

به نام خدا

امیرمحمد کمیجانی ۹۹۵۲۲۰۳۲

توضیحات کد تمرین سری ۵

- تمام توضیحات سل ها آورده شده است و به ترتیب میباشد.

سل ۱ : کتابخانه های مورد نیاز ایمپورت شده اند

سل ۲ : متغیر های مورد نیاز تعریف شده اند.

ابتدا ۲۰ رمز ارز برتر را در متغیری تعریف کرده ایم (میتوانستیم ۱۰۰ رمز ارز را تعریف کنیم اما با همین تعداد به مطلوب مان میرسیم)

تاریخ شروع و پایان بازه مد نظر را تعریف نموده ایم.

متغیر `my_cryptos` بعد از این است که تغییرات مورد نظر خود را بر روی رمزارز های خود داده ایم و متغیری است که از آن بعدا استفاده خواهیم کرد.

متغیر `stationaries` برای ذخیره کردن دیتای سری زمانی هایی است که طبق شرایط گفته شده در صورت سوال یعنی با اطمینان بیشتر از ۹۵ درصد مانا هستند.

متغیر `results` ، متغیری است که نتایج تمام مراحل در آن ذخیره میشود

سل ۳ : در این سل برای `resample` کردن دیتای ۱ ساعته به ۴ ساعته استفاده کرده ایم.

همانطور که میدانیم، هنگامی که دیتا را از یاهو فایننس فچ میکنیم ؛ تایم فریم ۴ ساعته در آن وجود ندارد و دیتا را به صورت ۱ ساعته از یاهو فایننس دریافت کردیم سپس آنرا تبدیل به ۴ ساعته کردیم.

مراحل تبدیل آن به ۴ ساعته به این صورت است : قیمت `open` برابر است با قیمت اوپن اولین روز ؛ قیمت `close` برابر با قیمت کلوز آخرین روز؛ برای قیمت `high` بالاترین قیمت ها

رو در ۴ ساعت در نظر میگیریم و برای low پایین ترین قیمت در ۴ ساعت. مقدار حجم هم مجموع حجم در ۴ ساعت.

سل ۴ : اطلاعات رمزارز ها را به صورت یک ساعته دریافت میکنیم سپس از تابع Resample استفاده میکنیم و اطلاعات را مطابق توضیحات قسمت قبل به اطلاعات در تایم فریم ۴ ساعته بدست می آوریم. سپس قیمت close تمام رمز ارز ها به دست آورده و در متغیر my_cryptos ذخیره میکنیم.

توضیحات سل ۵ تا ۷ :

از تابع coint_johansen برای استفاده از تست یوهانسن استفاده میکنیم که نشان دهنده میزان cointegration در سری زمانی موجود میباشد. که ابتدا ترکیب دو یا چند سری زمانی را انجام میدهیم (با استفاده از دستور pd.concat) و سپس ولیو ترکیب ها را به تابع یوهانسن میدهیم. از مزایای تست یوهانسن این است که میتوانیم از آن در ترکیب بیشتر از دو سری زمانی استفاده کنیم در حالی که در مابقی تست ها چنین امکانی وجود ندارد. در اینجا دو راه استفاده کرده ایم :

(۱) در سل ۵ که کامنت شده است ؛ مقادیر critical value, tracestat را به دست آوردیم و سپس به مقایسه این دو پرداختیم.

به طور دقیق از این دو برای تعیین سطحی برای ریزکت کردن null hypothesis میشویم یا به طور ساده تر سطح اطمینان (confidence level) برای relationship cointegrating و به طور کلی cointegration میباشد. و سطوح اطمینان آن به سطوح ۱درصد و ۵درصد و ۱۰ درصد تقسیم میشود. اما از این مورد در کد استفاده نکردیم چون که این مورد مقدار دقیق p-value را به ما نمیداد.

(۲) روشی که روش اصلی ما است به این صورت است که ، بعد از اینکه از تست یوهانسن استفاده کردیم وزن ها را داریم و با استفاده از وزن ها و ترکیب دو سری زمانی قبلی سری زمانی جدید و ترکیب شده از آنها را میسازیم سپس سری زمانی ساخته شده را به آزمون ADF میدهیم تا مقدار دقیق p-value را بدست آوریم.

تابع ADFی که کد آنرا در سل ۶ میبینید به این صورت است که اگر شرایط سوال یعنی ۹۵ درصد مانایی رعایت شده بود مقدار True را برگرداند. در ادامه سل ۷ اگر سری زمانی ایجاد شده ۹۵ درصد مانایی داشت به همراه p-value و وزن ها آنرا برمیگردانیم.

سل ۸ : در این سل ترکیبات دوتایی از رمزارز ها میسازیم و الگوریتم مرحله قبل را روی آنها اجرا میکنیم و اگر ۱۰ ترکیب با شرایط خواسته شده داشتیم متوقف میکنیم.

*در اینجا فقط ترکیبات دوتایی را امتحان کردیم و چون تعداد ۱۰ ترکیب را در همین مرحله به دست آوردیم دیگر ترکیبات ۳تایی و ۴تایی را امتحان نکردیم.

سل ۹ : در این سل نتایجی که در قسمت اول خواسته شده ، نام و وزن و p-value را پرینت کردیم.

پارت ۲ : مقدار hurst exponent برای مشخص کردن وضعیت سری زمانی مان است که مشخص کنیم در کدام یک از وضعیت های random walk , mean reverting , trending قرار داریم.

سل ۱۰ : در این قسمت خواسته دوم سوال یعنی hurst exponent را با استفاده از تابع آماده compute_Hc بدست آوردیم.

سل ۱۱ : مقادیر constant , hurst exponent برای هر سری زمانی جدید را چاپ کردیم.

پارت ۳ :

سل ۱۲ : تابعی را برای محاسبه مقدار half-life قرار داده ایم ؛ برای محاسبه این تابع از Ornstein-Uhlenbeck استفاده شده است که در کد به صورت sm.OSL آورده شده است. محاسبات این قسمت همانند اسلاید ها میباشد:

formula for mean-reverting process:

$$dy_t = (\alpha + \delta y_{t-1})dt + d\varepsilon$$

10

Hurst Exponent and Variance Ratio Test ...

- The advantage of writing the equation in the differential form is that it allows for an **analytical solution for the expected value of y_t** .

$$E(y_t) = y_0 e^{\delta t} - \frac{\alpha}{\delta} (1 - e^{\delta t})$$

- Remembering that δ is **negative for a mean-reverting process**, this tells us that the expected value of the price decays exponentially to the value $-\frac{\alpha}{\delta}$ with the **half-life of decay equals to $-\frac{\log(2)}{\delta}$** .
- The coefficient δ is called the speed of mean reversion and **the half-life of the mean-reversion, $t_{1/2}$, is the average time it will take the process to get pulled half-way back to the mean.**

سل ۱۳ : در این سل مقادیر half-life برای هر ترکیب سری زمانی جدید را محاسبه نمودیم.

قسمت آخر:

سل ۱۴ : نمایش تمام اطلاعات از مرحله اول تا سوم در متغیر results

سل ۱۵ : قرار دادن اطلاعات در یک فایل csv