

L3.4 — Pengujian Strategi: Backtest, Walk-Forward, dan Overfitting

Kurikulum Invesmentpedia · Level 3 Pengelola Portofolio · Modul
Inti 4 (Penutup Inti)

Invesmentpedia
2026

Daftar Isi

- L3.4 — Pengujian Strategi: Backtest, Walk-Forward, dan Overfitting
 - - 1. Identitas Modul
 - 2. Capaian Pembelajaran Mata Kuliah (CPMK)
 - 3. Peta Konsep
 - 4. Materi
 - 5. Contoh Terhitung
 - 6. Lembar Kerja Peserta
 - 7. Bank Soal
 - 8. Rubrik Penilaian
 - 9. Panduan Fasilitator
 - 10. Penafian

L3.4 — Pengujian Strategi: Backtest, Walk-Forward, dan Overfitting

1. Identitas Modul

Atribut	Keterangan
Kode	L3.4
Nama	Pengujian Strategi: Backtest, Walk-Forward, dan Overfitting
Jenjang	Level 3 — Pengelola Portofolio dan Sistem

Atribut	Keterangan
Jalur	Inti (penutup modul inti Level 3)
Beban	18 JP (15 jam) — 6 JP teori, 8 JP praktik, 4 JP asesmen
Prasyarat	L3.1–L3.3; L2.3 Statistik Terapan
Pilar	P4 Analisis & Riset (dominan), P5 Risiko, P6 Operasional & Etika
CPL terkait	CPL-3, CPL-4, CPL-5, CPL-8, CPL-9
Kanon	K-04, K-05, K-10, K-11, K-12

Posisi modul. L2T.4 §4.8 memperkenalkan overfitting dan menjanjikan pembahasan penuhnya di sini. Janji itu ditagih sekarang.

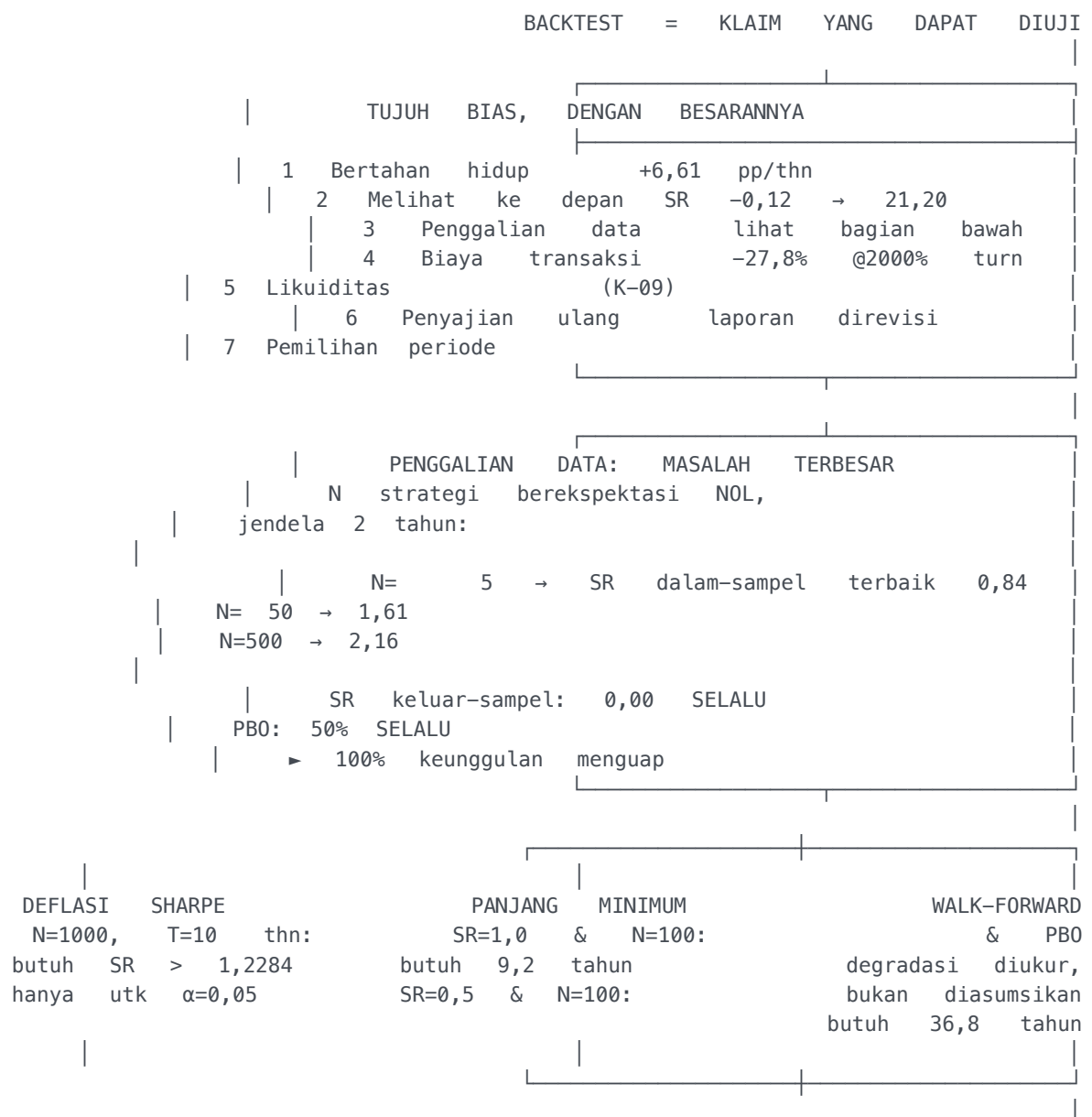
Modul ini adalah modul paling merusak dalam seluruh kurikulum, dan itu disengaja. Ia menunjukkan, dengan angka, bahwa **backtest yang tampak sangat bagus adalah hasil yang paling mungkin diperoleh dari strategi yang sama sekali tidak memiliki keunggulan**. Peserta yang menyelesaikan modul ini akan memandang setiap kurva ekuitas yang pernah ditunjukkan kepadanya dengan cara yang berbeda — termasuk kurva ekuitasnya sendiri.

2. Capaian Pembelajaran Mata Kuliah (CPMK)

Kode	Rumusan CPMK	Bloom
CPMK-L3.4.1	Mengidentifikasi tujuh bias backtest dan mengukur besaran masing-masing	C4
CPMK-L3.4.2	Menghitung Sharpe maksimum yang diharapkan dari N percobaan berekspektasi nol	C3
CPMK-L3.4.3	Menghitung panjang backtest minimum yang dibutuhkan untuk mendukung sebuah klaim Sharpe	C3
CPMK-L3.4.4	Menghitung ambang Sharpe terdeflasi setelah memperhitungkan jumlah percobaan	C4
CPMK-L3.4.5	Merancang dan menjalankan analisis walk-forward serta membaca degradasinya	C5
CPMK-L3.4.6	Mengestimasi Probabilitas Overfitting Backtest (PBO) dan	C5

Kode	Rumusan CPMK	Bloom
	menafsirkannya	
CPMK-L3.4.7	Mengevaluasi sebuah backtest yang disajikan pihak lain dan menyatakan apa yang kurang	C5
CPMK-L3.4.8	Menyusun Protokol Pengujian tertulis yang mencatat jumlah percobaan sejak awal	C6

3. Peta Konsep



4. Materi

4.1 Backtest adalah klaim, bukan bukti

Sebuah backtest menyatakan: “*bila strategi ini dijalankan pada periode tersebut, hasilnya begini.*” Pernyataan itu, bila dihitung dengan benar, adalah fakta tentang masa lalu.

