآمد خانوار» و «پس‌انداز خانوار» را در نظر بگیرید. در خانوارهای با درآمد پایین، همگی ناچارند تمام درآمد را صرف هزینه‌های معیشتی پایه کنند و پراکندگی پس‌انداز بسیار ناچیز است (Var کوچک). اما در خانوارهای با درآمد بالا، برخی پس‌انداز زیادی دارند و برخی دیگر مصرف بالایی می‌کنند؛ بنابراین پراکندگی پس‌انداز به شدت زیاد می‌شود (Var بزرگ).

۲۷.۲. پیامدهای ناهمسانی واریانس بر برآوردگرهای OLS

هنگامی که ناهمسانی واریانس رخ می‌دهد، چه اتفاقی برای برآوردگرهای OLS می‌افتد؟

۱. ناارییی حفظ می‌شود: برآوردکننده OLS همچنان نااریب (Unbiased) است، زیرا اثبات ناارییی ($E(\hat{\beta}) = \beta$) متکی بر فرض همسانی واریانس نبود و صرفاً به $E(u_i) = 0$ و برون‌زایی وابستگی داشت.

۲. کارایی از دست می‌رود (از دست رفتن ویژگی BLUE): برآوردگر OLS دیگر دارای کمترین واریانس در میان تمامی برآوردگرهای خطی نیست؛ یعنی دیگر کاردآمد (Efficient) نیست.

۳. اریبی در محاسبه خطاهای استاندارد: فرمول سنتی واریانس برآوردگر OLS یعنی $Var(\hat{\beta}) = s^2(X'X)^{-1}$ تحت ناهمسانی واریانس دیگر معتبر نیست و مقادیر خطای استاندارد (SE) محاسبه‌شده به شدت اریب خواهند بود.

۴. بی‌اعتباری آزمون‌های t و F: از آنجا که خطاهای استاندارد اریب هستند، آماره‌های t و F از توزیع‌های استاندارد خود پیروی نمی‌کنند؛ در نتیجه استنتاج آماری، فواصل اطمینان و تصمیم‌گیری‌های مبتنی بر معناداری کاملاً نامعتبر و گمراه‌کننده خواهند شد.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۹: «ناهمسانی واریانس: ماهیت و پیامدها»، صص ۴۰۱–۴۱۸.
- ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی (گجراتی)، جلد دوم، فصل ۱۱: «ناهمسانی واریانس»، صص ۵۲۱–۵۴۵.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics: A Modern Approach, ۷th ed., Chapter ۸: “Heteroscedasticity”, pp. ۲٦۰–۲۷۸.
- Greene, W. H. (۲۰۱۸). Econometric Analysis, ۸th ed., Pearson, Chapter ۹: “Heteroscedasticity and Autocorrelation”, pp. ۳۲۰–۳٤٥.

موضوع ۲۸: آزمون‌های تشخیص ناهمسانی واریانس (بروش–پگان و وایت)

۲۸.۱. آزمون بروش–پگان–گادفری (Breusch-Pagan-Godfrey Test - BPG)

آزمون BPG بررسی می‌کند که آیا واریانس جزء اخلال (u_i^2) تابع خطی از یک یا چند متغیر توضیحی مدل است یا خیر.

مراحل اجرای آزمون:

۱. مدل اصلی را با OLS برآورد کرده و پسماندها (e_i) را ذخیره می‌کنیم.

۲. متغیر مربعات پسماند یعنی e_i^2را بر روی متغیرهای توضیحی اصلی مدل رگرسیون می‌کنیم:

$$e_i^2 = \alpha_0 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \cdots + \alpha_k X_{ki} + v_i$$

۳. ضریب تعیین این رگرسیون کمکی ($R_{e^2}^2$) را استخراج می‌کنیم.

٤. آماره آزمون BPG بر اساس ضریب تعیین و حجم نمونه (n) محاسبه می‌شود:

$$LM = n \cdot R_{e^2}^2 \sim \chi^2_{(k-1)}$$

که در آن $k-1$تعداد متغیرهای توضیحی مدل است.

قاعده تصمیم‌گیری:

- فرض صفر (H_0): واریانس پسماندها همسان است (عدم ناهمسانی واریانس).
- اگر مقدار آماره LMاز مقدار بحرانی کای‌دو بزرگتر باشد یا $p\text{-value} < 0/05$، فرض صفر رد شده و وجود ناهمسانی واریانس تأیید می‌گردد.

۲۸.۲. آزمون عمومی وایت (White's General Test)

مزیت بزرگ آزمون وایت نسبت به BPG این است که نیازی به فرض نوع خاصی از رابطه ناهمسانی واریانس ندارد و علاوه بر متغیرهای اصلی، توان‌های دوم (مربعات) و اثرات متقابل متغیرهای توضیحی را نیز در مدل کمکی وارد می‌کند.

مراحل آزمون وایت:

۱. پسماندهای رگرسیون اصلی (e_i) را به دست می‌آوریم.

۲. معادله کمکی وایت را برآورد می‌کنیم (برای مدلی با دو متغیر X_2, X_3):

$$e_i^2 = \alpha_0 + \alpha_1 X_{2i} + \alpha_2 X_{3i} + \alpha_3 X_{2i}^2 + \alpha_4 X_{3i}^2 + \alpha_5 (X_{2i} \cdot X_{3i}) + v_i$$

۳. آماره آزمون بر اساس ضریب تعیین این مدل محاسبه می‌شود:

$$\text{White Statistic} = n \cdot R^2_{e^2} \sim \chi^2_{(df)}$$

که درجه آزادی آن برابر با تعداد کل ضرایب تخمین‌زده‌شده در معادله کمکی (بدون احتساب عرض از مبدأ) است.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۹: «آزمون‌های بروش-پگان و وایت»، صص ۴۲۰–۴۳۵.
- White, H. (۱۹۸۰). “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”, *Econometrica*, ٤٨(٤), pp. ٨١٧–٨٣٨.
- Stock, J. H., & Watson, M. W. (٢٠١٩). *Introduction to Econometrics*, Chapter ١١: “Regression with a Single Regressor: Assessing Studies Based on Multiple Regression”, pp. ٣٨٠–٣٩٥.

موضوع ۲۹: راهکارهای رفع ناهمسانی واریانس (خطاهای استاندارد استوار و GLS)

۲۹.۱. خطاهای استاندارد استوار در برابر ناهمسانی (Robust Standard Errors / White-Huber SE)

پیش از رایج شدن نرم‌افزارهای مدرن، محققان ناچار بودند شکل دقیق ناهمسانی واریانس را کشف کنند. اما امروزه استاندارد طلایی در پژوهش‌های تجربی استفاده از **خطاهای استاندارد همساز با ناهمسانی واریانس (Heteroscedasticity-Robust Standard Errors)** است که توسط وایت، هابر و همکاران توسعه یافته است.

منطق کار:

این روش به جای تغییر دادن ساختار ضرایب برآوردشده ($\hat{\beta}$ که همچنان ناریب است)، ماتریس واریانس–کوواریانس ضرایب را با استفاده از مقادیر واقعی مربعات پسماند نمونه اصلاح می‌کند:

$$Var_{Robust}(\hat{\beta}) = (X'X)^{-1}\left(\sum_{i=1}^{n} e_i^2\, x_i x_i'\right)(X'X)^{-1}$$

با استفاده از این فرمول، خطاهای استاندارد اصلاح شده و در نتیجه آماره‌های tو فواصل اطمینان کاملاً قابل اتکا می‌شوند (در دستورات EViews کافی است گزینه *White* انتخاب شود).

۲۹.۲. حداقل مربعات تعمیم‌یافته (Generalized Least Squares - GLS)

اگر شکل دقیق ناهمسانی واریانس مشخص باشد (مثلاً بدانیم واریانس خطا متناسب با مجذور یک متغیر توضیحی مانند $Var(u_i) = \sigma^2 X_i^2$است)، روش بهینه‌تر اعمال **حقل یا حداقل مربعات تعمیم‌یافته (GLS)** است.

مراحل روش GLS (حداقل مربعات وزنی - WLS):

اگر واریانس هر مشاهد مشخص یا متناسب با یک متغیر وزن‌دهی مثل W_iباشد، طرفین معادله اصلی رگرسیون بر مقدار ریشه واریانس (h_i) تقسیم می‌شوند تا مدل جدید دارای واریانس همسان گردد:

$$\frac{Y_i}{h_i} = \beta_0\left(\frac{1}{h_i}\right) + \beta_1\left(\frac{X_i}{h_i}\right) + \left(\frac{u_i}{h_i}\right)$$

مدل تبدیل‌شده جدید دارای پسماند همسان با واریانس ثابت است و اعمال OLS بر روی آن، برآوردگری کارآمد (BLUE) به نام **WLS (Weighted Least Squares)** تولید می‌کند.

منابع

- سوری، علی (۱۴۰۲). اقتصادسنجی (جلد اول)، فصل ۹: «روش خطاهای استاندارد مقاوم و GLS»، صص ۴۳۶–۴۵۵.

- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ۸: “Handling Heteroscedasticity with Robust Standard Errors and WLS”, pp. ۲۷۵–۲۹۰.

موضوع ۳۰: ماهیت و پیامدهای خودهمبستگی (Autocorrelation) در سری‌های زمانی

۳۰.۱. تعریف خودهمبستگی در پسماندها

یکی دیگر از فروض پنج‌گانه گاوس–مارکوف فرض عدم وجود کوواریانس میان اجزای اخلال دوره‌های مختلف است:

$$Cov(u_i, u_j \mid X) = E(u_i u_j) = ۰ \quad \forall i \neq j$$

اگر این فرض نقض شود و پسماندهای دوره‌های مختلف با یکدیگر همبستگی داشته باشند، به آن **خودهمبستگی (Autocorrelation یا Serial Correlation)** می‌گویند. خودهمبستگی عمدتاً در داده‌های سری زمانی مشاهده می‌شود.

متداول‌ترین شکل خودهمبستگی (فرآیند مرتبه اول)AR۱)):

$$u_t = \rho u_{t-۱} + \varepsilon_t$$

که در آن:

- ρضریب خودهمبستگی مرتبه اول است و شرط پایداری آن $| \rho | < ۱$است.
- ε_tیک خطای سفید (White Noise) با میانگین صفر و واریانس ثابت $\sigma_\varepsilon^۲$است.

۳۰.۲. پیامدهای خودهمبستگی بر برآوردگرهای OLS

۱. **ناااریبی حفظ می‌شود:** برآوردگر OLS همچنان ناااریب باقی می‌ماند.

۲. **از دست رفتن کارایی (عدم کارایی OLS):** واریانس برآوردگرهای OLS به شدت افزایش می‌یابد و دیگر کارآمد نیستند.

۳. **اریبی شدید در خطاهای استاندارد:** تحت خودهمبستگی (به‌ویژه با $\rho > ۰$)، واریانس پسماندها کمتر از مقدار واقعی برآورد می‌شود؛ در نتیجه خطاهای استاندارد بسیار کوچک‌تر از حد واقعی گزارش می‌شوند.

٤. **تورم آماره tو خطای نوع اول (پندار معناداری کاذب):** کوچکی خطاهای استاندارد باعث می‌شود آماره‌های tو Fبه صورت کاذب بزرگ شوند؛ در نتیجه پژوهشگر به اشتباه تصور می‌کند یک متغیر تأثیر معناداری دارد، در حالی که این معناداری صرفاً ناشی از خودهمبستگی سری زمانی است ($R^۲$ کاذب بالا).

