فصلنامه رویکردهای پژوهشی نوین در
مدیریت و حسابداری
سال پنجم، شماره ۸۳، زمستان ۱۴۰۰، ص ۱۳۲۴-۱۳۱۳

Journal of New Research Approaches
in Management and Accounting
ISSN: 2588-4573
http://www.majournal.ir

شیوه آسان برآورد صنعت-سال مدل رگرسیون

الهه زراعتی نوقابی [*1]
حبیب‌الله نخعی [2]

تاریخ دریافت: ۱۴۰۰/۱۰/۲۴ تاریخ چاپ: ۱۴۰۰/۱۱/۲۴

چکیده

همان گونه می‌دانید برآورد مدل رگرسیون به تفکیک هر صنعت و در هر سال کاری زمانبر است و به دقت بسیاری نیاز دارد چرا که می‌بایست به شمار حاصل ضرب تعداد صنعت در تعداد سال، مدل رگرسیون برای مشاهدات شرکت‌های مربوط به هر صنعت-سال برآورد شود و هر بار نتایج به نرم‌افزار اکسل منتقل شود و این مسئله شاید باعث شده که برخی پژوهشگران کمتر به تحقیقاتی که مستلزم برآورد صنعت-سال مدل رگرسیون است بپردازند. در این مقاله آموخته می‌شود که چگونه می‌توان با نوشتن دستورهایی شامل دو حلقه تودرتوی forvalues برآورد صنعت-سال مدل رگرسیون همراه با آزمون‌های فروض کلاسیک و همخطی را در زمانی بسیار کوتاه انجام داد. مثال‌های به کار رفته در این نوشتار برای نرم‌افزار استاتا انجام شده است که می‌توان همین روش را برای دیگر نرم‌افزارهای آماری نیز به کار برد امید است با به کار بردن این شیوه راه برای انجام پژوهش‌هایی که مستلزم برآورد صنعت-سال مدل رگرسیون‌اند هموارتر گردد.

واژگان کلیدی

رگرسیون، صنعت-سال، حلقه forvalues

[1] دانشجوی دکترای حسابداری، دانشگاه آزاد اسلامی واحد بیرجند، بیرجند، ایران. (*نویسنده مسئول: eligolzn@gmail.com)

[2] استادیار گروه حسابداری، دانشگاه آزاد اسلامی واحد بیرجند، بیرجند، ایران. (h.nakhaei48@yahoo.com)

۱. مقدمه

در پژوهش‌های حسابداری بسیار پیش می‌آید برای محاسبه مقادیر برخی از متغیرها می‌بایست مدل رگرسیون در سطح صنعت-سال برآورد شود. برای نمونه برآورد اقلام تعهدی اختیاری را در نظر بگیرید.

بنی مهد، عربی و حسن‌پور (۱۳۹۸) در کتاب پژوهش‌های تجربی و روش‌شناسی در حسابداری می‌گویند: برآورد مدل‌های مبتنی بر اقلام تعهدی اختیاری نیازمند صرف زمان کافی و دقت زیاد می‌باشد. همچنین در جای دیگر همین کتاب در باره مدل‌های برآوردکننده اقلام تعهدی اختیاری برای مدیریت سود اظهار می‌دارند: برای برآورد مدل باید داده‌های مربوط به متغیرها به صورت مقطعی (Cross Section) و به تفکیک هر صنعت و هر سال چیده شوند زیرا شکل مدیریت سود و نحوه اعمال آن در صنایع مختلف به دلیل نوع فعالیت کاملاً با یکدیگر متفاوت است. همچنین روند اعمال مدیریت سود از سالی به سال دیگر غیر یکسان است؛ بنابراین، به منظور افزایش قابلیت مقایسه داده‌ها و اندازه‌گیری صحیح مقادیر خطا باید داده‌های مربوط به هر صنعت در یک صفحه (Sheet) جداگانه از نرم‌افزار Excel چیده شوند و مدل مزبور به طور جداگانه در هر صنعت و هر سال برآورد و مقادیر خطا محاسبه شود. تخمین مدل در هر صنعت و هر سال به این معنی است که اگر در پژوهش خود ۱۲ صنعت طی دوره زمانی ۱۳۸۸ تا ۱۳۹۳ (۶ سال) داشته باشید، می‌بایست مدل مزبور را ۷۲ مرتبه (۱۲*۶) برآورد کنید.

بنی مهد و همکاران (۱۳۹۸) در جای دیگری در همان کتاب می‌گویند: مقادیر خطا در حالتی که مدل یک بار در بین کل صنایع برآورد شود و یا این که در هر صنعت یک بار برآورد شود با هم تفاوت خواهند داشت. محاسبه مقادیر خطا به تفکیک هر صنعت و در هر سال موجب کاهش ایجاد نقاط پَرت در مقادیر خطا می‌شود.

همان طور که در مثال بالا دیدید برای محاسبه مدیریت سود ناشی از دستکاری اقلام تعهدی اختیاری می‌بایست مدل رگرسیون را به شمار عدد حاصل از ضرب تعداد سال‌های پژوهش در تعداد صنایع برآورد کرد و سپس مقادیر خطای به دست آمده برای هر صنعت-سال را به صفحه اکسل مربوط به صنعت مورد نظر انتقال دهیم که انتقال درست مقادیر خطای به دست آمده نیاز به دقت و توجه بسیار دارد و گر نه ممکن است سهواً مقادیر در جای درست خود قرار نگیرند و بنابراین نتایج پژوهش معتبر نخواهد بود علاوه بر این انجام این همه بار برآورد مدل و انتقال درست مقادیر خطا به نرم‌افزار اکسل کاری زمانبر است به ویژه که در پژوهش‌هایی از این دست معمولاً تعداد برآورد مدل از مثال فوق هم بیشتر است.

