JIIC: JURNAL INTELEK INSAN CENDIKIA
https://jicnusantara.com/index.php/jiic
Vol : 3 No: 11, November 2026
E-ISSN : 3047-7824

Penilaian Kinerja Portofolio melalui Ukuran Return yang Disesuaikan dengan Risiko dan Model Multifaktor

Portfolio Performance Evaluation through Risk Adjusted Return Measures and Multifactor Models

Khairul Bashar[1*], Mufti[2], Irmaningsih[3], Kairunnisa[4]
Program Studi Magister Akuntansi, Ekonomi dan Bisnis, Universitas Malikussaleh
Email : khairul.256210101017@mhs.unimal.ac.id[1], mufti.256210101014@mhs.unimal.ac.id[2], irmaningsih.256210101012@mhs.unimal.ac.id[3], khairunnisa.256210101008@mhs.unimal.ac.id[4]*

Article Info

Article history :
Received : 10-11-2026
Revised : 12-11-2026
Accepted : 14-11-2026
Published : 16-11-2026

Abstract

Portfolio performance evaluation is an essential component of investment management because it assesses whether realized returns adequately compensate investors for the risks undertaken. This review maps conceptual and empirical approaches to portfolio evaluation, compares risk-adjusted performance measures, identifies performance determinants, and synthesizes evidence from international and Indonesian capital markets. This structured review uses an initial literature mapping and a proposed systematic review protocol informed by PRISMA 2020; database searches and full-text screening remain to be completed. The literature is organized around portfolio theory and CAPM; Sharpe, Treynor, Jensen's alpha, M², Sortino, and information ratios; multifactor models; diversification and benchmarks; performance persistence; and Indonesian empirical evidence. Classical studies establish the importance of risk-adjusted returns, while contemporary research highlights benchmark sensitivity, omitted risk factors, fees, survivorship bias, and market characteristics. Indonesian evidence indicates that evaluation outcomes vary across measures, periods, asset classes, and benchmarks. Key research gaps include inconsistent methodologies, limited cross-market-cycle testing, uneven adoption of multifactor models, and insufficient integration of downside risk, costs, and Sharia-compliant investment characteristics. The review concludes that portfolio evaluation should employ complementary measures, appropriate benchmarks, and robustness tests.

Keywords: mutual funds, portfolio performance, risk-adjusted return

Abstrak

Penilaian kinerja portofolio merupakan bagian penting dari proses manajemen investasi karena membantu investor dan manajer investasi menilai apakah tingkat pengembalian yang diperoleh sepadan dengan risiko yang ditanggung. Literature review ini bertujuan memetakan pendekatan konseptual dan empiris dalam evaluasi kinerja portofolio, membandingkan ukuran risk-adjusted return, mengidentifikasi faktor yang memengaruhi kinerja, serta menelaah bukti dari pasar modal internasional dan Indonesia. Kajian menggunakan pemetaan awal literatur dan rancangan protokol SLR dengan acuan PRISMA 2020; penelusuran database dan seleksi teks lengkap belum selesai. Literatur dikelompokkan ke dalam tema teori portofolio dan CAPM; ukuran Sharpe, Treynor, Jensen's alpha, M², Sortino, dan information ratio; model multifaktor; diversifikasi dan benchmark; persistensi kinerja; serta bukti empiris Indonesia. Literatur klasik menunjukkan bahwa return perlu dinilai relatif terhadap risiko, sementara riset kontemporer menekankan sensitivitas hasil terhadap benchmark, faktor risiko yang dihilangkan, biaya, survivorship bias, dan karakteristik pasar. Bukti Indonesia memperlihatkan bahwa hasil evaluasi dapat berbeda menurut metode, periode, jenis aset, dan indeks pembanding. Review ini mengidentifikasi kesenjangan berupa keterbatasan konsistensi metodologi, minimnya pengujian lintas siklus pasar, penggunaan model multifaktor yang belum merata, serta perlunya integrasi risiko downside, biaya, dan karakteristik syariah. Kajian menyimpulkan

bahwa evaluasi portofolio sebaiknya menggunakan beberapa ukuran yang saling melengkapi, benchmark yang sesuai, dan pengujian ketahanan hasil.

