

شناسایی شوک‌ها در قیمت ربع، نیم و تمام سکه در ایران با استفاده از سری‌های زمانی چند متغیره

رحیم چینی پرداز، الهام مختاری و بهزاد منصوری *

تاریخ وصول: ۱۳۹۴/۲/۲۳ تاریخ پذیرش: ۱۳۹۴/۴/۱۰

چکیده:

در این تحقیق به تأثیر و شناسایی انواع نقاط پرت نوساز، جمع پذیر، تغییر سطح، تغییر موقت در تعیین مدل و برآورد پارامترهای سری زمانی چند متغیره پرداخته می‌شود. برای شناسایی انواع نقاط پرت در مدل سری‌های زمانی چند متغیره روش تکرار پانکراز و همکاران (۲۰۰۰) مورد توجه قرار گرفته و توانایی این روش نسبت به روش کلاسیک یک متغیره سری زمانی مورد بررسی قرار می‌گیرد. از آنجا که قیمت سکه تحت تأثیر قیمت جهانی طلا و همچنین به دلیل استفاده آن در اعیاد و مراسم ملی، تحت تأثیر تغییرات فصلی مختلف قرار است، می‌تواند الگوی مناسبی برای سری‌های زمانی همراه با نقاط پرت و شوک‌های اقتصادی باشد. روش پیشنهادی برای به دست آوردن شوک‌های اقتصادی قیمت ربع، نیم و تمام سکه آزادی ایران از فروردین ۱۳۷۸ تا آذر ۱۳۸۷ مورد استفاده قرار گرفته و نتایج آن با روش کلاسیک یک متغیره مقایسه شده است. مقایسه نتایج عددی بیانگر حساسیت روش پیشنهادی نسبت به روش معمول است.

طبقه‌بندی JEL: 62M10، 91B84، 37M10

واژه‌های کلیدی: نقطه‌ی پرت نوساز، نقطه پرت جمع پذیر، نقطه پرت تغییر سطح، نقطه‌ی پرت تغییر موقت، سری زمانی چند متغیره، VARMA

* به ترتیب، استاد آمار دانشکده ریاضی و علوم کامپیوتر دانشگاه شهید چمران اهواز، کارشناسی ارشد آمار دانشگاه آزاد مرودشت و استادیار دانشکده ریاضی و علوم کامپیوتر دانشگاه شهید چمران اهواز.
(chinipardaz_r@scu.ac.ir)

۱- مقدمه

بسیاری از سری‌های زمانی در عمل تحت تأثیر برخی رویدادهای خارجی نظیر: اعتصاب‌ها، جنگ‌ها، بحران‌های سیاسی یا اقتصادی ناگهانی، قرار می‌گیرند. نتیجه این رویدادهای بازدارنده، ظهور مشاهدات تصنعی است که با سایر مشاهدات سری زمانی سازگاری ندارد که به آنها نقاط پرت (دور افتاده) می‌گویند.

نادیده گرفتن این اختلال‌ها می‌تواند در شناسایی و برآورد پارامترهای مدل و در نتیجه بر پیش‌بینی سری زمانی اثر نموده و زمینه ساز بروز خطاهای آماری گردد. لذا بایستی در مرحله‌ی اول، مکان و نوع این اختلال‌ها شناسایی شود. در مرحله‌ی بعد اثر آن‌ها برآورد شود و پس از تعدیل، پارامترهای مدل برآورد و پیش‌بینی‌ها انجام شود.

تشخیص نقاط پرت در سری‌های زمانی چند متغیره نخستین بار توسط پانکراز و همکاران (۲۰۰۰) بر اساس آماره‌ی نسبت درست نمایی مورد توجه قرار گرفت. سپس گالیانو و همکاران (۲۰۰۴) به وسیله‌ی الگوریتمی بر پایه‌ی ماتریس تصویر، اوتی (۲۰۰۵) با به کار بردن اندازه‌ی نامتجانسی بر اساس مؤلفه‌های اصلی، ونگ و شن (۲۰۰۶) الگوریتمی بر پایه ضریب پراکندگی موضعی و در کار بعدی در سال (۲۰۰۸) از طریق نرم فروبنیوس تعمیم یافته به تشخیص نقاط پرت در سری‌های زمانی چند متغیره پرداختند. در این مقاله روش پانکراز و همکاران (۲۰۰۰) مورد نظر قرار گرفته است. با توجه به اینکه توزیع آماره آزمون تشخیص نقاط پرت نامعلوم است با روش شبیه‌سازی و بر اساس تعداد نمونه مورد نیاز در این مقاله به دست آمده است. مقایسه دو روش تشخیص نقاط پرت یک متغیره و چند متغیره که از الگوریتم‌های متفاوت استفاده می‌کنند، مورد بررسی و برای قیمت سکه (ربع سکه، نیم سکه و تمام سکه) انجام شده است. مقاله در شش بخش تنظیم شده است. در بخش دوم مقاله مدل اتورگرسیو چند متغیره همراه با نقاط پرت نشان داده شده است. در بخش سوم مقاله تأثیر انواع نقاط پرت بر روی باقیمانده‌های مدل که در آنالیز تشخیص نقاط پرت نقش اساسی دارد نشان داده شده است تا در بخش بعدی آزمون تشخیص نقطه پرت و نوع آن به دست آورده شود. در این میان دو آماره آزمون داده شده است. از آنجا که این آماره‌ها دارای توزیع آماری معینی نیستند، لازم است نقاط بحرانی توزیع بر اساس تعداد نمونه معینی به دست آورده شود. شبیه‌سازی نقاط بحرانی برای تعداد نمونه $n=117$ که تعداد نمونه مورد نیاز

بخش بعدی است، انجام شده است. در دو بخش بعدی مقاله (بخش ۵ و ۶) داده‌های قیمت سکه (ربع سکه، نیم سکه و تمام سکه) مورد بررسی و الگوریتمی برای تشخیص نقطه و نوع آن پیشنهاد شده است. در بخش ۵ الگوریتم یک متغیره و در بخش ۶ سری زمانی چند متغیره مورد استفاده قرار گرفته است. در بخش نهمی مقاله، بخش ۷، تفاوت دو روش تشخیص (تک متغیره برای هر مؤلفه و چند متغیره به صورت توأم مؤلفه‌ها) آورده شده است.

۲- مدل سری‌های زمانی چند متغیره همراه با نقاط پرت

فرض کنید یک سری زمانی K بعدی، ایستا و وارون پذیر $VARMA(p,q)$ به صورت $\Phi(B)Y_t = \Theta(B)v_t$ باشد. که در آن v_t فرایند اغتشاش خالص، دارای توزیع نرمال با بردار میانگین صفر و ماتریس کواریانس Σ و $\Phi(B)$ و $\Theta(B)$ چند جمله‌ای‌های ماتریسی به ترتیب از مرتبه p و q بر حسب عملکرد تفاضلی پس رو می‌باشند.

$$\Phi(B) = I - \Phi_1 B - \dots - \Phi_p B^p, \quad \Theta(B) = I - \Theta_1 B - \dots - \Theta_q B^q$$

فرض کنید سری زمانی چند متغیره $\{Y_t\}$ تحت تأثیر یک ضربه یا رویداد خارجی قرار گرفته و مدل آن به صورت $Z_t = Y_t + f(t)$ تغییر یافته است که در آن $\{Z_t\}$ سری مشاهده شده و $f(t)$ تابع پارامتری و نشان دهنده‌ی اثرهای مداخله‌ای می‌باشد. با توجه به نوع اثرهای مداخله‌ای این تابع ممکن است به صورت ثابت یا تصادفی باشد که در عمل توسط تحلیل‌گر داده‌ها و براساس اطلاع از اثرهای مداخله‌ای بر فرآیند $\{Z_t\}$ مشخص می‌شود. در مدل ثابت، اثر مداخله‌ی خارجی به صورت:

$$Z_t = Y_t + r(B) \check{S}_t^{(h)} \quad (1)$$

می‌باشد. که در آن $\check{S} = (\check{S}_1, \check{S}_2, \dots, \check{S}_k)'$ اثر تداخل اولیه (میزان تأثیر نقاط پرت روی سری زمانی Y_t)، $r(B)$ چند جمله‌ای ماتریسی بر حسب B ، بیانگر نیروی تداخل و $\check{S}_t^{(h)}$ متغیر شاخصی که نشان دهنده‌ی درستی تداخل در زمان h می‌باشد و به صورت زیر تعریف می‌شود.

$$\langle_i^{(h)} = \begin{cases} 1 & t = h \\ 0 & t \neq h \end{cases}$$

با توجه به معادله (۱)، اگر $\Gamma(B) = \Psi(B)$ باشد نقطه‌ی پرت نوساز گفته می‌شود. اگر در

$$\Psi(B) = \{\Phi(B)\}^{-1} \Theta(B) = \left(I + \sum_{i=1}^{\infty} \Psi_i B^i \right) v_t \quad (2)$$

$\alpha(B) = I$ باشد نقطه‌ی پرت جمع پذیر گفته می‌شود. اگر

$$\Gamma(B) = (1-B)^{-1} I = (I + IB^2 + IB^3 + \dots) \quad (3)$$

باشد، نقطه‌ی پرت تغییر سطح تعریف می‌شود. اگر $\Gamma(B) = \{D(u)\}^{-1}$ باشد یک نقطه‌ی پرت تغییر موقت خواهد بود. در اینجا $\{D(u)\}$ یک ماتریس قطری با عناصر $(1-u_1B), (1-u_2B), \dots, (1-u_kB)$ به طوری که $0 < u_i < 1$ است و معمولاً برای سادگی $u_1 = u_2 = \dots = u_k = u$ در نظر می‌گیرند. برای سادگی در علامت گذاری از اینجا به بعد نقطه‌ی پرت جمع پذیر را با AO ، نقطه‌ی پرت نوساز را با IO ، نقطه‌ی پرت تغییر سطح را با LS ، نقطه‌ی پرت تغییر موقت را با TC ، نشان می‌دهیم. برای توضیح مطلب نقاط پرت مطرح شده را در یک مدل اتورگرسیو برداری مرتبه ۱، (۱) VAR ، به صورت $v_t \sim N(0, \Sigma)$ $Y_t = (I - \Phi B)^{-1} v_t$ ، بررسی می‌کنیم.