از آنجا که انجام پژوهش‌هایی که نیاز به برآورد مدل برای هر سال و در سطح هر صنعت به طور جداگانه دارند مستلزم صرف زمان همراه با دقت فراوان است ممکن است برخی از پژوهشگران کمتر تمایل به انجام این گونه پژوهش‌ها داشته باشند. در حالی که به شیوه‌ای که گفته خواهد شد برآورد هر چند بار مدل در زمانی اندک و بدون خطای ناشی از انتقال مقادیر برآوردی میسر می‌شود بنابراین، این نوآوری باعث سهولت بسیار در انجام چنین پژوهش‌هایی می‌گردد که انتظار می‌رود این امر موجب گردد مقالات پژوهشی بیشتری در زمینه‌هایی که نیازمند برآورد مدل رگرسیون صنعت-سال است در آینده منتشر شود.

۲. برآورد صنعت-سال مدل رگرسیون با این شیوه

۲-۱. نکات مقدماتی

همان گونه که می‌دانید نخست باید پس از این که داده‌های مربوط به متغیرها در نرم‌افزار اکسل به صورت مقطعی و به تفکیک هر صنعت چیده شدند برای هر شرکت و نیز هر صنعت یک کد عددی داده شود تا نرم‌افزار آماری‌ای مانند استاتا بتواند با این کدهای عددی مقاطع و صنایع را از هم باز شناسد.

بنی مهد و همکاران (۱۳۹۸) در همان کتاب می‌گویند: به دلیل این که تخمین مدل بر روی داده‌های یک صنعت انجام می‌شود، لازم است تعداد شرکت‌های هر صنعت به میزان کافی باشد (حداقل ۱۵ شرکت). البته در تخمین مدل رگرسیون هر چه تعداد مشاهدات (شرکت‌ها) بیشتر باشد نتایج قابل اتکاتری حاصل می‌شود. روی چاوداری (۲۰۰۶) حداقل تعداد ۱۵ شرکت را در زمان محاسبه مقادیر خطا برای هر شرکت ضروری می‌داند. البته در مقالات فارسی تعداد شرکت‌های کمتری را به دلیل محدودیت در شرکت‌های نمونه در نظر می‌گیرند.

و باز در جای دیگری در همان کتاب اظهار می‌دارند: هر گاه لازم باشد مقادیر خطای مدل به عنوان یک شاخص یا متغیر حسابداری محاسبه شود می‌بایست داده‌ها به صورت مقطعی برای هر صنعت در هر سال چیده شوند. تعداد شرکت‌های هر صنعت برای این منظور به گفته روی چاوداری (۲۰۰۶) و کوتاری و همکاران (۲۰۰۵) نباید کمتر از ۱۵ شرکت باشد.

برای نمونه در مورد مثال بالا اگر حداقل ۱۵ شرکت در هر صنعت در نظر بگیریم که البته نیازی نیست تعداد شرکت‌های هر صنعت با هم برابر باشند این را برای ساده‌تر بودن مثال در نظر گرفته‌ایم بنابراین باید به شمار ۱۸۰ (۱۵*۱۲) کد عددی متمایز به شرکت‌های مختلف و ۱۲ کد عددی متمایز به صنایع گوناگون داده شود که برای تعیین متغیر مورد نظر مثلاً برای محاسبه مدیریت سود ناشی از دستکاری اقلام تعهدی اختیاری می‌بایست مدل رگرسیون را به شمار عدد حاصل از ضرب تعداد سال‌های پژوهش در تعداد صنایع که برای یک دوره ۶ ساله ۷۲ بار (۶*۱۲) برای یک دوره ۱۰ ساله ۱۲۰ بار (۱۰*۱۲) می‌شود را برآورد کرد که همان طور که گفته شد کاری است که زمان و دقت بسیاری را می‌طلبد.

پیش از پرداختن به شیوه پیشنهادی این مقاله خوب است این نکته گفته شود که نرم‌افزارهای آماری‌ای مانند استاتا می‌توانند با داده‌های رشته‌ای شامل کاراکترهای یونیکد[1] و اعداد کار کنند تنها لازم است در هنگام نوشتن دستورها آن‌ها را در میان دو علامت نقل قول دوتایی (")[2] نهاد تا نرم‌افزار آماری بتواند آن‌ها را به عنوان رشته بشناسد که از این توانایی نرم‌افزارهای آماری می‌توان تنها با نوشتن چند دستور ساده برای کددهی به برخی از متغیرهای پژوهش بدون صرف زمان زیاد و بی هیچ اشتباهی بهره برد فرض کنید در پژوهشی بخواهید برای متغیر دو ارزشی BIG برای شرکت‌هایی که نام حسابرس آن‌ها سازمان حسابرسی است عدد یک و غیر از سازمان حسابرسی عدد صفر را در نظر بگیرید راه معمول این است که در نرم‌افزار اکسل مثلاً ستونی با نام BIG کنار ستونی که عنوان آن نام حسابرس است باز کرده و اعداد ۱ یا ۰ را بسته به نام حسابرس در نرم‌افزار وارد کنید و در صورت مفقود بودن خالی بگذارید که برای مثال پیشین می‌بایست برای یک دوره ۶ ساله در ۱۰۸۰ خانه (۶*۱۲*۱۵) برای یک دوره ۱۰ ساله در ۱۸۰۰ خانه (۱۰*۱۲*۱۵) اعداد ۱ یا ۰ را در نرم‌افزار اکسل وارد کنید و یا خالی رها کنید اگر بگیریم که ویرایش هر خانه یک ثانیه زمان ببرد برای یک دوره ۶ ساله

