

نقض فروض رگرسیون

Relaxing the Assumption of the Classical Model

مدل کلاسیک رگرسیون خطی نرمال بر پایه تعدادی فروض بنیادی استوار است که در نتیجه نقض هر یک از آنها، نتایج به دست آمده از رگرسیون گمراه کننده خواهد بود. از جمله این فروض عبارتند از: (۱) میانگین شرطی جزء اخلاص جامعه (u_i)، صفر است. (۲) واریانس شرطی u_i ثابت یا همسان است. (۳) خودهمبستگی در اجزاء اخلاص وجود ندارد. (۴) متغیرهای توضیحی غیرتصادفی اند و میان آنها هم خطی مرکب وجود ندارد. (۵) اجزاء اخلاص جامعه دارای توزیع نرمال با میانگین صفر و واریانس ثابت است. (۶) مدل رگرسیون به طور صحیح تصریح شده است، به این معنی که تورش تصریح وجود ندارد. (۷) تعداد مشاهدات باید از تعداد متغیرهای توضیحی بیشتر بوده و به اندازه کافی دارای تغییرپذیری باشند.

• بنابراین با این فروض، تخمین زن های حداقل مربعات معمولی ضرایب رگرسیون (Ordinary least squares)، بهترین تخمین زن های بدون تورش خطی و در اصطلاح BLUE (Best Linear Unbiased Estimator) هستند.

به طوری که با فرض نرمال بودن، امکان محاسبه صحیح تخمین زن های فاصله ای و آزمون فرضیه درست بودن ضرایب رگرسیون جامعه، فراهم می شود. در این زمینه لازم به ذکر است، ممکن است به دلایل مختلفی، یک یا چند فرض از فروض فوق تأمین نشود، لذا در ادامه به بررسی پیامدها و عواقب نقض فروض اساسی خودهمبستگی، واریانس ناهمسانی، هم خطی و خطای تصریح و همچنین آزمون های شناسایی آنها پرداخته می شود.

خودهمبستگی (Auto- Correlation)

خودهمبستگی به معنی وجود همبستگی میان جملات اخلاص در طول دوره های زمانی است. به عبارت دیگر، اگر جزء

اخلاص مربوط به مشاهده ای تحت تأثیر جزء اخلاص مشاهده دیگری قرار گیرد، دال بر خودهمبستگی است. در این حالت، کواریانس بین u_t و وقفه های آن به علت وجود رابطه میان آنها صفر نیست.

خودهمبستگی میان جملات اخلاص دارای اشکال گوناگونی از جمله فرآیند اتورگرسیو (Autoregressive) مرتبه اول AR(1) و مراتب بالاتر آن AR(k)، فرآیند میانگین متحرک (Moving-Average) مرتبه اول MA(1) و مراتب بالاتر آن MA(r) و همچنین تلفیقی از فرآیندهای اتورگرسیو و میانگین متحرک ARIMA است (کیانی، ۱۳۹۴: ۲۰۰).

دلایل ایجاد خودهمبستگی

از جمله عوامل مؤثر در ایجاد خودهمبستگی عبارتند از: فرم نادرست معادله رگرسیون، اثر نوسانات ادواری در متغیرهای اقتصادی، پدیده تار عنکبوتی (Phenomenon Cobweb)، تغییر و تبدیل داده های آماری، سکون (Inertia) یا وابستگی مشاهدات پی در پی داده های سری زمانی، نامانایی سری زمانی، وقفه ها و مدل های خود رگرسیونی (Autoregressive) که در آنها حداقل یکی از متغیرهای توضیحی، وقفه های متغیر وابسته است (کیانی، ۱۳۹۴: ۱۹۹-۲۰۰ و Gujarati, 2004: 443-448).