Kata kunci: kinerja portofolio, reksa dana, return yang disesuaikan dengan risiko

PENDAHULUAN

Investasi pada dasarnya melibatkan keputusan alokasi dana ke dalam aset dengan karakteristik return, risiko, likuiditas, dan horizon waktu yang berbeda. Investor tidak cukup hanya mengetahui besarnya return nominal; investor juga perlu menilai risiko yang diambil untuk memperoleh return tersebut. Portofolio yang menghasilkan return tinggi belum tentu berkinerja lebih baik apabila return itu dicapai melalui volatilitas atau risiko sistematis yang jauh lebih besar. Karena itu, penilaian kinerja portofolio (portfolio performance evaluation) menjadi bagian integral dari proses perencanaan, pembentukan, pemantauan, dan evaluasi investasi.

Landasan modern pengelolaan portofolio berangkat dari gagasan diversifikasi Markowitz (1952), yang menunjukkan bahwa risiko portofolio ditentukan bukan hanya oleh risiko masing-masing aset, tetapi juga oleh kovarians atau hubungan pergerakan antar-aset. Perkembangan Capital Asset Pricing Model (CAPM) kemudian menyediakan kerangka untuk menghubungkan expected return dengan risiko sistematis (Sharpe, 1964). Berdasarkan kerangka tersebut, sejumlah ukuran kinerja dikembangkan untuk membandingkan return portofolio dengan risiko yang ditanggung, antara lain Sharpe ratio, Treynor ratio, dan Jensen's alpha.

Dalam praktik, tidak ada satu ukuran yang selalu tepat untuk seluruh portofolio. Sharpe ratio menggunakan risiko total sehingga relevan bagi investor yang memegang portofolio sebagai keseluruhan. Treynor ratio menggunakan beta dan lebih tepat ketika portofolio telah terdiversifikasi dengan baik. Jensen's alpha mengestimasi kelebihan return relatif terhadap return yang diprediksi model CAPM. Perbedaan konsep tersebut dapat menghasilkan peringkat portofolio yang tidak sama. Literatur selanjutnya memperluas pengukuran melalui M^2, Sortino ratio, information ratio, model multifaktor, serta pendekatan berbasis holdings dan market timing (Elton & Gruber, 2020).

Penilaian juga menghadapi persoalan metodologis. Pemilihan benchmark yang tidak sesuai dapat mengubah kesimpulan tentang outperformance. Penggunaan CAPM satu faktor dapat mengatribusikan premi faktor lain sebagai alpha. Biaya pengelolaan, turnover, biaya transaksi, dan survivorship bias juga dapat memengaruhi hasil. Carhart (1997), misalnya, menunjukkan bahwa faktor umum return saham dan biaya investasi menjelaskan sebagian besar persistensi kinerja reksa dana yang sebelumnya terlihat sebagai bukti kemampuan manajer.

Konteks Indonesia menarik untuk dikaji karena pasar modal berkembang, struktur investor berubah, serta terdapat variasi produk investasi konvensional dan syariah. Penelitian domestik telah menerapkan ukuran Sharpe, Treynor, Jensen, M^2, dan Sortino pada reksa dana saham, saham indeks, sektor industri, serta indeks syariah. Namun, periode pengamatan, sampel, benchmark, frekuensi data, dan model yang digunakan berbeda-beda. Pangestuti et al. (2017), misalnya, menggunakan enam ukuran kinerja untuk reksa dana saham Indonesia, sedangkan Riandini dan Risman (2022) membandingkan tiga ukuran klasik pada periode 2018–2020.

Permasalahan kajian mencakup perkembangan teori dan metode evaluasi portofolio; perbedaan konseptual ukuran Sharpe, Treynor, Jensen's alpha, M^2, Sortino, dan information ratio;

pengaruh diversifikasi, benchmark, biaya, dan kondisi pasar; bukti empiris internasional dan Indonesia; serta kesenjangan penelitian yang masih terbuka.

Secara akademik, kajian mengintegrasikan teori portofolio, model penetapan harga aset, dan pengukuran kinerja dalam satu kerangka. Secara praktis, review memberi panduan bahwa pemilihan ukuran perlu disesuaikan dengan karakteristik portofolio, tingkat diversifikasi, benchmark, horizon evaluasi, dan tujuan investor. Bagi bidang akuntansi, evaluasi kinerja juga relevan dengan transparansi pelaporan, pengukuran hasil pengelolaan dana, akuntabilitas manajer investasi, serta komunikasi risiko kepada pemangku kepentingan.