Yang dibaca hampir semua orang justru pernyataan yang jauh lebih kuat: “*strategi ini bekerja.*” Lompatan dari yang pertama ke yang kedua memerlukan tiga syarat yang hampir tidak pernah diperiksa:

1. **Perhitungannya benar** — tidak ada bias yang mencemari hasilnya (§4.2)
2. **Hasilnya tidak dapat dijelaskan oleh kebetulan**, mengingat berapa banyak varian yang dicoba (§4.3)
3. **Datanya cukup panjang** untuk mendukung klaim sebesar itu (§4.4)

Modul ini memeriksa ketiganya, dan menemukan bahwa sebagian besar backtest gagal pada syarat kedua — bukan karena penyusunnya tidak jujur, melainkan karena **jumlah percobaan hampir tidak pernah dicatat**.

K-04 — proses dinilai terpisah dari hasil. Backtest yang bagus dengan proses yang buruk lebih berbahaya daripada backtest yang buruk, karena ia mengundang penempatan modal.

4.2 Tujuh bias, dengan besarannya

Bias 1 — Bertahan hidup (survivorship)

Menguji strategi hanya pada instrumen yang **masih ada hari ini** menghapus seluruh instrumen yang gagal.

Simulasi: 200 saham, 20 tahun, kemungkinan delisting 2,5% per tahun dengan kerugian 60% pada tahun delisting.

Sampel	CAGR rata-rata
Seluruh saham (termasuk yang delisting)	−1,454%
Hanya yang bertahan 20 tahun	+5,153%
Bias	+6,607 pp per tahun

Hanya **60,3%** saham bertahan dua puluh tahun; 39,7% menghilang dari sampel. Atas 20 tahun, bias itu menggelembungkan kekayaan akhir sebesar **266,1%**.

Bias ini mengubah strategi yang merugi menjadi strategi yang tampak menguntungkan. Ia adalah bias paling merusak dalam daftar ini dan paling sering tidak disadari, karena basis data komersial sering hanya memuat instrumen yang masih diperdagangkan.

Bias 2 — Melihat ke depan (look-ahead)

Menggunakan informasi yang belum tersedia pada saat keputusan dibuat. Bentuk paling umum: memakai harga penutupan hari ini untuk menghasilkan sinyal **dan** mengeksekusi pada harga penutupan yang sama.

Simulasi: 5.000 hari, strategi mengikuti arah pergerakan.

Sinyal	Sharpe	CAGR
Sah (memakai $t - 1$)	-0,1166	-3,93%
Look-ahead (memakai t)	+21,1992	+994,29%

Pergeseran satu hari mengubah Sharpe dari -0,12 menjadi 21,20.

Angka ini layak direnungkan. Sharpe 21 tidak ada di dunia nyata — tidak pernah. **Setiap backtest yang menghasilkan Sharpe di luar batas kewajaran hampir pasti memuat kebocoran informasi, dan itulah yang pertama harus dicari,** bukan dirayakan.

Bentuk look-ahead yang lebih halus dan lebih sering lolos:

- Memakai data laporan keuangan pada tanggal periode, bukan tanggal publikasi (jeda dua sampai tiga bulan)
- Memakai anggota indeks versi hari ini untuk periode masa lalu
- Menormalkan data memakai rata-rata atau simpangan baku seluruh periode
- Memakai harga penyesuaian yang sudah memuat aksi korporasi yang belum diumumkan

Bias 3 — Penggalan data (data snooping)

Mencoba banyak varian dan melaporkan yang terbaik. **Ini masalah terbesar dalam daftar ini,** dan §4.3 membahasnya tersendiri.

Bias 4 — Biaya transaksi

Hambatan biaya tumbuh linear terhadap perputaran. Pada strategi dengan Sharpe kotor 1,20 dan volatilitas 12%:

Perputaran/tahun	Biaya 10 bps	Sharpe bersih	Perubahan
50%	0,10%/thn	1,1917	-0,7%
250%	0,50%/thn	1,1583	-3,5%
1.000%	2,00%/thn	1,0333	-13,9%
2.000%	4,00%/thn	0,8667	-27,8%

Pada asumsi 10 bps, perputaran yang menghapus seluruh Sharpe adalah **7.200% per tahun.** Itu terdengar menenangkan — dan menyesatkan, karena **10 bps hanya berlaku untuk instrumen paling likuid.** Pada

saham berkapitalisasi kecil, aset digital di luar yang terbesar, atau periode tertekan, biaya efektif berkali lipat lebih tinggi, dan perputaran yang mematikan turun secara proporsional.

Dua komponen yang hampir selalu hilang dari backtest:

- **Dampak pasar** — order besar menggerakkan harga melawan Anda; efeknya tumbuh lebih cepat daripada linear terhadap ukuran
- **Biaya kesempatan** — order yang tidak terisi pada harga yang diasumsikan

Bias 5 — Likuiditas

Backtest mengasumsikan setiap order terisi pada harga yang tercatat. Dalam kenyataan, volume yang dapat Anda transaksikan tanpa menggerakkan harga terbatas — dan paling terbatas justru ketika Anda paling ingin keluar (**K-09**).

Uji yang wajib: untuk setiap transaksi dalam backtest, periksa apakah ukuran posisi melampaui proporsi wajar volume harian instrumen itu pada hari itu. Strategi yang lolos pada modal Rp100 juta sering gagal total pada Rp10 miliar.

Bias 6 — Penyajian ulang (restatement)

Data fundamental direvisi setelah publikasi pertama. Backtest yang memakai angka versi hari ini menggunakan informasi yang tidak tersedia pada saat keputusan diambil. Basis data yang memelihara **arsip versi asli** (*point-in-time*) diperlukan untuk menghindarinya, dan jarang tersedia bagi peneliti ritel.

Bias 7 — Pemilihan periode

Memilih titik awal dan akhir yang menguntungkan. Ini bentuk penggalan data yang paling mudah dilakukan dan paling sulit dideteksi, karena jumlah “percobaan” yang dilakukan tidak terlihat di hasil akhirnya.

Uji sederhana: jalankan strategi pada seluruh subperiode lima tahun yang mungkin. Bila hasilnya hanya baik pada beberapa jendela, Anda menemukan jendela, bukan strategi.

4.3 Penggalan data: masalah terbesar, diukur

Ini bagian terpenting modul ini.

Pertanyaan yang harus diajukan pada setiap backtest: *berapa banyak varian yang dicoba sebelum yang ini dipilih?*

4.3.1 Sharpe maksimum yang diharapkan dari kebetulan

Bila N strategi, semuanya tanpa keunggulan sejati, diuji pada data yang sama, Sharpe terbaik yang diharapkan adalah:

$$E[\max S R_N] \approx \hat{\sigma}_{SR} \left[(1 - \gamma) Z^{-1}(1 - 1/N) + \gamma Z^{-1}(1 - 1/N e) \right]$$

dengan $\gamma \approx 0,5772$ (tetapan Euler–Mascheroni) dan $\hat{\sigma}_{SR} \approx 1/\sqrt{T}$ untuk T tahun data.

N percobaan	$T = 5$ thn	$T = 10$ thn	$T = 20$ thn
10	0,7042	0,4979	0,3521
50	1,0180	0,7198	0,5090
100	1,1317	0,8002	0,5659
500	1,3651	0,9653	0,6826
1.000	1,4557	1,0294	0,7279
5.000	1,6492	1,1662	0,8246

Bacalah baris 1.000 pada kolom 5 tahun: Sharpe 1,4557. Itu backtest yang akan meyakinkan hampir setiap investor — dan ia diperoleh dari seribu strategi yang **seluruhnya berekspektasi nol**.

