

## مطالعه و بررسی روندها در یک سری زمانی از داده‌های تابعی

زهرة شیشه بر<sup>۱</sup>، سید مرتضی نجیبی<sup>۱</sup>، سکینه رمضانی<sup>۱</sup>

تاریخ دریافت: ۹۸/۳/۲۹

تاریخ پذیرش: ۹۹/۳/۱

چکیده:

دنباله‌ای از توابع (منحنی‌ها) که در طول زمان جمع‌آوری می‌شوند را یک سری زمانی تابعی گویند. فراوانی چنین مشاهداتی تحلیل سری‌های زمانی تابعی را به یکی از شاخه‌های محبوب تحقیقاتی در علم آمار تبدیل کرده است. هدف اصلی از تحلیل سری زمانی تابعی، پیش‌بینی و توصیف کمی مکانیسم‌های تصادفی است که منجر به تولید توابع شده است. در این راستا نیاز است سری زمانی تابعی به مؤلفه‌های روند، دوره‌های زمانی و خطا تجزیه شود. اما قبل از تجزیه نیاز به شناسایی و تشخیص این گونه مؤلفه‌ها داریم. از این رو در این مقاله یک روش ناپارامتری برای بررسی و تشخیص وجود روند در یک سری زمانی تابعی با استفاده از توابع رکورد معرفی شده است. سپس با پیاده‌سازی و استفاده از این روش در یک سری زمانی تابعی نحوه کاربرد آن مورد تحقیق قرار گرفته است. در پایان نیز کارایی این روش برای تشخیص روند در یک مجموعه از داده‌های واقعی نرخ باروری مورد بررسی قرار گرفته است.

**واژه‌های کلیدی:** داده‌های تابعی، سری زمانی تابعی، روند، تابع رکورد، تابع ماکسیمم.

### ۱ مقدمه

مدنظر قرار می‌گیرد. ۱- تشخیص مؤلفه روند، ۲- تشخیص مؤلفه دوره‌های منظم، ۳- مؤلفه تصادفی (خطا).

در سال‌های اخیر مطالعات زیادی در حوزه FDA صورت گرفته‌اند (برای مطالعه بیشتر [۱۱، ۱۳، ۱۶] را ببینید). اما نسبتاً مطالعات بسیار کمتری در حوزه FTS انجام شده‌اند. به طور مثال مدل اتورگرسیو تابعی از درجه ۱ توسط [۳] معرفی شد و سپس توسط سایرین مورد مطالعه قرار گرفت [۴، ۱۰]. در عمده تحقیقات پیشین FTS به مبحث پیش‌بینی سری‌های زمانی با استفاده از مدل‌های پارامتری و استنباط و آزمون فرضیه در خصوص پارامترهای مدل پرداخته شده است. در حالی که محققین کمتری به شناسایی و تشخیص روند و دوره‌های زمانی در FTS پرداخته‌اند. برای این منظور اولین بار [۷] روند در FTS را با استفاده از تابع رکورد تعریف کرده‌اند و آن را برای تشخیص روند در ایستگاه‌های قطب جنوب اسپانیا بکار برده‌اند. همچنین [۸] آزمون‌های ناپارامتری به منظور شناسایی دوره‌های زمانی یک مجموعه داده سری زمانی تابعی را معرفی کرده‌اند. هدف اصلی این مقاله تحلیل روند در FTS با استفاده از آزمون فرضیه پیشنهادی [۷] است. برای این منظور با روش محاسبه آماره آزمون و مقادیر بحرانی به کمک یک مثال شبیه‌سازی شده آزمون فرضیه پیشنهادی و مزیت‌ها و معایب آن مورد بررسی قرار می‌گیرد.

در ادامه این مقاله، در بخش دوم مفاهیم و تعاریف مورد نیاز برای تعریف روند

بسیاری از پدیده‌ها ذاتاً توابع پیوسته‌ای هستند که مدل‌سازی آن‌ها به کمک روش‌های چند متغیره کلاسیک به خوبی انجام نمی‌گیرد. از این رو به جای تحلیل آن‌ها در فضای برداری با بعد متناهی، لازم است آن‌ها را در فضای تابعی با بعد نامتناهی در نظر بگیریم. چنین اقدامی به رویکرد جدیدی در آمار تحت عنوان تحلیل داده‌های تابعی<sup>۲</sup> (FDA) منتهی می‌شود [۱۳]. در این رویکرد داده‌ها به فضای تابعی با بعد نامتناهی متعلق هستند و لذا برای تحلیل آن‌ها ناگزیر به کاهش بعد هستیم. یکی از روش‌ها برای این کار استفاده از توابع پایه متناهی برای توسیع یک فضای نامتناهی است. از آنجا که در عمل عموماً داده‌ها به صورت گسسته اندازه‌گیری می‌شوند، اولین گام در FDA برازش منحنی یا توابعی مناسب است که خواص داده‌ها را به روشی خاص خلاصه کند. برازش منحنی‌های مناسب به داده‌های گسسته با استفاده از روش‌های رگرسیون ناپارامتری مثل اسپلاین‌های هموارساز<sup>۳</sup> امکان‌پذیر است.

در بسیاری از پدیده‌های تابعی، یک سری زمانی از توابع مشاهده شده است که مدل کردن وابستگی زمانی آن‌ها با استفاده از روش‌های تحلیل سری‌های زمانی تابعی<sup>۴</sup> (FTS) انجام می‌پذیرد. هدف اصلی از تحلیل سری زمانی و در حالت خاص FTS، پیش‌بینی و توصیف کمی مکانیسم‌های تصادفی است که منجر به تولید داده شده‌اند. در این راستا در تحلیل یک FTS سه موضوع

<sup>۱</sup>دانشگاه شیراز، شیراز، ایران

<sup>۲</sup>Functional Data Analysis

<sup>۳</sup>Smoothing Splines

<sup>۴</sup>Functional Time Series

## ۱.۲ روند افزایشی برای FTS

دنباله توابع  $\chi = \{x_t(s), s \in [a, b]\}_{t=1}^T$  از  $T$  تابع حقیقی که بر فاصله‌ی  $[a, b]$  تعریف شده‌اند را در نظر بگیرید که برای راحتی کار از فاصله‌ی  $[0, 1]$  به جای  $[a, b]$  استفاده می‌کنیم. گوییم دنباله  $\{x_t(s)\}_{t=1}^T$  دارای روند افزایشی است اگر

$$x_1(s) \leq x_2(s) \leq \dots \leq x_T(s), s \in [0, 1].$$

همان‌طور که می‌بینید انتظار مشاهده چنین یکنواختی در هر نقطه از  $s$  برای تعیین روند در کل فاصله‌ی زمانی، در عمل بسیار دور از واقعیت است. جهت تعدیل این مفهوم با استفاده از تعریف تابع رکورد برای همه منحنی‌ها تعریف واقعی‌تری از روند افزایشی ارائه می‌دهیم. کلاس توابع پیوسته که روی  $[0, 1]$  تعریف می‌شوند را با  $C[0, 1]$  نشان می‌دهیم. تابع

