به نام یزدان پاک

دانشکده‌ی علوم

پایان‌نامه کارشناسی آمار ریاضی

برآوردگرهای نیرومند در سری‌های زمانی خودبازگشتی همبسته دوره‌ای با نقاط پرت جمعی

به کوشش:

امیرحسین شبانی

استاد راهنما:

دکتر زهره شیشه‌بر

بهمن ۱۳۹۰

بسم الله الرحمن الرحيم

به نام خدا

اظهارنامه

اینجانب امیرحسین شبانی (۸۸۰۵۷۰) دانشجوی رشته آمـار گـرایش آمـار محـض دانشکده‌ی علوم اظهار می‌کنم که این پایان نامه حاصل پـژوهش خـودم بـوده و در جاهایی که از منابع دیگران استفاده کرده‌ام نشانی دقیـق و مشخصـات کامـل آن را نوشته‌ام.

همچنین اظهار می‌نمایم که تحقیق و موضوع پایان‌نامـه‌ام تکـراری نیسـت و تعهـد می‌نمایم که بدون مجوز دانشگاه دستاوردهای آن را منتشر ننمـوده و یـا در اختیـار غیر قرار ندهم. کلیه حقوق این اثر مطابق بـا آیـین‌نامـه مالکیـت فکـری و معنـوی متعلق به دانشگاه شیراز است.

نام و نام خانوادگی: امیرحسین شبانی

تاریخ و امضاء: ۹۰/۱۱/۰۳

به نام خداوند بزرگ

برآوردگرهای نیرومند در سریهای زمانی خودبازگشتی همبسته دوره ای با نقاط پرت جمعی

به وسیله‌ی

امیرحسین شبانی

پایان نامه

ارائه شده به تحصیلات تکمیلی دانشگاه بعنوان بخشی از فعالیت‌های تحصیلی برای اخذ درجه کارشناسی ارشد

در رشته‌ی

آمار ریاضی

از دانشگاه شیراز

شیراز

ایران

ارزیابی و تصویب توسط کمیته پایان نامه با درجه : عالی

دکتر زهره شیشه بر ، دانشیار بخش آمار (رئیس کمیته).....

دکتر احمد رضا سلطانی ، استاد بخش آمار (استاد مشاور).....

دکتر عبدالرسول برهانی ، استادیار بخش آمار(استاد مشاور).....

تقدیم به:

سربازان ایران زمین

که تا پای جان در راه شکوه و پایداری

سرزمین جاویدان پارس

کوشیدند

سپاس‌گزاری

اکنون که این پایان نامه به پایان رسیده بر این هیچ فرض است خداوندی را سپاس گذارد که او را وجود ، سلامتی و شادکامی بخشید و در سرزمین پارس دیده بر جهان گشود تا بر نیای پارسی خود ببالد. پس از یزدان پاک فرض است ، زحمات خانواده‌اش به ویژه پدر و مادر را ارج نهد که آموختند نخستین آموخته‌های زندگی را. همچنین بر این هیچ فرض است زحمات اساتیدی را ارج نهد که وی را آموختند چگونه آموختن را. سپاس باد گرامی استادی را که به کار کشید این کوچک شاگردش را . و دو صد سپاس زحمات اساتید مشاوری را که در طول دوره تحصیل از هر کمکی دریغ نکردند شاگردشان را .

و هزاران سپاس کسانی را که با جان و مال و آبروی خویش در راه آبادی و آزادی این سرزمین کوشیدن تا مردم این سرزمین توانا باشند بر امور خویش.