Perhatikan juga bahwa kolom bergerak ke kiri lebih cepat daripada baris bergerak ke bawah: **data yang pendek jauh lebih berbahaya daripada jumlah percobaan yang banyak**. Seribu percobaan atas 20 tahun menghasilkan Sharpe semu 0,7279; sepuluh percobaan atas 5 tahun menghasilkan 0,7042. Hampir sama.

4.3.2 Berapa Sharpe yang benar-benar dibutuhkan

Bila N percobaan dilakukan, ambang signifikansi harus dinaikkan. Ambang Sharpe pada $\alpha = 0,05$ dengan $T = 10$ tahun:

N	Ambang SR
1	0,5201
10	0,8120
100	1,0383
1.000	1,2284

Dengan seribu percobaan, Sharpe 1,20 tidak signifikan. Ia berada di bawah ambang 1,2284 — meskipun pada satu percobaan tunggal, Sharpe 1,20 akan sangat meyakinkan.

Konsekuensi praktisnya keras: menjalankan lebih banyak percobaan **menaikkan** standar bukti yang harus dipenuhi hasil terbaik Anda. Mencari lebih keras membuat temuan Anda lebih sulit dipercaya, bukan lebih mudah.

4.3.3 Panjang backtest minimum

Bailey dan rekan menurunkan panjang data minimum untuk mendukung klaim Sharpe tertentu setelah N percobaan:

$$\text{MinBTL} \approx \frac{2 \ln N}{SR^2} \text{ tahun}$$

SR yang diklaim	$N = 10$	$N = 100$	$N = 1.000$
0,5	18,4 thn	36,8 thn	55,3 thn
1,0	4,6 thn	9,2 thn	13,8 thn
1,5	2,0 thn	4,1 thn	6,1 thn
2,0	1,2 thn	2,0 thn	3,5 thn

Dua bacaan.

Pertama: klaim Sharpe rendah jauh lebih sulit dibuktikan. Untuk mendukung $SR=0,5$ setelah 100 percobaan dibutuhkan **36,8 tahun** data. Ini sering mengejutkan — Sharpe sederhana terasa seperti klaim yang rendah hati, padahal membuktikannya memerlukan data yang nyaris tidak ada.

Kedua: inilah alasan strategi frekuensi rendah sulit divalidasi. Strategi yang menghasilkan 12 transaksi per tahun tidak akan pernah mengumpulkan data yang cukup dalam karier seorang peneliti.

4.4 Walk-forward dan degradasi keluar-sampel

Cara terbaik mengukur overfitting bukan mendebatnya, melainkan **mengukurnya**.

Analisis walk-forward: bagi data menjadi jendela berurutan. Pada setiap jendela, optimalkan pada bagian dalam-sampel, lalu ukur kinerja pada bagian keluar-sampel **berikutnya** yang belum pernah disentuh. Geser maju. Ulangi.

Yang diukur adalah degradasi:

$$\text{Degradasi} = \frac{SR_{\text{keluar-sampel}} - SR_{\text{dalam-sampel}}}{SR_{\text{dalam-sampel}}}$$

Simulasi menentukan. Ambil N strategi yang **seluruhnya berekspektasi nol**. Pilih yang terbaik pada jendela dalam-sampel dua tahun. Ukur pada dua tahun keluar-sampel.

N	SR dalam-sampel (terpilih)	SR keluar-sampel	PBO
5	+0,8177	+0,0013	39,2%
20	+1,3400	-0,0567	52,2%
50	+1,6041	-0,0213	51,0%
100	+1,7857	+0,0573	43,8%
500	+2,1614	+0,0462	48,2%

Bandingkan kolom kedua dengan prediksi analitis §4.3.1 (d disesuaikan untuk jendela dua tahun): 0,8433 / 1,3440 / 1,6096 / 1,7894 / 2,1585. **Simulasi dan teori cocok sampai dua desimal.** Ini bukan kebetulan — ini menunjukkan bahwa seluruh Sharpe dalam-sampel dapat dijelaskan sepenuhnya oleh proses seleksi.

Dan kolom ketiga adalah kesimpulannya: Sharpe keluar-sampel adalah nol, pada setiap N .

Dengan $N=500$, backtest menunjukkan Sharpe **2,16** — kinerja yang akan menarik modal institusional. Keluar-sampel: **0,05**. Degradasi **100%**.

Probabilitas Overfitting Backtest (PBO)

PBO adalah probabilitas bahwa strategi yang terpilih sebagai terbaik dalam-sampel berkinerja **di bawah median** keluar-sampel.

$$PBO = P\left(SR_{\text{terpilih}}^{\text{keluar}} < \text{median}\left(SR^{\text{keluar}}\right)\right)$$

Tafsirannya:

PBO	Arti
Mendekati 0%	Seleksi dalam-sampel benar-benar mengidentifikasi strategi yang lebih baik
Mendekati 50%	Seleksi tidak membawa informasi — sama saja memilih acak
Di atas 50%	Seleksi secara aktif merugikan — Anda memilih yang paling ter-overfit

Pada simulasi di atas, PBO berkisar **39,2%–52,2%**, berpusat pada 50% sebagaimana yang harus terjadi bila tidak ada keunggulan sejati. **Proses seleksi tidak membawa informasi sama sekali.**

PBO adalah angka yang seharusnya menyertai setiap backtest yang disajikan, dan hampir tidak pernah ada.

Peringatan tentang walk-forward

Walk-forward **mengurangi** overfitting; ia tidak menghapusnya. Bila Anda menjalankan seluruh prosedur walk-forward berkali-kali dengan pengaturan berbeda dan memilih yang terbaik, Anda telah melakukan penggalan data pada tingkat yang lebih tinggi. **Jumlah percobaan tetap harus dicatat**, termasuk percobaan atas rancangan walk-forward itu sendiri.

4.5 Menilai backtest yang disajikan pihak lain

Kemampuan ini bernilai praktis langsung: peserta Level 3 akan terus-menerus ditunjukkan backtest oleh penyedia produk, penjual sistem, dan rekan.

Daftar periksa — sepuluh pertanyaan:

#	Pertanyaan	Bila tidak terjawab
1	Berapa varian yang diuji sebelum ini dipilih?	Hasil tidak dapat ditafsirkan
2	Berapa panjang data? Bandingkan dengan MinBTL	Klaim tidak didukung

#	Pertanyaan	Bila tidak terjawab
	untuk N itu	
3	Apakah instrumen yang delisting termasuk?	Curigai bias bertahan hidup
4	Kapan tepatnya sinyal tersedia dan kapan eksekusi diasumsikan?	Curigai look-ahead
5	Berapa perputaran, dan berapa asumsi biaya?	Sharpe kotor, bukan bersih
6	Apakah ukuran posisi diuji terhadap volume harian?	Tidak dapat diskalakan
7	Apakah data fundamental <i>point-in-time</i> ?	Curigai bias penyajian ulang
8	Berapa hasilnya pada seluruh subperiode 5 tahun?	Curigai pemilihan periode
9	Berapa PBO-nya?	Overfitting tidak terukur
10	Berapa drawdown maksimum dan lama pemulihannya?	Risiko tidak dinyatakan (L3.3)

Pertanyaan 1 adalah yang menentukan, dan jawaban yang paling sering diterima adalah “saya tidak tahu” atau “tidak banyak”. Keduanya berarti sama: N tidak dicatat, sehingga hasilnya tidak dapat ditafsirkan.

K-10 — jika kedengarannya terlalu bagus untuk jadi kenyataan, biasanya memang begitu. Backtest dengan Sharpe tinggi, drawdown kecil, dan kurva ekuitas mulus adalah **hasil yang paling mungkin diperoleh dari pencarian yang cukup besar atas strategi tanpa keunggulan**. Kemulusan itu sendiri adalah gejala, bukan bukti.