$$r_t : C[0, 1]^t \rightarrow C[0, 1]$$

را ماکسیمم  $t$  تابع تعریف می‌کنیم. در واقع مقدار تابع  $r_t$  در نقطه  $s$  که با نماد  $r_t(s)$  نشان داده می‌شود برابر است با

$$r_t(s) \equiv \max(x_1(s), x_2(s), \dots, x_t(s)), s \in [0, 1]$$

و نسبت زمانی که منحنی  $t$  منطبق بر تابع ماکسیمم بوده است را با  $\omega_t$  نشان می‌دهیم. همان‌طور که مشخص است این مقدار متناسب است با طول فاصله‌ی زمانی که نمودار  $\lambda$  روی ماکسیمم مقدار تا تابع  $t$  قرار می‌گیرد و به صورت زیر قابل تعریف است:

$$\omega_t \equiv \int_0^1 I_{\{r_t(s)=x_t(s)\}} ds. \quad (3)$$

در ادامه با استفاده از سه تعریف متفاوت روند قوی تا ضعیف، نحوه انجام آزمون زیر را پایه‌ریزی می‌کنیم:

(۴)

$$\begin{cases} H_0 : \text{"در دنباله‌ای از سری‌های زمانی تابعی روند افزایشی وجود ندارد"} \\ H_1 : \text{"در دنباله‌ای از سری‌های زمانی تابعی روند افزایشی وجود دارد"} \end{cases}$$

مفهوم فرض صفر این است که بخش‌هایی از دنباله‌های سری زمانی از توابع موجود در دنباله  $\chi$ ، تا لحظه  $t$  مشابه یکدیگر و با احتمال یکسان برابر مقدار  $r_t(s)$  هستند. بنابراین انتظار می‌رود تحت فرض صفر مقدار،  $\omega_t = \frac{1}{T}$  برای تمام  $t$  ها ( $t = 1, \dots, T$ ) باشد. در مقابل اگر دنباله‌ای از منحنی‌ها بیشتر از زمان مورد انتظار منطبق بر منحنی ماکسیمم باشند، نشان‌دهنده وجود یک روند افزایشی است. به زبان ریاضی اگر  $\omega_t > \frac{1}{T}$  برای  $t = 1, \dots, T$  رخ دهد فرض صفر رد می‌شود. در ادامه به منظور انجام آزمون فرضیه ناپارامتری فوق سه تعریف متفاوت از روند افزایشی ارائه شده است.

در FTS ارائه می‌شود. در بخش سوم با ارائه آزمون فرضیه‌های تشخیص روند در FTS و معرفی ناحیه‌های بحرانی متناسب آن‌ها، روش پیشنهادی فریدمن و همکاران معرفی شده است. در بخش چهارم با استفاده از یک مثال شبیه‌سازی شده و داده‌های واقعی نرخ باروری تجزیه و تحلیل روند در FTS بررسی و نتایج حاصل از آن ارائه شده است. در پایان در بخش پنجم یافته‌های تحقیق جمع‌بندی و پیشنهادهایی به منظور ادامه تحقیق ارائه خواهد شد.

## ۲ تحلیل روند در FTS

به‌طور کلی روند در سری‌های زمانی به صورت تغییرات متغیر مورد مطالعه در یک دوره‌ی طولانی بدون در نظر گرفتن تغییرات دوره‌ای فصلی و نامنظم تعریف می‌شود [۱]. اما تاکنون تعریف مشخصی برای روند سری زمانی از توابع ارائه نشده است. در این بخش به معرفی روند در یک سری زمانی از داده‌های تابعی بر اساس تابع رکورد می‌پردازیم. بدون از دست دادن کلیت مسئله در اینجا تنها به معرفی روند افزایشی در یک سری زمانی تابعی می‌پردازیم. جهت معرفی روند افزایشی ابتدا لازم است تابع رکورد را به عنوان تابع ماکسیمم در نظر بگیریم. بسط این نظریه برای بررسی روند کاهشی یا دوطرفه نیز به راحتی قابل انجام است. به‌طور خاص بررسی روند کاهشی با استفاده از تابع مینیمم به عنوان تابع رکورد قابل تعریف است. در ادامه سه نوع روند افزایشی که از قوی تا ضعیف طبقه‌بندی می‌شوند ارائه خواهد شد.

فرض کنید  $T$  تابع حقیقی داریم که هر کدام از توابع در  $N$  نقطه مشاهده شده‌اند و به صورت  $(x_t(s_1), \dots, x_t(s_N))$  برای  $t = 1, \dots, T$  ثبت شده‌اند. توجه شود که فاصله نقاط  $s_1, \dots, s_N$  می‌تواند غیر یکسان باشد. به منظور استفاده از FDA ابتدا لازم است مشاهدات را به صورت تابع پیوسته‌ای در نظر بگیریم. برای این منظور عموماً با استفاده از توابع پایه و هموارسازی مشاهدات هر تابع را به کمک ترکیب خطی از تعدادی توابع پایه منتهای به صورت زیر می‌نویسیم [۱۳]:

$$\mu_j(s_j) \approx \sum_{m=1}^M c_{tm} B_m(s_j), 1 \leq t \leq T, \quad (1)$$

که در آن  $\{B_m(s), m = 1, \dots, M\}$  مجموعه‌ای از توابع پایه مانند پایه‌های موجک، اسپلاین، فوریه و... هستند و  $c_{tm}$  بردار ضرایب پایه است که با استفاده از روش کمترین توان‌های دوم خطا در معادله رگرسیونی ناپارامتری زیر برآورد شده است:

$$x_t(s_j) = \mu_t(s_j) + e_t(s_j), t = 1, \dots, T, j = 1, \dots, N \quad (2)$$

که در آن  $\{e_t(s_j), j = 1, \dots, N\}$  متغیرهای تصادفی مستقل و هم توزیع با میانگین صفر، نشان‌دهنده‌ی نوسانات تصادفی در اطراف میانگین  $\mu_t(s_j)$  هستند.

**تعریف ۳.۲.** روند افزایشی  $k$ -ضعیف دنباله  $X$  را گوئیم دارای روند افزایشی  $k$ -ضعیف است اگر  $k \in \mathbb{C}$  و وجود داشته باشد به طوری که

$$\omega_{[k]} > 0, \forall k \geq k_0$$

که در آن،  $\omega_{[1]} < \omega_{[2]} < \dots < \omega_{[T]}$ ،  $t = 1, \dots, T$  و  $\omega_t \equiv \omega_t - \frac{1}{t}$  آماره‌های ترتیبی هستند.

