بسم الله الرحمن الرحیم

دانشگاه پیام نور

دانشکده علوم پایه

واحد تهران شرق

پایان نامه

برای دریافت مدرک کارشناسی ارشد

رشته آمار ریاضی

گروه آمار

عنوان پایان نامه:

برآوردگر استوار فرآیندهای ARFIMA تحت نقاط دورافتاده جمع پذیر

حمیده درستکار

استاد راهنما:

دکتر مسعود یارمحمدی

استاد مشاور:

دکتر علی شادرخ

شهریور ۱۳۹۱

تقدیم به پدر و مادر عزیزم

که بعد از یگانه خالق هستی، همواره پشتیبان من در تمام عرصه های زندگی به ویژه

تحصیل علم بوده اند.

سپاس گزاری...

سپاس خداوندی را که سخنوران از ستودن او عاجزند، و حسابگران از شمارش نعمت های او ناتوان، و تلاشگران از ادای حق او درمانده اند.

باتقدیر و سپاس از استاد گرانقدر، جناب آقای دکتر مسعود دیارمحمدی که همواره با راهنمائی های ارزشمند خویش مرا در تدوین این پایان نامه یاری نمودند.

از استاد گرامی جناب آقای دکتر علی شادرخ که زحمت مشاوره ی این پایان نامه را بر عهده گرفتند کمال تشکر را دارم.

همچنین از استاد محترم جناب آقای دکتر عادل محمدپور که زحمت مطالعه و داوری این اثر را تقبل نمودند تشکر می نمایم.

درپایان بر خود واجب می دانم از زحمات خانواده و دوستانم در طول دوران تحصیل تشکر و قدردانی نمایم.

حمیده درستکار

شهریور ۱۳۹۱

چکیده

در این پایان نامه برآوردگر نیمه‌پارامتری از خانواده برآوردگرهای لگاریتم دوره نگار رگرسیون (GPH) برای برآورد پارامتر تفاضل‌گیری کسری در مدل اتورگرسیو میانگین متحرک جمعی کسری (ARFIMA) که در برابر نقاط دورافتاده جمع‌پذیر استوار می‌باشد، معرفی می‌کنیم. شیوه یافتن این برآوردگر مبتنی بر برآوردگر استواری از تابع اتوکواریانس ام آ وای و جنتون (۲۰۰۰) می‌باشد که برای بدست آوردن یک دوره نگار استوار شده مورد استفاده قرار گرفته است. برآوردگر پارامتر حافظه بلند مدت استوار یک نوع دیگری از برآوردگر معروف پیشنهاد شده بوسیله گویک و پورتر–هداک (۱۹۸۳) به نام GPH می‌باشد. نتایج شبیه‌سازی مونت کارلو نشان می‌دهند که برآوردگر پیشنهادی برای پارامتر تفاضل‌گیری موقعی که داده‌ها حاوی نقاط دورافتاده جمع‌پذیر باشند استوار می‌باشد.

واژه‌های کلیدی: نقاط دورافتاده جمع‌پذیر، مدل ARFIMA، حافظه بلندمدت، استواری

فهرست مطالب

فهرست جداول و شکل ها

مقدمه

اغلب اوقات سریهای زمانی اقتصادی و مالی با مشاهدات نابهنجار و غیرمعمول (نقاط دورافتاده[1]) آلوده می‌شوند. اثرات مربوط به این داده‌ها بر برآورد پارامترهای مدل‌های میانگین متحرک جمع بسته اتورگرسیو (ARIMA) توسط محققین زیادی بررسی شده است. به عنوان مثال لدالتر[2] (۱۹۸۹) نشان داد که پیش بینی‌های فاصله‌ای برخلاف پیش بینی‌های نقطه‌ای به طور قابل ملاحظه‌ای نسبت به نقاط دورافتاده جمع‌پذیر حساس‌ترند. شانگ و همکارانش[3] (۱۹۸۸) و چن و لیو[4] (۱۹۹۳) نشان دادند که برآورد پارامترها در مدل‌های ARMA، زمانی که حاوی نقاط دورافتاده هستند دارای اریبی بیشتری است. دوچ وهمکارانش[5] (۱۹۹۰) و شان[6] (۱۹۹۵-۱۹۹۲) اریبی استخراج شده از خودهمبستگی‌های نمونه را معلول نقاط دورافتاده دانستند. نقاط دورافتاده در سریهای زمانی چند متغیری، توجه تعداد زیادی از محققین را بخود جلب کرده است، به عنوان مثال تی سی و همکارانش[7] (۲۰۰۰) اثر نقاط دورافتاده را در بحث چند متغیری روی مدل‌های حاشیه‌ای و توام بررسی نموده‌اند.

