



وزن‌دهی کالبیده برای جبران مقادیر کرانگین، بی‌پاسخی و ناپوشانش در طرح نیروی کار

آرمان بیداربخت‌نیا[†]* و حمیدرضا نواب‌پور[‡]

[†] مرکز آمار ایران

[‡] دانشگاه علامه طباطبائی

چکیده. استفاده از متغیرهای کمکی مناسب در محاسبه‌ی برآوردهای موزون می‌تواند کارایی این برآوردها را بهبود بخشد. این نوع برآوردها به‌منظور جبران اثر احتمال‌های انتخاب نابرابر، بی‌پاسخی، ناپوشانش، یا نوسان‌های حاصل از نمونه‌گیری حول مقادیر معلوم جامعه به کار می‌روند. ساده‌ترین برآوردهایی که از اطلاعات کمکی استفاده می‌کنند برآوردهای رگرسیونی و نسبی هستند. دیویل و سارندال (۱۹۹۲) نشان دادند که برآوردهای رگرسیونی را می‌توان به شکل یک برآوردهای موزون به دست آورد که وزن‌های آن با مینیمم کردن یک تابع فاصله و به شرط برقراری یک سلسله معادلات کالبیده به دست آمده‌اند. به‌طور کلی وزن‌هایی را که به این شکل محاسبه می‌شوند، وزن‌های کالبیده، و برآوردهای حاصل را برآوردهای کالبیده گویند. در این مقاله ضمن معرفی صورت عام برآوردهای کالبیده، حالت‌های خاصی از این برآوردها که حاصل به‌کارگیری تابع‌های فاصله‌ی متفاوت است نیز مورد بررسی قرار می‌گیرد. سپس با استفاده از داده‌های طرح نیروی کار، برآورد نرخ بیکاری توسط نظام‌های مختلف وزن‌دهی محاسبه شده، دقت و بار محاسباتی این روش‌های برآورد نقطه‌ای با یکدیگر مقایسه می‌شوند.

واژگان کلیدی. وزن‌های طرح؛ بیش‌پوشانش؛ کم‌پوشانش؛ بی‌پاسخی؛ روش‌های وزن‌دهی؛ آمارگیری نیروی کار.

۱ مقدمه

نقص چارچوب، بی‌پاسخی و احتمال‌های انتخاب نابرابر همواره نتایج آمارگیری‌ها و به‌خصوص آمارگیری‌های خانواری را که در آن‌ها از طرح‌های پیچیده و چندمرحله‌ای استفاده می‌شود، دچار مشکل می‌کند. تلاش‌هایی که در مرحله‌ی طراحی و اجرای آمارگیری انجام می‌شود، فقط باعث کم شدن مشکلات بالا می‌شود و هیچ‌گاه آثار ناشی از آن‌ها کاملاً از بین نمی‌رود. آمارشناسان همواره در تلاش هستند تا با روش‌های برآورد مناسبی که به کار می‌گیرند، اثر حاصل از نقص چارچوب، بی‌پاسخی و احتمال‌های انتخاب نابرابر را جبران کنند. اغلب برای این کار از روش‌های تعدیل وزنی استفاده می‌شود. در این روش‌ها وزن‌های پایه‌ای طرح با استفاده از اطلاعات کمکی موجود به‌گونه‌ای تعدیل می‌شوند که توزیع‌های نمونه‌ای با توزیع‌های جامعه همگون شوند. ساده‌ترین برآوردهایی که از اطلاعات کمکی استفاده می‌کنند برآوردهای رگرسیونی و نسبتی هستند. دوایل و سارندال (۱۹۹۲) خانواده‌ای از برآوردها به نام برآوردهای کالیبره (calibrated) را معرفی کردند. در این برآوردها وزن‌هایی جستجو می‌شوند که بر اساس یک تابع فاصله‌ی مشخص، کم‌ترین فاصله تا وزن‌های پایه‌ی طرح را حفظ و در عین حال در یک سلسله معادلات کالیبره صدق کنند؛ به عبارت دیگر، برآوردهایی تولید کنند که با مقدار کل معلوم جامعه برای متغیرهای کمکی برابر باشند. در حال حاضر، حالت خاصی از برآوردهای کالیبره در طرح آمارگیری نیروی کار مورد استفاده قرار می‌گیرد. در بخش ۲ به معرفی اجمالی طرح نیروی کار می‌پردازیم که توسط مرکز آمار ایران اجرا می‌شود. در بخش ۳ پس از معرفی برآوردهای رگرسیونی تعمیم‌یافته، صورت عام برآوردهای کالیبره شرح داده می‌شود. سپس حالت‌های خاصی از برآوردهای کالیبره که بر اثر استفاده از تابع فاصله‌های متفاوت به دست می‌آیند، مورد بررسی قرار می‌گیرند و در ادامه، در بخش‌های ۴ و ۵ به بررسی نکات عملی و محاسباتی، علاوه بر نتایج حاصل از مقایسه‌ی روش‌های مختلف، پرداخته می‌شود.

