جمهوری اسلامی ایران
مرکز آمار ایران
پژوهشکده‌ی آمار

پژوهشکدهٔ آمار

روش آماری مناسب برای ساخت مدل عرضه‌ی نیروی کار با توجه به ویژگی‌های متغیر پاسخ

۱۳۸۴

روش آماری مناسب برای ساخت مدل عرضه‌ی نیروی کار با توجه به ویژگی‌های متغیر پاسخ

فاطمه هرندی

فرشید جمشیدی

علیرضا فرهادی‌کیا

مریم جوادی

گروه پژوهشی طرحهای فنی و روش‌های آماری

پژوهشکده‌ی آمار

۱۳۸۴

به نام خداوند جان و خرد

پیشگفتار

هر چند در طول سه دهه گذشته با اجرای طرح‌های نمونه‌گیری علمی و دقیق و تهیه اطلاعات خرد در سطح افراد و خانوارها، ادبیات وسیعی در زمینه بررسی عرضه‌ی نیروی کار در سطح خرد شکل گرفته و عواملی که بر تصمیم‌گیری افراد برای حضور در بازار کار و میزان این حضور مؤثر هستند، در قالب مدل‌های رگرسیونی مورد مطالعه قرار گرفته است اما در اغلب این بررسی‌ها از مدل‌های رگرسیونی معمولی و در نتیجه از برآوردگرهای کمترین توان‌های دوم معمولی، OLS[1] و یا حداکثر از کمترین توان‌های دوم تعمیم یافته، GLS[2] استفاده شده است که با توجه به محدودیت‌ها و ویژگی‌های متغیر وابسته مدل عرضه (ساعت کار یا حضور در نیروی کار)، وجود اریبی در نتایج بسیار محتمل است. برای مثال چنانچه ساعت کار به عنوان متغیر وابسته برای ساخت مدل در نظر گرفته شود، نپذیرفتن مقادیر منفی و صفر بودن آن برای تعداد قابل ملاحظه‌ای از واحدهای مورد مطالعه، از جمله مسایلی است که در صورت عدم توجه می‌تواند منجر به بروز اریبی در نتایج شود.

با توجه به این مسائل، پژوهشکده‌ی آمار با تعریف یک طرح پژوهشی، ضمن مطالعه روش‌های مختلف ساخت مدل‌های رگرسیونی با توجه به ویژگی‌های متغیر وابسته مدل، اقدام به ارائه مدل رگرسیونی مناسب برای عرضه‌ی نیروی کار افراد نمود. طرح پژوهشی فوق‌الذکر شامل دو مرحله «مطالعاتی» و «محاسباتی» بود. در مرحله «مطالعاتی»، ضمن مروری مختصر بر مدل رگرسیون خطی و مفروضات آن، مدل‌های رگرسیونی برای متغیرهای وابسته دو حالتی و محدود شده نیز مورد بررسی قرار گرفت که ماحصل آن در فصول اول تا سوم گزارش حاضر ارائه شده است. مبانی نظریه اقتصادی عرضه‌ی نیروی کار و مطالعات انجام شده در سایر کشورها و ایران در زمینه عرضه‌ی نیروی کار در سطح خرد نیز مورد بررسی قرار گرفت که اهم موارد آن در فصول ٤ و ٥ گزارش، ارائه شده است.

---

[1] Ordinary Least Squares
[2] Generalized Least Squares

در مرحله «محاسباتی»، ابتدا به بررسی عرضه‌ی نیروی کار در ایران از سال ۱۳۳۵ تاکنون پرداخته شـده اسـت؛ سـپس با روش‌های مختلف آماری، مدل عرضه‌ی نیروی کار کشور در سطح خرد که در آن سـاعت کـار بـه‌عـنوان متغـیر وابسـته و ویژگـی‌هـای اجتماعی اقتصادی فرد یا خانوار او به‌عنوان متغـیرهای مستقل در مدل‌ها حضور یافته‌اند، ساخته شده؛ و در انتها با مقایسه مدل‌های مختلف، مدل مناسـب عرضـه‌ی نـیروی کـار کشـور انتخاب و ضرایب آن تحلیل شده است. ماحصل این مرحله از طرح در فصل‌های ٦ و ٧ گزارش، ارائه شده است.