Review ini bertujuan: (1) mengidentifikasi pendekatan teoritis dan empiris dalam penilaian kinerja portofolio; (2) membandingkan ukuran risk-adjusted return; (3) menyintesis faktor yang memengaruhi kinerja; (4) memetakan bukti empiris internasional dan Indonesia; serta (5) merumuskan kesenjangan penelitian dan rekomendasi metodologis.

METODE PENELITIAN

Kajian ini merupakan literature review terstruktur berbasis penelusuran awal, dengan rancangan Systematic Literature Review (SLR) dan PRISMA 2020 sebagai acuan pelaporan. Pemetaan substantif telah disusun, tetapi pencarian multi-database, deduplikasi, dan penilaian teks lengkap belum diselesaikan. Oleh karena itu, hasil kajian diperlakukan sebagai sintesis awal, bukan SLR final.

Kajian dirancang sebagai Systematic Literature Review (SLR) dengan PRISMA 2020 sebagai pedoman pelaporan. PRISMA 2020 menyediakan checklist dan flow diagram untuk meningkatkan transparansi identifikasi, penyaringan, penilaian kelayakan, dan inklusi studi (Page et al., 2021). PRISMA merupakan pedoman pelaporan, bukan pengganti keputusan metodologis tentang database, kriteria eligibilitas, atau teknik sintesis.

Fokus review adalah penilaian kinerja portofolio saham dan reksa dana, terutama ukuran risk-adjusted return dan model atribusi kinerja. Studi tentang portofolio obligasi atau aset campuran dapat dimasukkan jika relevan dengan pengukuran kinerja portofolio secara umum. Artikel yang membahas hanya prediksi harga saham tanpa evaluasi kinerja portofolio tidak menjadi fokus utama.

Pertanyaan Review (PICO/PEO)

Tabel 1. Kerangka Pertanyaan Review

Elemen	Definisi operasional
Population / konteks	Portofolio investasi, reksa dana, dana saham, indeks saham, portofolio konvensional dan syariah.
Exposure / konsep	Metode pengukuran kinerja, risk-adjusted return, diversifikasi, benchmark, market timing, selectivity, persistensi.
Comparator	Benchmark pasar, kelompok dana lain, metode evaluasi alternatif, atau periode pasar yang berbeda.
Outcomes	Return, volatilitas, beta, alpha, peringkat kinerja, outperformance/underperformance, persistensi.

Sumber Informasi

Untuk pelaksanaan final, pencarian disarankan pada Scopus, Web of Science Core Collection, ScienceDirect, SpringerLink, Wiley Online Library, Taylor & Francis, Emerald, Google Scholar, GARUDA, dan portal SINTA. Google Scholar dan GARUDA berfungsi sebagai sumber penemuan tambahan, sedangkan status indeks jurnal harus diverifikasi secara terpisah melalui Scopus Sources dan SINTA. Penelusuran awal untuk draf ini juga menggunakan mesin pencarian literatur akademik Consensus untuk menemukan artikel dan abstrak relevan.

Tanggal pencarian final perlu dicatat saat penulis benar-benar menjalankan pencarian database. Rentang utama yang disarankan adalah 2010–2026 untuk menangkap perkembangan kontemporer, dengan pengecualian artikel seminal sebelum 2010 seperti Markowitz (1952), Sharpe (1964, 1966), Jensen (1968), Treynor dan Mazuy (1966), serta Carhart (1997).

Strategi Pencarian

String pencarian perlu disesuaikan dengan sintaks masing-masing database. Contoh string utama:

("portfolio performance" OR "mutual fund performance" OR "investment performance") AND ("Sharpe ratio" OR "Treynor ratio" OR "Jensen alpha" OR "Sortino ratio" OR "M2" OR "information ratio" OR "risk-adjusted return")

String tambahan untuk konteks Indonesia:

("kinerja portofolio" OR "kinerja reksa dana" OR "kinerja saham") AND (Sharpe OR Treynor OR Jensen OR Sortino OR "risk adjusted") AND (Indonesia OR Indonesian)

