

آزمون برابری بردار ضرایب رگرسیونی مدل‌های پانلی در چندین گروه

احد ملک‌زاده^۱، اصغر اسمعیلی عیان^۲، سیدمهدی محمودی^۳

^۱گروه علوم کامپیوتر و آمار، دانشکده ریاضی، دانشگاه صنعتی خواجه نصیرالدین طوسی.

^۲گروه آمار، واحد اهواز، دانشگاه آزاد اسلامی، اهواز، ایران.

^۳دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه سمنان.

تاریخ دریافت: ۱۳۹۹/۰۱/۲۲ تاریخ پذیرش و انتشار: ۱۳۹۹/۰۶/۳۰

چکیده: مدل داده پانلی در بسیاری از شاخه‌های علمی همانند اقتصاد، علوم اجتماعی، پزشکی و اپیدمیولوژی مورد استفاده قرار می‌گیرد. در دهه‌های گذشته، استنباط روی ضرایب رگرسیون در مدل‌های داده پانلی توسعه یافته است. در این مقاله، به معرفی روش‌هایی به منظور انجام آزمون فرضیه برابری مدل پانلی در بین گروه‌های موجود در مجموعه داده‌های پانلی پرداخته می‌شود. ابتدا یک کمیت تصادفی معرفی می‌شود که توزیع آن را به دو روش تقریب و بوت استرپ پارامتری برآورد خواهد شد. همچنین یک کمیت محوری برای انجام این آزمون فرضیه معرفی می‌شود. در یک مطالعه شبیه‌سازی، رویکردهای پیشنهادی با روش موجود بر اساس خطای نوع اول و توان آزمون مورد مقایسه قرار می‌گیرد. همچنین مجموعه داده‌های پانل بنزینی با روش ارائه شده مورد تحلیل قرار می‌گیرد.

واژه‌های کلیدی: داده‌های پانلی، ضریب رگرسیونی، بوت استرپ پارامتری، روش دقیق.

۱ مقدمه

مجموعه داده‌های پانلی، مجموعه‌ای است که در دوره‌های زمانی مختلف، هر نمونه را بررسی کرده و بنابراین مشاهدات متعددی را در مورد هر عضو نمونه ارائه می‌نماید (شیائو، ۲۰۱۴). از جمله مثال‌های معروف از مطالعات داده‌های پانلی می‌توان به مطالعه پانلی پویایی درآمد خانوارهای آمریکا از مرکز تحقیقات دانشگاه میشیگان، پانل اقتصادی-اجتماعی خانوارهای آلمانی برای تحقیقات اقتصادی در دانشگاه اُهایو و درآمد خانوار و پویایی نیروی کار در استرالیا اشاره کرد. برای مشاهده مثال‌های بیشتر می‌توان به یک منبع از داده‌های پانلی (وب سایت دانشگاه نیویورک، ۲۰۰۴) مراجعه نمود. تحلیل مدل‌های رگرسیونی داده‌های پانلی، توسط محققان زیادی در علوم آمار و اقتصادسنجی از جمله بالسترا و نرلاو (۱۹۶۶)، نرلاو (۱۹۷۱)، فولر و بتس (۱۹۷۴)، بالتاجی (۱۹۸۱)، کریازیدو (۱۹۹۷)، کائو و چیانگ (۲۰۰۰)، فیلیپس و مون (۲۰۰۰)، آرلانو و هونور (۲۰۰۱)، بالتاجی و همکاران (۲۰۰۳)، کانای (۲۰۱۱) و منابع موجود در آنها مورد توجه قرار گرفته است. صادقی و کاظمی (۱۳۸۸) به بحث در خصوص استنباط آماری پارامترهای مدل رگرسیونی داده‌های پانلی با اثرات ثابت و تصادفی به وسیله روش‌های ماکسیمم درستنمایی و بیزی پرداخته‌اند.

ژائو (۲۰۰۷) بازه اطمینان تعمیم یافته و p -مقدار تعمیم یافته را برای انجام استنباط روی ضرایب رگرسیونی مدل داده‌های پانلی، شامل ایجاد فاصله اطمینان و آزمون فرضیه روی ترکیب خطی از ضرایب رگرسیونی تحت مدل‌های رگرسیونی مولفه خطای تک عاملی و دو عاملی به کار گرفته است. اخیراً بالتاجی (۲۰۱۳) در کتاب خود به مرور کلی بر تحلیل داده‌های پانلی در اقتصادسنجی پرداخته است. یو و همکاران (۲۰۱۷) رویکردی را برای آزمون فرضیه ترکیب ناهمگن از ضرایب رگرسیونی مدل‌های رگرسیونی مولفه خطای دو عاملی ارائه داده‌اند. زو و تیان (۲۰۱۷) با استفاده از رویکرد بوت استرپ پارامتری، آزمون فرضیه و برآورد فاصله‌ای ترکیب خطی ضرایب رگرسیونی و مولفه‌های واریانس اثرات تصادفی مدل داده پانلی متعادل تک عاملی و دو عاملی را ارائه نمودند. اسمعیلی و همکاران (۲۰۲۰) همچنین با به کارگیری رویکرد بوت استرپ پارامتری، به انجام آزمون فرضیه و ایجاد ناحیه اطمینان برای بردار ضرایب رگرسیونی مدل داده‌های پانلی پرداخته‌اند.

یافتن تفاوت بین گروه‌های مختلف در یک مجموعه داده پانلی از جمله مسائل مهمی است که تاکنون اشاره و توجه خاصی به آن در مقالات مرتبط با مدل‌های پانلی نشده است. به عنوان نمونه‌ای از کارهای انجام شده در این مورد می‌توان تنها به مقاله ژائو (۲۰۰۷) اشاره کرد که در آن با بهره‌گیری از روش p -مقدار تعمیم یافته به آزمون فرضیه برابری ضرایب رگرسیونی مدل داده‌های پانلی در چندین گروه پرداخته شده

است. انگیزه ما در این مقاله ارائه سه روش به منظور انجام آزمون فرضیه برابری ضرایب رگرسیونی مدل داده‌های پانلی در چندین گروه مختلف است که این روش‌ها دارای سه ویژگی الف- کنترل خطای نوع اول آزمون، ب- عدم پیچیدگی محاسباتی که منجر به سادگی آنها می‌شود و ج- توان بالای آزمون است. در این مقاله، سه روش ارائه شده عبارتند از روش تقریب، روش بوت استرپ پارامتری و روش آزمون دقیق که در دو روش اول با معرفی یک کمیت تصادفی با توزیع نامعلوم و پیچیده تلاش می‌شود که توزیع کمیت تصادفی تقریب زده شود. در روش سوم که همان روش آزمون دقیق است، یک کمیت محوری معرفی شده است. با یک مطالعه شبیه‌سازی، به مقایسه عملکرد این سه روش با روش ارائه شده توسط ژائو (۲۰۰۷) خواهیم پرداخت. همچنین عملکرد هریک از روش‌های بالا را در یک مثال واقعی نشان خواهیم داد. ساختار این مقاله به این شرح است: در بخش ۲ به معرفی و نمادگذاری مدل رگرسیونی داده‌های پانلی پرداخته می‌شود. روش‌ها و رویکردهای تقریبی و دقیق برای آزمون فرضیه برابری ضرایب رگرسیونی، زمانی که چندین مدل داده پانلی وجود دارد، در بخش ۳ بیان شده است. به منظور بررسی عملکرد روش‌های جدید با روش ارائه شده توسط ژائو (۲۰۰۷)، نتایج مطالعه عددی حاصل از یک شبیه‌سازی وسیع، در بخش ۴ آمده است. به نمونه‌ای از داده‌های واقعی در بخش ۵ به عنوان یک مثال کاربردی پرداخته می‌شود. نهایتاً نتایج در بخش بحث و نتیجه‌گیری آورده شده است.