### ۳ آزمون‌های ناپارامتری برای روند افزایشی در FTS

در این قسمت ما آزمون‌های ناپارامتری را جهت بررسی روند افزایشی برای یک دنباله از FTS بر اساس سه تعریف ذکر شده در بخش قبل ارائه می‌دهیم. دنباله  $\{X_t(s), s \in [0, 1]\}_{t=1}^T$  که در آن  $X_t(s)$  یک متغیر تصادفی که مقادیر آن در فضای  $L^1_{[0,1]}$  است را در نظر بگیرید. متغیرهای تصادفی  $R_t(s)$  را به صورت زیر در نظر بگیرید:

$$R_t(s) = \max(x_1(s), x_2(s), \dots, x_t(s)), 0 \leq s \leq 1,$$

$$t \in \mathbb{C} \equiv \{2, \dots, T\}.$$

سپس به منظور محاسبه نسبت زمانی که تابع  $X_t(s)$  برابر  $R_t(s)$  است متغیر تصادفی  $W_t$  را به صورت زیر تعریف می‌کنیم:

$$W_t \equiv \int_0^1 I_{\{R_t(s)=X_t(s)\}} ds.$$

این کمیت، با مقیاس استاندارد شده  $0 \leq s \leq 1$ ، نسبتی از زمان است که  $X_t$  با تابع ماکسیمم برای داده  $t$  ام منطبق می‌شود و مقدار مورد انتظار آن به صورت زیر محاسبه می‌شود:

$$J_t \equiv E(W_t) = \int_0^1 P(X_t(s) = R_t(s)) ds.$$

در آزمون فرضیه (۴)، تحت فرض صفر هیچ روندی در دنباله  $\{X_t(s), s \in [0, 1]\}_{t=1}^T$  وجود ندارد، به عبارتی منحنی‌ها در دنباله قابل تعویض اند. بنابراین برای هر  $t \in \mathbb{C}$ ،  $J_t = \frac{1}{t}$  است. زیرا  $t$  تابع داریم که هر کدام در بازه  $[0, 1]$  مقدار می‌گیرند و به طور متوسط انتظار داریم تنها مقدار  $\frac{1}{t}$  از هر تابع برابر مقدار ماکسیمم شود. بنابراین برای آزمون این فرضیه که منحنی خاص  $X_t$  یک تابع رکورد تابع  $t$  است، تنها کافی است آزمون زیر را برای تابع مشاهده شده در زمان  $t$  انجام دهیم.

$$\begin{cases} H_{0,t} : J_t = 1/t \\ H_{1,t} : J_t > 1/t \end{cases} \quad (۸)$$

**تعریف ۱.۲.** روند افزایشی قوی: دنباله  $X$  دارای روند افزایشی قوی است اگر

$$0 < \omega_2 - \frac{1}{2} < \omega_3 - \frac{1}{3} < \dots < \omega_T - \frac{1}{T} \quad (۵)$$

و

$$0 > \omega_2 - \frac{1}{2} > \omega_3 - \frac{1}{3} > \dots > \omega_T - \frac{1}{T} \quad (۶)$$

برقرار باشند که در آن مشابه رابطه (۳) و بر اساس تابع مینیمم به صورت  $\omega_t \equiv \min(x_1(s), x_2(s), \dots, x_t(s)), s \in [0, 1]$  و  $\omega_t \equiv \int_0^1 I_{\{x_t(s)=x_i(s)\}} ds$  تعریف می‌شود.

به دلیل تقارن در نامعادله‌های (۵) و (۶) در اینجا فقط با معادله‌های (۵) کار خواهیم کرد و آزمون‌های مربوطه برای نامعادله (۶) می‌تواند به طور مشابه به دست آید. در نهایت به کمک تصحیح بونفرونی انجام هر دو آزمون به طور هم‌زمان برای دو نامعادله در نظر گرفته می‌شود. برای آزمون وجود یک روند افزایشی در یک دنباله از توابع، فرضیه  $H_a$  را به صورت زیر در نظر می‌گیریم.

”روندی مطابق با نامعادله (۵) وجود دارد.“  $H_a$

یا به عبارت دیگر، در این آزمون بررسی می‌شود که آیا دنباله  $\{\omega_t - \frac{1}{t}\}$  برای  $t = 1, \dots, T$  یک دنباله کاملاً مثبت افزایشی از  $t$  است؟ در این آزمون شرط کاملاً مثبت افزایشی بودن، بسیار محدودکننده و سخت‌گیرانه است. به عبارتی دیگر این روش ایجاب می‌کند هر تابع رکوردی نسبت به تمام توابع قبل داشته باشد. از این رو می‌توان روش یا تعریفی متعادل‌تر و با محدودیت کمتر به صورت زیر ارائه کرد:

**تعریف ۲.۲.** روند افزایشی ضعیف: دنباله  $X$  تابعی  $X$  دارای روند افزایشی ضعیف است اگر فرضیه مقابل را به صورت زیر در نظر بگیریم:

$$H_b : \min_{t \in \mathbb{C}} (\omega_t - \frac{1}{t}) > 0, \mathbb{C} \in \{2, \dots, T\} \quad (۷)$$

از نامعادله (۷) داریم به ازای هر  $t \in \mathbb{C}$  این عبارت برای یک  $t$  ثابت بدین معناست که منحنی  $t$ م بیش از  $\frac{1}{t}$  اوقات بر روی منحنی ماکسیمم منطبق می‌شود. مشخص است که این شرط، شرط ضعیف‌تری از روند افزایشی نسبت به تعریف ۱.۲ است و محدودیت کمتری نسبت به معادله (۵) دارد. علی‌رغم اینکه شرط  $\min_{t \in \mathbb{C}} (\omega_t - \frac{1}{t}) > 0$  در معادله (۷) محدودیت کمتری نسبت به (۵) دارد، اما هنوز هم می‌تواند برای بسیاری از مسائل کاربردی به سختی به کار گرفته شود. از آنجا که شرط  $(\omega_t - \frac{1}{t}) > 0$  برای هر  $t$  بایستی برقرار باشد، نیاز است که منحنی  $t$ م در دنباله، به معنای ضعیف، با توجه به تمام توابع گذشته یک تابع رکورد باشد، به جای آن در ادامه الگوی معنی‌داری از روند افزایشی را به کمک تعریف ضعیف‌تری از روند افزایشی ارائه می‌دهیم.

