



دانشگاه صنعتی امیرکبیر
دانشکده مهندسی کامپیوتر

پروژه درس یادگیری ماشین آماری مسأله‌ی دوم: پیش‌بینی تغییرات شاخص بورس ایران

دانشجو:

سید احمد نقوی نوزاد

استاد:

دکتر نیک آبادی

۱. شرح مسأله:

این سال‌ها با ناپایداری‌های برجسته‌ای در زمینه‌ی بازار مالی جهانی شناخته شده است و این مسئله موجب نگرانی فزاینده‌ای در مورد مدیریت خطر یا بحران گشته است. در ریاضیات مالی و مدیریت خطر مالی، معیار سرمایه در خطر (VaR) یک مقیاس خطر پراستفاده برای خطر ضرر در مورد یک دارائی خاص از دارائی‌های پولی به حساب می‌رود. در مورد یک دارائی خاص، احتمال و افق زمانی، VaR به عنوان یک حد آستانه تعریف می‌شود به طوری که احتمال آن که ضرر نسبت دادن یک ارزش به یک کالا بر اساس قیمت فعلی بازار برای آن کالای خاص (mark-to-market) با توجه به افق زمانی ارائه شده، از یک مقدار خاص (با فرض وجود بازارهای نرمال و عدم وجود تجارت در مورد این دارائی) تجاوز می‌نماید، همان حد احتمال ارائه شده است.

VaR می‌تواند به عنوان ابزاری برای توصیف خطر یک رخداد خاص که در تمامی زمینه‌های مدیریت خطر مانند بازار سرمایه، اعتبارات و نیز خطرات عملیاتی یا بیمه‌ای وجود دارد، به کار رود. شمار فزاینده‌ای از رویکردها برای تخمین معیار سرمایه در خطر (VaR) وجود دارند. پژوهشگران متعددی دسته‌بندی‌های گوناگونی از انواع این روش‌ها را ارائه نموده‌اند. با توجه به تحقیق پریگنان و اسمیت (2006)، ۷۳ درصد از بانک‌ها در میان ۶۰ بانک آمریکائی، کانادائی و بانک‌های بین‌المللی بزرگ در بین سال‌های 1996-2005 گزارش داده‌اند که روش مورد استفاده‌ی آن‌ها برای تخمین VaR، از نوع شبیه‌سازی تاریخی (Historical Simulation) بوده است. روش شبیه‌سازی مونتو کارلو (Monte Carlo Simulation) به عنوان دومین روش محبوب شناخته شده بوده است.

در این مسأله نیز هدف پیش‌بینی تغییرات شاخص بورس سهام در بازار بورس ایران است. مقصود نهائی در این پروژه پیش‌بینی دارائی در خطر یا همان VaR قیدشده در بالا می‌باشد که مقدار آن حد پایین لگاریتم بازگشت می‌باشد که در ستون سوم فایل داده‌ی پیوست شده لیست شده است. در این جا قصد ما این است که با اطمینان $1-\alpha$ درصد بیان کنیم که مقدار لگاریتم بازگشت در روز t ام بالاتر از VaR پیش‌بینی شده برای این روز است.

۲. مروری بر کارهای انجام شده (Literature Review) به صورت خلاصه

میورا و او (2001) یک دسته‌بندی کاربردی برای اندازه‌گیری VaR با توجه به ابعاد دوگانه‌ی فرضیات توزیعی و فرضیات وابستگی ارائه نمودند که می‌توان آن را در یک جدول سه در دو با

عناوین نرمال، غیر نرمال و غیر پارامتریک در طول یک بعد و i.i.d و وابستگی زمانی در طول بعد دیگر نشان داد. جدول زیر نشانگر دسته‌بندی مربوطه به صورت گفته شده می‌باشد:

جدول یک: دسته‌بندی VaR از نوع I

	Normal	Non-Normal	Non-Parametric
i.i.d	Variance-Covariance (VC) & Equally Weighted Moving Average (EQMA)		Historical Simulation
Time Dependence	Exponentially Weighted Moving Average (EXMA) & GARCH	Exponentially Weighted Moving Average (EXMA) & GARCH	

مانگانلی و انگل (2001) نتیجه گرفتند که سه دسته‌بندی عریض از رویکردهای تخمین VaR شامل رویکردهای پارامتریک، غیر پارامتریک و نیمه پارامتریک می‌باشند. RiskMetrics و Garch که می‌توانند تحت هر دو فرضیه‌ی نرمال و غیر نرمال به کار روند، رویکردهای پارامتریک هستند. شبیه‌سازی تاریخچه‌ای (Historical Simulation) یک روش غیر پارامتریک می‌باشد که تابع توزیع تجربی مربوط به مقادیر لگاریتم بازگشت پیشین را برای تولید VaR به کار می‌گیرد. رویکردهای نیمه پارامتریک شامل Extreme Value Theory، CAViaR و quasi-maximum likelihood می‌باشند. جدول ۲ در پایین بیانگر همین موضوع می‌باشد.

