



Specyfikacja Testów

Poniższe opracowanie zawiera zestawienie kluczowych elementów procesu testowania strategii inwestycyjnych. Choć nie są one bezpośrednio związane ze specyfiką samej strategii, mają istotny wpływ na wyniki końcowe:

- **Portfel instrumentów** – jeśli nie zaznaczono inaczej, strategię są testowane na portfelu kontraktów futures obejmującym różne klasy aktywów. Instrumenty te są notowane na światowych giełdach, głównie w USA. Niektóre strategię mogą być testowane na rynku „**Equity**”, co będzie wyraźnie zaznaczone.

Kontrakty futures:

- Corn (C)
- Cocoa (CC)
- EUA Emissions (CFI)
- Crude Oil (CL)
- Cotton (CT)
- U.S. Dollar Index (DX)
- Euro German Bund (EBL)
- E-mini S&P 500 (ES)
- DAX (FDX)
- Long Gilt (FLG)
- Gold (GC)
- Copper (HG)
- Hang Seng (HSI)
- Coffee (KC)
- Brent Crude (LCO)
- Low Sulphur Gasoil (LGO)
- Henry Hub Natural Gas (NG)
- E-mini Nasdaq 100 (NQ)
- Nikkei225 (NIY)
- Frozen Concentrated Orange Juice (OJ)
- Palladium (PA)
- Soybeans (S)
- Sugar (SB)
- 10-year U.S. T-Note (TY)
- Wheat (W)

Equity:

- U.S. Dollar Index (DX)
- Euro German Bund (EBL)
- E-mini S&P 500 (ES)
- DAX (FDX)
- Long Gilt (FLG)



- Gold (GC)
 - Hang Seng (HSI)
 - E-mini Nasdaq 100 (NQ)
 - Nikkei225 (NIY)
 - 10-year U.S. T-Note (TY)
- **Skala czasowa testów** – o ile nie wskazano inaczej, strategie są testowane w skali dziennej.
 - **Źródło danych historycznych** – wszystkie dane historyczne wykorzystane w testach pochodzą od dostawcy Commodity Strategias, Inc (CSI). Dane są korygowane o historyczne rolowania (back-adjusted) metodą close-to-close w oparciu o Open-Interest.
 - **Oprogramowanie** – Trading Blox Builder II.
 - **Test window** – dane historyczne pochodzą z okresu od 01.01.1995 do 31.12.2024.
 - **Podział danych historycznych** – zgodnie z najlepszymi praktykami rynkowymi, stosunek ilości danych (dni) out-of-sample do in-sample powinien wynosić nie więcej niż 1:8 i nie mniej niż 1:3. Oznacza to, że na każdy 1 rok danych out-of-sample powinno przypadać co najmniej 3 lata danych in-sample, jednak nie więcej niż 8 lat. Analogiczna zasada stosowana jest w analizie Walk-Forward (Walk-Forward Analysis).

W naszym przypadku testy przeprowadzane są zgodnie z poniższym podziałem:

- In-Sample – na danych z okresu od 01.01.2007 do 31.12.2024;
 - Out-of-Sample – na danych z okresu od 01.01.1998 do 31.12.2006;
 - Walk-forward analysis (WFA) – na danych z okresu od 01.01.1995 do 31.12.2024. Okno optymalizacji WFA (in-sample) ustawiane jest na 3 do 6 lat, natomiast okno testowania WFA (out-of-sample) na 25%-35% tego czasu.
- **Koszty transakcyjne** – slippage dla transakcji otwarcia/zamknięcia pozycji wynosi 7,5% ATR (Average True Range) z dnia otwarcia/zamknięcia pozycji. Ponadto uwzględniono prowizję w wysokości 4 USD za round-trip (tzn. kupno i sprzedaż kontraktu), jak również roll slippage w wysokości 7,5% ATR w przypadku rolowania pozycji na kolejną serię kontraktów.
Konserwatywne koszty transakcyjne obniżają roczną stopę zwrotu (CAGR%) nawet o kilka punktów procentowych, w zależności od strategii inwestycyjnej.
 - **Wielkość pozycji** – wielkość pozycji jest ustalana w oparciu o zmienność instrumentu (volatility adjusted) mierzoną ATR (Average True Range) i/lub odległość zlecenia obronnego (będącego częścią strategii bądź hipotetyczną wartością takiego zlecenia stop loss) metodą Percent Risk (zwaną również Fixed Fractional Money Manager).
 - **Funkcja celu (The Objective Function)** – to miara, względem której dokonuje się optymalizacji strategii inwestycyjnej oraz WFA. Innymi słowy, jest to miara, której wartość chcemy mieć jak najwyższą i jak najstabilniejszą (robust) dokonując optymalizacji oraz WFA.
W naszym przypadku jest to MAR (CAGR%/Maksymalny Drawdown), czyli miara zysku do ryzyka. Dodatkowo niezwykle istotny w całej analizie jest maksymalny drawdown.
Należy pamiętać, że dokonując optymalizacji parametrów, funkcja celu jest jednym z kilku elementów branych pod uwagę przy wyborze ostatecznego zestawu parametrów. Podczas określania najlepszego zestawu parametrów należy wziąć pod uwagę:
 - wartość funkcji celu;
 - zestaw parametrów, który jest najbardziej stabilny (robust) w zakresie +/-20%;
 - zestaw parametrów, który charakteryzuje się najbardziej stabilnym rozkładem drawdown w czasie oraz krzywej kapitału.



