کاربرد بوت‌استرپ در روش‌گشتاورهای‌تعمیم‌یافته در سری‌های‌زمانی

محبوبه ربیعه، نصراله ایران‌پناه
گروه آمار، دانشگاه اصفهان

چکیده

یکی از مسائل مهم و متداول در مباحث آماری، تحلیل سری‌های زمانی است که در اقتصادسنجی کاربرد دارد. در تحلیل داده‌های سری‌زمانی شناسایی و برازش مدل از گام‌های اساسی به شمار می‌رود و باید پارامترهای مدل به روشی مناسب برآورد شوند. روش گشتاورهای تعمیم یافته، به عنوان یک روش مهم برآوردیابی در اقتصاد معرفی شده است. این روش به دلیل کاربرد مستقیم و محدودیت‌های کم در فرآیند تولید داده‌ها، یک ابزارمهم در اقتصادسنجی می‌باشد. روش‌های بازنمونه‌گیری بوت‌استرپ براساس نمونه‌ی مشاهده شده و بدون در نظرگرفتن فرضیاتی مانند مشخص بودن توزیع باقیمانده‌ها عمل می‌کند.

واژه‌های کلیدی: روش گشتاورهای تعمیم‌یافته، بوت‌استرپ بلوک متحرک، بوت‌استرپ باقیمانده‌ها، شبیه‌سازی مونت‌کارلو.

۱ مقدمه

اکثر روش‌های آماری برای تحلیل، نیاز به بعضی اطلاعات از توزیع نمونه‌ای دارند. در یک مسأله‌ی برآورد به علت اینکه هر برآوردکننده ممکن است یک خطای برآورد داشته باشد، داشتن یک اندازه‌ی دقت برای برآوردکننده بسیار مهم است. بنابراین انتخاب مدل و برازش آن نقش مهمی در تحلیل داده‌ها در مراکز و سازمان‌های کشور دارا می‌باشند. یکی از قدیمی‌ترین روش‌های برآوردیابی، روش برآورد گشتاوری است که مبتنی بر به کارگیری گشتاورهای جمعیتی (توزیع) و گشتاورهای نمونه می‌باشد. بسیاری از مدل‌های اقتصادسنجی تنها توسط شرایط گشتاوری برازش داده

می‌شوند، از این رو GMM [1] رایج‌ترین روشی است که در استنباط این مدل‌ها کاربرد دارد. هانسن [۶] روش GMM را به عنوان یک روش مهم برآوردیابی در اقتصاد معرفی نمود. هال و هارویتز [۴] کاربرد روش بوت‌استرپ بلوکی را برای برآوردگر GMM ارائه نمودند. براون و نووی[۱] کاربرد روش بوت‌استرپ نیم‌پارامتری را در محاسبه برآوردگر GMM پیشنهاد نمودند. هاردلی و همکاران [۵] روش‌های بوت‌استرپ را در سری‌های‌زمانی معرفی نمودند. نووی و اسمیت [۸] خواص حدی برآوردگرهای GMM را بررسی نمودند. اینو و شینتانی [۷] خواص مجانبی برآوردگرهای GMM را در سری‌های زمانی معرفی نمودند. چی‌اوسه [۲] چگونگی محاسبه‌ی برآوردگرهای GMM را با نرم‌افزار R ارائه نمود.

در این مقاله ابتدا، روش گشتاورهای تعمیم یافته ارائه می‌شود. سپس با معرفی یک روش بوت‌استرپ بلوکی در یک مطالعه‌ی شبیه‌سازی احتمال پوشش در فاصله اطمینان صدکی را در طول بلوک‌های متفاوت مورد مقایسه قرارمی‌گیرد و در نهایت با استفاده از بوت‌استرپ نیم‌پارامتری به برآورد ضرایب اتورگرسیون مرتبه اول می‌پردازد.

۲ روش گشتاورهای تعمیم یافته (GMM)