در گروه مطالعاتی طرح مذکور از پژوهشکده‌ی آمار، سرکار خانم‌ها فاطمه هرندی (مجری طرح) و مـریم جوادی و آقای فرشید جمشیدی و از دفتر اقتصاد کلان سازمان مدیریت و برنامه‌ریزی کشور آقـای علیرضـا فـرهادی کیا در گروه عضویت داشتند که به این وسیله از زحمات بی‌دریغ یکایک این افراد تشکر و قدردانی می شود.

در طـول انجـام این طرح، سرکار خانم الهام علیزاده با دقت و حوصله بسیار، زحمت حروفچینی تمـام مـدارک و مسـتندات طـرح را به عهده داشته‌اند که در این‌جا لازم است از تلاش بی‌وقفه ایشان تشکر و قدردانی شود.

در پایـان از خوانـندگان محـترم تقاضـا مـی‌شـود، نظـرات اصلاحی خود را در ارتباط با محتوی گزارش، به گروه پژوهشی طرح‌های فنی و روش‌های آماری پژوهشکده منعکس نمایند.

گروه پژوهشی طرح‌های فنی و روش‌های آماری

پژوهشکده‌ی آمار

فهرست مطالب

عنوان **صفحه**

مقدمه ۱

فصل ۱- کلیات ۵

فصل ۲- مدل‌های رگرسیونی برای متغیرهای وابسته دو حالتی ۱۳

فصل ۳- مدل‌های رگرسیونی برای متغیرهای وابسته محدود شده ۷۵

فصل ٤- مبانی نظری عرضه‌ی نیروی کار ۱۰۹

فصل ۵- مروری بر تحقیقات انجام شده ۱۳۵

فصل ٦- بررسی توصیفی شاخص‌های نیروی کار از سال ۱۳۳۵ تا سال ۱۳۸۲ ..... ۱٦۵

فصل ۷- مدل عرضه‌ی نیروی کار کشور ۱۷۹

منابع و ماخذ ۲۰۵

پیوست‌ها ۲۱۱

مقدمه

در بـازار نـیروی کار در سطح کلان، عوامل متعددی از قبیل رشد جمعیت و ساختار سنی و جنسی آن، الگـوی مشارکت (تعجیل در سن ورود به نیروی کار، تأخیر در سن خروج از نیروی کار، نحوه مشـارکت زنان و ...)، میزان مشارکت، سطح تحصیلات و ... بر عرضه مؤثر است. در ایران با توجه بـه پدیده انفجار جمعیت و شرایط ویژه اقتصادی، نرخ بیکاری در سالهای اخیر رشد نسبتاً بالایی را تجـربه کـرده اسـت و امروزه ایجاد فرصت‌های شغلی یکی از دغدغه‌های مسئولان کشور می‌باشد. بـرای تعییـن مـیزان فرصت‌های شغلی مورد نیاز، پیش‌بینی جمعیت فعال اقتصادی و ترکیب آن (از نظـر جـنس، سـن، سـطح تحصـیلات و ...) ضـروری است. برای پیش‌بینی این جمعیت می‌بایست عواملـی کـه موجـب ورود و خروج افراد به نیروی کار می‌شود، شناسایی شده و بر مبنای آن مدل عرضه‌ی نیروی کار ساخته شود. این عوامل که ویژگی‌های فردی یا خانواری افراد را نشان می‌دهد، مجموعـه‌ای از متغـیرهای اجتماعـی- اقتصـادی هستند که بر تصمیم‌گیری افراد مبنی بر پیوستن به نـیروی کـار یـا خـروج از آن تأثـیر می‌گذارد. با شناخت این عوامل از طریق ساخت مدل عرضه‌ی نـیروی کـار کشـور، مـی‌تـوان ابـزار مناسـب بـرای تصـمیم‌گـیری و سیاسـت‌گذاری را