

## ارائه‌ی نمودار کنترل مبتنی بر مدل رگرسیون ریج در حضور هم‌خطی چندگانه

شروین اسدزاده<sup>۱\*</sup>، ولی‌اله مهدوی<sup>۲</sup>

۱. استادیار گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد تهران شمال، تهران، ایران

۲. دانشجوی کارشناسی ارشد گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد تهران شمال، تهران، ایران

### خلاصه

نمودارهای کنترل انتخاب عامل انحراف مهم‌ترین ابزار کنترل کیفیت آماری در فرایندهای چندمرحله‌ای هستند. استقرار این نمودارهای کنترل که از مدل‌های رگرسیون و باقیمانده‌ی آن‌ها برای حذف تأثیر مراحل قبل بر روی مرحله‌ی جاری استفاده می‌کنند، دارای شرط‌های مهمی از جمله استقلال خطی بین متغیرهای ورودی تأثیرگذار هستند. در صورت فقدان این شرط که در دنیای واقعی نیز بسیار رایج بوده و به آن هم‌خطی چندگانه می‌گویند، مدل رگرسیونی حاصل دچار پدیده‌ی تورم واریانس در ضرایب مدل می‌شود. در نتیجه، این عدم اطمینان در برآورد ضرایب رگرسیونی منجر به اشکالات اساسی در عملکرد نمودارهای کنترل انتخاب عامل انحراف سنتی می‌شود. برای حل این مشکل روش‌های آماری و اجرایی مختلفی وجود دارد که استفاده از رگرسیون ریج یکی از مؤثرترین رویکردهاست. در این مقاله بعد از مدل‌سازی و ارائه‌ی نمودار کنترل جدید مبتنی بر رگرسیون ریج، شبیه‌سازی گسترده به منظور ارزیابی عملکرد رویکرد پایش پیشنهادی صورت پذیرفته است و نتایج حاصل از آن با نمودار کنترل مبتنی بر رگرسیون خطی معمولی موجود در ادبیات مقایسه شده است. شواهد نمایانگر برتری کامل رویکرد پایش پیشنهادی نسبت به رویکرد موجود در حضور هم‌خطی چندگانه است.

### اطلاعات مقاله

تاریخچه مقاله:

دریافت ۱۳۹۷/۹/۱۹

پذیرش ۱۳۹۸/۱۰/۲۱

کلمات کلیدی:

فرایندهای چندمرحله‌ای

نمودار کنترل انتخاب عامل

انحراف

هم‌خطی چندگانه

تورم واریانس

رگرسیون ریج

### ۱- مقدمه

مدل‌های رگرسیونی شناخته می‌شود [۱]. در روش پایش مبتنی بر مدل، از دانش فرایند برای برقراری ارتباط مابین مشخصه کیفی موردنظر و هر نوع متغیر تصادفی کمکی مؤثر بر آن استفاده می‌شود. مهم‌ترین مزایای روش پایش مبتنی بر مدل، افزایش حساسیت نمودار کنترل نسبت به تغییر در پارامترهای فرایند و همچنین بهبود قابلیت تشخیص نمودار که به‌ویژه برای ریشه‌یابی مسائل فرایند حائز اهمیت است برشمرده شده است. این تکنیک خصوصاً در زمانی که خاصیت آبخاری در فرایندهای چندمرحله‌ای به چشم می‌خورد، کاربرد فراوانی دارد.

وید و وودال [۲] به‌مرور و تحلیل نمودار انتخاب عامل انحراف مبتنی بر رگرسیون خطی ساده و بررسی مفاهیم پایه و اساسی در فرایندهای چندمرحله‌ای با مشخصه‌ی کیفی نرمال پرداخته‌اند. اسکینر و همکاران

امروزه در صنایع تولیدی و خدماتی، به‌ندرت فرایند تک‌مرحله‌ای و ساده وجود دارد و بیشتر فرایندهای تولید محصولات و ارائه‌ی خدمات، شامل چندین مرحله بوده که از هم مستقل نیستند. این بدان معناست که تغییر در مشخصه کیفی ورودی بر وضعیت تحت کنترل بعضی یا همه متغیرهای کیفی خروجی اثر می‌گذارد. ویژگی ذکرشده خاصیت آبخاری نام دارد که ویژگی اصلی فرایندهای چندمرحله‌ای است. در پایش فرایندهای چندمرحله‌ای نیز یافتن منشاء خارج از کنترل، مسأله‌ی مهمی است و به این منظور ابزارها و روش‌های آماری مختلفی پیشنهاد شده است. یکی از کاراترین رویکردها، به‌کارگیری نمودارهای کنترل انتخاب عامل انحراف است که به‌عنوان رویکرد پایش مبتنی بر

\* نویسنده مسئول: شروین اسدزاده

تلفن ثابت: ۰۲۱-۷۷۰۰۹۷۴۸ ؛ پست الکترونیکی: [sh\\_asadzadeh@iau-tnb.ac.ir](mailto:sh_asadzadeh@iau-tnb.ac.ir)

و مدل‌های شکنندگی به‌منظور انتخاب و لحاظ کردن تأثیر متغیرهای غیرقابل‌اندازه‌گیری، پایش داده‌های قابلیت اطمینان خودهمبسته را مورد مطالعه قرار دادند.

اما به‌کارگیری مدل‌های رگرسیون دارای پیش‌فرض‌ها و شروطی است که در دنیای واقعی همیشه برقرار نیست و این مسأله منجر به اُفت کارایی و یا غیرقابل استفاده شدن این ابزار مهم می‌شود. یکی از این شروط که تخطی از آن به‌ویژه مدل رگرسیون خطی را با مشکل مواجه می‌کند، استقلال متغیرهای ورودی و یا تأثیرگذار است. در دنیای واقعی گاهی شرایطی پیش می‌آید که این متغیرها دارای وابستگی خطی هستند. این هم‌خطی می‌تواند چهار منشاء اصلی داشته باشد. شیوه‌ی غیردقیق جمع‌آوری داده‌ها، ارتباط ذاتی بین متغیرهای مدل در جامعه، اختصاص دادن مدل اشتباه و مدل‌سازی با تعداد متغیرهای رگرسیونی بیش‌ازحد، می‌توانند منجر به هم‌خطی شوند [۱۷]. به‌رحال این پدیده منجر به بروز خطاهای اساسی در مدل می‌شود که مهم‌ترین آن تورم در واریانس ضرایب برآورد شده است. این تورم واریانس منجر به عدم اطمینان در ضرایب برآورد شده و دقت پایین آن‌ها می‌شود و این خطر بزرگی برای ناکارآمد شدن نمودارهای کنترل حاصل از آن خواهد بود.

با توجه به‌مرور ادبیات، پایش فرایندهای چندمرحله‌ای در حضور هم‌خطی چندگانه تاکنون انجام نگرفته است، درحالی‌که وجود این شرایط به‌ویژه در فرایندهای شیمیایی بسیار رایج است [۱۷]. هدف این مقاله ارائه‌ی رویکرد پایش برای این فرایندها به کمک رگرسیون رنج است تا مدل‌سازی با قابلیت اعتماد بیشتر و در نتیجه نمودار انتخاب عامل انحراف با دقت بیشتری به‌دست آید.

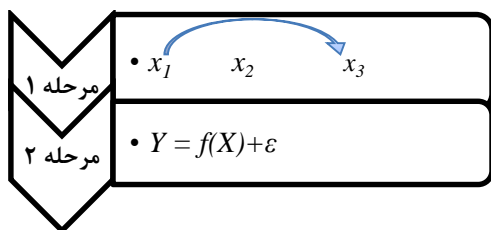
ساختار مقاله بدین‌صورت طراحی شده است: در بخش ۲ به بیان مسأله پرداخته شده است. در بخش ۳ مفاهیم هم‌خطی چندگانه و رگرسیون رنج شرح داده شده و در بخش ۴ روش پایش پیشنهادی مطرح گردیده است. در بخش ۵ به ارزیابی عملکرد نمودارهای کنترل و تحلیل آن‌ها پرداخته و در بخش ۶ نتیجه‌گیری و پیشنهادهایی برای پژوهش‌های آتی مطرح شده است.

## ۲- بیان مسأله

اگر یک فرایند دارای دو مرحله‌ی کاری بهم وابسته باشد؛ متغیر تصادفی  $X$  نشان‌دهنده‌ی مشخصه‌ی کیفی مرحله‌ی اول فرایند بوده که در زمان تحت کنترل بودن آن مرحله، متغیر تصادفی فوق از توزیع نرمال با پارامترهای میانگین  $\mu_x$  و انحراف معیار  $\sigma_x$  پیروی می‌کند. همچنین فرض شود متغیر تصادفی  $Y$  خروجی مرحله‌ی دوم باشد که از توزیع نرمال پیروی می‌کند. به‌این‌ترتیب نمونه‌ی تصادفی  $(X_i, Y_i)$  در پایان مرحله‌ی دوم این فرایند به دست می‌آید که  $i$  شمارشگر دفعات نمونه‌گیری است. برای ارائه‌ی نمودار انتخاب عامل انحراف، نخست باید

