

پیش‌بینی نرخ بیکاری با استفاده از روش‌های یادگیری ماشین

مجری

عباس مرادی

همکاران

عاطفه واثی

رضا هادی‌زاده

محمدرضا محمودی



پژوهشکده‌ی آمار

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی

بهار ۱۴۰۳

بسم الله الرحمن الرحيم

کد شناسه: RP-0302

عنوان: پیش‌بینی نرخ بیکاری با استفاده از روش‌های یادگیری ماشین

مجری: عباس مرادی

همکاران: عاطفه واثی، رضا هادی‌زاده، محمدرضا محمودی

گروه پژوهشی: پردازش داده‌ها و اطلاع‌رسانی

تاریخ انتشار: بهار ۱۴۰۳

نوبت انتشار: اول

صفحه‌آرا: ساناز مهندسی

طراح جلد: ساناز مهندسی

❖ حق مالکیت معنوی این طرح پژوهشی متعلق به پژوهشکده‌ی آمار است و نقل مطالب فقط با ذکر مأخذ مجاز است.



پژوهشکده‌ی آمار

تهران، خیابان دکتر فاطمی، خیابان باباطاهر، خیابان سرتیپ فکوری،

شماره‌ی ۱۴۵

۳- ۸۸۶۳۰۴۴۰ ۰۲۱

www.srtc.ac.ir

پیش‌گفتار

در این پژوهش با توجه به اهمیت نرخ بیکاری در اقتصاد کلان، با استفاده از مدل‌های کلاسیک و نوین نرخ بیکاری پیش‌بینی می‌شود. نرخ بیکاری تابعی از نرخ بیکاری سال‌های گذشته، تولید ناخالص داخلی، نقدینگی، نرخ مشارکت و نرخ تورم می‌باشد. در این پژوهش داده‌های فصلی نرخ بیکاری در دوره بهار ۱۳۸۴ تا تابستان ۱۴۰۱ مورد بررسی قرار گرفته است. با توجه به آنکه نرخ بیکاری به صورت ماهانه موجود نیست و از آنجا که ارتباط بین نرخ بیکاری و تورم مد نظر این پژوهش است، نرخ بیکاری در دوره مورد بررسی با استفاده از روش‌های linear spline, cubic و nearest-neighbor درونیابی و به ماهانه تبدیل شده‌اند که در نهایت با توجه به نتایج بهتر مدل linear داده‌های حاصل از این روش مورد استفاده قرار گرفته‌اند. در این پژوهش برای پیش‌بینی نرخ بیکاری از مدل‌های غیر خطی شبکه عصبی پس انتشار خطا، روش‌های سری‌زمانی فازی و سایر الگوریتم‌های مهم از جمله الگوریتم هیبرید استفاده شده است. همچنین برای ارزیابی مدل‌ها از معیار میانگین مربعات خطا، میانگین قدر مطلق انحراف و ریشه میانگین مربع خطا استفاده شده است. نتایج پژوهش نشان می‌دهد در صورت ترکیب کردن شبکه‌های عصبی و هیبریدی نتایج بهتری نسبت به سایر الگوریتم‌ها دارد. همچنین استفاده از روش‌های یادگیری جمعی مناسب‌تر هستند.

این پژوهش در گروه پردازش داده‌ها و اطلاع‌رسانی پژوهشکده‌ی آمار توسط آقای دکتر عباس مرادی به عنوان مجری طرح و آقایان رضا هادی زاده و دکتر محمدرضا محمودی به عنوان همکاران اصلی طرح پژوهشی انجام شده است که بدین وسیله از همه این عزیزان صمیمانه تشکر و قدردانی می‌شود.

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی
پژوهشکده‌ی آمار

فهرست مطالب

۱	کلیات پژوهش.....
۱-۱	مقدمه.....
۲-۱	بیان مساله.....
۳-۱	اهمیت و ضرورت پژوهش.....
۴-۱	اهداف تفصیلی پژوهش.....
۵-۱	تعاریف و مفاهیم کلی مورد نیاز.....
۱-۵-۱	سری‌زمانی.....
۲-۵-۱	سری‌های زمانی پیوسته و گسسته.....
۳-۵-۱	سری‌زمانی مانا.....
۴-۵-۱	انواع سری‌های زمانی.....
۶-۱	خصوصیات کلی سری‌های زمانی.....
۱-۶-۱	مؤلفه‌های سری‌زمانی.....
۷-۱	تکنیک‌های اساسی برآورد.....
۱-۷-۱	خصوصیات برآوردگرها.....
۲-۷-۱	روش گشتاورها.....
۳-۷-۱	روش درست‌نمایی ماکزیم.....
۴-۷-۱	برآورد پیوسته‌ی پارامترها.....
۱۳	مروری بر ادبیات پژوهش.....
۱-۲	مقدمه.....
۲-۲	طرح آمارگیری نیروی کار.....
۳-۲	مدل‌های پیش‌بینی.....
۱-۳-۲	روش مدل‌سازی باکس و جنکینز (اتورگرسیو میانگین متحرک جمعی (ARIMA)).....
۲-۳-۲	روش مدل‌سازی فضا-حالت هموارساز نمایی.....
۳-۳-۲	روش ترکیبی مدل‌سازی تبدیل موجک گسسته- اتورگرسیو میانگین متحرک جمعی.....
۴-۳-۲	روش k-نزدیک‌ترین همسایگی (k-Nearest Neighbors).....
۴-۲	تدقیق پیش‌بینی نرخ بیکاری با استفاده از سایر سری‌های زمانی.....