در اوایل سال ۱۹۸۰، گرنجر و جویکس[8] (۱۹۸۰) و هاسکینگ[9] (۱۹۸۱) تعمیمی از فرآیند ARIMA را که در آن پارامتر تفاضل‌گیری می‌تواند مقادیر ناصحیح فرض شود به عنوان فرآیند ARFIMA پیشنهاد کردند. هاسکینگ (۱۹۸۱) نشان داد که سریهای زمانی با نمایش ARFIMA که در آن $d\in(۰,۰/۵)$ مانا و با حافظه بلندمدت می باشند. قسمت اخیر بوسیله خودهمبستگی‌های معنی‌دار آماری در تاخیرهای طولانی یا بعبارت دیگر بوسیله منحصربفرد بودن یک چگالی طیفی با فرکانس صفر بیان شده است.

---

1 - Outliers
2 - Ledolter
3 - Chang et al
4 - Chen & Liu
5 - Deutsch et al
6 - Chan
7 - Tsay et al
8 - Granger & Joyeux
9 - Hosking

در این پایان نامه، استراتژی برآورد پارامتر، نیمه‌پارامتری برای رتبه تفاضل‌گیری d در مدل‌های ARFIMA پیشنهاد شده است. روش‌های پارامتری، بطور نمونه حداکثر درستنمایی، برای برآورد توام پارامتر تفاضل‌گیری و پارامترهای شاخص در نمایش MA و AR ، نیز استفاده می‌شوند.(براون (۱۹۹۵)).

در روش نیمه‌پارامتری، برآورد پارامتر در دو مرحله محاسبه می‌شود:

در مرحله نخست برآورد رتبه تفاضل‌گیری d با استفاده از لگاریتم دوره نگار رگرسیون [۱] محاسبه شده و سپس پارامترهای MA و AR بعد از تفاضل‌گیری سری برآورد می‌شوند. معروفترین برآوردگر نیمه پارامتری بوسیله گویک و پورتر- هداک [۲] (۱۹۸۳) پیشنهاد شد، انواع دیگری از این برآوردگر توسط ریزن [۳] (۱۹۹۴) و رابینسون [۴] (۱۹۹۵) و اشخاص دیگری ارائه گردید.

از جمله کاربردهای عملی مدل ARFIMA می توان به موارد زیر اشاره نمود:

- در اقتصاد و زمینه مالی و سرمایه‌گذاری توسط ریزن و همکارانش (۲۰۰۳); باهاردویج و سانسون [5] (۲۰۰۶)
- کاربرد آن در هواشناسی توسط بایلی و شانگ [6] (۲۰۰۲)

هالدراپ و نیلسین (۲۰۰۷) تاثیر کمیت‌های خطا [۷]، نقاط دورافتاده و گسستگی‌های ساختاری [۸] دیگر را روی برآورد پارامتر با حافظه بلند مدت ارزیابی کردند. نتایج نشان می‌دهند که چنین تاثیراتی می‌توانند کاملا اساسی و نیازمند بررسی باشند. یک نقطه دورافتاده جمع‌پذیر در داده‌ها اساسا می‌تواند یک اریبی در برآورد پارامتر تفاضل‌گیری ایجاد کند.

[1] - a log-periodogram regression
[2] - Gewek & Porter-Hudak
[3] - Reisen
[4] – Robinson
[5] - Bhardwaj & Swanson
[6] - Baillie & Chang
[7] – Measurement errors
[8] – Structural breaks

آن‌ها نتیجه گرفتند که برآوردگرهای نیمه‌پارامتری بر مبنای رگرسیون، وقتی که پهنای باند[1] -که برابر است با تعداد فرکانس استفاده شده در برآورد- کوچک است، اریبی کمتری دارند. محققین شیوه‌های پیشنهاد شده توسط فیلیپس و سان[2] (۲۰۰۳) را که افزودن یک عامل غیرخطی به لگاریتم دوره نگار رگرسیون به عنوان راهی برای مینیمم کردن هر اریبی موجود، توصیه می‌کنند. بیاسگلی و آگوستینلی[3] (۲۰۰۳) استفاده از شیوه حداکثر درستنمایی وزن دار[4] را به عنوان یک تعدیل از برآوردگر ارائه شده به وسیله براون[5] (۱۹۹۴) پیشنهاد کردند.