[1] Unicode
[2] Double Quotation Mark

۱۸ دقیقه (۱۰۸۰/۶۰) و برای یک دوره ۱۰ ساله ۳۰ دقیقه (۱۸۰۰/۶۰) زمان خواهد برد که می‌توان به آسانی همین کار را با نوشتن سه دستور ساده مانند زیر در نرم‌افزار آماری‌ای مانند استاتا انجام داد:

gen BIG = 1 if نام حسابرس == "سازمان حسابرسی"

replace BIG = 0 if نام حسابرس =! "سازمان حسابرسی"

replace BIG = . if نام حسابرس == " "

معنی این سه دستور بالا این است که به نرم‌افزار آماری فرمان می‌دهد که متغیری به نام BIG تولید کند در صورتی که مقدار متغیر نام حسابرس برابر با رشته "سازمان حسابرسی" باشد به آن عدد یک و اگر رشته‌ای غیر از رشته "سازمان حسابرسی" باشد عدد صفر را نسبت دهد و اگر داده‌ای وجود ندارد آن را مفقود در نظر بگیرد.

۲-۲. برآورد مدل رگرسیون در سطح هر صنعت-سال

اکنون برای رهیافت به شیوه‌ای که در این مقاله عرضه خواهد شد بیایید ابتدا شیوه سنتی را بررسی کنیم برای انجام اولین برآورد باید داده‌های نخستین سال شرکت‌های صنعت یک را از نرم‌افزار اکسل را به نرم‌افزار آماری وارد کرده و نتایج به دست آمده را کپی کرده و به ستون مربوط در نرم‌افزار اکسل منتقل کنیم و همین کار را برای سال‌های دیگر صنعت یک و همچنین برای کل سال‌های صنعت دو و همین طور بقیه صنایع انجام دهیم که تعداد کل برآوردها برابر با شمار عدد حاصل از ضرب تعداد سال‌های پژوهش در تعداد صنایع خواهد بود.

حال بیایید در نظر بگیرید که چگونه می‌توان بدون نیاز به انتقال نتایج تک‌تک برآوردها به ستون مربوط در نرم‌افزار اکسل کل نتایج را در نرم‌افزار آماری‌ای مانند استاتا به دست آورد. همان گونه که می‌دانید در نرم‌افزار استاتا دو پنجره به نام‌های Data Editor (Edit) و Data Editor (Browse) مشابه صفحه نرم‌افزار اکسل ورودی به استاتا با همان داده‌ها هستند که این دو پنجره کاملاً مانند هم‌اند تنها تفاوتشان این است که پنجره نخست ویراستنی و پنجره دوم ناویراستنی است پنجره دوم برای این غیر قابل ویرایش در نظر گرفته شده است تا در هنگامی که نمی‌خواهیم ویرایشی انجام دهیم به پنجره Data Editor (Browse) مراجعه کنیم تا به طور ناخواسته موجب تغییر در داده‌ها نشویم هر گاه که متغیر جدیدی ایجاد می‌کنیم علاوه بر این که نام متغیر در پنجره Variables ظاهر می‌شود در پنجره‌های Data Editor (Edit) و Data Editor (Browse) نیز ستونی به همین نام همراه با داده‌های این متغیر ایجاد می‌شود پس اگر بتوان کاری کرد که نرم‌افزار آماری‌ای مانند استاتا به طور خودکار هر بار برآورد مدل رگرسیون را به ترتیب برای داده‌های قرارگرفته در بازه مورد نظر انجام دهد یعنی نخست برای صنعت یک و سال یکم و بعد برای صنعت یک و سال دوم و همین طور برای صنعت یک تا سال آخر انجام داده و به همین ترتیب مدل را برای دیگر صنایع در همه سال‌ها برآورد کند و سپس نتایج برآورد هر مرتبه را به متغیرهایی با نام‌های متمایز نسبت دهد در این صورت تعدادی ستون به شمار برآوردهای انجام‌گرفته خواهیم داشت که در هر ستون با نام متمایز به تعداد شرکت‌های صنعت مربوط داده تولید خواهد شد و برای دیگر خانه‌های آن ستون داده‌ای تولید نخواهد شد و مفقود خواهند بود؛ یعنی اگر صنعت یک دارای ۳۰ شرکت باشد ستون مربوط به سال نخست صنعت یک دارای ۳۰ داده در ۳۰ سطر مربوط به سال نخست صنعت یک خواهد بود و ستون مربوط به سال دوم دارای ۳۰ داده در ۳۰ سطر مربوط به سال دوم صنعت یک و به همین گونه ستون مربوط به سال آخر صنعت یک دارای ۳۰ داده در ۳۰ سطر مربوط به سال آخر صنعت یک خواهد داشت و همین طور اگر صنعت دو دارای ۲۵ شرکت باشد ستون مربوط به سال نخست صنعت دو دارای ۲۵ داده در ۲۵ سطر مربوط به سال نخست صنعت دو خواهد بود و ستون مربوط به سال دوم