[۳] یک فرایند دو مرحله‌ای را در شرایطی که مشخصه‌ی کیفی در مرحله‌ی دوم از نوع پواسون است را مورد بررسی قرار دادند و جهت کنترل فرایند، از آماره‌ی باقیمانده‌های انحراف، استفاده کردند. جرکاپورن و همکاران [۴] با این فرض که متغیر خروجی از توزیع گاما پیروی می‌کند، از مدل خطی تعمیم‌یافته‌ی استوار استفاده کرده‌اند. در این مقاله از باقیمانده‌های انحراف که یک آماره‌ی نسبت درستنمایی است، جهت کنترل فرایند استفاده شده است. اسدزاده و همکاران [۵] به ارائه‌ی رویکردی استوار به‌منظور پایش فرایندهای چندمرحله‌ای مشخصه‌ی کیفی نرمال در حضور داده‌های پرت پرداخته‌اند. نورالسنا و شکاری [۶] نمودارهای کنترل انتخاب عامل انحراف<sup>۱</sup> با اندازه‌ی نمونه و فاصله‌های اندازه‌گیری تطبیقی ارائه کرده‌اند. عسگری و همکاران [۷] آماره‌ی باقیمانده استاندارد را بر مبنای یک تابع ارتباط جدید زمانی که متغیر پاسخ در مرحله‌ی دوم از توزیع پواسون پیروی می‌کند، پیشنهاد کردند و برای پایش آماره از نمودارهای کنترل شوهارت و میانگین متحرک موزون نمایی استفاده نمودند. امیری و همکاران [۸] پایش فرایند دو مرحله‌ای درحالی‌که مرحله‌ی دوم دارای توزیع بینم و آماره‌ی این آزمون مقادیر حاصل از مدل خطی تعمیم‌یافته است را بررسی کرده و برای برقراری ارتباط بین مشخصه کیفی مراحل اول و دوم یک تابع ارتباط جدید ارائه نمودند. اسدزاده و همکاران [۹] فرایندهای چندمرحله‌ای را با حضور متغیرهای تأثیرگذار مشاهده‌شده و مشاهده‌نشده مورد مطالعه قرار دادند و از تلفیق مدل‌های رگرسیونی مخاطره متناسب کاکس و مدل‌های شکنندگی<sup>۲</sup> استفاده کردند. نان پن و همکاران [۱۰] به دنبال گسترش یک نمودار کنترل جدید مناسب برای بازرسی کیفیت فرایند سیستم‌های تولید چندمرحله‌ای با در نظر گرفتن خروجی‌های خودهمبسته بوده‌اند. خدمتی و نیازی [۱۱] یک رویکرد کنترل جدید برای پایش پروفایل‌های ساده‌ی خطی در فرایندهای چندمرحله‌ای پیشنهاد داده‌اند. سپس این تحقیق را بهرامی و همکاران [۱۲] برای پروفایل‌های پیچیده‌تر چندمتغیره در فرایندهای چندمرحله‌ای توسعه دادند. کیم و همکاران [۱۳] فرایندهای چندمرحله‌ای را برای شرایطی که مشاهدات دارای خودهمبستگی قابل‌توجهی هستند مورد مطالعه قرار داده و برای پایش آن نمودار جدیدی براساس نمودار میانگین متحرک موزون نمایی چندمتغیره ارائه نموده و رویکرد خود را با رویکردهای موجود مقایسه کردند. اسدزاده و کیادلیری [۱۴] به ارائه‌ی نمودارهای کنترل جمع تجمعی و میانگین متحرک موزون نمایی به‌منظور پایش پارامتر مقیاس داده‌های قابلیت اطمینان با توزیع ویبول در حضور سانسورشدگی نوع دو در فرایندهایی با خاصیت آبخاری<sup>۳</sup> پرداخته‌اند. کیم [۱۵] طراحی یک نمودار کنترل میانگین متحرک موزون نمایی چندمتغیره با در نظر گرفتن روش اجرایی انتخاب متغیرها را مورد بررسی قرار داده است. در نهایت، کشاورز و همکاران [۱۶] به کمک تلفیق مدل‌های رگرسیونی زمان شکست تسریع یافته

متغیرها باید دائم آن‌ها را پایش کرد و برای استفاده از نمودارهای انتخاب عامل انحراف باید ابتدا روابط بین آن‌ها را مدل‌سازی نمود. اما مدل رگرسیون این مسأله را نمی‌توان طبق روش رایج حداقل مربعات به‌دست آورد چرا که استقلال خطی بین متغیرهای ابتدایی ( $X_j$ ) برقرار نیست. در شکل (۱) روابط موجود بین متغیرها نشان داده شده است. توضیح این‌که هرچه مدت زمان ( $X_1$ ) بیشتری در کوره صرف شود دمای آن ( $X_3$ ) نیز کاهش می‌یابد. این عدم استقلال منجر می‌شود که رویکردهای معمول، ناتوان از ارائه‌ی مدل رگرسیونی دقیق بین آن‌ها بوده و استفاده از رگرسیون حداقل مربعات دچار اختلال گردد. بنابراین نمی‌توان همانند قبل از نمودار انتخاب عامل انحراف سنتی برای پایش فرایند استفاده نمود.



شکل (۱): فرایند دومرحله‌ای با ارتباط خطی بین متغیرهای مرحله‌ی اول

پیشنهاد این پژوهش استفاده از رویکرد رگرسیون ریج برای حل این مسأله است چرا که در این نوع رویکرد، پیش‌شرط استقلال متغیرهای مرحله‌ی اول ضروری نیست و باوجود اینکه نسبت به رویکرد معمول در برآورد پارامترها کمی اریب است، ولی در نهایت مدل به‌دست‌آمده صحیح‌تر و بیشتر قابل‌اعتماد خواهد بود. در ادامه به جزئیات بیشتر این مسأله پرداخته می‌شود.

### ۳- رگرسیون ریج و تورم واریانس

به‌طور کلی  $p$  متغیر را هم‌خط<sup>۲</sup> گویند اگر بردارهای متناظر با هر یک از این متغیرها بر روی یک زیرفضا قرار گیرند طوری که دارای بُعد کمتر از  $p$  باشد، یعنی یکی از بردارها ترکیب خطی از بردارهای دیگر شود. هنگامی که بین متغیرهای مستقل در رگرسیون چندگانه وابستگی وجود داشته باشد، بین آن‌ها هم‌خطی چندگانه وجود دارد [۱۷].

اگر  $j$  آمین ستون ماتریس  $X$  با  $x_j$  نمایش داده شود، در این صورت؛ اگر  $X = [x_1, x_2, \dots, x_p]$  که  $x_j$  شامل  $n$  سطح از  $j$  آمین متغیر رگرسیونی است. می‌توان هم‌خطی چندگانه را به‌صورت وابستگی خطی ستون‌های ماتریس  $X$  نیز تعریف کرد. بردارهای  $x_1, x_2, \dots, x_p$  به‌صورت خطی وابسته‌اند هرگاه مجموعه‌ای از مقادیر ثابت  $t_1, t_2, \dots, t_p$  که همگی صفر نیستند وجود داشته باشند، طوری که:

$$\sum_{j=1}^p t_j x_j = 0 \quad (۶)$$

رابطه‌ی بین دو مشخصه‌ی  $X_i$  و  $Y_i$  مشخص شود. به‌طور کلی چنین رابطه‌ی بین این دو متغیر حاکم است:

$$Y_i = f(X_i) + \varepsilon_i \quad (۱)$$

به‌طوری که در زمان تحت کنترل آماری بودن مرحله‌ی دوم فرایند، خطای تصادفی  $\varepsilon_i$  نرمال مستقل است:

$$\varepsilon_i \sim N(0, \sigma^2); \text{ i. i. d} \quad (۲)$$

حال اگر رابطه‌ی بین متغیرها از مدل رگرسیون خطی چندگانه پیروی کند آنگاه بر این اساس در  $n$  بار نمونه‌گیری، برای نمونه‌گیری  $i$  ام می‌توان نوشت:

$$Y_i = \beta_0 + \beta_j X_{ij} + \varepsilon_i; \quad i = 1, 2, \dots, n \quad \text{و} \quad j = 1, 2, \dots, p \quad (۳)$$

$\beta_0$  و  $\beta_j$  مقادیر ثابت رگرسیون هستند و  $j$  شمارشگر متغیرهای کیفی در مرحله‌ی اول بوده که تعداد آن‌ها  $p$  تا است. برای پایش مرحله‌ی دوم باید تأثیر مرحله‌ی قبل را از آن حذف کرد. برای این کار روی اختلاف مقادیر نمونه با آنچه مدل رگرسیون ارائه کرده است تمرکز می‌شود. یعنی باقیمانده‌های مدل، نشانگر عملکرد در مرحله‌ی دوم است.

$$e_i = Y_i - \hat{Y}_i \quad (۴)$$

که در اینجا  $\hat{Y}_i$  مقدار تخمین زده شده برای  $Y_i$  بشرط  $X_i$  است.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_j X_{ij} \quad i = 1, 2, \dots, n \quad \text{و} \quad j = 1, 2, \dots, p \quad (۵)$$

نمودار انتخاب عامل انحراف یک نمودار کنترل شوهارت و یا دیگر انواع نمودار است که مقدار باقیمانده‌ی بالا را کنترل می‌کند [۱]. اما چنانکه گفته شد شرط ابتدایی آن، استقلال بین متغیرهای کیفی در مرحله‌ی اول است. وقتی استقلال خطی بین  $X_j$ ها به‌عنوان پیش‌نیاز این نمودار، تأمین نشود و نوعی از وابستگی خطی بین متغیرهای مرحله‌ی اول وجود داشته باشد، مدل رگرسیونی به‌دست‌آمده دچار اختلالاتی شده که منجر به برآورد نادرست مدل و در نهایت نمودار کنترل نامطمئن می‌گردد. این اختلال که تورم واریانس<sup>۱</sup> نامیده می‌شود، در ضرایب برآورد شده‌ی  $\hat{\beta}$  واریانس بالایی ایجاد کرده و امکان برآورد نادرست مدل را افزایش می‌دهد. همچنین در هر بار نمونه‌گیری ممکن است ضرایب متفاوتی را ارائه کند [۱۷]. از آنجاکه این خصوصیت در صنایع به‌طور مکرر دیده می‌شود، باید برای پایش فرایندهای چندمرحله‌ای، این مسأله را موردتوجه قرار داده تا به صحت مدل‌سازی خدشه‌ای وارد نباشد. بر این اساس در صورت بروز شرایط هم‌خطی، نباید رویکردهای موجود را بکار گرفت.