منابع

- سوری، علی (۱۴۰۲). *اقتصادسنجی (جلد اول)*، فصل ۱۰: «خودهمبستگی: ماهیت، پیامدها و آزمون‌ها»، صص ۴۷۱–۴۹۵.
- ابریشمی، حمید (۱۴۰۱). *مبانی اقتصادسنجی*، جلد دوم، فصل ۱۲: «خودهمبستگی»، صص ۵۸۱–۶۱۰.
- Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, Chapter ۱۲: “Serial Correlation and Heteroscedasticity in Time Series Regressions”, pp. ۳۶۰–۳۸۵.
- Enders, W. (۲۰۱٤). Applied Econometric Time Series, Chapter ۳, pp. ۱۱۵–۱٤۰.

موضوع ۳۱: آزمون‌های تشخیص خودهمبستگی (دوربین–واتسون و بروش–پگان / LM)

۳۱.۱. آزمون دوربین–واتسون (Durbin-Watson - DW Test)

متداول‌ترین آزمون برای تشخیص خودهمبستگی مرتبه اول ($AR(۱)$) آماره دوربین–واتسون است.

فرمول آماره DW:

$$DW = \frac{\sum_{t=۲}^{n}(e_t - e_{t-۱})^۲}{\sum_{t=۱}^{n} e_t^۲}$$

با بازآرایی این رابطه، ارتباط نزدیکی میان آماره DW و ضریب خودهمبستگی پسماندها (ρ) برقرار می‌شود:

$$DW \approx ۲(۱ - \hat{\rho})$$

دامنه تغییرات و قواعد قضاوت:

- دامنه آماره DW بین ۰تا ۴است:
  - اگر $\hat{\rho} = ۰ \implies DW \approx ۲$(عدم خودهمبستگی).
  - اگر $\hat{\rho} > ۰ \implies DW < ۲$(خودهمبستگی مثبت).
  - اگر $\hat{\rho} < ۰ \implies DW > ۲$(خودهمبستگی منفی).

جدول بحرانی دوربین–واتسون:

این آزمون دارای دو مقدار بحرانی پایین (d_L) و بالا (d_U) است و با یک **منطقه خاکستری (Inconclusive Zone)** مواجه است:

۱. اگر $۰ < DW < d_L$: رد فرض صفر و وجود خودهمبستگی مثبت.

۲. اگر $d_L \le DW \le d_U$: منطقه ابهام (نمی‌توان با قطعیت نظر داد).

۳. اگر $d_U < DW < ۴ - d_U$: عدم رد فرض صفر (پذیرش عدم خودهمبستگی).

۳۱.۲. آزمون براوش–گادفری (Breusch-Godfrey LM Test)

آزمون دوربین–واتسون دارای محدودیت‌هایی است (از جمله عدم قابلیت استفاده در مدل‌هایی که وقفه متغیر وابسته در آن‌ها حضور دارد). **آزمون ضریب لاگرانژ بروش–گادفری (LM Test)** آزمونی بسیار قدرتمندتر است که خودهمبستگی‌های مرتبه بالاتر ($AR(p)$) را نیز پوشش می‌دهد.

مراحل آزمون بروش–گادفری:

۱. رگرسیون اصلی را برآورد کرده و پسماندها (e_t) را استخراج می‌کنیم.

۲. پسماند e_tرا بر روی متغیرهای توضیحی اصلی و مقادیر باوقفه پسماند تا مرتبه p($e_{t-۱}, e_{t-۲}, \dots, e_{t-p}$) رگرسیون می‌کنیم:

$$e_t = \beta_۰ + \beta_۱ X_t + \rho_۱ e_{t-۱} + \rho_۲ e_{t-۲} + \cdots + \rho_p e_{t-p} + v_t$$

۳. ضریب تعیین این رگرسیون کمکی ($R_e^۲$) را استخراج کرده و آماره آزمون را محاسبه می‌کنیم:

$$LM = n \cdot R_e^۲ \sim \chi_{(p)}^۲$$

قاعده تصمیم‌گیری:

اگر مقدار آماره LMاز مقدار بحرانی کای‌دو با pدرجه آزادی بیشتر باشد، فرض صفرِ عدم خودهمبستگی رد می‌شود.

منابع

- سوری، علی (۱۴۰۲). *اقتصادسنجی (جلد اول)*، فصل ۱۰: «آزمون دوربین–واتسون و بروش–گادفری»، صص ۴۹۶–۵۲۰.
- Durbin, J., & Watson, G. S. (۱۹۵۰). "Testing for Serial Correlation in Least Squares Regression", *Biometrika*, ۳۷(۳/٤), pp. ٤۰۹–٤۲۸.
- Greene, W. H. (۲۰۱۸). *Econometric Analysis*, Chapter ۹, pp. ۳٦۰–۳۸۰

موضوع ۳۲: ماهیت، تعریف، انواع و ریشه‌های همخطی چندگانه (Multicollinearity)

۳۲.۱. جایگاه نظری موضوع

فرض چهارم از فروض کلاسیک گاوس–مارکوف (در صورت ماتریسی) بدین شرح است که هیچ‌یک از متغیرهای توضیحی نتوانند به‌صورت ترکیب خطی از یکدیگر بیان شوند؛ به بیان دیگر، بردارهای ستونی ماتریس Xباید **مستقل خطی (Linearly Independent)** باشند. نقض این فرض، پدیده‌ای است که به آن **همخطی چندگانه** گفته می‌شود. این پدیده نخستین بار توسط **راگنار فریش (Ragnar Frisch، برنده نوبل اقتصاد ۱۹۶۹)** در دهه ۱۹۳۰ مطرح و نام‌گذاری شد؛ او در مطالعات خود درباره هم‌نوسانی متغیرهای اقتصادی کلان مشاهده کرد که بسیاری از آن‌ها به‌طور هم‌زمان و همراه یکدیگر حرکت می‌کنند و جداسازی اثر خالص هر یک از دیگری دشوار است.

۳۲.۲. تعریف و تفکیک دو نوع همخطی

مدل رگرسیون خطی چندمتغیره را در نظر بگیرید:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \ , \ i = 1,2, \ldots, n$$

الف) همخطی کامل (Perfect Multicollinearity)

اگر ضرایب ثابتی ($c_1, c_2, \ldots, c_k$) وجود داشته باشند که همگی صفر نباشند و برای **تمامی مشاهدات** رابطه زیر برقرار باشد:

$$c_1 X_{1i} + c_2 X_{2i} + \cdots + c_k X_{ki} = 0$$

در این حالت ستون‌های ماتریس Xوابسته خطی‌اند، رتبه ماتریس از $k + 1$کمتر شده و ماتریس $X'X$**منفرد (Singular)** می‌شود:

$$\det(X'X) = 0$$

در نتیجه وارون ماتریس وجود ندارد و برآوردگر OLS اصلاً قابل محاسبه نخواهد بود:

$$\hat{\beta} = (X'X)^{-1}X'Y \ \Rightarrow \ \text{تعریف‌نشده}$$

مثال شهودی: اگر در مدل تولید، متغیر «تعداد کارگران» (L) و متغیر «تعداد کل نیروی انسانی» ($L + M$، شامل مدیران) را همزمان وارد کنیم، چون Lترکیب خطی از دیگری است، OLS عملاً فرومی‌پاشد.

ب) همخطی ناقص (Imperfect Multicollinearity)

حالت رایج در داده‌های واقعی؛ در آن متغیرها همبستگی **بالا اما نه برابر با یک** دارند:

$$c_1 X_{1i} + c_2 X_{2i} + \cdots + c_k X_{ki} = v_i$$

که در آن v_iیک جزء تصادفی است. ماتریس $X'X$وارون‌پذیر است ($\det \neq 0$) ولی بسیار کوچک است؛ اصطلاحاً ماتریس «بدشرطی» (Ill-conditioned) است. برآوردگر OLS محاسبه می‌شود اما بسیار **ناپایدار (Unstable)** و دارای واریانس بالا خواهد بود.

۳۲.۳. ریشه‌های شکل‌گیری همخطی در پژوهش‌های تجربی

۱. **هم‌حرکتی متغیرهای کلان (Comovement):** تولید ناخالص داخلی، مصرف خصوصی، سرمایه‌گذاری، حجم پول و شاخص قیمت‌ها همگی با روند زمانی مشترک حرکت می‌کنند.

۲. **متغیرهای هم‌مفهوم یا مشتق از هم:** مثل «درآمد ماهانه» و «درآمد سالانه» یا «تورم» و «نرخ رشد سطح قیمت‌ها».

۳. **رگرسیون‌های چندجمله‌ای:** ورود هم‌زمان X، X^2و X^3(مثل مدل‌های منحنی کوزنتس زیست‌محیطی) که ذاتاً همخطی ایجاد می‌کند.

٤. **تله متغیر مجازی:** ورود همه دسته‌های یک متغیر مقوله‌ای همراه با عرض از مبدأ (مانند ۴ متغیر مجازی فصول سال)، که همخطی کاملاً دقیق می‌سازد.

٥. **محدودیت داده و نمونه‌گیری:** دامنه تغییرات کم یک متغیر یا نمونه کوچک باعث پدیدار شدن همخطی کاذب می‌شود (وارد کردن شرط به‌جای قید).

٦. **تعریف متغیرهای تعاملی و توان دوم** بدون مرکزسازی داده‌ها.

۳۲.۴. شهود اقتصادی: چرا جداسازی ضرایب دشوار می‌شود؟

در همخطی شدید، داده‌ها به ما «اطلاعات کافی برای تفکیک اثر خالص» نمی‌دهند. فرض کنید هزینه زندگی (Y) تابعی از درآمد (X_1) و ثروت (X_2) باشد، اما در جامعه، درآمد و ثروت تقریباً با هم تغییر کنند. با هر واحد افزایش هم‌زمان در هر دو، تنها «اثر ترکیبی» آن‌ها مشاهده می‌شود، نه اثر خالص هر یک. به بیان ریاضی، با تغییر جزئی در داده‌ها، مبنای تجربی برای «چرخاندن» صفحه برازش (Hyperplane) حول داده‌ها نامشخص می‌شود و جهت‌گیری صفحه به‌شدت حساس به نویز خواهد بود. این همان مفهوم ناپایداری نمونه‌ای برآوردگرهاست.

منابع

۱. سوری، علی (۱۴۰۲). اقتصادسنجی، جلد اول: مفاهیم پایه، تهران: فرهنگ مهران، فصل ۸: «همخطی چندگانه: ماهیت و منابع»، صص ۳۵۵–۳۷۰.

۲. ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی (تألیف دامتار گجراتی)، جلد دوم، تهران: انتشارات دانشگاه تهران، فصل ۱۰، صص ۴۶۵–۴۸۰.

۳. نوفرستی، محمد (۱۳۹۸). ریشه واحد و هم‌جمعی در اقتصادسنجی، تهران: رسا، مقدمه فصل ۱ (بحث هم‌حرکتی متغیرهای سری زمانی).

٤. Frisch, R. (١٩٣٤). Statistical Confluence Analysis by Means of Complete Regression Systems. Oslo: University Institute of Economics.

٥. Gujarati, D. N., & Porter, D. C. (٢٠٠٩). Basic Econometrics, ٥th ed., McGraw-Hill, Chapter ١٠: “Multicollinearity”, pp. ٣٢١–٣٣٥.

٦. Wooldridge, J. M. (٢٠٢٠). Introductory Econometrics: A Modern Approach, ٧th ed., Cengage, Chapter ٣, Section ٣A (Appendix), pp. ١١٠–١١٥.