پیوسته در نظر بگیریم. به طور مشخص اگر منحنی  $t$  ام،  $t = 1, \dots, T$  تنها در  $N$  نقطه  $s_1, \dots, s_N$  مشاهده شود، آنگاه  $T$  بردار  $N$  بعدی از داده‌های اولیه به صورت زیر در اختیار داریم:

$$(x_1(s_1), x_1(s_2), \dots, x_1(s_N))$$

$$(x_2(s_1), x_2(s_2), \dots, x_2(s_N))$$

$$(x_T(s_1), x_T(s_2), \dots, x_T(s_N))$$

که بر روی آن‌ها یک منحنی رگرسیونی ناپارامتری به صورت زیر برازش شده است:

$$X_t(s_j) = \mu_t(s_j) + e_t(s_j), t = 1, \dots, T, j = 1, \dots, N \quad (11)$$

تحت فرض  $H_0$ ،  $\mu_t(s)$  مستقل از  $t$  فرض شده است. به عبارتی دیگر تحت این فرض روندی وجود ندارد و برای هر  $t$   $\mu_t(s) = \mu(s)$  است. از این رو نسبتی از زمان که منحنی  $t$  ام، بر منحنی ماکسیمم توابع منطبق می‌شود را می‌توان به سادگی با نوسانات تصادفی  $\{e_t(s_1), e_t(s_2), \dots, e_t(s_N)\}$  محاسبه کرد. در این صورت توابع  $R_t(s)$  و  $e^{(t)}(s)$  که برابر ماکسیمم مقدار تابع و خطا از  $t$  منحنی اول در نقطه مکانی خاص  $s_j$  است به صورت  $R_t(s_j) \equiv \max(X_1(s_j), X_2(s_j), \dots, X_T(s_j))$  و  $e^{(t)}(s_j) \equiv \max(e_1(s_j), e_2(s_j), \dots, e_T(s_j))$  تعریف می‌شود.

توجه شود که فرض مستقل و هم توزیع بودن  $\{e_t(s_j), j = 1, \dots, N\}$ ، سبب می‌شود که  $\{e^{(t)}(s_j), j = 1, \dots, N\}$  نیز مستقل و هم توزیع شود. بنابراین با نماد گذاری فوق آماره مناسب برای بررسی آزمون  $(\lambda)$ ، مدت زمانی است که منحنی  $\lambda$  ام در بردار ماکسیمم تا منحنی  $\lambda$  ام صرف می‌کند که آن را  $\tau_\lambda$  می‌نامیم و به صورت زیر تعریف می‌کنیم.

$$\tau_\lambda \equiv \sum_{j=1}^N I_{\{R_\lambda(s_j) = X_\lambda(s_j)\}},$$

تحت فرض صفر آزمون  $(\lambda)$  مبنی بر عدم وجود روند در معادله رگرسیونی  $(11)$ ، دارای توزیع دوجمله‌ای  $B(N, p)$   $N$  آزمایش و احتمال پیروزی  $p$  به صورت:

$$p \equiv P(R_t(s_j) = X_t(s_j)) = P(e_t(s_j) = e^{(t)}(s_j)) \\ = 1/t, t = 1, \dots, T$$

است. بنابراین آزمون فرضیه  $(\lambda)$  را می‌توان با در نظر گرفتن نسبتی از زمان که  $\{X_t(s_j), j = 1, \dots, N\}$  بر حداکثر منحنی تا تابع  $t$  منطبق می‌شود، به صورت  $J_t \equiv \frac{\tau_t}{N}$  آزمون کرد. در این آزمون اگر  $J_t - \frac{1}{t} > c\alpha$  باشد، فرض صفر رد می‌شود.

به بیانی دقیق تر مقدار ناحیه بحرانی  $c\alpha$  را می‌توان بر اساس سطح معنی داری از

همچنین اگر بخواهیم آزمون هم‌زمان برای همه توابع  $\{X_t, t = 1, \dots, T\}$  را برای فرض  $(\lambda)$  انجام دهیم، کافی است از تصحیح بونفرونی استفاده کنیم. همان‌طور که می‌دانیم تصحیح بونفرونی با فرض استقلال بین آزمون‌های هم‌زمان، سطح معنی داری را کنترل می‌کند. درحالی‌که در واقعیت و به خصوص در سری زمانی، همبستگی بین آزمون‌ها سبب بسیار محافظه کارانه رفتار کردن این نوع روش می‌شود. از این رو این نوع آزمون‌ها در عمل از کارایی لازم برخوردار نیستند و لازم است یک آزمون کلی برای ارزیابی وجود روند افزایشی در تعریف  $2.2$  و  $3.2$  بر اساس یک آماره آزمون مشخص ارائه شود. برای این منظور آزمون پیچیده‌تر زیر را در نظر بگیریم:

$$\begin{cases} H_0 : \min_{t \in \mathbb{C}} (J_t - 1/t) = 0 \\ H_A : \min_{t \in \mathbb{C}} (J_t - 1/t) > 0 \end{cases} \quad (9)$$

که در آن  $\mathbb{C} \equiv \{2, \dots, T\}$  است. فرضیه مقابل در این آزمون نشان‌دهنده این موضوع است که روند افزایشی ضعیفی بر اساس تعریف  $2.2$  برای بخشی از این سری زمانی تابعی وجود داشته است. توجه شود که ساختمان این آزمون با آزمون‌های هم‌زمان  $(\lambda)$  هیچ تفاوتی ندارد و تنها در اینجا سعی به ارائه‌ی یک آماره‌ی آزمون داریم. همچنین اگر  $\bar{J} \equiv (\bar{J}_2, \bar{J}_3, \dots, \bar{J}_T)$  که در آن  $\bar{J}_t = (J_t - 1/t)$  را در نظر بگیریم، آنگاه می‌توان آماره‌های ترتیبی آن را به صورت زیر تعریف نمود:

$$\bar{J}_{[2]} < \bar{J}_{[3]} < \dots < \bar{J}_{[T]}.$$

بدیهی است که آزمون فرضیه  $(9)$  به آزمون  $H_0 : \bar{J}_{[2]} = 0$  مقابل  $H_A : \bar{J}_{[2]} > 0$  تبدیل خواهد شد. بر همین اساس می‌توان این آزمون را برای آزمون فرضیه تعریف  $3.2$  که فرضی ضعیف تر است، به صورت زیر مدل بندی نمود:

$$\begin{cases} H_0 : \bar{J}_{[k]} = 0, \quad \forall k \in \mathbb{C} \\ H_{A1} : \bar{J}_{[k_0]} > 0, \quad k \text{ برای برخی} \end{cases} \quad (10)$$

دقت کنیم هرگاه  $\bar{J}_{[k_0]} > 0$  برای یک  $k_0$  آنگاه  $k_0 > k, \forall k$  لازم به ذکر است که آزمون بر مبنای تعریف  $1.2$  ارائه نشده است و این یک مسئله باز است.

## ۴ محاسبه آماره آزمون و مقادیر بحرانی

در عمل چون هر تابع در تعداد محدودی از نقاط مشاهده شده است، برای انجام آزمون‌های ارائه شده در بخش قبل لازم است ابتدا با استفاده از مدل  $(2)$  که در بخش دوم به آن پرداخته شد، مشاهدات گسسته را به صورت منحنی‌های

رتبه مربوط به  $X_t(s_j)$  را  $\lambda_{tj}$  نشان می‌دهیم. این روش را برای تمام  $s_j$ ها و  $j = 1, \dots, N$  و  $t \in \{1, \dots, T\}$  تکرار می‌کنیم. بردار رتبه‌های مربوط به  $t$  امین منحنی مشاهده شده  $(X_t(s_1), X_t(s_2), \dots, X_t(s_N))$  را به صورت  $\Lambda_t = (\lambda_{t1}, \lambda_{t2}, \dots, \lambda_{tN}) \in \{1, \dots, T\}^N$  تعریف می‌کنیم.