یک مثال که این شرایط در آن مشهود بوده عبارت است از فرایند تولید استیلن از هپتان نرمال در مجاورت هیدروژن که در داخل یک کوره قرار می‌گیرند. متغیرهای کیفی در مرحله‌ی نخست عبارت‌اند از مدت زمان صرف‌شده در کوره ( $X_1$ )، نسبت هیدروژن به هپتان نرمال ( $X_2$ ) و درجه حرارت رآکتور ( $X_3$ ) و نیز متغیر کیفی مرحله‌ی دوم؛ درصد تبدیل هپتان نرمال به استیلن (۷) است [۱۷]. برای کنترل این

عامل تورم واریانس وجود دارد که از این بین عامل تورم واریانس، محدودیت‌های زوجی ماتریس همبستگی را نداشته و محاسبه و تحلیل آن از شاخص مقادیر ویژه نیز بسیار ساده‌تر است [۱۷]. بنابراین در این پژوهش از عامل تورم واریانس با تعریف زیر استفاده می‌شود:

$$VIF_j = \frac{1}{1-R_j^2} \quad (14)$$

که در آن  $R_j^2$  ضریب تعیین چندگانه است که از رگرسیون  $x_j$  نسبت به دیگر متغیرهای رگرسیونی به دست می‌آید. اگر  $x_j$  تقریباً به‌طور خطی با بعضی متغیرهای رگرسیونی دیگر ارتباط داشته باشد در این صورت  $R_j^2$  به عدد یک نزدیک می‌شود و  $VIF_j$  بزرگ‌تر خواهد شد. مرسوم است که وقتی  $VIF$  بالاتر از ۱۰ باشد به آن توجه کرده و مشکلات هم‌خطی را در آن قابل بررسی اعلام می‌کنند [۱۷].

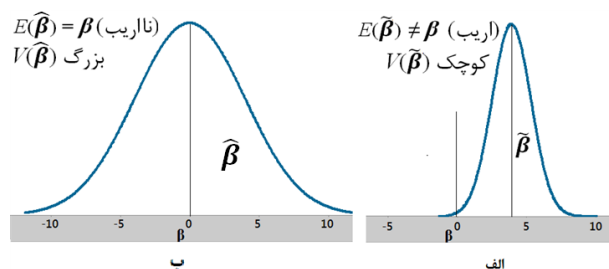
اما برای رفع این مشکل می‌توان از رگرسیون ریج استفاده کرد. اگر از ناریب بودن برآوردگر  $\hat{\beta}_j$  چشم‌پوشی شود، می‌توان یک برآوردگر اریب برای  $\beta_j$  مانند  $\tilde{\beta}_j$  پیدا نمود که دارای واریانس کمتری نسبت به برآوردگر ناریب  $\hat{\beta}_j$  باشد. در این حالت میانگین مربعات خطای برآوردگر جدید به این صورت تعریف می‌شود:

$$MSE(\tilde{\beta}_j) = E(\tilde{\beta}_j - \beta_j)^2 = V(\tilde{\beta}_j) + [E(\tilde{\beta}_j) - \beta_j]^2 \quad (15)$$

یعنی؛

$$MSE(\tilde{\beta}_j) = V(\tilde{\beta}_j) + (\text{bias in } \tilde{\beta}_j)^2 \quad (16)$$

که  $MSE$  امید ریاضی مربع فاصله‌ی  $\tilde{\beta}_j$  تا  $\beta_j$  است. پس با پذیرفتن مقدار کمی اریبی برای  $\tilde{\beta}_j$ ، واریانس آن را می‌توان طوری کوچک کرد که  $MSE$  برای  $\tilde{\beta}_j$  از واریانس برآوردگر ناریب  $\hat{\beta}_j$  کمتر باشد. در نتیجه مقدار برآورد در این حالت پایدارتر خواهد بود و تحت تأثیر تغییرات جزئی در داده‌ها واقع نمی‌شوند.



شکل (۲): توزیع نمونه‌ای برآوردگرها برای ضرایب  $\beta$  (الف) اریب و (ب) ناریب

شکل (۲-الف) وضعیتی را نشان می‌دهد که در آن واریانس برآوردگر اریب در حد قابل توجهی کوچک‌تر از واریانس برآوردگر ناریب در بخش (ب) است. در نتیجه فاصله‌ی اطمینان  $\beta$  کوچک‌تر شده، یعنی ضرایب پایدارتر خواهند شد.

در رگرسیون ریج علاوه بر حداقل‌سازی مربعات خطا، طول بردار ضرایب که دچار تورم واریانس هستند نیز حداقل و یا کمتر می‌شود:

اگر رابطه‌ی (۶) برای زیرمجموعه‌ای از ستون‌های  $X$  برقرار باشد، در این صورت رتبه‌ی ماتریس  $X^T X$  کمتر از  $p$  بوده و  $(X^T X)^{-1}$  وجود نخواهد داشت چرا که درمیان این ماتریس صفر شده و در تعریف معکوس، در مخرج کسر صفر خواهیم داشت. وقتی این وابستگی وجود داشته باشد، گفته می‌شود هم‌خطی چندگانه وجود دارد. زمانی که ماتریس  $(X^T X)^{-1}$  وجود ندارد رگرسیون را نمی‌توان برآورد کرد. اما این هم‌خطی همیشه به‌طور کامل نبوده و نسبی است یعنی ضریب همبستگی تنها ۰ یا ۱ نبوده و می‌تواند عددی بین این‌ها باشد. در این شرایط میزان اختلال در برآورد ضرایب رگرسیون باید اندازه‌گیری شود [۱۷].

فرض می‌شود دو متغیر مستقل  $x_1$ ،  $x_2$  وجود داشته باشند. با این فرض که  $x_1$ ،  $x_2$  و  $y$  استاندارد شده باشند، مدل به این صورت خواهد بود؛

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (7)$$

معادلات نرمال حداقل مربعات عبارتند از:

$$(X^T X) \hat{\beta} = X^T y \quad (8)$$

با توجه به اینکه متغیرهای مستقل و پاسخ استاندارد شده‌اند، ماتریس  $X^T X$ ، به یک ماتریس از ضرایب همبستگی بین متغیرهای رگرسیونی و  $X^T y$  به یک بردار از ضرایب همبستگی بین متغیرهای رگرسیونی و پاسخ تبدیل می‌شود یعنی؛

$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix} \quad (9)$$

که در آن  $r_{12}$  همبستگی ساده بین  $x_1$  و  $x_2$  و  $r_{1y}$  همبستگی ساده بین  $x_1$  و  $y$  است. معکوس ماتریس  $(X^T X)$  عبارت است از:

$$(X^T X)^{-1} = \begin{bmatrix} \frac{1}{(1-r_{12}^2)} & \frac{-r_{12}}{(1-r_{12}^2)} \\ \frac{-r_{12}}{(1-r_{12}^2)} & \frac{1}{(1-r_{12}^2)} \end{bmatrix} \quad (10)$$

و برآوردهای ضرایب رگرسیون عبارت است از:

$$\begin{aligned} \hat{\beta}_1 &= \frac{r_{1y} - r_{12} r_{2y}}{(1-r_{12}^2)} \\ \hat{\beta}_2 &= \frac{r_{2y} - r_{12} r_{1y}}{(1-r_{12}^2)} \end{aligned} \quad (11)$$

اگر بین  $x_1$  و  $x_2$  هم‌خطی شدید وجود داشته باشد، در این صورت مقدار  $r_{12}$  بزرگ خواهد بود و با توجه به ماتریس  $(X^T X)^{-1}$ ، اگر  $|r_{12}| \rightarrow 1$ ، آن‌گاه:

$$V(\hat{\beta}_i) = \frac{1}{(1-r_{12}^2)} \times \sigma^2 \rightarrow +\infty \quad (12)$$

بسته به اینکه  $r_{12} \rightarrow 1$  یا  $r_{12} \rightarrow -1$ ، خواهیم داشت:

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \frac{-r_{12}}{(1-r_{12}^2)} \times \sigma^2 \rightarrow \pm\infty \quad (13)$$

این تورم واریانس منجر به این می‌شود که در هر بار برآورد با نمونه‌های مختلف، ضرایب مختلفی به دست آید.

برای اندازه‌گیری میزان هم‌خطی، معیارهای مختلفی از جمله ماتریس همبستگی، شاخص سیستم مقادیر ویژه ماتریس  $X^T X$  و

## ۴- رویکرد پایش پیشنهادی

آنچه مبنای اصلی رویکرد پیشنهادی این تحقیق است، از این چالش آغاز شد که فرض استقلال خطی متغیرهای مرحله اول در فرایندهای چندمرحله‌ای، نقض شده باشد. این شرایط منجر به رسیدن به یک مدل اشتباه و پر ریسک می‌شود. پیشنهاد این پژوهش استفاده از رگرسیون ریبج برای مدل‌سازی است تا طبق آنچه در بخش (۳) بیان شد، واریانس ضرایب برآورد شده ( $\tilde{\beta}$ ) را کاهش داده و مدل به‌دست آمده مطمئن‌تر و به واقعیت نزدیک‌تر گردد. به عبارت دیگر گام-های زیر برای ارائه‌ی نمودار انتخاب عامل انحراف (CSC) مبتنی بر رگرسیون ریبج پیشنهاد می‌گردد:

- ۱- در ابتدا  $n$  نمونه  $(X_i, Y_i)$  انتخاب می‌گردد که  $X_i$  بردار متغیرهای مرحله‌ی نخست  $(x_{i1}, x_{i2}, \dots, x_{ip})$  و  $i$  شمارنده‌ی نمونه است.
- ۲- رابطه‌ی رگرسیون چندگانه بین  $X_i$  و  $Y_i$  ها به دست آورده می‌شود؛ الف. در صورتی که شاخص‌های تورم واریانس، بیانگر عدم وجود این مشکل باشد، طبق روال رگرسیون عادی و نمودار CSC عادی عمل می‌شود.
- ب. در صورتی که شاخص‌های تورم واریانس، مشکل هم‌خطی را جدی‌گزارش دادند (طبق بخش ۳)، آنگاه گام بعدی اجرا شود.
- ۳- رابطه‌ی بین نمونه‌های  $(X_i, Y_i)$  با استفاده از رگرسیون ریبج، به‌گونه‌ای به‌دست آورده شود که شاخص‌های  $VIF$  آن قابل قبول باشد.
- ۴- باقیمانده‌های  $e_i = Y_i - \hat{Y}_i$  محاسبه گردد.
- ۵- بر اساس  $\sigma$  مربوط به باقیمانده‌ها، حدود کنترل بالا و پایین  $(UCL, LCL)$  محاسبه گردد.

$$\begin{aligned} UCL &= l\sigma \\ CL &= 0 \\ LCL &= -l\sigma \end{aligned} \quad (24)$$

که در این رابطه، مقدار  $l$  ضریبی است که در نمودارهای کنترل به منظور رسیدن به مقدار متوسط طول دنباله‌ی تحت کنترل خاصی تعیین می‌گردد [۱۸].