پیش‌بینی نرخ بیکاری به روش‌های کلاسیک.....	۳۳
۱-۳- مقدمه.....	۳۳
۲-۳- آمار توصیفی.....	۳۳
۳-۳- بررسی مانایی متغیرها.....	۳۶
۳-۴- پیش‌بینی نرخ بیکاری و نرخ تورم به روش کلاسیک.....	۳۸
۱-۴-۳- پیش‌بینی.....	۴۴

پیش‌بینی نرخ بیکاری به روش‌های نوین.....	۴۷
۱-۴- مقدمه.....	۴۷
۲-۴- ارتباط بین نرخ بیکاری و نرخ تورم.....	۴۷
۳-۴- پیش‌بینی نرخ بیکاری با استفاده از سری‌های زمانی فازی.....	۵۳
۴-۴- پیش‌بینی نرخ بیکاری و نرخ تورم با استفاده از جنگل تصادفی.....	۵۶
۵-۴- پیش‌بینی نرخ بیکاری و نرخ تورم با استفاده از حافظه طولانی کوتاه مدت.....	۵۸
۶-۴- پیش‌بینی نرخ بیکاری و نرخ تورم با استفاده از شبکه‌های عصبی بازگشتی مبتنی بر دروازه.....	۵۹
۷-۴- پیش‌بینی نرخ بیکاری و نرخ تورم با استفاده از گرادیان تقویتی شدید.....	۶۰
۸-۴- پیش‌بینی نرخ بیکاری و نرخ تورم با استفاده از یادگیری جمعی.....	۶۰

نتیجه‌گیری.....	۶۳
۱-۵- مقدمه.....	۶۳
۲-۵- تحلیل نتایج طرح.....	۶۳
۳-۵- پیشنهادات.....	۶۴

مرجع‌ها.....	۶۵
--------------	----

فهرست جدول‌ها

جدول ۳-۱- مقدار آماره‌های توصیفی نرخ بیکاری، نرخ تورم، ... و شاخص فلاکت (فروردین ۱۳۸۴ تا شهریور ۱۴۰۱).....	۳۴
جدول ۳-۲- آزمون نرمالیتی.....	۳۴
جدول ۳-۳- نتیجه آزمون دیکی-فولر تعمیم یافته.....	۳۶
جدول ۳-۴- نتیجه آزمون فیلیپس-پرون.....	۳۷
جدول ۳-۵- نتیجه آزمون KPSS.....	۳۷
جدول ۳-۵- نتایج مدل‌سازی با مدل‌های نوع آرما.....	۴۳
جدول ۳-۶- نتایج مدل‌سازی با ETS.....	۴۳
جدول ۳-۷- نتایج مدل‌سازی سری تقریب و جزییات با استفاده از در روش مدل‌سازی WaveletArima.....	۴۳
جدول ۳-۸- مقادیر پیش‌بینی شده سری‌های زمانی نرخ بیکاری، نرخ تورم و شاخص فلاکت.....	۴۴

فهرست شکل ها

- شکل ۲-۱- تعداد دفعات ماهانه مقالات خبری، وبلاگ ها و توییت هایی که به زبان کره ای نوشته شده اند. ۲۴.....
- شکل ۲-۲- جملات مثالی برای برچسب گذاری POS و تحلیل احساسات. ۲۵.....
- شکل ۲-۳- نتایج تجزیه و تحلیل CCF کلمات کلیدی SKI و GI آنها با UI مقایسه می شود. ۲۸.....
- شکل ۲-۴- مقایسه نرخ بیکاری واقعی (UI) و داده های مناسب پیش بینی شده توسط مدل ها. ۳۲.....
- شکل ۳-۱- نمودار نرمالیتی سری زمانی نرخ بیکاری. ۳۴.....
- شکل ۳-۲- نمودار نرمالیتی سری زمانی نرخ تورم. ۳۵.....
- شکل ۳-۳- نمودار نرمالیتی سری زمانی نسبت نرخ تورم به نرخ بیکاری. ۳۵.....
- شکل ۳-۴- نمودار نرمالیتی سری زمانی شاخص فلاکت. ۳۶.....
- شکل ۳-۵- نمودار سری زمانی بالا، نمودار ACF چپ و نمودار PACF راست سری زمانی نرخ بیکاری. ۳۹.....
- شکل ۳-۶- نمودار سری زمانی بالا، نمودار ACF چپ و نمودار PACF راست سری زمانی نرخ تورم. ۴۰.....
- شکل ۳-۷- نمودار سری زمانی بالا، نمودار ACF چپ و نمودار PACF راست سری زمانی نسبت تورم به بیکاری. ۴۱.....
- شکل ۳-۸- نمودار سری زمانی بالا، نمودار ACF چپ و نمودار PACF راست سری زمانی شاخص فلاکت. ۴۲.....
- شکل ۳-۹- مقایسه مقادیر پیش بینی سری زمانی نرخ بیکاری با روشهای کلاسیک (ARIMA, ETS, WaveletArima, KNN). ۴۵.....
- شکل ۳-۱۰- مقایسه مقادیر پیش بینی سری زمانی نرخ تورم با روشهای کلاسیک (ARIMA, ETS, WaveletArima, KNN). ۴۵.....
- شکل ۳-۱۱- مقایسه مقادیر پیش بینی سری زمانی نرخ با روشهای کلاسیک (ARIMA, ETS, WaveletArima, KNN). ۴۵.....
- شکل ۳-۱۲- مقایسه مقادیر پیش بینی سری زمانی شاخص فلاکت با روشهای کلاسیک (ARIMA, ETS, WaveletArima, KNN). ۴۶.....
- شکل ۴-۱- ضریب همبستگی نرخ بیکاری و نرخ ترم. ۴۸.....
- شکل ۴-۲- ضریب همبستگی نرخ بیکاری و نرخ ترم (نرمال سازی شده). ۴۸.....
- شکل ۴-۳- تعیین نقش رهبر و پیرو نرخ های بیکاری و تورم. ۴۹.....
- شکل ۴-۴- نقشه حرارتی گسسته، تعیین نقش رهبر و پیرو نرخ های بیکاری و تورم. ۵۰.....
- شکل ۴-۵- نقشه حرارتی پیوسته، تعیین نقش رهبر و پیرو نرخ های بیکاری و تورم. ۵۰.....
- شکل ۴-۶- تاب خوردگی نرخ های بیکاری و تورم. ۵۱.....
- شکل ۴-۷- یافتن ارتباط بین نرخ های بیکاری و تورم- برحسب تجزیه مولفه ها. ۵۲.....