**گام ۲.** تصویر تابعی با جایگشت رتبه‌ها روی شبکه مشبک: در این گام بدون در نظر گرفتن توابع مشاهده شده  $\{X_t(s_1), X_t(s_2), \dots, X_t(s_N)\}_{t=1}^T$  به‌طور مستقیم بردارهای رتبه تحقق‌یافته از روی شبکه مشبک  $\{1, \dots, T\}^N$  که با استفاده از الگوریتم زیر ساخته می‌شود را به دست می‌آوریم:

- $\Lambda_1$  را با انتخاب یک مقدار برای  $\lambda_{1j}$  به صورت تصادفی از  $\{1, \dots, T\}$  برای هر  $j = 1, \dots, N$  می‌سازیم.
- $\Lambda_{1(i>1)}$  را با انتخاب یک مقدار برای  $\lambda_{1j}$  به‌طور تصادفی از باقیمانده توابع  $\{\lambda_{1j}, \dots, \lambda_{(i-1)j}\} \setminus \{1, \dots, T\}$  هر  $j = 1, \dots, N$  می‌سازیم.
- این کار را برای ساختن تمام  $\Lambda_1, \Lambda_2, \dots, \Lambda_T$  تا زمانی که مجموعه مسیرها به‌طور کامل شبکه  $\{1, \dots, T\}^N$  را پر کنند، ادامه می‌دهیم.

**گام ۳.** محاسبه مقادیر آماره آزمون بر اساس بردار رتبه تحقق‌یافته تحت فرض صفر: برای هر بردار تحقق‌یافته رتبه  $(\Lambda_1, \Lambda_2, \dots, \Lambda_T)$ ، می‌توان مقدار آماره آزمون،  $\bar{J}_{[k]}$  یا  $\bar{J}_{[r]}$  را محاسبه کنیم. به‌عنوان مثال فرض کنید مطابق شکل ۱ (الف) تابع ۴ تابع بر روی ۵ نقطه مشاهده شده است در اختیار داریم. با انجام گام‌های ۱ و ۲، بردارهای رتبه تحقق‌یافته به صورت کاملاً تصادفی و به صورت جدول ۱ دیده خواهند شد. همان‌طور که در این مثال مشخص است بعد از محاسبه مقادیر  $t = 1, \dots, T$   $\bar{J}_t - \frac{1}{t}$  برای تمام منحنی‌ها می‌توان مقادیر  $\bar{J}_{[k]}$  یا  $\bar{J}_{[r]}$  را بر اساس سه گام بالا و با استفاده از بردار رتبه‌های تصادفی تولید شده، به دست آورد. با  $n$  بار تکرار سه گام بالا و تولید رتبه‌های تصادفی در هر مرحله، تعداد  $n$  آماره آزمون تحت فرض صفر به صورت تصادفی تولید شده است. در صورتی که  $n$  به اندازه کافی بزرگ باشد می‌توان به خوبی توزیع آماره آزمون تحت فرض صفر را تقریب زد. بنابراین تنها کافی است برای آزمون فرضیه (۹) یا (۱۰) در سطح  $\alpha$  چندک  $1 - \alpha$  توزیع آماره آزمون تحت فرض را استخراج کنیم و آن را با مقدار آماره آزمون برای داده‌های مشاهده شده، مقایسه کنیم. اگر مقدار آماره آزمون در ناحیه رد فرض صفر قرار گرفت، آزمون را رد و در غیر این صورت می‌پذیریم.

پیش تعیین شده  $\alpha$  برای آزمون دوجمله‌ای  $t = 1, \dots, T$ ،  $\bar{J}_t - \frac{1}{t} = 0$  تعیین کرد.

در ادامه به منظور به دست آوردن آماره مناسبی برای آزمون (۹)، اولین آماره ترتیبی از بردار  $(\bar{J}_1 - \frac{1}{1}, \bar{J}_2 - \frac{1}{2}, \dots, \bar{J}_T - \frac{1}{T})$  را به صورت زیر تعریف می‌کنیم.

$$\bar{J}_{[r]} \equiv \min_{t \in C} (\bar{J}_t - \frac{1}{t})$$

به‌طور مشابه با توسع همین روش می‌توان آماره آزمون مناسب برای آزمون فرضیه (۱۰) را به صورت اولین  $k$ ای که  $\bar{J}_{[k]} > 0$  شود در نظر گرفت. در نهایت برای به دست آوردن ناحیه بحرانی آزمون‌های (۹) یا (۱۰)، فرضیه صفر زمانی رد می‌شود که به ترتیب  $\bar{J}_{[r]} > 0$  یا  $\bar{J}_{[k]} > 0$  بزرگ‌تر از یک مقدار بحرانی به دست آمده بر اساس خطای نوع اول  $\alpha$  باشند. در بخش بعد نحوه محاسبه مقدار بحرانی این آزمون‌ها ارائه خواهد شد.

## ۱.۴ محاسبه مقادیر بحرانی

در بخش‌های قبلی آزمون‌های متفاوتی برای آزمون روند در FTS ارائه شدند. اما با توجه به عدم استقلال متغیرهای تصادفی  $\{\bar{J}_t - \frac{1}{t}, t \in T\}$  محاسبه مقادیر دقیق بحرانی به دلیل پیچیدگی‌های زیادی که در ساختمان ترکیباتی مسئله وجود دارد، دشوار است. لذا لازم است جهت استفاده از آماره آزمون در انجام آزمون فرضیه‌های معرفی شده، ابتدا توزیع هر یک را با استفاده از شبیه‌سازی از مقدار آماره آزمون تحت فرض صفر تقریب بزنیم. در این مقاله بر اساس جایگشت بین توابع و نقاط آن‌ها با رویکردی که مشابهت زیادی با آزمون ناپارامتری جایگشت دارد، توزیع آماره آزمون فرضیه‌های معرفی شده را تقریب می‌زنیم.

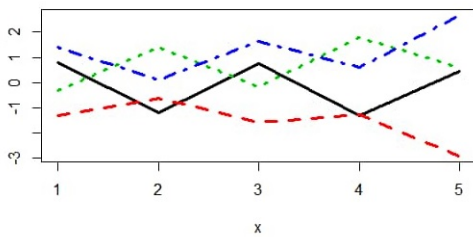
همان‌طور که دیدیم مقادیر داده‌ها تأثیری در آماره آزمون‌های معرفی شده در بخش سوم ندارند و تنها رتبه‌های آن‌ها در محاسبه آماره آزمون نقش دارند. از این رو در ادامه بر اساس مقادیر رتبه‌بندی شده  $X_t(s_j)$  به صورت  $\{X_1(s_j), X_2(s_j), \dots, X_T(s_j)\}$  برای هر  $j = 1, \dots, N$ ، گام‌های لازم جهت شبیه‌سازی از یک مقدار آماره آزمون تحت فرض را ارائه می‌دهیم.