عواقب ناشی از خودهمبستگی

در حالت وجود خودهمبستگی برآوردکننده های OLS، با وجود اینکه خطی، بدون تورش و سازگارند اما کارا نیستند. به عبارت دیگر، واریانس این برآوردکننده ها حداقل نیستند. بنابراین، پیامدهای استفاده از تخمین زن های حداقل مربعات معمولی، در مواقع عدم استقلال جملات اخلاص، شامل موارد (۱) طولانی بودن فاصله اطمینان و کاهش قدرت آزمون های معنی داری. (۲) برآورد کمتر از حد واریانس واقعی σ^2 و برآورد بیش از حد R^2 ، (۳) نتایج غلط در معنی داری ضرایب به علت بی اعتباری آزمون ها t و F ، (۴) حساس بودن برآوردکننده های OLS به تغییرات در مقادیر نمونه است

(سوری، ۱۳۹۷: ۳۱۲؛ عرب-مازار، ۱۳۶۹: ۲۵۰-۲۴۹ و (Gujarati, 2004: 455).

کشف و آزمون خودهمبستگی

از جمله مهم‌ترین روش‌های تشخیص خودهمبستگی میان جملات خطا می‌توان به روش ترسیمی (رسم مقادیر e_t در مقابل e_{t-1})، آزمون دوربین-واتسون (Durbin-watson)، آزمون بروش-گادفری (Breush-Godfery)، آزمون Q، آزمون اتورگرسیون در داده‌های فصلی-آزمون والیز، آزمون نسبت فون-نیومن، آزمون برنبلوت-وب، آزمون h دوربین، آزمون LM، اشاره کرد (صادقی‌شاهدانی و اشرف‌زاده، ۱۳۹۶: ۵۰۸-۴۹۲؛ کیانی، ۱۳۹۴: ۲۰۸-۲۲؛ حقیقت و موسوی، ۱۳۹۶: ۶۹-۷۴؛ Ljung & box, 1978: 267-303؛ waliss, 1972: 617-636).

واریانس ناهمسانی (Heteroscedasticity)

واریانس ناهمسانی زمانی رخ می‌دهد که همه اجزای اخلال u_i که در تابع رگرسیون ظاهر می‌شود، دارای واریانس یکسان و برابر عدد ثابت σ^2 نباشند. گفتنی است این نقض فرضیه برای واریانس‌های سری‌های مقطع عرضی نسبت به سری زمانی رایج‌تر است.

دلایل ایجاد واریانس ناهمسانی

دلایل متعددی برای وجود واریانس‌های ناهمسان اجزای اخلال u_i وجود دارد که از جمله از آن‌ها عبارتند از: (۱) الگوهای یادگیری-خطا (Error-Learning Models) که به معنی کاهش σ^2 در طول زمان ناشی از کاهش خطا در رفتار افراد به دلیل آموزش و یادگیری است. (۲) گسترش ابعاد انتخاب مصرفی افراد ناشی از افزایش درآمد که منجر به افزایش σ^2 خواهد شد. (۳) بهبود روش‌های گردآوری

داده‌ها که منتج به کاهش σ^2 می‌شود. (۴) داده‌های پرت (Outliers Data). (۵) تصریح و فرم نادرست الگو. (۶) چولگی در توزیع یک یا چند متغیر توضیحی موجود در الگو. (۷) تبدیل نادرست داده‌ها (Gujarati, 2004: 389-391).

عواقب ناشی از واریانس ناهمسانی

وجود واریانس ناهمسانی برآوردکننده‌های OLS را غیرکارا می‌کند، به این معنی که دیگر دارای حداقل واریانس نیستند، اما همچنان بدون تورش و به طور کلی سازگارند. در این شرایط، با بزرگ شدن فاصله اطمینان، آزمون فرضیه‌هایی که با F ، t و χ^2 انجام می‌شود، قابل اعتماد نیست و در نتیجه ضرایب از لحاظ آماری، معنی‌دار نخواهند بود (سوری، ۱۳۹۷: ۲۹۲؛ عباسی‌نژاد، ۱۳۹۵: ۱۰۴؛ Gujarati, 2004: 398).