در اختیار دست‌اندرکاران برنامه‌ریزی سیاست‌های اشتغال قرار داد. هر چند در طول سه دهه گذشته با اجرای طرح‌های نمونه‌گیری علمی و دقیق و تهیه اطلاعات خرد در سطح افراد و خانوارها، ادبیات وسیعی در زمیـنه بررسی عرضه‌ی نیروی کار در سطح خرد شکل گرفته و عواملی که بر تصمیم‌گیری افراد بـرای حضـور در بـازار کـار و مـیزان ایـن حضور مؤثر هستند، در قالب مدل‌های رگرسیونی مورد مطالعـه قـرار گرفـته است اما در اغلب این بررسی‌ها از مدل‌های رگرسیونی معمولی و در نتیجه از برآوردگـرهای کمتریـن تـوان‌های دوم معمولی، OLS[1] و یا حداکثر از کمترین توان‌های دوم تعمیم یافـته، GLS[2] استفاده شده است که با توجه به محدودیت‌ها و ویژگی‌های متغیر وابسته مدل عرضه (سـاعت کار یا حضور در نیروی کار)، وجود اریبی در نتایج بسیار محتمل است. برای مثال چنانچه

[1] Ordinary Least Squares
[2] Generalized Least Squares

ساعت کار به عنوان متغیر وابسته برای ساخت مدل در نظر گرفته شود، نپذیرفتن مقادیر منفی و صفر بودن آن برای تعداد قابل ملاحظه‌ای از واحدهای مورد مطالعه، از جمله مسایلی است که در صورت عدم توجه می‌تواند منجر به بروز اریبی در نتایج شود.

موارد استفاده از مدل‌های رگرسیونی نیز طی دو دهه گذشته، گسترش قابل توجهی یافته است. در ابتدا محققین تلاش می‌نمودند از طریق مدل‌های خطی عمومی (GLM)[1] شامل آنالیز واریانس، آنالیز کواریانس و رگرسیون چندگانه، بین مجموعه‌ای از متغیرهای توضیحی و یک متغیر پاسخ ارتباط برقرار کنند. این ابزارها زمانی مفید بودند که متغیر وابسته (متغیر پاسخ)، با مقیاس فاصله‌ای مساوی اندازه‌گیری شده بود. اما در علوم اجتماعی و زیستی، تنها موارد محدودی از متغیرهای پاسخ دارای چنین ویژگی هستند، و در اکثر موارد، متغیر وابسته، دو حالتی، ترتیبی، اسمی، ممیزی شده، بریده شده یا شمارشی است. خوشبختانه طی دهه گذشته نرم‌افزارهای آماری قدرتمندی برای برآورد یک چنین مدل‌هایی گسترش یافته که پاسخی به نیاز روز افزون کاربرد این مدل‌ها بوده است. نکته حائز اهمیت در ارتباط با این مدل‌ها پیچیدگی تفسیر نتایج در مقایسه با مدل رگرسیون خطی است که معمولاً این موضوع کمتر مورد توجه قرار می‌گیرد و اغلب تنها به معنی‌دار بودن آماری، و بررسی علامت پارامترهای برآورد شده، اکتفا می‌شود. متأسفانه علی‌رغم پیچیدگی این مدل‌ها و نحوه تفسیر نتایج آن‌ها، در اغلب کتاب‌های رگرسیون نیز، مباحث مربوط به این‌گونه مدل‌ها بسیار به اختصار آمده است که موجب تشدید مشکل درک و کاربرد این مدل‌ها توسط محققین می‌شود.