چکیده

برآوردگرهای نیرومند در سری‌های زمانی خودبازگشتی همبستهء دوره‌ای با نقاط پرت جمعی

به کوشش

امیرحسین شبانی

همانطور که می‌دانیم حضور داده‌های پرت در یک مجموعه از مشاهدات هموارد مورد بحث بوده است، در سری‌های زمانی نیز حضور این گونه داده‌ها غیر قابل اجتناب است و از آنجا که حذف داده‌های پرت از یک مجموعه داده‌ی سری زمانی کاری نسبتا نامعقول است، پس لازم است به برآوردگرهایی دست یابیم که اثرپذیری کمتی از داده‌های پرت نسبت به برآوردگرهای کلاسیک داشته باشند. از آنجا که الگوی خودبازگشتی همبسته دوره‌ای دارای کاربردهای زیادی است، در این پایان‌نامه تحلیل این سری‌ها در حضور داده‌های پرت جمعی مورد بحث قرار می-گیرد. فصل اول شامل مقدمات و مباحث مربوط به برآوردگرهای نیرومند و خصوصیات آنها می‌باشد. فصل دوم شامل تعاریف و خصوصیات برآوردگر نیرومند کواریانس دو متغیر و تابع اتوکواریانس یک سری زمانی است. فصل سوم به خصوصیات یک سری *PCAR* پرداخته می‌شود. فصل چهارم شامل اثرات ناشی از نقاط پرت روی برآورد پارامترهای مدل فوق و برآوردگر نیرومند این پارامترها تحت معادلات یول-والکر است و در فصل پنجم به شبیه‌سازی مدل *PAR(1)* پرداخته و مدل‌بندی میانگین دمای فصلی مینیمم مرکز سینوپتیک شیراز، مدل‌بندی تعرق و تبخیر مرجع را برای همین مرکز و مدل بندی دبی فصلی رودخانه *Fraser* را ارائه می‌دهیم.

فهرست مطالب

عنوان **صفحه**

فصل اول: مقدمه..... ۲

فصل دوم: برآوردگرهای نیرومند مکان و مقیاس..... ۴

۲-۱- تعیین نقاط پرت یا غربالگری ۵

۲-۲- الگوهای ایجاد کننده نقاط پرت ۵

۲-۳- برآوردگر نیرومند..... ۷

۲-۴- معیارهای تعیین برآوردگرهای نیرومند..... ۷

۲-۴-۱- نقطه شکست..... ۷

۲-۴-۲- تابع تأثیر ۹

۲-۴-۳- حساسیت رشد خطا..... ۱۰

۲-۴-۴- منحنی اریبی..... ۱۰

۲-۴-۵- سازگاری و کارایی ۱۱

۲-۵- معرفی آماره‌های خاص..... ۱۱

۲-۵-۱ آماره‌های R,U,L..... ۱۱

۲-۵-۲- آماره‌های L تعمیم یافته..... ۱۲

۲-۵-۳- برآوردگرهای M ۱۴

۲-۶- برآوردگر نیرومند مکان ۱۶

۲-۶-۱- میانه نمونه‌ای..... ۱۶

۲-۶-۲- میانگین بریده شده α% ۱۷

۲-۶-۳- برآوردگر مکان هاجز- لهمن..... ۱۸

۲-۶-۴- برآوردگر مکان هاجز- لهمن تعمیم یافته..... ۱۸

۲-۷- برآوردگر نیرومند مقیاس..... ۲۰

۲-۷-۱- دامنه میان چارکی..... ۲۰

۲-۷-۲- میانه قدر مطلق انحرافات از میانه..... ۲۰

۲-۷-۳- برآوردگر مقیاس S_n ۲۲
۲-۷-۴- برآوردگر مقیاس Q_n ۳۱

فصل سوم: برآوردگر نیرومند کواریانس و تابع اتوکواریانس و تابع خودهمبستگی

۳-۱- برآوردگر نیرومند کواریانس ۳۷
۳-۱-۱- موارد استفاده از برآوردگرهای نیورمند ۳۸
۳-۱-۲- برآوردگر نیرومند کواریانس بر اساس برآوردگر نیرومند مقیاس ۳۹
۳-۲- برآوردگر نیرومند تابع اتوکواریانس ۴۱
۳-۲-۱- نقطه شکست برآوردگر تابع اتوکواریانس ۴۲
۳-۲-۲- تابع تأثیر برآوردگر تابع اتوکواریانس و واریانس تقریبی ۴۷
۳-۳- استفاده از برآوردگر M برای محاسبه برآوردگر مقیاس ۵۱
۳-۴- برآورد پارامترهای مدل ARMA با کمک برآوردگر M مقیاس ۵۲