روش گشتاوری تعمیم یافته ابتدا توسط هانسن [۶] به عنوان یک راهکار ساده در صورت پیچیده بودن تابع درستنمائی ارائه شد. روش GMM با به کارگیری گشتاورهای جمعیتی و نمونه، به برآورد پارامترهای توزیع جامعه می‌پردازد. در روش GMM، y_i مشاهدات مستقل و هم توزیع از بردار داده‌های Y هستند و θ یک بردار k بعدی از پارامترهای مورد علاقه و $h(Y;\theta)$ یک بردار s بعدی از توابع متغیرهای مشاهده شده و پارامترهای مورد علاقه است به طوریکه $s > k$ و برای بردار پارامتر واقعی θ_0 به طور منحصر به فرد شرایط گشتاوری $E_F[h(Y,\theta_0)] = 0$ برقرار باشد. به‌عبارت دیگر، امیدریاضی تابع $h(Y,\theta_0)$ تحت توزیع F برابر صفرشود. با تعریف $h_n(\theta) \equiv n^{-1}\sum_{i=1}^{n} h(y_i,\theta)$، برآوردگر GMM با مینیمم سازی $Q_n = h_n(\theta)[V_n(\tilde{\theta})]^{-1}h_n(\theta)$ به دست می آید که در آن $\tilde{\theta}$ یک مقدار اولیه برای θ_0 و $\tilde{V}_n \equiv V_n(\tilde{\theta})$ یک مقدار برآوردگر برای $V \equiv E_F[h(y_i,\theta_0)h(y_i,\theta_0)']$ است. برآوردگرهای GMM، به انتخاب شاخص گشتاوری وابسته و برآوردگرهایی سازگار، کارا، اریب و توزیعی مجانباً نرمال دارند.

برای مثال فرض کنید $x_1, x_2, \cdots, x_n$ یک نمونه ی تصادفی n تایی از توزیع $N(\mu,\sigma^2)$ باشد. روش گشتاوری پارامترهای μ و σ^2 را به صورت زیر می‌باشد.

$$\begin{aligned}\mu_1 &= E_{\mu,\sigma^2}(X_1) = \mu,\\ \mu_2 &= E_{\mu,\sigma^2}(X_1^2) = \mu^2 + \sigma^2.\end{aligned}$$

بنابراین داریم:

$$\begin{cases}\tilde{\mu} = \bar{x},\\ \tilde{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.\end{cases}$$

مثال دیگری را ملاحظه کنید. فرض کنید $y_1, y_2, \ldots, y_n$ یک نمونه تصادفی از جامعه ای نامعلوم با میانگین λ و واریانس معلوم λ باشد. با در نظر گرفتن تابع $h(y_i;\lambda) = y_i - \lambda$، قید گشتاوری جامعه به صورت $E[h(y_i;\lambda)] = E(y_i) - \lambda = 0$ و قید گشتاوری نمونه به

[1] Generalized method of moments

صورت

$$\frac{1}{n}\sum_{i=1}^{n} h(y_i;\lambda) = \frac{1}{n}\sum_{i=1}^{n} y_i - \lambda = 0,$$

می‌باشد که با حل آن $\hat{\lambda} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

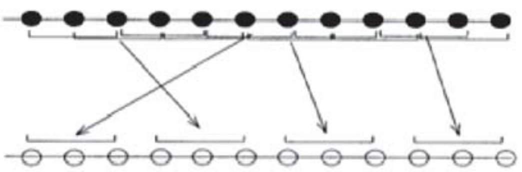
بنابراین $\bar{y}$ یک برآوردگر GMM برای λ است. حال روش متغیرهای ابزاری (IV)[2] را که در شبیه‌سازی‌ها کاربرد دارد را معرفی می‌کنیم. اگر شرط ناهمبسته بودن متغیر توضیحی و خطا در مدل رگرسیونی $Y = X\beta + u$ که در آن X یک ماتریس $n\times k$ و Y یک بردار k بعدی برقرار نباشد یعنی $E(X_i u_i) \neq 0$ آنگاه متغیر ابزاری H طوری معین می‌شود که شرط ناهمبسته بودن یعنی $E(u_i H_i) = 0$ برقرار شود و با حل معادله‌ی $\frac{1}{n}H'u(\beta) = 0$، β برآورد می‌شود چنین روشی را روش متغیر ابزاری گویند.

۳ بوت استرپ

بوت‌استرپ این امکان را برای یک نفر فراهم می‌سازد که تعداد زیادی نسخه‌ی جایگزین از یک آماره را که به طور معمول از یک نمونه محاسبه می‌شود را جمع‌آوری کند. برای به دست آوردن نمونه‌ی بوت‌استرپ در داده هایی با ساختاری وابسته مانند سری‌های زمانی بدون نیاز به فرض معلوم بودن ساختار دقیق این وابستگی می توان از روش‌های بازنمونه گیری بلوکی استفاده نمود.