Catatan etika. Sebagian besar backtest yang menyesatkan **tidak dibuat untuk menipu**. Penyusunnya mencoba banyak varian, menemukan satu yang bagus, dan lupa menghitung berapa banyak yang dicoba. Ketika Anda menilai backtest orang lain, tanyakan N tanpa menuduh. Ketika Anda menyusun backtest sendiri, catat N **sejak awal** — karena Anda juga akan lupa.

4.6 Protokol Pengujian

Seluruh modul ini bermuara pada satu dokumen yang harus dibuka **sebelum** pengujian dimulai.

PROTOKOL PENGUJIAN STRATEGI — Templat

Bagian 1 — Ditulis sebelum data dibuka (*K-11*)

1.1 Hipotesis dalam satu kalimat: ____ 1.2 **Mekanisme ekonomi:** mengapa keunggulan ini seharusnya ada? ____ 1.3 Ruang parameter yang akan diuji: ____ $\times$ ____ $\times$ ____ = N = ____ 1.4 Panjang data yang tersedia:

____ tahun 1.5 $\text{MinBTL} = 2 \ln N / SR^2$ untuk SR yang diharapkan ____: ____ tahun 1.6 Apakah $1.4 \geq 1.5$? ☐ Ya ☐ Tidak → kurangi N atau turunkan klaim 1.7 Ambang SR terdeflasi untuk N dan T saya: ____ 1.8 Data keluar-sampel yang disisihkan dan tidak akan disentuh: ____ sampai ____

Bagian 2 — Catatan percobaan (diisi terus-menerus, tidak pernah dihapus)

#	Tanggal	Varian diuji	yang Hasil	Catatan
1				
...				
		TOTAL	N	____
		AKHIR		

Butir ini adalah inti protokol. N akhir hampir selalu jauh lebih besar daripada N yang direncanakan di 1.3, dan selisihnya hanya diketahui bila dicatat saat terjadi.

Bagian 3 — Pemeriksaan bias

Bias	Diperiksa?	Cara pemeriksaan	Hasil
Bertahan hidup	☐		
Melihat ke depan	☐		
Biaya transaksi	☐		
Likuiditas (ukuran vs volume harian)	☐		
Penyajian ulang	☐		
Pemilihan periode (seluruh subperiode 5 thn)	☐		

Bagian 4 — Hasil

4.1 SR dalam-sampel: ____ · SR keluar-sampel: ____ · **Degradasi:** ____% 4.2 **PBO:** ____% 4.3 Perputaran: ____%/thn · Biaya diasumsikan: ____ bps · SR bersih: ____ 4.4 Drawdown maksimum: ____% · Lama pemulihan: ____ (L3.3) 4.5 Kapasitas: modal maksimum sebelum dampak pasar berarti: Rp ____

Bagian 5 — Keputusan

5.1 Apakah SR bersih keluar-sampel melampaui ambang terdeflasi (1.7)? ☐ Ya ☐ **Tidak** 5.2 Apakah PBO di bawah 30%? ☐ Ya ☐ **Tidak** 5.3 Apakah mekanisme ekonomi (1.2) masih dapat dipertahankan setelah melihat hasil? ☐ Ya ☐ Tidak 5.4 **Bila salah satu dari 5.1–5.3 adalah “Tidak”: strategi ditolak.** ☐ Ditolak ☐ Diterima 5.5 Bila ditolak, data keluar-sampel **sudah terpakai** dan tidak dapat dipakai lagi untuk varian berikutnya

Tanda tangan dan tanggal: ____

Catatan tentang butir 5.5. Ini yang paling sering dilanggar. Sekali Anda melihat hasil keluar-sampel, data itu menjadi bagian dari proses seleksi Anda. Menguji varian kedua pada data yang sama berarti N bertambah dan seluruh perhitungan §4.3 harus diulang dengan N yang baru.

K-12 — Anda adalah risiko terbesar dalam portofolio Anda. Protokol ini tidak melindungi Anda dari pasar. Ia melindungi Anda dari keinginan Anda sendiri untuk menemukan sesuatu.

5. Contoh Terhitung

Contoh 1 — Mengukur bias bertahan hidup dan melihat ke depan

Bagian A — Bias bertahan hidup

Langkah 1 — rancangan. 200 saham, 20 tahun, imbal hasil tahunan $\mu = 9\%$, $\sigma = 28\%$, kemungkinan delisting 2,5% per tahun dengan kerugian 60% pada tahun delisting. 2.000 percobaan.

Langkah 2 — hitung proporsi yang bertahan.

$$(1 - 0,025)^{20} = 0,975^{20} = 60,3\%$$

39,7% saham menghilang dari sampel — dan mereka adalah yang terburuk.

Langkah 3 — bandingkan kedua sampel.

Sampel	CAGR rata-rata
Seluruh saham	-1,454%
Hanya yang bertahan	+5,153%
Bias	+6,607 pp/tahun

Langkah 4 — hitung dampak kumulatif atas 20 tahun.

$$\left(\frac{1,05153}{0,98546}\right)^{20} - 1 = 266,1\%$$

Sampel yang bias menunjukkan kekayaan akhir **3,66 kali lipat** kenyataan.

Langkah 5 — bacaan yang menentukan. Sampel penuh menunjukkan strategi ini **merugi** (-1,454%); sampel bias menunjukkan ia **menguntungkan** (+5,153%). Bias ini tidak melebih-lebihkan keunggulan — ia menciptakan keunggulan yang tidak ada.

Bagian B — Bias melihat ke depan

Langkah 1 — rancangan. 5.000 hari imbal hasil acak (μ harian 0,03%, σ 1,2%). Strategi: ambil posisi searah pergerakan.

Langkah 2 — versi sah. Sinyal dari $t - 1$, eksekusi pada t :

$$SR = -0,1166 \text{ CAGR} = -3,93\%$$

Sebagaimana seharusnya — data ini acak, jadi tidak ada keunggulan.

Langkah 3 — versi look-ahead. Sinyal dari t , eksekusi pada t :

$$SR = +21,1992 \text{ CAGR} = +994,29\%$$

Langkah 4 — ukur kerusakannya. Pergeseran satu hari mengubah Sharpe dari $-0,12$ menjadi $21,20$, dan CAGR dari $-3,93\%$ menjadi $+994,29\%$.

Langkah 5 — aturan praktis yang mengikuti. Sharpe di atas sekitar 3 pada strategi yang dapat diakses investor ritel nyaris selalu menandakan kebocoran informasi. **Ketika backtest Anda menghasilkan angka luar biasa, langkah pertama adalah mencari kebocorannya — bukan merayakannya.**

Verifikasi yang wajib: untuk setiap sinyal, catat **stempel waktu ketersediaan data** dan pastikan eksekusi terjadi setelahnya.

Contoh 2 — Sharpe semu dari penggalian data

Seorang peneliti menguji **500 kombinasi parameter** pada **5 tahun** data dan melaporkan strategi terbaik dengan Sharpe **1,35**.

Langkah 1 — hitung galat baku Sharpe.

$$\hat{\sigma}_{SR} \approx \frac{1}{\sqrt{T}} = \frac{1}{\sqrt{5}} = 0,4472$$

Langkah 2 — hitung Sharpe maksimum yang diharapkan dari kebetulan.

$$E[\max SR_{500}] = 0,4472 \left[(1 - 0,5772) Z^{-1}(0,998) + 0,5772 Z^{-1}(0,999264) \right] = 1,3651$$

Langkah 3 — bandingkan.

$$SR_{\text{dilaporkan}} = 1,35 \quad E[\max SR_{500}] = 1,3651$$

Sharpe yang dilaporkan berada di bawah nilai yang diharapkan dari kebetulan murni. Strategi ini tidak memberikan bukti keunggulan apa pun. Bila seluruh 500 strategi berekspektasi nol, hasil terbaik yang **diharapkan** adalah 1,3651 — lebih baik daripada yang benar-benar ditemukan.