با توجه به این مسائل، پژوهشکده‌ی آمار با تعریف یک طرح پژوهشی، ضمن مطالعه روش‌های مختلف ساخت مدل‌های رگرسیونی با توجه به ویژگی‌های متغیر وابسته مدل، اقدام به ارائه مدل رگرسیونی مناسب برای عرضه‌ی نیروی کار افراد نمود. طرح پژوهشی فوق‌الذکر شامل دو مرحله «مطالعاتی» و «محاسباتی» بود. در مرحله «مطالعاتی»، ضمن مروری مختصر بر مدل رگرسیون خطی و مفروضات آن، مدل‌های رگرسیونی برای متغیرهای وابسته دو حالتی و محدود شده نیز مورد بررسی قرار گرفت که ماحصل آن در فصول اول تا سوم گزارش حاضر ارائه شده

[1] General Linear Model

اســت. مبانی نظریه اقتصادی عرضه‌ی نیروی کار و مطالعات انجام شده در سایر کشورها و ایران در زمیـنه عرضه‌ی نیروی کار در سطح خرد نیز مورد بررسی قرار گرفت که اهم موارد آن در فصول ٤ و ٥ گزارش، ارائه شده است.

در مـرحله «محاسـباتی»، ابـتدا بـه بررسـی عرضـه‌ی نیروی کار در ایران از سال ١٣٣٥ تاکنون پرداخـته شـده اسـت؛ سپس با روش‌های مختلف آماری، مدل عرضه‌ی نیروی کار کشور در سطح خـرد کـه در آن ساعت کار به‌عنوان متغیر وابسته و ویژگی‌های اجتماعی اقتصادی فرد یا خانوار او بـه‌عـنوان متغـیرهای مسـتقل در مـدل‌ها حضور یافته‌اند، ساخته شده؛ و در انتها با مقایسه مدل‌های مخـتلف، مـدل مناسب عرضه‌ی نیروی کار کشور انتخاب و ضرایب آن تحلیل شده است. ماحصل این مرحله از طرح در فصل‌های ٦ و ٧ گزارش، ارائه شده است.

فصل ۱

کلیات

یک مدل اقتصادی مناسب، مدلی است که تا حد امکان ساده باشد؛ برای یک مجموعه داده، برآوردی یکتا از پارامترها را حاصل نماید؛ مقدار زیادی از تغییرات متغیر وابسته را بیان کند؛ با فرضیات تئوری، بخصوص در مورد علایم پارامترها، سازگار باشد و قدرت پیشگویی آن در عمل خوب باشد. در فرموله کردن یک مدل اقتصادی، علاوه بر توجه به ویژگی‌های یک مدل خوب، باید به خطاهایی که ممکن است به‌صورت بالقوه در تشخیص مدل رخ دهد، میزان اهمیت و نحوه تصحیح یا کاهش آن توجه نمود. از جمله این خطاها، حذف یک متغیر مهم در مدل، ورود متغیر غیرمربوط به مدل، فرم تابعی غلط، عدم برقراری مفروضات آماری مدل و وجود نقاط دور افتاده و مؤثر در مجموعه داده‌ها می‌باشد. در ادامه مروری مختصر بر هریک از این خطاها، نحوه تشخیص و نحوه رفع آن داریم.

در صورت همبستگی متغیر مستقل حذف شده از مدل با متغیرهای مستقل موجود در مدل، برآورد ضرایب اریب و ناسازگار خواهند بود و در صورت وجود نداشتن این همبستگی، ضریب عرض از مبدا اریب می‌شود (بجز وقتی که میانگین متغیر حذف شده، صفر باشد). در هر دو حالت، واریانس خطاها و ضرایب مدل، اریب بوده (اریبی مثبت) و در نتیجه بازه اطمینان و آزمون فرض‌ها واقعی نیستند.