فصل چهارم: معرفی مدل PCAR و برآورد پارامترها در داده‌هایی با نقاط پرت

۴-۱- تعریف سری زمانی همبستهء متناوب ۵۴
۴-۱-۱- خواص سری‌های همبستهء متناوب ۵۴
۴-۲- روش برآورد نیرومند پارامترهای PCAR در حضور نقاط پرت ۵۷

فصل پنجم: برآوردگر نیرومند پارامترهای مدل PCAR در داده‌های با نقاط پرت

۵-۱- مدل PCAR با نقاط پرت جمعی ۶۳
۵-۱-۱- تأثیر نقاط پرت روی برآورد پارامترهای مدل ۶۳
۵-۱-۲- تاثیر نقاط پرت روی روش‌های انتخاب مدل ۶۶
۵-۲- برآورد PEACV بر اساس یک برآوردگر نیرومند مقیاس ۶۷
۵-۳- برآورد نیرومند پارامترهای مدل PAR بوسیله معادلات یول- والکر ۷۴

فصل ششم: شبیه‌سازی و کاربردها

۶-۱- شبیه‌سازی مدل ARMA ۷۷
۶-۲- شبیه‌سازی مدل $PAR_{۴}(۱)$ با نقاط پرت جمع ۷۸
۶-۳- شبیه‌سازی مدل (۱) AR با نقاط پرت جمعی ۸۳
۶-۴- مدل‌بندی دمای مینیمم مرکز سینوپتیک شیراز ۸۵
۶-۵- تبخیر و تعرق مرجع ۸۸

۶-۶- مقدار دبی رودخانه Fraser ۹۱
پیوست‌ها ۹۳
ضرایب میکس ۹۶
برنامه‌های کامپیوتری ۱۰۲
منابع ۱۱۰

فهرست جدول ها

عنوان و شماره صفحه

جدول ۲-۱- مشخصات برآوردگرهای نیرومند مکان تحت توزیع گوسی ۱۹
جدول ۲-۲- مشخصات برآوردگرهای نیرومند مقیاس تحت توزیع گوسی ۳۴
جدول ۶-۱- میانگین برآورد پارامترهای مدل (۱) PAR_4 با و بدون نقاط پرت ۸۲
جدول ۶-۲- برآورد کلاسیک و نیرومند تابع اتوکواریانس برای (۱) AR ۸۴
جدول ۶-۳- میانگین دمای مینیمم فصلی مرکز سینوپتیک شیراز ۸۶
جدول ۶-۴- برآورد اتوکواریانس دوره‌ای کلاسیک ۸۷
جدول ۶-۵- برآورد اتوکواریانس دوره‌ای نیرومند ۸۷
جدول ۶-۶- میانگین فصلی ET_0 ۸۹
جدول ۶-۷- برآورد کلاسیک تابع اتوکواریانس ET_0 ۹۰
جدول ۶-۸- برآورد نیرومند تابع اتوکواریانس ET_0 ۹۰
جدول ۶-۹- برآورد کلاسیک و نیرومند پارامترهای رودخانه Fraser ۹۲

فهرست شکل ها

عنوان و شماره صفحه

شکل ۲-۱- نمودار تابع تاثیر برآوردگرهای مقیاس ۳۵
شکل ۳-۱- تعداد مجموع و تفاضل‌های آشفته برای یک داده پرت ۴۲
شکل ۳-۲- تعداد مجموع و تفاضل‌های آشفته برای یک داده پرت ۴۳
شکل ۳-۳- تعداد مجموع و تفاضل‌های آشفته برای چند داده پرت ۴۳
شکل ۳-۴- نمودار نقطه شکست زمانی و کران‌های آن ۴۶
شکل ۳-۵- نمودار تابع تاثیر کواریانس بین دو متغیر تحت توزیع گوسی ۵۰
شکل ۶-۱- نمودار سری (۱) PAR_4 بدون نقاط پرت جمعی ۷۹
شکل ۶-۲- نمودار سری (۱) PAR_4 با نقاط پرت جمعی ۸۰
شکل ۶-۳- نمودار سری (۱) $PCAR_4$ با نقاط پرت جمعی ۸۰
شکل ۶-۴- نمودار نرمال بودن برآوردگر مقیاس نیرومند ۸۱
شکل ۶-۵- نمودار سری (۱) AR با داده‌های پرت جمعی ۸۳
شکل ۶-۶- نمودار (۱) AR بدون نقاط پرت ۸۳
شکل ۶-۷- نمودار نرمال بدون برآوردگر نیرومند مقیاس Q_n در (۱) AR ۸۵
شکل ۶-۸- نمودار داده‌های فصلی رودخانه Fraser ۹۱
شکل ۶-۹- نمودار تابع اتوکواریانس برای دوره‌های رودخانه ۹۱