Kriteria Inklusi dan Eksklusi

Tabel 2. Kriteria Inklusi dan Eksklusi

Kriteria	Inklusi	Eksklusi
Jenis publikasi	Artikel jurnal peer-reviewed; review; studi empiris kuantitatif; SLR/bibliometrik relevan.	Berita, opini, blog, materi promosi, skripsi yang tidak diterbitkan sebagai artikel, abstrak tanpa informasi memadai.
Substansi	Mengukur atau mengulas kinerja portofolio/reksa dana dengan return dan/atau risiko.	Hanya membahas prediksi harga, sentimen, atau pemilihan saham tanpa evaluasi portofolio.
Bahasa	Bahasa Inggris atau Indonesia.	Bahasa lain jika tidak tersedia terjemahan yang dapat diverifikasi.
Ketersediaan	Metadata, abstrak, dan informasi bibliografis dapat diverifikasi; full text diupayakan.	Rekaman duplikat atau metadata yang tidak dapat diverifikasi.
Periode	Utama 2010–2026; artikel klasik seminal tanpa batas tahun.	Di luar periode utama dan tidak memiliki kontribusi teori/metode yang jelas.

Kriteria	Inklusi	Eksklusi
Indeksasi	Scopus, WoS, SINTA, atau jurnal ilmiah bereputasi yang statusnya dapat diperiksa.	Klaim indeksasi tidak dapat dibuktikan atau jurnal bermasalah.

HASIL DAN PEMBAHASAN

Teori Portofolio Modern

Markowitz (1952) memperkenalkan kerangka pemilihan portofolio berdasarkan trade-off antara expected return dan varians. Investor dapat mengurangi risiko tidak sistematis melalui kombinasi aset yang korelasinya tidak sempurna. Diversifikasi tidak menghilangkan risiko pasar secara keseluruhan, tetapi dapat mengurangi risiko spesifik aset. Konsep efficient frontier menggambarkan kombinasi portofolio yang memberikan return tertinggi untuk tingkat risiko tertentu atau risiko terendah untuk return tertentu.

Implikasi bagi evaluasi kinerja adalah bahwa return satu aset atau portofolio harus dibaca bersama kontribusi risiko dan diversifikasinya. Portofolio yang tidak terdiversifikasi dapat memiliki risiko total tinggi meskipun beta pasar rendah. Oleh karena itu, penggunaan beta sebagai satu-satunya ukuran risiko perlu dipertimbangkan secara hati-hati.

CAPM dan Risiko Sistematis

CAPM menghubungkan expected return aset dengan tingkat bebas risiko dan premi risiko pasar. Secara umum, model dapat ditulis $E(Ri) = Rf + \beta i[E(Rm) - Rf]$. Beta menunjukkan sensitivitas return aset terhadap perubahan return pasar. Dalam evaluasi portofolio, CAPM menjadi dasar Jensen's alpha dan Treynor ratio. Namun, CAPM merupakan model teoritis yang bergantung pada asumsi tertentu, termasuk ekspektasi homogen, pasar modal ideal, dan hubungan linear antara return dan beta.

Return dan Risiko

Return portofolio dapat dihitung sebagai perubahan nilai portofolio ditambah distribusi kas selama periode pengukuran, disesuaikan dengan nilai awal. Untuk portofolio berbobot tetap, return periode dapat dihitung sebagai jumlah bobot aset dikalikan return aset. Pengukuran harus konsisten antara return time-weighted atau money-weighted, frekuensi data, perlakuan dividen, pajak, biaya, serta perubahan komposisi portofolio.

Risiko total sering diukur dengan varians atau standar deviasi return. Risiko sistematis diukur dengan beta terhadap benchmark. Risiko downside berfokus pada penyimpangan return di bawah target minimum yang dapat diterima. Pemilihan ukuran risiko harus sesuai dengan karakter distribusi return dan preferensi investor.

Kinerja Absolut dan Relatif

Kinerja absolut menilai return atau alpha portofolio tanpa semata-mata membandingkan peringkat antarportofolio. Kinerja relatif membandingkan hasil terhadap benchmark atau kelompok pembanding. Keduanya menjawab pertanyaan berbeda: apakah portofolio memberikan hasil positif dan memadai, serta apakah portofolio mengungguli alternatif yang relevan. Evaluasi yang kuat sebaiknya menyampaikan kedua dimensi tersebut.

Evolusi Pendekatan Penilaian Kinerja

Literatur penilaian kinerja berkembang dari ukuran return relatif sederhana menuju ukuran risk-adjusted, model faktor, serta pendekatan yang menilai holdings dan kemampuan timing. Markowitz (1952) menyediakan fondasi diversifikasi dan trade-off return-risiko. Sharpe (1964) mengembangkan CAPM, sementara Sharpe (1966) memperkenalkan ukuran reward-to-variability untuk mengevaluasi reksa dana. Ukuran Treynor menekankan return relatif terhadap risiko pasar, sedangkan Jensen (1968) mengukur kemampuan menghasilkan return di atas prediksi CAPM.