ورود متغیر غیرمربوط به مدل، تنها موجب افزایش واریانس برآورد ضرایب و در نتیجه کاهش دقت برآوردها می‌شود و این امکان به‌وجود می‌آید که فرض صفر بودن ضرایبی را که واقعاً معنی‌دار هستند به دلیل بازه اطمینان بزرگ، بپذیریم. بنابراین هرچند، علی‌رغم وجود یک چنین متغیرهایی در مدل، برآوردها همچنان نااریب و واریانس خطاها و در نتیجه آزمون‌های فرض صحیح می‌باشند، بهتر است تمام متغیرهایی که تئوری اقتصادی مجاز می‌داند (متغیرهای اصلی) را در مدل بیاوریم و حتی اگر معنی‌دار نشوند در مدل حفظ شوند. متغیرهای کنترلی و یا متغیرهایی که تئوری اقتصادی در مورد آن‌ها تصریحی ندارد، می‌توانند فقط در صورت معنی‌دار شدن در مدل باقی بمانند.

با توجه به این‌که تئوری اقتصادی معمولاً فقط متغیرهای مرتبط را مشخص می‌کند و نه شکل تابعی ارتباط را، امکان تشخیص نادرست فرم تابعی مدل و در نتیجه ناکارآیی مدل وجود دارد. شکل تابعی صحیح را می‌توان براساس مشاهدات تجربی و مقدار ضریب تعیین[1]، R^2، تعیین نمود (با تغییر شکل تابعی با متغیرهای مستقل یکسان و مقایسه R^2 حاصل از مدل‌های مختلف، مشروط بر این‌که شکل تابعی y تغییر نکند).

همان‌گونه که می‌دانید متداولترین روش برآورد در مدل‌های رگرسیون خطی کلاسیک (CLRM)[2]، روش OLS است. برآوردگر OLS برای β مقداری است که مجموع توان‌های دوم باقی‌مانده‌ها را می‌نیمم می‌کند که در صورت برقراری فرضیات مدل، این برآوردگر، بهترین برآوردگر خطی نااریب (BLUE)[3] برای β است، یعنی در بین کلیه برآوردگرهای خطی نااریب ممکن، دارای کمترین واریانس است.

در مدل رگرسیون خطی کلاسیک، برآوردگرهای OLS پارامترها در صورتی BLUE هستند که: متغیرهای مستقل با باقی‌مانده‌ها همبسته نباشند؛ باقی‌مانده‌ها دارای امید ریاضی صفر و واریانس ثابت باشند؛ بین باقی‌مانده‌ها همبستگی وجود نداشته باشد؛ و بین متغیرهای مستقل نیز همبستگی خطی قابل توجه وجود نداشته باشد. همچنین در صورتی که توزیع باقی‌مانده‌ها نرمال باشد، توزیع پارامترها نیز نرمال خواهد بود در نتیجه به سادگی می‌توان فروض مربوط به معنی‌دار بودن پارامترها را آزمون و بازه اطمینان آن‌ها را برآورد نمود. در این حالت، برآوردگرهای OLS، برآوردگرهای ماکسیمم درست‌نمایی[4] نیز هستند.

عدم برقراری هر یک از این فروض ممکن است باعث ناکارآیی مدل و گمراه کننده بودن نتایج حاصل از آن شود که در ادامه در مورد هر یک از این مفروضات و مشکلات ناشی از عدم برقراری آن به طور خلاصه مطالبی عنوان می‌شود.