- شکل ۴-۸- بررسی علیت گرنجری نرخ تورم و نرخ بیکاری (اثرگذاری سری‌ها نسبت به همدیگر)..... ۵۲
- شکل ۴-۹- سری‌های زمانی نرخ تورم- نرخ بیکاری- نرخ فلاکت..... ۵۳
- شکل ۴-۱۰- تقسیم بندی جهت ایجاد درجه عضویت مثالی..... ۵۴
- شکل ۴-۱۱- تابع عضویت مثالی..... ۵۴
- شکل ۴-۱۲- تقسیم بندی سری‌زمانی نرخ بیکاری..... ۵۵
- شکل ۴-۱۳- قواعد پیوند نرخ بیکاری با استفاده از قاعده چن..... ۵۵
- شکل ۴-۱۴- پیش‌بینی نرخ بیکاری بر اساس سری زمانی فازی..... ۵۵
- شکل ۴-۱۵- قواعد پیوند نرخ تورم با استفاده از قاعده چن..... ۵۶
- شکل ۴-۱۶- پیش‌بینی نرخ تورم بر اساس سری زمانی فازی..... ۵۶
- شکل ۴-۱۷- پیش‌بینی نرخ بیکاری و نرخ تورم با استفاده از جنگل تصادفی..... ۵۸
- شکل ۴-۱۸- پیش‌بینی نرخ بیکاری و نرخ تورم با استفاده از حافظه طولانی کوتاه مدت..... ۵۹
- شکل ۴-۱۹- پیش‌بینی نرخ بیکاری و نرخ تورم با استفاده از شبکه‌های عصبی بازگشتی..... ۵۹
- شکل ۴-۲۰- پیش‌بینی نرخ بیکاری و نرخ تورم با استفاده از گرادیان تقویتی شدید..... ۶۰
- شکل ۴-۲۱- میزان RMSE نرخ بیکاری و نرخ تورم با استفاده از یادگیری جمعی..... ۶۱
- شکل ۴-۲۲- پیش‌بینی بهینه نرخ بیکاری و نرخ تورم با استفاده از یادگیری جمعی (پشته‌سازی)..... ۶۲

کلیات پژوهش

۱-۱- مقدمه

بدون شک داشتن یک تصویر و ذهنیت درست از وضعیت و شاخص‌های اقتصاد کلان هر کشور می‌تواند سیاست‌گذاران را در تصمیم‌گیری‌ها و تدقیق سیاست‌ها یاری نماید. آمار و اطلاعات دقیق و پیش‌بینی‌های متناسب با آن، پایه و اساس برنامه‌ریزی‌های کلان هر کشور است. در این پژوهش با توجه به اهمیت نرخ بیکاری در اقتصاد کلان با استفاده از مدل‌های کلاسیک و نوین نرخ بیکاری پیش‌بینی می‌شود. نرخ بیکاری تابعی از نرخ بیکاری سال‌های گذشته، تولید ناخالص داخلی، نقدینگی، نرخ مشارکت و نرخ تورم می‌باشد. در این پژوهش داده‌های فصلی نرخ بیکاری از بهار ۱۳۸۴ تا تابستان ۱۴۰۱ مورد بررسی قرار گرفته است. از آنجا که طبق تعاریف بین‌المللی، سن کار از سال ۱۳۹۸ از ۱۰ سال به ۱۵ سال تغییر یافت. در دوره مورد بررسی برای یکسان شدن سن کار به صورت ۱۵ سال به بالا در محاسبات لحاظ گردید. همچنین با توجه به اینکه ارتباط بین نرخ بیکاری و نرخ تورم مد نظر این پژوهش است و این دو نرخ هم فرکانس نسیتند در واقع نرخ بیکاری فصلی است ولی تورم ماهانه است، نرخ بیکاری در دوره مورد بررسی با استفاده از روش‌های linear, cubic, spline و nearest-neighbor درونیایی و به ماهانه تبدیل شدند که در نهایت با توجه به نتایج بهتر مدل linear داده‌های حاصل از این روش مورد استفاده قرار گرفت. در این پژوهش برای پیش‌بینی نرخ بیکاری از مدل‌های غیر خطی شبکه عصبی پس انتشار خطا، روش‌های سری‌زمانی فازی و سایر الگوریتم‌های مهم از جمله الگوریتم هیبرید استفاده شده است. همچنین برای ارزیابی مدل‌ها از معیار میانگین مربعات خطا، میانگین قدر مطلق انحراف و ریشه میانگین مربع خطا استفاده شده است. نتایج پژوهش نشان می‌دهد در صورت ترکیب کردن شبکه‌های عصبی و هیبریدی نتایج بهتری نسبت به استفاده از سایر الگوریتم‌ها دارد، روش یادگیری جمعی نیز نتایج پیش‌بینی را بهتر کرد. در این طرح به مطالعه انواع روش‌های پیش‌بینی نرخ بیکاری و تاثیرات سایر