Elton dan Gruber (2020) merangkum perkembangan model evaluasi reksa dana, mulai single-index dan multi-index models, pengukuran portofolio pasif, ukuran berbasis holdings, market timing, hingga evaluasi dana obligasi. Mereka menekankan bahwa data, bias, faktor yang tidak dimasukkan, dan kualitas benchmark merupakan isu sentral dalam interpretasi kinerja.

Sharpe Ratio

Sharpe ratio mengukur excess return per unit risiko total. Rumusnya adalah:

$$\text{Sharpe}_p = (R_p - R_f) / \sigma_p$$

R_p adalah return portofolio, R_f adalah return bebas risiko, dan σ_p adalah standar deviasi return portofolio. Semakin tinggi nilai Sharpe, semakin besar excess return yang dihasilkan untuk setiap unit volatilitas. Ukuran ini berguna ketika investor menilai portofolio sebagai investasi tunggal atau ketika portofolio belum sepenuhnya terdiversifikasi.

Keterbatasannya adalah penggunaan standar deviasi yang memperlakukan volatilitas positif dan negatif secara simetris. Distribusi return yang menceng, berekor tebal, atau mengandung downside risk besar dapat membuat interpretasi Sharpe kurang memadai. Nilai Sharpe juga sensitif terhadap periode, frekuensi observasi, dan pemilihan risk-free rate.

Treynor Ratio

$$\text{Treynor}_p = (R_p - R_f) / \beta_p$$

Treynor ratio membandingkan excess return dengan risiko sistematis yang diukur melalui beta. Ukuran ini lebih sesuai untuk portofolio yang telah terdiversifikasi sehingga risiko spesifik telah ditekan. Namun, beta bergantung pada benchmark dan periode estimasi. Ketika beta mendekati nol atau negatif, rasio dapat menjadi tidak stabil atau sulit ditafsirkan.

Jensen's Alpha

$$R_p - R_f = \alpha_p + \beta_p(R_m - R_f) + \varepsilon_p$$

Jensen's alpha adalah intersep regresi excess return portofolio terhadap excess return pasar. Alpha positif menunjukkan return di atas prediksi CAPM pada tingkat beta tertentu, sedangkan alpha negatif menunjukkan underperformance relatif terhadap model. Jensen (1968) menerapkan ukuran ini pada reksa dana dan membedakan kemampuan prediksi manajer dari risiko portofolio.

Alpha bukan bukti otomatis atas skill. Alpha dapat timbul akibat benchmark yang kurang tepat, faktor risiko yang dihilangkan, perubahan eksposur, kesalahan estimasi, atau data yang bias. Karena itu, alpha sebaiknya diuji menggunakan model faktor yang lebih kaya dan periode yang memadai.

M², Sortino, dan Information Ratio

Modigliani dan Modigliani (1997) mengembangkan ukuran M² dengan mengubah volatilitas portofolio agar setara dengan volatilitas benchmark, sehingga kinerja dapat dibaca dalam satuan return. M² memudahkan komunikasi kepada investor karena hasilnya dapat dibandingkan langsung dengan return pasar. Sortino ratio mengganti standar deviasi dengan downside deviation, sehingga lebih fokus pada return di bawah target minimum yang ditentukan. Information ratio membandingkan active return terhadap tracking error dan lazim digunakan untuk menilai konsistensi pengelolaan aktif relatif terhadap benchmark.

Tabel 4. Perbandingan Ukuran Kinerja Portofolio

Ukuran	Pembilang	Penyebut/penyesuaian	Kegunaan utama
Sharpe	Excess return	Standar deviasi total	Portofolio sebagai keseluruhan
Treynor	Excess return	Beta pasar	Portofolio terdiversifikasi
Jensen alpha	Abnormal return	Estimasi CAPM/regresi	Selektivitas relatif terhadap model
M²	Return setara risiko	Volatilitas disetarakan dengan pasar	Komunikasi selisih return
Sortino	Excess return di atas target	Downside deviation	Risiko penurunan
Information ratio	Active return	Tracking error	Pengelolaan aktif terhadap benchmark

Model Multifaktor

Model CAPM satu faktor dapat gagal menangkap eksposur terhadap ukuran perusahaan, value, momentum, kualitas, likuiditas, atau faktor lain. Fama dan French (1993) mengembangkan model tiga faktor dengan faktor pasar, size, dan value. Carhart (1997) menambahkan faktor momentum menjadi model empat faktor dalam evaluasi persistensi reksa dana. Model multifaktor membantu memisahkan alpha manajer dari kompensasi atas eksposur terhadap faktor sistematis.