۲ طرح آمارگیری نیروی کار

طرح آمارگیری نیروی کار یکی از طرح‌های مهم مرکز آمار ایران است که اطلاعات اساسی نیروی کار کشور را تولید می‌کند. پیش از این، شاخص‌های عمده‌ی نیروی کار به‌وسیله‌ی طرح آمارگیری از ویژگی‌های اشتغال و بیکاری خانوار به دست می‌آمد که از سال ۱۳۸۳ به بعد طرح آمارگیری نیروی کار جایگزین آن شد. این طرح که از روش نمونه‌گیری چرخشی بهره می‌گیرد، به‌منظور محاسبه‌ی برآوردهای فصلی و سالانه‌ی نیروی کار و نیز برآورد تغییرات آن در سال‌ها و فصل‌های متوالی و همچنین فصل‌های مشابه از دو سال متوالی طراحی شده است. نخست نمونه‌ی پایه، مشتمل بر بلوک‌ها و آبادی‌هایی که از چارچوب سرشماری کارگاهی ۱۳۸۱

انتخاب شده‌اند، ساخته شد و خانوارهای این بلوک‌ها و آبادی‌ها طی یک عملیات فهرست‌برداری، فهرست شدند. سپس خانوارهای نمونه برای هر دوره‌ی آمارگیری از بین این خانوارهای فهرست‌شده انتخاب شدند. به‌طورکلی در این طرح از یک خانوار نمونه به‌صورت چرخشی در طی فصل‌های مختلف، چندین بار (حد اکثر ۴ بار) آمارگیری می‌شود.

فرایند وزن‌دهی طرح نیروی کار در چهار مرحله‌ی جداگانه انجام می‌شود. در ابتدا وزن پایه اعمال می‌شود که عبارت است از وارون احتمال انتخاب هر خانوار در نمونه‌ی متعلق به استان. دومین مرحله‌ی وزن‌دهی برای جبران اثر خوشه‌های بی‌پاسخ انجام می‌شود. در مرحله‌ی سوم وزن‌دهی، تعدیل بی‌پاسخی برای خانوار در خوشه انجام می‌شود و نهایتاً در آخرین مرحله‌ی وزن‌دهی، وزن‌های حاصل از مرحله‌ی سوم، برای پیش‌بینی‌های جمعیتی تعدیل می‌شوند. عامل تعدیل در مرحله‌ی دوم عبارت است از وارون نرخ خوشه‌های آمارگیری شده در استان که روی وزن پایه اعمال می‌شود و عامل تعدیل مرحله‌ی سوم، وارون نرخ پاسخ خانوار در خوشه است که روی وزن حاصل از مرحله‌ی دوم اعمال می‌شود. در مرحله‌ی آخر، وزن‌دهی یک حالت خاص از براوردگرهای کالبدی (برآوردگر پس‌طبقه‌بندی کامل) مورد استفاده قرار می‌گیرد. برای این کار با استفاده از پیش‌بینی‌های جمعیتی، جمعیت خانوارهای معمولی ساکن در ۱۲ زیرگروه حاصل از شهری/روستایی، مرد/زن و سه گروه سنی کمتر از ۱۰ سال، ۱۰ تا ۶۴ سال و ۶۵ سال و بیش‌تر در فصل مورد نظر محاسبه می‌شوند. همچنین مجموع وزن‌های حاصل از مرحله‌ی قبل در این ۱۲ زیرگروه نیز محاسبه می‌شود. خارج قسمت جمعیت پیش‌بینی‌شده و مجموع وزن‌های حاصل از مرحله‌ی قبل در هر زیرگروه، عامل تعدیل وزن برای افراد متعلق به زیرگروه مورد نظر است که روی وزن حاصل از مرحله‌ی قبل اعمال می‌شود.

۳ براوردگرهای کالبدی

سارندال و دیگران (۱۹۹۲) با استخراج یک عامل تعدیل وزنی و تغییر صورت معمولی براوردگر رگرسیونی به یک براوردگر موزون، براوردگر رگرسیونی تعمیم‌یافته (GREG) را معرفی کردند که خود حالت خاصی از یک رده‌ی وسیع از براوردگرهایی است که از متغیرهای کمکی استفاده می‌کنند. این خانواده از براوردگرها را براوردگرهای کالبدی گویند.

سارندال و دیگران (۱۹۹۲) برای دستیابی به مقاصد مورد نظر، براوردگر رگرسیونی را به‌صورت یک

برآوردگر موزون از مقادیر مشاهده شده y_k نوشتند:

$$(۱) \quad \hat{t}_{y\text{GREG}} = \sum_s a_k g_k y_k.$$

وزن کلی که به y_k داده می شود عبارت است از حاصل ضرب دو مؤلفه a_k و g_k ، یکی وزن های پایه ی طرح (a_k وزن پایه ی طرح برای عضو k ام نمونه و برابر با وارون احتمال شمول آن، π_k ، است) و دیگری g_k که هم به عضو k ام و هم به تمام اعضای نمونه ی s که عضو k ام به آن تعلق دارد، وابسته است. مؤلفه ی g_k به صورت زیر محاسبه می شود:

$$(۲) \quad g_k = 1 + c_k \left(\sum_U X_k - \sum_s a_k X_k \right)' \left(\sum_s a_k c_k X_k X_k' \right)^{-1} X_k.$$