سری‌های زمانی مانند نرخ تورم بر آن پرداخته می‌شود. در این فصل ضمن بیان مساله و ضرورت اجرای طرح به معرفی کلیات دیگر پژوهش از قبیل تعاریف و مفاهیم کلی پرداخته می‌شود.

۱-۲- بیان مساله

یکی از مهمترین مسائلی که پایه و اساس برنامه‌ریزی‌ها می‌باشد، داشتن تصویر روشنی از آینده است. یکی از مهمترین طرح‌های آماری که نقش بالقوه‌ای در برنامه‌ریزی‌های اقتصادی کشورها دارد پیش‌بینی نرخ بیکاری در کشور است. این آمارها اغلب براساس تعریف‌ها و استانداردهای بین‌المللی و با محاسبه‌ی شاخص‌های شناخته شده به‌دست می‌آیند. رایج‌ترین منبع اطلاعاتی که برای برآورد شاخص‌های نیروی کار مورد استفاده قرار می‌گیرد اطلاعات حاصل از آمارگیری‌های نمونه‌ای است، که به دلیل اهمیت موضوع، در اغلب کشورها به صورت یک آمارگیری مستمر در دوره‌های زمانی متوالی اجرا می‌شود. طراحی آمارگیری‌های نیروی کار معمولاً به گونه‌ای صورت می‌گیرد که برآوردهای حاصل از اجرای آن در سطح کل کشور و استان‌های مختلف بهینه باشند و خطاهای نسبی برآوردها از مقدار تعیین شده‌ای بیشتر نباشند. آمارگیری نیروی کار ایران که از سال ۱۳۸۴ با طراحی جدید به صورت فصلی اجرا می‌شود برآوردهای قابل اعتمادی در سطح کل کشور و استان‌ها ارائه می‌کند، برای برنامه‌ریزی‌های بهتر و اجرای سیاست‌های مناسب‌تر توسط برنامه‌ریزان و سیاست‌گذاران، پیش‌بینی نرخ بیکاری بسیار حائز اهمیت است. از آنجایی که آمارگیری نیروی کار ایران برای محاسبه‌ی برآوردهای مورد نظر در سطح کل کشور و استان‌ها بهینه شده است، بکارگیری روش‌های پیش‌بینی برای نرخ بیکاری ضروری است. هدف این مطالعه آن است که با استفاده از سری‌زمانی نرخ بیکاری را از بهار ۱۳۸۴ تا تابستان ۱۴۰۱ مورد بررسی قرار داده و پیش‌بینی نرخ بیکاری پاییز ۱۴۰۱ را انجام دهد.

۱-۳- اهمیت و ضرورت پژوهش

نیروی انسانی یکی از مهمترین عوامل پیشرفت استان‌ها، شهرستان‌ها و مناطق کشور است و کشورها در تلاشند که از این مزیت خود به هر نحو که امکان‌پذیر است استفاده کنند. کشورها نیز با اتکا به توانایی‌های نیروی کار بالقوه و موجود در تلاشند وضعیت اقتصادی خود را بهبود دهند. برای اطلاع از وضعیت نیروی انسانی نیاز به گردآوری اطلاعات نیروی انسانی به روش‌های مختلف است. در واقع یک مساله متداول در همه جوامع بشری، پایش و کنترل شاخص‌های مهم اقتصادی-اجتماعی است. اشتغال و بیکاری، از جمله موضوعات اساسی اقتصاد هر کشور است، به گونه‌ای که افزایش و کاهش بیکاری، به عنوان یکی از شاخص‌های توسعه یافتگی جوامع تلقی می‌شود. نرخ بیکاری یکی از شاخص‌هایی است که برای ارزیابی شرایط اقتصادی کشور مورد استفاده قرار می‌گیرد. پایش و اعمال سیاست در خصوص کنترل یا ایجاد تغییر دلخواه در نرخ بیکاری از دغدغه‌های دولتمردان محسوب می‌شود. طرح آمارگیری نیروی کار با هدف برآورد شاخص‌های نیروی کار از جمله نرخ بیکاری به صورت فصلی و سالانه و همچنین تغییرات آن در کل کشور و استان‌ها با روش نمونه‌گیری چرخشی^۱ اجرا می‌شود.