در این پایان نامه، یک برآوردگر استوار برای پارامتر حافظه بلندمدت در فرآیند ARFIMA که در مقابل نقاط دورافتاده جمع‌پذیر استوار می‌باشد، ارائه می‌شود. شیوه یافتن این برآوردگر مبتنی بر برآوردگر استواری از تابع اتوکواریانس (ACOVF) ام آ وای و جنتون[6] (۲۰۰۰) می باشد که برای بدست آوردن یک دوره نگار[7] استوار شده مورد استفاده قرار گرفته است. برآوردگر پارامتر حافظه بلند مدت استوار یک نوع دیگری از برآوردگر معروف پیشنهاد شده بوسیله گویک و پورتر- هداک (۱۹۸۳) به نام GPH می‌باشد. استواری این برآوردگر از طریق شبیه‌سازی مونت کارلو[8] در فصل ۴ نشان داده می‌شود.

این پایان نامه شامل فصل‌های زیر می‌باشد:

- در فصل۱، برخی مفاهیم سریهای زمانی مورد استفاده در فصل‌های ۲، ۳ و ۴ ارائه می‌شود.
- در فصل ۲، سریهای زمانی با حافظه بلندمدت، فرآیندهای ARFIMA، تفاضل‌گیری کسری، روش برآورد پارامتر تفاضل‌گیری شرح داده شده است.

[1] - Bandwidth
[2] - Phillips & Sun
[3] - Bisaglia & Agostinelli
[4] - Weighted maximum likelihood
[5] - Beran
[6] - Ma & Genton
[7] - Periodogram
[8] - Monte Carlo Simulation

- در فصل ۳، نقاط دورافتاده در سریهای زمانی، تاثیر نقاط دورافتاده جمع‌پذیر بر برآورد پارامترهای مدل‌های مانا، برخی خواص تابع اتوکواریانس استوار، برآوردگرهای استوار برای توابع اتوکواریانس، خودهمبستگی و در آخر یک برآوردگر استوار برای رتبه تفاضل‌گیری کسری d در مدل ARFIMA(p,d,q) محاسبه شده است.
- در فصل ۴، روش‌های شبیه‌سازی مونت کارلو جهت بررسی این برآوردگر استوار ارائه و مورد بحث واقع شده است.

فصل اول

مروری بر برخی مفاهیم سریهای زمانی

مقدمه

برای مطالعه فرآیندهای با حافظه بلندمدت لازم است مروری بر سری زمانی داشته باشیم که قسمت‌هایی از نمایش طیفی را نیز شامل می‌شود. در بخش ۱-۱ با نگاهی کلی به تجزیه و تحلیل سری زمانی، به مباحث مانایی، انباشتگی، توابع خودهمبستگی و خودهمبستگی جزیی اشاره کرده و در بخش ۱-۲ به معرفی نمایش طیفی سری زمانی خواهیم پرداخت. در بخش ۱-۳ نیز مدل‌های مختلف در مدل‌بندی‌های سریهای زمانی را معرفی خواهیم کرد. در بخش ۱-۴ نیز برخی تعاریف کاربردی مورد استفاده در این پایان نامه مورد مطالعه قرار گرفته‌اند.

۱-۱ شرح و بررسی مفاهیم کلی

۱-۱-۱ سری زمانی

یک سری زمانی مجموعه‌ای از مشاهدات است که بر حسب زمان مرتب شده‌اند. به عبارت دیگر می‌توان گفت یک سری زمانی عبارت است از داده‌هایی که از مشاهده یک پدیده در طول زمان بدست آمده‌اند. در یک تقسیم بندی کلی سریهای زمانی را به پیوسته و گسسته تقسیم می‌کنند. یک سری زمانی را پیوسته گویند هر گاه مشاهدات بطور پیوسته در طول زمان ایجاد شده باشند. یک سری زمانی را گسسته گویند هرگاه مشاهدات فقط در زمان‌های معینی که معمولا به فواصل مساوی از یکدیگر قرار دارند اخذ شده باشند.