همبستگی خطی کامل بین متغیرهای مستقل مدل یا به بیان دیگر وجود هم‌خطی کامل، می‌تواند مشکلات قابل توجهی ایجاد کند. هرچند وجود هم‌خطی کامل در عمل بسیار غیرمحتمل است اما هم خطی قابل توجه به‌خصوص در صورتی که داده‌ها حاصل از نمونه‌گیری باشند، زیاد دیده می‌شود. در حالت هم‌خطی کامل، به‌دست آوردن برآوردهای یکتا از همه پارامترها میسر نیست اما در هم خطی قابل توجه، برآوردهای OLS همچنان نااریب اما واریانس آن‌ها بزرگ است که موجب کاهش دقت برآوردها، بزرگی بازه اطمینان، قبول نادرست فرض صفر بودن ضرایب یا معنی‌دار نشدن پارامترهایی که در تئوری انتظار داریم معنی‌دار شوند، می‌گردد. علاوه بر آن هرچند

---

[1] Coefficient of Determination
[2] Classic Linear Regression Models
[3] Best Linear Unbiased Estimator
[4] Maximum Likelihood Estimators

مقدار R^2 قابل توجه است اما ضریب تعداد زیادی از متغیرهای مستقل معنی‌دار نمی‌شود. از طرفی برآوردهای OLS بسیار حساس به داده‌ها و ناپایدار بوده و احتمال به دست آوردن علامت نادرست برای ضرایب (با توجه به تئوری اقتصادی) وجود خواهد داشت و در نهایت ارزیابی سهم هریک از متغیرهای مستقل در بیان تغییرات متغیر وابسته (y)، مشکل است.

همان‌طور که اشاره شد در داده‌های حاصل از نمونه‌گیری، بحث اصلی بر سر وجود یا عدم وجود هم خطی نیست بلکه تعیین درجه هم‌خطی است که حائز اهمیت می‌باشد. برای تشخیص قابل توجه بودن میزان هم خطی بین متغیرهای مستقل مدل، لازم است مواردی از قبیل: مقدار R^2 قابل توجه $(R^2 > ۰/۸)$ علی‌رغم اندک بودن تعداد ضرایب معنی‌دار در مدل؛ و بالا بودن ضریب همبستگی بین دو بدوی متغیرهای مستقل (حداقل یک مورد بیش از ۰/۸)، مورد بررسی قرار گیرد. علاوه بر آن، بررسی همبستگی جزئی بین متغیرهای مستقل و ساخت مدل‌های رگرسیونی بین هر یک از x ها با بقیه x ها و بررسی مقدار R^2 آن‌ها، می‌تواند در تشخیص قابل توجه بودن هم‌خطی، کمک کننده باشد. فاکتور تورم واریانس (VIF)[1]، شاخصی از میزان هم‌خطی است که مقدار ۱ آن نشانه نبودن هم‌خطی و بسیار بزرگ بودن آن نشان‌دهنده هم‌خطی کامل است.

برای تشخیص هم‌خطی در عمل، چون راه حل قطعی برای تشخیص آن وجود ندارد و با بررسی‌های مختلف تنها می‌توان تا حدودی به وجود آن پی برد، شاید بهتر باشد حداقل چند مورد از موارد فوق مورد بررسی قرار گیرد تا از وجود و اهمیت میزان هم‌خطی اطمینان بیشتری حاصل شود. به طور کلی وجود هم‌خطی قابل توجه در صورتی که برآورد هر یک از پارامترهای مدل به تنهایی حائز اهمیت باشد، خوب نیست چون همان‌گونه که قبلاً ذکر شد در صورت وجود هم‌خطی قابل توجه، دقت برآوردها کم است. اما در صورتی‌که از مدل فقط برای پیش‌بینی استفاده شود، مشروط بر عدم تغییر ارتباط بین متغیرهای مستقل تا زمان مورد نظر در پیش‌بینی، وجود هم‌خطی، مشکل اساسی ایجاد نمی‌کند. پس از تشخیص هم‌خطی، برای کاهش یا رفع آن راه‌های مختلفی را بسته به مورد می‌توان انتخاب نمود که عبارتند از: تغییر مدل تا جایی که فرضیات اقتصادی مجاز می‌داند، استفاده از اطلاعات دقیق‌تر قبلی در مورد پارامترهای مورد نظر، تغییر متغیر، ترکیب داده‌های پانل و مقطعی، آنالیز مؤلفه‌های اصلی[2] یا رگرسیون برجسته[3].