(۱) نقطه‌ی پرت نوساز: مدل (۱) VAR توام با نقطه‌ی پرت نوساز در زمان h ، به صورت $Z_t = (I - \Phi B)^{-1} \{v_t + \check{S}_t^{(h)}\}$ است که در آن $E(B) = (I - \Phi B)^{-1}$ بنابراین $E_0 = I$ و برای $i = 1, 2, \dots$ ، $E_i = \Phi^i$. بنابراین از نقطه‌ای به بعد E ها به صفر نزدیک می‌شوند و تأثیر این نقطه‌ی پرت کاهش می‌یابد تا اینکه بعد از مدت زمانی کوتاه این تأثیر به صفر می‌رسد.

(۲) نقطه‌ی پرت جمع پذیر: در این حالت $Z_t = (I - \Phi B)^{-1} v_t + \check{S}_t^{(h)}$ ، نقطه‌ی پرت جمع پذیر فقط در زمان $t=h$ سطح سری دو متغیره را به اندازه‌ی S تحت تأثیر قرار می‌گیرد و بقیه‌ی نقاط سری بدون تغییر باقی می‌مانند.

(۳) نقطه‌ی پرت تغییر سطح: فرآیند (۱) VAR توام با نقطه‌ی پرت تغییر سطح در زمان $t=h$ به صورت $Z_t = (I - \Phi B)^{-1} v_t + (1-B)^{-1} \check{S}_t^{(h)}$ می‌باشد. به دلیل تأثیر $(1-B)^{-1} \check{S}_t^{(h)} = \check{S}_t^{(h)} + \check{S}_{t-1}^{(h)} + \check{S}_{t-2}^{(h)}$ بر فرآیند (۱) VAR ، از زمان $t=h$ تا

پایان سری تغییر سطحی به اندازه S در سری ایجاد می‌شود. در نتیجه اثر این نقطه، به طور پیوسته و دائمی تا انتهای سری ادامه می‌یابد.

(۴) نقطه‌ی پرت تغییر موقت: سری زمانی دو متغیره (۱) VAR توام با نقطه‌ی پرت تغییر موقت، در زمان $t=h$ عبارت است از:

$$Z_t = (I - \Phi B)^{-1} v_t + (1 - uB)^{-1} \check{S}'_t^{(h)} \quad (0 < u < 1) \quad (۴)$$

که $(1 - uB)^{-1} \check{S}'_t^{(h)} = \check{S}'_t^{(h)} + u\check{S}'_{t-1}^{(h)} + u^2\check{S}'_{t-2}^{(h)} + \dots$. بنابراین در این حالت از زمان $t=h$ به بعد سری با یک تغییر روبرو که با افزایش زمان با نرخ u کاهش می‌یابد تا جایی که در نقاط انتهای سری اثر این نقطه از بین می‌رود. نقاط پرت جمع پذیر و تغییر سطح موارد محدود شده‌ی تغییر موقت می‌باشند.

۳- اثر نقاط پرت بر باقیمانده‌ها و آزمون وجود نقاط پرت

با توجه به اینکه روش‌های کلاسیک بررسی مدل در سری‌های زمانی براساس رفتار باقیمانده‌ها انجام می‌شود، باقیمانده‌ها در شناسایی و برآورد پارامترهای مدل اهمیت ویژه‌ای دارند. فرض کنید $\{Y_t\}$ فرآیند VARMA(p,q) به صورت $\Phi(B)Y_t = \Theta(B)v_t$ یا $\Pi(B)Y_t = v_t$ باشد. که در آن پارامترها معلوم فرض می‌شوند و رابطه‌ی $\Pi(B) = \{\Theta(B)\}^{-1} \Phi(B)$ برقرار می‌باشد. بنابراین با استفاده از (۱) سری باقیمانده‌ها به صورت:

$$A_t = \Pi(B)Z_t = \Pi(B)Y_t + \Pi(B)r(B)\check{S}'_t^{(h)} = v_t + \Pi(B)r(B)\check{S}'_t^{(h)} \quad (۵)$$

حاصل می‌شود که در آن $\Pi(B)r(B) = y(B) = I - y_1B - y_2B^2 - y_3B^3 - \dots$ خواهد بود. همچنین عملگرهای $x(B) = \Pi(B)(1-B)^{-1}$ و $s(B) = \Pi(B)\{D(B)\}^{-1}$ در شناسایی نقاط پرت تغییر سطح و تغییر موقت مفید می‌باشند. برای $i = 1, 2, \dots$ با

استفاده از $(1-B)x(B) = \Pi(B)$ ، وزن‌های X_i به صورت، $X_i = \sum_{j=1}^i (\Pi_j - I)$

و با توجه به $(1-uB)s(B) = \Pi(B)$ ، وزن‌های S_i به صورت،

$$S_i = \sum_{j=1}^i \Pi_j^{u^{i-j}} - u^i I$$

صورت $A_t = v_t + \check{S}'_t^{(h)}$ ، نقطه پرت جمع پذیر به صورت

$$A_t = v_t + \Pi(B)\check{S}'_t^{(h)}$$

نقطه پرت تغییر سطح به صورت

و برای نقطه پرت تغییر موقت به صورت $A_t = v_t + \chi(B) \check{S}'_t^{(h)}$ خواهد بود. $A_t = v_t + s(B) \check{S}'_t^{(h)}$

برای تأثیر نقاط پرت بر باقیمانده‌های مدل سری زمانی، نشان دادن این باقیمانده‌ها از نظر نموداری مناسب می‌باشد. اگر سری زمانی با نقطه‌ی پرت نوساز آمیخته شود، در این حالت $\Psi(B) = I + \Psi_1 B + \Psi_2 B^2 + \Psi_3 B^3 + \dots$ با توجه به رابطه‌ی $\Psi(B) \Pi(B) = I$ ، سری باقیمانده‌ها به صورت:

$$r_t = \begin{cases} v_t & t \neq h \\ v_t - \check{S} & t = h \end{cases}$$

خواهد بود. بنابراین نقطه‌ی پرت نوساز در نقطه‌ی زمانی h بر سری باقیمانده‌ها به اندازه‌ی S تأثیر می‌گذارد و در بقیه‌ی نقاط هیچ تغییری در سری رخ نمی‌دهد. شکل‌های ۱ و ۲ اثر این نقطه‌ی پرت بر سری $\{r_t\}$ و $\{r_t - v_t\}$ را نشان می‌دهند. اگر نقطه‌ی پرت از نوع جمع‌پذیر باشد، $\alpha(B) = I$ و $r_t = \Pi(B) z_t = v_t + \Pi(B) \check{S}'_t^{(h)} = v_t + \check{S}'_t^{(h)} - \Pi_1 \check{S}'_{t-1}^{(h)} - \dots$ پرت بر سری باقیمانده‌ها در زمان $t=h$ و پس از آن، تأثیر می‌گذارد، این تأثیر از $\Pi(B)$ نتیجه می‌شود. پس سری باقیمانده‌ها به صورت:

$$r_t = \begin{cases} v_t & t < h \\ v_t - \Pi_i \check{S} & t = h + i, i \geq 0 \end{cases}$$

می‌باشد. به طوری که $\Pi_0 = -I$ و بقیه Π_i ها همان ضرائب مدل چند متغیره VAR می‌باشند. برای مثال در یک مدل VAR (۱)، وزن‌های Π_i به صورت $\Pi(B) = I - \Pi_1 B$ خواهد بود. بنابراین در این مدل $r_t = (I - \Pi_1 B) Y_t = \Pi_1 Y_{t-1}$. اگر VAR (۱) با یک نقطه‌ی پرت جمع‌پذیر ترکیب شود، چون وزن‌های Π_0 و Π_1 مخالف صفر و بقیه‌ی Π_i صفر می‌باشند، این نقطه فقط در نقاط $t=h$ و $t=h+1$ بر سری باقیمانده‌ها تأثیر می‌گذارد. شکل‌های ۳ و ۴ اثر این نقطه را روی سری باقیمانده‌ها و سری هموار شده‌ای که از تفاضل $\{r_t\}$ و $\{v_t\}$ بدست آمده را نشان می‌دهند.

در صورت وجود یک نقطه‌ی پرت تغییر سطح در سری چند متغیره، $r_t = v_t + \chi(B) \check{S}'_t^{(h)}$ می‌باشد. بنابراین

$$r_t = \begin{cases} v_t & t < h \\ v_t - \left(\sum_{j=1}^i \Pi_j - I \right) \tilde{S} & t = h+i, \quad i \geq 0 \end{cases}$$

و $X_0 = -I$ خواهد بود.