۱-۱-۲ فرآیند تصادفی

خانواده‌ای از متغیرهای تصادفی $\{Z_t, t \in T\}$ است که روی یک فضای احتمال تعریف شده باشد به طوری که T مجموعه اندیس‌هاست.

۱–۱–۳ مانایی [1]

مانایی بحث بسیار مهمی در مدل‌سازی سریهای زمانی می‌باشد زیرا بسیاری از مدل‌های احتمال سریهای زمانی بر مبنای مانایی سری استوار است. یک سری زمانی را مانای اکید گوئیم، هر گاه به ازای همه مقادیر $h, t_1, t_2, \ldots t_n$ توزیع توأم $Z_{t_1}, \ldots, Z_{t_n}$ مانند توزیع توأم $Z_{t_1+h}, \ldots, Z_{t_n+h}$ باشد یعنی:

$$F_{Z_{t_1}, \ldots, Z_{t_n}}(Z_1, \ldots, Z_n) = F_{Z_{t_1+h}, \ldots, Z_{t_n+h}}(Z_1, \ldots, Z_n) \quad (۱–۱)$$

همچنین یک سری زمانی را مانای ضعیف یا مانای مرتبه دوم نامند، هر گاه میانگین آن ثابت بوده و تابع اتوکواریانس آن فقط به تأخیر k بستگی داشته باشد. به عبارت دیگر میانگین وتابع اتوکواریانس آن به زمان بستگی نداشته باشد.

۱–۱–۴ اغتشاش خالص (نوفه سفید) [2]

به دنباله‌ای از متغیرهای تصادفی $\{\varepsilon_t\}$ که مستقل و هم‌توزیع [3] (iid) باشند اطلاق می‌شود. اهمیت این فرآیند به واسطه این است که فرآیندهای مفیدی را از آن می‌توان ساخت.

۱–۱–۵ تابع اتوکواریانس، خودهمبستگی و خودهمبستگی جزئی در تأخیر k

تابع اتوکواریانس، همبستگی بین مشاهداتی را که k واحد زمانی با یکدیگر اختلاف دارند اندازه می‌گیرد و به صورت زیر تعریف می‌شود:

$$\gamma_k = \mathrm{cov}(z_t, z_{t+k}) = E(z_t - \mu)(z_{t+k} - \mu) \quad (۲–۱)$$

تابع خودهمبستگی، عبارت است از همبستگی بین مشاهداتی که k واحد زمانی با یکدیگر فاصله دارند. تابع خودهمبستگی نظری که آن را با ρ_k نشان می‌دهند به شکل زیر تعریف می‌شود:

---

[1] - Stationary
[2] - White noise
[3] - independently and identically distribution

$$\rho_k = \frac{\text{cov}(z_t, z_{t+k})}{\text{var}(z_t)} = \frac{\gamma(k)}{\gamma(0)} \qquad (۳-۱)$$

ρ_k ضریب اتوکواریانس در تأخیر k است. معمولا تابع خودهمبستگی را با ACF[1] نشان می‌دهند.

همچنین تابع خودهمبستگی جزئی، همبستگی بین z_t و z_{t+k} بعد از حذف اثر متغیرهای $z_{t+1},z_{t+2},...,z_{t+k-1}$ را ضریب خودهمبستگی جزئی می‌نامند. این خودهمبستگی‌ها را با φ_{kk} نشان می‌دهند. این تابع که آن را با PACF[2] نیز نشان می‌دهند در تشخیص الگوی احتمالی مولد داده‌ها استفاده می‌شود.

$$\varphi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

۱-۱-۶ مفهوم انباشتگی[3]

اگر یک سری زمانی یک بار تفاضلی شود و سری تفاضلی شده مانا باشد، آنگاه سری زمانی اصلی انباشته از مرتبه اول نامیده و به صورت $I(1)$ نشان داده می‌شود. در صورتی که یک سری زمانی پس از دو بار تفاضل‌گیری حالت مانایی پیدا کند انباشته از مرتبه دوم می‌باشد و آن را با $I(2)$ نشان می‌دهند. بطور کلی اگر از یک سری زمانی d مرتبه تفاضل‌گیری شود، آن سری انباشته از مرتبه d یا $I(d)$ است. بدین ترتیب هرگاه یک سری زمانی انباشته از مرتبه یک یا بالاتر باشد، آن سری نامانا است. بنابراین فرآیند $I(0)$ نشان‌دهنده یک فرآیند مانا می‌باشد.