فرضیه دیگری که لازمه کارآیی برآوردگرهای OLS می‌باشد، یکسان بودن واریانس باقی‌مانده‌ها، برای مشاهدات مختلف است. علت پدیده ناهم‌واریانسی، همبستگی باقی‌مانده‌ها با حداقل یکی از متغیرهای مستقل موجود در مدل و یا عدم حضور یکی از متغیرهای مستقل مهم،

---

[1] Variance Inflation Factor
[2] Principal Component Analysis
[3] Ridge Regression

در مدل است. در صورت وجود ناهم‌واریانسی، هرچند برآوردهای OLS هم‌چنان نااریب و خطی هستند اما حتی با افزایش تعداد نمونه، کارا نیستند یعنی دیگر حداقل واریانس را ندارند. علاوه بر آن فرمول معمول در برآورد واریانس پارامترها اریب بوده و در نتیجه استفاده از روش‌های قراردادی در آزمون فرض به نتایج گمراه کننده منجر می‌شود. با توجه به این‌که در داده‌های مقطعی اغلب وجود ناهم‌واریانسی مورد انتظار است، بررسی آن حائز اهمیت ویژه می‌باشد. تشخیص ناهم‌واریانسی نیز همانند هم‌خطی راه‌حل قطعی آماری ندارد، چون معمولاً داده‌ها حاصل از یک نمونه‌گیری است و واریانس واقعی باقی‌مانده‌ها به‌ازای x های مشاهدات مختلف، مشخص نیست (چون معمولاً برای هر x فقط یک مقدار y را داریم). ابزارهایی که تا حدودی در تشخیص ناهم‌واریانسی کمک کننده هستند، عمدتاً مبتنی بر بررسی رابطه بین باقی‌مانده‌ها و هر یک از متغیرهای مستقل یا مقادیر پیش‌بینی (در صورتی‌که تعداد متغیرهای مستقل زیاد باشد) از طریق نمودار یا ساخت مدل و آزمون معنی‌دار بودن پارامترهای مدل یا ضریب تعیین آن می‌باشد که از جمله اهم موارد آن می‌توان به رسم نمودار باقی‌مانده‌ها یا مجذور آن‌ها در مقابل هر یک از متغیرهای مستقل (یا پیش‌بینی‌ها که خود ترکیب خطی از متغیرهای مستقل هستند)، اشاره نمود. در صورت مشاهده روند خاصی در این نمودار، می‌توان تا حدودی به وجود ناهم‌واریانسی پی برد. آزمون‌های مهم در این زمینه شامل آزمون پارک[1]، آزمون گِلج سِر[2]، آزمون وایت[3] و ... است. پس از تشخیص وجود ناهم‌واریانسی در مدل، چنانچه مشاهدات در سطح کل جامعه باشد و در نتیجه مقادیر واریانس باقی‌مانده‌ها (σ_i^2) مشخص باشد با استفاده از روش WLS[4] و با وزن‌های $\frac{1}{\sigma_i}$ ، مشکل ناهم‌واریانسی مرتفع می‌شود. اما چون در عمل ما معمولاً مدل را براساس داده‌های نمونه می‌سازیم و مقادیر (σ_i^2) را نمی‌دانیم، مشکل ناهم‌واریانسی را با تغییر متغیر متناسب حل می‌کنیم. به این ترتیب که در حالت رگرسیون دو متغیره، چنانچه براساس نمودار باقی‌مانده‌ها برحسب x ، $|e_i|$ با افزایش x ، افزایش یابد، تغییر متغیر مناسب $\frac{1}{\sqrt{x_i}}$ است که در هر دو طرف معادله رگرسیون ضرب می‌شود و معادله حاصل بدون عرض از مبدا برآورد می‌شود. البته در صورتی‌که ارتباط باقی‌مانده‌ها و متغیر مستقل از درجه ۲ باشد، هر دو طرف مدل را در $\frac{1}{x_i}$ ضرب می‌کنیم. در حالت چند متغیره نیز چنانچه با توجه به نمودار باقی‌مانده‌ها، برحسب هر یک از x ها، بتوان مناسب‌ترین x را برای تغییر متغیر تعیین نمود، همانند فوق عمل می‌شود و در صورتی که $|e_i|$ با