۲ معرفی مدل پانلی و نمادگذاری

مدل رگرسیونی تک عاملی داده‌های پانلی به صورت

$$Y_{it} = \alpha + X'_{it}\beta + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad (1)$$

نمایش داده می‌شود، که در آن α یک اسکالر، β یک بردار K -بعدی مجهول، Y_{it} مقدار پاسخ و X_{it} بردار مشاهدات حاصل از K متغیر تبیینی مقطع i ام در زمان t ام هستند. در مدل رگرسیونی تک عاملی داده‌های پانلی که آن را مدل رگرسیونی با مؤلفه خطای یک عاملی نیز گویند، $u_{it} = \mu_i + \nu_{it}$ است که در آن μ_i اثر غیرقابل مشاهده هر واحد نمونه‌ای و ν_{it} جمله خطا است. فرم ماتریسی مدل (۱) به صورت

$$Y = \alpha \mathbf{1}_{NT} + X\beta + u = Z\delta + u \quad (2)$$

است، که در آن Y و بردارهای NT بعدی، X یک ماتریس $NT \times K$ ، $Z = [I_{NT}, X]$ ، $Z_\mu = u = Z_\mu \mu + \nu$ ، $\delta = (\alpha, \beta)'$ بردار ضرایب و ν یک بردار r بعدی با درایه‌های ν ، $I_N \otimes I_T$ (عملگر ضرب کرونگر)، I_s یک ماتریس همانی از مرتبه s ، $\mu = (\mu_1, \dots, \mu_N)'$ و $\nu = (\nu_{11}, \dots, \nu_{NT})'$ است. معمولاً فرض می‌شود ν_{it} ها، از یکدیگر مستقل و دارای توزیع نرمال با میانگین صفر و واریانس σ_ν^2 و μ_i ها از یکدیگر مستقل و دارای توزیع نرمال با میانگین صفر و واریانس σ_μ^2 و مستقل از ν_{it} ها هستند. به این ترتیب با استفاده از مدل (۱)، Y دارای توزیع نرمال با بردار میانگین $E(Y) = Z\delta$ و ماتریس کواریانس

$$\text{Cov}(Y) = \Sigma = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\nu^2(I_N \otimes I_T) = \sigma_\mu^2 P + \sigma_\nu^2 Q$$

است، که در آن J_T یک ماتریس $T \times T$ با درایه‌های ν ، $\sigma_\mu^2 = T\sigma_\mu^2 + \sigma_\nu^2$ ، $P = I_N \otimes \bar{J}_T$ ، $\sigma_\nu^2 = T\sigma_\mu^2 + \sigma_\nu^2$ ، $E_T = I_T - \bar{J}_T$ ، $\bar{J}_T = \frac{1}{T}J_T$ است که $Q = I_N \otimes E_T$ و $PQ = 0$. **وانسیک و کاپتین (۱۹۸۲)** با استفاده از ویژگی‌های P و Q نشان دادند

$$\Sigma^r = (\sigma_\mu^2)^r P + (\sigma_\nu^2)^r Q, \quad (۳)$$

که در آن r مقدار دلخواه عددی غیر صفر است. بنابراین $\Sigma^{-1} = (\sigma_\mu^2)^{-1} P + (\sigma_\nu^2)^{-1} Q$. برآوردگر کمترین توان‌های دوم تعمیم یافته δ توسط **سوامی و آورا (۱۹۷۲)** به صورت

$$\hat{\delta} = (Z' \Sigma^{-1} Z)^{-1} Z' \Sigma^{-1} Y. \quad (۴)$$

بیان شده است. به سادگی می‌توان نشان داد $(Z' \Sigma^{-1} Z)^{-1} \sim N(\delta, (Z' \Sigma^{-1} Z)^{-1})$. برای بدست آوردن برآوردگرهای σ_μ^2 و σ_ν^2 می‌توان مدل (۲) را به صورت

$$\begin{pmatrix} QY \\ PY \end{pmatrix} = \begin{pmatrix} QZ \\ PZ \end{pmatrix} \delta + \begin{pmatrix} Qu \\ Pu \end{pmatrix} = \begin{pmatrix} QX\beta \\ PZ\delta \end{pmatrix} + \begin{pmatrix} Q\nu \\ Pu \end{pmatrix}, \quad (۵)$$

نوشت، که در آن $QY \sim N(QX\beta, \sigma_\nu^2 Q)$ و $PY \sim N(PZ\delta, \sigma_\mu^2 P)$ به طوری که QY و PY متقابلاً مستقلند. بنابراین می‌توان تعریف کرد $SS_1 = Y'PY - Y'PZ(Z'PZ)^{-1}Z'PY$ و

$SS_{\nu} = Y'QY - Y'QX(X'QX)^{-1}X'QY$ و $\frac{SS_{\nu}}{\sigma_{\nu}^2}$ مستقلاً به ترتیب دارای توزیع‌های $\chi_{(N(T-1)-K)}^2$ و $\chi_{(N-K-1)}^2$ هستند، که $\chi_{(m)}^2$ متغیر تصادفی کای-دو مرکزی با m درجه آزادی را نشان می‌دهد. بنابراین برآوردهای نااریب σ_{ν}^2 و σ_{μ}^2 به ترتیب به صورت

$$\tilde{\sigma}_{\nu}^2 = \frac{SS_{\nu}}{N(T-1)-K}, \quad \tilde{\sigma}_{\mu}^2 = \frac{SS_{\nu}}{N-K-1}, \quad \tilde{\sigma}_{\mu}^2 = \frac{1}{T}(\tilde{\sigma}_{\nu}^2 - \tilde{\sigma}_{\mu}^2). \quad (6)$$

به دست می‌آیند. مطابق با (۳)، برآوردهای Σ و Σ^{-1} را می‌توان به صورت

$$\tilde{\Sigma} = \tilde{\sigma}_{\nu}^2 P + \tilde{\sigma}_{\mu}^2 Q, \quad \tilde{\Sigma}^{-1} = (\tilde{\sigma}_{\nu}^2)^{-1} P + (\tilde{\sigma}_{\mu}^2)^{-1} Q. \quad (7)$$

بدست آورد. فرض کنید مدل داده‌های پانلی $G \geq 2$ گروه با مولفه خطای تک عاملی به صورت

$$Y_{git} = \alpha^g + X_{git}\beta^g + u_{git}, \quad g = 1, \dots, G, \quad i = 1, \dots, N_g, \quad t = 1, \dots, T$$

موجود باشند، که در آن β^g یک بردار K بعدی است. شکل ماتریسی این مدل به صورت

$$Y_g = Z_g \delta_g + u_g \quad (8)$$

نشان داده می‌شود. بنابراین خواهیم داشت

$$\begin{aligned} \text{Cov}(u_g) = \Sigma_g &= \sigma_{\nu}^2 (I_{N_g} \otimes \bar{J}_T) + \sigma_{\mu}^2 (I_{N_g} \otimes (I_T - \bar{J}_T)) \\ &= \sigma_{\nu}^2 P_g + \sigma_{\mu}^2 Q_g, \quad g = 1, \dots, G. \end{aligned}$$

برآوردگر کمترین توان‌های دوم تعمیم یافته δ_g عبارت است از $\hat{\delta}_g = (Z_g' \Sigma_g^{-1} Z_g)^{-1} Z_g' \Sigma_g^{-1} Y_g$ تحت فرضیه برابری ضرایب رگرسیونی مدل داده‌های پانلی در G گروه (یعنی، $\delta^1 = \dots = \delta^G = \delta$).

