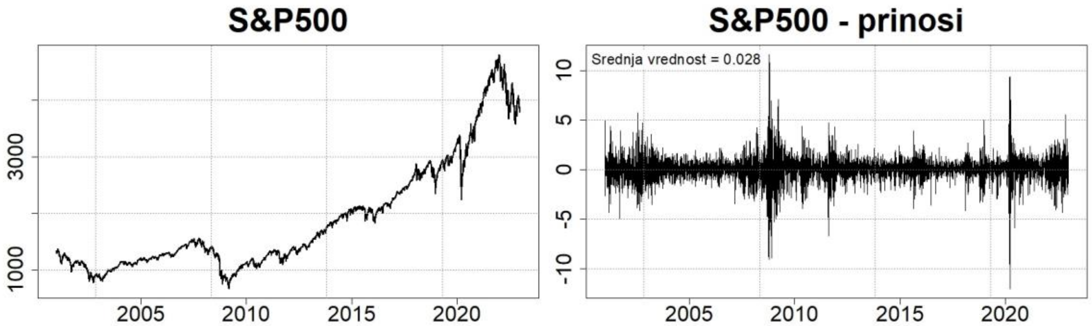
2. VARIJANSA, KOVARIJANSA I KORELACIJA

Da bismo sagledali određene karakteristike pojedinačnih HoV ili portfolija od više HoV, onda je bitno razumevanje određenih statističkih kategorija, kao što su varijansa, kovarijansa i korelacija. Kada se analiziraju prethodno spomenute statističke osobine vremenske serije, potrebno je naglasiti da se ne posmatraju njene cenovne vrednosti, nego prinosi, tj. dnevne stope prinosa koje predstavljaju stopu promene između dve uzastopne cenovne vrednosti. Stope prinosa se izražavaju u procentima. Drugim rečima, izraz preko koga se računaju stope prinosa je $r = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100$, gde je r stopa prinosa nekog instrumenta, a P_t i P_{t-1} su cenovne vrednosti nekog instrumeta u vremenu t i vremenu t-1. Na slici 2.1 je dat vizuelni prikaz empirijskog kretanja cene američkog indeksa S&P500 (P, leva slika) i njegovih izračunatih stopa prinosa (r, desna slika). Stope prinosa nam govore koliko su se cene menjale između uzastopnih dana, i možemo primetiti da su te promene bile izrazito velike za vreme svetske finansijske krize od 2008-2010, kao i u vreme početka pandemije COVID19 u 2020 godini. Sve vrednosti stopa promene prikazane na desnoj slici imaju pozitivne ili negativne vrednosti, a ako bi se one grupisale u vidu distribucije, onda bismo dobili tzv. diskretan raspored verovatnoće.

Slika 2.1. Grafički prikaz vrednosti američkog indeksa S&P500 i njegove stope prinosa

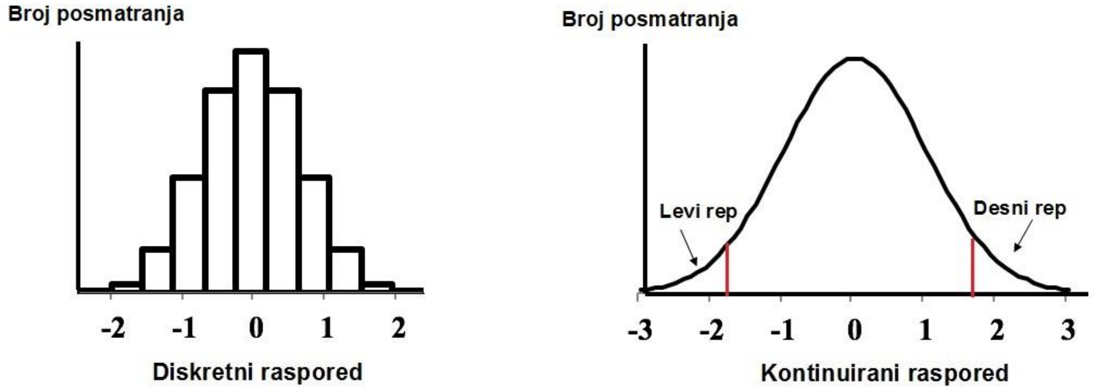


Obavešenje: Y osa prikazuje vrednosti cene ili stope prinosa, a X osa prikazuje godine. Uzorak je 22 godine dnevnih posmatranja, što znači ukupno nešto više od 5500 posmatranja.
Izvor: delo autora.