## مراحل شبیه‌سازی آماره آزمون

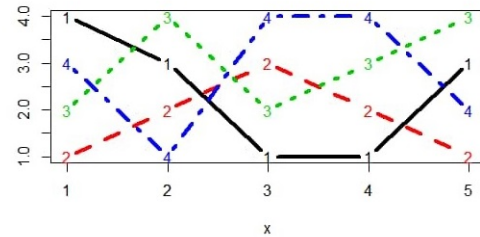
**گام ۱.** تبدیل هر تابع به بردار رتبه متناظر خود: ابتدا در هر نقطه  $s_j$  مقادیر مشاهده شده  $\{X_1(s_j), X_2(s_j), \dots, X_T(s_j)\}$  را رتبه‌بندی می‌کنیم و سپس

	$f_1(s_j)$	$f_2(s_j)$	$f_3(s_j)$	$f_4(s_j)$
$s_1$	۴	۱	۲	۳
$s_2$	۳	۲	۴	۱
$s_3$	۱	۳	۲	۴
$s_4$	۱	۲	۳	۴
$s_5$	۳	۱	۴	۲

جدول ۱: رتبه منحنی نام  $(f_i)$  در مکان  $(s_j)$



الف



ب

شکل ۱: مثال مفهومی تولید مقادیر تحقق یافته رتبه‌ها: الف: چهار تابع که هر یک در پنج نقطه مشاهده شده است. ب: توابع تحقق یافته از بردار رتبه‌های تولید شده در گام ۱ و ۲ شبیه‌سازی

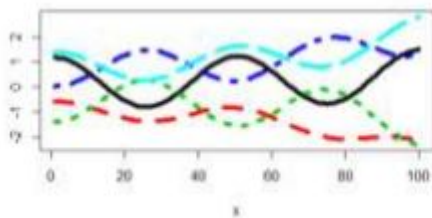
## ۵ مثال از داده‌های واقعی

### ۱.۵ مثال شبیه‌سازی شده

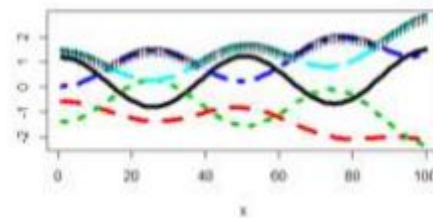
منحنی تابع رکورد (ماکسیمم) در بین پنج تابع با خط شکسته مشخص شده است. در این مثال آماره‌های ترتیبی بردار  $\{\bar{J}_1 - \frac{1}{5}, \bar{J}_2 - \frac{1}{5}, \bar{J}_3 - \frac{1}{5}, \bar{J}_4 - \frac{1}{5}, \bar{J}_5 - \frac{1}{5}\}$  به صورت زیر به دست آمده‌اند:

$$(\bar{J}_{[1]} = 0.41, \bar{J}_{[2]} = 0.34, \bar{J}_{[3]} = -0.03, \bar{J}_{[4]} = -0.05, \bar{J}_{[5]} = 0.41)$$

ابتدا جهت درک بهتر مطالب بالا با مثالی ساده به بررسی وجود روند در بین پنج تابع شبیه‌سازی شده در شکل ۱ (الف) می‌پردازیم. در شکل ۱ (ب)



الف



ب

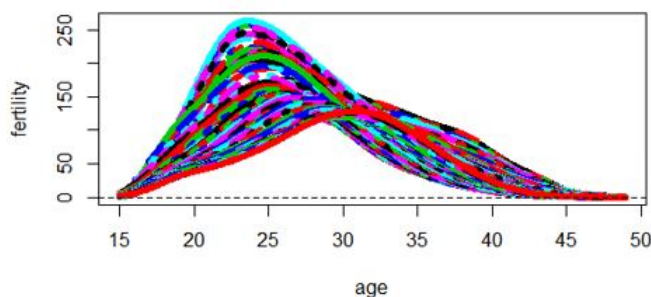
شکل ۲: الف: پنج تابع که هر یک در ۱۰۰ نقطه مشاهده شده‌اند. ب: منحنی تابع ماکسیمم برای این پنج منحنی

$${}^5 f(s, t) = e^{s^2} \cos(2\pi t \omega) + \cos\left(\frac{2\pi s}{2}\right) \sin(2\pi t \omega), \quad 0 \leq s \leq 100, \quad t = 1 : 5, \quad \omega = 0.22$$

## ۲.۵ داده‌های واقعی

در این بخش، با استفاده از روش‌های گفته شده در بخش چهارم به بررسی وجود روند و تشخیص آن در نرخ باروری بر حسب سن در استرالیا، با استفاده از نرم‌افزار  $R$  می‌پردازیم. داده‌های مربوط به میزان باروری بر حسب سن در استرالیا برای سال‌های ۱۹۲۱ تا ۲۰۰۶ برای افراد بین ۱۵ تا ۴۹ سال جمع‌آوری شده که در بسته  $fts_a$  در نرم‌افزار  $R$  موجود می‌باشند. در شکل ۲ منحنی‌های مربوط به این داده‌ها برای ۸۹ سال رسم شده‌اند که منحنی  $t$  ام مربوط به داده‌ی سال  $t$  ام می‌باشد که با استفاده از ۲۱ تابع پایه  $B$ -اسپلاین هموار شده است.