حال می‌توان از این نمودار در فاز II پایش آماری فرایند دو مرحله‌ای مذکور استفاده نمود. ادعا می‌شود نمودار فوق از نمودار CSC عادی در حالتی که هم‌خطی وجود دارد، کارایی بیشتری خواهد داشت. به‌عنوان یک اثبات اجمالی می‌توان بیان داشت که در شرایط هم‌خطی به‌علت تورم واریانس در ضرایب برآورد شده، در هر نمونه‌گیری که در فاز نخست انجام می‌شود یک مدل به‌طورکلی متفاوت به‌دست می‌آید. پس مدل در بسیاری مواقع با ضرایب نادرست برآورد می‌شود و مدل حاصل‌شده در زمان به‌کارگیری، خطای زیادی دارد و در نتیجه باقیمانده‌ها افزایش یافته و هشدارهای خطای بیشتری رخ می‌دهد. ولی با استفاده از رگرسیون ریبج با وجود کمی انحراف و اما به دلیل واریانس کمتر، دقت ضرایب بیشتر شده و مدل برآورد شده به واقعیت نزدیک‌تر است. به عبارت دیگر باقیمانده‌ها کوچک‌تر شده و بیشتر در محدود کنترل قرار می‌گیرد و خطای کمتری رخ می‌دهد.

$$Q_n(\beta) = (y - X\beta)^T(y - X\beta) + k\beta^T\beta = \sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 + k \sum_{j=1}^p \beta_j^2 \quad (17)$$

و با صفر قرار دادن مشتق آن:

$$\frac{\partial Q_n(\beta)}{\partial \beta} = -2X^T(y - X\beta) + 2k\beta = 0 \quad (18)$$

معادله‌ی استاندارد برآورد ضرایب به این شکل در خواهد آمد:

$$\tilde{\beta} = (X^T X + kI_p)^{-1} X^T y \quad (19)$$

که در آن  $k$  ضریب اریبی و نوعی جریمه است که منجر می‌شود برآوردگر جدید، ناریبی خود را از دست بدهد. این ضریب باید طوری توسط تحلیلگر انتخاب شود که در ازای کمی اریبی، واریانس ضرایب برآوردشده کمتر شوند.  $I$  ماتریس واحد است و اگر  $k = 0$  آنگاه برآوردگر ریبج، همان برآوردگر حداقل مربعات است. به عبارت دیگر رویکرد حداقل مربعات حالت خاصی از رگرسیون ریبج است. به‌جای استفاده از معادله‌ی استاندارد، به‌سادگی با تغییر ماتریس  $X$  و  $Y$  به‌صورت زیر و حل معادله ماتریسی رگرسیون ساده، در واقع جواب‌های رگرسیون ریبج را به‌دست آورده‌ایم:

$$\tilde{X} = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \\ \sqrt{k_1} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{k_2} & \sqrt{k_3} & \dots & 0 \\ 0 & 0 & \dots & \ddots & 0 \\ \vdots & \vdots & \dots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{k_p} \end{pmatrix}_{(n+p) \times p} \quad (20)$$

$$= \begin{pmatrix} X \\ \sqrt{k_p} I_p \end{pmatrix}$$

$$\tilde{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{(n+p) \times 1} = \begin{pmatrix} Y \\ \mathbf{0}_p \end{pmatrix} \quad (21)$$

که در آن‌ها  $n$  و  $p$  همان ابعاد ماتریس  $X$  است که قبلاً بیان شد. مزیت این معادله برای به‌دست آوردن رگرسیون ریبج این است که می‌توان به‌سادگی با تغییر ماتریس  $X$  و  $Y$  به  $\tilde{X}$  و  $\tilde{Y}$  از داده‌های اولیه، آن‌ها را در نرم‌افزارهای رایج مثل مینی‌تب قرار داده داده و با استفاده از دستور رگرسیون عادی، جوابی معادل رگرسیون ریبج به دست آورد [۱۷].

برای یافتن  $k$  بهینه نیز می‌توان از رابطه‌های (۲۲) و (۲۳) استفاده نمود [۱۷]:

$$k_j = \frac{\sigma^2}{\alpha_j^2} ; \quad j = 1, 2, \dots, p \quad (22)$$

که در آن  $\sigma^2$  واریانس  $Y$  بوده و اگر  $T$  ماتریس متعامدی از مقادیر ویژه‌ی  $X'X$  باشد، آنگاه  $\alpha$  نیز این‌گونه به‌دست خواهد آمد:

$$\alpha = T'\beta \quad (23)$$

## ۵- ارزیابی عملکرد رویکرد پایش پیشنهادی و مقایسه با

## رویکرد پایش موجود

در این بخش، مطالعات شبیه سازی به منظور ارزیابی عملکرد رویکرد پیشنهادی در حضور هم خطی چندگانه صورت پذیرفته است. در این مطالعات، از متوسط طول دنباله<sup>۱</sup>  $(ARL)$  به عنوان متداول ترین شاخص برای ارزیابی عملکرد نمودارهای کنترل استفاده شده است. به این ترتیب که ابتدا بررسی عملکرد رویکرد موجود در حضور هم خطی چندگانه مورد مطالعه قرار گرفته و سپس با نشان دادن ضعف رویکرد موجود در ادبیات، مطالعات گسترده ای به منظور مقایسه عملکرد رویکرد پایش پیشنهادی با رویکرد موجود صورت پذیرفته است. این شبیه سازی از یک فرایند دو مرحله ای استفاده نموده است که با  $10,000$  بار تکرار، حداکثر اطمینان از کارآمدی رویکرد پیشنهادی را کسب نموده و میزان کارایی آن سنجیده شده است.

به این منظور ابتدا فرایندی دو مرحله ای با ۴ متغیر در نظر گرفته می شود که در مرحله اول آن سه متغیر  $j = 1, 2, 3$ :  $X_j$  وجود دارد و در مرحله دوم متغیر کیفی  $Y$  که متأثر از مرحله اول است. یعنی خاصیت آبخاری در این فرایند حاکم است. شرط کلیدی در مرحله اول این است که  $X_j$  ها از یکدیگر مستقل نبوده و روابط ذیل حاکم است:

$$\begin{aligned} X_1 &\sim N(5,2) \\ X_2 &\sim N(15,3) \end{aligned} \quad (25)$$

$$\begin{aligned} X_3 &= 6X_1 + 2 + \varepsilon_1; & \varepsilon_1 &\sim N(0,1) \\ Y &= 15X_1 - 10X_2 + 5X_3 + \varepsilon_2; & \varepsilon_2 &\sim N(0,6) \end{aligned}$$

در این نمونه، ضرایب واقعی  $X$  ها بردار  $\beta$  است:

$$\beta = (15, -10, 5)$$

داده های به دست آمده از شبیه سازی اولیه ای این فرایند در جدول (۱) آورده شده اند. نمودارهای پراکندگی ارتباط های بین متغیرهای مرحله اول ( $X_j$ ) نیز در شکل (۳) مشاهده می شوند.

جدول (۱): سی نمونه از یک فرایند دو مرحله ای با متغیرهای وابسته

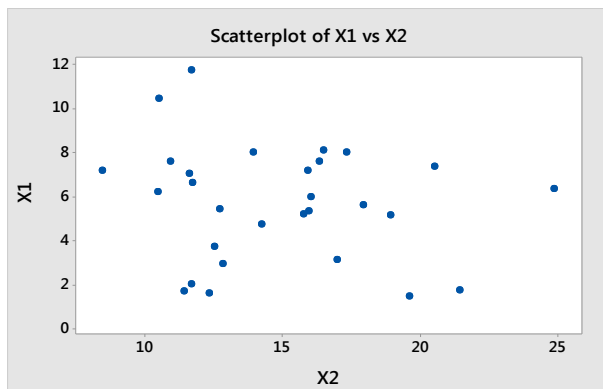
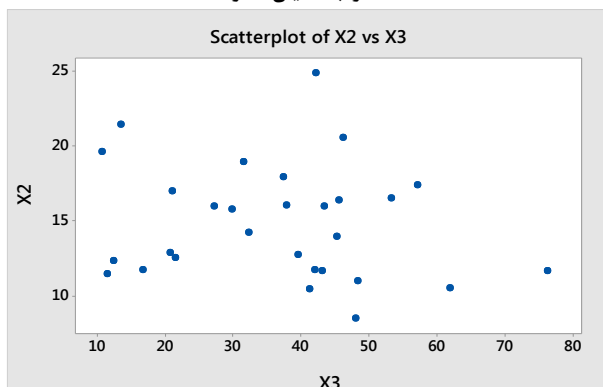
## در مرحله اول

مشاهده	Y	X <sub>3</sub>	X <sub>2</sub>	X <sub>1</sub>
۱	۲۵/۲۶۸	۲۰/۸۸۸۹	۱۲/۸۸۴۰	۲/۹۸۲۳
۲	۲۷۹/۶۸۷	۴۸/۱۲۳۳	۸/۴۸۸۱	۷/۲۰۹۴
۳	-۱۴۹/۱۸۹	۱۳/۶۶۹۳	۲۱/۴۶۹۲	۱/۷۹۰۱
۴	۹۱/۳۰۴	۳۷/۹۶۳۵	۱۶/۰۵۶۵	۶/۰۰۴۱
۵	۱۶۷/۸۵۴	۴۱/۳۴۳۷	۱۰/۴۹۱۴	۶/۲۳۵۶
۶	۱۶۵/۵۲۵	۳۹/۶۹۸۷	۱۲/۷۶۳۳	۵/۴۶۲۴
۷	۱۹۹/۱۸۱	۴۲/۱۵۸۸	۱۱/۷۶۴۶	۶/۶۶۳۷
۸	-۱۳۰/۹۷۰	۱۰/۷۵۷۰	۱۹/۶۴۴۰	۱/۴۸۱۵
۹	۲۵۵/۳۷۶	۵۷/۲۴۶۷	۱۷/۳۶۸۴	۸/۰۲۲۸
۱۰	۵۹/۱۸۶	۲۷/۳۴۴۲	۱۵/۹۷۱۰	۵/۳۴۰۶
۱۱	۱۴۱/۹۲۸	۴۳/۴۹۹۴	۱۵/۹۶۱۹	۷/۱۹۰۲