Šta je to distribucija? Distribucija je skup ili grupisanje svih empirijskih posmatranja, pri čemu svako empirijsko posmatranje ima određen broj ponavljanja. Što je broj ponavljanja veći, verovatnoća (ili broj) takvih vrednosti u ukupnom skupu (distribuciji) je veći, i obrnuto. Dnevne stope prinosa koje su umerene veličine se ponavljaju više puta i one su grupisane oko centralne tendencije skupa ili srednje vrednosti skupa (aritmetička sredina). Na desnoj slici, slike 2.1, srednja vrednost svih empirijskih posmatranja je 0,028. Najveći broj empirijskih posmatranja se kreće oko srednje vrednosti. Sa druge strane, ekstremne vrednosti skupa se javljaju retko (najčešće u kriznim periodima) i one su odaljene dalje od srednje vrednosti ili aritmetičke sredine skupa. Na desnom delu slike 2.1 se može videti da se u kriznim periodima, kao što su svetka ekonomska kriza i COVID19, dnevni prinosi kreću čak i preko ±10%, što je jako mnogo, ali broj tih ekstremnih dnevnih prinosa je jako mali.

U suštini se mogu razlikovati dva osnovna tipa distribucije (rasporeda) – diskretni i kontinuirani. Osnovna karakteristika diskretnog rasporeda je da je broj posmatranja konačan, odnosno da ima neki ceo broj. U našem primeru, taj broj bi bio nešto više od 5500 posmatranja. Nasuprot diskretnom rasporedu, kontinuirani raspored predstavlja oivičeno kretanje diskretnog rasporeda, a teorijski broj posmatranja ove distribucije je neograničen, zato što vrednosti posmatranja kontinuiranog rasporeda mogu biti beskonačno mali. Na slici 2.2 su prikazane slike diskretnog i kontinuiranog rasporeda i može se primetiti da je oblik kontinuiranog rasporeda u obliku zvona. To je najčešći oblik izgleda distribucije empirijskih uzastopnih posmatranja u prirodi. Teorijska distribucija u obliku zvona se naziva Gausova ili normalna distribucija. Na desnoj figuri slike 2.2 su prikazani tzv. levi i desni rep distribucije koji predstavljaju ekstremna posmatranja ili ekstremne vrednosti distribucije. Te ekstremne vrednosti smo videli na desnoj figuri, slike 2.1, u vreme svetske ekonomske krize i COVID19 pandemije, i kao što je rečeno, broj tih ekstremnih posmtranja je veoma mali, dok je broj posmatranja srednje vrednosati ili vrednosti centralne tendencije skupa najveći.

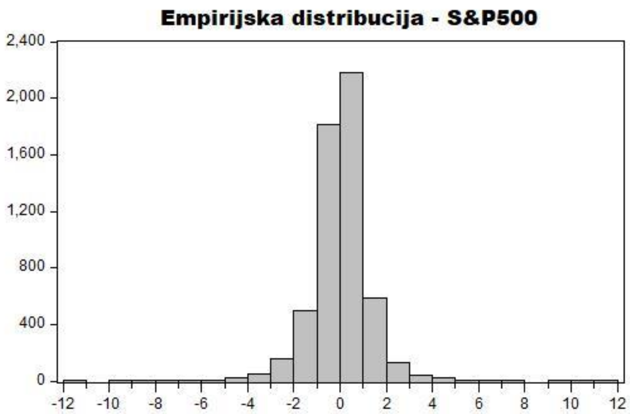
Slika 2.2. Diskretni i kontinuirani raspored stopa prinosa



Izvor: delo autora.