موضوع ۳۳: پیامدهای همخطی چندگانه بر برآوردگر OLS و استنباط آماری

۳۳.۱. آنچه حفظ می‌شود

نکته‌ای نظری بسیار مهم: همخطی ناقص، نقض فرض‌های $E(u \mid X) = ۰$یا همسانی واریانس **نیست**؛ بنابراین:

$$E(\hat{\beta} \mid X) = \beta$$

برآوردگر OLS همچنان **نااریب، سازگار (Consistent) و BLUE** باقی می‌ماند. بنابراین مشکل همخطی، «مشکل داده» است نه «مشکل مدل»؛ گفته می‌شود همخطی یک **مشکل نمونه‌ای (Sample Problem)** است، نه یک مشکل ساختاری.

۳۳.۲. تورم واریانس ضرایب

فرمول دقیق واریانس برآوردگر OLS عبارت است از:

$$Var(\hat{\beta}) = \sigma^{۲}(X'X)^{-۱}$$

در حالت دو متغیر توضیحی، واریانس برآوردگر شیب $\hat{\beta}_{۱}$عبارت است از:

$$Var(\hat{\beta}_{۱}) = \frac{\sigma^{۲}}{\sum x_{۱i}^{۲}\,(۱ - r_{۱۲}^{۲})}$$

که در آن $r_{۱۲}$همبستگی نمونه‌ای دو متغیر توضیحی است. عامل $\frac{۱}{۱-r_{۱۲}^{۲}}$دقیقاً همان **VIF** است:

r_{12}	$VIF = \frac{1}{1-r_{12}^2}$
0.0	1.00
0.5	1.33
0.8	2.78
0.9	5.26
0.95	10.26
0.99	50.25

به‌عبارت دیگر، با همبستگی ۹۹/۰ میان دو متغیر، واریانس برآوردگر شیب **۵۰ برابر** حالت بدون همبستگی می‌شود و خطای استاندارد حدود $\sqrt{۵۰} \approx ۷$برابر بزرگ می‌گردد.

۳۳.۳. پیامدها بر آماره‌های t و F

آماره tبرای آزمون معناداری هر ضریب:

$$t = \frac{\hat{\beta}_j - \beta_j^{\cdot}}{SE(\hat{\beta}_j)}$$

با بزرگ شدن SE، آماره tکوچک می‌شود و در نتیجه معناداری آماری متغیر رد می‌شود (خطای نوع دوم). پدیده معروف و متمایزکننده همخطی این است:

آزمون Fکلی مدل معنادار است، اما اکثر آماره‌های tتک‌تک متغیرها غیرمعنادارند؛ همراه با R^2و $\bar{R}^2$بالا.

این وضعیت «پارادوکس همخطی» نامیده می‌شود و در آزمون‌های تشخیص، نشانه‌ای کلیدی است.

۳۳.۴. ناپایداری ضرایب و حساسیت به مشخصات مدل

دترمینان کوچک $X'X$باعث می‌شود برآوردگرها به نویز نمونه‌ای فوق‌العاده حساس باشند:

- با حذف یا افزودن تنها **چند مشاهده**، مقدار و حتی **علامت** ضرایب معکوس می‌شود.
- ضرایب ممکن است علامت خلاف نظریه اقتصادی بگیرند (مثلاً اثر منفی درآمد بر مصرف!) که نشان‌دهنده به‌دام افتادن در همخطی است.
- قدرت پیش‌بینی برون‌نمونه‌ای مدل به‌شدت کاهش می‌یابد، زیرا ضرایب قابل اتکا نیستند.

۳۳.۵. تفسیر اقتصادی مختل می‌شود

از آنجا که رگرسیون چندمتغیره اثر **خالص (Ceteris Paribus)** هر متغیر را می‌سنجد، در حضور همخطی، داده‌ها اساساً اطلاعاتی برای تفکیک این اثر خالص ندارند. ضرایب برآوردی دیگر بازتاب پارامتر واقعی ساختاری نیستند بلکه ترکیب ناپایداری از اثرات هم‌بسته است.

منابع

۱. سوری، علی (۱۴۰۲). اقتصادسنجی، جلد اول، تهران: فرهنگ مهران، بخش «پیامدهای همخطی بر استنباط»، صص ۳۷۰–۳۸۵.

۲. ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی، جلد دوم، تهران: انتشارات دانشگاه تهران، فصل ۱۰، صص ۴۸۰–۵۰۰.

۳. Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, ۷th ed., Chapter ۳, pp. ۹۵–۱۱۵.

٤. Greene, W. H. (۲۰۱۸). Econometric Analysis, ۸th ed., Pearson, Chapter ۳: “The Algebra of Least Squares” (Ill-conditioning), pp. ۳۲–٤٥.

٥. Stock, J. H., & Watson, M. W. (۲۰۱۹). Introduction to Econometrics, ٤th ed., Pearson, Chapter ٦, Section on Imperfect Multicollinearity, pp. ۲۱٤–۲۱۸.

موضوع ۳۴: روش‌های تشخیص همخطی چندگانه

۳۴.۱. نشانه‌های استقرایی (Non-Statistical Indicators)

پیش از آزمون‌های رسمی، نشانه‌های زیر زنگ خطر محسوب می‌شوند:

۱. $R^۲$ و $\bar{R}^۲$ بسیار بالا همراه با آماره‌های t غیرمعنادار.

۲. علامت‌های ضرایب مغایر با انتظارات تئوری اقتصاد.

۳. تغییر شدید ضرایب با حذف/افزودن یک متغیر یا چند مشاهده.

۳۴.۲. ماتریس همبستگی جفتی

محاسبه همبستگی ساده r_{ij} میان همه متغیرهای توضیحی. مقادیر بالای ۰/۸ هشدار است. **محدودیت:** همخطی می‌تواند چندبعدی باشد (مثلاً $X_۱$ ترکیب خطی از $X_۲$ و $X_۳$) بدون آنکه هیچ r_{ij} بالایی وجود داشته باشد؛ پس این روش ناقص است.

۳۴.۳. آزمون فریش (Frisch Confluence Analysis)

رویکرد تاریخی فریش: بررسی رفتار ضرایب برآوردی در مدل‌های با k، $k-۱$، ... متغیر («رگرسیون‌های توده‌ای» – Bundled Regressions). اگر با افزودن هر متغیر، ضرایب سایر متغیرها دچار نوسان شدید شوند، نشانه همخطی است.

۳۴.۴. عامل تورم واریانس (Variance Inflation Factor - VIF)

رایج‌ترین شاخص استاندارد. برای هر متغیر X_j، رگرسیون کمکی زیر برآورد می‌شود:

$$X_j = \gamma_۰ + \gamma_۱ X_۱ + \cdots + \gamma_{j-۱} X_{j-۱} + \gamma_{j+۱} X_{j+۱} + \cdots + \gamma_k X_k + r_j$$

و با ضریب تعیین آن ($R_j^۲$):

$$VIF_j = \frac{۱}{۱ - R_j^۲}$$

- $VIF_j = ۱$: هیچ همخطی وجود ندارد.
- $VIF_j \geq ۵$ (برخی منابع ۴): همخطی مشکل‌ساز.
- $VIF_j \geq ۱۰$: همخطی شدید (قاعده سرانگشتی مرسوم).

توجه: با افزایش تعداد متغیرها در رگرسیون کمکی، R_j^2ذاتاً بالا می‌رود؛ لذا در مدل‌های بزرگ آستانه‌های ۵ یا ۱۰ محافظه‌کارانه‌اند و برخی پژوهشگران آستانه $VIF > 2.5$را برای علوم اجتماعی مناسب می‌دانند.

مثال عددی کوچک: فرض کنید $R_1^2 = 0.85$:

$$VIF_1 = \frac{1}{1 - 0.85} = \frac{1}{0.15} \approx 6.67$$

یعنی واریانس $\hat{\beta}_1$تقریباً ۶/۷ برابر حالت بدون همخطی متورم شده است.

۳۴.۵. شاخص وضعیت و تجزیه مقدار ویژه (Condition Number)

بر پایه مقادیر ویژه λماتریس $X'X$(پس از نرمال‌سازی متغیرها):

$$CN = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

قواعد مرسوم بلسی (Belsley) و همکاران:

- $CN < 10$: همخطی ضعیف.
- $10 \leq CN \leq 30$: همخطی متوسط.
- $CN > 30$یا ۵۰ <: همخطی شدید.

مزیت این شاخص نسبت به VIF آن است که همخطی‌های چندبعدی (شامل عرض از مبدأ) را نیز آشکار می‌کند؛ «تجزیه واریانس» بلسی (Variance Decomposition Proportions) نشان می‌دهد همخطی از کدام ترکیب متغیرها ناشی می‌شود.

۳۴.۶. آزمون کلاین (Klein's Rule of Thumb)

اگر همبستگی جفتی r_{ij}از R^2**کل مدل** بزرگ‌تر باشد، همخطی جدی تلقی می‌شود:

$$| r_{ij} | > R^2_{\text{model}} \Rightarrow \text{همخطی جدی}$$

منابع

۱. سوری، علی (۱۴۰۲). *اقتصادسنجی، جلد اول*، تهران: فرهنگ مهران، فصل ۸: «روش‌های تشخیص همخطی»، صص ۳۸۶–۴۰۰.

۲. ابریشمی، حمید (۱۴۰۱). *مبانی اقتصادسنجی، جلد دوم*، فصل ۱۰، صص ۵۰۰–۵۱۵.

۳. Belsley, D. A., Kuh, E., & Welsch, R. E. (۱۹۸۰). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.

٤. Gujarati, D. N., & Porter, D. C. (۲۰۰۹). Basic Econometrics, ٥th ed., Chapter ۱۰, pp. ۳۳٥–۳٥۰.

٥. Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, ۷th ed., Chapter ۳, Appendix ۳A.

موضوع ۳۵: راهکارها و درمان همخطی چندگانه

۳۵.۱. مقدمه‌ای بر منطق درمان

از آنجا که همخطی مشکل «داده» است نه «مدل»، هیچ درمان بی‌هزینه‌ای وجود ندارد؛ هر راهکار، نوعی مبادله (Trade-off) است. پیش از هر اقدام باید پرسید: **آیا همخطی اصلاً مشکل‌ساز است؟** اگر هدف صرفاً پیش‌بینی درون‌نمونه‌ای است، همخطی متوسط قابل تحمل است؛ اما اگر هدف تفسیر ضرایب یا آزمون فرض درباره آن‌هاست، درمان ضروری است.

۳۵.۲. راهکارهای مبتنی بر داده

۱. **افزایش حجم نمونه:** چون $Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$، افزایش n عناصر $X'X$ را بزرگ‌تر و واریانس را کوچک‌تر می‌کند. اگر ساختار همخطی در جامعه پایدار باشد، مشاهدات جدید تنوع لازم را فراهم می‌کنند.

۲. **افزایش تنوع داده (Cross-sectional pooling):** ترکیب داده‌های سری زمانی و مقطعی (Pooled Data) می‌تواند واریانس متغیرها را افزایش دهد؛ البته باید مراقب تفاوت‌های ساختاری باشد.

۳. **داده‌های جدید یا آزمایشگاهی:** در اقتصادسنجی تجربی مدرن، طراحی داده به‌گونه‌ای که همبستگی متغیرها کم باشد (مثل آزمایش‌های کنترل‌شده).