^۱ Rotation Sampling

سیاست‌گذاران همواره شاخص‌ها و نماگرهای اقتصادی کلان کشور را بررسی می‌کنند و براساس آن تصمیم‌گیری می‌کنند پیش‌بینی این شاخص‌ها می‌تواند منجر به تدقیق سیاست‌ها و تصمیم‌های کلان کشور شود. داشتن اطلاعات دقیق و تصور درست از نرخ تورم و نرخ بیکاری به صورت توأم برای سیاست‌گذاران بسیار حائز اهمیت است. براساس مطالعات انجام شده تاکنون در کشور، پژوهشی در زمینه پیش‌بینی نرخ بیکاری با استفاده از سایر شاخص‌ها و با استفاده از یادگیری ماشین انجام نشده است. هدف اصلی این پژوهش انجام این مهم به عنوان اولین مطالعه علمی در این زمینه است.

۱-۴- اهداف تفصیلی پژوهش

مهمترین هدف این پژوهش این است به پیش‌بینی نرخ بیکاری برای فصول بعدی در دوره مورد مطالعه بپردازد. در این پژوهش علاوه به این هدف اصلی به اهداف فرعی زیر نیز پرداخته می‌شود. در واقع به بررسی اثرات متقابل نرخ تورم و نرخ بیکاری نسبت به هم پرداخته می‌شود. داده‌های سری‌زمانی بین مقدار نرخ تورم و نرخ بیکاری از فروردین ۱۳۸۴ لغایت تابستان ۱۴۰۱ گردآوری شده است. هدف از این بررسی این است که مشخص شود کدام سناریوی زیر رخ خواهد داد:

- آیا نرخ تورم بر نرخ بیکاری اثرگذار است؟
- آیا نرخ بیکاری بر نرخ تورم اثرگذار است؟
- آیا هر دو بر هم اثرگذار هستند؟
- آیا هیچ ارتباطی با هم ندارند؟

۱-۵- تعاریف و مفاهیم کلی مورد نیاز

در این بخش برخی از مفاهیم کلی که در این طرح مطرح می‌شوند را معرفی می‌کنیم. مفاهیم تخصصی و فنی نیز در بخش‌های بعدی معرفی خواهند شد.

۱-۵-۱- سری‌زمانی

داده‌هایی که از مشاهده یک فرایند به طور پی در پی بدست می‌آیند را به صورت $x_1, x_2, x_3, \dots$ در نظر بگیرید، اندیس $1, 2, 3, \dots$ می‌تواند نشان دهنده‌ی بُعد زمان، فاصله و غیره باشد. چنین دنباله‌ای را سری می‌نامند [۳]. اگر اندیس معرف زمان باشد، سری را سری‌زمانی می‌نامند. پس بطور کلی می‌توان گفت یک سری‌زمانی مجموعه مشاهداتی است که بر حسب زمان مرتب شده باشد. سری‌های زمانی یا داده‌هایی که از مشاهدات یک فرایند در طول زمان بدست می‌آیند بسیار متداول‌اند. به طور روزمره می‌توان نمونه‌هایی از سری‌های زمانی را در علوم مختلف مانند اقتصاد، آمار، جمعیت‌شناسی، هواشناسی و... مشاهده کرد. چون مقادیر سری‌زمانی در زمان‌های متفاوت اندازه‌گیری می‌شود این امکان وجود دارد که همبستگی زمانی بین مقادیر سری‌زمانی مشاهده شده، وجود داشته باشد. با استفاده از علم آمار، یادگیری ماشین و سایر قوانین احتمالاتی حاکم بر سری‌زمانی، می‌توان همبستگی زمانی را مورد تجزیه و تحلیل قرار داد. هدف از تجزیه و تحلیل سری‌های زمانی معمولاً درک

یا مدل‌سازی مکانیزم‌های تصادفی آن است. پس از آن می‌توان داده‌های مربوط به اجزای سری‌زمانی را تولید کرد و با استفاده از آن یافته‌ها، رفتار سری مزبور را بر اساس مقادیر گذشته سری پیش‌بینی کرد.

۱-۵-۲- سری‌های زمانی پیوسته و گسسته

هرگاه مشاهدات یک فرایند به طور پیوسته نسبت به زمان صورت گیرد به سری حاصل شده از چنین فرایندی سری‌زمانی پیوسته می‌گویند. در صورتی که اگر مشاهدات یک فرایند فقط در زمان‌های معینی معمولاً به فواصل مساوی از یکدیگر صورت بگیرند، سری حاصل شده را یک سری‌زمانی گسسته می‌نامند، حتی اگر متغیر اندازه‌گیری شده، یک متغیر پیوسته باشد. معمولاً سری‌های زمانی گسسته را با نماد x_t و سری‌های زمانی پیوسته را با $x(t)$ نمایش می‌دهند. معمولاً در آمار رسمی، با سری‌های زمانی گسسته که مشاهدات آن‌ها در فواصل زمانی یکسانی (ماهانه، فصلی، سالانه و ...) محاسبه شده است سروکار داریم.