مدل حاکم بر تمام مشاهدات به صورت $Y_A = Z_A \delta + u_A$ است که در آن

$$Y_A = \begin{pmatrix} Y_1 \\ \vdots \\ Y_G \end{pmatrix}, \delta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, u_A = \begin{pmatrix} u_1 \\ \vdots \\ u_G \end{pmatrix}, Z_A = \begin{pmatrix} 1_{N_1 T} & X_1 \\ \vdots & \vdots \\ 1_{N_G T} & X_G \end{pmatrix}, \nu_A = \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_G \end{pmatrix},$$

و $N_A = \sum_{g=1}^G N_g$ است. بنابر این تعاریف، برآوردگر کمترین توان‌های دوم تعمیم یافته بردار δ برابر با $\hat{\delta} = (Z_A' \Sigma_A^{-1} Z_A)^{-1} Z_A' \Sigma_A^{-1} Y_A$ نتیجه می‌شود. مدل تبدیل یافته (۸) را می‌توان مطابق با مدل (۵) به صورت

$$\begin{aligned} Q_g Y_g &= Q_g X_g \beta^g + Q_g \nu_g, \\ P_g Y_g &= P_g Z_g \delta^g + P_g u_g, \end{aligned} \quad (9)$$

نوشت. همچنین مدل‌های تبدیل یافته $Y_A = Z_A \delta + u_A$ عبارت است از

$$\begin{aligned} Q_A Y_A &= Q_A X_A \beta + Q_A \nu_A, \\ P_A Y_A &= P_A Z_A \delta + P_A u_A, \end{aligned} \quad (10)$$

که در آن

$$Q_A = \begin{pmatrix} Q_1 & & \\ & \ddots & \\ & & Q_G \end{pmatrix}, \quad P_A = \begin{pmatrix} P_1 & & \\ & \ddots & \\ & & P_G \end{pmatrix},$$

با درایه‌های صفر روی عناصر غیر قطر اصلی هستند. معمولاً فرض می‌شود که σ_ν^2 و σ_μ^2 در همه گروه‌ها برابر باشند. بنابراین برآوردگرهای σ_ν^2 و σ_μ^2 مبتنی بر مدل‌های (۱۰) به صورت

$$\begin{aligned} \hat{\sigma}_\nu^2 &= \frac{SS_{\nu A}}{N_A(T-1)-K} = \frac{Y_A'(Q_A - Q_A X_A (X_A' Q_A X_A)^{-1} X_A' Q_A) Y_A}{N_A(T-1)-K}, \\ \hat{\sigma}_\mu^2 &= \frac{SS_{\mu A}}{N_A-1-K} = \frac{Y_A'(P_A - P_A Z_A (Z_A' P_A Z_A)^{-1} Z_A' P_A) Y_A}{N_A-1-K}. \end{aligned} \quad (11)$$

به دست می‌آیند. همچنین برآوردهای Σ_A^{-1} و Σ_g^{-1} ، تحت فرضیه برابری δ_g ها برابر هستند با

$$\begin{aligned}\tilde{\Sigma}_g^{-1} &= (\tilde{\sigma}_g^2)^{-1} P_g + (\tilde{\sigma}_g^2)^{-1} Q_g; g = 1, \dots, G \\ \tilde{\Sigma}_A^{-1} &= (\tilde{\sigma}_A^2)^{-1} P_A + (\tilde{\sigma}_A^2)^{-1} Q_A.\end{aligned}\quad (12)$$

۳ آزمون برابری ضرایب رگرسیونی مدل داده‌های پانلی در چندین گروه

در مساله آزمون فرضیه برابری بردار ضرایب رگرسیونی مدل داده‌های پانلی در چندین گروه، یعنی

$$H_0: \delta^1 = \dots = \delta^G \stackrel{def}{=} \delta. \quad (13)$$

داریم $\hat{Y}_A = Z_A \hat{\delta} \sim N(Z_A \delta, Z_A (Z_A' \Sigma_A^{-1} Z_A)^{-1} Z_A')$ همچنین بردار مانده‌های مدل کلی به صورت $e = Y_A - \hat{Y}_A \sim N(0, \Sigma_A^{1/2} B \Sigma_A^{1/2})$ ، که در آن $B = I - \Sigma_A^{-1/2} Z_A (Z_A' \Sigma_A^{-1} Z_A)^{-1} Z_A' \Sigma_A^{-1/2}$ یک ماتریس تصویر از رتبه $N_A T - (K + 1)$ است. با تعریف $e^* = \Sigma_A^{-1/2} e \sim N(0, B)$ ، مجموع توان‌های دوم مانده‌های مدل به صورت

$$SSE_T = e^{*'} e^* = e' \Sigma_A^{-1} e, \quad (14)$$

است، که دارای توزیع کای-دو با $N_A T - (K + 1)$ درجه آزادی است. همچنین مجموع توان‌های دوم مانده‌ها برای مدل داده پانلی گروه g ام عبارت است از

$$SSE_g = e_g' \Sigma_g^{-1} e_g \sim \chi_{(N_g T - (K + 1))}^2, \quad (15)$$

که در آن $\hat{Y}_g = Z_g \hat{\delta}_g$ و $e_g = Y_g - \hat{Y}_g$

الف-روش تقریب: وقتی σ_ν^2 و σ_μ^2 معلوم هستند، یک آماره آزمون مناسب برای آزمون برابری δ_g ها در (۱۳) عبارت است از

$$F^* = \frac{SSE_T - \sum_{g=1}^G SSE_g}{\sum_{g=1}^G SSE_g} \frac{(N_A T - GK - G)}{(G - 1)(K + 1)}. \quad (16)$$

قضیه ۱. اگر σ_μ^2 و σ_ν^2 معلوم باشند آنگاه F^* دارای توزیع فیشر با درجات آزادی $(K+1)(G-1)$ و $(N_A T - GK - G)$ است.

برهان: با دوباره نویسی SSE_T و SSE_g به صورت

$$\begin{aligned} SSE_T &= Y'_A \Sigma_A^{-1} Y_A - Y'_A \Sigma_A^{-1} Z_A (Z'_A \Sigma_A^{-1} Z_A)^{-1} Z'_A \Sigma_A^{-1} Y_A \\ &= Y'_A \Sigma_A^{-1} Y_A - \hat{\delta}' (Z'_A \Sigma_A^{-1} Z_A) \hat{\delta}, \\ SSE_g &= Y'_g \Sigma_g^{-1} Y_g - Y'_g \Sigma_g^{-1} Z_g (Z'_g \Sigma_g^{-1} Z_g)^{-1} Z'_g \Sigma_g^{-1} Y_g \\ &= Y'_g \Sigma_g^{-1} Y_g - \hat{\delta}'_g (Z'_g \Sigma_g^{-1} Z_g) \hat{\delta}_g, \end{aligned}$$

خواهیم داشت

$$\begin{aligned} \sum_{g=1}^G SSE_g &= \sum_{g=1}^G Y'_g \Sigma_g^{-1} Y_g - \sum_{g=1}^G \hat{\delta}'_g (Z'_g \Sigma_g^{-1} Z_g) \hat{\delta}_g \\ &= Y'_A \Sigma_A^{-1} Y_A - \sum_{g=1}^G \hat{\delta}'_g (Z'_g \Sigma_g^{-1} Z_g) \hat{\delta}_g. \end{aligned}$$

حال داریم

$$\begin{aligned} SSE_T - \sum_{g=1}^G SSE_g &= \sum_{g=1}^G \hat{\delta}'_g (Z'_g \Sigma_g^{-1} Z_g) \hat{\delta}_g - \hat{\delta}' (Z'_A \Sigma_A^{-1} Z_A) \hat{\delta} \\ &= \sum_{g=1}^G Y'_g \Sigma_g^{-1} Z_g \hat{\delta}_g - \sum_{g=1}^G Y'_g \Sigma_g^{-1} Z_g \hat{\delta} \\ &= \sum_{g=1}^G \hat{\delta}'_g (Z'_g \Sigma_g^{-1} Z_g) (\hat{\delta}_g - \hat{\delta}) \\ &= \sum_{g=1}^G (\hat{\delta}_g - \hat{\delta})' (Z'_g \Sigma_g^{-1} Z_g) (\hat{\delta}_g - \hat{\delta}). \end{aligned}$$