Langkah 4 — hitung panjang backtest minimum.

$$\text{MinBTL} = \frac{2 \ln 500}{1,35^2} = \frac{2(6,2146)}{1,8225} = 6,82 \text{ tahun}$$

Tersedia 5 tahun. **Kurang 1,82 tahun**, bahkan bila Sharpe-nya nyata.

Langkah 5 — hitung ambang yang sebenarnya dibutuhkan. Pada $T = 5$, $N = 500$, $\alpha = 0,05$, kuantil ke-95 dari maksimum 500 normal:

$$S R_{\text{ambang}} = \Phi^{-1}(0,95^{1/500}) \times 0,4472 = 1,6603$$

Sharpe 1,35 berada **jauh di bawahnya**.

Langkah 6 — kesimpulan dan tanggapan yang benar.

“Sharpe 1,35 dari 500 percobaan atas 5 tahun data tidak memberikan bukti keunggulan. Nilai yang diharapkan dari 500 strategi tanpa keunggulan apa pun adalah 1,3651 — lebih tinggi daripada yang Anda temukan. Untuk mendukung klaim Sharpe 1,35 setelah 500 percobaan, Anda memerlukan 6,82 tahun data; Anda memiliki 5. Ambang signifikansi pada jumlah percobaan itu adalah 1,6603.”

Langkah 7 — bandingkan dengan skenario yang lebih baik. Peneliti yang sama, tetapi mencatat bahwa ia hanya menguji **10 varian** yang ditetapkan sebelumnya, atas **20 tahun** data:

$$E[\max S R_{10}] = 0,2236[\dots] = 0,3521 \text{ MinBTL} = \frac{2 \ln 10}{1,35^2} = 2,53 \text{ tahun}$$

Sharpe 1,35 yang sama kini sangat berarti: ia jauh melampaui 0,3521, dan 20 tahun data jauh melebihi MinBTL 2,53 tahun.

Angka Sharpe-nya identik. Yang berbeda adalah N dan T — dan itulah yang menentukan apakah angka itu berarti.

Contoh 3 — Walk-forward: mengukur degradasi

Langkah 1 — rancangan. N strategi, seluruhnya berekspektasi **nol**. Jendela dalam-sampel 2 tahun; jendela keluar-sampel 2 tahun berikutnya. Pilih terbaik dalam-sampel; ukur keluar-sampel. 400 percobaan.

Langkah 2 — prediksi analitis sebelum simulasi (§4.3.1, $\hat{\sigma}_{SR} = 1/\sqrt{2} = 0,7071$):

N	$E[\max S R]$ prediksi
5	0,8433
20	1,3440
50	1,6096
100	1,7894
500	2,1585

Langkah 3 — hasil simulasi.

N	$S R$ dalam-sampel	Prediksi	$S R$ keluar-sampel	PBO
5	+0,8177	0,8433	+0,0013	39,2%

N	SR dalam-sampel	Prediksi	SR keluar-sampel	PBO
20	+1,3400	1,3440	-0,0567	52,2%
50	+1,6041	1,6096	-0,0213	51,0%
100	+1,7857	1,7894	+0,0573	43,8%
500	+2,1614	2,1585	+0,0462	48,2%

Langkah 4 — periksa kecocokan teori dan simulasi. Selisih terbesar antara kolom kedua dan ketiga adalah **0,0256** (pada $N = 5$); pada $N = 500$ selisihnya hanya **0,0029**.

Kecocokan ini adalah temuannya. Seluruh Sharpe dalam-sampel dapat diprediksi **sepenuhnya** dari jumlah percobaan dan panjang data — tanpa satu pun asumsi tentang strateginya. Karena strategi-strategi itu memang tidak memiliki keunggulan, tidak ada yang tersisa untuk dijelaskan.

Langkah 5 — baca kolom keluar-sampel. Seluruhnya berkisar nol: **+0,0013** sampai **-0,0567**. **Degradasi 100% pada setiap N .**

Langkah 6 — soroti baris $N = 500$.

$$SR_{\text{dalam-sampel}} = 2,1614 \quad SR_{\text{keluar-sampel}} = 0,0462$$

Backtest dengan Sharpe 2,16 akan menarik modal institusional. Ia diperoleh dari 500 strategi yang **seluruhnya berekspektasi nol**.

Langkah 7 — baca PBO. Rentang 39,2%–52,2%, berpusat pada 50%. Tafsirannya: **proses seleksi dalam-sampel tidak membawa informasi apa pun tentang kinerja keluar-sampel**. Memilih strategi terbaik dalam-sampel sama bergunanya dengan memilih acak.

Langkah 8 — apa yang akan terlihat berbeda bila keunggulan itu nyata. Pada strategi dengan keunggulan sejati:

- SR keluar-sampel akan **positif dan berarti**, meski lebih rendah daripada dalam-sampel
- PBO akan **jauh di bawah 50%** — idealnya di bawah 30%
- Degradasi akan berupa persentase, bukan 100%

Ketiga tanda itu yang harus dicari, dan ketiganya dapat dihitung.

Contoh 4 — Menilai backtest yang disajikan

Sebuah penyedia sistem menyajikan:

“Strategi kami menghasilkan CAGR 34% dengan Sharpe 2,8 dan drawdown maksimum hanya 8% selama backtest 2019–2023. Diuji pada 50 saham teratas.”

Langkah 1 — periksa Sharpe terhadap batas kewajaran. Sharpe 2,8 dengan drawdown maksimum 8% atas lima tahun. Bandingkan dengan L3.3 §4.2: portofolio dengan volatilitas rendah pun mengalami drawdown berarti.

Volatilitas tersirat dari Sharpe 2,8 dan imbal hasil berlebih sekitar 29% ($34\% - 5\%$):

$$\sigma = \frac{0,29}{2,8} = 10,36\%$$

Drawdown maksimum yang diharapkan untuk $\sigma = 10,36\%$ atas 5 tahun berada di kisaran 14%–17% (interpolasi tabel L3.3 §4.2). **Klaim 8% berada jauh di bawah yang wajar** — indikasi pertama bahwa ada yang tidak beres.

Langkah 2 — periksa “50 saham teratas”. Teratas menurut kapan? Bila menurut kapitalisasi hari ini, seluruh backtest memuat bias bertahan hidup **dan** bias melihat ke depan sekaligus: saham yang menjadi terbesar pada 2023 dipilih untuk diperdagangkan sejak 2019.

Besaran biasanya dari Contoh 1: bias bertahan hidup saja bernilai **+6,607 pp per tahun**. Memilih pemenang secara retrospektif jauh lebih besar lagi.

Langkah 3 — periksa periode. 2019–2023 adalah **lima tahun**. Ajukan pertanyaan 2 daftar periksa: berapa N ?

Andaikan penyedia menjawab jujur “sekitar 200 kombinasi”:

$$E[\max S R_{200}] \approx 1,2368 \text{ pada } T=5$$

Sharpe 2,8 **melampaui** nilai itu — sehingga penggalan data saja tidak menjelaskan seluruhnya. Tetapi:

$$\text{MinBTL} = \frac{2 \ln 200}{2,8^2} = \frac{10,5966}{7,84} = 1,35 \text{ tahun}$$

Lima tahun melampaui 1,35 tahun. **Pada dimensi ini, klaimnya lolos** — yang menunjukkan bahwa bias pemilihan sampel (Langkah 2), bukan penggalan data, adalah masalah utamanya.