Međutim, normalni ili Gausov raspored u finansijama je pre izuzetak nego pravilo, odnosno finansijske vremenske serije najčešće ne prate normalni raspored, nego su distribucije često više izdužene na centru uz relativno velik broj ekstremnih posmatranja, što se onda naziva debeli rep. Na slici 2.3 je prikazana empirijska diskretna distribucija američkog indeksa sa slike 2.1.

Slika 2.3. Empirijska diskretna distribucija američkog indeksa S&P500



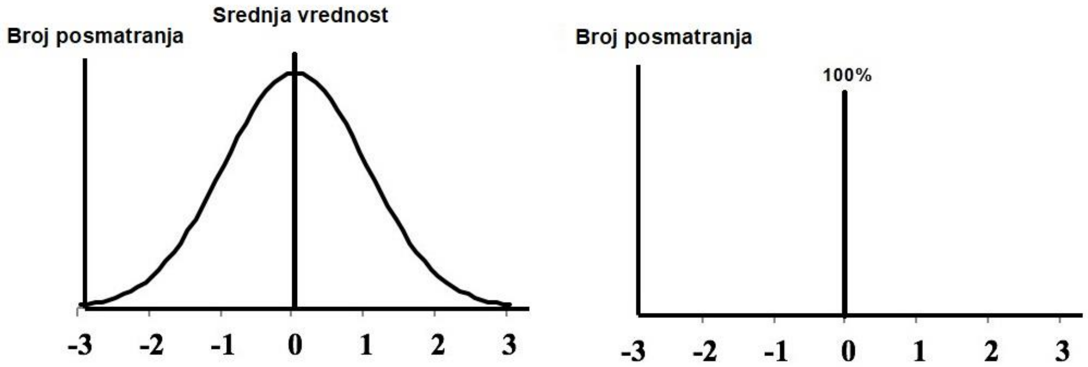
Izvor: delo autora.

Kao što se može videti na slici 2.3, empirijska distribucija indeksa S&P500 nema klasičan oblik zvona, nego je više izdužena na sredini, što znači da se najveći broj prinosa kreće oko nula. Takođe, moguće je primetiti postojanje vrednosti na repovima, koje su prilično udaljene od nule, što nam govori da su ekstremne vrednosti prisutne u empirijskoj distribuciji S&P500 indeksa, a njih vezujemo za već pomenute dve globalne krize.

Ovde se postavlja pitanje zašto je važno znati kakva je empirijska distribucija? Važno je zato što vizuelni prikaz distribucije može da nam ukaže kolika je prosečna očekivana stopa prinosa i rizik određene distribucije. Očekivana stopa prinosa se opisuje tzv. centralnom tendencijom skupa (leva figura na slici 2.4). Centralna tendencija skupa ili aritmetička sredina skupa nam pokazuje na kojoj vrednosti je lociran centar rasporeda verovatnoće oko koga su grupisane sve ostale vrednosti. U suštini, centralna tendencija skupa pokazuje koju vrednost najveći broj posmatranja u distribuciji ima. Ako posmatramo sa aspekta finansijskih vremenskih serija, onda nam centralna tendencija skupa govori koja stopa prinosa se u najvećem broju pojavljuje u distribuciji.