طبق روابط بالا وقتی مدل (۱) VAR با نقطه‌ی پرت تغییر سطح ترکیب می‌شود سری باقیمانده‌ها تا نقطه زمانی $t=h-1$ با v_t یکسان و از این زمان به بعد این نقطه بر سری باقیمانده‌ها تأثیر می‌گذارد.

همان طور که از شکل‌های ۵ و ۶ دیده می‌شود، اثر نقطه‌ی پرت در زمان $t=h$ آغاز و تا پایان سری ادامه می‌یابد. و چون در مدل (۱) VAR، $\Pi(B)$ دارای جملات متناهی و همی X_i ‌ها برای $i \geq 1$ با هم برابر می‌باشند، از نقطه‌ی $t=h+1$ تا پایان سری این اثر ثابت می‌ماند. اثر نقطه‌ی پرت جمع پذیر بر r_t پس از مدت زمانی متوقف می‌شود اگر $\Pi(B)$ دارای جملات متناهی باشد.

در حالتی که نقطه‌ی پرت تغییر موقت باشد، سری باقیمانده‌ها به صورت $r_t = v_t + S(B) \tilde{S}'^{(h)}$ خواهد بود. طبق رابطه‌ی بالا اگر Y_t دارای فرآیند (۱) VAR باشد که در زمان $t=h$ با یک نقطه‌ی پرت تغییر موقت آمیخته شود آن گاه سری باقیمانده‌ها، $\{r_t\}$ ، تا زمان $t=h-1$ با v_t یکسان می‌باشد و از زمان $t=h$ به بعد این نقطه بر سری تأثیر می‌گذارد. نقطه‌ی پرت تغییر موقت از زمان $t=h$ بر سری باقیمانده‌ها تأثیر می‌گذارد.

با توجه به مقادیر وزنی S_i ‌ها می‌بینیم که با افزایش زمان این مقادیر کاهش می‌یابند به طوری که از جایی به بعد به صفر نزدیک می‌شوند و میزان تأثیر این نقطه بر سری باقیمانده‌ها کاهش می‌یابد تا جایی که در اواخر سری این اثر صفر می‌شود. البته مدت زمان این تأثیر به مقدار u بستگی دارد. هر چه u مقدار بزرگتری اختیار کند مدت زمان اثر این نقطه‌ی پرت بیشتر می‌شود.

در شکل‌های ۷ و ۸، اثر این نقطه پرت در نظر گرفته شده است. همان طور که شکل‌ها نیز نشان می‌دهند، علاوه بر تأثیر u ، هر چه اندازه‌ی S بزرگتر باشد تأثیر نقطه‌ی تغییر موقت بیشتر خواهد بود.

برخلاف تغییر سطح، تغییر موقت بر سطح سری مشاهده شده یا سری باقیمانده‌ها تأثیر دائمی ندارد. زیرا اثر آن در نهایت کاهش می‌یابد که در این حالت عامل کاهش u معرفی می‌شود. با فرض اینکه S اندازه‌ی اثر نقطه‌ی پرت می‌باشد از مجموع مربعات باقیمانده‌ها و سپس از روش حداقل مربعات برآورد اثر نقطه‌ی پرت برای هر یک از نقاط پرت به دست می‌آید. در عمل پارامترهای مدل سری زمانی چند متغیره‌ی VARMA، معلوم نبوده و اطلاعاتی از مقادیر A_h ، وزن‌های X_i و Π_i وجود ندارد. با استفاده از روش‌های برآورد پارامترها در سری‌های زمانی چند متغیره، پارامترها با این فرض که در سری نقطه‌ی پرت وجود ندارد برآورد می‌شوند و در نهایت اثر نقاط پرت به صورت زیر برآورد می‌شوند:

$$\check{S}_{IO,h} = \hat{A}_h \quad (6)$$

$$\check{S}_{AO,h} = -\left(\sum_{i=0}^{n-h} \hat{\Pi}_i' \sum^{-1} \hat{\Pi}_i\right)^{-1} \left(\sum_{i=0}^{n-h} \hat{\Pi}_i' \sum^{-1} \hat{A}_{h+i}\right) \quad (7)$$

$$\check{S}_{LS,h} = -\left(\sum_{i=0}^{n-h} \hat{\chi}_i' \sum^{-1} \hat{\chi}_i\right)^{-1} \left(\sum_{i=0}^{n-h} \hat{\chi}_i' \sum^{-1} \hat{A}_{h+i}\right) \quad (8)$$

$$\check{S}_{TC,h} = -\left(\sum_{i=0}^{n-h} \hat{S}_i' \sum^{-1} \hat{S}_i\right)^{-1} \left(\sum_{i=0}^{n-h} \hat{S}_i' \sum^{-1} \hat{A}_{h+i}\right) \quad (9)$$

۴- آزمون وجود نقطه پرت

فرض کنید H_0 فرض عدم وجود نقطه‌ی پرت در سری زمانی چند متغیره در زمان $t=h$ و فرض مقابل وجود یکی از انواع نقاط پرت به صورت:

$$\begin{cases} H_0: \check{S}_i = 0 \\ H_1: \check{S}_i \neq 0 \end{cases}$$

باشد که \check{S}_i برای $i=IO, AO, LS, TC$ ، نشان دهنده‌ی اثر یکی از چهار نوع نقطه پرت می‌باشد.

زمانی که داده‌ها وابستگی نداشته باشند پانکراز و همکاران (۲۰۰۰) دو آماره

$J_{i,h}$ ، $C_{i,h}$ را به صورت زیر معرفی کرده‌اند.

ابتدا آماره‌ی $J_{i,h} = \check{S}_{i,h}' \sum_{i,h}^{-1} \check{S}_{i,h}$ که دارای توزیع خی دو غیرمرکزی با پارامتر

غیرمرکزی $\check{S}_i = \check{S}_{i,h}' \sum_{i,h} \check{S}_{i,h}$ می‌باشد. آماره‌ی دیگر برای آزمون فرضیه‌ها استفاده

از

$$C_{i,h} = \max_{1 \leq j \leq k} |\check{S}_{j,i,h}| / \sqrt{t_{j,i,h}} \quad (10)$$

می‌باشد. در این جا $\check{S}_{j,i,h}$ ، j امین عنصر از بردار $\check{S}_{i,h}$ و $t_{j,i,h}$ ، (j, i) امین عنصر از $\sum_{i,h}$ می‌باشد.

تحت فرض H_0 و با فرض اینکه پارامترهای مدل VARMA معلوم باشند این آماره دارای توزیع خی دو مرکزی با K درجه آزادی است از آنجا که برای داده‌های مشاهده شده، h معلوم نیست و به دنبال یافتن نقطه‌ی پرت هستیم لازم است ماکزیمم C و J برای مشاهدات به دست آید. پس در زمان $t=h$ دو آماره‌ی آزمون توام به صورت زیر محاسبه می‌شوند:

$$J_{\max}(i, h_i) = \max_h (J_{i,h}) \quad (11)$$

$$C_{\max}(i, h_i^*) = \max_h (C_{i,h}) \quad (12)$$

که در آن h_i نقطه‌ی زمانی می‌باشد که آماره‌ی آزمون $J_{i,h}$ در آن ماکزیمم و h_i^* نقطه‌ای است که آماره‌ی آزمون $C_{i,h}$ در آن ماکزیمم می‌شود. تحت فرض صفر آماره $J_{\max}(I, h_i)$ ماکزیمم یک نمونه‌ی تصادفی n تایی از توزیع خی دو با k درجه‌ی آزادی می‌باشد. با توجه به وابستگی مشاهدات سری زمانی، به جز نقطه پرت IO که پرت بودن یک نقطه مستقل از دیگر مشاهدات است، توزیع پارامتری و معینی برای این آماره‌ها وجود ندارد و بنابراین توزیع مجانبی آماره و مقادیر بحرانی توزیع $J_{\max}(I, h_i)$ با استفاده از شبیه سازی به دست آورده شده‌اند. هر یک از سه آماره‌ی آزمون توام $J_{\max}(i, h_i)$ برای $i=AO, LC, TC$ ماکزیمم یک نمونه‌ی وابسته‌ی n تایی می‌باشند. توزیع مجانبی این آماره‌ها بسیار پیچیده می‌باشد. در عمل مدل صحیح سری ناشناخته می‌باشد و توزیع‌های مجانبی به صورت تقریبی به دست می‌آیند. بنابراین برای کارا بودن روش شناسایی نقاط پرت، مقادیر بحرانی هر یک از آماره‌های توام برای انواع نقاط پرت با توجه به مدل برازش داده شده، با استفاده از شبیه سازی، به دست می‌آیند.

در این تحقیق برای دقت بیشتر با استفاده از شبیه‌سازی این مقادیر بحرانی برای مدل $VAR(2)$ با اندازه‌ی نمونه $n=117$ محاسبه شده است. در این شبیه سازی 15 تکرار و در هر تکرار یک بردار $VAR(2)$ با 117 مشاهده تولید می‌شود. سپس مقادیر آماره‌های J_{\max} و C_{\max} برای چهار نوع نقطه‌ی پرت به دست می‌آید. این مقادیر بحرانی برای سطوح مختلف در جدول 1 خلاصه شده است.