---

[1] - Autocorrelation Function
[2] - Partial Autocorrelation Function
[3] - Integration

۱-۲ نمایش طیفی سری زمانی

۱-۲-۱ دوره نگار

راه دیگر تحلیل یک سری زمانی بر این فرض استوار است که سری از موج‌های سینوسی و کسینوسی با فرکانس‌های متفاوت تشکیل شده است. ابزاری که این ایده را بکار می‌گیرد دوره نگار است که برای اولین بار توسط اسکاستر[۱] (۱۸۹۸) معرفی شده است. دوره نگار در اصل برای جدا کردن و برآورد نمودن دامنه مؤلفه سینوسی با فرکانس‌های معلوم بکار می‌رود.

۱-۲-۲ طیف و نمایش طیفی

طیف یک تبدیل فوریه از تابع اتوکواریانس است. تبدیل فوریه، نامیده شده به اسم ریاضیدان فرانسوی ژوزف فوریه، یک تبدیل انتگرالی است که هر تابع $f(t)$ را به یک تابع دیگر $F(\lambda)$ منعکس می‌کند. در این صورت، به $F(\lambda)$ تبدیل فوریه تابع $f(t)$ می‌گویند. حالت خاص تبدیل فوریه، سری فوریه نام دارد و آن زمانی کاربرد دارد که تابع $f(t)$ متناوب باشد، یعنی: $f(t+T)=f(t)$. چنانچه تابع متناوب نباشد و یا به عبارتی، تناوب آن برابر بی‌نهایت باشد ($T \to \infty$)، از سری فوریه عبارت زیر به دست می‌آید:

$$F(\lambda)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(t)e^{-i\lambda t}dt$$

$$f(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} F(\lambda)e^{i\lambda t}d\lambda$$

در واقع، سری فوریه، تابعی است که با استفاده از آن می‌توان هر تابع متناوب را به صورت جمعی از توابع نوسانی ساده (سینوسی، کسینوسی و یا تابع نمایی مختلط) نوشت. با بسط هر تابع به صورت

[1] - Schuster

سری فوریه، مولفه‌های بسامدی آن به دست می‌آید. برای حل بسیاری از مسائل بهتر است که تابع در دامنه فرکانس تعریف شده باشد زیرا این دامنه ویژگی‌هایی دارد که به راحتی محاسبات می‌انجامد. فوریه بر این باور بود که هر گونه تابع متناوب را می‌توان به صورت جمعی از توابع سینوسی نوشت، این مطلب درست نمی‌باشد. شرایط لازم برای هر تابع متناوب برای اینکه به صورت فوریه نوشته شود به صورت زیر است:

- تابع در هر دوره تناوبی انتگرال پذیر باشد:

$$\int_a^{a+T} |f(z)| dz < \infty$$

- تابع فقط شمار محدودی بیشینه و کمینه دارد.
- تابع فقط شمار محدودی ناپیوستگی دارد.

نمایش طیفی، یک فرآیند مانای $\{Z_t\}$ را به مجموعی از مؤلفه‌های سینوسی با ضرایب تصادفی ناهمبسته تجزیه می‌کند.

۱-۲-۳ تابع توزیع طیفی و تابع چگالی طیفی

برای تابع اتوکواریانس معلوم γ_k، می‌توانیم نمایش طیفی را به صورت زیر بنویسیم:

$$\gamma_k = \int_{-\pi}^{\pi} e^{i\lambda k} dF(\lambda) \qquad (۴-۱)$$

که در آن $F(\lambda)$ تابع توزیع طیفی، نامیده می‌شود.

برای یک فرآیند مانای گسسته تصادفی، تابع توزیع طیفی، تابعی پیوسته (کراندار یکنوا) بر بازه $(0,\pi)$ است و بنابراین می‌تواند نسبت به λ در این بازه مشتق‌پذیر باشد و مشتق آن را با $f(\lambda)$ نمایش می‌دهند.