Langkah 4 — periksa yang tidak disebutkan. Daftar periksa §4.5 memuat sepuluh pertanyaan; penyajian ini menjawab **dua** (periode dan semesta, itu pun tidak lengkap). Yang tidak dinyatakan:

Tidak dinyatakan	Mengapa penting
N percobaan	Tanpa ini hasil tidak dapat ditafsirkan
Perputaran dan biaya	Sharpe kotor atau bersih?
Saham yang delisting	Bias bertahan hidup
Waktu ketersediaan sinyal	Bias melihat ke depan
Uji ukuran vs volume harian	Kapasitas
PBO	Overfitting tidak terukur
Lama pemulihan drawdown	Risiko tidak lengkap
Hasil per subperiode	Pemilihan periode

Langkah 5 — susun tanggapan.

“Terima kasih. Sebelum saya dapat menilai ini, saya memerlukan empat hal. Pertama, berapa varian yang diuji sebelum yang ini dipilih — tanpa angka itu, Sharpe 2,8 tidak dapat ditafsirkan. Kedua, apakah ‘saham teratas’ ditentukan menurut data pada setiap tanggal historis atau menurut data hari ini; bila yang kedua, backtest ini memuat bias yang pada simulasi kami bernilai lebih dari 6 poin persentase per tahun. Ketiga, berapa perputaran dan asumsi biayanya. Keempat, drawdown 8% pada volatilitas tersirat 10,4% lebih rendah daripada yang wajar untuk periode lima tahun; saya ingin melihat perhitungannya.”

“Saya tidak menuduh apa pun — sebagian besar backtest yang menyesatkan disusun dengan itikad baik oleh orang yang lupa mencatat berapa varian yang mereka coba. Keempat angka itu akan menyelesaikan persoalan.”

Langkah 6 — keputusan. Tanpa keempat jawaban itu, **backtest ini tidak dapat ditafsirkan** — bukan salah, melainkan tidak dapat ditafsirkan. Perbedaan itu penting dan harus disampaikan demikian.

Langkah 7 — aturan umum yang dapat dibawa pulang. Backtest dengan kurva ekuitas mulus, Sharpe tinggi, dan drawdown kecil bukan bukti strategi yang bagus. Ia adalah **profil hasil yang paling mungkin muncul dari pencarian yang cukup besar**. Kemulusannya adalah gejala.

6. Lembar Kerja Peserta

Bagian A — Hitung kontaminasi backtest Anda sendiri

Ambil satu backtest yang pernah Anda buat.

- N varian yang **benar-benar** Anda coba (jujurlah; hitung juga yang gagal): ____
- Panjang data T : ____ tahun · $\hat{\sigma}_{SR} = 1/\sqrt{T} =$ ____
- $E[\max S R_N] =$ ____
- $S R$ yang saya laporkan: ____
- Apakah $S R$ saya melampaui $E[\max S R_N]$? ☐ Ya ☐ Tidak → tidak ada bukti keunggulan
- $\text{MinBTL} = 2 \ln N / S R^2 =$ ____ tahun · Apakah $T \geq \text{MinBTL}$? ☐ Ya ☐ Tidak
- Ambang $S R$ terdeflasi: ____ · Apakah $S R$ saya melampauinya? ☐ Ya ☐ Tidak

Bagian B — Audit tujuh bias

Bias	Ada?	Besaran yang saya estimasi	Cara saya memeriksanya
Bertahan hidup	☐		

Bias	Ada?	Besaran yang saya estimasi	Cara saya memeriksanya
Melihat ke depan	☐		
Penggalian data	☐		
Biaya transaksi	☐		
Likuiditas	☐		
Penyajian ulang	☐		
Pemilihan periode	☐		
- Bias terbesar dalam backtest saya: _____ SR setelah saya koreksi bias yang dapat saya koreksi: _____			

Bagian C — Walk-forward

Jendela	Periode dalam-sampel	SR dalam	Periode keluar	SR keluar
1				
2				
...				
Rata-rata				
- Degradasi = _____% PBO = _____% - Tafsiran PBO saya: ☐ <30% baik ☐ ~50% seleksi tidak informatif ☐ >50% seleksi merugikan				

Bagian D — Nilai backtest pihak lain

Ambil satu backtest yang pernah ditunjukkan kepada Anda (produk, sistem, atau kursus).

#	Pertanyaan daftar periksa §4.5	Terjawab?	Jawabannya
1	Berapa <i>N</i> ?	☐	
2	Panjang data vs MinBTL	☐	
...		☐	
10	Drawdown dan pemulihan	☐	
- Jumlah pertanyaan terjawab: _____ dari 10 Kesimpulan saya: ☐ Dapat ditafsirkan ☐ Tidak dapat ditafsirkan - Tanggapan tertulis saya kepada penyaji (gunakan nada Contoh 4 Langkah 5): _____			

Bagian E — Protokol Pengujian

Isi templat §4.6 secara lengkap untuk strategi berikutnya yang akan Anda uji. **Bagian 1 harus diisi sebelum data dibuka**, dan Bagian 2 diisi terus-menerus.

Dokumen ini menjadi Bagian 4 Pernyataan Kebijakan Investasi pada **Proyek Mandat Level 3**.

7. Bank Soal

Butir disusun mengikuti urutan materi §4 agar dapat dipakai per sesi, bukan menaik menurut tingkat kognitif. Setiap butir diberi label kognitif, sehingga fasilitator dapat menyusun kuis pendek menurut bagian §4 yang sedang diajarkan, menurut tingkat kognitif, atau keduanya.

Sebaran modul ini: **C1** Mengingat 7 · **C3** Menerapkan 4 · **C4** Menganalisis 10 · **C5** Mengevaluasi 4.

1. Sebutkan tiga syarat agar backtest mendukung klaim “strategi ini bekerja”. (C1) (Kunci: perhitungan bebas bias; hasil tidak dapat dijelaskan kebetulan mengingat N ; data cukup panjang)
2. Sebutkan tujuh bias backtest. (C1) (Kunci: bertahan hidup; melihat ke depan; penggalian data; biaya transaksi; likuiditas; penyajian ulang; pemilihan periode)
3. Hitung proporsi saham yang bertahan 20 tahun bila kemungkinan delisting 2,5%/tahun. (C3) (Kunci: $0,975^{20} = 60,3\%$)
4. Sampel penuh memberi CAGR $-1,454\%$, sampel bertahan $+5,153\%$. Hitung bias dan dampak kumulatif 20 tahun. (C3) (Kunci: $+6,607$ pp/tahun; $(1,05153 / 0,98546)^{20} - 1 = 266,1\%$)
5. Apa yang paling mengkhawatirkan dari hasil soal 4? (C5) (Kunci: bias tidak melebihi-lebihkan keunggulan — ia menciptakan keunggulan yang tidak ada; strategi merugi tampak menguntungkan)
6. Sinyal sah memberi Sharpe $-0,1166$; look-ahead memberi 21,1992. Apa aturan praktis yang mengikuti? (C4) (Kunci: Sharpe luar biasa nyaris selalu menandakan kebocoran informasi; langkah pertama adalah mencari kebocorannya, bukan merayakannya)
7. Sebutkan empat bentuk look-ahead yang halus. (C1) (Kunci: memakai tanggal periode alih-alih tanggal publikasi; anggota indeks versi hari ini; normalisasi memakai statistik seluruh periode; harga penyesuaian yang memuat aksi korporasi belum diumumkan)
8. Strategi Sharpe kotor 1,20, vol 12%, perputaran 2.000%/thn @10 bps. Hitung Sharpe bersih. (C3) (Kunci: hambatan 4,00%/thn; $1,20 - 0,04 / 0,12 = 0,8667$, turun 27,8%)
9. Mengapa tabel biaya transaksi pada §4.2 dapat menyesatkan? (C5) (Kunci: 10 bps hanya untuk instrumen paling likuid; pada kapitalisasi kecil, aset digital kecil, atau periode tertekan, biaya berkali lipat dan perputaran yang mematikan turun proporsional)
10. Dua komponen biaya apa yang hampir selalu hilang dari backtest? (C4) (Kunci: dampak pasar — tumbuh lebih cepat daripada linear; biaya kesempatan order tidak terisi)
11. Tuliskan rumus $E[\max S R_N]$. (C1) (Kunci: $\hat{\sigma}_{SR}[(1 - \gamma)Z^{-1}(1 - 1/N) + \gamma Z^{-1}(1 - 1/(Ne))]$, $\gamma \approx 0,5772$)
12. Hitung $E[\max S R]$ untuk $N=1.000$ dan $T=5$ tahun. Apa artinya? (C4) (Kunci: 1,4557; backtest yang sangat meyakinkan diperoleh dari 1.000 strategi berekspektasi nol)