## 1. Average Run Length

۱۲	-۱۷/۷۴۹	۱۶/۷۸۷۳	۱۱/۷۱۴۰	۲/۰۴۹۵
۱۳	۲۵/۴۴۴	۳۱/۷۲۱۱	۱۸/۹۷۲۴	۵/۱۵۶۱
۱۴	۱۹۳/۲۵۱	۴۵/۶۵۸۴	۱۶/۳۸۵۶	۷/۶۳۲۸
۱۵	۱۹۸/۶۷۸	۴۵/۳۲۲۷	۱۳/۹۵۵۵	۸/۰۴۲۴
۱۶	۷۲/۸۳۳	۳۲/۳۶۳۸	۱۴/۲۶۱۱	۴/۷۴۷۰
۱۷	-۱۶۶/۸۵۷	-۰/۷۷۲۲	۱۴/۵۹۳۱	-۰/۵۶۰۶
۱۸	-۴۳/۵۵۲	۱۱/۵۷۳۹	۱۱/۴۵۱۸	۱/۷۰۹۵
۱۹	۱۰۵/۶۷۱	۳۷/۵۳۸۷	۱۷/۹۶۵۵	۵/۶۵۵۹
۲۰	۱۴۳/۶۶۸	۴۶/۲۹۱۱	۲۰/۵۶۸۸	۷/۳۸۲۷
۲۱	۶۳/۰۶۷	۴۲/۳۲۹۵	۲۴/۸۹۵۶	۶/۳۸۹۴
۲۲	-۱۰/۶۲۶	۲۱/۱۷۲۶	۱۷/۰۱۵۷	۳/۱۶۲۱
۲۳	۴۴۱/۵۱۰	۷۶/۳۹۳۶	۱۱/۷۱۰۱	۱۱/۷۳۳۲
۲۴	۵۵/۷۰۴	۲۹/۹۴۴۰	۱۵/۸۰۳۹	۵/۲۱۷۰
۲۵	۲۵۴/۳۹۲	۴۸/۴۰۷۶	۱۰/۹۷۱۸	۷/۵۹۶۵
۲۶	۲۶/۳۹۳	۲۱/۵۷۷۹	۱۲/۵۴۷۳	۳/۷۵۲۹
۲۷	-۵۱/۹۹۵	۱۲/۵۷۹۴	۱۲/۳۶۴۱	۱/۶۵۵۲
۲۸	۱۸۹/۲۴۹	۴۳/۲۳۸۹	۱۱/۶۶۷۱	۷/۰۵۵۸
۲۹	۲۳۶/۲۹۴	۵۳/۳۳۶۴	۱۶/۵۱۲۷	۸/۱۱۳۰
۳۰	۳۷۶/۲۶۴	۶۱/۹۸۷۱	۱۰/۵۳۸۶	۱۰/۴۶۶۶

همان طور که در شکل (۳- الف، ب) دیده می شود متغیرهای  $X_1$  و  $X_2$  و نیز  $X_2$  و  $X_3$  از یکدیگر مستقل بوده و شکل (۳- ج) نشان از وابستگی خطی زیاد بین متغیرهای  $X_1$  و  $X_3$  دارد.

الف) ارتباط میان  $X_2$  و  $X_1$ ب) ارتباط میان  $X_3$  و  $X_2$

جدول (۴): برآورد ضرایب به کمک رویکرد رگرسیون ریج

Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
$X_1$	۱۲/۵۲	۴/۱۳	۳/۰۳	۰/۰۵	۲۱/۲۴
$X_2$	۱۰/۳۳	۰/۷۰۴	-۱۴/۶۷	۰	۳/۷۱
$X_3$	۵/۴۲	۰/۶۸۹	۷/۸۸	۰	۲۳/۳۹

Regression Equation					
$Y = 12/52 X_1 - 10/334 X_2 + 5/425 X_3$					

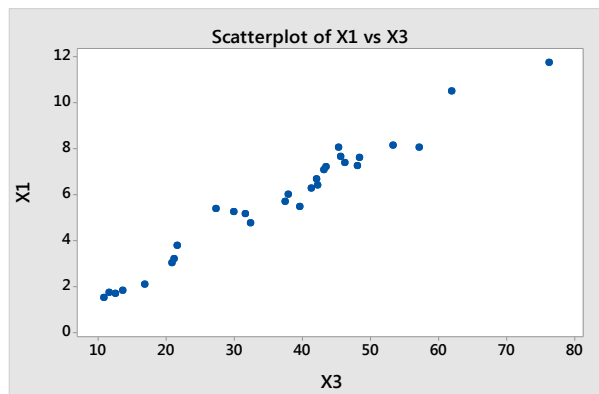
همان‌طور که مشاهده می‌شود تغییرات زیادی هم‌سو با بهبود مدل حاصل شده است؛

۱- بهبود  $P$ -value: در رویکرد ریج سطح معناداری مربوط به متغیر  $X_1$  حدوداً ۰/۱۰ بود که با اغماض پذیرفته شد، ولی اینجا به ۰/۰۵ کاهش یافته است که سطح اطمینان را بسیار بالاتر می‌برد.  
 ۲- کاهش قابل توجه عامل تورم واریانس: در رویکرد ریج، عامل تورم واریانس عددی نزدیک به ۲۰۰ بود که رویکرد جدید آن را به حدود ۲۰ کاهش داده است. هرچند بهتر است که این شاخص از ۱۰ بالاتر نباشد ولی همین مقدار کاهش، بهبود بسیار قابل توجهی را نشان می‌دهد.

۳- ضرایب نزدیک‌تر به واقعیت: مهم‌ترین مزیت رویکرد پیشنهادی، میزان صحت مدل‌سازی است. با توجه به اینکه ضرایب واقعی داده‌ها عبارت بودند از  $(15, -10/5)$  و  $\beta$  و رگرسیون عادی آن را به شکل  $(9/62, -10/92, 6/14)$  برآورد کرده بود؛ قدر مطلق بردار اختلاف آن  $(5/38, 0/92, 1/14)$  حاصل شد ولی اکنون بردار برآورد ریج  $(12/48, -10/33, 5/43)$  شده و بردار قدر مطلق اختلاف آن نیز  $(2/48, 0/33, 0/43)$  شده است که کوچک‌تر از بردار اختلاف ضرایب رگرسیون عادی است. می‌توان بهبود ضرایب برآورد شده را نیز با محاسبه اختلاف آن‌ها و به نام بردار  $Im = (2/9, 0/59, 0/71)$  نشان داد که در اینجا  $Im$  است.

برای بررسی دقیق‌تر عملکرد رویکرد پایش پیشنهادی با استفاده از باقیمانده‌ها که در جدول (۵) دیده می‌شود، مقادیر  $ARL$  برای هر دو رویکرد طی شبیه‌سازی‌هایی با ۱۰,۰۰۰ بار تکرار مورد بررسی قرار گرفت؛ طبق پژوهش وید و وودال [۲] اگر  $l = 3$  آنگاه مقدار  $ARL$  باید به  $370/37$  برسد. با شبیه‌سازی که در نرم‌افزار متلب اجرا شده است، مقدار متوسط طول دنباله برای رویکرد موجود، عدد  $89/274$  با انحراف معیار  $0/89$  به‌دست آمد که این عدد بسیار کمتر از  $370/37$  است که از قبل هم به‌خاطر اختلاف فاحش بین رگرسیون به‌دست‌آمده با رابطه‌ی واقعی بین متغیرها، قابل پیش‌بینی بود.

همین روال برای رویکرد پایش پیشنهادی تکرار شد که مقدار متوسط طول دنباله تحت کنترل برابر  $266/928$  با انحراف معیار  $2/7278$  به‌دست آمد. هرچند مقدار متوسط طول دنباله به‌طور دقیق به مقدار مطلوب نرسیده است ولی افزایش ۳ برابری آن در مقایسه با حالت قبل حاکی از نتیجه‌ی مورد انتظار و بهبود قابل توجه جواب در رویکرد پیشنهادی این پژوهش است.



(ج) ارتباط میان  $X_3$  و  $X_1$

شکل (۳): نمودارهای پراکندگی متغیرهای مرحله‌ی اول

اگر از رگرسیون عادی (روش حداقل‌مربعات) برای توضیح ارتباط بین متغیرها استفاده شود، خروجی مطابق با جدول (۲) می‌شود؛

جدول (۲): برآورد ضرایب به کمک رویکرد حداقل‌مربعات

Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
$X_1$	۹/۶۲	۵/۷۶	۱/۶۷	۰/۱۰۷	۱۸۳/۰۵
$X_2$	-۱۰/۹۲	۰/۳۹۹	-۳۲/۲۲	۰/۰۰۰	۳/۹۲
$X_3$	۶/۱۴	۰/۹۳۰	۶/۶۰	۰/۰۰۰	۱۹۲/۶۹

Regression Equation					
$Y = 9/62 X_1 - 10/916 X_2 + 6/140 X_3$					

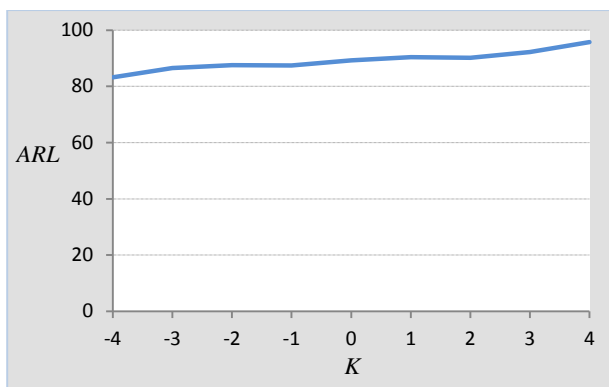
در این مدل‌سازی، ضرایب با  $P$ -value مناسب یا نسبتاً مناسب به‌دست‌آمده است. ولی شاخص تورم واریانس، به‌طور قابل توجهی بزرگ است و این به معنای افزایش ریسک در عدم صحت ضرایب برآورد شده است. توجه به ضرایب برآورد شده نیز نشان می‌دهد که بخاطر تورم واریانس، رگرسیون حداقل‌مربعات نتوانسته است مدل موجود بین داده‌ها را درست برآورد نماید.