۱.۳ بوت استرپ بلوک متحرک

در روش بوت استرپ بلوکی[3] برای اینکه ساختار وابستگی حفظ شود، ابتدا بلوک‌هایی از مشاهدات را در نظر گرفته سپس بلوک‌های بوت‌استرپ را با نمونه گیری تصادفی با جایگذاری از بلوک‌ها به دست آورده و در انتها نمونه‌ی بوت‌استرپ از به هم پیوستن بلوک‌های بوت‌استرپ حاصل می شوند. یکی از روش‌های بوت‌استرپ بلوکی، روش بوت‌استرپ بلوک متحرک است. همانطورکه در شکل ۱ مشاهده می‌کنید. دایره های سیاه، سری زمانی اصلی را مشخص می‌کند. برای تولید نمونه‌های بوت‌استرپ سری زمانی (دایره های سفید) یک طول بلوک، انتخاب شده و تمام بلوک‌های هم جوار ممکن از این طول در نظر گرفته می‌شوند. سپس نمونه گیری با جایگذاری از این بلوک ها انجام شده و با کنار هم قرار دادن آن ها سری‌زمانی جدید ساخته می شود. به عبارت دیگر اگر فرض کنیم که n مشاهدات اولیه داشته باشیم و b اندازه‌ی بلوک باشد آنگاه $k = n/b$ تعداد بلوک‌هایی است که باید با روش تصادفی ساده باجایگذاری از $n - b + 1$ بلوک ساخته شده، گرفته شود. اگر k صحیح نباشد، کوچکترین عدد صحیحی که در رابطه‌ی $kb \geq n$ صدق کند را به عنوان تعداد بلوک انتخابی در نظر می‌گیریم. لازم به ذکر است که پس از کنار هم قرار دادن بلوک‌ها تعداد نمونه‌ی بوت‌استرپ برابر تعداد مشاهدات اولیه یعنی n می‌باشد.



شکل ۱: نمودار بوت‌استرپ بلوک متحرک برای سری‌های زمانی

[2] Instrumental variables

[3] Block bootstrap

۲.۳ بوت استرپ نیم‌پارامتری

یک سری‌زمانی، دنباله‌هایی از مشاهدات یک متغیر بر حسب زمان است. در مدل های سری‌زمانی خودرگرسیو (AR) تحلیل داده های سری‌زمانی معمولاً مبتنی بر فرض نرمال بودن باقیمانده‌ها است. فرایندی را $AR(p)$ گوئیم که شامل p مرتبه‌ی خود رگرسیون باشد. مدل آن به صورت زیر تعریف می شود:

$$x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + \cdots + \theta_p x_{t-p} + a_t.$$

روش بوت‌استرپ برای برآورد اریبی، واریانس و توزیع نمونه‌ای آماره‌ها ارائه شده از جمله موارد کاربردی در سری‌های‌زمانی می‌باشد. روش بوت‌استرپ در مشاهدات مستقل با مشاهدات وابسته یکسان نمی‌باشد و با توجه به اینکه ساختمان وابستگی در مشاهدات وابسته معلوم (مانند سری‌های‌زمانی) و یا نامعلوم باشد، تفاوت دارد. روش بوت‌استرپ در مشاهدات وابسته که بوت‌استرپ باقیمانده‌ها نام دارد، تنها در مرحله‌ی اول یعنی تولید نمونه‌ی بوت‌استرپ تغییر خواهد کرد و دو مرحله‌ی بعدی مشابه الگوریتم بوت‌استرپ است.
در این قسمت روش بوت‌استرپ باقیمانده‌ها را در مدل $AR(1)$ بیان می‌کنیم. در مدل سری‌زمانی،

$$z_t = \beta z_{t-1} + \varepsilon_t,$$

که $t = u, u+1, \ldots, v$، β به صورت زیر برآورد می شود:

$$RSE(\hat{\beta}) = \min_b \sum_{t=u}^{v} (z_t - b z_{t-1})^2.$$

با تعریف $\hat{\varepsilon}_t = z_t - \hat{\beta} z_{t-1}$ و مرکزی کردن $\hat{\varepsilon}_t$ها به صورت $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - (\frac{1}{T}\sum_{t=u}^{v}\hat{\varepsilon}_t)$، نمونه‌ی تصادفی ساده باجایگذاری را از $\tilde{\varepsilon}_t$ها به صورت $\varepsilon_u^*, \cdots, \varepsilon_v^*$ به دست آورد. با در نظر گرفتن $z_1^* = z_1$ و $z_t^* = \hat{\beta} z_{t-1}^* + \varepsilon_t^*$؛ $t = u, u+1, \cdots, v$، β^* با روش کمترین توان دوم خطا محاسبه و با B بار تکرار کردن این روش انحراف معیار و اریبی به دست خواهد آمد. در دو بخش بعد، با استفاده از نرم‌افزار R و پکیج‌های $gmm, mvtnorm, FitAR, FitARMA$ به ارائه‌ی مثال‌های مورد نظر می‌پردازیم.