کشف و آزمون واریانس ناهمسانی

آزمون‌های مختلفی برای تشخیص واریانس ناهمسانی وجود دارد. از جمله از مهم‌ترین آن‌ها می‌توان به روش ترسیمی، آزمون پارک (park)، آزمون وایت، آزمون LM (Lagrange Multiple Test)، آزمون گلدفلد-کوانت (Goldfeld-Quant)، آزمون گلچسر (Glesjer)، آزمون بروش-پاگان، آزمون اسپیرمن (spearman's rank correlation test)، آزمون هاروی (Harvey)، آزمون بارتلت و غیره اشاره کرد (سوری، ۱۳۹۷: ۲۹۶-۳۰۲؛ کیانی، ۱۳۹۴: ۲۳۳-۲۴۱؛ حقیقت و موسوی، ۱۳۹۶: ۶۶-۶۸؛ johonston, 1984: 301؛ White, 1980: 817-838؛ Bresch & Pagan, 1979: 1287-1294).

هم‌خطی (Multicollinearity)

بودن ضریب تعیین (R^2) الگو با وجود ضرایب اندک معنی‌دار (Gujarati, 2004: 350).

کشف و آزمون هم‌خطی

با توجه به اینکه در اکثر موارد درجه‌ای از هم‌خطی وجود دارد، برخلاف فرض‌های اساسی دیگر که آزمون وجود یا عدم وجود مشکل مطرح است، درخصوص پیامد هم‌خطی، باید حاد بودن یا نبودن آن تشخیص داده شود. از جمله مهم‌ترین معیارهای تشخیص هم‌خطی حاد می‌توان به محاسبه R^2 حذفی (بر اساس تعیین تفاوت ضریب تعیین معادله قبل و بعد از حذف یک متغیر مستقل)، روش فارا-گلور (Farrar-Glauber) بر اساس آزمون ضرایب همبستگی جزئی، محاسبه دترمینان ماتریس xx' مقایسه آماره‌های t و F (عدم معنی‌داری ضرایب ناشی از کوچک بودن آماره t و در عین حال معنی‌داری F محاسبه‌شده)، معیار رگرسیون معین و متغیر وابسته قرار دادن متغیرهای مستقل، معیار شاخص وضعیت (Condition Index) و مقادیر ویژه (Eigenvalues)، بررسی ضریب همبستگی مرتبه صفر میان دو متغیر و بررسی عامل تورم واریانس (Variance Inflation Factor) اشاره کرد (کیانی، ۱۳۹۴: ۲۸۱-۲۸۴؛ عباسی‌نژاد، ۱۳۹۵: ۱۱۷-۱۲۱).

در زمینه آزمون‌ها، باید به این نکته اشاره شود که روش‌های مطرح‌شده دارای کاربرد در هر حالتی نیستند. همچنین، با توجه به اینکه هم‌خطی از جمله مشکلات نمونه است، لذا پژوهشگر نمی‌تواند آن را کنترل کند. بنابراین، قاعده مشخص و دقیقی برای چاره‌جویی مشکل در این زمینه وجود نخواهد داشت.

اقدامات رفع هم‌خطی

اصطلاح هم‌خطی به معنای وجود رابطه کامل یا دقیق خطی میان همه یا برخی از متغیرهای توضیحی الگوی رگرسیون است (Gujarati, 2004: 246). در شرایط هم‌خطی، به علت وجود روابط متقابل و قوی میان متغیرهای مستقل، تفکیک اثر مجزای آن‌ها بر روی متغیر وابسته امکان‌پذیر نیست. هم‌خطی هم شامل ارتباط خطی کامل و هم شامل حالتی است که متغیرهای توضیحی به صورت ناقص همبسته هستند. در این زمینه گفتنی است اگر هدف از تحلیل رگرسیون پیش‌بینی باشد، وجود هم‌خطی به عنوان مشکل مهمی مطرح نخواهد بود. این در حالی است که اگر هدف از برآورد الگو، تخمین و تحلیل دقیقی از ضرایب باشد، هم‌خطی با توجه به بزرگی انحراف معیارهای برآوردگرها، مشکل اساسی محسوب می‌شود.