وقتی وزن $a_k g_k$ برای بردار کمکی X_k به کار می رود و روی عناصر $k \in s$ جمع بسته می شود، یک برآورد از مقدار کل صفت کمکی، t_x ، در جامعه به دست می آید. این برآورد دقیقاً برابر با مقدار معلوم مقدار کل t_x در جامعه است (دویل و سارندال، ۱۹۹۲). لذا داریم:

$$(۳) \quad \sum_s a_k g_k X_k = \sum_U X_k.$$

این نتیجه بسیار مطلوب است؛ زیرا اگر این برآورد، مقداری غیر از آن چیزی که ما از مقدار معلوم جامعه در دست داریم، داشته باشد، وزن های مورد نظر مستدل نخواهند بود. این وزن های $a_k g_k$ را با w_k نشان داده، وزن های کالبدی می نامیم. علاوه بر ویژگی بالا وزن های w_k فاصله ی خ ی دو، $\sum_s \frac{(w_k - a_k)^2}{a_k}$ ، را نیز مینیمم می کنند. به طور کلی در همه ی برآوردگرهای کالبدی به صورت $\sum_s w_k y_k$ با وزن های w_k که در معادلات کالبدی (۳) صدق می کنند و یک تابع فاصله را مینیمم می کنند سر و کار داریم. روش کالبدن با این تابع فاصله را که منجر به GREG می شود تعدیل خطی می نامند. همان طور که از رابطه ی (۲) مشخص است، عامل تعدیل g_k می تواند مقدار منفی بگیرد که نتیجه ی مطلوبی نیست. برای رفع این مشکل می توان وزن ها را به مقادیر مثبت محدود کرد یا تابع فاصله ای به کار گرفت که وزن های منفی تولید نکند.

۳/۱ صورت عام برآوردگرهای کالبدی

در قسمت قبل، فاصله ی بین وزن های اصلی a_k و وزن های جدید w_k ، با اندازه ی $\frac{(w_k - a_k)^2}{a_k}$ مشخص شد؛ اما طبیعی است که اندازه های دیگری برای این فاصله باید مورد بررسی قرار گیرند. این اندازه ها باید در

بعضی ویژگی‌های پایه‌ای مشترک باشند. دوئل و سارندال (۱۹۹۲) تابع فاصله‌ی $D_k(w, a)$ را با فرض تعدادی ویژگی اولیه تعریف کردند. با در نظر گرفتن $d_k(w, a) = \frac{\partial D_k(w, a)}{\partial w}$ ، میانگین فاصله با استفاده از $E_p \{ \sum_s D_k(w, a_k) \}$ اندازه‌گیری می‌شود. مینیم کردن این کمیت به منظور برقراری مقایسه‌ای (۳) برای نمونه‌ی s معادل است با جستجوی w_k ‌ای که برای هر s ویژه، مجموع $\sum_s D_k(w_k, a_k)$ را تحت قیده‌ای مقایسه‌ای (۳) مینیم کند. اگر λ نشان‌دهنده‌ی بردار j -بعدی ضرایب لاگرانژ باشد، خواهیم داشت:

$$(۴) \quad d_k(w_k, a_k) - X'_k \lambda = 0.$$

فرض‌ها تضمین می‌کنند که جواب معادله، در صورت وجود، واحد باشد. به این ترتیب همیشه می‌توان نوشت:

$$(۵) \quad w_k = a_k F_k(X'_k \lambda),$$

که $a_k F_k(\cdot)$ تصویر وارون $d_k(\cdot, a_k)$ است.

همان‌طور که ملاحظه شد، λ در رابطه‌ی (۵) یک مقدار نامعلوم است. با توجه به رابطه‌ی (۳)، معادله‌های کالیده (قیده‌ای مقایسه‌ای) لازم برای تعیین $\lambda = (\lambda_1, \dots, \lambda_j, \dots, \lambda_J)'$ عبارت‌اند از:

$$(۶) \quad t_x = \sum_s w_k X_k = \sum_s a_k F_k(X'_k \lambda) X_k.$$

در رابطه‌ی (۶) فقط مقدار نامعلوم λ است. پس از تعیین λ با استفاده از رابطه‌ی (۵) مقدار w_k به دست می‌آید و به این ترتیب براوردگرهای کالیده‌ی حاصل برای t_y به شکل زیر محاسبه می‌شوند:

$$(۷) \quad \hat{t}_{yw} = \sum_s w_k y_k = \sum_s a_k F_k(X'_k \lambda) y_k.$$

در ادامه با به کار بردن تابع‌های فاصله‌ی مختلف، چهار نوع براوردگر کالیده مورد بررسی قرار می‌گیرد.

۳/۲ رگرسیون خطی

با انتخاب تابع فاصله‌ی $\frac{(w_k - a_k)^T}{2a_k}$ داریم $F_k(X'_k \lambda) = 1 + X'_k \lambda$ (راو، ۲۰۰۳). وزن‌های کالیده به صورت $w_k = a_k(1 + X'_k \lambda)$ به دست می‌آیند. بردار ضرایب لاگرانژ λ با استفاده از رابطه‌ی $\sum_s w_k X_k = t_x$ به شکل زیر تعیین می‌شود:

$$\lambda = T_s^{-1}(t_x - \hat{t}_x),$$

که در آن فرض می‌شود $T_s = \sum_s a_k X_k X'_k$ وارون‌پذیر باشد. برآوردگر حاصل برای t_y عبارت است از:

$$(۸) \quad \hat{t}_{y\text{GREG}} = \sum_s w_k y_k = \hat{t}_y + (t_x - \hat{t}_x)' \hat{B}_s,$$

که $\hat{t}_x = \sum_s a_k X_k$ برآوردگر هورویتز-تامپسون برای بردار متغیرهای کمکی X است و $\hat{B}_s = T_s^{-1} \sum_s a_k X_k y_k$.