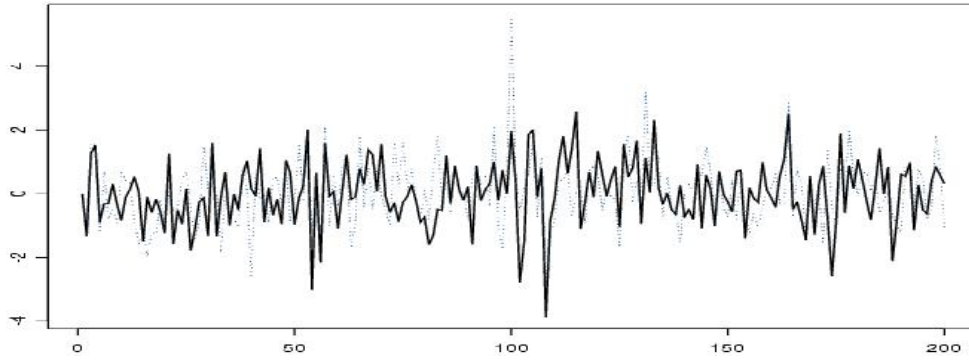
جدول ۱: نتایج شبیه سازی مقادیر بحرانی برای آماره‌های توام

	(i, h_i)	۵۰٪	۹۵٪	۹۷/۵٪	۹۹٪
	(I, h_I)	۱۲/۲	۱۸/۰۴	۱۸/۲۲	۲۱/۴
$J_{\max}(i, h_i)$	(A, h_A)	۱۲/۷	۱۷/۵۴	۱۷/۹۴	۱۹/۶۸
	(L, h_L)	۸/۲۹	۱۳/۶۵	۱۴/۰۲	۱۴/۴۱
	(T, h_T)	۱۱/۱۱	۱۵/۷۱	۱۶/۶۹	۱۷/۱
	(I, h_I)	۳/۰۷	۳/۶۶	۳/۷	۳/۹۶
$C_{\max}(i, h_i)$	(A, h_A)	۳/۱۲	۳/۶۷	۳/۹	۴/۱۷
	(L, h_L)	۲/۶۹	۳/۳۶	۳/۶۶	۳/۶۹
	(T, h_T)	۳/۰۱	۳/۷۸	۳/۹	۴/۰۳

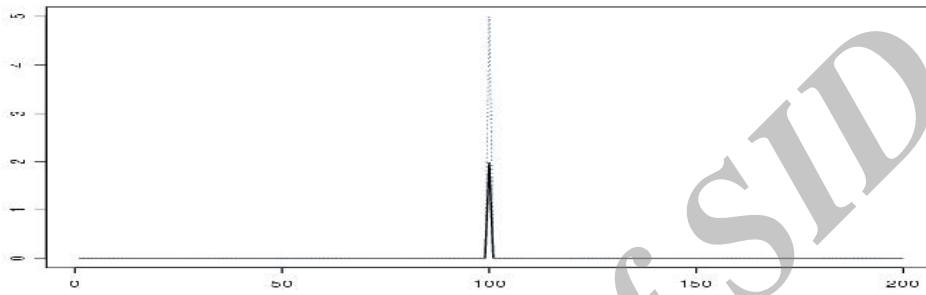
۵- شناسایی نقاط پرت در داده های قیمت سکه

به چند دلیل قیمت سکه می‌تواند الگوی مناسبی برای استفاده از نتایج بالا باشد، اول اینکه به دلیل کیفیت جنس با نوسانات قیمت طلا که منشا جهانی دارد دارای نوسانات زمانی است. از طرف دیگر چون به عنوان هدیه در زمان‌های خاصی در کشور ایران مورد استفاده قرار می‌گیرد دارای تغییرات فصلی است. تغییرات فصلی در قیمت سکه نیز منشا یکسانی ندارد چون خرید سکه هم در اعیاد ملی مورد استفاده قرار می‌گیرد هم در اعیاد مذهبی. تقارن این اعیاد ممکن است باعث تشدید یا کاهش قیمت سکه گردد. در کشورهایی مانند ایران که ثبات قیمت‌ها دستخوش حوادث اجتماعی و اقتصادی می‌باشند ممکن است تغییراتی در قیمت سکه به وجود آورند. شناسایی نقاط پرت و نوع آن‌ها در قیمت سکه از این جهت می‌تواند بسیار مفید باشد. بعضی از این اثرات بیرونی و تغییر قیمت به صورت موقتی (نقطه‌ی پرت جمع پذیر و یا تغییر موقت) و بعضی از این اثرات به صورت پایدار (نقطه‌ی پرت نوساز و یا تغییر سطح) است.

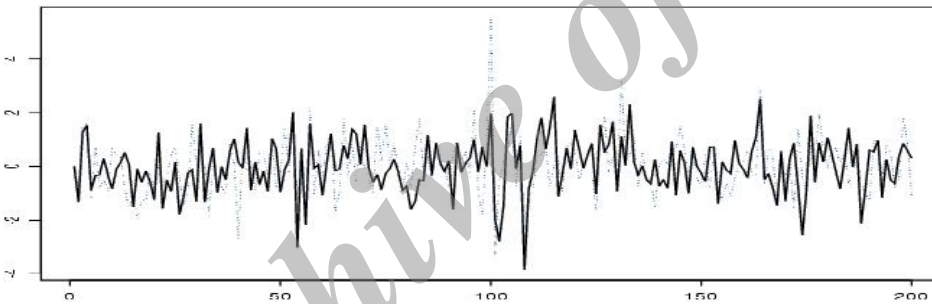
شکل (۱) باقیمانده‌های فرآیند $VAR(1)$ توأم با نقطه‌ی پرت نوساز، $\{a_t\}$.



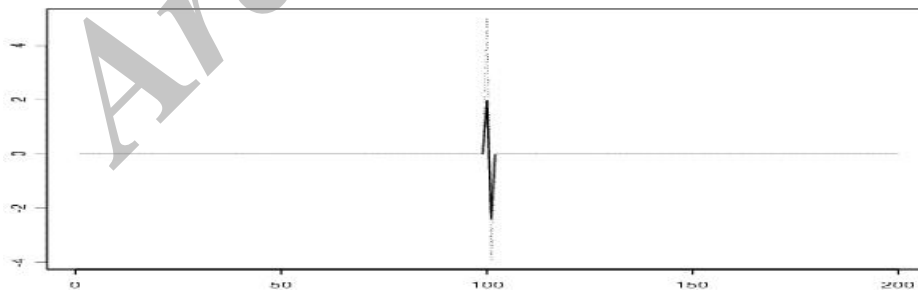
شکل (۲) سری $\{a_t - \varepsilon_t\}$ برای نقطه‌ی پرت نوساز.



شکل (۳) باقیمانده‌های $VAR(1)$ توأم با نقطه‌ی پرت جمع‌پذیر،

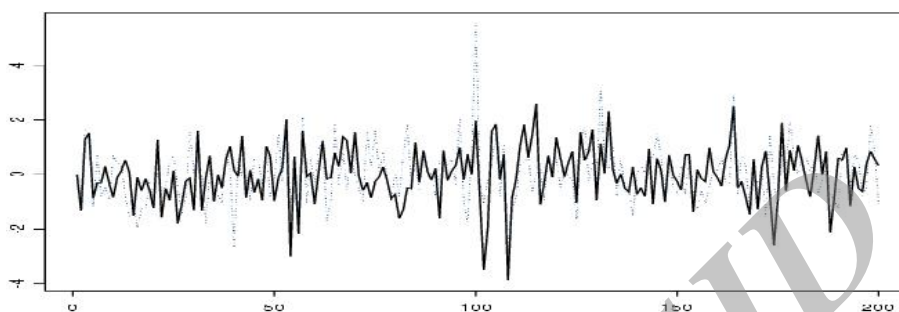


شکل (۴) سری $\{a_t - \varepsilon_t\}$ برای نقطه‌ی پرت جمع‌پذیر.

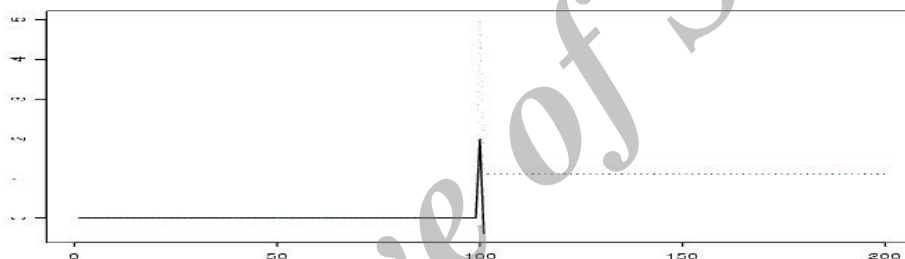


داده‌های قیمت سکه طلا به مدت ۱۱۷ ماه از فروردین ۱۳۷۸ تا آذر ۱۳۸۷ توسط بانک مرکزی جمهوری اسلامی ایران جمع آوری شده است. داده‌ها به صورت ماهانه بوده که از میانگین حسابی روزهای هر ماه به دست آمده است. با توجه به توصیه‌ی متعارف $\{a_t\}$ بانک مرکزی ایران برای روزهای تعطیل که قیمت رسمی وجود ندارد قیمت روز قبل از روز تعطیل در نظر گرفته می‌شود.