۳۵.۳. راهکارهای مبتنی بر مشخصات مدل

۱. **حذف متغیر زائد:** اگر دو متغیر عملاً یک مفهوم را می‌سنجند (مثل «مصرف» و «درآمد» برای سنجش اندازه خانوار)، بر مبنای نظریه یکی را حذف کنیم. **هشدار:** حذف متغیر مرتبط با مدل، **اریبی متغیر محذوف (Omitted Variable Bias)** ایجاد می‌کند؛ پس این تصمیم باید تئوری‌محور باشد، نه صرفاً آماری.

۲. **تبدیل متغیرها:**

- **تفاضل‌گیری (First Difference):** در سری زمانی، $\Delta X_t = X_t - X_{t-1}$ روند مشترک را حذف می‌کند و همخطی را کاهش می‌دهد (البته ممکن است خودهمبستگی جدید ایجاد کند).
- **نسبت‌سازی:** استفاده از نسبت‌ها مانند X_1/X_2 (مثلاً سرمایه‌گذاری به تولید).
- **مرکزسازی (Mean-centering):** $\tilde{X} = X - \bar{X}$ به‌ویژه برای مدل‌های تعاملی ($X_1 \cdot X_2$) که همخطی ذاتاً ایجاد می‌کنند؛ ضریب تعاملی تغییر معنایی نمی‌کند ولی $\hat{\beta}_1$ و $\hat{\beta}_2$ به «اثر در میانگین» تفسیر می‌شوند.

۳. **ترکیب متغیرها (Index Construction):** اگر چند متغیر یک سازه واحد (مثل «رفاه اقتصادی») را می‌سنجند، با تحلیل مؤلفه‌های اصلی (PCA) یا میانگین استانداردشده، شاخص واحدی بسازیم.

٤. **افزودن مشاهدات از منابع مکمل یا استفاده از پانل:** در پانل، تغییرات درونی و بین‌گروهی (Within/Between) به تفکیک اثرات کمک می‌کند.

۳۵.۴. راهکارهای مبتنی بر روش برآورد

۱. **رگرسیون ریج (Ridge Regression):** با افزودن جریمه $\ell_۲$:

$$\hat{\beta}_{Ridge} = \arg\min_{\beta} \left\{ \sum_{i=۱}^{n} (y_i - \beta_. - x_i'\beta)^۲ + \lambda \sum_{j=۱}^{k} \beta_j^۲ \right\}$$

ریج اریب است اما MSE کل را (به‌ویژه در همخطی شدید) کاهش می‌دهد و برای **پیش‌بینی** بسیار مناسب است. انتخاب λ با اعتبارسنجی متقاطع (Cross-Validation) انجام می‌شود.

۲. **رگرسیون مؤلفه‌های اصلی (PCR):** به‌جای متغیرهای اصلی، مؤلفه‌های اصلی (که متعامدند و همخطی صفر دارند) وارد مدل شوند.

۳. **تحلیل مسیر یا مدلیابی معادلات ساختاری (SEM):** در پژوهش‌های علوم اجتماعی، روابط هم‌بسته در قالب شبکه‌ای از معادلات تفکیک می‌شوند.

٤. **اقتصادسنجی بیزی (Bayesian Econometrics):** با اعمال «اطلاعات قبلی» (Prior)، واریانس برآوردگرها مهار می‌شود.

۳۵.۵. راهکار نرم‌افزاری

- **EViews:** View > Coefficient Diagnostics > Variance Inflation Factors (در برخی نسخه‌ها از طریق افزونه Add-ins).
- **Stata:** پس از برآورد، دستور vif را اجرا کنید؛ برای Condition Number از ۲coldiag (بسته کمکی).
- **R:** بسته car با تابع vif()؛ برای Condition Index از kappa(zX) یا colldiag().

۳۵.۶. جمع‌بندی و جدول تصمیم‌گیری

وضعیت پژوهش	راهکار پیشنهادی
هدف = پیش‌بینی	رگرسیون ریج یا PCR؛ همخطی متوسط قابل تحمل
هدف = تفسیر ضرایب	حذف/ترکیب متغیر، تبدیل‌ها، افزایش داده
همخطی از جمله تعاملی	مرکزسازی متغیرها
داده سری زمانی رونددار	تفاضل‌گیری یا مدل‌سازی هم‌جمعی
چند متغیر هم‌مفهوم	ساخت شاخص (PCA)

منابع

۱. سوری، علی (۱۴۰۲). اقتصادسنجی، جلد اول، تهران: فرهنگ مهران، فصل ۸: «راهکارهای درمان همخطی»، صص ۴۰۰–۴۱۵.

۲. ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی، جلد دوم، تهران: انتشارات دانشگاه تهران، فصل ۱۰، صص ۵۱۵–۵۳۰.

۳. محمدپور، احمد (۱۳۹۷). مدل‌سازی معادلات ساختاری در پژوهش‌های اجتماعی، تهران: جامعه‌شناسان (برای رویکرد SEM و ترکیب متغیرها).

٤. Hoerl, A. E., & Kennard, R. W. (۱۹۷۰). “Ridge Regression: Biased Estimation for Nonorthogonal Problems”, Technometrics, ۱۲(۱), pp. ۵۵–٦۷.

٥. Gujarati, D. N., & Porter, D. C. (۲۰۰۹). Basic Econometrics, ٥th ed., Chapter ۱۰, pp. ۳۵۰–۳۵۷.

٦. Wooldridge, J. M. (۲۰۲۰). Introductory Econometrics, ۷th ed., Chapter ۳ and Appendix on “Remedies for Multicollinearity”.

۷. James, G., Witten, D., Hastie, T., & Tibshirani, R. (۲۰۲۱). An Introduction to Statistical Learning, ۲nd ed., Springer, Chapter ٦ (Shrinkage Methods).

موضوع ۳۶: مدل‌های با وقفه توزیع‌شده (Distributed Lag Models) و مفاهیم پایه‌ای آن

۳۶.۱. جایگاه نظری و ضرورت تحلیل زمانی

در اقتصاد و علوم اجتماعی، اثر تغییر متغیرهای توضیحی بر متغیر وابسته همیشه در **همان دوره زمانی** (t) بروز نمی‌کند. برای مثال، افزایش نرخ سود بانکی (X) تأثیر خود را بر سرمایه‌گذاری یا تقاضای وام (Y) طی چند فصل آینده به‌تدریج اعمال می‌کند. مدل‌هایی که علاوه بر مقدار جاری متغیر توضیحی، مقادیر با وقفه (Lagged Values) آن را نیز وارد معادله می‌کنند، **مدل‌های با وقفه توزیع‌شده (Distributed Lag Models)** نامیده می‌شوند.

۳۶.۲. ساختار ریاضی مدل

شکل کلی یک مدل با وقفه توزیع‌شده با طول وقفه متناهی (q) به صورت زیر است:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_q X_{t-q} + u_t$$

که در آن:

- Y_t: متغیر وابسته در زمان t.
- $X_t, X_{t-1}, \ldots$: مقادیر جاری و وقفه‌دار متغیر مستقل.
- β_0: ضریب اثر فوری یا کوتاه‌مدت (Impact Multiplier).
- $\beta_j (j = 1,2,\ldots,q)$: ضرایب تأخیر یا اثرات وقفه.
- u_t: جزء اخلال سفید (White Noise).

۳۶.۳. ضرایب اثر کوتاه‌مدت، بلندمدت و تجمعی

۱. **اثر کوتاه‌مدت (Short-run Effect):** ضریب β_0 که نشان‌دهنده واکنش فوری Y به یک تغییر واحد در X در همان دوره وقوع است.

۲. **اثرات بین راهی (Intermediate Effects):** ضرایب $\beta_1, \beta_2, \ldots$ که میزان تأثیرگذاری را در دوره‌های بعدی ($t-1, t-2, \ldots$) نشان می‌دهند.

۳. **مجموع ضرایب یا اثر بلندمدت (Long-run Propensity - LRP):** فرض کنید سیستم به تعادل بلندمدت رسیده باشد ($X_t = X_{t-1} = \cdots = X_{t-q} = X^*$ و $u_t = 0$). اثر کل تغییر پایدار در X بر Y برابر است با جمع تمامی ضرایب:

$$\text{LRP} = \sum_{j=۰}^{q} \beta_j = \beta_۰ + \beta_۱ + \beta_۲ + \cdots + \beta_q$$

۳۶.۴. چالش‌های برآورد با OLS

تخمین مستقیم مدل‌های وقفه توزیع‌شده با OLS با دو مشکل اساسی روبه‌روست:

۱. **همخطی شدید (Multicollinearity):** متغیرهای $X_t, X_{t-۱}, X_{t-۲}, \dots$ معمولاً همبستگی بسیار بالایی با یکدیگر دارند که موجب تورم واریانس (VIF) و غیرمعنادار شدن ضرایب t می‌شود.

۲. **درجه آزادی محدود (Loss of Degrees of Freedom):** با افزایش طول وقفه q، تعداد مشاهدات قابل استفاده در نمونه $(T - q)$ به‌شدت کاهش می‌یابد.

منابع

۱. سوری، علی (۱۴۰۲). اقتصادسنجی پیشرفته: مباحثی در سری زمانی، تهران: فرهنگ مهران، فصل ۴: «مقدمه‌ای بر مدل‌های با وقفه توزیع‌شده»، صص ۱۸۵–۲۰۵.

۲. ابریشمی، حمید (۱۴۰۱). مبانی اقتصادسنجی (گجراتی)، جلد دوم، تهران: انتشارات دانشگاه تهران، فصل ۱۷، صص ۷۵۵–۷۷۵.

۳. Gujarati, D. N., & Porter, D. C. (۲۰۰۹). Basic Econometrics, ۵th ed., McGraw-Hill, Chapter ۱۷: “Time Series Econometrics: Distributed-Lag Models”, pp. ۶۵۵–۶۷۵.

۴. Greene, W. H. (۲۰۱۸). Econometric Analysis, ۸th ed., Pearson, Chapter ۲۰: “Models for Time Series Data”, pp. ۸۵۰–۸۷۰.

موضوع ۳۷: مدل‌های خودرگرسیونی و وقفه توزیع‌شده (ARDL) و مدل‌های کوک – دلک

۳۷.۱. جایگاه نظری و مزیت مدل‌های ARDL

برای غلبه بر محدودیت‌های مدل‌های وقفه توزیع‌شده متناهی (مانند همخطی و از دست رفتن درجات آزادی)، **اروین کوک (Nerlove)** و سپس **پاسران و همکاران (Pesaran et al.)** مدل‌های **خودرگرسیونی با وقفه توزیع‌شده (Autoregressive Distributed Lag - ARDL)** را توسعه دادند. در این مدل‌ها، علاوه بر وقفه‌های متغیر توضیحی X، وقفه متغیر وابسته Y_{t-1}نیز به عنوان متغیر توضیحی وارد معادله می‌شود.

۳۷.۲. ساختار ریاضی مدل ARDL(p, q)

یک مدل ARDL با pوقفه از متغیر وابسته و qوقفه از متغیر مستقل به صورت زیر نوشته می‌شود:

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=0}^{q} \beta_j X_{t-j} + u_t$$

۳۷.۳. شهود اقتصادی و فرآیند تعدیل (Nerlove's Partial Adjustment)

مدل تعدیل جزئی نشان می‌دهد که خانوارها یا بنگاه‌ها به دلیل هزینه‌های تعدیل (Adjustment Costs)، نمی‌توانند به‌صورت آنی به تغییرات متغیرهای مطلوب واکنش نشان دهند. فرم تعادل بلندمدت این مدل با محاسبه حالت ایستا ($Y_t = Y_{t-1} = Y^*$) بدست می‌آید:

$$Y^* = \frac{\alpha_0}{1 - \sum \phi_i} + \frac{\sum \beta_j}{1 - \sum \phi_i} X^*$$

که در آن ضریب بلندمدت (LRP) برابر است با:

$$LRP = \frac{\sum_{j=0}^{q} \beta_j}{1 - \sum_{i=1}^{p} \phi_i}$$

۳۷.۴. مدل‌های هندسی وقفه توزیع‌شده (Almon / Koyck)

- **مدل کوک (Koyck Transformation):** فرض می‌کند ضرایب وقفه‌ها با یک الگوی هندسی نزولی کاهش می‌یابند ($\beta_k = \beta_0 \lambda^k$ که $0 < \lambda < 1$). با جایگذاری در مدل اصلی، به یک مدل ARDL ساده‌شده تبدیل می‌شود.