همان‌گونه که ملاحظه می‌شود رابطه‌ی (۸) همان برآوردگر رگرسیونی تعمیم‌یافته است که با استفاده از روش کالبدین (calibration) حاصل شده است.

۳/۳ چنگک‌زنی (Raking)

در این روش که معمولاً به کار گرفته می‌شود، با به کار بردن تابع فاصله‌ی $D_k(w_k, a_k) = w_k \log(\frac{w_k}{a_k}) - w_k + a_k$ مدل عامل تعدیل $F_k(X'_k \lambda)$ به صورت $\exp(X'_k \lambda)$ در نظر گرفته می‌شود که ضرورتاً عامل تعدیل را نامنفی می‌کند. با استفاده از روش عددی نیوتون-رافسون برای مینیمم کردن تابع فاصله‌ی $D_k(w_k, a_k)$ با هدف برقراری قیدهای مقایسه‌ای، وزن‌های کالبدیه به صورت زیر به دست می‌آیند:

$$(۹) \quad w_k = a_k \exp(X'_k \lambda).$$

۳/۴ چنگک‌زنی محدود شده

همان‌طور که اشاره شد، روش چنگک‌زنی هرچند از تولید وزن‌های منفی جلوگیری می‌کند، امکان تولید وزن‌های با مقدار زیاد را نیز فراهم می‌سازد. به این ترتیب با استفاده از همان تابع فاصله‌ای که در روش چنگک‌زنی به کار گرفته شد، در این روش به دنبال عامل‌های تعدیل $F_k(X'_k \lambda)$ هستیم که محدود به یک دامنه‌ی از پیش تعیین شده باشند. به این ترتیب با تعیین دامنه‌ی $[L, U]$ ، مقدار عامل تعدیل را به گونه‌ای به دست می‌آوریم که همواره داشته باشیم $La_k \leq w_k \leq Ua_k$.

۳/۵ روش دوئل و سارندال (D&S)

ممکن است کسی بخواهد از تابع $F_k(X'_k \lambda)$ که می‌تواند وزن‌های خیلی بزرگ بسازد، اجتناب کند؛ زیرا به کار بردن این وزن‌ها در محاسبه‌ی برآوردها برای ساختن چندین زیرجامعه (حوزه‌ها)، ممکن است برای

برخی حوزه‌ها برآوردهای غیر واقعی تولید کنند. بنا بر این، توابع دیگر با این ویژگی جالب که وزن‌ها را به دامنه‌ای محدود می‌کنند که آمارشناس بتواند این دامنه را از پیش معلوم کند مورد ملاحظه قرار می‌دهیم. این توابع در سال ۱۹۹۲ توسط دوایل و سارندال معرفی شدند. آن‌ها از تابع فاصله‌ی $D_k(w_k, a_k)$ در روش چنگک‌زنی استفاده کردند و یک تابع فاصله‌ی جدید به دست آوردند. همان‌طور که ملاحظه شد، در روش چنگک‌زنی، مقادیر $F_k(X'_k \lambda) = \exp(X'_k \lambda)$ در دامنه‌ی $(0, \infty)$ قرار دارند. برای محدود کردن وزن‌ها و به‌طور خاص برای اجتناب از وزن‌های خیلی بزرگ، دو ثابت L و U به گونه‌ای تعیین شدند که $L < 1 < U$ و فرض می‌کنیم $A = \frac{U-L}{(1-L)(U-1)}$. با تعریف

$$(10) \quad F_k(X'_k \lambda) = \frac{L(U-1) + U(1-L) \exp(AX'_k \lambda)}{(U-1) + (1-L) \exp(AX'_k \lambda)}$$

داریم $F_k(0) = 1$; $F_k(\infty) = U$; $F_k(-\infty) = L$

نتیجه می‌شود که وزن‌های $w_k = d_k F_k(X'_k \lambda)$ به صورت $Ld_k < w_k < Ud_k$ محدود شده‌اند. توجه به این نکته مهم است که تابع فاصله‌ی $G_k(w_k, d_k)$ برای این مورد، عبارت است از:

$$(11) \quad D_k(w_k, a_k) = A^{-1} \sum_{k=1}^n a_k [(F_k - L) \log \{(1-L)^{-1}(F_k - L)\} + (U - F_k) \log \{(U-1)^{-1}(U - F_k)\}]$$

اگر L یک مقدار منفی بزرگ و U یک مقدار مثبت بزرگ باشد، به روش رگرسیون خطی و اگر $L = 0$ و U بزرگ باشد، به روش چنگک‌زنی نزدیک می‌شویم. همان‌طور که قبلاً نیز اشاره شد، به کار بردن تابع‌ها فاصله‌ی مختلف، منجر به برآوردهای مختلف می‌شود.