حال برای مدل‌سازی به‌کمک رگرسیون ریج، با توجه به آنچه در رابطه‌های (۲۰) تا (۲۳) بیان شد، مقادیر ماتریس  $k$  را محاسبه می‌کنیم تا با افزودن آن به ماتریس‌های  $X$  و  $Y$ ، پارامترهای رگرسیون ریج را برآورد نماییم.

جدول (۳): برآورد ماتریس  $k$  برای استفاده در رگرسیون ریج

	$X_1$	$X_2$	$X_3$	$Y$
	۵/۷۲	۰	۰	۰
	۰	۳/۵	۰	۰
	۰	۰	۲۲/۲۵	۰

در همین شرایط وقتی از رویکرد رگرسیون ریج برای مدل‌سازی استفاده گردد، نتیجه‌ی مدل‌سازی مطابق جدول (۴) می‌شود:

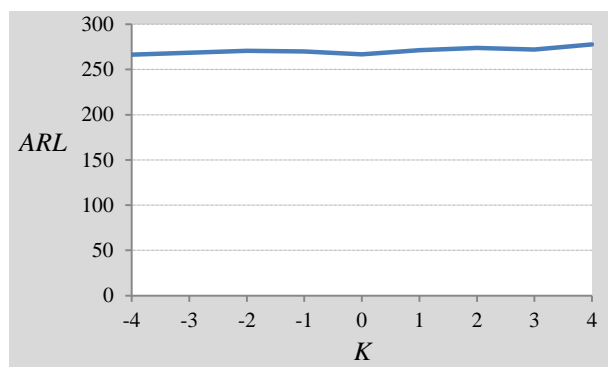
شکل (۴):  $ARL$  حاصل از رویکرد موجود با  $K\sigma_x$  شیفت در  $X$ 

همان‌طور که در شکل (۴) مشاهده می‌شود، نمودار تقریباً توانسته است تأثیر شیفت در  $X$  را حذف کند. البته طبق داده‌های موجود در جدول (۴)، ۱۴٪ افزایش  $ARL$  (طی  $-4\sigma_x$  تا  $4\sigma_x$  شیفت در  $X$ ) ایجاد شده است که باید آن را به حساب اختلال در مدل ساده در حالت تورم واریانس گذاشت.

جدول (۶):  $ARL$  حاصل از رویکرد موجود با  $K\sigma_x$  شیفت در  $X$ 

$K$	$ARL$
-۴	۸۳/۲۵۵
-۳	۸۶/۵۷۱
-۲	۸۷/۵۲۸
-۱	۸۷/۴۳۲
۰	۸۹/۲۷۴
۱	۹۰/۳۶۵
۲	۹۰/۲۲۴
۳	۹۲/۱۷۹
۴	۹۵/۷۴۶

اگر از رویکرد رگرسیون رنج برای مدل‌سازی استفاده شود، هنگامی که  $K\sigma_x$  شیفت در  $X$  اعمال شود بعد از ۱۰,۰۰۰ تکرار برای شبیه‌سازی  $ARL$ . نتایج مطابق شکل (۵) و جدول (۷) می‌باشد.

شکل (۵):  $ARL$  حاصل از رویکرد پیشنهادی با  $K\sigma_x$  شیفت در  $X$ 

طبق این یافته‌ها در طول  $-4\sigma_x$  تا  $4\sigma_x$  شیفت در  $X$ ، به میزان ۴٪ تغییر در  $ARL$  ایجاد شده است که به‌طور نسبی بهتر از خروجی رویکرد پایش رایج شده است. پس می‌توان گفت که نمودار انتخاب عامل

جدول (۵): باقیمانده‌های رگرسیون‌های حداقل مربعات و رنج

مشاهده	عادی	رنج
۱	۸/۵۶	۱۳/۶۵
۲	۴/۶۵	۸/۶
۳	۳/۱۷	-۴/۸۸
۴	۱۶/۳۱	۲۳/۸۳
۵	۱۴/۲۷	۹/۰۱
۶	-۲۰/۰۷	-۱۹/۱۵
۷	-۱۲/۶۶	-۱۳/۴۳
۸	-۱۱/۸۱	-۱۵/۱۵
۹	۱۸/۳۶	۱۹/۳
۱۰	-۴/۶۲	-۳/۷
۱۱	-۵/۲۲	-۸/۰۰
۱۲	-۱۰/۱۸	-۱۳/۲
۱۳	-۱۶/۷۴	-۱۳/۱۱
۱۴	۱۱/۰۳	۱۵/۹۹
۱۵	۱۰/۰۳	۱۷/۸۲
۱۶	۸/۹۷	۷/۷۴
۱۷	۷/۵۲	۱۶/۰۵
۱۸	-۱۵/۹۸	-۲۳/۹
۱۹	-۲۴/۲۷	-۲۳/۹۱
۲۰	-۳۱/۴۵	-۲۶/۱۱
۲۱	-۱۵/۸۶	-۱۴/۸۱
۲۲	۲/۵۸	-۴/۸۴
۲۳	-۶/۰۵	-۹/۴۱
۲۴	۱۶/۹	۱۶/۸۵
۲۵	۱۲/۹۶	۱۲/۶۴
۲۶	۱۳/۴۷	۱۰/۶۸
۲۷	۱۴/۷۱	۱۰/۷۵
۲۸	-۱۲/۵۸	۱/۱۵
۲۹	-۵/۸۱	-۸/۷۶
۳۰	۳/۸۷	۱۰/۰۳

#### ۵-۱- مقایسه دو رویکرد پایش در شرایط ایجاد شیفت در

##### میانگین $X$

مبنای نمودارهای انتخاب عامل انحراف، جدا کردن عامل انحراف در مرحله‌ی اول از مرحله‌ی دوم است [۱]. به‌عبارت‌دیگر انحراف و شیفت در  $X$  ها نباید تغییری در نمودار باقیمانده‌ها ایجاد کند. بر این اساس برای ارزیابی کارآمدی رویکرد پیشنهادی، در داده‌های مذکور به اندازه  $K\sigma_x$  در  $X_1$  و  $X_2$  شیفت ایجاد کرده و دوباره باقیمانده‌ها محاسبه شده‌اند. شبیه‌سازی با ۱۰,۰۰۰ بار تکرار انجام و مقادیر  $ARL$  را محاسبه کرده‌ایم که در شکل (۴) و جدول (۶) گزارش شده است.



جدول (۸):  $ARL$  حاصل از رویکرد پیش موجود ( $ARL_1$ ) و رویکرد

پیشنهادی ( $ARL_2$ ) با  $K\sigma_Y$  شیفت در  $Y$

$K$	$ARL_1$	$ARL_2$	$K$	$ARL_1$	$ARL_2$
۰	۲۰۱/۱۹	۲۰۰/۱۸	-۱/۵	۲۴/۴۱	۸/۰۶
۰/۱	۱۵۹/۲۵	۱۷۸/۰۹	-۱/۴	۲۹/۱۸	۱۰/۳۲
۰/۲	۱۲۷/۸۱	۱۵۵/۸۷	-۱/۳	۳۶/۳۱	۱۳/۵۳
۰/۳	۹۵/۷۶	۱۲۱/۶	-۱/۲	۴۶/۸۵	۱۷/۹۷
۰/۴	۷۸/۱۸	۹۹/۰۷	-۱/۱	۵۸/۱۲	۲۱/۹۶
۰/۵	۶۲/۹۷	۷۷/۳۵	-۱	۷۵/۹۹	۲۹/۲۴
۰/۶	۴۸/۷	۵۹/۶۲	-۰/۹	۹۳/۰۳	۳۶/۷۲
۰/۷	۳۹/۴	۴۷/۳۶	-۰/۸	۱۱۹/۴۷	۴۷/۹
۰/۸	۳۱/۶۳	۳۶/۱۸	-۰/۷	۱۵۳/۱۳	۶۳/۴۴
۰/۹	۲۶/۵۶	۲۷/۹۴	-۰/۶	۱۸۷/۲۱	۷۹/۶۸
۱	۲۰/۷۲	۲۱/۰۱	-۰/۵	۲۲۶/۶۴	۱۰۳/۲۴
۱/۱	۱۶/۹۳	۱۵/۹۱	-۰/۴	۲۴۹/۱۲	۱۳۰/۵۸
۱/۲	۱۳/۶۷	۱۲/۷۹	-۰/۳*	۲۷۷/۷	۱۵۹/۰۸
۱/۳	۱۱/۳۴	۱۰/۱۴	-۰/۲	۲۶۳/۰۵	۱۸۴/۶۷
۱/۴	۹/۹۹	۷/۵۷	-۰/۱	۲۲۸/۳۶	۲۰۰/۰۹
۱/۵	۸/۶۱	۵/۹۲			

$ARL_1$  \* نقطه ماکزیمم

انحراف پیشنهادی توانسته است تأثیر مرحله‌ی اول را به‌وسیله‌ی باقیمانده‌ها حذف نموده و کارآمدی خود را اثبات نماید.