می‌توان نشان داد $\hat{\delta}_g = (\sum_{g=1}^G Z'_g \Sigma_g^{-1} Z_g)^{-1} \sum_{g=1}^G (Z'_g \Sigma_g^{-1} Z_g) \hat{\delta}_g$ یعنی $\hat{\delta}$ میانگین وزنی $\hat{\delta}_g$ ها است. حال تعریف می‌کنیم $\hat{Y}_g^* = \Sigma_g^{-1/2} \hat{Y}_g \sim N(\Sigma_g^{-1/2} Z_g \delta_g, H_g)$ که در آن $e_g^* = \Sigma_g^{-1/2} e_g$ مستقل از \hat{Y}_g^* است. $H_g = I - \Sigma_g^{-1/2} Z_g (Z'_g \Sigma_g^{-1} Z_g)^{-1} Z'_g \Sigma_g^{-1/2}$ است چون $\text{Cov}(\hat{Y}_g^*, e_g^*) = 0$. بنابراین $SSE_T - \sum_{g=1}^G SSE_g$ مستقل از SSE_g است. همچنین $\sum_{g=1}^G SSE_g \sim \chi^2_{(N_A T - G(K+1))}$ و تحت فرضیه H_0 ، $SSE_T \sim \chi^2_{(N_A T - (K+1))}$ است و چون $SSE_T = \sum_{g=1}^G (\hat{\delta}_g - \hat{\delta})' (Z'_g \Sigma_g^{-1} Z_g) (\hat{\delta}_g - \hat{\delta}) + \sum_{g=1}^G SSE_g$ ، نتیجه گرفته می‌شود $\sum_{g=1}^G (\hat{\delta}_g - \hat{\delta})' (Z'_g \Sigma_g^{-1} Z_g) (\hat{\delta}_g - \hat{\delta})$ دارای توزیع $\chi^2_{((G-1)(K+1))}$ است. معمولاً در عمل σ_μ^2 و σ_ν^2 نامعلوم هستند که در این صورت آماره آزمون F^* غیرقابل کاربرد خواهد

بود. بنابراین با جایگذاری برآوردهای این مقادیر از رابطه (۱۱) خواهیم داشت

$$F = \frac{S\tilde{S}E_T - \sum_{g=1}^G S\tilde{S}E_g}{\sum_{g=1}^G S\tilde{S}E_g} \frac{(N_A T - GK - G)}{(G-1)(K+1)}, \quad (17)$$

که در آن

$$\begin{aligned} S\tilde{S}E_T &= Y'_A \tilde{\Sigma}_A^{-1} Y_A - Y'_A \tilde{\Sigma}_A^{-1} Z_A (Z'_A \tilde{\Sigma}_A^{-1} Z_A)^{-1} Z'_A \tilde{\Sigma}_A^{-1} Y_A, \\ S\tilde{S}E_g &= Y'_g \tilde{\Sigma}_g^{-1} Y_g - Y'_g \tilde{\Sigma}_g^{-1} Z_g (Z'_g \tilde{\Sigma}_g^{-1} Z_g)^{-1} Z'_g \tilde{\Sigma}_g^{-1} Y_g. \end{aligned}$$

به عنوان اولین پیشنهاد، توزیع F را با توزیع فشر با درجات آزادی $r_1 = (G-1)(K+1)$ و $r_2 = N_A T - GK - G$ تقریب زده می‌شود. بدین ترتیب فرضیه H_0 در سطح γ رد می‌شود، هرگاه $F_0 > F_{r_1, r_2, \gamma}$ باشد که F_0 مقدار مشاهده شده F در (۱۷) و $F_{m, n, \gamma}$ ، چنک γ ام توزیع فشر با درجات آزادی m و n هستند.

ب- بوت استرپ پارامتری: به دلیل اینکه کمیت تصادفی F در (۱۷) وابسته به پارامترهای مزاحم بوده و بدست آوردن توزیع آن پیچیده است، برای انجام آزمون (۱۳) می‌توان توزیع تقریبی آن را با استفاده از رویکرد بوت استرپ پارامتری بدست آورد و p -مقدار بوت استرپ پارامتری را محاسبه نمود. فرض کنید $\tilde{\Sigma}_A$ و $\tilde{\delta}$ به ترتیب، مقادیر مشاهده شده برآوردهای $\tilde{\Sigma}_A$ و $\tilde{\delta}$ باشند. Y_A^B از توزیع نرمال با میانگین $\tilde{\delta}$ و ماتریس کواریانس $\tilde{\Sigma}_A$ تولید می‌شود. سپس متغیر محوری بوت استرپ پارامتری مبتنی بر کمیت تصادفی (۱۷) به صورت

$$F^B = \frac{S\tilde{S}E_T^B - \sum_{g=1}^G S\tilde{S}E_g^B}{\sum_{g=1}^G S\tilde{S}E_g^B} \frac{(N_A T - GK - G)}{(G-1)(K+1)}, \quad (18)$$

محاسبه می‌شود، که در آن

$$\begin{aligned} S\tilde{S}E_T^B &= Y_A'^B \tilde{\Sigma}_A^{-1B} Y_A^B - Y_A'^B \tilde{\Sigma}_A^{-1B} Z_A (Z_A' \tilde{\Sigma}_A^{-1B} Z_A)^{-1} Z_A' \tilde{\Sigma}_A^{-1B} Y_A^B, \\ S\tilde{S}E_g^B &= Y_g'^B \tilde{\Sigma}_g^{-1B} Y_g^B - Y_g'^B \tilde{\Sigma}_g^{-1B} Z_g (Z_g' \tilde{\Sigma}_g^{-1B} Z_g)^{-1} Z_g' \tilde{\Sigma}_g^{-1B} Y_g^B, \end{aligned}$$

به طوری که $g = 1, \dots, G$ و

$$\begin{aligned}\tilde{\Sigma}_A^{-1B} &= (\tilde{\sigma}_{1B}^2)^{-1} P_A + (\tilde{\sigma}_{\nu B}^2)^{-1} Q_A, \quad \tilde{\Sigma}_g^{-1B} = (\tilde{\sigma}_{1B}^2)^{-1} P_g + (\tilde{\sigma}_{\nu B}^2)^{-1} Q_g, \\ \tilde{\sigma}_{1B}^2 &= \frac{S_{1A}^{2B}}{N_A - K - 1}, \quad \tilde{\sigma}_{\nu B}^2 = \frac{S_{\nu A}^{2B}}{N_A(T-1) - K}, \\ S_{1A}^{2B} &= Y_A'^B (P_A - P_A Z_A (Z_A' P_A Z_A)^{-1} Z_A' P_A) Y_A^B, \\ S_{\nu A}^{2B} &= Y_A'^B (Q_A - Q_A X_A (X_A' Q_A X_A)^{-1} X_A' Q_A) Y_A^B.\end{aligned}$$

فرض کنید F_0 مقدار مشاهده شده F باشد، آنگاه p -مقدار بوت استرپ پارامتری عبارت است از

$$P_b = P(F^B \geq F_0). \quad (19)$$

بنابراین فرضیه صفر در (۱۳) در سطح γ داده شده رد می‌شود هرگاه $P_b < \gamma$ باشد.