۳/۶ واریانس و برآورد واریانس

دوایل و سارندال (۱۹۹۲) به این نتیجه رسیدند که \hat{t}_{yw} به‌طور مجانبی معادل با \hat{t}_{yGREG} می‌باشد که یک مورد خاص از \hat{t}_{yw} با $\hat{t}_{yw} = 1 + X'_k \lambda$ است. به این ترتیب، رابطه‌ی زیر را برای برآورد واریانس برآوردگر کالبدی که معرفی کرده بودند، پیشنهاد دادند:

$$(12) \quad \hat{V}(\hat{t}_{yw}) = \sum_s \sum_k \left(\frac{\Delta_{kl}}{\pi_{kl}} \right) (w_k e_k)(w_l e_l),$$

که $e_k = y_k - X'_k \hat{B}_{ws}$ در معادلات نرمال مبتنی بر نمونه صدق می‌کند، $\pi_{kk} = \pi_k$ و $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$. وزن‌های کالبدی w_k در رابطه‌ی (۱۲) به منظور وزن دادن به مانده‌های e_k به کار رفته است.

در این مقاله با توجه به محدودیت‌های نرم‌افزای و سخت‌افزاری، برای محاسبه‌ی واریانس برآوردهای کالبدیه از روش گروه‌های تصادفی استفاده شده است. فان و دیگران (۱۹۸۰) در یک بررسی که بر اساس داده‌های سرشماری ۱۹۷۰ آمریکا انجام گرفت، با مقایسه‌ی چهار روش خطی‌سازی سری تیلور، جک‌نایف، تکرار مکرر متعادل (BRR) و گروه‌های تصادفی، روش گروه‌های تصادفی با ۲۴ زیرنمونه را برای برآورد واریانس برآوردها نسبت چنگ‌زنی در سرشماری ۱۹۸۰ آمریکا پیشنهاد دادند.

۳/۶/۱ روش گروه‌های تصادفی

در روش گروه‌های تصادفی، m زیرنمونه به روش سیستماتیک از نمونه‌ی اصلی انتخاب می‌شود. در صورتی که $m = ۲$ باشد، برای هر زیرنمونه، وزن‌های جدید با استفاده از روش تعدیل وزنی مورد نظر محاسبه می‌شود، با این تفاوت که کنترل‌های مقایسه‌ای در این‌جا برابر با حاصل ضرب $\frac{1}{m}$ در مقادیر معلوم جامعه است. برای m ‌های بزرگ‌تر از ۲ نیازی به محاسبه‌ی وزن‌های جدید در هر زیرنمونه نیست و در تمام زیرنمونه‌ها از وزن‌های تولید شده با کل نمونه استفاده می‌شود. به این ترتیب اگر $\hat{X}_{(j)}$ برآورد مقدار کل متغیر مورد بررسی X در زیرنمونه‌ی j ام، ($j = ۱, ۲, \dots, m$)، باشد، برآورد واریانس برآوردها \hat{X} عبارت است از:

$$(۱۳) \quad \hat{V}_{RG} = (1 - f) \frac{m}{m-1} \sum_{j=1}^m \left(\hat{X}_j - \sum_{j=1}^m \frac{\hat{X}_j}{m} \right)^2,$$

که در آن f عبارت است از کسر نمونه‌گیری.

۴ نکات عملی و محاسباتی

به‌منظور انجام محاسبات لازم برای انجام تعدیل وزنی، برنامه‌ای با استفاده از نرم‌افزار SAS/IML برای هر یک از روش‌ها نوشته شد. همان‌طور که اشاره شد روش‌های مورد نظر، به‌جز روش رگرسیون خطی، روش‌های مبتنی بر تکرارند؛ لذا یک مقدار ثابت δ به‌عنوان معیار توقف تکرارها در برنامه در نظر گرفته شد. البته برای رسیدن به همگرایی، علاوه بر ثابت δ معیار دیگری نیز به‌عنوان حد اکثر تعداد تکرارها منظور شد. این تعداد نیز برای روش چنگ‌زنی ۵ تکرار و برای سایر روش‌ها ۱۰ تکرار در نظر گرفته شد.

۴/۱ تعدیل وزنی رگرسیون خطی

برنامه‌ای که برای انجام تعدیل وزنی با استفاده از روش رگرسیون خطی تهیه شده، بدون تکرار است و فقط در یک مرحله وزن‌های نهایی را برای هر یک از افراد با پاسخ نمونه به‌گونه‌ای تولید می‌کند که معادلات کالیده برای تمام کنترل‌ها در سطح استانی و کل کشور برقرار شود. در این روش، تمامی عوامل‌های تعدیل مقدارهای مثبت اختیار می‌کنند و در این مورد، مشکل وزن‌های منفی مشاهده نمی‌شود. دلیل این امر، این است که در تمامی موارد، برآورد هورویتز-تامپسون جمعیت در زیرگروه مورد نظر کم‌تر از جمعیت زیرگروه از پیش‌بینی جمعیتی است. به عبارت دیگر با فرض درست بودن اطلاعات کمکی، وزن‌های پایه‌ی طرح در تمامی موارد دچار کم‌برآورد هستند.

۴/۲ تعدیل وزنی چنگک‌زنی

همان‌طور که توضیح داده شد، روش چنگک‌زنی مبتنی بر تکرار است و در چند مرحله انجام می‌شود؛ به این ترتیب نسبت به روش رگرسیون خطی از حجم محاسباتی بیشتری برخوردار است و اجرای برنامه‌ی رایانه‌ای آن نیز با توجه به اندازه‌ی زیاد داده‌ها، زمان بیشتری را صرف می‌کند. این روش پس از ۵ تکرار و با معیار $\delta = 10$ به همگرایی رسیده است. در واقع δ فاصله‌ی بردار برآوردهای نهایی متغیرهای کمکی با استفاده از آخرین وزن‌های به دست آمده و بردار برآوردهای حاصل از وزن‌های پایه‌ای برای متغیرهای کمکی است.