جدول دو: دسته‌بندی VaR از نوع II

Parametric Normal & Non-Normal	Nonparametric	Semiparametric
RiskMetrics GARCH	Historical Simulation	Extreme Value Theory, CAVia Quasi-maximum Likelihood

روش Variance-Covariance ساده‌ترین روش تخمین VaR می‌باشد که گاهی اوقات روش delta-normal نامیده می‌شود. این روش فرض می‌کند که افشائیات سرمایه خطی بوده و به دنبال آن معیارهای خطر به صورت پیوسته و نرمال توزیع شده‌اند. از آنجائی که بازگشت دارائی یک ترکیب خطی از متغیرهای نرمال می‌باشد، به خودی خود دارای توزیع نرمال می‌باشد. بنابراین ناپایداری سرمایه می‌تواند با استفاده از ماتریس کوواریانس و بردار وزن به سادگی محاسبه گردد. ارزش اصلی روش variance-covariance همان سادگی آن است، چرا که یک راه حل ساده را ارائه می‌نماید. هرچند که فرضیه‌ی خطی بودن به کار رفته در آن خود یک نقص به شمار می‌آید و به عبارتی قادر به اندازه‌گیری اثرات غیر خطی تعبیه شده در داده‌های ارائه شده نمی‌باشد. علاوه بر این، فرضیه‌ی توزیع نرمال آن نیز ممکن است خروجی‌های بینهایت ناخواسته را دست کم بگیرد.

شبیه‌سازی تاریخچه‌ای (Historical Simulation) در واقع فرایندی برای پیش‌بینی سرمایه در خطر به وسیله‌ی «شبیه‌سازی» یا ساختن تابع توزیع تجمعی مربوط به بازگشت‌های دارائی در طول زمان می‌باشد. این روش نیازی به هیچ فرضیه‌ی آماری‌ای مابرای لایتغیربودن توزیع مقادیر لگاریتم بازگشت، یا به طور ویژه بی‌ثباتی آن‌ها ندارد. خاصیت محدودسازی شبیه‌سازی تاریخچه‌ای (HS) به فرضیه‌ی i.i.d (مستقل و دارای توزیع یکسان) بودن مقادیر لگاریتم بازگشت آن مربوط است. با توجه به مشاهدات تجربی، شناخته شده که بازگشت‌های سرمایه به طور آشکارا از یکدیگر مستقل نیستند از آنجا که الگوهای خاصی را مانند خوشه‌بندی ناپایدار به نمایش می‌گذارند. شوربختانه شبیه‌سازی تاریخچه‌ای (HS) چنین الگوهائی را شامل نمی‌شود.

۳. شرح دقیق روش‌های پیاده‌سازی شده

در روش اول که همان روش شبیه‌سازی تاریخچه‌ای (HS) می‌باشد، ابتدا برای یک مقدار خطای خاص، متغیری به نام اندازه پنجره تعریف می‌کنیم که مشخص‌کننده‌ی تعداد مقادیر لگاریتم بازگشت روزهای قبل بوده و بنا بر آن است که مقدار لگاریتم بازگشت برای روز بعدی را با استفاده از این مقادیر تخمین بزنیم، بدین‌گونه که برای یک مقدار خطای خاص مثلاً α ، ابتدا مقادیر موجود در پنجره را مرتب کرده و سپس چارک α این مقادیر را به دست آورده و همین مقدار، تخمینی برای مقدار لگاریتم بازگشت در روز بعدی خواهد بود. حال باید بعد از هر بار تخمین این پنجره را یک مقدار به جلو حرکت داده و دوباره همین عملیات را تکرار نمائیم؛ لذا