ج- آزمون دقیق: به منظور انجام آزمون فرضیه برابری بردارهای ضرایب رگرسیونی مدل داده‌های پانلی در G گروه، یک کمیت محوری به عنوان سومین روش به صورت

$$F_{exact} = \max(F_P, F_Q) \quad (20)$$

پیشنهاد می‌شود، که در آن

$$\begin{aligned}F_P &= (SSE_{TP} - \sum_{g=1}^G SSE_{P_g}) / \sum_{g=1}^G SSE_{P_g} \\ F_Q &= (SSE_{TQ} - \sum_{g=1}^G SSE_{Q_g}) / \sum_{g=1}^G SSE_{Q_g}.\end{aligned}$$

همچنین

$$\begin{aligned}SSE_{TP} &= Y_A' (P_A - P_A Z_A (Z_A' P_A Z_A)^{-1} Z_A' P_A) Y_A, \\ SSE_{TQ} &= Y_A' (Q_A - Q_A X_A (X_A' Q_A X_A)^{-1} X_A' Q_A) Y_A, \\ SSE_{P_g} &= Y_g' (P_g - P_g Z_g (Z_g' P_g Z_g)^{-1} Z_g' P_g) Y_g, \\ SSE_{Q_g} &= Y_g' (Q_g - Q_g X_g (X_g' Q_g X_g)^{-1} X_g' Q_g) Y_g.\end{aligned}$$

قضیه ۰۲. F_P و F_Q مستقل هستند و

$$F_P F_{((K+1)(G-1), N_A - G(K+1))} \frac{(K+1)(G-1)}{N_A - G(K+1)},$$

$$F_Q F_{(K(G-1), N_A(T-1) - GK)} \frac{K(G-1)}{N_A(T-1) - GK}.$$

برهان: با دوباره نویسی SSE_{T_P} و SSE_{P_g} به صورت

$$\begin{aligned} SSE_{T_P} &= \mathbf{Y}'_A \mathbf{P}_A \mathbf{Y}_A - \mathbf{Y}'_A \mathbf{P}_A \mathbf{Z}_A (\mathbf{Z}'_A \mathbf{P}_A \mathbf{Z}_A)^{-1} \mathbf{Z}'_A \mathbf{P}_A \mathbf{Y}_A \\ &= \mathbf{Y}'_A \mathbf{P}_A \mathbf{Y}_A - \hat{\boldsymbol{\delta}}' (\mathbf{Z}'_A \mathbf{P}_A \mathbf{Z}_A) \hat{\boldsymbol{\delta}}, \\ SSE_{P_g} &= \mathbf{Y}'_g \mathbf{P}_g \mathbf{Y}_g - \mathbf{Y}'_g \mathbf{P}_g \mathbf{Z}_g (\mathbf{Z}'_g \mathbf{P}_g \mathbf{Z}_g)^{-1} \mathbf{Z}'_g \mathbf{P}_g \mathbf{Y}_g \\ &= \mathbf{Y}'_g \mathbf{P}_g \mathbf{Y}_g - \hat{\boldsymbol{\delta}}'_g (\mathbf{Z}'_g \mathbf{P}_g \mathbf{Z}_g) \hat{\boldsymbol{\delta}}_g, \end{aligned}$$

خواهیم داشت

$$\begin{aligned} \sum_{g=1}^G SSE_{P_g} &= \sum_{g=1}^G \mathbf{Y}'_g \mathbf{P}_g \mathbf{Y}_g - \sum_{g=1}^G \hat{\boldsymbol{\delta}}'_g (\mathbf{Z}'_g \mathbf{P}_g \mathbf{Z}_g) \hat{\boldsymbol{\delta}}_g \\ &= \mathbf{Y}'_A \mathbf{P}_A \mathbf{Y}_A - \sum_{g=1}^G \hat{\boldsymbol{\delta}}'_g (\mathbf{Z}'_g \mathbf{P}_g \mathbf{Z}_g) \hat{\boldsymbol{\delta}}_g. \end{aligned}$$

بنابراین

$$\begin{aligned} SSE_{T_P} - \sum_{g=1}^G SSE_{P_g} &= \sum_{g=1}^G \hat{\boldsymbol{\delta}}'_g (\mathbf{Z}'_g \mathbf{P}_g \mathbf{Z}_g) \hat{\boldsymbol{\delta}}_g - \hat{\boldsymbol{\delta}}' (\mathbf{Z}'_A \mathbf{P}_A \mathbf{Z}_A) \hat{\boldsymbol{\delta}} \\ &= \sum_{g=1}^G \mathbf{Y}'_g \mathbf{P}_g \mathbf{Z}_g \hat{\boldsymbol{\delta}}_g - \sum_{g=1}^G \mathbf{Y}'_g \mathbf{P}_g \mathbf{Z}_g \hat{\boldsymbol{\delta}} \\ &= \sum_{g=1}^G \hat{\boldsymbol{\delta}}'_g (\mathbf{Z}'_g \mathbf{P}_g \mathbf{Z}_g) (\hat{\boldsymbol{\delta}}_g - \hat{\boldsymbol{\delta}}) \\ &= \sum_{g=1}^G (\hat{\boldsymbol{\delta}}_g - \hat{\boldsymbol{\delta}})' (\mathbf{Z}'_g \mathbf{P}_g \mathbf{Z}_g) (\hat{\boldsymbol{\delta}}_g - \hat{\boldsymbol{\delta}}). \end{aligned}$$

به سادگی می‌توان نشان داد که $\hat{\boldsymbol{\delta}} = (\sum_{g=1}^G \mathbf{Z}'_g \mathbf{P}_g \mathbf{Z}_g)^{-1} \sum_{g=1}^G (\mathbf{Z}'_g \mathbf{P}_g \mathbf{Z}_g) \hat{\boldsymbol{\delta}}_g$ یعنی $\hat{\boldsymbol{\delta}}$ میانگین وزنی $\hat{\boldsymbol{\delta}}_g$ ها است. حال تعریف می‌کنیم $\hat{\mathbf{Y}}_g^* = \mathbf{P}_g \hat{\mathbf{Y}}_g \sim N(\mathbf{P}_g \mathbf{Z}_g \boldsymbol{\delta}_g, \mathbf{H}_g)$ که در

آن $H_g = I - P_g Z_g (Z_g' P_g Z_g)^{-1} Z_g' P_g$ چون $\text{Cov}(\hat{Y}_g^*, e_g^*) = 0$ ، بنابراین \hat{Y}_g^* مستقل از $e_g^* = P_g e_g$ و همچنین $SSE_{TP} - \sum_{g=1}^G SSE_{P_g}$ مستقل از $\sum_{g=1}^G SSE_{P_g}$ است. بنابراین $\sum_{g=1}^G SSE_{P_g} \sim \sigma^2 \chi_{(N_A - G(K+1))}^2$ و تحت فرضیه H_0 داریم $SSE_{TP} \sim \chi_{(N_A - (K+1))}^2$ چون

$$SSE_{TP} = \sum_{g=1}^G (\hat{\delta}_g - \hat{\delta})' (Z_g' P_g Z_g) (\hat{\delta}_g - \hat{\delta}) + \sum_{g=1}^G SSE_{P_g},$$

نتیجه می‌شود $\sum_{g=1}^G (\hat{\delta}_g - \hat{\delta})' (Z_g' P_g Z_g) (\hat{\delta}_g - \hat{\delta})$ دارای توزیع $\chi_{((G-1)(K+1))}^2$ است. به طور مشابه نشان داده می‌شود F_Q نیز دارای توزیع فیشر است و اثبات کامل می‌شود. بنابراین p -مقدار دقیق برای آزمون فرضیه عبارت است از

$$P_e = 1 - F_{FP} \left(\frac{N_A - G(K+1)}{(K+1)(G-1)} F_{exact}^* \right) F_{FQ} \left(\frac{N_A(T-1) - GK}{K(G-1)} F_{exact}^* \right), \quad (21)$$

که در آن $F_{FP}(\cdot)$ و $F_{FQ}(\cdot)$ توابع توزیع فیشر با درجات آزادی $((K+1)(G-1), N_A - G(K+1))$ و $(K(G-1), N_A(T-1) - GK)$ هستند و F_{exact}^* مقدار مشاهده شده حاصل از رابطه (۲۰) است. بنابراین برای یک سطح آزمون داده شده γ فرضیه H_0 رد می‌شود اگر $P_e < \gamma$.