Penggunaan model multifaktor perlu mempertimbangkan relevansi faktor terhadap pasar dan jenis portofolio. Faktor yang dibangun dari pasar Amerika Serikat tidak selalu dapat dipindahkan secara langsung ke Indonesia tanpa pengujian. Peneliti perlu menjelaskan definisi faktor, sumber data, frekuensi, pembentukan portofolio faktor, dan risiko multikolinearitas.

Diversifikasi, Benchmark, dan Biaya

Diversifikasi memengaruhi pilihan ukuran risiko. Ketika portofolio belum terdiversifikasi, standar deviasi lebih representatif untuk risiko total. Ketika portofolio telah terdiversifikasi, beta dan tracking error dapat menjadi lebih relevan, tergantung tujuan evaluasi. Benchmark harus mencerminkan universe investasi, gaya, ukuran, mata uang, dan batasan portofolio. Membandingkan reksa dana saham dengan indeks yang tidak sesuai dapat menghasilkan kesimpulan yang bias.

Biaya merupakan determinan penting dari net performance. Carhart (1997) menunjukkan bahwa expense ratio dan turnover berhubungan negatif dengan kinerja dan bahwa faktor return saham serta biaya menjelaskan sebagian besar persistensi yang diamati. Dengan demikian, evaluasi sebaiknya menyajikan gross return dan net return apabila data memungkinkan, serta membedakan biaya manajemen, biaya transaksi, pajak, dan biaya distribusi.

Persistensi Kinerja dan Skill Manajer

Persistensi berarti kinerja relatif suatu portofolio pada periode sebelumnya berkaitan dengan kinerja pada periode berikutnya. Persistensi jangka pendek tidak serta-merta membuktikan skill karena dapat dipengaruhi momentum, eksposur faktor, biaya, survivorship, dan keberuntungan. Carhart (1997) menunjukkan bahwa sebagian besar persistensi return reksa dana dapat dijelaskan oleh faktor umum dan biaya, sementara underperformance yang persisten terutama terkonsentrasi pada dana dengan kinerja terburuk.

Penelitian lanjutan perlu membedakan persistensi return mentah, persistensi risk-adjusted alpha, dan persistensi peringkat. Uji out-of-sample, survivorship-bias-free data, dan pengendalian multiple testing diperlukan untuk mengurangi risiko kesimpulan yang terlalu optimistis.

Bukti Internasional

Bukti internasional memperlihatkan bahwa hasil evaluasi dipengaruhi oleh model dan karakteristik pasar. Carhart (1997), menggunakan data reksa dana saham Amerika Serikat yang memperhitungkan dana yang tidak bertahan, menemukan bahwa faktor return saham dan biaya investasi menjelaskan hampir seluruh persistensi kinerja rata-rata dan risk-adjusted return. Temuan tersebut menantang interpretasi sederhana bahwa dana yang menjadi pemenang pada satu periode pasti dikelola oleh manajer dengan skill superior.

Elton dan Gruber (2020) menekankan bahwa evaluasi modern tidak cukup berhenti pada single-factor alpha. Pengukuran perlu memperhatikan model multi-index, passive portfolio, holdings-based measures, market timing, dan kualitas benchmark. Artikel ini merupakan review konseptual yang penting untuk menjelaskan perkembangan metode dan isu pengukuran.

Kiymaz dan Simsek (2017) menilai reksa dana pasar berkembang yang berbasis di Amerika Serikat. Mereka melaporkan bahwa sebagian dana saham emerging markets menghasilkan alpha signifikan pada periode tertentu, tetapi hasilnya tidak konsisten antar-subperiode; dana obligasi emerging markets umumnya tidak menghasilkan alpha positif signifikan. Temuan ini menunjukkan pentingnya analisis subperiode dan perbedaan kelas aset.