برای تعداد دفعاتی که پنجره را حرکت می‌دهیم یک مقدار شمارنده تعریف کرده و با هر بار حرکت به جلوی پنجره مقدار این شمارنده را افزایش می‌دهیم. همین‌طور یک مقدار شمارنده دیگر را نیز برای تعداد دفعات موفقیت در تخمین صحیح را تعریف می‌نمائیم. حال اگر مقدار تخمینی برای روز t ام از مقدار واقعی آن با توجه به داده‌های مسئله کمتر باشد، مقدار شمارنده اول و دوم هر دو را افزایش داده و در نهایت بعد از آن که پنجره‌ی لغزان به انتهای داده‌های مسئله رسید مقدار خطای جدید در تخمین صحیح را از تفاضل تعداد موفقیت‌ها از تعداد دفعات جابجائی پنجره، تقسیم بر مقدار دوم به دست می‌آوریم. حال اگر مقدار تخمینی جدید برای خطا با مقدار اولیه‌ی در نظر گرفته‌شده برای آن تفاوت چندانی نداشته باشد، اندازه‌ی پنجره‌ی در نظر گرفته‌شده برای مسئله مناسب خواهد بود و به عبارتی باید با اندازه پنجره‌های مختلف مقادیر مختلفی برای نرخ خطا و یا برعکس نرخ موفقیت حاصل نمود و شباهت این مقادیر را با مقادیر مطلوب اولیه سنجید تا بالاخره به یک اندازه‌ی پنجره‌ی مطلوب دست پیدا کرد.

و اما در مورد روش دوم که با استفاده از فرضیه‌ی نرمال بودن توزیع داده‌های موجود در مسئله، آن را حل می‌کند باید گفت که اساس کار مشابه روش اول می‌باشد با این تفاوت که در اینجا برای داده‌های موجود در مسئله ابتدا مقادیر میانگین (μ) و واریانس (σ^2) را با استفاده از MLE محاسبه کرده و در نهایت با استفاده از فرمول زیر مقدار لگاریتم بازگشت برای روز t ام را تخمین می‌زنیم:

$$X = \mu + \sigma * \theta^{-1}(\alpha)$$

در اینجا نیز شمارنده‌های قیدشده برای روش اول موجود بوده و مقادیر آن‌ها تا آخر مسئله تغییر نموده و در پایان نیز مانند روش اول مقدار نرخ خطا و یا موفقیت را به دست آورده و با مقدار مطلوب مقایسه می‌نمائیم. اگر این مقادیر تخمینی و مطلوب با یکدیگر تفاوت چندانی نداشته باشند اندازه‌ی پنجره‌ی در نظر گرفته‌شده مطلوب می‌باشد و گرنه باید اندازه‌ی پنجره را همچنان تا رسیدن به یک تخمین مطلوب تغییر داد و مراحل بالا را دوباره تکرار نمود.

۴. ارائه‌ی نتایج به دست آمده و مقایسه‌ی آن‌ها

برای هر روش آزمایش‌های مختلفی انجام شده و نتایج متفاوتی حاصل شده‌اند. در اینجا فقط به قید نتایج مطلوب نهائی بسنده می‌نمائیم:

روش اول:

Moving Window Length	Shifting Counter	Success Counter	Desired Alpha (Error Rate)	Gained Alpha
201	1269	1204	.0500	.0505
201	1269	1248	.0100	.0158

روش دوم:

Moving Window Length	Shifting Counter	Success Counter	Desired Alpha (Error Rate)	Gained Alpha
350	1120	1060	.0500	.0527
350	1120	1098	.0100	.0188

۵. جمع‌بندی و نتیجه‌گیری

همان‌طور که قابل مشاهده است بعد از انجام آزمایش‌های متعدد و با در نظر گرفتن نرخ خطاهای متفاوت، مقادیر حاصله برای اندازه‌ی پنجره جهت کسب نرخ خطای مطلوب در دو روش متفاوت می‌باشد. در روش اول یا همان HS اندازه‌ی پنجره‌ی ۲۰۱ و در روش دوم یا همان فرض نرمال بودن داده‌های مسئله، اندازه‌ی پنجره‌ی ۳۵۰ ما را به جواب مطلوب هدایت می‌نماید.

۶. مراجع

- [1] Ryoza Miura and Shingo Oue (2000), Statistical Methodologies for Market Risk Measurement, Asia Pacific Financial Markets, 7, 305-319.
- [2] Daniel B. Nelson (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, Econometrica, 59, 347-370.
- [3] Tim Bollerslev (1986), Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics, 31, 307-327.
- [4] Roger B. Nelsen (1999), An Introduction to Copulas, Springer.