- **روش آلمون (Almon Polynomial Lag):** فرض می‌کند ضرایب β_j روی یک چندجمله‌ای با درجه مشخص قرار دارند که مشکل همخطی را از طریق کاهش پارامترهای تخمینی حل می‌کند.

منابع

۱. سوری، علی (۱۴۰۲). اقتصادسنجی پیشرفته، تهران: فرهنگ مهران، فصل ۵: «مدل‌های ARDL و پویایی‌های کوتاه‌مدت و بلندمدت»، صص ۲۱۰–۲۴۰.

۲. مهرگان، نادر (۱۳۹۸). اقتصادسنجی سری‌های زمانی، تهران: انتشارات برادران، فصل ۶، صص ۱۴۵–۱۷۰.

۳. Pesaran, M. H., Shin, Y., & Smith, R. J. (۲۰۰۱). “Bounds Testing Approaches to the Analysis of Level Relationships”, Journal of Applied Econometrics, ۱۶(۳), pp. ۲۸۹–۳۲۶.

٤. Gujarati, D. N., & Porter, D. C. (۲۰۰۹). Basic Econometrics, ٥th ed., Chapter ۱۷, pp. ٦٧٥–٦٩٥.

موضوع ۳۸: ایستایی (Stationarity)، روندها و انشعابات سری‌های زمانی

۳۸.۱. مفهوم نظری ایستایی

یک سری زمانی (Y_t) زمانی **ایستا (Stationary)** یا پایا نامیده می‌شود که میانگین، واریانس و کوواریانس‌های باطنی (Autocovariances) آن در طول زمان **ثابت** بمانند؛ یعنی با تغییر مبدأ زمانی (t) تغییر نکنند.

به زبان ریاضی برای سری ایستا (از مرتبه ضعیف یا کوواریانس-پایا):

۱. $E(Y_t) = \mu$ (میانگین ثابت)

۲. $Var(Y_t) = E[(Y_t - \mu)^2] = \sigma^2$ (واریانس متناهی و ثابت)

۳. $Cov(Y_t, Y_{t-k}) = \gamma_k$ (کوواریانس به فاصله زمانی k بستگی دارد نه به زمان t)

۳۸.۲. پیامدهای سری‌های ناایستا (Non-Stationary Series)

اگر سری ناایستا باشد:

- نتایج رگرسیون دچار **رگرسیون کاذب (Spurious Regression)** می‌شود (R^2 بسیار بالا و آماره‌های t فریبنده در حالی که هیچ رابطه واقعی وجود ندارد).
- واریانس در طول زمان واگرا می‌شود و مفروضات استنباط آماری کلاسیک نقض می‌گردند.

۳۸.۳. انواع ناایستایی

۱. **ناایستایی مانا در روند (Trend-Stationary - TS):** سری دارای یک روند قطعی (Deterministic Trend) است:

$$Y_t = \alpha_0 + \alpha_1 t + u_t$$

که با حذف روند قطعی ($Y_t - \alpha_1 t$) سری ایستا می‌شود.

۲. **ناایستایی مانا در تفاضل یا ریشه واحد (Difference-Stationary - DS):** سری دارای روند تصادفی (Stochastic Trend) است و با یک‌بار تفاضل‌گیری ایستا می‌شود.

منابع

۱. نوفرستی، محمد (۱۳۹۸). *ریشه واحد و هم‌جمعی در اقتصادسنجی*، تهران: رسا، فصل ۱ و ۲: «مفهوم ایستایی و فرآیندهای تصادفی»، صص ۱۵–۵۵.

۲. سوری، علی (۱۴۰۲). *اقتصادسنجی پیشرفته*، فصل ۶: «مفاهیم ایستایی و ناایستایی»، صص ۲۴۵–۲۷۵.

۳. Hamilton, J. D. (۱۹۹٤). Time Series Analysis. Princeton University Press, Chapter ۱۵: “Stationary Time Series Models”, pp. ٤۳۰–٤٦۰.

٤. Enders, W. (۲۰۱٤). Applied Econometric Time Series, ٤th ed., Wiley, Chapter ۱, pp. ۱–۵۵.

موضوع ۳۹: آزمون‌های ریشه واحد (Unit Root Tests): دیکی-فولر و دیکی-فولر تعمیم‌یافته (ADF)

۳۹.۱. مبانی نظری آزمون ریشه واحد

مدل خودرگرسیونی ساده مرتبه اول AR(۱) را در نظر بگیرید:

$$Y_t = \rho Y_{t-۱} + u_t \ , \ -۱ \le \rho \le ۱$$

- اگر $\rho = ۱$ باشد، ریشه واحد وجود دارد و سری ناایستا (از نوع گام تصادفی یا Random Walk) است.
- اگر $|\rho| < ۱$ باشد، سری ایستا است.

برای آزمون فرض آماری $H_۰ : \rho = ۱$ در برابر $H_۱ : \rho < ۱$)، مدل را به صورت تفاضلی بازنویسی می‌کنند:

$$\Delta Y_t = \delta Y_{t-۱} + u_t \ (\text{که در آن } \delta = \rho - ۱)$$

آزمون فرض به صورت زیر تبدیل می‌شود:

$$H_۰ : \delta = ۰ \ (\text{وجود ریشه واحد / ناایستایی})$$

$$H_۱ : \delta < ۰ \ (\text{عدم وجود ریشه واحد / ایستایی})$$

۳۹.۲. آزمون دیکی-فولر (Dickey-Fuller - DF)

دیکی و فولر نشان دادند که آماره آزمون این فرض از توزیع استاندارد t پیروی نمی‌کند؛ زیرا تحت فرض صفر، آماره واگرا است. آن‌ها توزیع حدی جدیدی به نام **توزیع τ (Tau)** را استخراج و مقادیر بحرانی آن را جدول‌بندی کردند. سه حالت اصلی برای رگرسیون آزمون DF وجود دارد:

۱. بدون عرض از مبدأ و بدون روند: $\Delta Y_t = \delta Y_{t-۱} + u_t$

۲. با عرض از مبدأ (با رانش): $\Delta Y_t = \alpha_۰ + \delta Y_{t-۱} + u_t$

۳. با عرض از مبدأ و روند زمانی: $\Delta Y_t = \alpha_۰ + \alpha_۲ t + \delta Y_{t-۱} + u_t$

۳۹.۳. آزمون دیکی–فولر تعمیم‌یافته (Augmented Dickey-Fuller - ADF)

چنانچه جزء اخلال u_tدچار خودهمبستگی باشد، آزمون DF اعتبار خود را از دست می‌دهد. برای رفع این مشکل، **دیکی و فولر (۱۹۷۹, ۱۹۸۱)** وقفه‌های تفاضلی متغیر وابسته را به سمت راست معادله اضافه کردند تا خطاهای اخلال سفید شوند:

$$\Delta Y_t = \alpha_0 + \alpha_1 t + \delta Y_{t-1} + \sum_{i=1}^{m} \gamma_i \, \Delta Y_{t-i} + u_t$$

تعداد وقفه‌های کمکی (m) با استفاده از معیارهای اطلاع Akaike (AIC) یا Schwarz (SIC) تعیین می‌شود.

۳۹.۴. سایر آزمون‌های مهم ریشه واحد

- **آزمون فیلیپس–پرون (Phillips-Perron - PP):** رویکرد ناپارامتری برای اصلاح ناهمسانی واریانس و خودهمبستگی خطاها.
- **آزمون KPSS:** با فرض صفرِ «ایستایی» (برخلاف ADF که فرض صفرش ناایستایی است).

منابع

۱. نوفرستی، محمد (۱۳۹۸). ریشه واحد و هم‌جمعی در اقتصادسنجی، تهران: رسا، فصل ۳: «آزمون‌های ریشه واحد»، صص ۶۱–۱۱۰.

۲. سوری، علی (۱۴۰۲). اقتصادسنجی پیشرفته، فصل ۷: «آزمون‌های ریشه واحد»، صص ۲۸۰–۳۱۵.

۳. Dickey, D. A., & Fuller, W. A. (۱۹۷۹). “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”, Journal of the American Statistical Association, ۷٤(۳٦٦), pp. ٤۲۷–٤۳۱.

٤. Phillips, P. C. B., & Perron, P. (۱۹۸۸). “Testing for a Unit Root in Time Series Regression”, Biometrika, ۷٥(۲), pp. ۳۳٥–۳٤٦.

موضوع ۴۰: مفهوم هم‌جمعی (Cointegration) و رویکرد انگل-گرنجر

۴۰.۱. جایگاه نظری هم‌جمعی

در حالی که رگرسیون میان دو یا چند سری زمانی ناایستا معمولاً کاذب است، **انگل و گرنجر (Granger, & Engle ۱۹۸۷)** و **برنده نوبل اقتصاد (کلایو گرنجر)** مفهوم **هم‌جمعی** را معرفی کردند: اگر دو سری زمانی Y_t و X_t هر دو در سطح ناایستا و از مرتبه یکم باشند ($I(۱)$)، ممکن است ترکیب خطی از آن‌ها ایستا باشد ($I(۰)$). در این صورت، گفته می‌شود این دو متغیر با یکدیگر **هم‌جمع** هستند و یک **رابطه تعادلی بلندمدت** واقعی میان آن‌ها برقرار است.

۴۰.۲. تعریف ریاضی هم‌جمعی

فرض کنید:

$$Y_t \sim I(۱)\ ,\ X_t \sim I(۱)$$

اگر برداری مانند $\beta = (۱, -\beta_۱)'$ وجود داشته باشد به‌طوری که جزء اخلال حاصل از رگرسیون زیر ایستا باشد:

$$u_t = Y_t - \beta_۰ - \beta_۱ X_t \sim I(۰)$$

آنگاه متغیرهای Y_t و X_t هم‌جمع از مرتبه $(۱, ۱)$ هستند. این پدیده نشان می‌دهد نیروهای اقتصادی مانع از واگرایی دائم این دو متغیر در بلندمدت می‌شوند.

۴۰.۳. رویکرد دو مرحله‌ای انگل-گرنجر (Engle-Granger Two-Step Method)

این روش سنتی شامل دو مرحله اصلی است:

۱. **مرحله اول (برآورد رابطه بلندمدت):** برآورد رگرسیون تعادلی با OLS و استخراج اجزای اخلال (باقی‌مانده‌ها):

$$Y_t = \hat{\beta}_۰ + \hat{\beta}_۱ X_t + \hat{e}_t \ \Rightarrow\ \hat{e}_t = Y_t - \hat{\beta}_۰ - \hat{\beta}_۱ X_t$$

۲. **مرحله دوم (آزمون ریشه واحد روی باقی‌مانده‌ها):** اجرای آزمون ADF روی باقی‌مانده‌های برآورد شده ($\hat{e}_t$) **بدون عرض از مبدأ و روند**:

$$\Delta \hat{e}_t = \delta \hat{e}_{t-۱} + \sum_{i=۱}^{m} \gamma_i \, \Delta \hat{e}_{t-i} + v_t$$

اگر مقدار آماره آزمون از مقادیر بحرانی ویژه انگل-گرنجر کوچکتر (منفی‌تر) باشد، فرض صفر (عدم هم‌جمعی / ریشه واحد در باقی‌مانده‌ها) رد شده و هم‌جمعی تأیید می‌گردد.