Borges (2021) membandingkan Sharpe, Sortino, Jensen, dan Treynor pada 1.901 reksa dana saham Brasil. Penelitian tersebut melaporkan hubungan positif antarukuran dan tidak menemukan perbedaan signifikan pada peringkat kinerja yang dihasilkan. Hasil ini memberi indikasi bahwa beberapa ukuran dapat memberikan urutan yang serupa pada konteks tertentu, tetapi tidak berarti ukuran-ukuran tersebut identik secara konseptual.

Ali et al. (2021) mengevaluasi reksa dana saham di Pakistan menggunakan beberapa ukuran risiko, termasuk Ulcer/ Martin ratio. Rata-rata ukuran risk-adjusted yang digunakan dilaporkan negatif, dan penelitian tidak menemukan bukti kuat bahwa manajer mengungguli pasar. Malhotra et al. (2023) pada reksa dana sektor keuangan juga menemukan bahwa meskipun beberapa ukuran

kinerja tampak relatif tinggi terhadap benchmark tertentu, alpha tidak berbeda dari nol dan tidak terdapat bukti kemampuan timing atau seleksi yang konsisten.

Bukti Indonesia

Pangestuti et al. (2017) mengevaluasi 42 reksa dana saham Indonesia menggunakan Sharpe, Treynor, Jensen, Adjusted Sharpe, Adjusted Jensen, dan Sortino. Studi tersebut menyimpulkan bahwa SAM Indonesian Equity merupakan reksa dana dengan kinerja terbaik dalam periode kajian dan bahwa sebagian besar reksa dana yang dianalisis telah terdiversifikasi dengan baik. Kontribusi penting studi ini adalah memperluas evaluasi di luar tiga ukuran klasik.

Qur'anitasari et al. (2019) menguji portofolio saham LQ-45 yang dibentuk dengan single index model dan membandingkan hasil Sharpe, Treynor, serta Jensen. Ketiganya menunjukkan pola pergerakan yang relatif serupa dan berfluktuasi. Penulis menekankan perlunya ketelitian dalam menentukan slope dan melakukan robustness test; dalam konteks sampel penelitian, Sharpe dinilai paling sesuai.

Andreas dan Basana (2021) mengevaluasi reksa dana saham Indonesia menggunakan Sharpe, Treynor, Jensen, dan M^2 dengan benchmark LQ-45 dan IHSG. Jumlah dana yang mengungguli benchmark berubah menurut metode dan indeks pembanding. Temuan tersebut menegaskan bahwa penentuan benchmark bukan keputusan teknis kecil, melainkan bagian substantif dari kesimpulan evaluasi.

Sari dan Risman (2021) membandingkan empat metode pada reksa dana saham Indonesia periode 2016–2018. Hasil uji mereka menunjukkan perbedaan antara return reksa dana dan IHSG serta perbedaan konsistensi hasil di antara metode Sharpe, Treynor, Jensen, dan M^2. Hal ini memperkuat alasan untuk melaporkan lebih dari satu ukuran.

Riandini dan Risman (2022) menguji tujuh reksa dana saham Indonesia periode 2018–2020 dengan Sharpe, Treynor, dan Jensen. Dalam sampel yang digunakan, Sucorinvest Equity dilaporkan memiliki kinerja terbaik. Ukuran sampel yang relatif kecil dan periode yang mencakup gejolak pasar mengingatkan bahwa generalisasi harus dilakukan secara hati-hati.

Hilmy dan Shafira (2022) menerapkan Sharpe, Treynor, dan Sortino pada konstituen IDX BUMN20 periode 2018–2021. Penelitian melaporkan bahwa kinerja yang diukur didominasi nilai negatif dan belum mampu melampaui instrumen bebas risiko yang digunakan. Studi ini memperlihatkan relevansi downside risk dan evaluasi indeks sektoral.

Azizah et al. (2024) membandingkan indeks saham syariah Indonesia dan Malaysia selama dan setelah pandemi COVID-19 menggunakan Sharpe, Treynor, dan Jensen. Hasilnya menunjukkan bahwa kesimpulan perbandingan dapat berbeda menurut ukuran, dan perbedaan kinerja antarindeks tidak selalu signifikan secara statistik. Abidah et al. (2024) membandingkan portofolio saham bank konvensional dan syariah Indonesia periode 2021–2023 dan melaporkan kinerja portofolio bank konvensional lebih tinggi pada tiga ukuran yang digunakan. Kedua studi memperlihatkan bahwa jenis portofolio dan konteks syariah merupakan tema penting, tetapi kesimpulan tidak dapat digeneralisasi ke seluruh saham syariah atau konvensional.