Druga bitna informacija koju možemo dobiti posmatrajući distribuciju je tzv. disperzija, tj. odstupanje od centralne tendencije skupa. Zašto je disperzija bitna? Disperzija je bitna zato što nam ona govori o riziku određene vremenske serije. Ovde se može postaviti pitanje, a šta je rizik? Rizik je generalno apstraktan pojam i ekonomisti su dugo vremena imali problema kako da ga kvantifikuju. Došlo se na ideju da je rizik u stvari odstupanje empirijskih posmatranja od centralne tendencije ili proseka. U tom smislu, što je to odstupanje veće, i rizik je veći, i obrnuto. Prema tome, što je više empirijskih vrednosti udaljeno dalje od proseka (centralne tendencije), gledano i sa leve i sa desne strane, to je rizik te HoV ili portfolija (kao skupa više HoV) veći. Ako sva posmatranja u empirijskoj distribuciji imaju istu vrednost, onda ne postoji odstupanje od proseka, pa distribucija onda izgleda kao vertikalna linija (desna figura na slici 2.4). U tom slučaju rizik ne postoji, jer će prinos uvek biti jednak proseku, tj. vrednosti centralne tendencije u kojem god vremenu izvršili ulaganje u tu HoV, jer su sva empirijska posmatranja ista.

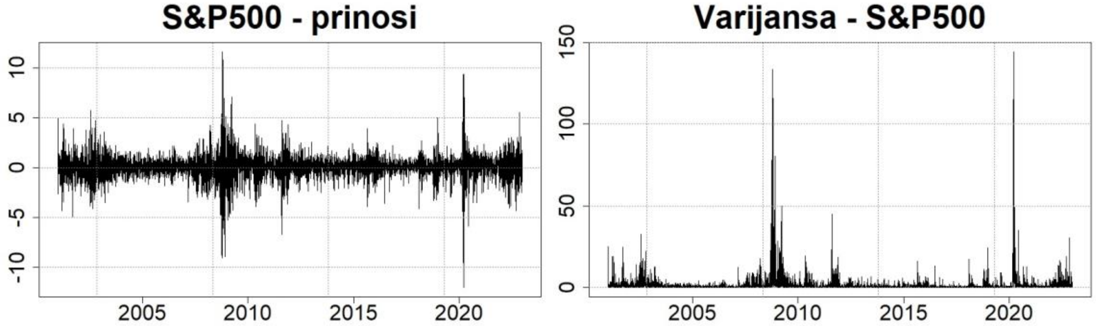
Slika 2.4. Izgled distribucije kad se empirijske vrednosti razlikuju i kad su iste



Izvor: delo autora.

Ako smo razumeli šta je rizik, onda može da se postavi pitanje kako ga izmeriti? Pošto smo rekli da je rizik u suštini odstupanje od centralne tendencije, odnosno proseka, onda je potrebno izračunati koliko je odstupanje svih empirijskih posmatranja od proseka i izračunati prosek tih odstupanja. Međutim, ovde imamo problem ako bismo samo oduzimali empirijske vrednosti od proseka, jer jednu distribuciju čine i pozitivna i negativna odstupanja od proseka. Drugim rečima, to bi značilo da se pozitivna i negativna odstupanja potiru kada bismo hteli da ih saberemo i izračunamo prosek odstupanja, pa bi rizik onda bio 0 ili blizu 0, što nije tačno. Da bi potiranje pozitivnih i negativnih odstupanja bilo izbegnuto, onda se svako odstupanje od proseka kvadrira kako bi minus ispred negativnih odstupanja bio eliminisan. Posle kvadriranja postoje samo kvadrati odstupanja od proseka i oni sada mogu da se saberu i podele sa brojem posmatranja, tj. da se izračuna prosek. Posle ovog postupka, dobijamo prosek kvadrata odstupanja, što se u finansijama, naziva **varijansa**. Što je varijansa veća i rizik je veći, i obrnuto. Varijansa se obeležava malim grčkim slovom sigma na kvadrat (σ^2), a računa se kao prosečna vrednost kvadrata odstupanja svih empirijskih posmatranja, tj. $\sigma^2 = \frac{\sum_{i=1}^{n}(R_i - E(R_i))^2}{n}$. R_i prikazuje empirijska posmatranja, čiji je broj od 1 do n, $E(R_i)$, je očekivano empirijsko posmatranje, odnosno prosek, a n je broj posmatranja u empirijskoj distribuciji. Pošto je varijansa kvadrat odstupanja od proseka, ona ne može biti nikada negativna. Na slici 2.5 su prikazani empirijski prinosi S&P500 indeksa i kreirana varijansa iz tih prinosa. Može se primetiti da kada su stope prinosa velike, onda je varijansa još veća, jer ona predstavlja kvadrirano odstupanje od proseka. Očigledni su veliki skokovi varijanse za vreme svetske finansijske krize i COVID19 krize, i ona ne može biti negativna, kao što je rečeno, jer je ona kvadrat odstupanja.