13. Bandingkan $N=1.000$ atas 20 tahun dengan $N=10$ atas 5 tahun. Apa pelajarannya? (C4) (Kunci: $0,7279$ vs $0,7042$ — hampir sama; data pendek jauh lebih berbahaya daripada banyak percobaan)
14. Pada $T=10$ tahun dan $N=1.000$, berapa ambang SR untuk $\alpha=0,05$? Apa konsekuensinya bagi Sharpe 1,20? (C4) (Kunci: 1,2284; Sharpe 1,20 tidak signifikan)
15. Apa konsekuensi praktis dari soal 14 bagi cara meneliti? (C4) (Kunci: lebih banyak percobaan menaikkan standar bukti — mencari lebih keras membuat temuan lebih sulit dipercaya, bukan lebih mudah)
16. Tuliskan rumus MinBTL dan hitung untuk $SR=0,5$ dengan $N=100$. (C3) (Kunci: $2 \ln N / SR^2$; $2 \ln 100 / 0,25 = 36,8$ tahun)
17. Mengapa klaim Sharpe rendah lebih sulit dibuktikan daripada Sharpe tinggi? (C4) (Kunci: MinBTL berbanding terbalik dengan SR^2 ; $SR=0,5$ butuh 36,8 tahun sementara $SR=2,0$ butuh 2,3 tahun pada N yang sama)
18. Tuliskan definisi degradasi walk-forward. (C1) (Kunci: $(SR_{keluar} - SR_{dalam}) / SR_{dalam}$)
19. Pada $N=500$, SR dalam-sampel 2,1614 dan keluar-sampel 0,0462. Hitung degradasinya dan jelaskan. (C4) (Kunci: $-97,9\% \approx 100\%$; seluruh Sharpe dalam-sampel berasal dari seleksi, bukan keunggulan)
20. Prediksi analitis untuk $N=500$ adalah 2,1585; simulasi memberi 2,1614. Mengapa kecocokan ini penting? (C4) (Kunci: seluruh Sharpe dalam-sampel dapat diprediksi dari N dan T saja, tanpa asumsi tentang strateginya — tidak ada yang tersisa untuk dijelaskan)
21. Tuliskan definisi PBO dan tafsiran tiga rentangnya. (C1) (Kunci: $P(SR_{terpilih}^{keluar} < \text{median})$; $\sim 0\%$ seleksi informatif; $\sim 50\%$ seleksi tidak membawa informasi; $> 50\%$ seleksi merugikan)
22. PBO simulasi berkisar 39,2%–52,2%. Apa kesimpulannya? (C4) (Kunci: berpusat pada 50% — proses seleksi tidak membawa informasi; memilih terbaik dalam-sampel sama bergunanya dengan memilih acak)
23. Tiga tanda apa yang menunjukkan keunggulan sejati pada walk-forward? (C1) (Kunci: SR keluar-sampel positif dan berarti; PBO jauh di bawah 50% (idealnya $< 30\%$); degradasi berupa persentase, bukan 100%)
24. Peneliti melaporkan Sharpe 1,35 dari 500 percobaan atas 5 tahun. Hitung $E[\max SR]$, MinBTL, dan ambangnya ($T = 5$). Apa kesimpulannya? (C5) (Kunci: 1,3651; 6,82 tahun; 1,6603 — Sharpe yang dilaporkan di bawah nilai harapan kebetulan, data kurang 1,82 tahun, jauh di bawah ambang 1,6603; tidak ada bukti keunggulan)
25. Sharpe 1,35 yang sama dari 10 percobaan atas 20 tahun. Hitung ulang dan bandingkan. (C5) (Kunci: $E[\max SR_{10}] = 0,3521$; MinBTL = 2,53 tahun; kini sangat berarti — angka Sharpe identik, yang menentukan adalah N dan T)

8. Rubrik Penilaian

Dim	Aspek	4 — Sangat Baik	3 — Baik	2 — Cukup	1 — Kurang
		Baik	Menguasai	Menyebut bias	Memperlakuka
D1	Penguasaan	Menjelaskan	Menguasai	Menyebut bias	Memperlakuka

Dim	Aspek	4 — Sangat Baik	3 — Baik	2 — Cukup	1 — Kurang
	konsep	mengapa backtest yang tampak sangat bagus adalah hasil paling mungkin dari pencarian besar tanpa keunggulan	tujuh bias dan kerangka deflasi	tanpa memahami besarannya	n backtest sebagai bukti
D2	Ketepatan kuantitatif	$E[\max S R]$, MinBTL, ambang deflasi, degradasi, dan PBO benar dan diperiksa silang terhadap prediksi analitis	Benar dengan satu kekeliruan kecil	Beberapa kekeliruan	Tidak dapat direproduksi
D3	Penalaran & kerangka	Protokol pengujian utuh; menghubungkan N , T , bias, dan degradasi menjadi satu penilaian	Kerangka koheren sebagian besar	Komponen ada tetapi terpisah	Tidak ada kerangka
D4	Perlakuan risiko (knockout)	Kapasitas, likuiditas, drawdown, dan lama pemulihan dihitung dan dinyatakan	Risiko backtest dinilai memadai	Risiko disebut tanpa dihitung	Strategi diusulkan tanpa uji likuiditas/kapasitas, atau tanpa drawdown → GAGAL
D5	Disiplin proses (knockout)	N dicatat sejak awal; data keluar-sampel disentuh sekali; strategi yang gagal ambang benar-benar ditolak	Protokol diikuti	Protokol diikuti sebagian	N tidak dicatat, atau data keluar-sampel dipakai berulang, atau strategi gagal ambang tetap diterima →

Dim	Aspek	4 — Sangat Baik	3 — Baik	2 — Cukup	1 — Kurang
					GAGAL
D6	Komunikasi & etika	Melaporkan <i>N</i> tanpa diminta; menilai backtest pihak lain tanpa menuduh; membedakan “salah” dari “tidak dapat ditafsirkan”	Pelaporan jujur	Pelaporan kurang lengkap	Menyajikan backtest tanpa <i>N</i> , atau menuduh penyaji beritikad buruk tanpa dasar

Ambang lulus: rata-rata $\geq 3,0$ dan $D4 \geq 2$ dan $D5 \geq 2$.

D4 (perlakuan risiko) dan D5 (disiplin proses) bersifat menggugurkan: nilai 1 pada salah satunya membuat tugas harus diulang, berapa pun nilai dimensi lain.

Catatan penilai. Peserta yang melaporkan backtest sendiri dengan jujur dan menyimpulkan “*tidak ada bukti keunggulan*” memperoleh nilai **lebih tinggi** daripada peserta yang menyajikan hasil mengesankan tanpa mencatat *N*. Ini dinyatakan kepada peserta sebelum tugas diberikan.