روش چنگک‌زنی هرچند از تولید وزن‌های منفی جلوگیری می‌کند، ممکن است وزن‌های با مقدار کرانگین تولید کند. به این ترتیب با در نظر گرفتن مرزهایی برای عامل‌های تعدیل، این عامل‌ها علاوه بر این‌که وزن‌هایی تولید می‌کنند که در معادلات کالیده صدق کنند، در حین ساخت به‌گونه‌ای محدود می‌شوند که از مرزهای تعیین‌شده بیش‌تر نباشند.

۴/۳ چنگک‌زنی محدود شده

با توجه به مقادیر مینیمم و ماکسیمم عامل‌های تعدیل حاصل از تعدیل چنگک‌زنی برای هر یک از استان‌ها و برای جلوگیری از تولید وزن‌های کرانگین، در هر استان یک مرز پایینی (L) و یک مرز بالایی (U) برای عامل‌های تعدیل در نظر گرفته شد. برای این کار نخست یک مجموعه مرزهای بالایی و پایینی با شرط $L \leq U \leq 1$ و با دامنه‌ی نسبتاً کم ($L = 0.95$ و $U = 1.2$) در نظر گرفته شد. در اجرای برنامه با استفاده از این مرزها، همان‌طور که انتظار می‌رفت به دلیل کوچک بودن دامنه‌ی مورد قبول برای

جدول ۱. مرزهای تعیین شده برای روش چنگک زنی محدود شده

استان	L	U	استان	L	U
مرکزی	۰/۹	۱/۵۲	لرستان	۰/۸۵	۱/۴۸
گیلان	۰/۸۵	۱/۵۶	ایلام	۰/۹	۱/۷۲
مازندران	۰/۸۵	۱/۵۸	کهگیلویه و بویراحمد	۰/۸۵	۱/۹
آذربایجان شرقی	۰/۸۵	۱/۳۹	بوشهر	۰/۸۵	۱/۴۷
آذربایجان غربی	۰/۹	۱/۴۸	زنجان	۰/۸۵	۱/۴۹
کرمانشاه	۰/۹	۱/۴۷	سمنان	۰/۹	۱/۵۳
خوزستان	۰/۹	۱/۵۳	یزد	۰/۸۷	۱/۶۱
فارس	۰/۹	۱/۵۸	هرمزگان	۰/۸۵	۱/۵۴
کرمان	۰/۸۵	۱/۶	تهران	۰/۸	۱/۷۴
خراسان رضوی	۰/۷	۱/۲۸	اردبیل	۰/۸۵	۱/۴۹
اصفهان	۰/۸۵	۱/۵۶	قم	۰/۷۵	۱/۴۸
سیستان و بلوچستان	۰/۹	۱/۶۷	قزوین	۰/۸۵	۱/۵۷
کردستان	۰/۹	۱/۷۶	گلستان	۰/۹	۱/۵۷
همدان	۰/۸	۱/۵۲	خراسان شمالی	۰/۸۵	۱/۵۱
چهارمحال و بختیاری	۰/۸۱	۱/۵۱	خراسان جنوبی	۰/۸	۱/۴۹

وزن های نهایی، در بیش تر استان ها همگرایی حاصل نشد. سپس دامنه ی مرزهای هر یک از استان ها تا جایی که همگرایی حاصل شود، آزاد شد و مرزهای نهایی به صورتی که در جدول ۱ قابل مشاهده است، به دست آمدند. به دلیل کوچک بودن دامنه ی عامل های تعدیل در روش چنگک زنی، وزن های حاصل از تعدیل چنگک زنی و چنگک زنی محدود شده تفاوت بسیار کمی با یکدیگر دارند.

۴/۴ روش D&S

این روش در ابتدا با استفاده از مرزهای نهایی که برای روش چنگک زنی تعیین شده بود، اجرا شد. با توجه به این که در مقایسه با روش چنگک زنی محدود شده، رسیدن به همگرایی در این روش دشوارتر است، با استفاده از این مرزها در بیش تر استان ها همگرایی حاصل نشد. به این ترتیب برای دستیابی به مرزهای مناسب، به ناچار در هر استان با دست کاری مرزهای پایینی و بالایی، مقادیر مناسب برای مرزها به دست آمد که در جدول ۲ قابل مشاهده است.

۵ بحث و نتیجه گیری

با استفاده از چهار برآوردرگر کالبدیه که در بخش های قبل معرفی شدند و داده های مربوط به