Slika 2.5. Prinosi i varijansa indeksa S&P500



Izvor: delo autora.

Budući da varijansa predstavlja kvadrat odstupanja, a nama treba samo ostupanje od proseka, onda se do njega dolazi jednostavnim korenovanjem varijanse, i ta veličina se zove **standardna devijacija**, sa oznakom sigma (σ). Prema tome standardna devijacija se računa kao koren iz varijanse ($\sigma = \sqrt{\sigma^2}$). Zadatak 1 daje numeričku ilustraciju računanja varijanse i standardne devijacije.

ZADATAK 1:

U koju HoV biste uložili novac ako biste se odlučivali prema nivou rizika uz date sledeće pretpostavke i ako je verovatnoća svakog događaja ista:

	HOV - A	HOV - B
Stanje ekonomije	Prinos u %	Prinos u %
Recesija	5	6
Normalno stanje	10	9
Ekspanzija	12	15

Prosečan prinos A = $\frac{5+10+12}{3} = 9$

Prosečan prinos B = $\frac{6+9+15}{3} = 10$

Odstupanje od prosečnog prinosa A	Odstupanje od prosečnog prinosa B	Kvadrat odstupanja prinosa A	Kvadrat odstupanja prinosa B
-4	-4	16	16
1	-1	1	1
3	5	9	25
		$\sum = 26$	$\sum = 42$

Varijansa za instrumente A i B je:

σ^2 (A) = 26/3 = 8.66666
σ^2 (B) = 42/3 = 14

Standardna devijacija za instrumente A i B je:

σ (A) =$\sqrt{8{,}66666}$= 2,94
σ (B) =$\sqrt{14}$= 3,74

Odgovor: Odlučili bismo se za opciju A zato što ima manju standardnu devijaciju, odnosno manje je odstupanje empirijskih prinosa od očekivanog prinosa, tj. rizik je manji.

Sledeća vrednost koju je važno objasniti je **kovarijansa**, koja se obeležava simbolom $COV(A, B)$. Kovarijansa u oznaci sadrži slova A i B, što upućuje na dva instrumenta, odnosno HoV_A i HoV_B. Drugim rečina, kovarijansa izračunava međusobni odnos između dva instrumenta, odnosno govori nam u kojim smerovima se kreću stope prinosa dva instrumenta: da li u istom smeru ili u različitim smerovima. Kad kažemo u istom smeru, to znači da stope prinosa oba instrumenta rastu ili opadaju, odnosno oba instrumenta beleže ili pozitivne ili negativne prinose. Tada je kovarijansa pozitiva. Suprotno, ako se prinosi dva instrumenta kreću u suprotnim smerovima, onda jedan instrument beleži pozitivne prinose, a drugi negativne, i obrnuto. Tada je kovarijansa negativna. Matematički, kovarijansa se računa kao proizvod odstupanja dva instrumenta od njihovih očekivanih ili prosečnih vrednosti. Kovarijansa može da se računa kao u jednačini (2.1) ako je poznata verovatnoća (P_i) svakog ishoda instrumenta A i B.

$$COV(R_{A,}R_B) = \sum_{i=1}^{n}[(R_A - E(R_A) \times (R_B - E(R_B) \times P_i] \qquad (2.1)$$

Ako su verovatnoće ishoda jednake za sva posmatranja, onda se kovarijansa jednostavno računa kao aritmetička sredina ili prosek proizvoda odstupanja dva instrumenta, kao u jednačini (2.2).