عواقب ناشی از هم‌خطی

اگرچه هم‌خطی میان متغیرهای مستقل منجر به نقض هیچ یک از فروض کلاسیک رگرسیون نمی‌شود، اما در هم‌خطی کامل یکی از فروض اساسی برآورد مدل رگرسیون، یعنی کامل بودن مرتبه ماتریس xx' ، نقض می‌شود. در چنین شرایطی برآورد پارامترهای الگو امکان‌پذیر نخواهد بود (عباسی‌نژاد، ۱۳۹۵: ۱۱۳). بر این اساس، از جمله مهم‌ترین پیامدهای هم‌خطی عبارتند از: (۱) بزرگی واریانس کواریانس برآوردکننده‌های OLS، که در نتیجه آن برآورد دقیق الگو را دشوار می‌سازد، (۲) بزرگی انحراف معیار که منجر به بزرگتر شدن فواصل اطمینان برای پارامترهای جامعه و افزایش احتمال پذیرش یک فرضیه غلط می‌شود، (۳) کوچک شدن آماره t و عدم معنی‌داری آماری یک یا چند پارامتر، (۴) حساس بودن برآوردکننده‌های OLS و خطای معیار آن‌ها در مقایسه با تغییرات کوچک در داده‌ها، (۵) بالا

اگرچه هیچ دستورالعمل قطعی برای رفع مشکل هم خطی وجود ندارد، اما برخی قواعد تجربی با توجه به شدت همبستگی می‌تواند مشکلات وابستگی ناشی از آن را رفع کند. در این زمینه استفاده از دیگر منابع اطلاعاتی مبتنی بر نظریه‌های اقتصادی و مطالعات تجربی دیگران در الگو، ترکیب داده‌های سری زمانی و مقطع عرضی، حذف متغیر، تغییر و تبدیل متغیرها، استفاده از اطلاعات اضافی، جدید و افزایش حجم نمونه، استفاده از نسبت‌ها یا تفاضل مرتبه اول و متغیرهای توضیح‌دهنده تصادفی می‌تواند راهگشا باشد (Gujarati, 2004: 364-368؛ کیانی، ۱۳۹۴: ۲۸۴-۲۸۹).

خطای تصریح (Specification Error)

یکی دیگر از فروض کلاسیک این است که فرم تابع به درستی انتخاب شده و خطای تصریح در الگو وجود نداشته باشد. اما در عمل ممکن است به دلایل مختلفی مانند خطا در انتخاب و اندازه‌گیری متغیرها، خطا در تصریح فرم تابع، حذف متغیرهای اثرگذار و سایر موارد که در ادامه به آن پرداخته می‌شود، این فرض نقض شده و تورش تصریح رخ دهد. قابل ذکر است که این خطا می‌تواند ناشی از آن باشد که روابط واقعی میان متغیرها شناخته شده نیست.

در این زمینه ملاک‌های متعددی برای قضاوت راجع به کیفیت و خصوصیات یک الگوی خوب وجود دارد که از جمله می‌توان به قابل قبول بودن داده‌ها (به معنی دستیابی به نتایج و پیش‌بینی‌های منطقی ناشی از الگو)، سازگاری با تئوری، وجود کمترین همبستگی میان متغیرهای توضیحی و جزء اخلاص، ثبات پارامترهای برآوردی و قدرت پیش‌بینی، اشاره کرد (Gujarati, 2004: 507).

انواع خطای تصریح

در اکثر موارد خطاهای مربوط به مشخصات الگو به خطاهایی اطلاق می‌شود که در فرموله‌بندی معادله رگرسیون رخ می‌دهند. از جمله مهم‌ترین آن‌ها می‌توان به حذف متغیر توضیح‌دهنده اثرگذار و مناسب، لحاظ متغیر توضیح‌دهنده بی‌اثر و نامناسب، حذف متغیر مهم و وارد کردن متغیر اشتباه به صورت همزمان، تصریح غلط فرم تابعی رگرسیون، وجود همبستگی بین اجزای اخلاص و متغیرهای توضیحی، تصریح نادرست جمله اخلاص و خطای در اندازه‌گیری در متغیر وابسته و توضیحی اشاره کرد (کیانی، ۱۳۹۴: ۲۹۶؛ آستریو و جی هال، ۱۳۹۳: ۲۰۹-۲۱۱؛ محمدی و محمدزاده، ۱۳۸۹: ۱۲۰؛ Gujarati, 2004: 509).