فصل اول

۱- مقدمه

همواره در استنباط آماری با این مشکل روبرو هستیم که مجموعه داده‌هایی که با کمک آنها به برآورد و یا آزمون می‌پردازیم شامل نقاط پرت هستند، سری‌های زمانی نیز بعنوان زیرمجموعه‌ای از آمار از این حالت خارج نیست، به عبارت دیگر در یک مجموعه از داده‌های سری زمانی نیز نقاط پرت موجودند، پس در این پایان‌نامه قصد داریم به ارائه برآوردگرهایی بپردازیم که تاثیرپذیری کمتری از نقاط پرت داشته باشند.

در بسیاری از تحلیل‌های سری زمانی با مواردی روبرو می‌شویم که داده‌ها دارای مؤلفه فصلی هستند و معمولاً بعد از حذف روند، با کمک روش‌های حذف روند فصلی به مدلی ایستا دست می‌یابیم و با کمک آن به برازش و برآورد پارامترهای مدل می‌پردازیم. مشکل این روش این است تنها زمانی که تابع خود همبستگی بین تمام فصل‌های مختلف یکسان باشد قابل استفاده است و در بسیاری از موارد کاربردی حالاتی موجودند که این فرض برقرار نیست، در واقع یک تناوب ذاتی در سری موجود است که با روش‌های حذف روند و مؤلفه فصلی نمی‌توان آنها را حذف کرد. این تحلیل در مدل‌هایی صورت می گیرد که تابع اتوکواریانس و میانگین توابعی از دوره تناوب سری خواهد بود این کلاس از مدل‌های سری زمانی را همبسته‌ء متناوب (PC) نامیدند. اگرچه در این مدل‌ها سری نا مانا است ولی بسیاری از خواص فرایندهای مانا را می‌توان به آنها تعمیم داد.

در زمینه استفاده از برآوردگرهای نیرومند در مدل‌های رگرسیونی تحقیقات بسیاری انجام گرفته است و تألیفات زیادی انجام شده که کامل‌ترین آنها توسط *R.A.Marrona(1998)* انجام گرفته است، و *R.J. Serfling(1984)* خانواده آماره‌هایی را معرفی کرده که تقریباً تمام برآوردگرهای نیرومند را پوشش می‌دهد، همچنین برآوردگرهای نیرومند مقیاس بسیار کارا

توسط *P.J.Rausseeuw(1993)* ارائه گردید، ولی استفاده از این برآوردگرها در سری‌های زمانی اولین بار توسط *Genton ، Yanyuan(2000)* ارائه شد ، اگرچه این روش را برای کار با معادلات یول-والکر برای سری‌های خودبازگشتی ارائه دادند ولی اولین گام در ارائه روش‌های نیرومند در سری زمانی بود و از برآوردگرهای مقیاس نیرومندی که توسط *P.J.Rausseeuw (1993)* ارائه شده بود استفاده کرد. در ادامه این مطالعات *Q.Shao(2005)* برآوردگرهای نیرومند را برای سری‌های خودبازگشتی چند متغیره بکار گرفت و نشان داد که در حضور نقاط پرت این برآوردگرها کاراتر از برآوردگرهای کلاسیک هستند. در ادامه این مطالعات *A.J.Q. Sarnaglia (2010)* روش‌های نیرومند را در سری‌های خودبازگشتی همبستهء دوره‌ای با نقاط پرت جمعی بکار برد و نشان داد که این روش‌ها در سری‌های زمانی همبستهء دوره‌ای نیز مانند سری‌های مانا بسیار کاراتر از برآوردگرهای کلاسیک هستند.