۴ شبیه‌سازی

در این قسمت توابع مختلفی از هسته‌ها را معرفی می‌کنیم.
هسته بارتلت توسط نووی و وست [۹] به صورت زیر ارائه شد:

$$k(x) = \begin{cases} 1 - |x|, & |x| \leq 1 \\ 0, & \text{سایر نقاط} \end{cases}$$

گلنت[۳] هسته‌ی پارزن را به فرم زیر معرفی نمود:

$$k(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & |x| \leq \frac{1}{2} \\ 2(1-|x|)^3, & \frac{1}{2} \leq |x| \leq 1 \\ 0, & \text{سایر نقاط} \end{cases}$$

با استفاده از نرم‌افزار R، هسته‌های معرفی‌شده را اجرا کرده و به مقایسه‌ی آنها می‌پردازیم.

مثال ۱. *مدل زیر را در نظر بگیرید:*

$$z_t = \beta_1 + \beta_2 x_t + u_t, \quad t = 1, \ldots, T,$$

به طوریکه اختلال [۴] *و رگرسورها* [۵] *دارای فرایند* $AR(1)$ *با مقدار ضریب همبستگی* $\rho = \{0/5، 0/9، 0/95\}$ *مشترک باشند. به عبارت دیگر* $u_t = \rho u_{t-1} + \varepsilon_{1t}$ *و* $x_t = \rho x_{t-1} + \varepsilon_{2t}$ *به طوریکه* $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})' \sim N(0, I_2)$.

[۴] Disturbabce
[۵] Regressors

در این مدل متغیر ابزاری برابر $(x_t, x_{t-۱}, x_{t-۲})$ می‌باشد. ابتدا x_t و u_t را از $AR(۱)$ تولید کرده، سپس متغیر تصادفی z_t را به دست می‌آوریم. آن‌ها را به روش بلوک متحرک بلوکی کرده و در هر حلقه‌ی بوت‌استرپ از بلوک‌ها به روش تصادفی ساده باجایگذاری نمونه گرفته و با کنار هم قرار دادن بلوک‌های به دست‌آمده، نمونه بوت‌استرپ حاصل می‌شود.

حال با این نمونه‌ی جدید مدل را برازش داده سپس ضرایب مدل را به روش GMM برآورد می‌کنیم. با اتمام حلقه‌ی بوت‌استرپ، B ضریب برآورد شده موجود است که آن‌ها را مرتب کرده و فاصله‌اطمینان صدکی را محاسبه می‌کنیم. با m بار تکرار مونت‌کارلو برای تمامی این مراحل، m فاصله‌اطمینان به دست‌آمده و در پایان سطح پوشش برای این فاصله‌اطمینان حاصل می‌شود. تمامی این نکات با مقادیر مختلف ρ و طول بلوک اجرا و نتایج احتمال پوشش از فاصله اطمینان صدکی نود درصد برای برآوردگر گشتاورهای تعمیم‌یافته در جدول ۱ ارائه شده است.

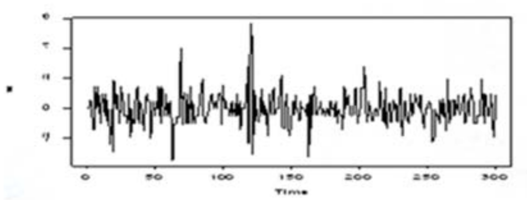
جدول ۱: نتایج عددی مثال شبیه‌سازی

هسته‌ها	طول بلوک	$\rho=۰/۹۵$	$\rho=۰/۹$	$\rho=۰/۵$
Bartlett	۶۴	۸۲/۴	۸۷/۹	۷۸/۵
	۱۲۸	۷۹/۶	۸۳/۴	۸۶/۳
Parzen	۶۴	۸۲/۳	۷۸/۴	۸۷/۵
	۱۲۸	۸۵/۱	۷۸/۷	۸۴/۲

همانطورکه در جدول بالا مشخص است، با افزایش طول بلوک، احتمال پوشش افزایش یافته و در طول بلوک بالاتر، با افزایش مقدار ρ این احتمال پوشش به طور ناچیزی افزایش پیدا می‌کند.

۵ مثال کاربردی

داده‌های روزانه از انتخاب انبارها در پنج شرکت مختلف درفرانسه از سال ۱۹۹۳ تا ۲۰۰۹ را در نظرگرفته شده است و سهام این شرکت‌ها را مورد بررسی قرار می‌دهیم. این شرکت‌ها عبارتند از: WMK، UIS، ORB، MAT و $ABAX$. این داده‌ها در نرم‌افزار R با عنوان $Finance$ موجود می‌باشد. نمودار سری‌زمانی آن به صورت زیر می‌باشد.