$$COV(R_{A,}R_B) = \frac{\sum_{i=1}^{n}[(R_A - E(R_A) \times (R_B - E(R_B)]}{n} \qquad (2.2)$$

Zašto je važno poznavati kovarijansu? Zato što nam ona govori o međusobnom odnosu prinosa dva instrumenta, što je jako bitna informacija prilikom sastavljanja portfolija, odnosno diverzifikacije. Diverzifikacija je postupak smanjenja rizika portfolija ulaganjem u dva ili više instrumenata čiji prinosi se kreću nesinhronizovano. Na primer, kada prinos jednog instrumenta raste, drugog instrumenta npr. raste, ali po nižoj stopi, trećeg opada po višoj stopi, četvrtog opada po nižoj stopi, itd., i onda bi se za takav portfolio moglo reći da je diverzifikovan. Na ovaj način, ako prinos jednog instrumenta raste, a drugog opada, onda u proseku prinos portfolija ostaje nepromenjen, pa je i rizik portfolija tada smanjen ili ga uopšte nema. Sa druge strane, ako bi oba instrumenta rasla ili opadala, onda bi se pozitivan ili negativan prinos portfolija duplirao, a samim tim i udaljenost od srednje vrednosti prinosa portfolija, što znači da bi se i rizik tada duplirao. Zato je najbolje sa aspekta smanjenja rizika portfolija da se kombinuju instrumenti koji imaju usklađen, odnosno istovremen rast i pad, tj. kad prinos jednog instrumenta raste, drugog opada, i obrnuto.

U praksi je poznato da npr. kada indeks akcija beleži rast onda cena zlata beleži pad jer u normalnim okolnostima investitori uglavnom ulažu u indeks koji ima tendenciju rasta, dok se investiranje u zlato tada zanemaruju. Ovo se dešava u periodima ekonomskog prosperiteta. Suprotno, u periodima recesije i krize, investitori napuštaju ulaganje u akcije i indekse akcija, jer indeksi beleže pad zbog lošeg stanja u ekonomiji i okreću se ulaganju u zlato, jer zlato ima i uvek će imati realnu vrednost, jer je potrebno mnogo rada da se napravi jedna unca (31,103 grama) zlata . Zbog činjenice da se u krizi svi okreću ulaganju u zlato, cena zlata tada raste jer mu raste tražnja. Prema tome, prinosi indeksa akcija i zlata se uglavnom kreću u suprotnim smerovima, što znači da imaju negativnu kovarijansu, a to je dobra karakteristika ova dva instrumenta da se nađu u istom portfoliju.

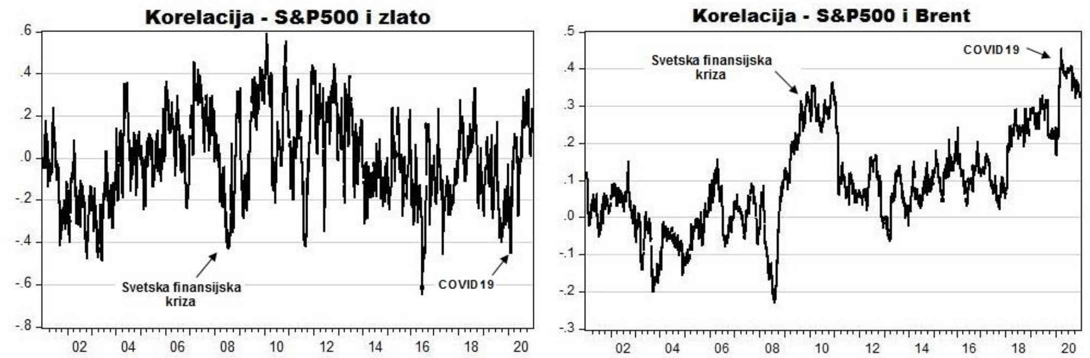
Sa druge strane, relativno usklađeno kretanje imaju indeks akcija i cena nafte. Drugim rečima, vrednost indeksa akcija i cena nafte često imaju upravo srazmeran odnos, odnosno kad vrednost indeksa raste, raste i cena nafte, i obrnuto. To je dešava iz sledećih razloga. Ako vrednost indeksa u svetu raste, to znači da globalnoj ekonomiji ide dobro, odnosno da i ona raste, a to onda znači i povećanu potrošnju energenata (u najvećem procentu nafte), što utiče na rast tražnje za naftom i porast njene cene. Sa druge strane, ako je globalna ekonomija u krizi, onda je i potrošnja nafte na niskom nivou, odnosno tražnja za naftom je mala. Zbog male globalne potražnje za naftom, njena cena pada. Prema tome, iz razloga što indeks akcija i cena nafte često imaju upravo srazmeran odnos, kombinacija ova dva instrumenta u istom portfoliju nije dobra ideja, jer ne doprinosi diverzifikaciji portfolija.