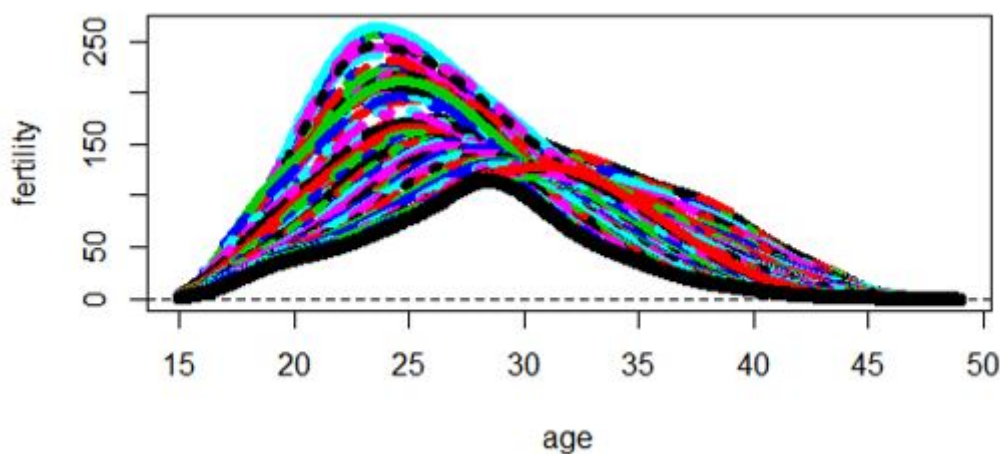
بنابراین مقدار آماره آزمون (۹) برای این پنج تابع برابر  $\bar{A}_{[7]} = -0.5$  است و به منظور انجام آزمون در سطح معنی‌داری  $\alpha = 0.05$  تنها کافی است مقدار چندک  $0.95$  تحت فرض صفر را به دست آوریم. برای این منظور با استفاده از روش سه مرحله‌ای بخش چهارم، مقدار چندک  $0.95$  برابر صفر به دست می‌آید. از آنجا که مقدار آماره آزمون کوچک‌تر از مقدار چندک  $0.95$  (صفر) است، فرض عدم وجود روند افزایشی ضعیف در بین این پنج تابع پذیرفته می‌شود. حال آزمون (۱۰) را با  $k_0 = 4$  مورد بررسی قرار می‌دهیم. برای این آزمون چندک  $0.95$  مقدار  $0.06$  به دست آمده است که با مقایسه آن با مقدار آماره آزمون  $(10) \bar{A}_{[7]} = 0.34$  فرض وجود روند افزایشی  $k$ -ضعیف با  $k_0 = 4$  در این  $FTS(H_1)$  پذیرفته می‌شود.



شکل ۳: توابع هموار شده میزان باروری در استرالیا بر حسب سن از سال ۱۹۲۱ تا ۲۰۰۶ (هر رنگ نشان‌دهنده منحنی مربوط به یک سال است)

منحنی بر منحنی رکورد (در اینجا مینیمم) تا سال منطبق است را به دست آوریم. به عنوان مثال در شکل ۳ منحنی رکورد تا سال ۸۶ ام با خط ضخیم نشان داده شده است.

با توجه به داشتن روند تقریباً نزولی در منحنی‌های باروری بر حسب سن در طول زمان، حدس خود را جهت بررسی وجود روند نزولی در این  $FTS$  آزمون می‌کنیم. برای این منظور ابتدا لازم است تعداد نقاطی که هر



شکل ۴: منحنی تابع مینیمم برای ۸۶ سال

مربوط به آزمون (۹) را به صورت زیر تعریف می‌کنیم:

$$\bar{J}_{[2]} \equiv \min_{t \in T} (J_t - \frac{1}{t}), T = 2, \dots, 86$$

فرض صفر این آزمون رد می‌شود اگر  $\bar{J}_{[2]}$  بزرگ‌تر از یک مقدار بحرانی مطلوب باشد. مقدار آماره آزمون (۹) برای این داده‌ها در بین ۸۶ سال مورد نظر،  $-0.384$  و مقدار چندک  $0.99$  بر اساس هزار بار شبیه‌سازی از آماره آزمون با استفاده از تکرار روش سه مرحله‌ای توضیح داده‌شده در بخش چهارم،  $-0.45$  به دست آمده است. از این رو فرض صفر در آزمون (۹) در سطح معنی‌داری  $0.01$  رد می‌شود یعنی این داده‌ها دارای روند کاهشی ضعیف هستند.

$$\bar{J}_{[2]} \equiv \min_{t \in T} (J_t - \frac{1}{t}) = -0.348 > -0.45, T = 2, \dots, 86$$

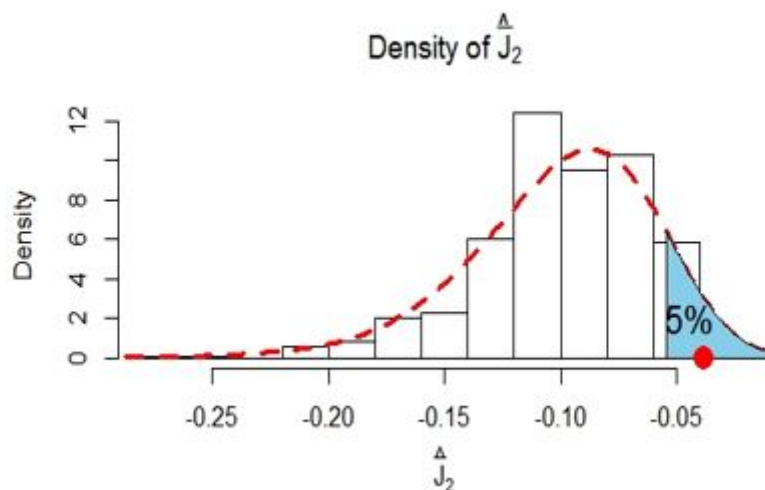
برای درک بیشتر در شکل ۴ تابع چگالی هسته‌ای برآورد شده از آماره‌های شبیه‌سازی شده در آزمون (۹) تحت فرض صفر رسم شده است و ناحیه بحرانی آزمون به صورت ناحیه آبی رنگ و مقدار آماره آزمون ( $-0.384$ ) به صورت یک نقطه قرمز رنگ مشخص شده است. همان‌طور که می‌بینیم مقدار آماره آزمون در ناحیه رد فرض صفر آزمون (۹) برای سطح معنی‌داری  $0.01$  قرار دارد.

حال فرضیه نزولی بودن روند در FTS بالا را با استفاده از آزمون فرضیه‌های (۸)، (۹) و (۱۰) که در بخش سه به توضیح آن پرداختیم، بررسی می‌کنیم.

در آزمون (۸)، به آزمودن تنها این فرضیه که منحنی خاص  $X_t$  یک تابع رکورد تا سال  $t$  است، پرداخته شد. به این منظور آزمون‌های تکی به فرم زیر را در نظر می‌گیریم:

$$\begin{cases} H_{0,t}: J_t = 1/t \\ H_{1,t}: J_t > 1/t \end{cases}, t = 1, \dots, 86$$

همان‌طور که گفته شد برای انجام آزمون فرضیه (۸) تنها کافی است نسبتی از زمان که  $\{X_t(s_j), j = 1, \dots, 35\}$  بر حداکثر منحنی تا سال منطبق می‌شود را به صورت  $\frac{t_j}{N} \equiv \bar{J}_t$  به دست آورد. سپس با انجام آزمون‌های دو جمله‌ای به صورت هم‌زمان برای تمام سال‌ها ( $t = 1, \dots, 86$ ) فرض صفر آزمون را رد یا پذیرش کرد. در اینجا بر اساس تابع رکورد مینیمم و با استفاده از تصحیح بونفرونی و مقایسه  $p$ -مقادیرها با سطح معنی‌داری اصلاح شده  $\frac{\alpha}{86}$  برای ۶۲ سال از ۸۶ سال فرض صفر آزمون رد می‌شود. بنابراین دلیلی برای رد هم‌زمان فرض صفر آزمون‌های تکی نداریم. از آنجا که رویکرد ترکیب نتایج حاصل از آزمون‌های تکی در عمل از کارایی لازم برخوردار نیستند، بنابراین به بررسی آزمون (۹) جهت تشخیص روند در این منحنی‌ها می‌پردازیم. آماره آزمون