شکل ۲: نمودار سری زمانی مثال داده واقعی

برای این داده‌ها، مدل سری‌زمانی $z_t = \theta z_{mt} + \varepsilon_t$ را در نظر بگیرید که در آن z_t ماتریسی از اختلاف داده‌ها با نرخ ریسک بازار سهام و z_{mt} برداری ازاختلاف بازده انتخاب ابزار و نرخ ریسک بازار سهام می‌باشد. به‌طوریکه z_t ماتریسی $a \times b$، z_{mt} برداری b بعدی و پارامتر $\theta = (\alpha', \beta')'$ ماتریسی $a \times ۱$ می‌باشد که α و β به ترتیب عرض از مبدا و شیب مدل است. متغیر ابزاری در این مدل به صورت $(۱, z_{mt})$ تعریف شده‌است. با در نظر گرفتن ۳۰۰ نمونه و تعداد تکرار بوت‌استرپ ۱۰۰، انحراف معیار و اریبی برای ضرایب α و β به روش GMM در جدول ۲ ارائه شده‌است.

در جدول ۳ فاصله اطمینان ۹۰ درصد، هر دو ضریب محاسبه شده‌اند.

جدول ۲: اریبی و انحراف‌معیار ضرایب در مثال واقعی

ضرایب	اریبی	انحراف معیار
α	۰/۳۲۰۹۸۱۳	۰/۰۶۱۱۲۷۴
β	۰/۰۱۵۹۷۶۷	۰/۱۱۴۱۹۲۶

جدول ۳: فاصله اطمینان ضرایب در مثال واقعی

ضرایب	متغیرها	کران بالا	کران پائین
α	WMK	۰/۰۹۳۲۰	-۰/۱۰۲۵۴
	UIS	۰/۳۰۴۰۲	-۰/۰۹۹۳۳
	ORB	۰/۵۰۷۲۸	-۰/۲۱۵۵۳
	MAT	۰/۲۱۶۵۳	-۰/۱۴۴۷۴
	ABAX	۰/۵۵۵۰۲	-۰/۳۷۲۳۴
β	WMK	۰/۵۳۹۸۲	۰/۰۹۴۵۶
	UIS	۱/۶۵۰۶۳	۰/۸۷۴۷۹
	ORB	۲/۲۱۱۳۵	۰/۷۷۶۴۶
	MAT	۱/۴۰۴۸۵	۰/۶۲۵۱۲
	ABAX	۲/۰۷۳۰۸	۰/۱۰۴۸۶

۶ بحث و نتیجه‌گیری

در این مقاله روش GMM را معرفی نموده و در سری‌های زمانی استفاده از روش‌های بوت‌استرپ نیم‌پارامتری و بلوکی را در شبیه سازی بیان نموده‌ایم. در تحلیل داده‌های سری‌های زمانی معمولاً فرض نرمال بودن باقیمانده‌ها الزامی است. برای این منظور از روش بوت‌استرپ نیم‌پارامتری استفاده کرده‌ایم. همچنین با استفاده از مطالعه‌ی شبیه‌سازی مونت‌کارلو، روش GMM معرفی شده را به‌صورت کاربردی مورد بررسی قرارداده‌ایم.

مراجع

[1] Brown, B.W. and Newey, W.K. (2002). Generalised method of moments, efficient bootstrapping, and improved inference. Journal of Business and Economic Statistics, 20, 507-517.

[2] Chausse, P. (2010). Computing generalized method of moments and generalized empirical likelihood with R. Journal of Statistical Software, 34, 11, 1-35.

[3] Gallant, A.R. (1987). Nonlinear Statistical Models. New York; Wiley.

[4] Hall, P. and Horowitz, J. (1996). Bootstrap critical values for tests based on generalizedmethod-of-moments estimators. Econometrica, 64, 891-916.

[5] Hardle, W., Horowitz, J. and Kreiss, J. (2003). Bootstrapping methods for time series. International Statistical Review, 71, 435-459.

[6] Hansen, L. (1982). Large sample properties of generalized method of moments estimators. Econometrica, 50, 1029-1054.

[7] Inoue, A. and Shintani, M. (2006). Bootstrapping GMM estimators for time series. Journal of Econometrics, 133, 531-555.

[8] Newey, W. and Smith, R. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica, 72, 219-256.

[9] Newey, W. and West, K.D. (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55, 703-708.