۱-۵-۳- سری‌زمانی مانا

به طور کلی هدف از تجزیه و تحلیل سری‌های زمانی، دریافت درک و بینش لازم جهت مدل‌سازی کردن روند تصادفی مشاهدات سری است. پس از دریافت این دانش، داده‌های جدید بر مبنای این مدل تولید می‌شود و پیش‌بینی مقادیر آینده‌ی آن سری بر پایه‌ی مقادیر گذشته‌ی آن صورت می‌گیرد. از آنجا که داده‌های سری‌زمانی مقادیر یک فرایند را در زمان‌های مختلف نشان می‌دهند، بنابراین ممکن است همبستگی زمانی بین این مقادیر وجود داشته باشد که باید با استفاده از قوانین احتمالی حاکم بر سری‌های زمانی، این همبستگی را کشف و مورد تجزیه و تحلیل قرار داد. بر این اساس یکی از مهمترین ویژگی‌هایی که باید مورد بررسی قرار گیرد، ایستایی یا مانا بودن فرایند است. بدین مفهوم که قوانین احتمالی حاکم بر فرایند نسبت به زمان تغییر نکند و یا به عبارت دیگر فرایند در تعادل آماری باشد.

۱-۵-۳-۱- انواع فرایندهای مانا

انواع فرایندهای مانا عبارتند از:

الف) مانای مرتبه‌ی اول

ب) مانای مرتبه‌ی دوم (یا مانای ضعیف)

ج) فرایند مانای قوی

د) فرایند مانای اکید

که به تفصیل به معرفی آن‌ها پرداخته می‌شود:

الف) مانای مرتبه‌ی اول: اگر امید ریاضی متغیر سری‌زمانی نسبت به زمان تغییر نکند یعنی:

$$(۱) \quad E[X_1] = E[X_2] = \dots = E[X_t] = \mu_t$$

آن فرایند را مانا از مرتبه‌ی اول می‌نامند.

(ب) مانای مرتبه‌ی دوم یا مانای ضعیف: هرگاه فرایندی علاوه بر میانگین نسبت به کواریانس نیز مانا باشد، یعنی:

$$(۲) \quad cov(X_t, X_{t-k}) = cov(k)$$

آن را مانا از مرتبه‌ی دوم می‌نامند. رابطه‌ی فوق نشان می‌دهد که مقدار کواریانس صرف نظر از موقعیت t به مقدار تأخیر زمانی (k) بستگی دارد. البته وقتی که فرایندی نسبت به کواریانس مانا باشد، خود به خود نسبت به واریانس نیز مانا خواهد بود.

(ج) فرایند مانای قوی: اگر علاوه بر میانگین و کواریانس، سایر خصوصیات آماری نیز نسبت به زمان ایستا باشد، چنین فرایندی را مانای قوی می‌نامند.

(د) فرایند مانای اکید: فرایند تصادفی $\{X_t\}$ را مانای اکید یا مطلق می‌گوئیم؛ اگر برای هر تأخیر k در هر نقطه‌ی $t_1, t_2, \dots, t_n$ توزیع $X(t_1), X(t_2), \dots, X(t_n)$ مانند توزیع $X(t_1 - k), X(t_2 - k), \dots, X(t_n - k)$ باشد [۹].

۱-۵-۴-انواع سری‌های زمانی

انواع سری‌هایی که مورد مطالعه قرار می‌گیرند را می‌توان به سری‌های یک متغیره در زمان و یا مکان (فاصله) و سری‌های چند متغیره در زمان و یا مکان (فاصله) تقسیم کرد. حال اگر مقادیر یک فرایند را نسبت به زمان مورد بررسی قرار دهیم سری حاصل شده را سری‌زمانی می‌نامند. در مطالعات بر اساس تعداد متغیرها، سری‌های زمانی را می‌توان به دو گروه زیر دسته‌بندی نمود:

۱-۴-۵-۱-سری‌های زمانی یک متغیره: مانند سری‌زمانی نرخ بیکاری

۱-۴-۵-۲-سری‌های زمانی چند متغیره: که خود به دو گروه ذیل می‌باشند:

۱-۴-۵-۱-۲-سری‌های زمانی چند مکانی: در این‌گونه سری‌ها یک پارامتر خاص را در چند مکان مختلف مانند چند استان مختلف اندازه‌گیری می‌کنیم.

۱-۴-۵-۲-۲-سری‌های زمانی چندتایی: این سری‌ها دارای چند پارامتر مختلف می‌باشند که در یک مکان خاص، اندازه‌گیری شده‌اند.

سری‌های زمانی تک متغیره و یا چند متغیره بیشتر بر اساس بازه‌ی زمانی مورد ارزیابی قرار می‌گیرند، چرا که خصوصیات کلی و روش‌های مدل‌سازی و برآورد پارامترهای مدل بستگی به بازه‌ی زمانی سری‌ها دارد، پس به طور کلی سری‌های زمانی را بر اساس بازه‌ی زمانی به گروه‌های زیر تقسیم می‌کنیم:

سری‌های زمانی پیوسته، مقادیر این سری‌ها به طور پیوسته ثبت می‌شوند (بازه‌ی زمانی این گونه سری‌ها صفر می‌باشد).

سری‌های زمانی که بازه‌ی زمانی بخشی از روز می‌باشد. (یک ساعت، دو ساعت، شش ساعت و دوازده ساعت)

سری‌های زمانی که بازه‌ی زمانی، بخشی از سال می‌باشد. (روز، هفته، ماه و فصل)

سری‌های زمانی سالانه و یا سری‌های زمانی با بازه‌ی زمانی بیشتر از سال (سال، دهه و قرن)

۱-۶- خصوصیات کلی سری‌های زمانی

سری‌های زمانی یک متغیره را به طور کلی با برآورد کردن خصوصیات آماری آنها مانند، میانگین، انحراف معیار، چولگی، توزیع احتمالاتی و ساختار وابستگی زمانی، مورد تجزیه و تحلیل قرار می‌دهند، در حالی که برای سری‌های زمانی چند متغیره لازم است که علاوه بر خصوصیات آماری فوق، ارتباط (وابستگی) بین سری‌های زمانی نیز برآورد شود.