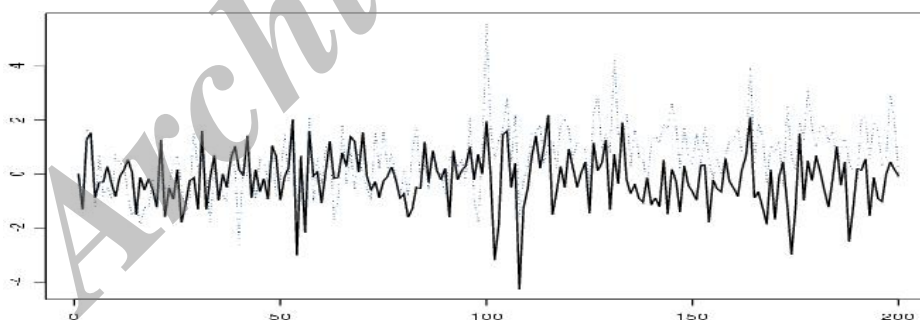
شکل ۵) باقیمانده‌های فرآیند $VAR(1)$ توأم با نقطه‌ی پرت تغییر سطح، $\{a_t\}$.

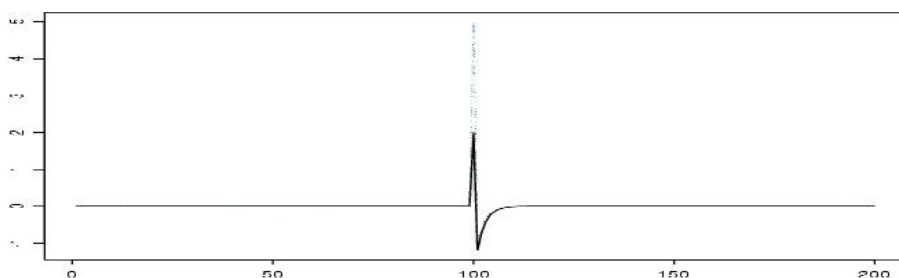


شکل ۶) سری $\{a_t - \varepsilon_t\}$ برای نقطه‌ی پرت تغییر سطح.



شکل ۷) باقیمانده‌های فرآیند $VAR(1)$ توأم با نقطه‌ی پرت تغییر موقت، $\{a_t\}$.



شکل ۸) سری $\{a_t - \varepsilon_t\}$ برای نقطه‌ی پرت تغییرموقت.

۵-۱- شناسایی و برآورد اثر نقاط پرت در داده‌های قیمت سکه به صورت یک متغیره

چن و لیو (۱۹۹۳) الگوریتمی براساس شیوه‌ی نسبت درست نمایی برای تشخیص نقاط پرت در سری‌های زمانی یک متغیره معرفی کرده‌اند. چینی پرداز و ذوالفقاری (۱۳۸۵) به شناسایی این الگوریتم پرداخته و آن را برای تشخیص نقاط پرت در داده‌های قیمت تمام سکه طرح قدیم به کار برده است. با توجه به اهمیت مقایسه این الگوریتم با الگوریتمی که برای تشخیص نقاط پرت در داده‌های سری زمانی به صورت چند متغیره در بخش‌های قبل معرفی شده در این قسمت با به کار بردن الگوریتم چن و لیو به شناسایی نقاط پرت در داده‌های سری‌های زمانی قیمت ربع سکه، نیم سکه و تمام سکه طرح جدید به صورت ماهانه پرداخته می‌شود. بنابراین تغییراتی که در قیمت سکه به صورت روزانه و یا هفتگی وجود داشته‌اند ولی تأثیر آن‌ها در قیمت سکه به صورت ماهانه نبوده است مورد بررسی این تحقیق نمی‌باشد. الگوریتم یک متغیره در هر تکرار، مشاهدات را به صورت جداگانه و به صورت یک به یک مورد بررسی قرار می‌دهد و آماره‌های آزمون برای چهار نوع نقطه‌ی پرت را محاسبه و پس از مقایسه این چهار آماره، آماره‌ای که در قدرمطلق ماکزیمم است را انتخاب، سپس با مقدار بحرانی مقایسه می‌کند چنان که این آماره از مقدار بحرانی بزرگتر باشد این مشاهده به عنوان نقطه‌ی پرت تشخیص داده می‌شود و اگر کوچکتر باشد مشاهده‌ی بعدی مورد بررسی قرار می‌گیرد. اگر این مشاهده به عنوان نقطه‌ی پرت شناسایی شود با توجه به نوع نقطه‌ی پرت، مقدار باقیمانده و مقدار مشاهده شده، اصلاح می‌شود. این مراحل برای تمام مشاهدات اجرا می‌گردد تا اینکه در پایان هر تکرار، چندین نقطه‌ی پرت شناسایی می‌شود و در تکرار بعدی داده‌ها و باقیمانده‌های اصلاح شده به کار می‌روند. الگوریتم زمانی خاتمه می‌یابد که

در تکراری هیچ نقطه‌ی پرتی شناسایی نشود. در پایان داده‌های اصلاح شده مورد بررسی و به آن‌ها مدل برازش داده می‌شود تا اثر نقاط پرت از برآورد پارامترها حذف شود. $i=IO, AO, LO, TO$ ، $Li = \frac{\hat{S}_{i,T}}{\sqrt{\hat{V}_{i,T}}}$ ، آماره‌ای است که الگوریتم تک

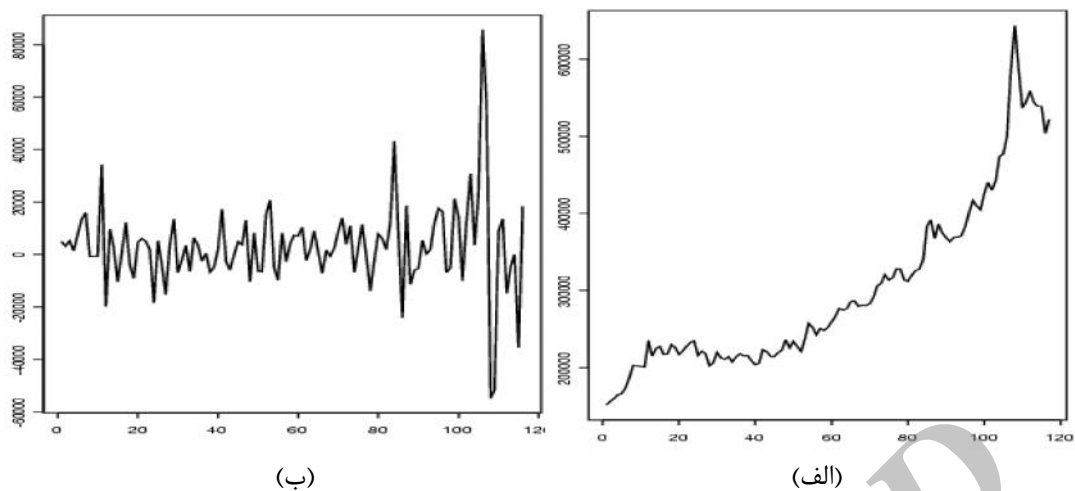
متغیره برای تشخیص نقاط پرت به کار می‌برد که در آن $\hat{S}_{i,T}$ برآورد اندازه‌ی نقطه‌ی پرت نوع i ام در زمان T و $\hat{V}_{i,T}$ برآورد واریانس $\hat{S}_{i,T}$ می‌باشد.

با توجه به نمودار سری زمانی و مقادیر خودهمبستگی و خودهمبستگی جزئی مشاهدات، سری زمانی داده‌های قیمت ربع سکه نایستا و دارای روند هستند. برای رفع نایستایی از روش تفاضلی کردن استفاده می‌شود. هر چند با تکیه به نمودار داده‌ها نمی‌توان به طور دقیق مکان نقاط پرت را تشخیص داد اما می‌توان به طور شهودی برخی از آن‌ها را تشخیص داد.

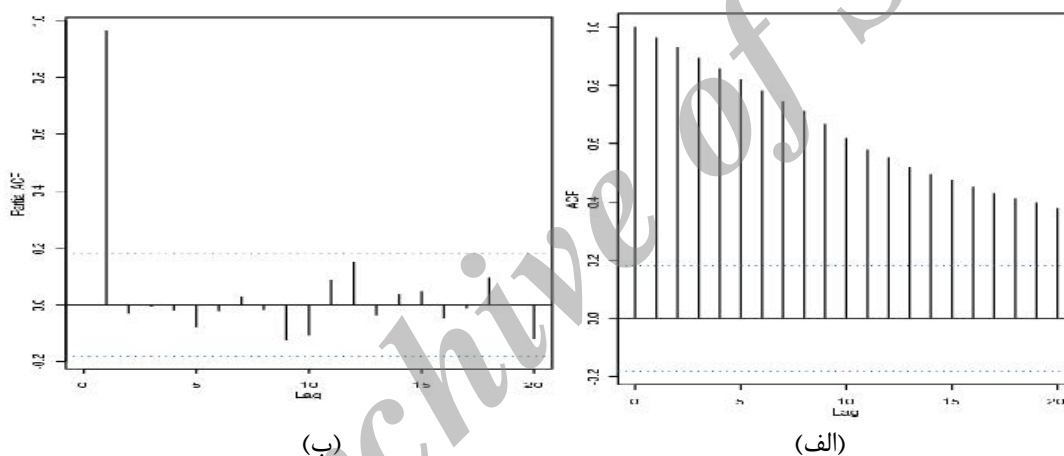
برای مثال با توجه به شکل ۹ قسمت الف، ممکن است در اواخر سال‌های ۱۳۷۸، ۱۳۸۶ و همچنین اوایل سال ۱۳۸۵ نقاط پرتی وجود داشته باشد. با توجه به نمودارهای ACF و PACF مدل‌های $ARIMA(3,1,0)$ ، $ARIMA(2,1,0)$ ، $ARIMA(3,1,1)$ برای برازش به داده‌ها مناسب می‌باشند.