جدول (۷):  $ARL$  حاصل از رویکرد پیشنهادی با  $K\sigma_X$  شیفت در  $X$

$K$	$ARL$
-۴	۲۶۶/۴۰۲
-۳	۲۶۸/۶۰۳
-۲	۲۷۰/۸۱۱
-۱	۲۶۹/۸۹۶
۰	۲۶۶/۹۲۸
۱	۲۷۱/۱۸۹
۲	۲۷۳/۸۸
۳	۲۷۱/۹۰۹
۴	۲۷۷/۸۰۱

### ۵-۲- مقایسه دو رویکرد پیش در شرایط ایجاد شیفت

#### در میانگین $Y$

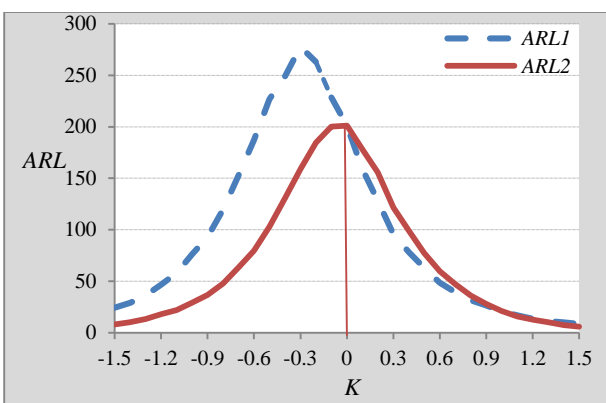
چنان‌که دیده شد در شرایط هم‌خطی، متوسط طول دنباله تحت کنترل در رویکرد پیش پیشنهادی تقریباً ۳ برابر بهتر از رویکرد پیش موجود بود و نمودار کنترل با رویکرد پیشنهادی برای پیش مرحله‌ی دوم، تقریباً اثر مرحله‌ی قبل را حذف کرده و کارآمدی خود را به اثبات رسانده است. اما برای بررسی‌های گسترده‌تر و مقایسه کارایی دو روش، باید توانایی کشف انحراف‌های مختلف در متغیر کیفی مرحله دوم را در نمودارهای کنترل موجود و پیشنهادی مورد مطالعه قرارداد. برای این کار نخست با تنظیم پارامتر حدود بالا و پایین هر دو نمودار کنترل به نقطه‌ای رسانده می‌شود که  $ARL$  تحت کنترل برای هر دو، مقدار تقریباً ۲۰۰ باشد. سپس مقدار  $Y$  در هر دو مدل به مقدار  $K\sigma_Y$  شیفت داده می‌شود به طوری که  $K$  بین  $-۱/۵$  تا  $+۱/۵$  با گام‌های  $۰/۱$  تغییر کند. در هر دو روش تکرارها ۱۰،۰۰۰ بار بوده است تا دقت قابل قبولی حاصل گردد.

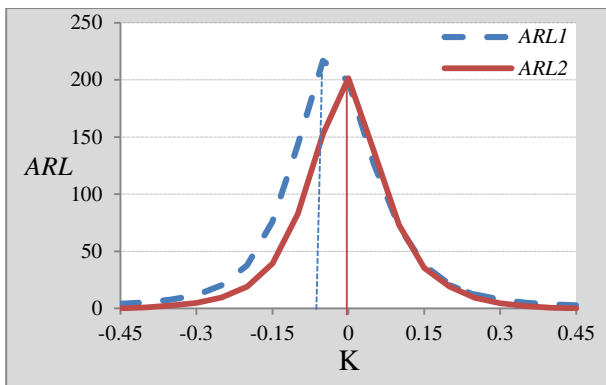
در نمودار کنترل با رویکرد موجود برای رسیدن به  $ARL$  مورد نظر، حدود بالا و پایین  $\pm ۴۵/۱۳۵$  شده بود که در نتیجه  $ARL$  تحت کنترل برابر  $۲۰۱/۱۹$  با انحراف معیار  $۲/۰$  به دست می‌آید. از طرفی در رویکرد پیشنهادی با رگرسیون رنج نیز حدود بالا و پایین  $\pm ۴۲/۴۱۵$  شد که در نتیجه  $ARL$  برابر  $۲۰۰/۱۸$  با انحراف معیار  $۲/۰$  به دست می‌آید. نتایج در جدول (۸) گزارش شده‌اند. طبق آنچه در جدول (۸) و شکل (۶) مشاهده می‌شود، با ایجاد شیفت مثبت در میانگین متغیر مرحله‌ی دوم ( $Y$ ) مقدار  $ARL$  در هر دو روش کاهش یافته است. البته نمودار در رویکرد موجود، مقدار اندکی سریع‌تر این خطا را اعلام می‌کند. اما در صورت اعمال شیفت در جهت منفی برای میانگین متغیر مرحله‌ی دوم،  $ARL$  روش پیشنهادی، شروع به کاهش کرده و در اعلام هشدار خارج از کنترل نسبت به رویکرد موجود (رویکرد مبتنی بر رگرسیون عادی) برتری قابل توجهی دارد.

شکل (۶): منحنی  $ARL$  حاصل از رویکرد پیش موجود ( $ARL_1$ ) و رویکرد

پیشنهادی ( $ARL_2$ ) با  $K\sigma_Y$  شیفت در میانگین  $Y$

آنچه در سمت چپ نمودار در شکل (۶) دیده می‌شود، علاوه بر برتری رویکرد پیشنهادی، افزایش مقادیر  $ARL$  و عدم تقارن نمودار مذکور برای رویکرد پیش موجود است. در این خصوص می‌توان بیان کرد که در واقع مدل به‌دست آمده از روش رگرسیون عادی در شرایط هم‌خطی، نماینده‌ی مناسبی برای داده‌ها برای ارائه‌ی نمودار کنترل CSC نیست. در واقع این مدل نتوانسته است رابطه‌ی واقعی بین داده‌های مراحل اول و دوم را به‌درستی نمایش دهد. به بیان دیگر وجود تورم واریانس در ضرایب منجر شده است که ضرایب  $\beta$  دورتر از واقعیت برآورد کرده و دچار اشتباه شود. در نتیجه مدل به‌دست آمده مربوط به شرایط دیگری از داده‌هاست. یعنی در مثال شبیه‌سازی پژوهش پیش‌رو، اگر  $Y$  ها حدوداً  $۰/۳\sigma_Y$  به سمت منفی شیفت داده می‌شدند مدل برآورد شده بیشتر به مدل واقعی منطبق شده و در نتیجه  $ARL$  حداکثر را در آن نقطه نشان می‌داد و با هرگونه انحراف از آن، هشدارها





شکل (۷):  $ARL$  حاصل رویکرد پایش موجود ( $ARL_1$ ) و رویکرد پیشنهادی ( $ARL_2$ ) با  $K\sigma_x$  شیفیت در  $\beta_1$

در این شرایط نیز نمودار  $ARL$  مشابه حالت شیفیت  $Y$  است. یعنی در سمت راست، عدم تفاوت معنادار و اما در سمت چپ برتری قابل توجه نمودار رویکرد پیشنهادی را شاهد هستیم. همچنین عدم تقارن و چولگی نمودار به سمت چپ، که مشابه وضعیتی است که در قسمت قبلی هم دیده شد، در رویکرد موجود به چشم می‌خورد. در ادامه در ضریب  $X_2$  یعنی  $\beta_2$  شیفیت ایجاد کرده و مقادیر  $ARL$  بررسی می‌شوند. با  $K\sigma_x$  که  $K$  بین  $-2$  تا  $+2$  تغییر می‌کند، مقادیر  $ARL$  مطابق با جدول (۱۰) و شکل (۸) خواهد بود.

جدول (۱۰):  $ARL$  حاصل از رویکرد پایش موجود ( $ARL_1$ ) و رویکرد پایش پیشنهادی ( $ARL_2$ ) با  $K\sigma_x$  شیفیت در  $\beta_2$

$K$	$ARL_1$	$ARL_2$
-2	1	1/76
-1/8	1/0.5	2/0.6
-1/6	1/12	2/65
-1/4	2/23	3/71
-1/2	3/85	5/4
-1	5/45	8/86
-0/8	7/89	15
-0/6	18/84	28
-0/4	40/24	59/68
-0/2	96/0.8	120/74
0	200/51	200/91
0/2	289/14*	160/0.7
0/4	201/13	82/47
0/6	98/63	39/47
0/8	44/17	20/39
1	21/29	10/94
1/2	10/93	6/74
1/4	6/69	4/4
1/6	4/1	3/0.5
1/8	3/58	2/29
2	1	1/83

بیشتر می‌شد. اما برای داده‌های فعلی بهترین مدل رگرسیون ریج است که وضعیتی قابل قبول داشته و با کوچک‌ترین شیفیت در مشخصه‌ی کیفی مرحله‌ی دوم از خودش واکنش نرمال نشان داده و هشدارهایش افزایش می‌یابد.

### ۵-۳- مقایسه‌ی عملکرد دو رویکرد در شرایط ایجاد شیفیت در ضرایب $\beta$

برای بررسی انحراف از حالت تحت کنترل در مرحله‌ی دوم، ممکن است تغییر فقط در متوسط  $Y$  صورت نگیرد، بلکه تغییری در ضرایب  $\beta$  صورت بگیرد. برای بررسی این شرایط در دو حالت ضرایب  $X_1$  و  $X_2$  جداگانه شیفیت داده شده و بررسی شده‌اند.

نخست در  $\beta_1$  شیفیت ایجاد کرده و مقادیر  $ARL$  مطالعه می‌شوند. با  $K\sigma_x$  که در آن  $K$  بین  $-0/5$  تا  $+0/5$  تغییر می‌کند، مقادیر  $ARL$  مطابق با جدول (۹) و شکل (۷) خواهد بود.

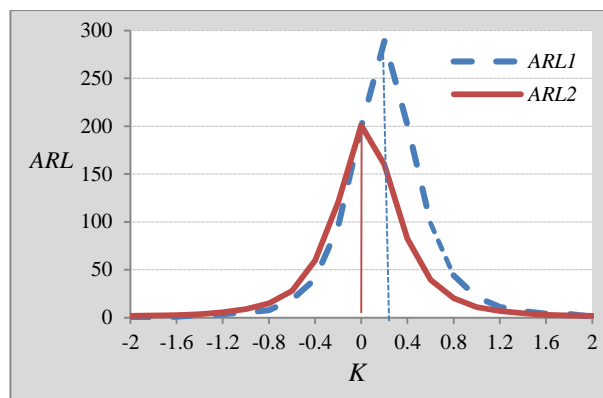
جدول (۹):  $ARL$  حاصل از رویکرد پایش موجود ( $ARL_1$ ) و رویکرد پیشنهادی ( $ARL_2$ ) با  $K\sigma_x$  شیفیت در  $\beta_1$

$K$	$ARL_1$	$ARL_2$
-0/50	3/27	1
-0/45	4/0.7	1
-0/40	5/43	1/14
-0/35	7/84	2/56
-0/30	11/86	4/9
-0/25	20/5	9/68
-0/20	37/76	19/3
-0/15	75/9	39/22
-0/10	142/84	82/78
-0/05	216/84*	153/16
0	199/12	201/61
0/05	126/45	138/0.9
0/10	71/32	73/18
0/15	37/66	35/42
0/20	21	19/28
0/25	12/31	9/31
0/30	7/87	4/64
0/35	5/0.7	2/19
0/40	3/65	1/14
0/45	2/84	1
0/50	2/33	1