۱-۶-۱- مؤلفه‌های سری زمانی

تغییرات هر سری زمانی می‌تواند به علت تغییرات بعضی از عوامل باشد که تعدادی از آنها طبیعی و بعضی ناشی از عوامل اقتصادی و اجتماعی هستند. این تغییرات را می‌توان به صورت اجزا یا مؤلفه‌های یک سری زمانی مشاهده کرد. هر سری زمانی به چهار مؤلفه به شرح زیر قابل تجزیه است:

روند^۲ (T): تغییرات دراز مدت در میانگین سری زمانی را روند می‌نامند. این تغییرات ناشی از عوامل طولانی مدت بر رفتار سری زمانی است. در واقع روند، تمایل سری زمانی به سیر صعودی داشتن، سیر نزولی داشتن یا تقریباً ثابت بودن است.

فصلی‌بودن^۳ (S): تغییرات درون یک سال که به طور ثابتی تکرار و یا سال به سال به شیوه‌ای یکسان جلو می‌رود را نشان می‌دهد.

تقویمی^۴ (C): تغییرات ناشی از ساختار تقویم مانند تعطیلات و تعداد روزهای کاری در ماه را نشان می‌دهد و نوعی عامل فصلی می‌باشد.

نامنظم^۵ (I): تغییرات تصادفی و غیر قابل پیش‌بینی را نشان می‌دهد که به طریقی نامنظم عمل می‌کند. پس از شناخت هر یک از مؤلفه‌های سری‌های زمانی، مسئله اساسی برآورد هر یک از این مؤلفه‌ها است. بدین منظور، فرض‌هایی در مورد هر یک از این مؤلفه‌ها که با جامعه مورد مطالعه مطابقت دارد در نظر گرفته می‌شود. این فرض‌ها را می‌توان در قالب مدل‌های سری زمانی بیان کرد. معمولاً مدل‌های سری زمانی بسته به اینکه مؤلفه-هوی به صورت جمعی یا ضربی بیان شوند، در یکی از این دو طبقه قرار می‌گیرند.

مدل جمعی: در مدل جمعی یا جمع پذیر هر مقدار سری زمانی در زمان مشخص حاصل جمع هر یک از مؤلفه-های سری زمانی است:

$$X_t = T + S + C + I$$

در مدل جمعی فرض بر این است که مؤلفه‌ها از یکدیگر مستقل هستند.

مدل ضربی: در مدل ضربی یا ضرب‌پذیر هر مقدار سری زمانی در زمان مشخص حاصل ضرب هر یک از مؤلفه-های سری زمانی است:

$$X_t = T * S * C * I$$

^۲ Trend

^۳ Seasonality

^۴ Cyclicity

^۵ Irregularity

در مدل ضربی فرض بر این اینست که مولفه با یکدیگر مرتبط هستند. با توجه به ماهیت متغیرهای اقتصادی، اقتصاددانان از مدل ضربی برای تحلیل سری‌های زمانی مشاهدات اقتصادی و بازرگانی استفاده می‌کنند. تجزیه و تحلیل مؤلفه‌های سری‌های زمانی دارای سابقه‌ای طولانی است که در قرنهای ۱۷ تا ۱۹ میلادی و در زمینه نجوم، هواشناسی و اقتصاد به کار گرفته شد و اولین بار به تحلیل فصلی توسط بایز-بالوت (۱۸۴۷) پرداخته شد [۴]. کار اولیه روی حذف ارتباط دوسویه بین دو متغیر متمرکز شده بود. اولزک (۱۸۸۴) و هوکر (۱۹۰۱) تلاش کردند مولفه روند و فصل را از طریق تفاضل‌گیری از میانگین قیمت‌ها طی چند سال حذف کنند [۸ و ۱۰].

۱-۷- تکنیک‌های اساسی برآورد

روش‌های منتج شده از ریاضیات آمار برای برآورد پارامترهای مدل، بر اساس سری‌زمانی نمونه را تکنیک‌های برآورد می‌نامند. سری‌زمانی نمونه $x_1, \dots, x_n$ را با مدلی با پارامترهای α و β بیان می‌کنیم و پارامترهای مدل را از روی سری‌زمانی نمونه برآورد می‌کنیم، یعنی $\hat{\alpha}$ و $\hat{\beta}$ مقادیر برآورد شده‌ی پارامترهای α و β مدل می‌باشند.

تکنیک‌های برآورد را به طور کلی می‌توان به سه روش عمده‌ی زیر تقسیم کرد [۶ و ۷]:

۱- روش گشتاورها

۲- روش حداقل مربعات

۳- روش درست‌نمایی ماکزیمم

معمولاً برای برآورد پارامترهای یک مدل، روشی را بایستی انتخاب کرد که کمترین مقدار خطای برآورد را داشته باشد. در ادامه‌ی بحث، معیاری را معرفی می‌کنیم که با استفاده از آن بتوان بین روش‌های برآورد، روشی را انتخاب کرد که کمترین خطای برآورد را داشته باشد. تعدادی از روش‌های برآورد فوق نسبت به روش‌های دیگر مزیت‌هایی دارند. در ادامه کارایی برآوردگرها تشریح می‌گردد.