- **Optymalizacja parametrów** – testowane i optymalizowane są tylko te parametry, które są kluczowe dla danej strategii inwestycyjnej. Dla pozostałych parametrów wykorzystuje się wartości stałe, niepodlegające optymalizacji. Na przykład, przy obliczaniu wskaźnika ATR, którego wartość wykorzystywana jest do kalkulacji odległości zlecenia stop loss, długość okna (liczba dni) do wyznaczenia ATR jest stała i wynosi 40 dni. Aby sprawdzić, czy dany wskaźnik jest kluczowy dla strategii, należy przetestować strategię na szerokim zakresie parametrów dla tego wskaźnika. Jeżeli wyniki strategii będą zbliżone, można uznać, że wskaźnik nie jest kluczowy, a dany parametr może mieć wartość stałą.

Kolejnymi elementami, o których należy wspomnieć przy optymalizacji, są zakresy okna (przedział) testowanego parametru (np. dla średniej kroczącej może to być zakres od 10 do 30 dni) oraz krok zmiany. Oba te elementy są ze sobą powiązane, dlatego należy je omówić równoległe. Jeżeli chodzi o testowany zakres okna dla parametru, musi on współgrać z ideą strategii. To znaczy, jeżeli strategia ma być krótkoterminowa i wyłapywać krótkie ruchy, to długość krótkiej średniej kroczącej (w strategii z przecięciem dwóch średnich) nie będzie testowana w obszarze 100–200 dni, ponieważ nie odpowiada to założeniom strategii. Prawdopodobny zakres to 10–20 dni.

Kolejna sprawa to krok zmiany. Dla krótkiej średniej kroczącej krok może wynosić 1, gdy testujemy zakres 10–20 dni (tj. 10, 11, 12 ... 20 dni). Natomiast, jeżeli długa średnia krocząca jest testowana w zakresie 100–200 dni, przyjęcie kroku 1 dnia może doprowadzić do over-fittingu. Rozsądniejsze jest zastosowanie kroku 10 dni, aby zachować relacje między obiema średnimi (w każdym przypadku po 10 kroków). Można także rozszerzyć pierwszy zakres do 10–30 dni z krokiem 1 dzień, a drugi z krokiem 5 dni dla zakresu 100–200 dni. Generalnie należy zachować relacje między parametrami, które wzajemnie na siebie wpływają.

- **Ilość transakcji testowych** – ogólna zasada mówi, że im więcej transakcji testowych, tym lepiej. Jednak nadrzędną zasadą jest, aby na każdy optymalizowany element strategii przypadało co najmniej 50 transakcji testowych. Na przykład, jeżeli strategia opiera się na przecięciu dwóch średnich kroczących oraz wykorzystuje zlecenie początkowe stop loss, mamy łącznie 3 elementy do optymalizacji. Potrzebujemy więc przynajmniej 150 transakcji testowych.
- **Miary** – podczas testowania i optymalizacji strategii zwracam uwagę na następujące miary:
 - **CAGR% (Compounded Annual Growth Rate)** – pokazuje średnioroczne tempo wzrostu inwestycji przez określony czas, uwzględniając efekt procentu składanego (średnia geometryczna).
 - **Max drawdown (Maksymalne całkowite obsunięcie kapitału)** – jednorazowe zdarzenie, które odzwierciedla największe cofnięcie kapitału w stosunku do poprzedniego szczytu kapitału w okresie testowania.
 - **MAR (MAR Ratio)** – wskaźnik oceny zysku do ryzyka, mierzący stopę zwrotu inwestycji (CAGR%) w stosunku do maksymalnego spadku wartości kapitału (Max Drawdown) w danym okresie.
 - **Sharpe Ratio** – dzieli zannualizowaną stopę zwrotu przez zannualizowane odchylenie standardowe stóp zwrotu, wykorzystując miesięczne dane. Stopa wolna od ryzyka nie jest uwzględniana w liczniku.
 - **Longest drawdown** – najdłuższy czas trwania obsunięcia kapitału w całym kapitale w całym okresie testowania, mierzony od poprzedniego szczytu kapitału do nowego szczytu kapitału, wyrażony w miesiącach.