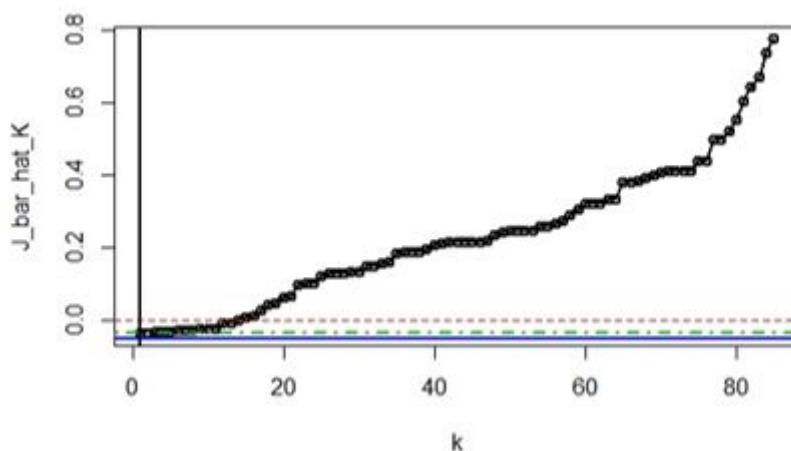


شکل ۵: توزیع آماره آزمون (۹) تحت فرض صفر

$k > 1$ ,  $k - 1$  امین آماره ترتیبی از عناصر  $\bar{J}$  است. فرض صفر هر دو آزمون به ترتیب زمانی رد می‌شود که  $\bar{J}_{[k]}$  یا  $\bar{J}_{[2]}$  از مقدار بحرانی به دست آمده از شبیه‌سازی بزرگ‌تر باشد. بنابراین زمانی که آزمون (۹) رد شود می‌توانیم

با استناد به مطالب بخش سوم، آزمون (۱۰) آزمونی ضعیف‌تر نسبت به آزمون (۹) است. به یاد آوردن آماره آزمون (۹) اولین آماره ترتیبی از بردار  $\bar{J} = \{J - \frac{1}{5}, J - \frac{1}{4}, J - \frac{1}{3}, J - \frac{1}{2}\}$  و آماره آزمون (۱۰)  $\bar{J}_{[k]}$  برای برخی





شکل ۶: مقادیر  $\hat{J}_{[k]}$

نتیجه بگیریم که آزمون (۱۰) نیز رد می‌شود. در شکل ۵ مقادیر  $\hat{J}_{[k]}$  را برای تمام سال‌ها با نقطه نشان داده شده‌اند. خط نقطه خط‌چین نشان‌دهنده آماره آزمون مربوط به آزمون‌های (۹) و (۱۰) یعنی  $\hat{J}_{[k]}$  و خط پیوسته مقدار چندک ۰.۹۹ برای این آزمون‌ها است. همان‌طور که می‌بینیم تمام مقادیر، بالاتر از مقدار چندک ۰.۹۹ برای این آزمون هستند و وجود روند کاهشی ضعیف را برای این داده‌ها نشان می‌دهد.

یکی از مهم‌ترین اجزاء FTS است که در این مطالعه با استفاده از آزمون‌های ناپارامتری روشی را برای تشخیص آن در FTS بیان کردیم. پس از تشخیص روند مسئله برآوردیابی آن مطرح می‌شود که این مهم زمینه‌ای برای تحقیقات آتی خواهد بود. اما تشخیص روند به‌تنهایی کافی نیست. نیاز است دوره توابع نیز آزمون شود که این موضوع در مقاله [۸] به آن پرداخته شده است. علاوه بر آزمون فرضیه روند و دوره در FTS نیازمند روشی برای تجزیه یک FTS به جزءهای روند و دوره هستیم که هنوز در تحقیقات این حوزه به آن پرداخته نشده است.

## ۶ بحث و نتیجه‌گیری

در این مقاله یک روش ناپارامتری برای بررسی و تشخیص وجود روند در FTS با استفاده از تعاریف متفاوت از روند و توابع رکورد معرفی شد و کارایی این روش با بررسی داده‌های واقعی میزان باروری در استرالیا بر حسب سن مورد بررسی قرار گرفت. هدف اصلی از تحلیل FTS، پیش‌بینی و توصیف کمی مکانیسم‌های تصادفی است که منجر به تولید توابع شده‌اند. در این راستا نیاز است FTS به مؤلفه‌های روند، دوره‌های زمانی و خطا تجزیه شود. اما قبل از تجزیه نیاز به شناسایی و تشخیص این گونه مؤلفه‌ها داریم. تشخیص روند

## مراجع

[۱] نیرومند، ح. و بزرگ‌نیا، س. (۱۳۹۰)، سری‌های زمانی، ناشر: دانشگاه پیام نور.

- [2] Berkes, I., Gabrys, R., Horváth, L., Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society, Series B*, 71, 927–946.
- [3] Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In *Nonparametric Functional Estimation and Related Topics*, Roussas, G. editor. NATO ASI Series, 335, Springer, Netherlands, 509–530.
- [4] Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York.

- [5] Comiso, J. C. (2000). Variability and trends in Antarctic surface temperatures from in situ and satellite infrared measurements. *Journal of Climate*, **13**, 1674–1696.
- [6] Ferraty, F. Vieu, P. (2006). *Nonparametric Modeling for Functional Data*. Springer, New York.
- [7] Fraiman, R., Justel, A., Liu, R., Llop, P. (2014). Detecting trends in time series of functional data: A study of Antarctic climate change. *The Canadian Journal of Statistics*, **42**, 597–609.
- [8] Hörmann, S., Kokoszka, P., Nisol, G. (2018). Testing for periodicity in functional time series. *The Annals of Statistics*, **46**(6A), 2960–2984.
- [9] Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, John Wiley & Sons, New York.
- [10] Kargin, V. Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, **99**, 2508–2526.
- [11] Ramsay, J. O. (1982). When the data are functions. *Psychometrika*, **47**, 379–396.
- [12] Ramsay, J. O., Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*, Springer, New York.
- [13] Ramsay, J. O. and Silverman, B. (2002). *Functional Data Analysis*, Springer, New York.
- [14] Santer, B. D., Wigley, T. M. L., Boyle, J. S., Gaffen, D. J., Hnilo, J. J., Nychka, D., Parker, D. E. Taylor, K. E. (2000). Statistical significance of trends and trend differences in layer—Average temperature time series. *Journal of Geophysical Research*, **105**, 7337–7356.