۱-۷-۱- خصوصیات برآوردگرها

زمانی که بحث در مورد دقت برآوردگرها می‌شود، دو کمیت اغلب مورد بررسی قرار می‌گیرد [۶]:
اریبی

میانگین مربع انحراف

هرگاه امید ریاضی برآوردگر $\hat{\alpha}$ ، برابر با مقدار پارامتر جامعه (α) باشد پس به $\hat{\alpha}$ یک برآوردگر نااریب α می‌گویند. در غیر این صورت $\hat{\alpha}$ برآوردگر α است با اریبی برابر $E[\hat{\alpha}] - \alpha$.

فرض کنید مدلی با پارامترهای α و β داریم ما می‌خواهیم m سری مصنوعی به طول N تولید کنیم، برای هر سری برآوردگر $\hat{\alpha}_i, i = 1, \dots, m$ را محاسبه می‌کنیم اگر میانگین برآوردگر $\hat{\alpha}$ از نظر آماری برابر پارامتر جامعه باشد پس $\hat{\alpha}$ یک برآوردگر نااریب α است. اختلاف ($\alpha - \hat{\alpha}$) بین پارامتر α و برآوردگر $\hat{\alpha}$ را خطای برآورد می‌نامند و امید ریاضی مربع این خطا را، میانگین مربع خطا (MSE) برآوردگر $\hat{\alpha}$ می‌نامند.

$$E[(\alpha - \hat{\alpha})^2] = Var(\hat{\alpha}) + [\alpha - E(\hat{\alpha})]^2 \quad (۳)$$

در رابطه‌ی (۳-۲)، MSE برابر با واریانس برآوردگر به اضافه‌ی مجذور اریبی است و اگر $\hat{\alpha}$ یک برآورد نااریب از α باشد یعنی؛ $\alpha = E[\hat{\alpha}]$ ، پس در این حالت MSE برابر با واریانس برآوردگر خواهد بود. در واقع هدف این است که یک برآوردگر، دارای حداقل MSE باشد. حال اگر $\hat{\alpha}$ نااریب و دارای مینیمم واریانس باشد می‌توان گفت $\hat{\alpha}$ یک برآوردگر کارآ خواهد بود. از این‌رو موقع انتخاب کردن روش برآورد، بایستی دو خصوصیت اریبی و MSE را مورد بررسی قرار داد و روش برآوردی را انتخاب کرد که دارای حداقل اریبی و MSE باشد.

۱-۷-۲-روش گشتاورها

امید ریاضی متغیر تصادفی X ، $E(X)$ را گشتاور اول جامعه‌ی X می‌نامند و امید ریاضی x_r را r امین گشتاور جامعه‌ی X نامیده می‌شود. برای محاسبه‌ی r امین گشتاور نمونه بر اساس سری‌زمانی نمونه‌ی $x_1, \dots, x_n$ از رابطه (۴-۲) استفاده می‌شود.

$$m_r = \left(\frac{1}{N}\right) \sum_{i=1}^N x_i^r \quad (۴)$$

اگر سری‌زمانی تعریف شده در بالا به وسیله‌ی مدلی با پارامترهای $\alpha_1, \alpha_2, \dots, \alpha_p$ معرفی شود پس در نتیجه گشتاورهای جامعه تابع پارامترهای مدل خواهند بود. بنابراین از طریق مساوی قرار دادن گشتاورهای جامعه با گشتاورهای نمونه و سپس حل معادلات حاصله، پارامترهای مدل را می‌توان برآورد کرد. اگر P تعداد پارامتر-هایی باشد که می‌خواهیم برآورد کنیم از مساوی قرار دادن P گشتاور جامعه و نمونه و سپس از حل همزمان آن-ها مقدار پارامترها را بدست می‌آوریم.

به عنوان مثال، مدل سری‌زمانی $x_t = a + bz_t$ که a و b پارامترهای مدل هستند و z_t یک متغیر تصادفی مستقل با میانگین صفر و واریانس یک می‌باشد را در نظر بگیرید. حال هدف ما در این بررسی برآورد پارامترهای a و b و بر مبنای سری‌زمانی نمونه‌ی $x_1, \dots, x_n$ است. از آنجایی که مدل معرفی شده دارای دو پارامتر است پس، از تساوی قرار دادن گشتاور اول و دوم جامعه با نمونه و حل آن‌ها، مقادیر a و b را می‌توان به دست آورد. ابتدا گشتاور اول و دوم جامعه را بدست می‌آوریم.

$$E[x] = a \quad E[x^2] = a^2 + b^2$$

و سپس گشتاور اول و دوم نمونه را بر اساس سری‌زمانی نمونه به دست می‌آوریم.

$$m_1 = \frac{1}{N} \sum_{i=1}^N x_i \quad m_2 = \frac{1}{N} \sum_{i=1}^N x_i^2$$

از تساوی قرار دادن معادلات بالا خواهیم داشت:

$$\hat{\alpha} = \frac{1}{N} \sum_{i=1}^N x_i \quad a^2 + b^2 = \frac{1}{N} \sum_{i=1}^N x_i^2$$