[1] Park Test
[2] Glejser Test
[3] White Test
[4] Weighting Least Squares

بیش از یک x هم‌بسته باشد، از $\hat{y}_i$ به‌عنوان مبنای تغییر متغیر استفاده می‌کنیم یعنی مدل را بر $\sqrt{\hat{y}_i}$ تقسیم می‌کنیم. لازم به ذکر است که پس از برآورد مدل اخیر، دو طرف آن در وزن مشخص شده ضرب می‌شود.

راه دیگری که برای کاهش ریسک ناهم‌واریانسی وجود دارد، تبدیل کلیه متغیرهای مدل به شکل لگاریتمی است (مشروط بر این‌که از نظر تئوری اقتصادی این تبدیل موجه باشد). در این حالت چون دامنه تغییرات مقادیر به‌جای مبنای ۱۰، مبنای ۲ می‌شود، میزان تغییرات بسیار کمتر خواهد بود.

برای برقراری فرض ناهم‌بسته بودن متغیرهای مستقل با باقی‌مانده‌های مدل، کافی است متغیرهای مستقل(xها) تصادفی نباشند یعنی مقادیر آن اعداد ثابت باشند و نه نمونه‌ای از کلیه مقادیر ممکن x ها.

فرض عدم وجود همبستگی بین باقی‌مانده‌ها نیز معمولا در داده‌های زمانی دیده می‌شود که به دلیل سیکل‌های موجود در این نوع داده‌ها به وجود می‌آید. البته در پاره‌ای از موارد، نامناسب بودن مدل برازش شده از جنبه عدم حضور متغیرهای مهم یا فرم تابعی غلط نیز می‌تواند موجب ایجاد خود همبستگی شود. با وجود خود همبستگی، برآورد OLS پارامترها همچنان نااریب و خطی است اما کارا نیست. واریانس برآوردها اریب و اغلب کم برآوردی قابل توجهی دارد در نتیجه مقدار آماره آزمون‌های t وF به صورت غیرواقعی بزرگ شده که ممکن است به غلط منجر به معنی‌دار شدن پارامترها گردد. اریبی برآورد واریانس نیز ممکن است منجر به غیرواقعی شدن مقدار R^2 گردد. همچنین واریانس مقادیر پیش‌بینی نیز ممکن است ناکارآ باشد. بنا بر این همانند ناهم‌واریانسی، خود همبستگی نیز اشکالات اساسی ایجاد می‌کند که باید به دقت مورد بررسی قرار گیرد. برای تشخیص خود همبستگی از نمودار باقی‌مانده‌ها در مقابل زمان و آزمون‌هایی چون دوربین واتسون[1] استفاده می‌شود. چنانچه در نمودار مزبور، همبستگی خاصی مشاهده شود و یا مقادیر آماره‌ی آزمون فوق نزدیک به ٤ (خود همبستگی منفی) یا صفر (خود همبستگی مثبت) باشد، مشکل خود همبستگی وجود دارد. مقادیر نزدیک به ۲ آماره دوربین واتسون، دلالت بر عدم وجود خود همبستگی دارد.

برای حل مشکل خود همبستگی پس از تشخیص آن، لازم است ابتدا میزان همبستگی بین باقی‌مانده‌ها برآورد گردد و با استفاده از آن، تغییر متغیر مناسب انجام گیرد.

[1] Durbin-Watson