۴ مطالعه شبیه‌سازی

در این بخش خطای نوع اول و توان آزمون سه روش ارائه شده و روش ژائو (۲۰۰۷) در یک مطالعه شبیه‌سازی مقایسه می‌شوند. این ۴ روش به ترتیب به اختصار AP، PB، ET و GPV نشان داده می‌شوند. فرض کنید SST_1 و SST_ν به ترتیب مجموع توان‌های دوم مانده‌ها تحت فرضیه (۱۳) برای مدل‌های (۱۰) باشند. همچنین $SSE_\nu = \sum_{g=1}^G SS_{g,\nu}$ و $SSE_1 = \sum_{g=1}^G SS_{g,1}$ است که برای $g = 1, \dots, G$

$$SS_{g,\nu} = Y_g' (Q_g - Q_g X_g (X_g' Q_g X_g)^{-1} X_g' Q_g) Y_g$$

$$SS_{g,1} = Y_g' (P_g - P_g Z_g (Z_g' P_g Z_g)^{-1} Z_g' P_g) Y_g.$$

p - مقدار تعمیم یافته ارائه شده توسط ژائو (۲۰۰۷) به صورت

$$p = 1 - E_R \left\{ F_{r_1, r_2} \left[\frac{r_2}{r_1} \left(R \frac{SST_\nu}{SSE_\nu} + (1 - R) \frac{SST_1}{SSE_1} - 1 \right) \right] \right\}, \quad (22)$$

است، که در آن $R \sim \text{Beta} \left(\frac{N_A(T-1)-GK}{2}, \frac{N_A-G(K+1)}{2} \right)$ و $F_{r_1, r_2}(\cdot)$ تابع توزیع تجمعی فیشر با درجات آزادی $r_1 = (G-1)(2K+1)$ و $r_2 = N_A T - 2GK - G$ هستند.

الگوریتم ۱. برآورد توان آزمون‌ها:

گام ۱- برای مقادیر داده شده $\sigma_\mu^2, \sigma_\nu^2$ و بردارهای δ_g و Z_g به ازای $Y_g, g = 1, \dots, G$ را تولید کرده و $\tilde{\delta}_g, \tilde{\sigma}_g, \tilde{\sigma}_\nu, \tilde{\sigma}_1, \tilde{\Sigma}_{A^*}, F, F_{exact}$ را محاسبه کنید.
گام ۲- مقادیر Y_A^B و R را از توزیع‌های زیر تولید نمایید.

$$Y_A^B \sim N(Z_A \tilde{\delta}_*, \tilde{\Sigma}_{A^*}), \quad R \sim \text{Beta} \left(\frac{N_A(T-1)-GK}{2}, \frac{N_A-G(K+1)}{2} \right)$$

گام ۳- گام ۲ به تعداد m مرتبه (۵۰۰۰ مرتبه) برای برآورد p -مقدار روش‌های GPV و PB تکرار شود.
گام ۴- گام‌های ۱ تا ۳ به تعداد M مرتبه (۵۰۰۰ مرتبه) برای بدست آوردن برآوردهای توان آزمون روش‌های GPV و ET, AP, PB تکرار شوند.

در این مطالعه شبیه‌سازی، به جای ماتریس X_A سه ستون از داده‌های پانلی ارائه شده توسط **بالتاجی و گریفین (۱۹۸۳)** استفاده شده است. نتایج شبیه‌سازی برای حالت $G = 2$ و مقادیر مختلف (N_1, N_2, T) در قالب سناریوهای جدول ۱ مورد مطالعه قرار گرفته است و برآوردهای آن‌ها در شکل‌های ۱ و ۲ آورده شده‌اند. به عنوان یک مثال، برای سناریوی ۱۶-۱۳، یعنی حالت $(N_1, N_2, T) = (5, 10, 4)$ ، از ۴ ردیف اول داده‌های $10 + 5$ کشوری که به تصادف انتخاب شده‌اند، استفاده شده است. فرض می‌شود بردار ضرایب رگرسیونی جامعه اول یعنی δ^1 برابر با $(2, 3, 1, 5)$ و بردار δ^2 مقادیر مختلفی اختیار می‌کنند. ملاک قابل قبول برای مقایسه روش‌ها بدین صورت است: اولاً روشی نسبت به سایر روش‌ها ترجیح داده می‌شود که مقدار خطای نوع اول برآورد شده آن اختلاف معنی داری با ۰/۰۵ نداشته باشد. چنین روشی را قابل اعتماد می‌نامیم. ثانیاً بهترین روش باید بزرگ‌ترین توان را میان روش‌های قابل اعتماد به خود اختصاص دهد (لی و همکاران، ۲۰۱۴؛ شیائو و همکاران، ۲۰۱۷؛ ملک‌زاده و خراطی، ۲۰۱۸، ۲۰۱۹). با استفاده از قضیه حد مرکزی، فاصله اطمینان ۹۵ درصدی به منظور پوشش سطح اسمی ۰/۰۵، بازه‌ای

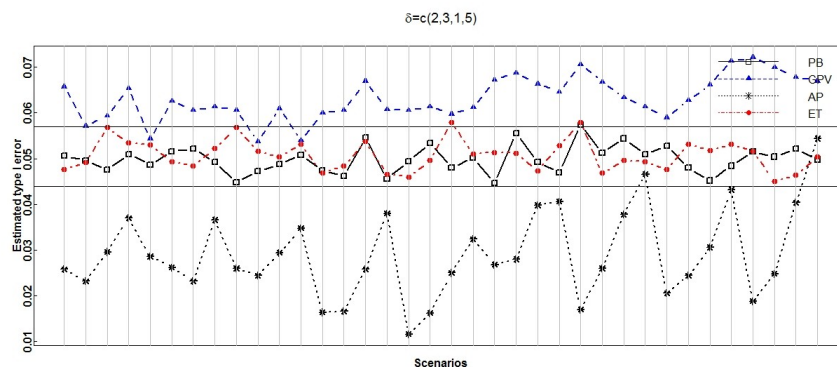
جدول ۰۱. سناریوهای در نظر گرفته شده در شبیه سازی‌ها		
سناریوها	σ_{μ}^2	(N_1, N_2, T)
۱ - ۴	$(0.01, 0.1, 1, 5)$	$(5, 13, 3)$
۵ - ۸	$(0.01, 0.1, 1, 5)$	$(10, 8, 3)$
۹ - ۱۲	$(0.01, 0.1, 1, 5)$	$(9, 9, 3)$
۱۳ - ۱۶	$(0.01, 0.1, 1, 5)$	$(5, 10, 4)$
۱۷ - ۲۰	$(0.01, 0.1, 1, 5)$	$(8, 7, 4)$
۲۱ - ۲۴	$(0.01, 0.1, 1, 5)$	$(6, 6, 5)$
۲۵ - ۲۸	$(0.01, 0.1, 1, 5)$	$(5, 7, 5)$
۲۹ - ۳۲	$(0.01, 0.1, 1, 5)$	$(7, 5, 5)$
۳۳ - ۳۶	$(0.01, 0.1, 1, 5)$	$(5, 5, 6)$

محدود به مقادیر ۰/۴۴۰ و ۰/۵۶۰ را نتیجه می‌دهد. به عبارتی اگر مقدار خطای نوع اول برآورد شده یک آزمون، کمتر از کران پایین یا بیش‌تر از کران بالای این بازه باشد می‌توان نتیجه گرفت که آن آزمون به ترتیب محافظه‌کارانه یا آزاد است.