Iako je kovarijansa koristan pokazatelj sa aspekta sagledavanja međusobnog odnosa dva instrumenta, ona nije praktična za upotrebu, jer često ima vrlo visoke pozitivne ili negativne vrednosti. Stoga, mnogo bolji pokazatelj je **korelacija**, a kovarijansa je međukorak kod izračunavanja korelacije. Korelacija se obeležava grčkim slovom ro (ρ). Kao i kovarijansa, korelacija nam takođe govori o međusobnom odnosu dva instrumenta, stim što korelacija nema neograničene vrednosti kao kovarijansa, nego je korelacija ograničena u intervalu od -1 do 1, i ona ne može imati vrednosti manje od -1 ili veće od 1. Ako je korelacija -1 ili -100% to znači da dva instrumenta imaju savršenu negativnu korelaciju, tj. kad jedan instrument raste, u istom iznosu drugi pada, i obrnuto. Kada je korelacija 1 ili 100% to znači da dva instrumenta imaju savršenu pozitivnu korelaciju, odnosno kad jedan instrument raste (pada), onda drugi u istom iznosu raste (pada). Kada je korelacija 0, dva instrumenta su savršeno nekorelirana, tj. njihova kretanja uopšte nisu povezana. Ako je korelacija do ±0.2 onda se kaže da je korelacija između dva instrumenta mala, ako je u intervalu od ±0.2 do ±0.6 onda je korelacija umereno visoka, a ako je preko ±0.6 onda se kaže da je korelacija visoka ili jaka. Za korelaciju se kaže da je standardizovana kovarijansa, odnosno kovarijansa koja je podeljena proizvodom standardnih devijacija dva instrumenta. Prema tome, korelacija se računa kao u izrazu (2.3).

$$\rho_{A,B} = \frac{COV(R_A, R_B)}{\sigma_A \cdot \sigma_B} \tag{2.3}$$

Na slici 2.6 su prikazane dinamičke korelacije (po danima) između američkog indeksa S&P500 i zlata, i S&P500 i Brent nafte. Kao što se može videti, obe dinamičke korelacije se kreću u intervalu -1 do +1.

Slika 2.6. Dinamička korelacija između parova S&P500-zlato i S&P500-nafta



Izvor: delo autora.

Na slikama je posebno označen period svetske finansijske krize (2008-2010), pa se može uočiti da je dinamička korelacija između S&P500 i zlata bila izrazito negativna u 2008. (oko -0.4), kada je kriza eruptirala, što nam govori da su investitori na početku krize, zbog velikih nepoznanica koliko će kriza trajati i koliko će biti duboka, napuštali indeks S&P500 i ulagali u zlato. To je izazvalo pad vrednosti indeksa i rast cene zlata, a visoka negativna korelacija od -40% nam govori koliko je taj obrnuto srazmeran odnos bio jak. Slično se desilo i početkom 2020 kada je izbila COVID19 pandemija.