در این پایان نامه علاوه بر مطالعه سری‌های زمانی خودبازگشتی همبستهء دوره‌ای به مطالعه برآوردگرهای نیرومند مکان و مقیاس پرداخته و معیارهایی برای شدت نیرومندی این برآوردگرها ارائه می‌دهیم و همچنین خانواده‌های بسیار گسترده‌ای از آماره‌ها و برآوردگرها را ارائه خواهیم داد که اکثر برآوردگرهای نیرومند جزء این خانواده‌ها هستند. علاوه بر این مدل-های *AR(1)* و *PCAR(1)* را با داده‌های پرت جمعی شبیه‌سازی کرده و نشان می‌دهیم برآوردگر نیرومند پارامترهای این مدل‌ها از برآوردگر کلاسیک کاراتر بوده و همچنین برازش بهتری به داده‌ها خواهند داشت، در ادامه چند کاربرد روش‌های نیرومند در سری‌های زمانی همبستهء دوره‌ای که شامل مدل‌بندی میانگین دمای مینیمم فصلی، مدل‌بندی تعرق و تبخیر مرجع و مدل‌بندی دبی آب رودخانه است را ارائه می‌دهیم.

فصل دوم

۲- برآوردگرهای نیرومند مکان و مقیاس

۲-۱-تعیین نقاط پرت یا غربالگری

از لحاظ کاربردی نقطه پرت (*Outlier*)، داده‌ای است که از دیگر داده‌ها فاصله داشته باشد. معیار عددی تعیین این فاصله تحت توزیع نرمال با کمک یک تابع بصورت $|\cdot - \hat{\mu}|/\hat{\sigma}$ است، در صورتی که این مقدار برای هر داده از نقطه برش(*Cut off point* ؛ مقداری ثابت که مقادیر بزرگتر از آن را نقطه پرت در نظر بگیریم) از پیش تعیین شده‌ای بین ۲/۵ تا ۳ تجاوز کند، داده مربوطه را پرت تلقی می‌کنیم. در این رابطه ابتدا لازم است که برآوردگر مکان و مقیاس داده‌ها را تعیین کنیم و از آنجا که این برآوردها نیز تحت تاثیر نقاط پرت هستند لازم است برای استنباط از برآوردگرهایی استفاده شود که نسبت به حضور درصدی از داده‌های پرت مقدار قابل قبول خود را حفظ کند. در صورتی که برای غربالگری داده‌های پرت از برآوردگرهای ذکر شده استفاده کنیم، طرح غربالگری مطلوبی خواهیم داشت.[34]

۲-۲- الگوهای ایجاد کننده نقاط پرت

نقاط پرت تولید شده در یک مجموعه داده، توسط یکی از الگوهای زیر ایجاد می‌شوند:

I نقاط پرت جمعی (*Additive Outliers*)

II نقاط پرت جابجایی (*replacement outliers*)

III نقاط پرت وابسته به عامل خطا (*innovation outliers*)

نقاط پرت جمعی در داده‌ها بصورتی هستند که با احتمال ε داده‌های مورد علاقه بعلاوه مقادیرمخدوش کننده و با احتمال $1-\varepsilon$ داده‌های مورد علاقه به تنهایی مشاهده می‌شوند. این فرآیند به تمام زیردنباله‌های داده‌ها اثر نمی‌گذارد، یعنی همواره زیر دنباله‌ای از داده‌ها موجود است که شامل نقاط پرت نباشد. در صورتی که V_t داده‌های مورد علاقه و B_t یک فرایند برنولی با احتمال موفقیت ε باشد و W_t فرایند مخدوش کننده مستقل از V_t و B_t باشد، نقاطی که مشاهده می‌شود بصورت زیر است.