با برآورد پارامترها و بررسی باقیمانده‌ها و هم چنین بررسی معیار اطلاع آکائیک و مقایسه مقدار احتمال هر یک از مدل‌ها، مناسب‌ترین مدل، $ARIMA(2,1,0)$ تشخیص داده می‌شود که در نتیجه برازش آن به داده‌ها قبل از حذف اثر نقاط پرت به صورت، می‌باشد. با توجه به نتایج جدول ۲ دیده می‌شود نقاط ۱۰۸ (اسفند ۱۳۸۶) ۸۵ (فروردین ۱۳۸۵)، ۱۲ (اسفند ۱۳۷۸) که به وسیله الگوریتم، نقطه‌ی پرت (که به طور شهودی از روی نمودار داده‌ها نیز معلوم هستند) و نیز نقاط ۸۷ (خرداد ۱۳۸۵)، ۱۱۶ (آبان ۱۳۸۷) (که از روی نمودار با داده‌های دیگر هماهنگ به نظر می‌رسند) به وسیله الگوریتم به عنوان نقطه‌ی پرت شناسایی شده‌اند. در نهایت مدل برازش داده شده به سری اصلاح شده بعد از حذف اثر نقاط پرت به صورت $Y_t = v_t + (1-B)(1-\alpha/0.5193B + \alpha/0.1033B^2)$ می‌باشد.

شکل ۹: قیمت ربع سکه الف: سری زمانی قیمت ربع سکه و ب: سری زمانی تفاضل اول قیمت ربع سکه



شکل ۱۰: الف: نمودار خود همبستگی قیمت ربع سکه و ب: نمودار خود همبستگی جزئی قیمت ربع سکه



نمودارهای ACF و PACF مربوط به باقیمانده‌های مدل برازش داده شده به داده‌های اصلاح شده بیانگر عدم همبستگی در فرآیند باقیمانده‌ها است. این مراحل برای نیم سکه و تمام سکه اجرا و نتایج حاصل از شناسایی نقاط پرت در این داده‌ها با استفاده از روش چن و لیو در جدول ۲ خلاصه شده است.

جدول ۲ نشان می‌دهد که نقاط ۱۰۷ (بهمن ۱۳۸۶) و ۱۱۶ (آبان ۱۳۸۷) در سری‌های ربع سکه و نیم سکه مشترک (نقطه‌ی ۱۰۷ در هر دو سری از نوع نقطه‌ی پرت نوساز و نقطه‌ی ۱۱۶ از نوع تغییر موقت) می‌باشند. شاید دلیل چنین تشابه‌ی تاثیر یکسان پیشامد خارجی بر دو نوع سری به صورت همزمان می‌باشد. نکته‌ی قابل توجه دیگری که از نتایج دیده می‌شود نزدیکی موقعیت نقاط پرت در این سری‌هاست. برای مثال نقطه‌ی ۱۲ (اسفند ۱۳۷۸) در سری‌های ربع سکه و نیم سکه و نقطه‌ی ۱۱ (بهمن ۱۳۷۸) در سری تمام سکه به طور متوالی رخ داده‌اند. با توجه به اینکه این نقاط پرت مربوط به ماه‌های پایانی سال می‌باشد. با توجه به مساله‌ی تورم که در پایان سال بیش تر رخ می‌دهد قیمت طلا و در نتیجه قیمت سکه تمام در این زمان دارای نوسانات بیشتری است.

افزایش قیمت طلا در بهمن ۱۳۷۸ و پرت شدن آن نقطه در تمام سکه و نه در ربع سکه و نیم سکه بیانگر این است که با افزایش قیمت طلا قیمت تمام سکه نسبت به قیمت ربع سکه و نیم سکه بیش تر تحت تأثیر قرار می‌گیرد.

جدول ۲: انواع نقاط پرت در داده‌های قیمت سکه در ایران به صورت یک متغیره

نوع سکه	تکرار	نوع نقطه ی پرت	آماره ی t	ω	زمان وقوع نقطه ی پرت
		نوساز	۲۰۵۰۵۵	۸۰۷۱۷/۹۵	۱۰۷ بهمن ۸۶
	اول	تغییر موقت	۳/۹۴۸۳۵۷	۴۴۲۳۴/۱۲	۱۰۸ اسفند ۸۶
		تغییر موقت	۳/۶۳۳۹۳۷	۳۹۴۴۸/۷۸	۱۱۶ آبان ۸۷
ربع سکه					
		جمع پذیر	۴/۳۶۴۵۳۱	۲۹۰۹۸/۲۷	۱۲ اسفند ۷۸
	دوم	نوساز	۳/۷۸۶۲۵۲	۴۰۵۰۲/۱	۸۵ فروردین ۸۵
		جمع پذیر	۳/۶۰۳۳۱۵	۲۰۴۶۸/۴	۸۷ خرداد ۸۵
		تغییر موقت	۴/۱۳۵۳۹۷	۱۱۰۷۸۳/۴	۱۲ اسفند ۷۸
	اول	جمع پذیر	۴/۳۲۳۹۸	۸۰۴۶۰/۱۳	۱۱۴ شهریور ۸۷
		تغییر موقت	۴/۰۲۴۲۹۴	۹۳۹۴۲/۸۸	۱۱۶ آبان ۸۷
نیم سکه					
	دوم	نوساز	۳/۹۱۲۴۷۸	۸۹۸۳۹/۲۴	۱۰۷ بهمن ۸۶
تمام سکه	اول	جمع پذیر	۴/۵۹۸۵۲۲	۰/۰۶۹۸۸۱۷۹۴	۷ مهر ۷۸
		جمع پذیر	-۴/۵۳۷۹۳۴	۰/۰۶۲۲۴۱۹۹	۱۱ بهمن ۷۸

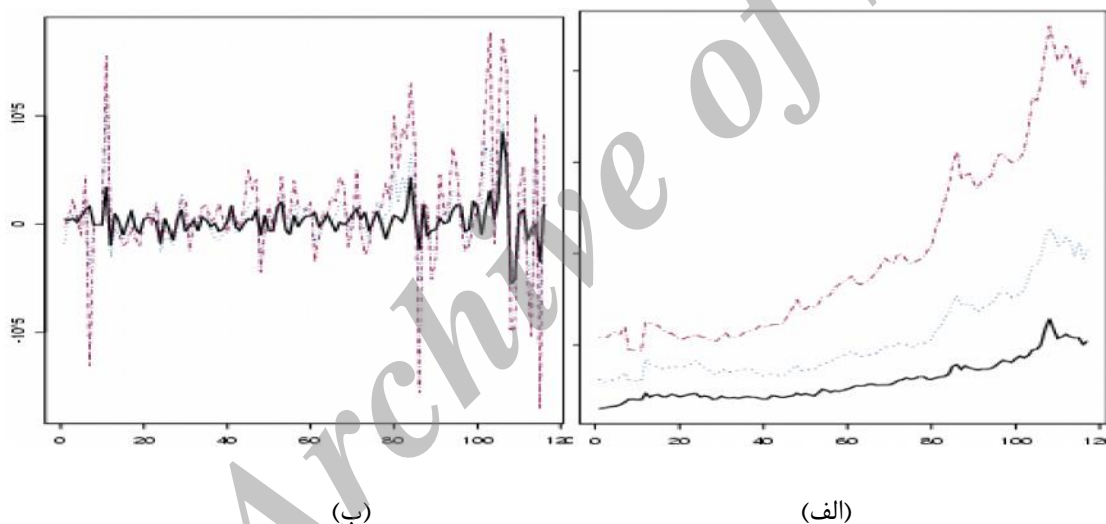
در مقابل تأثیر اوضاع بازار و فرهنگ خرید سکه به عنوان هدیه برای ربع سکه و نیم سکه بیشتر است.

با توجه به رابطه‌ی بین قیمت‌های ربع سکه، نیم سکه و تمام سکه و تأثیری که این قیمت‌ها بر هم دارند. بررسی آن‌ها به صورت یک متغیره باعث می‌شود که این ارتباط نادیده گرفته شود و محقق در تشخیص نقاط پرت دچار اشتباه شود. بنابراین در قسمت بعد به بررسی این سری‌ها به صورت توأم پرداخته می‌شود.

۶- تشخیص نقاط پرت در داده‌های قیمت سکه ی طلا به صورت چند متغیره

با توجه به رابطه‌ی بین قیمت‌های ربع سکه، نیم سکه و تمام سکه امکان تأثیر آنها بر یکدیگر، سری زمانی سه متغیره مورد بررسی قرار می‌گیرد. در شکل ۱۱ نمودار این سری سه متغیره به صورت ماهانه از فروردین ۱۳۷۸ تا آذر ۱۳۸۷ که در آن قیمت ربع سکه، مولفه‌ی اول با Y_1 ، قیمت نیم سکه، مولفه‌ی دوم با Y_2 و قیمت تمام سکه، مولفه‌ی سوم با Y_3 نشان داده شده است. ابتدا برای رفع ناپایداری از ماتریس تفاضل گیری یک مرتبه‌ای زیر استفاده شد (رینسل، ۱۹۹۷).