دارای ۲۵ داده در ۲۵ سطر مربوط به سال دوم صنعت دو و به همین گونه ستون مربوط به سال آخر صنعت دو دارای ۲۵ داده در ۲۵ سطر مربوط به سال آخر صنعت دو خواهد داشت به همین ترتیب ستون‌های دیگر به تعداد شرکت‌های صنعت مرتبط دارای داده خواهند بود و خانه‌های دیگر همه ستون‌ها مفقود خواهند بود که در این خانه‌ها در پنجره‌های Data Editor (Edit) و Data Editor (Browse) یک نقطه (.)[3] به نشانه مفقودی نمایش داده خواهد شد اکنون اگر بتوان داده‌های مفقود هر ستون را به صفر تبدیل کرد و همه ستون‌ها را با هم جمع کرد داده‌های ستون حاصل برابر با داده‌های ستونی خواهد شد که با روش سنتی در نرم‌افزار اکسل تولید می‌شد که این کار به آسانی با دستورهایی ساده با کمک دو حلقه تو در توی forvalues در نرم‌افزار استاتا انجام پذیر است برای تفهیم بهتر این شیوه در زیر دستورهای لازم برای تعیین مقادیر خطای به دست آمده از برآورد مدل کوتاری، لئون و ویزلی[4] (۲۰۰۵) برای محاسبه مدیریت سود آورده شده است.

همان طور که می‌دانید یکی از رایج‌ترین معیارهای استفاده شده در ادبیات برای سنجش دستکاری‌های سود مقدار تعهدات اختیاری است که نماینده‌ای برای مدیریت سود است در این مثال برای اندازه‌گیری تعهدات اختیاری، از مدل معروف به عملکرد ارائه شده توسط کوتاری و همکاران (۲۰۰۵) استفاده شده است.

این مدل برگرفته از مدل تعدیل شده جونز[5] (۱۹۹۱) توسط دیچو، اسلون و سوینی[6] (۱۹۹۵) فرض می‌کند که تغییرات در درآمدهای فروش منهای تغییرات در حساب‌ها و اسناد دریافتنی یعنی تغییرات در درآمدهای فروش نقدی و نیز جمع دارایی‌های ثابت ناخالص مستقل از اختیارات مدیریتی است بنابراین مقادیر خطای مدل بیانگر اقلام تعهدی اختیاری/ مدیریت سود است متغیر بازده دارایی‌ها (ROA_{it}) به باور کوتاری و همکاران به دلیل عملکرد متفاوت شرکت‌ها با کنترل عملکرد مالی آن‌ها، کارایی مدل تعدیل شده جونز را افزایش می‌دهد.

$$\frac{TACC_{it}}{Assets_{it-1}} = \beta_0 + \beta_1 \frac{1}{Assets_{it-1}} + \beta_2 \frac{\Delta REV_{it} - \Delta REC_{it}}{Assets_{it-1}} + \beta_3 \frac{PPE_{it}}{Assets_{it-1}} + \beta_4 ROA_{it} + \varepsilon_{it}$$

$TACC_{it}$: کل اقلام تعهدی شرکت i در پایان سال t

ΔREV_{it}: تغییر در درآمدهای فروش شرکت i در سال t

ΔREC_{it}: تغییر در حساب‌ها و اسناد دریافتنی تجاری شرکت i در سال t

PPE_{it}: ناخالص اموال، ماشین‌آلات و تجهیزات شرکت i در پایان سال t

$Assets_{it-1}$: مجموع دارایی‌های شرکت i در پایان سال t-1

ROA_{it}: بازده دارایی‌های شرکت i در سال t

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$: ضرایب مدل کوتاری و همکاران

ε_{it}: خطای مدل که بیان کننده مقدار مدیریت سود است.

مدل بالا در سطح هر صنعت و برای هر سال جداگانه برآورد می‌شود

مقادیر منفی جمله خطای مدل بیانگر مدیریت سود از نوع حداقل‌سازی سود

---

[3] Full Stop
[4] Kothari, Leone and Wasley.
[5] Jones
[6] Dechow, Sloan & Sweeney

و مقادیر مثبت جمله خطای مدل بیانگر مدیریت سود از نوع حداکثرسازی سود

و قدرمطلق جمله خطای مدل بیانگر مدیریت سود بدون توجه به نوع آن است که در این مثال ملاک قرار گرفته است.

چنانچه تغییرات مقادیر خطای مدل در یک دوره زمانی در دو سوی خط رگرسیون تقریباً مشابه و توزیع آن‌ها نرمال باشد مدیریت سود در آن دوره از نوع هموارسازی سود است.

در زیر پس از ایجاد متغیرهای جدید برابر خط توضیحات یعنی اولین خط، دستورهای لازم برای برآورد ۱۲۰ بار مدل بالا یعنی برای هر صنعت-سال یک بار در نمونه‌ای شامل ۱۲ صنعت در یک دوره ۱۰ ساله از ۱۳۹۰ تا ۱۳۹۹ در نرم‌افزار استاتا آورده شده است:

```
//TACCA=TACC/Assets A=1/Assets ΔREV_ΔRECA=(ΔREV-ΔREC)/Assets PPEA=PPE/Assets

pctrim  TACCA A ΔREV_ΔRECA PPEA ROA if 1389<YEAR<1400 , p(1 99) gen(wi_)
recode(bound)

forvalues i=1/12{
forvalues j=1390/1399{
qui reg wi_TACCA wi_A wi_ΔREV_ΔRECA wi_PPEA wi_ROA if IND==`i' & YEAR==`j'
predict res`i'`j' if IND==`i' & YEAR==`j', resid
gen EM`i'`j'=res`i'`j' if res`i'`j'>0
replace EM`i'`j'=-res`i'`j' if res`i'`j'<0
replace EM`i'`j'=0 if res`i'`j'==.
}
}
gen EM=EM11390+EM11391+…….+EM11398+EM11399
+EM21390+EM21391+…….+EM21398+EM21399
.
.
+EM111390+EM111391+…….+EM111398+EM111399
+EM121390+EM121391+…….+EM121398+EM121399
```