۴۰.۴. محدودیت‌های روش انگل-گرنجر

- در نمونه‌های کوچک کارایی پایینی دارد.
- تنها اجازه وجود یک رابطه هم‌جمعی را در سیستم می‌دهد (در حالی که در سیستم‌های چندمتغیره ممکن است چندین بردار هم‌جمعی وجود داشته باشد که رویکرد جوهانسون آن را حل می‌کند).

منابع

۱. نوفرستی، محمد (۱۳۹۸). *ریشه واحد و هم‌جمعی در اقتصادسنجی*، تهران: رسا، فصل ۴ و ۵: «مفهوم هم‌جمعی و آزمون انگل-گرنجر»، صص ۱۱۵–۱۶۵.

۲. سوری، علی (۱۴۰۲). *اقتصادسنجی پیشرفته*، فصل ۸: «هم‌جمعی و مدل‌های تصحیح خطا»، صص ۳۲۰–۳۶۰.

۳. Engle, R. F., & Granger, C. W. J. (۱۹۸۷). “Co-integration and Error Correction: Representation, Estimation, and Testing”, Econometrica, ۵۵(۲), pp. ۲۵۱–۲۷۶.

٤. Banerjee, A., Dolado, J., Galbraith, J. W., & Hendry, D. F. (۱۹۹۳). Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data. Oxford University Press.

موضوع ۴۱: رویکرد جوهانسون به هم‌جمعی و مدل تصحیح خطا (VECM)

۴۱.۱. ضرورت نظری رویکرد سیستم‌های چندمتغیره

روش انگل–گرنجر (موضوع ۴۰) دو محدودیت اساسی دارد: (الف) تنها یک رابطه هم‌جمعی را می‌تواند شناسایی کند، (ب) انتخاب متغیر وابسته (نرمال‌سازی) بر نتایج اثر می‌گذارد، زیرا در مرحله اول یک جهت علیت تحمیل می‌شود. **سورن جوهانسون (Johansen, ۱۹۸۸, ۱۹۹۱)** و **جوسیلیوس (Juselius)** رویکرد حداکثر درست‌نمایی (Maximum Likelihood) را در چارچوب یک سیستم **VAR** توسعه دادند که تمام متغیرها در آن درون‌زا فرض می‌شوند و تعداد بردارهای هم‌جمعی به‌صورت درون‌زا تعیین می‌گردد.

۴۱.۲. مدل خودرگرسیون برداری (VAR) و بازنمایی تصحیح خطا

یک VAR از مرتبه p برای بردار k متغیره Y_t:

$$Y_t = \mu + \Pi_1 Y_{t-1} + \cdots + \Pi_p Y_{t-p} + \varepsilon_t$$

با تبدیل به فرم تفاضلی، **مدل تصحیح خطا (Vector Error Correction Model – VECM)** حاصل می‌شود:

$$\Delta Y_t = \mu + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \Pi\, Y_{t-1} + \varepsilon_t$$

که در آن:

$$\Gamma_i = -\left(\Pi_{i+1} + \cdots + \Pi_p\right),\ \Pi = \sum_{i=1}^{p} \Pi_i - I_k$$

۴۱.۳. رتبه ماتریس Π: قلب آزمون جوهانسون

تمام اطلاعات مربوط به هم‌جمعی در ماتریس Π متمرکز است:

$$\Pi = \alpha \beta'$$

که β ماتریس $k \times r$ بردارهای هم‌جمعی (Cointegrating Vectors) و α ماتریس $k \times r$ سرعت تعدیل (Speed of Adjustment) است. تحلیل رتبه Π:

رتبه (r)	تفسیر
$r = 0$	$\mathbf{\Pi} = \mathbf{0}$ ؛ هیچ رابطه هم‌جمعی وجود ندارد (VAR در تفاضل‌ها)
$0 < r < k$	r رابطه هم‌جمعی مستقل؛ VECM معتبر
$r = k$	$\mathbf{\Pi}$ تمام‌رتبه؛ همه متغیرها ایستا هستند (تناقض با فرض $I(1)$)

۴۱.۴. دو آماره آزمون جوهانسون

۱) آماره اثر (Trace Statistic):

$$\lambda_{trace}(r) = -T \sum_{i=r+۱}^{k} \ln\left(۱ - \hat{\lambda}_i\right)$$

این آماره آزمون $H_۰$: «حداکثر rبردار هم‌جمعی وجود دارد» را در برابر فرض مقابل «بیش از rبردار» می‌سنجد.

۲) آماره مقدار ویژه حداکثر (Maximum Eigenvalue Statistic):

$$\lambda_{max}(r, r+۱) = -T\ln\left(۱ - \hat{\lambda}_{r+۱}\right)$$

که $H_۰$: «r بردار هم‌جمعی» را در برابر $H_۱$: «$r+۱$ بردار» می‌آزماید. در هر دو، $\hat{\lambda}_i$مقادیر ویژه ماتریس Π(مرتب‌شده نزولی) است.

۴۱.۵. تفسیر αو βو آزمون‌های تشخیص

- **بردار β:** ضرایب رابطه تعادلی بلندمدت. برای شناسایی، قید نرمال‌سازی (مثلاً $\beta_۱ = ۱$) اعمال می‌شود.
- **بردار α:** سرعت بازگشت به تعادل پس از شوک؛ علامت منفی و معناداری آن همان **ضریب تصحیح خطا (ECT)** در ادبیات کلاسیک است. مقدار آن نشان می‌دهد چه کسری از عدم تعادل در هر دوره اصلاح می‌شود.
- **آزمون‌های تشخیص لازم:** آزمون ریشه واحد روی باقی‌مانده‌ها، پایداری VAR (ریشه‌های خارج از دایره واحد)، آزمون خودهمبستگی LM روی باقی‌مانده‌ها، نرمال بودن (Jarque–Bera) و آزمون علیت گرنجر در چارچوب VECM.

۴۱.۶. مثال کاربردی گذرا

در مطالعه‌ای درباره تقاضای پول ایران با $Y = (m_t, y_t, i_t)$(حجم پول، درآمد ملی، نرخ تورم) و $T = ۶۰$، اگر $\hat{\lambda}_۱ = ۰.۲۸$، $\hat{\lambda}_۲ = ۰.۰۹$:

$$\lambda_{trace}(۰) = -۶۰[\ln(۰.۷۲) + \ln(۰.۹۱)] = -۶۰[-۰.۳۲۸۵ - ۰.۰۹۴۳] \approx ۲۵.۴$$

که آماره اثر برای $r = ۰$در سطح ۵٪ (مقادیر بحرانی حدود ۷/۲۹ برای $k = ۳$) **رد نمی‌شود**، اما با حذف فرض ثابت، مقدار بحرانی تعدیل‌شده به‌دست می‌آید؛ لذا به‌کارگیری جدول Osterwald-Lenum ضروری است.

منابع

۱. نوفرستی، محمد (۱۳۹۸). ریشه واحد و هم‌جمعی در اقتصادسنجی، تهران: رسا، فصل ۶ و ۷: «رویکرد جوهانسون و مدل‌های VECM»، صص ۱۷۰–۲۳۰.

۲. سوری، علی (۱۴۰۲). اقتصادسنجی پیشرفته، تهران: فرهنگ مهران، فصل ۹: «هم‌جمعی در سیستم‌های چندمتغیره»، صص ۳۶۵–۴۱۰.

۳. Johansen, S. (۱۹۸۸). “Statistical Analysis of Cointegration Vectors”, Journal of Economic Dynamics and Control, ۱۲(۲–۳), pp. ۲۳۱–۲۵٤.

٤. Johansen, S. (۱۹۹۱). “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models”, Econometrica, ۵۹(٦), pp. ۱۵۵۱–۱۵۸۰.

۵. Osterwald-Lenum, M. (۱۹۹۲). “A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics”, Oxford Bulletin of Economics and Statistics, ۵٤(۳), pp. ٤٦۱–٤۷۲.

موضوع ۴۲: روش حداقل مربعات پویا (DOLS)، ARDL با آزمون مرزی پسران و جمع‌بندی راهنمای کاربست

۴۲.۱. ضرورت روش‌های جایگزین در نمونه‌های کوچک

در نمونه‌های کوچک (نظیر داده‌های سالانه اقتصادی که به دلیل محدودیت دسترسی، به‌ندرت از ۵۰ مشاهده فراتر می‌روند)، برآوردگرهای ML جوهانسون و OLS انگل–گرنجر **اریب** هستند و توزیعشان به پارامترهای مزاحم (Nuisance Parameters) حساس است. دو رویکرد مقاومی توسعه یافته است.

۴۲.۲. روش حداقل مربعات پویا (Dynamic OLS – DOLS)

سایکونن (,Saikkonen۱۹۹۱) و **استاک و واتسون (Watson, & Stock۱۹۹۳)** نشان دادند که افزودن پیش‌وقفه‌ها و پس‌وقفه‌های تفاضل متغیرهای توضیحی، اریبی همبستگی درونی (Endogeneity Bias) را رفع می‌کند. مدل DOLS:

$$Y_t = \beta_0 + \beta_1 X_t + \sum_{j=-q}^{q} \gamma_j \, \Delta X_{t+j} + e_t$$

- مجموع ΔX_{t+j}و لگاریتم‌های آن نقش اصطلاحات **پویا** در جهت سازگاری را ایفا می‌کنند.
- ضریب β_1مستقیماً برآوردگر رابطه بلندمدت است.
- واریانس آن با روش‌های مقاوم به خودهمبستگی (Newey–West) محاسبه می‌شود تا استنباط معتبر باشد.
- در نمونه‌های کوچک، انتخاب $q = 1$یا $q = 2$کافی است (با AIC/SIC).

۴۲.۳. رویکرد ARDL و آزمون مرزی پسران (Bounds Testing)

پسران، شین و اسمیت (۲۰۰۱) رویکردی بر پایه مدل ARDL و تبدیل آن به فرم تصحیح خطا ارائه دادند که **مزیت کلیدی** آن این است که حتی اگر برخی متغیرها $I(0)$و برخی $I(1)$باشند (اما هیچ‌کدام $I(2)$نباشند)، معتبر باقی می‌ماند. آزمون مرزی دو مرحله دارد:

مرحله ۱: آزمون وجود رابطه بلندمدت با آماره F پسران

مدل شرطی تصحیح خطا برآورد می‌شود:

$$\Delta Y_t = \alpha_0 + \phi_1 Y_{t-1} + \phi_2 X_{t-1} + \sum_{i=1}^{p} \gamma_i \, \Delta Y_{t-i} + \sum_{j=0}^{q} \delta_j \, \Delta X_{t-j} + u_t$$

فرض صفر عدم رابطه بلندمدت: $H_0 : \phi_1 = \phi_2 = 0$. توزیع آماره Fغیراستاندارد است و دو کران‌بندی ارائه شده است:

وضعیت مقایسه	نتیجه
$F >$ کران بالای (Upper Bound، همه متغیرها $I(1)$)	رد H_0 ؛ وجود رابطه بلندمدت تأیید می‌شود
$F <$ کران پایین (Lower Bound، همه متغیرها $I(0)$)	عدم رد H_0 ؛ رابطه‌ای وجود ندارد
کران پایین $\leq F \leq$ کران بالا	نتیجه **نامعین** (Inconclusive)

مرحله ۲: برآورد رابطه بلندمدت (اگر مرحله ۱ تأیید شود) با فرم ARDL و محاسبه ضرایب بلندمدت.