شکل ۱۱: سری سه متغیره: الف: نمودار سری سه متغیره قیمت سکه و ب: نمودار سری سه متغیره قیمت سکه پس از یک بار تفاضل گیری



$$D(B)Y_t = \begin{pmatrix} (1-B) & \circ & \circ \\ \circ & (1-B) & \circ \\ \circ & \circ & (1-B) \end{pmatrix} Y_t$$

با توجه به نمودارهای خودهمبستگی و همبستگی متقابل، چندین مدل برای داده‌های تفاضل گیری شده مورد مقایسه قرار گرفته و نتایج حاصل از برآزش دادن این مدل‌ها در جدول ۳ خلاصه شده است.

در اینجا $\left| \sum_m \right|$ دترمینان واریانس باقیمانده های مدل، AIC_m معیار اطلاع آکائیک مرتبه‌ی m می‌باشد، که از $AIC_m = \log \left(\left| \sum_m \right| \right) + 2mk^2 / (n - m)$ به دست می‌آید و هر چه این معیار کوچکتر باشد مدل برآزش داده شده بهتر است. آماره‌ی M_m به صورت:

$$M_m = -(n - m - mk - 1.5) \log (\det(S_m) / \det(S_{m-1}))$$

جدول ۳: نتایج حاصل از برآزش مدل‌های مختلف VAR به داده‌های تفاضل گیری شده

M(مرتبه ی VAR)	۱	۲	۳	۴
$\left \sum_m \right \times e^{-25}$	۳/۲۷	۲/۲۹	۲/۱۷	۱/۸۵
M_m	۲۰/۴۸	۳۸/۱۲	۵/۵۳۴	۱۶/۰۶
AIC_m	۵۸/۹۰	۵۸/۷۰	۵۸/۸	۵۸/۸
$X(2m \cdot 0 / 975)$	۱۴/۱۷	۱۶/۷۵	۱۸/۷۳	۲۰/۷۰

که در آن $S_m = \sum_{t=m+1}^n \hat{V}_t \hat{V}_t'$ دارای توزیع خی دو با $2m$ درجه‌ی آزادی و معنی‌دار بودن ضریب m ام در مدل $VAR(m)$ را مورد آزمون قرار می‌دهد. اگر مقدار این آماره از چندک مرتبه‌ی $2m$ ام توزیع خی دو بزرگتر باشد ضریب m ام در مدل VAR وجود دارد (رینسل، ۱۹۹۷). با توجه به جدول ۳ و آزمون $\Phi_2 = 0$ در مدل $VAR(2)$ مقدار آماره‌ی $M_2 = 38/12$ می‌باشد که از چندک توزیع خی دو با ۴ درجه‌ی آزادی بیشتر اشد بنابراین مدل $VAR(2)$ به عنوان مدل مناسب انتخاب می‌شود و برآورد کمترین مربعات خطا برای این مدل به صورت:

$$F_1 = \begin{pmatrix} 0/32 & -0/06 & 0/01 \\ 0/57 & -0/05 & 0/2 \\ 0/82 & -0/45 & 0/21 \end{pmatrix} \Phi_r = \begin{pmatrix} -0/75 & 0/35 & -0/03 \\ -1/06 & 0/31 & 0/05 \\ -2/18 & 0/72 & 0/076 \end{pmatrix}$$

خواهد بود. در نتیجه مدل برازش داده شده $VAR(2)$ ، یعنی $Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + V_t$ می‌باشد.

در این داده‌ها در تکرار ۱۵ مقادیر آماره J_{max} برای چهار نوع نقطه‌ی پرت از مقادیر بحرانی مربوطه کوچکتر و در نتیجه الگوریتم مسیر خود را با آماره C_{max} ادامه می‌دهد زیرا برای هیچ نوع نقطه‌ی پرتی آماره C_{max} در تکرار ۱۵ از مقادیر بحرانی مربوط به خود بیشتر نیست، الگوریتم خاتمه می‌یابد و نتایج حاصل از تشخیص نقاط پرت در جداول ۴ و ۵ آمده است. در مرحله بعد مکان، نوع و اندازه‌ی نقاط پرت با استفاده از نتایج جداول ۴ و ۵ تعیین و در جدول ۶ خلاصه می‌شود که در آن ستون‌های ۳، ۴ و ۵ اندازه‌ی نقطه‌ی پرت برای هر مولفه را به صورت توأم نشان می‌دهد.

جدول ۴: تشخیص نقاط پرت و نوع آن‌ها در داده‌های طلا با آماره J_{max}

تکرار	$J_{max}(Ih_I)$	$J_{max}(Ah_A)$	$J_{max}(Lh_L)$	$J_{max}(Th_T)$
۱	۳۳/۸۴(۱۰۷)	۲۵/۶۵(۸)	۹/۲۶(۱۰۹)	۱۳/۸۱(۸)
۲	۲۹/۲۹(۱۲)	۲۶/۹۴(۸)	۱۲/۸۴(۱۰۹)	۱۸/۱۶(۱۰۹)
۳	۲۳/۳۲(۸)	۲۷/۱۳(۸)	۱۳/۲۲(۱۰۹)	۱۸/۴۶(۱۰۹)
۴	۲۱/۶۵(۱۱۴)	۱۸/۸۳(۸۷)	۱۳/۷۱(۱۰۹)	۱۹/۳۸(۱۰۹)
۵	۱۹/۱(۱۱۵)	۱۹(۱۰۳)	۱۵/۸۳(۱۰۹)	۲۰/۱۷(۱۰۹)
۶	۱۸/۲۸(۱۱۵)	۱۹/۵۲(۸۷)	۸/۱۱(۱۰۳)	۱۸/۹۸(۱۰۳)
۷	۱۹/۹۲(۱۱۵)	۲۰/۹۳(۸۷)	۵/۱۵(۱۱۴)	۱۸/۲۸(۳۸)
۸	۱۹/۶۶(۱۱۵)	۲۰/۶۶(۱۰۵)	۵/۳۱(۱۱۴)	۱۸/۶۱(۳۸)
۹	۲۰/۳۰(۱۱۵)	۱۶/۱(۳۹)	۶/۱۷(۷۶)	۱۸/۶۹(۲۵)
۱۰	۲۰/۷۷(۱۰۸)	۱۷/۵۰(۳۹)	۶/۶(۳۵)	۲۰/۹۱(۳۸)
۱۱	۲۱/۶۱(۱۰۸)	۱۵/۹۲(۱۰۸)	۵/۴۵(۹۷)	۲۳/۳۵(۲۵)
۱۲	۲۱/۹۸(۱۰۸)	۱۷/۳۲(۱۳)	۵/۴۸(۹۷)	۱۲/۷۳(۱۳)
۱۳	۱۵/۱۹(۱۳)	۱۷/۸۷(۱۳)	۵/۱۵(۱۱۶)	۱۳/۵۸(۱۸)
۱۴	۱۵/۳۱(۱۳)	۱۸/۳۱(۱۳)	۶/۵۶(۷۶)	۱۴/۲۸(۱۳)
۱۵	۱۴/۵۸(۴۹)	۱۵/۸۲(۴۹)	۶/۸۲(۷۶)	۱۳/۹۳(۱۸)
مقادیر بحرانی	۱۸/۲۲	۱۷/۹۴	۱۴/۰۲	۱۶/۶۹

جدول ۵: تشخیص نقاط پرت و نوع آن‌ها در داده‌های طلا با آماره C_{\max}

تکرار	$C_{\max}(Ih_T)$	$C_{\max}(Ah_A)$	$C_{\max}(Lh_L)$	$C_{\max}(Th_T)$
۱۵	۳/۶۸(۴۹)	۳/۶۸(۴۹)	۲/۴۳(۷۶)	۳/۲۹(۸۴)
مقادیر بحرانی	۳/۷	۳/۹	۳/۶۶	۳/۹

همان طور که در جدول ۶ دیده می‌شود بعضی از نقاط پرت تعیین شده توسط این الگوریتم از نمودار قابل تشخیص و بعضی غیر قابل تشخیص‌اند. نکته جالب اینکه سال ۱۳۸۴ که با حدس اولیه از نمودار نقطه پرت به نظر می‌رسید با این الگوریتم پرت تشخیص داده نشد.

مدل برازش داده شده به داده‌ها پس از حذف اثر نقاط پرت به صورت:

$$Y_t = \begin{pmatrix} -0.32 & 0.053 & 0.054 \\ -0.33 & -0.03 & 0.12 \\ -1.35 & -0.6 & 0.14 \end{pmatrix} Y_{t-1} + \begin{pmatrix} -0.57 & 0.16 & -0.06 \\ -1.18 & 0.45 & -0.15 \\ -1.94 & 1.25 & -0.44 \end{pmatrix} Y_{t-2} + V_t$$

به دست می‌آید.