نخستین دستور در بالا یعنی pctrim برای ویرایش مشاهدات پَرت[۷] به کار رفته است که به نرم‌افزار آماری‌ای مانند استاتا دستور می‌دهد داده‌های پرت مربوط به متغیرهای لیست شده در بازه زمانی ۱۳۹۹-۱۳۹۰ را که کوچک‌تر از صدک ۱ و بزرگ‌تر از صدک ۹۹ هستند را به ترتیب با صدک‌های ۱ و ۹۹ جایگزین کند و پیشوند دلخواه _wi را به نام متغیرهای فهرست شده بیفزاید که مشخص شود که داده‌های پرت این متغیرها ویرایش شده‌اند. البته درست آن است که ویرایش داده‌ها در سطح مقاطع، جداگانه انجام بگیرد که جلوتر در باره چگونگی انجام این کار به طور خودکار در درون حلقه‌ها گفته خواهد شد.

افلاطونی (۱۳۹۵) در کتاب تحلیل آماری در پژوهش‌های مالی و حسابداری با نرم‌افزار Stata می‌گوید: مشاهداتی که از نظر اندازه بسیار بزرگ‌تر یا بسیار کوچک‌تر از سایر مشاهدات هستند، مشاهدات پرت نامیده می‌شوند. وجود مشاهدات پرت تأثیر منفی شدیدی روی نتایج دارد. در بیشتر اوقات، رد فروض کلاسیک رگرسیون (به ویژه فرض نرمال بودن باقیمانده‌ها) و تورش‌دار بودن ضرایب، ناشی از وجود مشاهدات پرت است. برای شناسایی مشاهدات پرت، راه دقیقی وجود ندارد و از روش‌های تقریبی برای این منظور استفاده می‌شود. در پژوهش‌ها حوزه مالی و حسابداری، مشاهداتی که

[7] Outlier

کمتر از صدک اول و بزرگ‌تر از صدک ۹۹ هستند، به عنوان مشاهدات پرت در نظر گرفته می‌شوند. البته زمانی که میزان پراکندگی مشاهدات زیاد باشد، استفاده از صدک‌های ۵ و ۹۵ نیز توصیه می‌شود، به طور معمول دو رفتار متفاوت با مشاهدات پرت صورت می‌گیرد که عبارتند از (۱) حذف مشاهدات پرت[8] و (۲) ویرایش[9] مشاهدات پرت. در روش اول مشاهدات به طور کلی از مجموعه مشاهدات مربوط به یک متغیر حذف می‌شود. زمانی که تعداد مشاهدات کم باشد، حذف مشاهدات پرت، باعث کاهش درجه آزادی مدل می‌شود و از قابلیت اتکای آماره‌های آزمون می‌کاهد. به علاوه، برخی پژوهشگران عقیده دارند که مشاهدات پرت، حاوی اطلاعاتی هستند که با حذف آن‌ها، اطلاعات مذکور نیز نادیده گرفته می‌شوند. در روش دوم، مشاهدات پرت حذف می‌شود و به جای آن‌ها اعداد دیگری جایگزین می‌گردد. این اعداد می‌توانند (۱) میانگین مشاهدات، (۲) میانه مشاهدات و یا (۳) برای مشاهدات کمتر از صدک iام عدد مربوط به صدک مذکور و برای مشاهدات بزرگ‌تر از صدک jام عدد مربوط به صدک اخیر و (۴) هر عدد دلخواه دیگری باشد.

اکنون به دستور دوم می‌پردازیم که آغاز حلقه forvalues بیرونی است پس از نوشتن واژه کلیدی forvalues و تعیین بازه تغییرات متغیر i برای صنعت که از ۱ تا ۱۲ است بدنه این حلقه با نوشتن ابروی/آکلاد[10] چپ آغاز می‌شود.

در دستور سوم حلقه درونی آغاز شده است که متغیر این حلقه j برای سال در بازه ۱۳۹۹-۱۳۹۰ به کار رفته است که بدنه این حلقه نیز با نوشتن ابروی چپ آغاز می‌شود.

دستور چهارم در داخل حلقه درونی به نرم‌افزار دستور می‌دهد که رگرسیون را برای مدل کوتاری و همکاران به شرطی که صنعت i و سال j باشد برآورد کند.

دستور پنجم به نرم‌افزار فرمان می‌دهد که مقادیر خطای حاصل از برآورد مدل در دستور قبلی را استخراج کرده و به متغیری به نام resij نسبت دهد توجه کنید که نام این متغیر در هر چرخه متفاوت است چرا که اگر متغیری با نام یکسان از پیش وجود داشته باشد نرم‌افزار در هنگام ایجاد متغیر همنام با متغیر پیشین با نمایش پیغام خطا از کار باز خواهد ایستاد. دو علامتی که متغیرهای دو حلقه یعنی i و j میان آن‌ها قرار گرفته‌اند به نرم‌افزار نشان می‌دهد که آن‌ها متغیرهای حلقه‌ها هستند که برای نوشتن آن‌ها باید در صفحه کلید استاندارد آمریکایی، به ترتیب کلید بالای کلید Tab در حالت انگلیسی(`)[11] و کلید حرف گ در حالت انگلیسی(')[12] را فشرد.