بالا تر و مقادیر متوسط طول دنباله‌ی خارج از کنترل به شدت کاهش یابد. در کل می‌توان اذعان داشت که نمودار کنترل انتخاب عامل انحراف مبتنی بر مدل رگرسیون ریح عملکرد بسیار بهتری در مقایسه با رویکرد مبتنی بر مدل رگرسیون معمولی در جهت کشف شیفت‌های مختلف دارد و با توجه به دارا بودن منحنی متوسط طول دنباله تقریباً متقارن، به‌کارگیری آن به‌ویژه در فرایندهایی با حضور هم‌خطی چندگانه به‌طور اکید توصیه می‌گردد. به‌عنوان پیشنهاد برای پژوهش‌های آتی می‌توان به به‌کارگیری دیگر روش‌های کنترل تورم واریانس مانند رویکرد LASSO و Elastic Net پرداخته و اثربخشی هر کدام را بر روی نمودار انتخاب عامل انحراف بررسی و بهترین رویکرد را معرفی کرد. همچنین با توجه به این‌که تاکنون نمودارهای انتخاب عامل انحراف به‌طور جداگانه برای دو وضعیت هم‌خطی چندگانه و نقاط پرت بررسی و ارائه شده‌اند، می‌توان شرایطی را در نظر گرفت که هر دو اختلال فوق به‌طور هم‌زمان رخ دهند و در این حالت به بررسی شرایط و ارائه‌ی رویکرد پیش جدید پرداخت. در نهایت توسعه رویکرد پیش پیشنهادی مبتنی بر رگرسیون ریح را می‌توان به کمک نمودارهای کنترل حافظه‌دار (مثل EWMA و CUSUM) بهبود بخشید.

#### مراجع

- [1] Zhang, G.X. (1984). "A new type of control charts and theory of diagnosis with control charts. World Quality Congress, Transactions. Milwaukee", *WI: American Society for Quality Control*, 3: 175-185.
- [2] Wade, MR., Woodall, WH. (1993). "A review and analysis of cause-selecting control charts". *Journal of Quality Technology*, 25(3): 161-169.
- [3] Skinner, K.R., Montgomery, D.C., Runger, G.C. (2003). "Process monitoring for multiple count data using generalized linear model-based control charts", *International Journal of Production Research*, 41(6): 1167-1180.
- [4] Jearkpaporn, D., Borrór, C.M., Montgomery, D.C., Runger, G.C. (2003). "Process monitoring for correlated gamma distributed data using generalized linear model control charts", *Quality and Reliability Engineering International*, 19(6): 477-491.
- [5] Asadzadeh, Sh., Aghaie, A., Shahriari, H. (2009). "Monitoring dependent process steps using robust cause selecting control charts", *Quality and Reliability Engineering International*, 25(7): 851-874.
- [6] Noorossana, R, Shekary, M, (2012). "Monitoring two dependent process steps using special variable sample sizes and sampling intervals cause-selecting control charts", *Quality and Reliability Engineering International*, 28(4): 437- 453.
- [7] Asgari, A., Amiri, A., Niaki, S. T. A. (2014). "A new link function in glm-based control charts to improve monitoring of two-stage processes with poisson response". *The International Journal of Advanced Manufacturing Technology*, 72 (9): 1243-1256.
- [8] Amiri, A., Yeh, A. B., Asgari, A. (2016). "Monitoring



شکل (۸): ARL حاصل از رویکرد پیش موجود ( $ARL_1$ ) و رویکرد پیش پیشنهادی ( $ARL_2$ ) با  $K\sigma$  شیفت در  $\beta_2$

نمودار تغییرات  $ARL$  که برای این وضعیت به‌دست‌آمده است، به‌ظاهر کمی متفاوت به نظر می‌رسد. در نمودار تغییرات  $ARL$  بر اثر شیفت در ضریب  $\beta_2$  مشاهده می‌شود که هنگام کاهش ضریب، تفاوت چشمگیری ایجاد نمی‌شود. ولی در سمت راست یعنی افزایش ضریب، برتری کاملاً محسوس در رویکرد پیشنهادی نسبت به رویکرد رایج دیده می‌شود. نکته دیگر عدم تقارن در نمودار  $ARL$  رویکرد موجود است که در بخش قبلی مورد بررسی قرار گرفت. تفاوتی که اینجا وجود دارد، چولگی نمودار  $ARL$  به سمت راست است که در واقع به ماهیت وابستگی ضرایب رگرسیون عادی در حالت هم‌خطی برمی‌گردد بدین ترتیب که در رویکرد رایج در شرایط تورم واریانس، بعضی ضرایب  $\beta$  را بالاتر و در مقابل، بعضی را پایین‌تر برآورد می‌کند و نوعی بالانس بین ضرایب شکل می‌گیرد تا همچنان باقیمانده‌ها حداقل شود ولی صحت این مدل قابل‌اعتماد نیست.

#### ۶- نتیجه‌گیری و جمع‌بندی

در این مقاله توسعه‌ی نمودارهای کنترل انتخاب عامل انحراف در حضور متغیرهای ورودی و تأثیرگذار وابسته، موردتوجه قرار گرفت. برای در نظر گرفتن خاصیت آشناری در این فرایندها و همچنین کاهش آثار مخرب هم‌خطی چندگانه از مدل رگرسیون ریح با هدف مدل‌سازی ارتباط میان مرحله‌های فرایند استفاده شد و باقی‌مانده‌های حاصل از این مدل به‌منظور حذف اثرات متغیرهای مراحل قبل مورد استفاده قرار گرفت. شبیه‌سازی‌های گسترده به‌منظور بررسی عملکرد نمودار کنترل انتخاب عامل انحراف پیشنهادی با رویکرد موجود در ادبیات صورت پذیرفت. نتایج نشان داد که در حضور هم‌خطی چندگانه، مدل رگرسیون خطی دچار تورم واریانس در ضرایب برآورد شده می‌شود و این مسأله منجر به افزایش احتمال خطا در ضرایب و در نتیجه خطا در مدل به‌دست‌آمده می‌شود که به‌شدت کارایی رویکرد پیش موجود را کاهش می‌دهد. همچنین در این شرایط منحنی متوسط طول دنباله متقارن نمی‌شود که از ضعف‌های رویکرد پیش موجود است. از طرفی به‌کارگیری رگرسیون ریح به برآوردهایی اریب اما با واریانس کمتر و در نتیجه، ثبات بیشتر منجر می‌شود که باعث می‌شود تا توان آزمون

- two-stage processes with binomial data using generalized linear model-based control charts”. *Quality Technology and Quantitative Management*, 13(3): 241-262.
- [9] Asadzadeh, Sh., Aghaie, A., Shahriari, H. (2014). “Using frailty models to account for heterogeneity in multistage manufacturing and service processes”, *Quality and Quantity*, 48(2): 593-604.
- [10] Pan, J. N., Li, C., Wu, J. J. (2016). “A new approach to detecting the process changes for multistage systems”, *Expert Systems with Applications*, 62: 293-301.
- [11] Khedmati, M., Niaki, S. T. A. (2016). “A new control scheme for phase –II monitoring of simple linear profiles in multistage processes”, *Quality and Reliability Engineering International*, 32(7): 2559-2571.
- [12] Bahrami, H., Niaki, S.T.A., Khedmati, M., (2019). “Monitoring multivariate profiles in multistage processes”. *Communications in Statistics - Simulation and Computation*, 48(1): 1-29
- [13] Kim, J., Jeong, M. K., Elsayed, A. E. (2017). “Monitoring multistage processes with autocorrelated observations”. *International Journal of Production Research*, 55(8), 2385-2396.
- [14] Asadzadeh, Sh., Kiadaliry, F. (2017). “Monitoring type-2 censored reliability data in multistage processes”. *Quality and Reliability Engineering International*, 33(8): 2551-2561.
- [15] Sangahn, K. (2019). “Variable-selection based SPC procedures for high-dimentional multistage processes”. *Journal of Systems Engineering and Electronics*, 30(1): 144-153.
- [16] Keshavarz, M., Asadzadeh, Sh., Niaki, S.T.A. (2019). “Controlling autocorrelated data in multistage manufacturing processes with an application to textile industry”. *Quality and Reliability Engineering International*, 35(7): 2314-2326.
- [17] Montgomery, D. C., Peck, E. A., Vining, G. G. (2012). *Introduction to linear regression analysis*, 5<sup>th</sup> Edition, Hoboken, NJ: Wiley.
- [18] Montgomery, D. C. (2009). *Introduction to statistical quality control*. 7<sup>th</sup> Edition, Hoboken, NJ: Wiley.



DOI: 10.22084/ier.2020.17875.1811

## A Ridge-Regression Based Control Chart in the Presence of Multicollinearity

Sh. Asadzadeh<sup>1\*</sup>, V. Mahdavi<sup>2</sup>

<sup>1,2</sup> Department of Industrial Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran

### ARTICLE INFO

#### *Article history:*

Received 10 December 2018

Accepted 11 January 2020

#### *Keywords:*

Multistage Processes

Cause Selecting Chart (CSC)

Multicollinearity

Variance Inflation

Ridge Regression

### ABSTRACT

Cause-selecting charts (CSCs) are the main tool for statistical quality control in multistage processes. The establishment of these control charts, which use regression models and residuals to remove the effect of previous stages from current stage, needs some critical assumptions like the linear independence among incoming quality variables. When this assumption is violated, which is quite common in real practice and called multicollinearity, the variance inflation in regression parameters occurs. Subsequently, this leads to some crucial problems in the performance of traditional CSCs. To tackle the mentioned problem, there exist some statistical and practical methods and the application of ridge regression is one of the most efficient approaches. In this paper, modeling and designing a novel control chart based on ridge regression has been addressed and extensive simulation studies have been conducted to investigate the performance of the suggested monitoring procedure compared with the traditional control chart in the literature. The results reveal that the proposed control chart outperforms the existing control chart in the presence of multicollinearity.

\* Corresponding author. Sh. Asadzadeh  
Tel.: 021-77009847; E-mail address: [sh\\_asadzadeh@iautnb.ac.ir](mailto:sh_asadzadeh@iautnb.ac.ir)