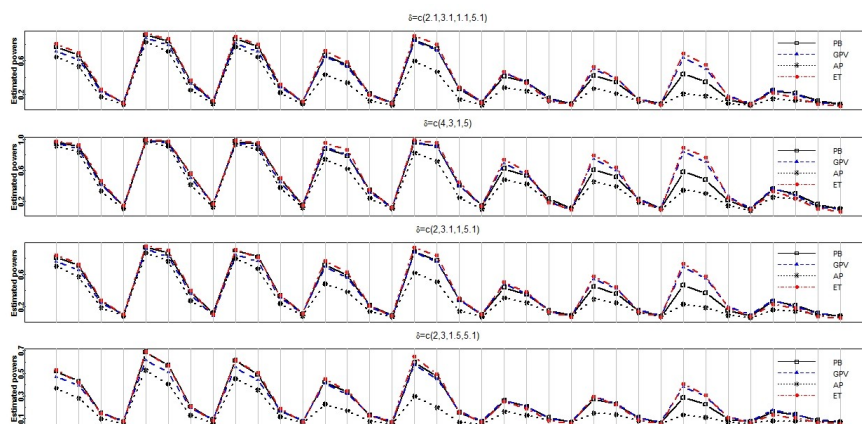
توان‌های برآورد شده از یک شبیه‌سازی به شبیه‌سازی دیگر اندکی تغییر می‌کنند (ملک‌زاده و خراطی، ۲۰۱۸؛ اسمعیلی و همکاران، ۲۰۲۰). بنابراین می‌توان از آزمون Z برای مقایسه توان‌های دو روش مختلف با یکدیگر استفاده نمود. بدین صورت که توان‌های دو آزمون زمانی اختلاف معنی داری در سطح α ۱۰۰ درصد دارند که $|\hat{p}_1 - \hat{p}_2| > Z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/5000}$ باشد که در آن $\hat{p} = (\hat{p}_1 + \hat{p}_2)/2$ و \hat{p}_1 و \hat{p}_2 توان‌های برآورد شده دو روش آزمون بر اساس ۵۰۰۰ نمونه هستند. در ادامه، نتایج حاصل از مقادیر برآورد شده از مطالعه شبیه‌سازی بیان شده است.

فرع ۰۱. در تمامی سناریوهایی که در این مطالعه شبیه سازی در نظر گرفته شده است، مقدار خطای نوع اول برآورد شده آزمون با رویکرد بوت استرپ پارامتری (PB) بین ۰/۴۴۶ و ۰/۵۵۶ و روش دقیق (ET) بین ۰/۴۵۰ و ۰/۵۶۰ قرار دارند که نشان می‌دهد هر دو روش، آزاد یا محافظه‌کارانه نیستند.

فرع ۰۲. مقدار خطای نوع اول برآورد شده روش GPV غالباً از کران بالای ۰/۵۶۰ بیشتر شده‌اند بنابراین این روش، آزاد در نظر گرفته می‌شود. همچنین مقدار خطای نوع اول برآورد شده روش AP غالباً از کران پایین ۰/۴۴۰ کمتر به دست آمده‌اند، پس این روش محافظه‌کارانه است. بنابراین این روش‌ها برای انجام آزمون فرضیه (۱۳) قابل اعتماد نیستند. این مطالب را با نگاهی به شکل ۱ که شامل مقادیر برآورد شده خطای نوع اول چهار آزمون است، می‌توان نتیجه گرفت.



شکل ۱. برآورد خطای نوع اول چهار روش آزمون



شکل ۲. برآورد توان آزمون چهار روش به ازای بردارهای پارامتری مختلف

فرع ۳.۰ برای مقایسه توان‌های برآورد شده، ملاک مذکور زمانی استفاده می‌شود که مقدار خطای نوع اول برآورد شده سایر روش‌ها نزدیک به ۵٪ باشند. به طور مثال، در سناریوی ۵، یعنی حالت $(N_1, N_2, T) = (10, 8, 3)$ با $(\sigma_\nu^2, \sigma_\mu^2) = (1, 0.1)$ ، می‌توان دید که توان‌های PB و ET به طور معنی‌داری در سطح ۱۰٪ بزرگتر از توان روش GPV هستند.

فرع ۴. شکل ۲ نتایج مربوط به برآورد توان این چهار روش به ازای مقادیر مختلف δ ‌های در نظر گرفته شده را نشان می‌دهد. همان‌طور که قبلاً بیان شد در این شبیه‌سازی‌ها مقدار σ_ν^2 برابر ۱ و مقادیر مختلفی

برای σ_μ^2 به ترتیب شامل 0.1% ، 1% و 5% در نظر گرفته شده است. نتایج شکل ۲ نشان می‌دهد که توان هر چهار روش به ازای افزایش σ_μ^2 کاهش می‌یابد. در تمامی نتایج گزارش شده، توان آزمون تقریب (AP) همان‌طور که انتظار داشتیم از بقیه روش‌ها کمتر است. نتایج روش تعمیم یافته (GPV) نیز به دلیل عدم کنترل خطای نوع اول (که همیشه مقداری بیشتر از 5% را از آن مشاهده نموده‌ایم) مقادیر غیر واقعی را ارائه نموده‌اند. در مقایسه دو روش دقیق و بوت استرپ پارامتری نیز می‌توان مشاهده نمود که روش دقیق، توان به مراتب بالاتری را نتیجه داده است. بنابراین روش دقیق را می‌توان روش پرتوان‌تر در بین این روش‌ها معرفی و پیشنهاد نمود.

فرع ۵. در مجموع به نظر می‌رسد که روش بوت استرپ پارامتری و روش دقیق نسبت به روش تقریبی و روش p -مقدار تعمیم یافته بر اساس نرخ خطای نوع اول و توان‌های آزمون، عملکرد بهتری دارند. به عبارتی روش بوت استرپ پارامتری و روش دقیق نسبت به سایر روش‌ها پرتوان هستند. روش PB نیازمند شبیه‌سازی است در حالی که روش ET یک کمیت محوری را تعریف می‌کند و محاسبه آن به مراتب ساده‌تر است. بنابراین در آزمون فرضیه برابری بردار ضرایب مدل‌های پانلی از روش ET بهره برده شود.

۵ مثال کاربردی

برای انجام آزمون فرضیه برابری ضرایب رگرسیون مدل داده‌های پانلی با روش‌های بیان شده مدل تقاضای بنزین بالتاجی و گریفین (۱۹۸۳) به صورت

$$\ln \frac{Gas}{Car} = \alpha + \beta_1 \ln \frac{Y}{N} + \beta_2 \ln \frac{PMG}{PGDP} + \beta_3 \ln \frac{Car}{N} + u$$

در نظر گرفته شده است، که در آن $\frac{Gas}{Car}$ مصرف بنزین موتور هر خودرو، $\frac{Y}{N}$ درآمد سرانه واقعی، $\frac{PMG}{PGDP}$ قیمت واقعی بنزین و $\frac{Car}{N}$ سهم سرانه اتومبیل را نشان می‌دهد. این مدل پانلی شامل مشاهدات سالانه در ۱۸ کشور سازمان همکاری و توسعه اقتصادی است که دوره‌ای از ۱۹۶۰ تا ۱۹۷۸ ($T = 19$) را پوشش می‌دهد. فرض می‌شود که مدل داده‌ها دارای مولفه خطای یک عاملی است. اینجا از کل داده با دوره زمانی $T = 19$ برای $G = 2, 3$ به صورتی که در جدول ۲ گزارش می‌شود، استفاده شده است. در بین این کشورها می‌توان گروه کشورهای صنعتی (شامل آمریکا، آلمان، فرانسه، ژاپن، انگلیس و کانادا)، کشورهای غیر اروپایی (شامل آمریکا، استرالیا، ژاپن، ترکیه، کانادا) و کشورهای اروپایی غربی شامل (هلند، دانمارک، سوئد، نروژ، بلژیک و سوئیس) را مشاهده نمود. در اینجا در نظر داریم مدل‌های پانلی حاکم بر کشورهای

اروپایی در مقابل سایرین (سناریوی اول)، صنعتی و سایرین (سناریوی دوم)، کشورهای اروپایی غربی، صنعتی و سایرین (سناریوی سوم)، و کشورهای غرب اروپا با سایر کشورهای اروپایی (سناریوی چهارم) را مورد بررسی قرار دهیم. بدین منظور آزمون فرضیه برابری ضرایب مدل‌های هر سناریو را بصورت جداگانه در سطح آزمون ۵٪ مورد ارزیابی قرار می‌دهیم. فرضیه صفر در تمامی سناریوها بیان می‌دارد که مدل پانلی

جدول ۲. p -مقدارهای گزارش شده از روش‌های آزمون برای مثال واقعی

سناریو	(N_1, N_2, \dots, N_G)	PB	GPV	AP	ET
اول	(۱۳, ۵)	۰/۰۰۰۸	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۲۱۶
دوم	(۶, ۱۲)	۰/۰۰۳۱	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۱۰۰
سوم	(۶, ۶, ۶)	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۱۰۱
چهارم	(۶, ۷)	۰/۰۰۱۴	۰/۰۰۰۲	۰/۰۰۰۰	۰/۰۱۸۸