در سه دستور بعدی به نرم‌افزار گفته می‌شود که داده‌های متغیر resij را با این شرایط که اگر داده مثبت است همان را و اگر منفی است منهای آن را و اگر مفقود است صفر را در نظر بگیرد و به متغیری به نام EMij نسبت دهد چرا که هدف ما در این مثال به دست آوردن قدر مطلق مقادیر خطای مدل برای مدیریت سود بدون توجه به نوع آن بود. در نظر گرفتن صفر برای داده‌های مفقود همان طور که پیش‌تر گفته شد برای این است که هنگام جمع کردن متغیرهای EMij تأثیری بر مقادیر داده‌های یکدیگر نداشته باشند.

در آخر در بیرون دو حلقه همه متغیرهای EMij را باهم جمع کرده و به متغیری به نام EM نسبت داده شده است. با توجه به توضیحات پیشین متغیر EM همان مدیریت سود بدون توجه به نوع آن است. برای جمع کردن ۱۲۰ متغیر EMij در این

[8] Trimming
[9] Winsorize
[10] Braces/Curly brackets/Squiggly brackets
[11] Grave accent
[12] Apostrophe

مثال می‌توان در پنجره Variables با گرفتن کلید کنترل و انتخاب این ۱۲۰ متغیر و کپی و پیست کردن آن‌ها در محیط برنامه‌نویسی نرم‌افزار یعنی پنجره Do-file Editor و نوشتن ۱۱۹ علامت جمع میان آن‌ها این عمل را انجام داد.

۳. ویرایش داده‌ها در سطح هر صنعت-سال با این شیوه

اکنون که بیشتر با شیوه پیشنهادی این مقاله آشنا شدید بیایید ببینیم چگونه می‌توان عمل ویرایش داده‌ها را در سطح مقاطع به طور خودکار در درون حلقه‌ها برای مثال بالا انجام داد؛ یعنی نخست ویرایش داده‌های سال نخست صنعت یک انجام گرفته و سپس برآورد مدل بر روی این داده‌های ویرایش شده در سطح شرکت‌های صنعت یک در سال نخست انجام بگیرد در مرحله دوم ویرایش داده‌های سال دوم صنعت یک انجام گرفته و سپس برآورد مدل بر روی این داده‌های ویرایش شده در سطح شرکت‌های صنعت یک در سال دوم انجام بگیرد و همین طور این کار در سطح مقاطع همه صنعت-سال‌ها صورت بگیرد. دستورهای مورد نیاز برای مثال قبلی در زیر آورده شده است:

```
forvalues i=1/12{
forvalues j=1390/1399{
pctrim  TACCA A ΔREV_ΔRECA PPEA ROA if IND==`i' & YEAR==`j' , p(1 99) gen(wi`i'`j'_)
recode(bound)
qui reg wi`i'`j'_TACCA wi`i'`j'_A wi`i'`j'_ΔREV_ΔRECA wi`i'`j'_PPEA wi`i'`j'_ROA
predict res`i'`j', resid
gen EM`i'`j'=res`i'`j' if res`i'`j'>0
replace EM`i'`j'=-res`i'`j' if res`i'`j'<0
replace EM`i'`j'=0 if res`i'`j'==.
}
}
gen EM=EM11390+EM11391+…….+EM11398+EM11399
+EM21390+EM21391+…….+EM21398+EM21399
.
.
+EM111390+EM111391+…….+EM111398+EM111399
+EM121390+EM121391+…….+EM121398+EM121399
```

همان گونه که می‌بینید در نخستین دستور در درون حلقه دوم ویرایش داده‌ها در سطح شرکت‌های هر صنعت-سال صورت گرفته و در دستور بعدی برآورد مدل رگرسیون در سطح شرکت‌های همان صنعت-سال انجام می‌شود. همان طور که پیش‌تر گفته شد اگر متغیری با نام یکسان از پیش وجود داشته باشد نرم‌افزار در هنگام ایجاد متغیر همنام با متغیر پیشین با نمایش پیغام خطا از کار باز خواهد ایستاد. زمانی که نرم‌افزار داده‌های پرت مربوط به متغیری را ویرایش می‌کند متغیر ویرایش شده‌ای تولید می‌کند که تنها در پیشوند با نام همان متغیر در حالت ویرایش نشده فرق می‌کند اگر ما مانند قبل می‌خواستیم عمل ویرایش را انجام دهیم نرم‌افزار پس از اولین چرخه با نمایش پیغام خطا از کار باز می‌ایستاد چرا که در چرخه دوم یعنی سال دوم از صنعت یک، متغیر ویرایش شده همان نام پیشین در چرخه اول یعنی سال اول از صنعت یک را می‌گرفت و تولید خطا می‌کرد که ما با نسبت دادن متغیرهای دو حلقه یعنی i و j به پیشوند یعنی wiij_ با تغییر پیشوند در هر چرخه از این خطا پیشگیری کرده‌ایم و عمل رگرسیون در هر بار تنها بر روی متغیرهایی با نامی متفاوت از چرخه‌های پیشین در سطح شرکت‌های هر صنعت-سال انجام می‌گیرد؛ یعنی متغیرهای مدل کوتاری و همکاران در چرخه نخست دارای پیشوند wi11390_ و در چرخه دوم دارای پیشوند wi11391_ و در چرخه پایانی دارای پیشوند wi121399_

می‌شوند و بنابراین همه متغیرهای مدل در هر چرخه دارای نامی متفاوت هستند و نرم‌افزار بی هیچ پیغام خطایی همه دستورها را اجرا خواهد کرد.