جدول ۲. مرزهای تعیین‌شده برای روش D&S

استان	L	U	استان	L	U
مرکزی	۰/۸۸	۱/۵۲	لرستان	۰/۸	۱/۴۸
گیلان	۰/۸۲	۱/۵۶	ایلام	۰/۸۹	۱/۷۲
مازندران	۰/۸۵	۱/۵۸	کهگیلویه و بویراحمد	۰/۸۲	۱/۹
آذربایجان شرقی	۰/۸۱	۱/۳۹	بوشهر	۰/۸۵	۱/۴۷
آذربایجان غربی	۰/۸۹	۱/۴۸	زنجان	۰/۸۵	۱/۴۹
کرمانشاه	۰/۸۹	۱/۴۷	سمنان	۰/۸۸	۱/۵۳
خوزستان	۰/۸۸	۱/۵۳	یزد	۰/۸۷	۱/۶۱
فارس	۰/۸۸	۱/۵۸	هرمزگان	۰/۸۳	۱/۵۴
کرمان	۰/۸۱	۱/۶	تهران	۰/۸	۱/۷۴
خراسان رضوی	۰/۷	۱/۲۸	اردبیل	۰/۸۵	۱/۴۹
اصفهان	۰/۸۵	۱/۵۶	قم	۰/۷۵	۱/۴۸
سیستان و بلوچستان	۰/۸۹	۱/۶۷	قزوین	۰/۸۵	۱/۵۷
کردستان	۰/۸۹	۱/۷۶	گلستان	۰/۸۹	۱/۵۷
همدان	۰/۸	۱/۵۲	خراسان شمالی	۰/۸۵	۱/۵۱
چهار محال و بختیاری	۰/۸	۱/۵۱	خراسان جنوبی	۰/۸	۱/۴۹

طرح آمارگیری نیروی کار در فصل پاییز سال ۱۳۸۴، وزن‌دهی کالبدی و برآوردهای حاصل از این وزن‌ها، به‌روشی که قبلاً توضیح داده شد، برای نرخ بیکاری و جمعیت افراد بیکار در نقاط شهری و روستایی هریک از ۳۰ استان محاسبه شده است. با توجه به این که برآوردهای کالبدی به‌منظور جبران اثرهای ناشی از بی‌پاسخی واحد، نقص چارچوب و احتمال‌های انتخاب نابرابر به کار می‌روند، طبیعی است که کارایی این روش‌ها در مواردی که فایده داده‌ها با مشکلات بیش‌تری از این قبیل مواجه باشد، محسوس‌تر است. با این حال هرچند در داده‌های مربوط به طرح نیروی کار، وزن‌های کرانگین کم‌تری در وزن‌های پایه‌ی طرح دیده می‌شوند و همچنین مشکل بی‌پاسخی واحد نیز در بسیاری از نقاط قابل توجه نبوده است، تفاوت بین روش‌های وزن‌دهی کالبدی را می‌توان در ابعاد کوچک‌تری مشاهده کرد.

برای مقایسه‌ی روش‌ها، علاوه بر آماره‌های توصیفی که برای وزن‌های کالبدی و عامل‌های تعدیل به کار رفته است، معیار اثر وزن‌دهی نابرابر (UWE) برای سنجش میزان فشاری که یک نظام وزن‌دهی به توزیع وزن‌های پایه‌ای اعمال کرده است، به کار می‌رود. علاوه بر این، اندازه‌ی UWE یک تقریب از میزان تورم واریانس حاصل از یک نظام وزن‌دهی نیز می‌باشد. این معیار از رابطه‌ی $UWE = 1 + \frac{n-1}{n} \{CV(w_k)\}^2$ محاسبه می‌شود که $CV(w_k)$ ضریب تغییرات وزن‌های w_k است (کالتون و فلورس سروانتس، ۲۰۰۳).

جدول ۳. انحراف استاندارد برای برآورد نرخ بیکاری و جمعیت بیکار در کل کشور

روش	نرخ بیکاری	جمعیت بیکار
رگرسیون خطی	۰٫۰۲۸۴۶	۳۲۰۴۹
چنگک‌زنی	۰٫۰۲۸۴۱	۳۲۰۴۵
چنگک‌زنی محدود شده	۰٫۰۲۸۳۸	۳۲۰۲۵
D&S	۰٫۰۲۸۴۰	۳۲۰۵۵

جدول ۴. مقدار UWE برای روش‌های وزن‌دهی

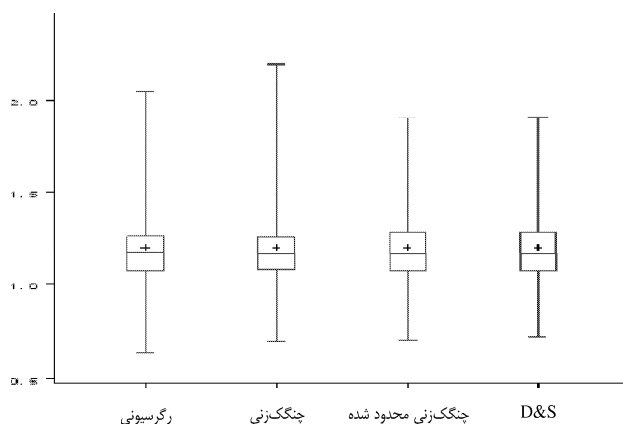
روش	UWE
رگرسیون خطی	۱٫۴۵۰۶
چنگک‌زنی	۱٫۴۵۰۱
چنگک‌زنی محدود شده	۱٫۴۵۲۴
D&S	۱٫۴۵۳۴

همان‌طور که در جدول ۳ مشاهده می‌شود، برآوردهای جمعیت بیکار و نرخ بیکاری حاصل از روش‌های محدود شده (چنگک‌زنی محدود شده و D&S) نسبت به سایر روش‌ها از لحاظ انحراف استاندارد مقدار کم‌تری دارند. در جدول ۴ ملاحظه می‌شود که روش چنگک‌زنی محدود شده به دلیل این که نسبت به روش D&S از ضریب تغییرات کم‌تری برخوردار است، مقدار UWE کم‌تری نیز دارد؛ به این مفهوم که مرزهای تعیین شده برای این روش، نسبت به روش D&S فشار کم‌تری به توزیع وزن‌های پایه‌ای وارد می‌کند. به عبارت دیگر، همان‌طور که از مقدار انحراف استاندارد نیز معلوم است، روش چنگک‌زنی محدود شده تورم کم‌تری در واریانس برآوردگر حاصل ایجاد می‌کند. توزیع عامل‌های تعدیل در چهار روش برآورد (شکل‌های ۱ و ۲) نیز مطلب بالا را تأیید می‌کند.