عواقب ناشی از خطای تصریح

حذف متغیرهای مستقل که نقش مهمی در توضیح متغیر وابسته دارند، باعث می‌شود که این متغیرها به عنوان بخشی از جزء اخلاص جامعه محسوب شوند که در نتیجه آن، برآوردکننده‌های الگو اریب و ناسازگار خواهند شد. همچنین، واریانس خطا به صورت نادرست برآورد می‌شود. نتیجه دیگر آن است که پیش‌بینی‌های مبتنی بر الگو، نادرست و غیرقابل اتکا است. افزون بر آن، فواصل پیش‌بینی (اطمینان) غیرقابل اعتماد خواهد بود و آزمون فرضیه‌ها نیز از اعتبار لازم برخوردار نیستند. در مقابل اگر معادله برآوردشده شامل متغیرهای مؤثر نباشد، مشکل به وجود آمده چندان جدی نخواهد بود. به عبارت دیگر در حالت ورود متغیر نامناسب، برآوردکننده‌های OLS اگرچه ممکن است کارا نباشند (دارای واریانس بزرگ‌تر از واریانس پارامترهای مدل حقیقی)، اما ناریب و سازگار هستند. با این حال در شرایط حضور متغیرهای غیراثرگذار به علت افزایش درجه آزادی مقدار R^2 کاهش می‌یابد، در حالی که RSS تقریباً تغییراتی نخواهد داشت. همچنین، در این حالت، آماره t و

در بسیاری موارد، خطای تصریح ناشی از عدم فرموله‌بندی مناسب الگو یا عدم دسترسی به اطلاعات صحیح برای آزمون الگو است (Gujarati, 2004: 514). با این حال با توجه به خصوصیات یک الگوی مناسب، لازم است معیارهای کلی از جمله مقادیر آماره t محاسباتی، R^2 و سازگاری علامت‌های ضرایب برآوردشده با مبانی نظری موردتوجه قرار گیرد.

عمومی‌ترین آزمون‌های خطای تصریح عبارت از آزمون reset رمزی (reset test Ramsey)، نرمال بودن پسماندها، آزمون باکس-کاکس، بررسی و گراف پسماندها و آزمون دوربین-واتسون است (Gujarati, 2004: 518؛ آستریو و جی هال، ۱۳۹۳، ص. ۲۲۴-۲۳۱). همچنین، از مهم‌ترین آزمون‌های کشف خطای تصریح در زمینه تشخیص انتخاب فرم تابعی درست، می‌توان به آزمون رنگین کمان آتس (Utts rainbow test)، آزمون تفاضل‌گیری پلاس-شوارت- وایت (Plosser-Schwert- White)، رهیافت عام به خاص هندری (Henry general to specific approach) و رهیافت میانگین اقتصادی (regression Average Economic) اشاره کرد (آستریو و جی هال، ۱۳۹۳: ۲۳۴-۲۳۵). قابل ذکر است، آزمون‌های ثابت ضرایب نیز شامل آزمون نقطه شکست چاو (chow breakpoint test)، پیش‌بینی چاو (chow forecast test)، آزمون بازگشتی و روش پیش‌بینی یک قدمی، آزمون مجموع تجمعی خطاهای بازگشتی (CUSUM)، آزمون مجموع مجذور تجمعی خطاهای بازگشتی (CUSUMQ)، استفاده از متغیرهای مجازی است (سوری، ۱۳۹۷: ۳۹۲-۴۰۵).

کتاب‌شناسی

آستریو، د.، هالا، س.ج. (۱۳۹۳). *اقتصادسنجی کاربردی: رهیافتی مدرن با استفاده از EViews و Microfit*، ترجمه علیرضا کرباسی، حسین محمدی و آزاده تعالی مقدم، مترجم، مشهد: دانشگاه فردوسی مشهد.