برای این که در عمل تفاوت نتایج در زمانی که ویرایش برای همه داده‌ها یعنی داده‌های ترکیبی یک بار در بیرون حلقه‌ها انجام می‌شود با زمانی که ویرایش داده‌ها به شمار حاصل ضرب تعداد صنایع در تعداد سال‌های پژوهش برای هر صنعت-سال یعنی داده‌های مقطعی انجام می‌گیرد سنجیده شود مدیریت سود را برابر دستورهای بالا بر روی داده‌های شرکت‌های بورس اوراق بهادار تهران انجام داده شد و دیده شد نتایج جز مواردی اندک با هم یکسان بودند. در هر صورت درست آن است عمل ویرایش داده‌ها در سطح شرکت‌های هر صنعت-سال انجام شود.

٤. رفع مشکل توقف نرم‌افزار به علت عدم کفایت داده در برخی از صنعت-سال‌ها

گاهی در عمل مواردی پیش می‌آید که نرم‌افزار به علت عدم کفایت داده برای برآورد مدل با نمایش پیغام خطا از پردازش باز می‌ایستد. برای این که دریابیم مشکل مربوط به کدام صنعت-سال است کافی است نگاهی به پنجره Data Editor (Browse) انداخته و ببینیم آخرین متغیر resij و یا EMij تولید شده توسط نرم‌افزار کدام است برای نمونه اگر دیده شود که آخرین متغیر EMij تولید شده توسط نرم‌افزار EM21393 است درمی‌یابیم که مشکل در صنعت ۲ و سال بعد یعنی ۱۳۹۴ است و یا با نگاه کردن به پیشوند آخرین متغیرهای ویرایشی تولیدی نرم‌افزار که در این مورد wi21394_ خواهد بود به همین نتیجه می‌رسیم. برای رفع این مشکل می‌توان پیش از حلقه‌های forvalues به متغیرهای مدل در این بازه مقادیر دلخواه داده و پس از اتمام کار نتیجه مربوط به صنعت ۲ و سال ۱۳۹۴ را مفقود در نظر بگیریم. یک راه ساده دادن مقادیر وقفه‌های مرتبه اول است چرا که اگر این مقادیر مشکل‌زا بودند چرخه پیشین انجام نمی‌شد که این کار برای صنعت ۵ و سال ۱۳۹۶ و نیز صنعت ۷ و سال ۱۳۹۱ با دادن مقادیر وقفه‌های مرتبه اول و در پایان مفقود در نظر گرفتن نتایج مربوط به این دو صنعت-سال با دستورهای زیر برای مثال پیشین انجام شده است:

```
replace TACCA=L.TACCA if IND==5 & YEAR==1396
replace TACCA=L.TACCA if IND==7 & YEAR==1391

replace A=L.A if IND==5 & YEAR==1396
replace A=L.A if IND==7 & YEAR==1391

replace ΔREV_ΔRECA=L.ΔREV_ΔRECA if IND==5 & YEAR==1396
replace ΔREV_ΔRECA=L.ΔREV_ΔRECA if IND==7 & YEAR==1391

replace PPEA=L.PPEA if IND==5 & YEAR==1396
replace PPEA=L.PPEA if IND==7 & YEAR==1391

replace ROA=L.ROA if IND==5 & YEAR==1396
replace ROA=L.ROA if IND==7 & YEAR==1391

forvalues i=1/12{
forvalues j=1390/1399{
pctrim  TACCA A ΔREV_ΔRECA PPEA ROA if IND==`i' & YEAR==`j' , p(1 99) gen(wi`i'`j'_)
recode(bound)
qui reg wi`i'`j'_TACCA wi`i'`j'_A wi`i'`j'_ΔREV_ΔRECA wi`i'`j'_PPEA wi`i'`j'_ROA
predict res`i'`j', resid
gen EM`i'`j'=res`i'`j' if res`i'`j'>0
```

```
replace EM`i'`j'=-res`i'`j' if res`i'`j'<0
replace EM`i'`j'=0 if res`i'`j'==.
}
}
gen EM=EM11390+EM11391+…….+EM11398+EM11399
+EM21390+EM21391+…….+EM21398+EM21399
.
.
+EM111390+EM111391+…….+EM111398+EM111399
+EM121390+EM121391+…….+EM121398+EM121399
replace EM=. if IND==5 & YEAR==1396
replace EM=. if IND==7 & YEAR==1391
```