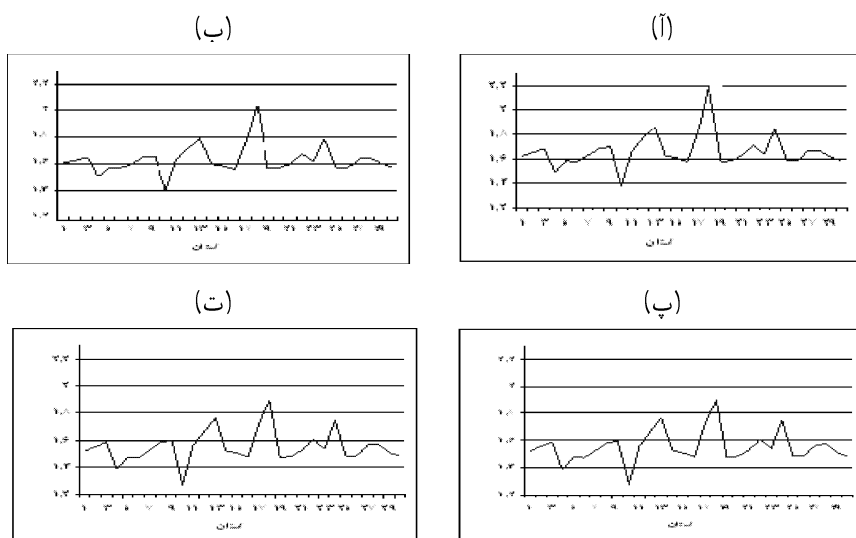
همان‌طور که از این نمودارها پیدا است، روش چنگک‌زنی نسبت به سایر روش‌ها دارای مقادیرهای کرانگین بیش‌تری است که موجب تورم بیش از حد برآورد واریانس این برآوردگر شده است.

به این ترتیب در طرح‌های آمارگیری که با مشکلاتی از قبیل بی‌پاسخی واحد اطلاعاتی، نقص چارچوب و احتمال‌های انتخاب نابرابر مواجه هستیم و نتایج حاصل تحت تأثیر این پدیده‌ها قرار می‌گیرند، برای جبران آثار ناشی از آن‌ها می‌توان از برآوردگرهای کالبدی‌ای که در این تحقیق مورد بررسی قرار گرفته‌اند، استفاده نمود. همان‌طور که در بالا اشاره شد، در بین روش‌های بررسی شده، روش‌های چنگک‌زنی محدود شده و D&S از دقت بالاتری برخوردارند. باید توجه داشت که تعیین مرزهای مورد نظر برای عامل‌های تعدیل وزنی، به گونه‌ای که همگرایی نیز حاصل شود، کار بسیار دشواری است. به این ترتیب، عوامل زمان و هزینه در انتخاب برآوردگر کالبدی‌ای مناسب بسیار با اهمیت است. به عبارت دیگر بهتر است از روش‌های

محدود شده در طرح‌هایی که به مقدار قابل توجهی با مشکلات مذکور مواجه هستند، استفاده شود؛ زیرا در موارد دیگر تفاوت خیلی زیادی در نتایج حاصل از برآوردهای کالبیده دیده نمی‌شود.



شکل ۱. نمودار شمایی عامل‌های تعدیل برای برآوردهای کالبیده



شکل ۲. مقدار حد اکثر عامل‌های تعدیل در استان‌ها به تفکیک برآوردهای کالبیده: (ا) چنگ‌زنی، (ب) رگرسیونی، (پ) D&S، (ت) رگرسیونی محدود شده.

مرجع‌ها

- Devill, J.C.; Särndal, C-E. (1992). Calibration Estimation in Survey Sampling. *JASA*, **87**, 376-382.
- Devill, J.C.; Särndal, C-E.; Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association* **88**, 1013-1020.
- Fan, M.; Woltman, H.; Miskura, S.; Thompson, J. (1980). Census Variance Estimation Procedure. Contributed paper presented at the 1981ASA meeting, August 1981.
- Kalton, G.; Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics* **19**, 81-97.
- Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, New York.
- Särndal, C-E.; Swensson, B.; Wretman, J.H. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

حمیدرضا نوابپور
گروه آمار، دانشکده‌ی اقتصاد،
دانشگاه علامه طباطبائی،
خیابان شهید بهشتی، نبش خیابان احمد قصیر،
تهران، ایران.
پایان‌نگار: h.navvabpour@srtc.ac.ir

آرمان بیداربختنیا
مرکز آمار ایران،
خیابان دکتر فاطمی، نبش رهی معیری،
تهران، ایران.
پیام‌نگار: bidar_a@yahoo.com