جدول ۶: نقاط پرت و نوع آن‌ها در داده‌های طلا

زمان	نوع	\hat{S}_1	\hat{S}_2	\hat{S}_3
۱۰۷ بهمن ۸۶	نوساز	۷۸۱۲۲/۹۳	۸۱۵۲۶/۱۷	۱۴۷۹۳۸/۹۸
۱۲ اسفند ۷۸	نوساز	۳۱۳۳۶/۸۶	۱۱۱۷۰۴/۶۶	۱۴۲۶۸۵/۷۵
۸ آبان ۷۸	جمع پذیر	۱۱۰۶۲/۴۷	۴۵۷۳۰/۷۰	-۱۴۱۹۴۰/۲۲
۱۱۴ شهریور ۸۷	نوساز	۴۰۱۱/۳۱۳	-۷۶۳۴۸/۸۸۸	-۱۰۱۱۱۷/۸۸
۱۰۹ فروردین ۸۷	تغییر موقت	۲۸۶۰۸/۴۶	-۱۹۳۱۱/۴۱	-۳۹۲۴۶/۷۸
۱۰۳ مهر ۸۶	تغییر موقت	۲۰۸۵۰/۱۲	۶۴۱۰۹/۴	۱۲۹۴۶۵/۲۸
۸۷ خرداد ۸۵	جمع پذیر	۲۳۵۴۱/۷۱	-۴۳۹۱۳/۱۵	-۱۳۳۵۶۲/۴۳
۱۰۵ آذر ۸۶	جمع پذیر	۲۷۴۷۷/۷۱	-۶۱۸۷۴/۱۰	-۱۴۰۵۰۳/۸۱
۱۱۵ مهر ۸۷	نوساز	۲۴۶۹/۴۵	۵۸۰۷۳/۱۸	۸۱۰۴۳/۱۳
۳۸ اردیبهشت ۸۱	تغییر موقت	۶۴۴۳/۸	-۲۳۴۰۴/۴۱	-۷۷۹۱/۸۷
۲۵ فروردین ۸۰	تغییر موقت	۱۷۵۵۴/۴۸	-۳۱۸۷۸/۲۵	-۲۷۱۴۴/۴۷
۱۰۸ اسفند ۸۶	نوساز	۳۲۲۵۴/۲۶	۱۳۹۲۹/۲	۵۹۲۸۵/۲۹
۱۱۶ آبان ۸۷	جمع پذیر	۱۱۳۱۹/۳۷	-۴۶۸۰۰/۱۸	-۱۱۰۶۳۶/۱۶
۱۳ فروردین ۷۹	جمع پذیر	۲۵۷۱۳/۱۷	-۲۸۷۵۷/۶۸	-۱۸۳۱۵/۵۴

۷- مقایسه نتایج چند متغیره با حالت یک متغیره

تفاوت اساسی الگوریتم‌های یک متغیره و چند متغیره در این است که در روش اول برای هر مشاهده ثابت، چهار نوع نقطه پرت بررسی می‌شوند. ماکزیمم آماره برای

انواع نقطه پرت محاسبه می‌شود. در صورت معنی‌دار بودن فرض که از مقایسه ماکزیمم آماره آزمون چهار نقطه پرت به دست می‌آید، نقطه پرت و نوع آن شناسایی می‌شود. هر گاه مشاهده توأم با نقطه پرت باشد داده‌های بعد از آن و باقیمانده‌ها اصلاح می‌شوند. در ادامه مشاهده بعدی مورد بررسی قرار می‌گیرد. این روند تا انتهای مشاهدات ادامه می‌یابد. در سری زمانی چند متغیره تمرکز ابتدا روی نوع نقطه پرت و نه مشاهده صورت می‌گیرد. ابتدا برای هر کدام از نقاط پرت آماره آزمون برای n مشاهده محاسبه و ماکزیمم آماره برای تمام مشاهدات که آماره توأم نامیده می‌شود، به دست می‌آید. بنابراین در هر تکرار ماکزیمم آماره توأم برای چهار نقطه پرت محاسبه و با مقایسه این آماره‌ها با مقدار بحرانی که از روش‌های شبیه سازی حاصل شده انواع نقاط پرت که آماره متناظر آنها از مقدار بحرانی بیشتر شده معلوم می‌گردد. در نهایت از بین نقاط پرت معنی‌دار شده در این تکرار ماکزیمم آماره‌ها به عنوان نقطه پرت اصلی همراه با نوع آن مشخص می‌شود. سپس داده‌های بعد از آن و باقیمانده‌ها اصلاح می‌گردند. این کار برای تکرارهای بعدی انجام می‌گردد تا زمانی که در یک تکرار هیچ نقطه پرتی تشخیص داده نشود. به عبارتی می‌توان گفت در سری زمانی یک متغیره آماره آزمون منفرد و در چند متغیره آماره توأم n تایی است.

اکنون می‌توان نتایج به دست آمده در یک سری زمانی یک متغیره و چند متغیره را مقایسه کرد. نقطه ۱۰۷ (بهمن ۱۳۸۶) قیمت ربع و نیم سکه در یک متغیره و چند متغیره نقطه پرت نوساز تشخیص داده شده‌اند. مشابهت نوع نقطه پرت بیانگر تأثیر یک پیشامد خارجی همزمان بر این دو نوع سکه است که اثر آن در تمام سکه زیاد نبوده است. در مقابل نقطه ۱۱۶ (آبان ۱۳۸۷) در هر دو روش نقطه پرت تشخیص داده شده اما در یک متغیره نقطه پرت تغییر موقت و در چند متغیره جمع پذیر است. احتمالاً علت این مسئله تأثیرپذیری پیشامد خارجی از رابطه بین انواع سکه است. با چنین استدلالی می‌توان با استفاده از جدول ۲ و ۶ نتایج دو روش را مقایسه نمود.

باید توجه کرد علاوه بر تشخیص نقطه پرت تشخیص نوع آن نیز دارای اهمیت است. برای خریدار تغییرات نقطه ۱۲ (اسفند ۱۳۷۸) موقتی است که بعد از مدت کوتاهی اثر آن از بین رفته و قیمت سکه متعادل خواهد شد. در حالی که با

استدلالی مبتنی بر چند متغیره تأثیر این افزایش قیمت تا پایان سری زمانی ادامه دارد.

نکته دیگر قابل ذکر در مقایسه دو روش تعداد نسبتاً بیشتر نقاط پرت در چند متغیره است (مقایسه جداول) است که نشان می‌دهد حساسیت تشخیص نقاط پرت در الگوریتم چند متغیره نسبت به یک متغیره بیشتر است. تنها استدلالی که در این رابطه می‌توان کرد رابطه بین قیمت انواع سکه‌ها است. تفاوت دیگر دو روش تشخیص نقاط پرت در یک متغیره و تشخیص نقاط مجاور آن به وسیله چند متغیره است (مهر و بهمن ۱۳۷۸ در یک متغیره و در مقابل آبان و اسفند ۱۳۷۸ در چند متغیره).

۸- بحث و نتیجه گیری

اگر مشاهده‌ی پرت در داده‌ها وجود داشته باشد برآوردهایی اریب از پارامترها حاصل می‌شود که منجر به نتایج نامطلوب و اشتباه می‌گردد و در نهایت پیش‌بینی‌های قابل اعتماد نخواهند بود. در داده‌های سری زمانی چند متغیره که مولفه‌های سری با هم در ارتباط هستند چنین مشکلی به دلیل نادیده گرفتن اثر متقابل ای مولفه‌ها بیشتر است. مقایسه‌ی الگوریتم‌های معرفی شده برای حالت یک متغیره و چند متغیره نشان داد الگوریتم چند متغیره، در تشخیص نقاط پرت نسبت به روش یک متغیره حساسیت بیشتری دارد و ممکن است نقاط پرتی که به دلیل نادیده گرفتن اثر مولفه‌های سری زمانی در تحلیل یک متغیره مغفول مانده‌اند.

فهرست منابع:

چینی پرداز، ر. و ذوالفقاری، پ. (۱۳۸۵). شناسایی نقاط پرت در سری‌های زمانی، مجله علم و فناوری، ۳ و ۴، ۱۴۳-۱۳۰.

مختاری، الف. (۱۳۸۹). بررسی نقاط پرت در داده‌های سری‌های زمانی چند متغیره، پایان نامه کارشناسی ارشد دانشگاه شهید چمران اهواز.

Box, G. E. P & Tiao, G.C. (1975). Intervention Analysis With application to Environmental and Economic Problems, Journal of American Statistical Association, 70, 70-79.

Chen. C. & Liu, L. M. (1993). Joint estimation of Model Parameters and Outlier Effects in Time Series, Journal of American Statistical Association, 88, 284-297.

Fox, A. J. (1972). Outliers in time series, Journal of the Royal Statistical Society, Ser. B, 34, 350-363.

Galeano, P., Pena, D. & Tsay, R.S. (2004). Outlier Detection in Multivariate Time Series Via Projection Pursuit, Tech. Report, University of Carlos III. <http://ideas.Repec.Org/p/cte/wsrepe/ws044211.html>.

Glendinning, R. H. (1998). Determining the Order of an ARMA Model from Outlier Contaminated Data, Communications in Statistics – Theory and Methods, 27, 13-40.

Otey, M. E. & Parthasarathy, S. (2005). A dissimilarity measure for comparing subset of data: Application to Multivariate Time Series, In: Proceedings of ICDM Workshop on Temporal Data Mining: Algorithms, Theory and Applications, The Fifth IEEE International Conference on Data Mining, 101-112.

Pankratz, A. E., Pena, D. & Tsay, R. S. (2000). Outliers in Multivariate Time Series, Journal of Biometrika, 87, 789-804.

Reinsel, G. (1997). Elements of Multivariate Time Series Analysis, Springer Series in Statistics, 2nd ed, New York.

Scot, S. (1987). The Impact of Outliers on Seasonal Adjustment, Office of Research and Forecasting, Bureau of Labor Statistics.

Weng, X. Q. & Shen, J. Y. (2008). Detecting Outlier Samples in Multivariate Time Series Dataset, Knowledge Based Systems, 21, 807-812.

Weng, X. Q. & Shen, J. Y. (2006). Outlier Mining for Multivariate Time Series Based on Local Sparsity Coefficient, in: Proceeding of the 6th World Congress on Intelligent Control and Automation, 5957-5961.

Archive of SID