$$\{X_t\ , t\epsilon\ \mathbb{Z}\}\ ,\ X_t = V_t + B_t W_t \qquad (۱.۲)$$

در الگوی نقاط پرت جابجایی، داده‌ها با احتمال ε بصورت مقادیر مخدوش کننده و با احتمال $1-\varepsilon$ بصورت داده‌های مورد علاقه مشاهده می‌شوند. این الگو نیز بر تمام زیردنباله‌های داده‌ها اثر نمی‌کند. در صورتی که V_t داده‌های مورد علاقه و B_t یک فرایند برنولی با احتمال موفقیت ε باشد و W_t فرایند مخدوش کننده مستقل از V_t و B_t باشد، فرآیندی که مشاهده می‌شود بصورت زیر است.

$$\{X_t\ , t\epsilon\ \mathbb{Z}\}\ ,\ X_t = (1 - B_t)V_t + B_t W_t \qquad (۲.۲)$$

در اکثر بررسی‌هایی که روی نقاط پرت شده است بیشتر به این دو الگو توجه شده‌است، بنا به نظر *Genton ، Yanyuan (2000)* این دو الگو بیشتر باعث ایجاد استنباط آماری نادرست می‌شوند. نقاط پرت وابسته به عامل خطا در حالتی ایجاد می‌شود که در یک مدل خطی یا سری زمانی توزیع خطا یک توزیع با دم‌های سنگین است، بعنوان مثال در یک مدل سری زمانی فرض کنید F توزیعی است که برای عامل خطا در نظر گرفته شده است و H توزیعی با پراکندگی بیشتر از F باشد درصورتی که خطای مشاهده شده از $G=(1-\varepsilon)\ F+\ \varepsilon\ H$ آمده باشد با الگوی نقاط پرت وابسته به عامل خطا مواجه هستیم. این الگو برخلاف دو مورد قبلی بر تمام داده‌ها اثر می‌گذارد. [23]

۲-۳- برآوردگر نیرومند

برآوردگر نیرومند به برآوردگری اطلاق می‌شود که در برآورد پارامتر نسبت به حضور داده‌های پرت تا حدی از خود مقاومت نشان می‌دهد، به این مفهوم که در برابر درصدی از داده‌ها که از دیگر داده‌ها فاصله داشته باشند، برآوردگر مقدار قابل قبول خود را حفظ می‌کند. استفاده از این برآوردگرها از آنجا ناشی می‌شود که نمی‌توان بطور مطلق از پرت بودن برخی داده‌ها اطلاع کسب کرد، یعنی علاوه بر اینکه پرت بودن داده‌ها وابسته به برآوردگر مکان و مقیاس بدست آمده توسط داده‌ها است، همچنین پرت بودن یک داده وابسته به نمونه‌ای است که داده‌ها در آن قرار دارند، به این مفهوم که یک داده ممکن است نسبت به برآورد مکان و مقیاسی که با یک نمونه‌ایجاد شده پرت بحساب نیاید ولی با برآوردهای فوق در نمونه‌ای دیگر پرت بحساب آید، همچنین پرت بودن داده‌ها به نقطه برش نیز وابسته است یعنی با تغییر نقطه برش پرت بودن یا نبودن داده‌ها تغییر می‌کند.

۲-۴- معیارهای تعیین برآوردگرهای نیرومند

در این بخش معیارهایی را معرفی می‌کنیم که با کمک آنها می‌توانیم در مورد نیرومندی یا عدم نیرومندی یک برآوردگر تصمیم بگیریم و همچنین شدت مقاومت برآوردگرها در برابر حضور نقاط پرت را با هم مقایسه کنیم به این مفهوم که، با کمک این معیارها می‌توان گفت که بین دو برآوردگر نیرومند کدامیک نیرومندتر است.

۲-۴-۱- نقطه شکست (***Breakdown point***)

اولین موردی که لازم است در مورد یک برآوردگر نیرومند بررسی شود این است که برآوردگر در حضور داده‌هایی پرت مقدار قابل قبول خود را حفظ کند، این بررسی به معیاری منجر می‌شود که آن را نقطه شکست نامیده و بصورت زیر تعریف می‌کنیم.[23]