9. Panduan Fasilitator

Alokasi waktu (18 JP)

Sesi	Isi	JP
1	§4.1–4.2 Bias 1–2; Contoh 1 — sesi pembuka yang menentukan	3
2	§4.2 Bias 3–7; besaran masing-masing	2
3	§4.3.1 Sharpe semu dari penggalan data	2,5
4	§4.3.2–4.3.3 Deflasi dan MinBTL; Contoh 2	2,5
5	§4.4 Walk-forward dan PBO; Contoh 3	3
6	§4.5 Menilai backtest pihak lain; Contoh 4	2

Sesi	Isi	JP
7	§4.6 Protokol Pengujian	1
8	Asesmen: Bank Soal + presentasi audit backtest sendiri	2

Urutan yang disarankan

Buka Sesi 1 dengan demonstrasi look-ahead secara langsung. Tulis strategi sepuluh baris di depan kelas, jalankan dengan sinyal dari $t - 1$ (Sharpe $-0,12$), lalu ubah **satu indeks** menjadi t dan jalankan lagi (Sharpe $21,20$). Tunjukkan kurva ekuitasnya.

Tidak ada penjelasan yang seefektif melihat kurva itu muncul dari perubahan satu karakter.

Sesi 8 harus diumumkan sejak awal modul: tugas akhirnya adalah **mengaudit backtest peserta sendiri**. Sebagian peserta akan menemukan bahwa pekerjaan berbulan-bulan mereka tidak memberikan bukti apa pun. Umumkan sejak awal bahwa **menemukan hal itu dan melaporkannya dengan jujur adalah hasil terbaik yang mungkin dari modul ini**, dan akan dinilai demikian.

Momen pengajaran kunci

Contoh 2 Langkah 3 adalah puncak modul ini. Sharpe yang dilaporkan ($1,35$) berada **di bawah** nilai yang diharapkan dari kebetulan murni ($1,3651$). Peneliti itu bekerja keras dan menemukan sesuatu yang lebih buruk daripada keberuntungan rata-rata. Diamkan sejenak sebelum melanjutkan.

Contoh 2 Langkah 7 mencegah nihilisme. Sharpe **$1,35$ yang sama** menjadi sangat berarti bila $N = 10$ dan $T = 20$. Ini penting untuk disampaikan: modul ini tidak mengatakan backtest tidak berguna. Ia mengatakan **backtest tanpa N tidak dapat ditafsirkan**.

Kecocokan teori-simulasi pada Contoh 3 Langkah 4 layak disorot eksplisit. Selisih $0,0029$ pada $N=500$. Ini bukan sekadar hasil simulasi yang rapi — ia menunjukkan bahwa tidak ada sisa yang memerlukan penjelasan lain.

Nada Contoh 4 Langkah 5 harus dilatih. Peserta cenderung menjadi sinis setelah modul ini dan menuduh setiap penyaji backtest beritikad buruk. Itu salah secara faktual dan merusak secara profesional. Latih kalimat: *“Saya tidak menuduh apa pun — sebagian besar backtest yang menyesatkan disusun dengan itikad baik oleh orang yang lupa mencatat berapa varian yang mereka coba.”*

Kesalahan umum peserta

Kesalahan	Koreksi
Merayakan Sharpe yang sangat tinggi	Contoh 1 Bagian B; cari kebocoran lebih dulu
Tidak mencatat N , lalu mengira-ngira belakangan	§4.6 Bagian 2; N akhir hampir selalu jauh lebih besar daripada yang direncanakan

Kesalahan	Koreksi
Menguji varian kedua pada data keluar-sampel yang sama	§4.6 butir 5.5; sekali dilihat, data itu bagian dari seleksi
Mengira walk-forward menghapus overfitting	§4.4; ia mengurangi, tidak menghapus — dan rancangannya sendiri dapat digali
Memakai semesta instrumen versi hari ini	Contoh 1 Bagian A dan Contoh 4 Langkah 2
Menyimpulkan “semua backtest tidak berguna”	Contoh 2 Langkah 7; yang menentukan adalah N dan T
Menuduh penyaji backtest beritikad buruk	§4.5 catatan etika; sebagian besar disusun dengan itikad baik

Aturan keras bagi fasilitator

1. **Dilarang menyajikan backtest apa pun di kelas tanpa menyebut N , panjang data, dan asumsi biayanya** — termasuk backtest yang dipakai modul ini sendiri sebagai bahan ajar.
2. **Dilarang menyebutkan sistem, produk, penyedia, atau penyelenggara kursus tertentu** sebagai contoh backtest yang menyesatkan. Seluruh contoh dalam modul ini bersifat konstruksi ajar, dan harus tetap demikian. Menyebut pihak nyata berpotensi menimbulkan konsekuensi hukum dan tidak diperlukan untuk mengajarkan materinya.
3. **Dilarang membantu peserta menyusun backtest untuk dipasarkan kepada pihak ketiga.** Modul ini mengajarkan menilai bukti, bukan menyiapkan bahan penjualan.
4. **Dilarang menyatakan bahwa suatu strategi “terbukti”** dalam bentuk apa pun.
5. **Dilarang menerima imbalan, komisi, atau biaya rujukan** dari penyedia data, platform backtest, penjual sistem, atau perusahaan *prop trading*. Kepentingan finansial apa pun wajib diungkapkan di awal sesi.
6. **Wajib menyampaikan** bahwa sebagian besar strategi yang lolos seluruh pemeriksaan dalam modul ini **tetap akan gagal di masa depan**, karena lolos pengujian berarti belum terbantah, bukan terbukti.

10. Penafian

Materi ini disusun untuk tujuan pendidikan dan peningkatan literasi keuangan. Materi ini **bukan** nasihat investasi, bukan rekomendasi untuk membeli, menjual, atau menahan instrumen keuangan apa pun, dan bukan ajakan untuk melakukan transaksi.

Seluruh backtest, nilai Sharpe, CAGR, drawdown, dan hasil simulasi dalam modul ini bersifat **ilustratif dan dihasilkan dari data simulasi** untuk keperluan pengajaran. Angka-angka tersebut tidak mewakili kinerja strategi, sistem, produk, atau penyedia nyata mana pun.

Backtest yang disajikan pada Contoh 4 adalah konstruksi ajar. Ia tidak merujuk pada penyedia, sistem, atau produk nyata mana pun, dan kemiripan dengan penyajian yang benar-benar ada adalah kebetulan.

Sebagaimana ditunjukkan sepanjang modul ini, **kinerja dalam pengujian historis tidak memberikan dasar yang memadai untuk memperkirakan kinerja masa depan**, dan strategi yang lolos seluruh pemeriksaan dalam modul ini tetap dapat gagal. Lolos pengujian berarti sebuah klaim belum terbantah; ia tidak berarti klaim itu terbukti.

Kinerja masa lalu tidak menjamin hasil di masa depan. Investasi dan perdagangan instrumen keuangan mengandung risiko kerugian, yang pada instrumen dengan leverage dapat melebihi modal yang ditempatkan.

Invesmentpedia dan fasilitatornya tidak memberikan nasihat investasi yang dipersonalisasi dan tidak berizin sebagai penasihat investasi. Untuk kebutuhan nasihat yang disesuaikan dengan keadaan pribadi Anda, hubungi pihak yang memiliki izin dari otoritas yang berwenang.

Keputusan investasi dan perdagangan sepenuhnya menjadi tanggung jawab masing-masing peserta.

Modul L3.4 — Kurikulum Invesmentpedia · Level 3 Pengelola Portofolio dan Sistem (penutup modul inti)