۵. آزمون فروض کلاسیک و همخطی در سطح هر صنعت-سال با این شیوه

پیش از برآورد هر مدل رگرسیون می‌بایست از برقرار بودن فروض کلاسیک و نیز عدم وجود همخطی اطمینان حاصل نمود و در صورت عدم برقراری هر یک از فروض کلاسیک و یا وجود همخطی باید مشکل را رفع کرد تا بتوان به نتایج برآورد مدل، اتکا کرد. در مورد برآورد صنعت-سال که مدل باید چند ده بار مثلاً در مورد بالا ۱۲۰ مرتبه تخمین زده شود آزمون فروض کلاسیک رگرسیون و همخطی برای این تعداد برآورد مدل رگرسیون کاری بسیار سخت است. چاره مانند پیش این است که آزمون برقراری هر یک از فروض کلاسیک و عدم وجود همخطی را در درون حلقه‌ها انجام داد که در مورد برآورد صنعت-سال چون که رگرسیون هر بار بر روی داده‌های مقطعی در سطح شرکت‌های یک صنعت انجام می‌شود نیازی به انجام آزمون برقراری فرض سوم یعنی همبستگی خوشه‌ای و نیز فرض کلاسیک چهارم در صورت برقراری فرض یکم نیست در زیر دستورهای لازم برای انجام ۱۲۰ بار آزمون فروض کلاسیک یک و دو و پنج و همخطی برای مثال پیشین در نرم‌افزار استاتا آورده شده است:

```
forvalues i=1/12{
forvalues j=1390/1399{
pctrim  TACCA A ΔREV_ΔRECA PPEA ROA if IND==`i' & YEAR==`j' , p(1 99) gen(wi`i'`j'_)
recode(bound)
qui reg wi`i'`j'_TACCA wi`i'`j'_A wi`i'`j'_ΔREV_ΔRECA wi`i'`j'_PPEA wi`i'`j'_ROA
predict res`i'`j' , resid
ttest res`i'`j' ==0
estat imtest, white
swilk res`i'`j'
sfrancia res`i'`j'
vif
gen EM`i'`j'=res`i'`j' if res`i'`j'>0
replace EM`i'`j'=-res`i'`j' if res`i'`j'<0
replace EM`i'`j'=0 if res`i'`j'==.
}
}
gen EM=EM11390+EM11391+…….+EM11398+EM11399
+EM21390+EM21391+…….+EM21398+EM21399
.
.
+EM111390+EM111391+…….+EM111398+EM111399
+EM121390+EM121391+…….+EM121398+EM121399
```

در دستورهای بالا فرض کلاسیک یک با آزمون تی‌استیودنت سنجیده شده است که اگر میانگین خطاها اختلاف معناداری از صفر نداشته باشد فرض یک برقرار است و فرض کلاسیک دو با آزمون وایت سنجیده شده است که اگر آماره کای دو آزمون در سطح ۵ درصد معنادار باشد فرض دو برقرار نیست و فرض کلاسیک پنج با آزمون‌های شاپیرو- ویلک و شاپیرو- فرانسیا سنجش شده است که اگر آماره این دو آزمون در سطح ۵ درصد معنادار باشد فرض پنج برقرار نیست و سرانجام با دستور vif آماره‌های تورم واریانس و عکس آن تولرانس محاسبه‌شده است که اگر عامل تورم واریانس vif برابر ۱ باشد مشکل همخطی وجود ندارد و اگر کمتر از ۵ و یا حتی ۱۰ باشد مشکل همخطی قابل اغماض است.

پس از اجرای دستورهای بالا به دلیل ویرایش داده‌های پرت تنها ممکن است برخی فروض کلاسیک و یا عدم همخطی برای شمار اندکی از صنعت-سال‌ها برقرار نباشد که اگر بگیریم که بررسی نتایج آزمون فروض کلاسیک یک و دو و پنج و همخطی هر چرخه ۵ ثانیه زمان ببرد بررسی همه نتایج ۱۰ دقیقه (۵*۱۲۰) زمان خواهد برد و از آنجا که به دلیل انجام ویرایش داده‌های پرت در هر چرخه و در نتیجه تولید متغیرهای جدید، نرم‌افزار در پنجره Results نام متغیرهای جدید را اعلام می‌کند مثلاً متغیری به نام wi121391_TACCA در چرخه دارای مشکل مشخص کننده صنعت ۱۲ و سال ۱۳۹۱ است که می‌توان نتایج چرخه‌های دارای اشکال را پس از رفع مشکل جداگانه به دست آورده و با نتایج متناظر حاصل از دستورهای فوق جایگزین کرد.

٤. منابع و مآخذ

۱. افلاطونی، عباس. (۱۳۹۵). تحلیل آماری در پژوهش‌های مالی و حسابداری با نرم‌افزار Stata. انتشارات ترمه، چاپ اول، تهران.

۲. بنی مهد، بهمن؛ عربی، مهدی و حسن‌پور، شیوا. (۱۳۹۸). پژوهش‌های تجربی و روش‌شناسی در حسابداری. انتشارات ترمه، چاپ ششم، تهران.

3. Dechow, P., Sloan, R., & Sweeney, A. (1995). Detecting earnings management. The Accounting Review, 70: 193-225.

4. Jones, J., (1991). Earning management during import relief investigations. Journal of Accounting Research, 29, 193-228.

5. Kothari, S.P., Leone, A. J., & Wasley, C. E. (2005). Performance matched discretionary accrual measures. Journal of Accounting and Economics, 39 (1): 165-197.

6. Roychowdhury, S. (2006). Earning management through real activities manipulation. Journal of Accounting and Economics, 42, 335-370.

Easy Way to Estimate Industry-Year Regression Model

Elaheh Zeraati Noughabi [*1]
Habibollah Nakhaei [2]

Date of Receipt: 2022/01/12 Date of Issue: 2022/02/12

Abstract

As you know, estimating the regression model separately for each industry and in each year is time consuming and requires a lot of care, because to the number of product the number of industries and the number of years the regression model for the observations of companies related to each industry-year must be estimated and each time the results are transferred to Excel software, and this issue may have led some researchers to do less studies that require to estimate industry-year regression model. In this article is taught how can estimate the industry-year regression model with classical hypothesis and collinearity tests by writing statements including two nested forvalues loops in a very short time. The examples used in this paper have done for Stata software, which can be used for other statistical software. It is hoped that using this method will lead road to be flatter for doing studies that require to estimate industry-year regression model.

Keyword

Regression, Industry-year, Forvalues loop

1. PhD Student in Accounting, Islamic Azad University, Birjand Branch, Birjand, Iran (*Corresponding Author: eligolzn@gmail.com).
2. Assistant Professor, Department of Accounting, Islamic Azad University, Birjand Branch, Birjand, Iran (h.nakhaei48@yahoo.com).