۴۲.۴. محدودیت‌های ARDL

۱. کارایی خودرگرسیون مشروط، مستقل بودن رگرسیورها از خطا را لازم دارد؛ در حضور شکست ساختاری و روند شدید ممکن است این فرض نقض شود.

۲. سناریوی «نامعین» در نمونه‌های کوچک شایع است و نیاز به تفسیر محتاطانه دارد.

۳. در حضور شکست ساختاری، باید پیش از آزمون، آزمون شکست (مثل Zivot–Andrews) اجرا و متغیر مجازی مناسب افزوده شود.

۴۲.۵. جدول مقایسه‌ای روش‌های آزمون روابط بلندمدت

روش	تعداد روابط	اعتبار با متغیرهای $I(0)$ و $I(1)$	کارایی در نمونه کوچک	حساسیت به جهت‌گیری
انگل–گرنجر (۲ مرحله‌ای)	یک	فقط $I(1)$	نسبتاً پایین	بالا
جوهانسون و VECM	چندگانه	فقط $I(1)$	متوسط	پایین
DOLS	یک	فقط $I(1)$	**بالا**	متوسط
ARDL / Bounds	یک	**هر دو**	**بالا**	پایین

۴۲.۶. جمع‌بندی مسیر عملی کاربست (راهنمای استادان و دانشجویان)

۱. بررسی نمودار زمانی، تعیین نوع روند (TS/DS).

۲. اجرای آزمونهای ADF/PP/KPSS برای هر متغیر (با وقفه بهینه بر پایه AIC/SIC).

۳. آزمون شکست ساختاری (Chow و Zivot–Andrews) در صورت وجود شکاف ظاهری در نمودار.

۴. انتخاب رویکرد: در نمونه بزرگ و سیستم چندمتغیره ← جوهانسون؛ در نمونه کوچک یا اختلاط $I(0)/I(1)$ ← ARDL Bounds یا DOLS.

۵. برآورد فرم ECM و بررسی ضریب α(ECT) منفی و معنادار.

۶. آزمون‌های تشخیص: ثبات ضرایب (CUSUM)، خودهمبستگی LM، ناهمسانی واریانس.

۷. پردازش و گزارش در EViews (View > Cointegration Test > Johansen System Cointegration Test یا ARDL Bounds Test)، Stata (ardl, vecrank, johansen) و R (urca, vars, ardl).

منابع

۱. نوفرستی، محمد (۱۳۹۸). ریشه واحد و هم‌جمعی در اقتصادسنجی، تهران: رسا، فصل ۸: «روش‌های جایگزین هم‌جمعی»، صص ۲۳۵–۲۶۵.

۲. سوری، علی (۱۴۰۲). اقتصادسنجی پیشرفته، تهران: فرهنگ مهران، فصل ۱۰: «مقایسه روش‌های هم‌جمعی»، صص ۴۱۵–۴۴۵.

۳. Saikkonen, P. (۱۹۹۱). “Asymptotically Efficient Estimation of Cointegration Regressions”, Econometric Theory, ۷(۱), pp. ۱–۲۱.

٤. Stock, J. H., & Watson, M. W. (۱۹۹۳). “A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems”, Econometrica, ٦۱(٤), pp. ۷۸۳–۸۲۰.

٥. Pesaran, M. H., Shin, Y., & Smith, R. J. (۲۰۰۱). “Bounds Testing Approaches to the Analysis of Level Relationships”, Journal of Applied Econometrics, ۱٦(۳), pp. ۲۸۹–۳۲٦.

منابع جامع جزوه اقتصادسنجی

۱. **ابریشمی، حمید** (۱۴۰۱). مبانی اقتصادسنجی (تألیف دامتار گجراتی)، جلد اول و دوم، تهران: انتشارات دانشگاه تهران.

۲. **سوری، علی** (۱۴۰۲). اقتصادسنجی، جلد اول: مفاهیم پایه، تهران: فرهنگ مهران.

۳. **سوری، علی** (۱۴۰۲). اقتصادسنجی پیشرفته: مباحثی در سری زمانی، تهران: فرهنگ مهران.

۴. **نوفرستی، محمد** (۱۳۹۸). ریشه واحد و هم‌جمعی در اقتصادسنجی، تهران: انتشارات رسا.

۵. **مهرگان، نادر** (۱۳۹۸). اقتصادسنجی سری‌های زمانی، تهران: انتشارات برادران.

۶. **محمدپور، احمد** (۱۳۹۷). مدل‌سازی معادلات ساختاری در پژوهش‌های اجتماعی، تهران: جامعه‌شناسان.

١. Banerjee, A., Dolado, J., Galbraith, J. W & ,.Hendry, D. F .(١٩٩٣) .Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data .Oxford University Press.

٢. Belsley, D. A., Kuh, E & ,.Welsch, R. E .(١٩٨٠) .Regression Diagnostics: Identifying Influential Data and Sources of Collinearity .New York: Wiley.

٣. Dickey, D. A & ,.Fuller, W. A“ .(١٩٧٩) .Distribution of the Estimators for Autoregressive Time Series with a Unit Root ,”Journal of the American Statistical Association ,(٣٦٦)٧٤ ,pp. ٤٢٧–.٤٣١

٤. Enders, W .(٢٠١٤) .Applied Econometric Time Series٤ ,th ed., Wiley.

٥. Engle, R. F & ,.Granger, C. W. J“ .(١٩٨٧) .Co-integration and Error Correction: Representation, Estimation, and Testing ,”Econometrica ,(٢)٥٥ ,pp. ٢٥١–.٢٧٦

٦. Frisch, R .(١٩٣٤) .Statistical Confluence Analysis by Means of Complete Regression Systems .Oslo: University Institute of Economics.

٧. Greene, W. H .(٢٠١٨) .Econometric Analysis٨ ,th ed., Pearson.

٨. Gujarati, D. N & ,.Porter, D. C .(٢٠٠٩) .Basic Econometrics٥ ,th ed., McGraw-Hill.

٩. Hamilton, J. D .(١٩٩٤) .Time Series Analysis .Princeton University Press.

١٠. Hoerl, A. E & ,.Kennard, R. W“ .(١٩٧٠) .Ridge Regression: Biased Estimation for Nonorthogonal Problems ,”Technometrics ,(١)١٢ ,pp. ٥٥–.٦٧

١١. James, G., Witten, D., Hastie, T & ,.Tibshirani, R .(٢٠٢١) .An Introduction to Statistical Learning٢ ,nd ed., Springer.

١٢. Johansen, S“ .(١٩٨٨) .Statistical Analysis of Cointegration Vectors ,”Journal of Economic Dynamics and Control ,(٣–٢)١٢ ,pp. ٢٣١–.٢٥٤

١٣. Johansen, S“ .(١٩٩١) .Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models ,”Econometrica ,(٦)٥٩ ,pp. ١٥٥١–.١٥٨٠

١٤. Osterwald-Lenum, M“ .(١٩٩٢) .A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics ,”Oxford Bulletin of Economics and Statistics ,(٣)٥٤ ,pp. ٤٦١–.٤٧٢

١٥. Pesaran, M. H., Shin, Y & ,.Smith, R. J“ .(٢٠٠١) .Bounds Testing Approaches to the Analysis of Level Relationships ,”Journal of Applied Econometrics ,(٣)١٦ ,pp. ٢٨٩–.٣٢٦

١٦. Phillips, P. C. B & ,.Perron, P“ .(١٩٨٨) .Testing for a Unit Root in Time Series Regression ,” Biometrika ,(٢)٧٥ ,pp. ٣٣٥–.٣٤٦

١٧. Saikkonen, P“ .(١٩٩١) .Asymptotically Efficient Estimation of Cointegration Regressions ,” Econometric Theory ,(١)٧ ,pp. ١–.٢١

١٨. Stock, J. H & ,.Watson, M. W“ .(١٩٩٣) .A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems ,”Econometrica ,(٤)٦١ ,pp. ٧٨٣–.٨٢٠

١٩. Stock, J. H & ,.Watson, M. W .(٢٠١٩) .Introduction to Econometrics٤ ,th ed., Pearson.

٢٠. Wooldridge, J. M .(٢٠٢٠) .Introductory Econometrics: A Modern Approach٧ ,th ed., Cengage

فرمول های به کار رفته در جزوه

۱. مدل رگرسیون ساده و پیش‌بینی (صفحات ۵ تا ۲۵)

- **تابع رگرسیون جامعه (PRF):**

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

- **تابع رگرسیون نمونه (SRF):**

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$$

- **محاسبه عرض از مبدأ و شیب در OLS:**

$$\hat{\beta}_1 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2},\ \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

- **رابطه گذشتن خط رگرسیون از نقطه میانگین:**

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

- **پیش‌بینی مقدار آتی:**

$$\hat{Y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{T+1}$$

۲. ضریب تعیین، تحلیل واریانس (ANOVA) و آزمون‌ها (صفحات ۳۲ تا ۴۳)

- **تجزیه مجموع مربعات (TSS = ESS + RSS):**

$$\sum (Y_i - \overline{Y})^2 = \sum (\hat{Y}_i - \overline{Y})^2 + \sum e_i^2$$

- **ضریب تعیین (R^2):**

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

- **ضریب تعیین تعدیل‌شده ($\bar{R}^2$):**

$$\bar{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)}$$

- **آماره آزمون t برای ضرایب:**

$$t = \frac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)}$$

- **آماره آزمون F (معناداری کلی مدل):**

$$F = \frac{(TSS - RSS)/k}{RSS/(n-k-1)}$$

۳. اقتصادسنجی ماتریسی و برآوردگر OLS (صفحات ۴۹ تا ۵۹)

- **مدل رگرسیون چندمتغیره در فرم ماتریسی:**

$$Y = X\beta + u$$

- **برآوردگر حداقل مربعات معمولی (OLS):**

$$\hat{\beta} = (X'X)^{-1}X'Y$$

- **واریانس-کوواریانس برآوردگر OLS:**

$$Var(\hat{\beta} \mid X) = \sigma^2(X'X)^{-1}$$

- **ماتریس‌های تصویر، پسماندساز و قضیه فریش–ووی–لاول (FWL):**

$$M = I - X(X'X)^{-1}X',\ P = X(X'X)^{-1}X'$$

- **آزمون والد (Wald Test) برای قیود خطی:**

$$W = (R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)$$

۴. متغیرهای مجازی و شکست ساختاری (صفحات ۶۱ تا ۶۹)

- **قید جمع متغیرهای مجازی (جلوگیری از تله متغیر مجازی):**

$$\sum_{j=1}^{m} D_{ji} = 1$$

- **مدل‌های تعاملی شیب و عرض از مبدأ:**

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (D_i \cdot X_i) + u_i$$

- **آماره آزمون چاو (Chow Test) برای شکست ساختاری:**

$$F = \frac{\left(RSS_R - (RSS_1 + RSS_2)\right)/k}{(RSS_1 + RSS_2)/(n_1 + n_2 - 2k)}$$

۵. سری‌های زمانی، وقفه توزیع‌شده و هم‌جمعی (صفحات ۹۵ تا ۱۱۰)

- **مدل وقفه توزیع‌شده و ضریب بلندمدت (LRP):**

$$Y_i = \alpha + \sum_{j=0}^{q} \beta_j X_{t-j} + u_t,\ LRP = \sum_{j=0}^{q} \beta_j$$

- **مدل AR(۱) و آزمون ریشه واحد دیکی–فولر (DF / ADF):**

$$\Delta Y_t = \rho Y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta Y_{t-i} + u_t$$

- **مدل تصحیح خطای انگل–گرنجر (EG-ECM):**

$$\Delta e_t = \gamma e_{t-1} + \sum \delta_i \Delta e_{t-i} + v_t$$

- **رویکرد جوهانسون و ماتریس Π:**

$$\Pi = \alpha \beta'$$

- **آماره اثر (Trace Statistic) در آزمون جوهانسون:**

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{k} \ln\left(1 - \hat{\lambda}_i\right)$$

- **آماره حداکثر مقدار ویژه (Maximum Eigenvalue):**

$$\lambda_{max}(r, r+1) = -T\ln\left(1 - \hat{\lambda}_{r+1}\right)$$

نمونه فرمول ها

مسئله اول: برآورد رگرسیون ساده و پیش‌بینی (مشابه مباحث فصل‌های اولیه جزوه)

صورت مسئله:

فرض کنید داده‌های زیر برای یک نمونه ۵ تایی از متغیرهای درآمد (X) و مصرف (Y) در دست است:

$$\sum X_i = ۳۰,\ \sum Y_i = ۴۰,\ \sum X_i^۲ = ۲۲۰,\ \sum Y_i^۲ = ۳۵۰,\ \sum X_i Y_i = ۲۷۰,\ n = ۵$$

الف) ضرایب خط رگرسیون نمونه ($\hat{\beta}_۰$ و خط $\hat{\beta}_۱$) را محاسبه کنید.