حاکم بر گروه‌های تعریف شده در هر سناریو یکسان هستند. نتایج جدول نشان می‌دهد که مدل مصرف بنزین در کشورهای اروپایی با سایر کشورها متفاوت است و همچنین این تفاوت در کشورهای صنعتی و سایر نیز قابل مشاهده است. بنابراین به بررسی مدل حاکم بر کشورهای صنعتی، غرب اروپا و سایر در سناریوی سوم پرداختیم که نشان می‌دهد مدل‌های حاکم بر این سه گروه نیز یکسان ناست. از جمله نتایج آنکه بین مدل‌های پانلی بین کشورهای غرب اروپا و سایر کشورهای اروپایی نیز مدل بطور معنادار متفاوت است. با نگاهی به دستاوردهای فوق می‌توان به اهمیت انجام آزمون فرضیه برابری مدل‌های پانلی حاکم در گروه‌های مشهود درون داده‌ها پی برد.

بحث و نتیجه گیری

روش تقریبی (AP)، روش بوت استرپ پارامتری (PB) و یک کمیت محوری با توزیع دقیق (ET) برای آزمون فرضیه برابری ضرایب رگرسیونی مدل داده‌های پانلی در چندین گروه ارائه شده است. مطالعه شبیه‌سازی نشان می‌دهد که برآورد سطح آزمون دو روش بوت استرپ پارامتری و دقیق، نزدیک به سطح اسمی (۵٪) هستند در حالی که دو روش دیگر قابل اعتماد نیستند. بدین ترتیب در حالت‌هایی که اندازه برآورد شده روش‌های دیگر نیز نزدیک به ۵٪ است، توان آزمون روش‌های PB و ET دارای اختلاف معنی داری با سایر روش‌ها هستند. همچنین توان آزمون روش ET غالباً بیش از PB است. بنابراین برای آزمون برابری ضرایب رگرسیونی، روش دقیق پیشنهاد می‌شود.

تقدیر و تشکر

از داوران محترم، هیئت تحریریه و ویراستار گرامی مجله به خاطر ارائه نقطه نظرات ارزشمندشان کمال تقدیر و تشکر را داریم.

مراجع

صادقی، س. و کاظمی، ا. (۱۳۸۸)، برازش مدل رگرسیونی با داده‌های پانلی پویا توسط روش‌های ماکسیمم درستنمایی و بیزی، مجله علوم آماری، ۱، ۷۹-۹۴.

Arellano, M. and Honoré, B. (2001), Panel Data Models: Some Recent Developments, *Handbook of Econometrics*, Elsevier, **5**, 29-32.

Balestra, P. and Nerlove, M. (1966), Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas, *Econometrica* (pre-1986), **34**, 585.

Baltagi, B. H. (1981), Pooling: An Experimental Study of Alternative Testing and Estimation Procedures in a Two-Way Error Component Model, *Journal of Econometrics*, **17**, 21-49.

Baltagi, B. H. (2013), *Econometric Analysis of Panel Data* (fifth).

Baltagi, B. H. and Griffin, J. M. (1983), Gasoline Demand in the OECD: An Application of Pooling and Testing Procedures, *European Economic Review*, **22**, 117-137.

Baltagi, B. H., Song, S. H. and Koh, W. (2003), Testing Panel Data Regression Models with Spatial Error Correlation, *Journal of Econometrics*, **117**, 123-150.

- Canay, I. A. (2011), A Simple Approach to Quantile Regression for Panel Data, *The Econometrics Journal*, **14**, 368-386.
- Esmaeli-Ayan, A., Malekzadeh, A. and Hormozinejad, F. (2020), Inferences on the Regression Coefficients in Panel Data Models: Parametric Bootstrap Approach, *Mathematical Sciences*, **14**, 67-73.
- Fuller, W. A. and Battese, G. E. (1974), Estimation of Linear Models with Crossed-Error Structure, *Journal of Econometrics*, **2**, 67-78.
- Hsiao, C. (2014), *Analysis of Panel Data*, No. 54, Cambridge University Press.
- Kao, C. and Chiang, M. H. (2000), On the Estimation and Inference of a Cointegrated Regression in Panel Data, *Advances in Econometrics*, **15**, 179-222.
- Kyriazidou, E. (1997), Estimation of a Panel Data Sample Selection Model, *Econometrica: Journal of the Econometric Society*, **65**, 1335-1364.
- Li, H. Q., Tang, M. L. and Wong, W. K. (2014), Confidence Intervals for Ratio of Two Poisson Rates Using the Method of Variance Estimates Recovery, *Computational Statistics*, **29**, 869-889.
- Malekzadeh, A. and Kharrati-Kopaei, M. (2018), Inferences on the Common Mean of Several Normal Populations Under Heteroscedasticity, *Computational Statistics*, **33**, 1367-1384.
- Malekzadeh, A. and Kharrati-Kopaei, M. (2019), Inferences on the Common Mean of Several Heterogeneous Log-Normal Distributions, *Journal of Applied Statistics*, **46**, 1066-1083.

Nerlove, M. (1971), A Note on Error Components Models, *Econometrica: Journal of the Econometric Society*, **39**, 383-396.

Panel Data Econometrics, Panel Data Sets (2004), <http://People.Stern.Nyu.Edu/Wgreene/Econometrics/PanelDataSets.htm>

Phillips, P. C. and Moon, H. R. (2000), Nonstationary Panel Data Analysis: An Overview of Some Recent Developments, *Econometric Reviews*, **19**, 263-286.

Swamy, P. A. V. B. and Arora, S. S. (1972), The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models, *Econometrica* **40**, 261-275.

Wansbeek, T. and Kapteyn, A. (1982), A Simple Way to Obtain the Spectral Decomposition of Variance Components Models for Balanced Data, *Communication in Statistics-Theory and Methods*, **11**, 2105-2112.

Xiao, M., Jiang, T., Zhang, H. and Shan, G. (2017), Exact One-Sided Confidence Limit for the Ratio of Two Poisson Rates, *Statistics in Biopharmaceutical Research*, **9**, 180-185.

Xu, L. and Tian, M. (2017), Parametric Bootstrap Inferences for Panel Data Models, *Communication in Statistics-Theory and Methods*, **46**, 5579-5594.

Yue, L. L., Shi, J. H. and Song, W. X. (2017), A Parametric Bootstrap Approach for Two-Way Error Component Regression Models, *Communications in Statistics-Simulation and Computation*, **46**, 3952-3961.

Zhao, H. B. (2007), Exact Tests in Panel Data Using Generalized P-values, *Communication in Statistics-Theory and Methods*, **37**, 18-36.

Testing the Equality of Regression Coefficients of Panel Models in Several Groups

Malekzadeh¹, A., Esmaeli-Ayan, A.², Mahmoudi, S. M.³

¹Department of Computer Science and Statistics, Faculty of Mathematics, Toosi University of Technology, Tehran, Iran.

²Department of Statistics, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran.

³Faculty of Mathematics, Statistics and Computer Sciences, Semnan University, Semnan, Iran.

Abstract: The panel data model is used in many areas, such as economics, social sciences, medicine, and epidemiology. In recent decades, inference on regression coefficients has been developed in panel data models. In this paper, methods are introduced to test the equality models of the panel model among the groups in the data set. First, we present a random quantity that we estimate its distribution by two methods of approximation and parametric bootstrap. We also introduce a pivotal quantity for performing this hypothesis test. In a simulation study, we compare our proposed approaches with an available method based on the type I error and test power. We also apply our method to gasoline panel data as a real data set.

Keywords: Panel data, Regression coefficient, Parametric bootstrap, Exact method.

Mathematics Subject Classification (2010): 62F03, 62F40, 62H15.