- **RAR% (Regressed Annual Return)** – stabilna miara stopy zwrotu, obliczana na podstawie regresji liniowej wszystkich punktów tworzących krzywą kapitału, odporna na zmiany dat rozpoczęcia i zakończenia testu.
- **R3 (R-cubed; Robust Risk/Reward Ratio)** – stabilna miara zysku do ryzyka. W liczniku wykorzystuje RAR%, a w mianowniku „maksymalne obsunięcie kapitału skorygowane ze względu na długość”, które jest iloczynem średniego maksymalnego obsunięcia kapitału oraz mnożnika uwzględniającego długość obsunięć.
Średnie maksymalne obsunięcie kapitału oblicza się, sumując pięć największych obsunięć i dzieląc wynik przez pięć. Mnożnik uwzględniający długość obsunięć to średnia maksymalna długość obsunięcia kapitału, która jest równa sumie długości pięciu najdłuższych obsunięć liczonych w dniach, podzielonej przez pięć, a następnie podzielonej przez 365.
- **R Sharp (Robust Sharpe Ratio)** – wskaźnik oceny zysku do ryzyka, dzielący RAR% przez odchylenie standardowe miesięcznej stopy zwrotu w skali roku.
- **Percent Profit Factor** – iloraz całkowitej wartości zyskowych transakcji do wartości stratnych transakcji w procentach. Wskaźnik musi wynosić co najmniej 1, a wartości powyżej 1,5 wskazują na dobrą strategię inwestycyjną.
- **Expectation ratio (R Ratio)** – czasami nazywane Wartością Oczekiwaną, to wskaźnik pokazujący, ile wynosi spodziewany zysk na każdego zaryzykowanego dolara w danej transakcji. Oblicza się go jako iloraz średniego zysku ze wszystkich transakcji i średniego ryzyka we wszystkich transakcjach. Przez średnie ryzyko rozumie się procent kapitału początkowego zaryzykowany w transakcji (stop loss).
Wartość wskaźnika większa niż 0 oznacza system zyskowy, natomiast wartości mniejsze niż 0 wskazują systemy stratne. Wskaźnik powinien wynosić co najmniej 0,1 (czyli na każdego zaryzykowanego dolara zyskujemy średnio 10 centów), natomiast odczyty powyżej 0,2 wskazują na dobrą strategię inwestycyjną.
- **SQN (System Quality Number)** – miara będąca ilorazem Expectation Ratio (R Ratio) i odchylenia standardowego Expectation Ratio (R Ratio). Następnie wynik jest mnożony przez pierwiastek kwadratowy z liczby transakcji. Miara ta pokazuje, na ile Expectation Ratio jest istotnie różne od zera – im wyższa wartość, tym lepiej:
 - Poniżej 1 – słaby system; istnieje niewielka szansa, że system w długim terminie będzie generował wyniki istotnie wyższe niż 0.
 - 1,01 – 2,00 – średni system; aby był statystycznie istotny, musi mieć wartość SQN co najmniej 1,70.
 - 2,01 – 3,00 – dobry system; jego zyskowość jest statystycznie istotnie wyższa niż 0.
 - 3,01 – 5,00 – świetny system.
 - 5,01 – 7,00 – fenomenalny system; istnieje zapewne tylko kilka takich na świecie.
 - 7,01 i więcej – „święty Graal”.
- **E-ratio** – wskaźnik stosunku wartości Maximum Favorable Excursion (MFE) do Maximum Adverse Excursion (MAE), skorygowanych o zmienność (ATR).
- **Wykresy** – standardowe prezentacje graficzne, takie jak krzywa kapitału (equity curve), drawdown, symulacje Monte Carlo, miesięczna stopa zwrotu, R-multiple.