ب) اگر درآمد جامعه در دوره آتی برابر ۷ شود ($X_۶ = ۷$)، مقدار پیش‌بینی‌شده مصرف ($\hat{Y}_۶$) چقدر خواهد بود؟

حل گام‌به‌گام:

۱. ابتدا میانگین متغیرها را حساب می‌کنیم:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{۳۰}{۵} = ۶,\ \bar{Y} = \frac{\sum Y_i}{n} = \frac{۴۰}{۵} = ۸$$

۲. محاسبه ضریب شیب ($\hat{\beta}_۱$):

$$\hat{\beta}_۱ = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n\sum X_i^۲ - (\sum X_i)^۲} = \frac{۵(۲۷۰) - (۳۰)(۴۰)}{۵(۲۲۰) - (۳۰)^۲} = \frac{۱۳۵۰ - ۱۲۰۰}{۱۱۰۰ - ۹۰۰} = \frac{۱۵۰}{۲۰۰} = ۰.۷۵$$

۳. محاسبه عرض از مبدأ ($\hat{\beta}_۰$):

$$\hat{\beta}_۰ = \bar{Y} - \hat{\beta}_۱\bar{X} = ۸ - (۰.۷۵)(۶) = ۸ - ۴.۵ = ۳.۵$$

بنابراین تابع رگرسیون نمونه عبارت است از:

$$\hat{Y}_i = ۳.۵ + ۰.۷۵X_i$$

۴. پیش‌بینی برای $X_۶ = ۷$:

$$\hat{Y}_۶ = ۳.۵ + ۰.۷۵(۷) = ۳.۵ + ۵.۲۵ = ۸.۷۵$$

مسئله دوم: تحلیل واریانس (ANOVA) و ضریب تعیین R^2

صورت مسئله:

در یک مدل رگرسیون با ۳ متغیر توضیحی ($k = ۳$) و حجم نمونه ۲۵ تایی ($n = ۲۵$)، مجموع مربعات کل (TSS) برابر ۱۰۰۰ و مجموع مربعات پسماند (RSS) برابر ۳۰۰ به‌دست آمده است.

الف) مقدار ضریب تعیین ($R^۲$) و ضریب تعیین تعدیل‌شده ($\bar{R}^۲$) را حساب کنید.

ب) آماره آزمون Fبرای معناداری کلی مدل را محاسبه کنید.

حل گام‌به‌گام:

الف) محاسبه $R^۲$و $\bar{R}^۲$:

$$ESS = TSS - RSS = ۱۰۰۰ - ۳۰۰ = ۷۰۰$$

$$R^۲ = \frac{ESS}{TSS} = \frac{۷۰۰}{۱۰۰۰} = ۰.۷۰ \ (۷۰٪)$$

$$\bar{R}^۲ = ۱ - \frac{RSS/(n-k-۱)}{TSS/(n-۱)} = ۱ - \frac{۳۰۰/(۲۵-۳-۱)}{۱۰۰۰/(۲۴)} = ۱ - \frac{۳۰۰/۲۱}{۱۰۰۰/۲۴} = ۱ - \frac{۱۴.۲۸۵}{۴۱.۶۶۷} \approx ۰.۶۵۷$$

ب) محاسبه آماره F:

$$F = \frac{ESS/k}{RSS/(n-k-۱)} = \frac{۷۰۰/۳}{۳۰۰/۲۱} = \frac{۲۳۳.۳۳}{۱۴.۲۸۵} \approx ۱۶.۳۳$$

مسئله سوم: اقتصادسنجی ماتریسی (محاسبه OLS)

صورت مسئله:

بردار داده‌های متغیر وابسته Yو ماتریس متغیرهای توضیحی Xبه شکل زیر داده شده‌اند:

$$Y = \begin{bmatrix} ۵ \\ ۷ \\ ۹ \end{bmatrix}, \ X = \begin{bmatrix} ۱ & ۱ \\ ۱ & ۲ \\ ۱ & ۳ \end{bmatrix}$$

برآوردگر حداقل مربعات معمولی ($\hat{\beta} = (X'X)^{-1}X'Y$) را محاسبه کنید.

حل گام‌به‌گام:

۱. محاسبه ماتریس ترانهاده X':

$$X' = \begin{bmatrix} ۱ & ۱ & ۱ \\ ۱ & ۲ & ۳ \end{bmatrix}$$

۲. محاسبه حاصل‌ضرب $X'X$:

$$X'X = \begin{bmatrix} ۱ & ۱ & ۱ \\ ۱ & ۲ & ۳ \end{bmatrix} \begin{bmatrix} ۱ & ۱ \\ ۱ & ۲ \\ ۱ & ۳ \end{bmatrix} = \begin{bmatrix} ۳ & ۶ \\ ۶ & ۱۴ \end{bmatrix}$$

۳. محاسبه معکوس ماتریس $(X'X)^{-1}$(دترمینان $۶ = ۳۶ - ۴۲ = (۶)(۶) - (۳)(۱۴) =$):

$$(X'X)^{-1} = \frac{۱}{۶} \begin{bmatrix} ۱۴ & -۶ \\ -۶ & ۳ \end{bmatrix} = \begin{bmatrix} ۱۴/۶ & -۱ \\ -۱ & ۳/۶ \end{bmatrix} = \begin{bmatrix} ۲.۳۳۳ & -۱ \\ -۱ & ۰.۵ \end{bmatrix}$$

۴. محاسبه حاصل‌ضرب $X'Y$:

$$X'Y = \begin{bmatrix} ۱ & ۱ & ۱ \\ ۱ & ۲ & ۳ \end{bmatrix} \begin{bmatrix} ۵ \\ ۷ \\ ۹ \end{bmatrix} = \begin{bmatrix} ۵ + ۷ + ۹ \\ ۵ + ۱۴ + ۲۷ \end{bmatrix} = \begin{bmatrix} ۲۱ \\ ۴۶ \end{bmatrix}$$

۵. محاسبه نهایی $\hat{\beta}$:

$$\hat{\beta} = \begin{bmatrix} ۲.۳۳۳ & -۱ \\ -۱ & ۰.۵ \end{bmatrix} \begin{bmatrix} ۲۱ \\ ۴۶ \end{bmatrix} = \begin{bmatrix} (۲.۳۳۳)(۲۱) + (-۱)(۴۶) \\ (-۱)(۲۱) + (۰.۵)(۴۶) \end{bmatrix} = \begin{bmatrix} ۴۹ - ۴۶ \\ -۲۱ + ۲۳ \end{bmatrix} = \begin{bmatrix} ۳ \\ ۲ \end{bmatrix}$$

نتیجه: $\hat{\beta}_۰ = ۳$ و $\hat{\beta}_۱ = ۲$(خط رگرسیون: $Y_i = ۳ + ۲X_i$).

مسئله چهارم: متغیرهای مجازی و آزمون تفاوت شیب و عرض از مبدأ

صورت مسئله:

مدل زیر برای بررسی تأثیر جنسیت (D که برای مرد ۱ و برای زن صفر است) بر درآمد (Y) بر حسب سال‌های تحصیل (X) تخمین زده شده است:

$$Y_i = ۲۰ + ۳X_i + ۵D_i + ۱.۵(D_i \cdot X_i)$$

الف) خط درآمد برای زنان و مردان را به تفکیک بنویسید.

ب) تفسیر ضریب متغیر تعاملی (۱.۵) چیست؟

حل گام‌به‌گام:

الف) تفکیک مدل برای هر گروه:

- **برای زنان ($D_i = ۰$):**

$$Y_i = ۲۰ + ۳X_i + ۵(۰) + ۱.۵(۰ \cdot X_i) \implies Y_i = ۲۰ + ۳X_i$$

- **برای مردان ($D_i = ۱$):**

$$Y_i = ۲۰ + ۳X_i + ۵(۱) + ۱.۵(۱ \cdot X_i) = (۲۰ + ۵) + (۳ + ۱.۵)X_i \implies Y_i = ۲۵ + ۴.۵X_i$$

ب) **تفسیر ضریب تعاملی (۱.۵):** این ضریب نشان‌دهنده **اختلاف شیب** خط درآمد-تحصیل مردان نسبت به زنان است؛ یعنی با افزایش هر یک سال تحصیل، شیب رشد درآمد مردان ۵/۱ واحد بیشتر از زنان است (عرض از مبدأ مردان نیز ۵ واحد بالاتر از زنان است).

مسئله پنجم: آزمون ریشه واحد (دیکی–فولر) و هم‌جمعی

صورت مسئله:

معادله زیر برای آزمون ریشه واحد روی متغیر تولید ناخالص داخلی (Y_t) تخمین زده شده است:

$$\Delta Y_t = ۰.۴ - ۰.۱۵Y_{t-۱} + ۰.۲۵\Delta Y_{t-۱},\ t = ۵۰$$

مقاره آماره tبرای ضریب $Y_{t-۱}$برابر **۳.۲**–به‌دست آمده است. با فرض اینکه مقدار بحرانی مک‌کینون در سطح ۵٪ برابر **۲.۹۳**–باشد، نتیجه آزمون ایستایی را تحلیل کنید.

حل گام‌به‌گام:

۱. فرضیه‌های آزمون دیکی–فولر:

- $H_0 : \rho = 0$ (متغیر دارای ریشه واحد است، یعنی ناایستا یا $I(1)$ است).
- $H_1 : \rho < 0$ (متغیر ایستا یا $I(0)$ است).

۲. مقایسه آماره تذکر داده‌شده با مقدار بحرانی:

- آماره محاسبه‌شده ($t = -3.2$) کوچکتر (منفی‌تر) از مقدار بحرانی مطلق ($\text{Critical Value} = -2.93$) است؛ بنابراین در ناحیه بحرانی قرار دارد.
- **نتیجه‌گیری:** فرض صفر (H_0) مبنی بر وجود ریشه واحد **رد می‌شود**. متغیر Y_t در سطح ایستا یعنی $I(0)$ است.