Suprotan odnos se može videti između S&P500 indeksa i nafte, odnosno kada je kriza uzela maha i kada se videlo koliko je duboka u 2009. cena nafte je drastično pala zbog značajnog smanjenja globalne tražnje za njom. Pad cene nafte je bio praćen padom indeksa S&P500, jer je globalna ekonomija ušla u recesiju, što je rezultovalo vrlo visokom pozitivnom korelacijom nešto manje od 40%. Između indeksa i cene nafte još veća pozitivna korelacija je zabeležena na početku COVID19 krize, dostignuvši vrednost veću od 40%.

Zadatak 2 daje praktičan primer računanja korelacije.

ZADATAK 2:

Posmatramo prinose dve kompanije A i B, u situaciji tri moguća ishoda kretanja privrede, kako sledi:

	Verovatnoća P_i	Prinos kompanije A (R_{Ai})	Prinos kompanije B (R_{Bi})
Ekspanzija	0,3	20	3
Normalan rast	0,4	10	35
Recesija	0,3	0	-5

Izračunajte koeficijent korelacije između dve akcije i protumačite ga.

Rešenje:

Kovarijansa koja nam je potrebna za računanje korelacije se računa preko formule:
$COV(R_{A,}R_B) = \sum_{i=1}^{n}[(R_A - E(R_A) \times (R_B - E(R_B) \times P_i]$

Za lakši račun možemo koristiti tabelu u kojoj se računaju elementi potrebni za izračunavanje kovarijanse i koeficijenta korelacije:

Verovatnoća P_i	Prinos R_A	Prinos R_B	Odstupanje od očekivanog prinosa R_A-$E(R_A)$	Odstupanje od očekivanog prinosa R_B-$E(R_B)$	Kvadrat odstupanja od očekivanog prinosa $(R_A-E(R_A))^2$	Kvadrat odstupanja od očekivanog prinosa $(R_B-E(R_B))^2$	$COV(R_A,R_B)$
0,3	20	3	10	-10,4	100	108,16	-31,2
0,4	10	35	0	21,6	0	466,56	0
0,3	0	-5	-10	-18,4	100	338,56	55,2
Prosek					60	320,64	24

Napomena: U poslednjem redu ove tabele je prikazan prosek, za razliku od tabele u prethodnom zadatku u kojoj je prikazan zbir (suma). To je zato što su u ovoj tabeli vrednosti ponderisane, tj. množene sa verovatnoćom, pa se onda automatski računa prosek.

Očekivani prinos akcije A i B:

$$E(R_A) = P_{Ai} \times R_{Ai} = 0{,}3 \times 20 + 0{,}4 \times 10 + 0{,}3 \times 0 = 10$$
$$E(R_B) = P_{Bi} \times R_{Bi} = 0{,}3 \times 3 + 0{,}4 \times 35 + 0{,}3 \times (-5) = 13{,}4$$

Kovarijansa je:
$COV(R_A,R_B)$= [0,3 × 10 × (-10,4)] + [0,4 × 0 × 21,6] + [0,3 × (-10) ×(-18,4)] = -31,2 + 0 + 55,2 = 24

Varijanse su:
σ_A^2 = 100 × 0,3 + 0 × 0,4 + 100 × 0,3 = 60
σ_B^2 = 108,16 × 0,3 + 466,56 × 0,4 + 338,56 × 0,3 = 32,45 + 186,62 + 101,57 = 320,64

Standardne devijacije instrumenata A i B su:
$\sigma_A = \sqrt{\sigma_A^2}$ =7,75

$\sigma_B = \sqrt{\sigma_B^2} = 17{,}91$

Koeficijent korelacije je onda:

$$\rho_{A,B} = \frac{COV(R_A, R_B)}{\sigma_A \cdot \sigma_B} = \frac{24}{7{,}75 \cdot 17{,}91} = 0{,}1729$$

Odgovor: Akcije dve kompanije su slabo korelisane, što znači da ih je poželjno imati u portfoliju u cilju diverzifikacije nesistemskog rizika, jer rast vrednosti jedne akcije se ne odražava značajnije na rast vrednosti druge akcije, i obrnuto.