معنی‌داری متغیرهای دیگر به صورت قابل توجهی تغییر نخواهند کرد، به عبارت دیگر پروسه‌های آزمون فرضیه و فاصله اعتماد، معتبر خواهند بود. در این خصوص لازم به ذکر است وجود متغیر غیرضروری که با یک یا چند متغیر توضیحی دیگر در الگو همبستگی بالایی داشته باشد، می‌تواند آماره t را تغییر دهد.

در مجموع، موارد ذکرشده مبتنی بر بهتر بودن ورود یک متغیر غیرمهم در برابر حذف متغیر مهم نیست. چراکه افزودن متغیر غیرلازم نه تنها منجر به کاهش کارایی تخمین‌زن‌ها می‌شود، بلکه احتمال رخداد عارضه هم‌خطی و کاهش درجه آزادی نیز افزایش خواهد داد (Gujarati, 2004: 510-514؛ آستریو و جی هال، ۱۳۹۳: ۲۱۰-۲۱۱). همچنین، اگر خطای اندازه‌گیری در متغیر وابسته با جزء اخلال ناهمبسته باشد، واریانس خطا بزرگ‌تر شده و در نتیجه واریانس ضرایب برآوردشده افزایش خواهد یافت. افزون بر آن، در شرایطی که خطای اندازه‌گیری متغیر وابسته با متغیرهای توضیحی ناهمبسته باشد، برآوردکننده‌های OLS برای ضرایب نارپ و ناسازگارند و برعکس. مضاف بر آن، اگر جزء اخلال و خطای اندازه‌گیری در متغیرهای توضیحی ناهمبسته باشد، خطای اندازه‌گیری واریانس را افزایش داده و برآوردکننده‌های OLS ارب‌دار و ناسازگار خواهند بود (آستریو و جی هال، ۱۳۹۳: ۲۲-۲۲۴).

فرم تبعی غلط در الگو نیز، مشکلاتی نظیر خودهمبستگی و واریانس ناهمسانی را ایجاد خواهد کرد. علاوه بر آن، تصریح نادرست جمله اخلال نیز می‌تواند منجر به برآورد ارب پارامترهای الگو شود (صادقی‌شاهدانی و اشرف‌زاده، ۱۳۹۶: ۲۸۷-۲۸۵).

آزمون کشف خطای تصریح

حقیقت، ج.، اکبرموسوی، ص. (۱۳۹۶). *اقتصادسنجی کاربردی*، همدان: نورعلم.

سوری، ع. (۱۳۹۷). *اقتصادسنجی ۱ مقدماتی همراه با کاربرد Eviews 8*، تهران: فرهنگ‌شناسی.

صادقی شاهدانی، م.، اشرف زاده، ع. (۱۳۹۶). *اقتصادسنجی عمومی*، تهران: دانشگاه امام صادق.

عرب مازار، ع. (۱۳۶۹). *اقتصادسنجی عمومی*. تهران: کویر.

عباسی نژاد، ح. (۱۳۹۵). *اقتصادسنجی (مبانی و روشها)*، تهران: دانشگاه تهران.

محمدی، ت.، محمدزاده، پ. (۱۳۸۹). *اقتصادسنجی*، تهران: ترمه.

هژیر کیانی، ک. (۱۳۹۴). *اقتصادسنجی و کاربرد آن*. تهران: جهاد دانشگاهی دانشگاه شهید بهشتی.

Breusch, T. S., Pagan, A. R. (1979). A Simple Test for Heteroskedasticity and Random Coefficient Variation, *Econometrica*, 47 (5): 1287–1294. doi:10.2307/1911963. JSTOR 1911963. MR 0545960.

Gujarati, D.N. (2004). *Basic Econometrics*. 4th Edition: McGraw-Hill Companies.

Johnston, J. (1984). *Econometric Methods*. 3rd Edition, New York: McGraw-Hill,

Ljung, G.M., Box, G.E.P. (19781). On a measure of lack of fit in time series models, *Biometrika*, 65, 297-303.

Wallis. K.F.(1972). *Introductory Econometrics*, London: Gray-Mills Publishing Ltd.

White H (1980). A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity, *Econometrica*, 48, 817–838.

سمانه عابدی

عضو هیئت علمی دانشکده اقتصاد، دانشگاه علامه طباطبائی