Matriks Sintesis Studi Terpilih

Tabel 5. Matriks Sintesis Studi Terpilih

Penelitian	Konteks dan metode	Temuan yang dilaporkan	Implikasi/keterbatasan
Carhart (1997)	AS; reksa dana saham; CAPM dan 4 faktor; kontrol survivorship	Faktor umum, biaya, dan transaksi menjelaskan sebagian besar persistensi.	Penting mengendalikan faktor dan biaya; hasil pasar AS tidak otomatis berlaku di Indonesia.
Elton & Gruber (2020)	Review metode kinerja reksa dana	Membahas single/multi-index, holdings, timing, data, bias, benchmark.	Sumber konseptual utama; bukan meta-analysis kuantitatif.
Kiymaz & Simsek (2017)	Dana emerging markets berbasis AS; alpha multifaktor	Alpha dana saham berubah antar-subperiode; dana obligasi umumnya tidak menghasilkan alpha positif signifikan.	Pentingnya subperiode dan perbedaan kelas aset.
Borges (2021)	1.901 reksa dana saham Brasil; empat rasio	Empat ukuran berkorelasi positif dan peringkat relatif tidak berbeda signifikan.	Kesamaan peringkat bersifat kontekstual.
Pangestuti et al. (2017)	42 reksa dana saham Indonesia; enam ukuran	SAM Indonesian Equity dilaporkan terbaik; mayoritas dana dinilai terdiversifikasi.	Contoh penggunaan ukuran lebih luas di Indonesia.
Qur'anitasari et al. (2019)	Saham LQ-45; Sharpe, Treynor, Jensen	Pola ketiga metode relatif serupa; Sharpe dinilai paling sesuai pada sampel.	Perlu robustness dan lebih banyak metode.
Andreas & Basana (2021)	Reksa dana saham Indonesia; Sharpe, Treynor, Jensen, M^2	Jumlah outperformer berubah sesuai benchmark dan metode.	Benchmark memengaruhi kesimpulan.
Sari & Risman (2021)	Reksa dana Indonesia; empat metode	Ditemukan perbedaan konsistensi hasil antarukuran.	Perlu pelaporan multi-metrik.
Riandini & Risman (2022)	Tujuh reksa dana saham; 2018–2020	Sucorinvest Equity terbaik pada sampel yang dikaji.	Sampel kecil membatasi generalisasi.
Hilmy & Shafira (2022)	IDX BUMN20; Sharpe, Treynor, Sortino	Nilai kinerja didominasi negatif terhadap acuan yang digunakan.	Downside risk perlu diperhatikan.

Penelitian	Konteks dan metode	Temuan yang dilaporkan	Implikasi/keterbatasan
Azizah et al. (2024)	JII-70 dan FBMS; tiga ukuran	Kesimpulan relatif berubah menurut ukuran; perbedaan tidak selalu signifikan.	Konteks lintas negara dan syariah penting.
Abidah et al. (2024)	Bank konvensional dan syariah; tiga ukuran	Portofolio bank konvensional unggul dalam sampel 2021–2023.	Sampel terbatas; hindari generalisasi luas.

Sintesis Lintas Studi

Pertama, literatur konsisten menunjukkan bahwa evaluasi kinerja harus mempertimbangkan risiko. Return mentah tidak cukup untuk menyimpulkan kualitas pengelolaan portofolio. Kedua, tidak ada ukuran universal: Sharpe, Treynor, dan Jensen memiliki basis risiko dan interpretasi berbeda. Ketiga, benchmark dan model faktor memengaruhi estimasi alpha serta jumlah portofolio yang dinilai outperform. Keempat, biaya dan bias data dapat menjelaskan sebagian persistensi kinerja. Kelima, bukti Indonesia mulai berkembang ke penggunaan Sortino, M^2, dan perbandingan syariah, tetapi masih terdapat ruang untuk memperluas penggunaan model multifaktor, data lebih panjang, dan robustness checks.

Hasil lintas studi tidak dapat disatukan sebagai satu estimasi kuantitatif karena perbedaan definisi sampel, periode, benchmark, metode pembentukan portofolio, dan ukuran outcome. Karena itu, sintesis naratif merupakan pendekatan yang lebih tepat untuk draf ini. Meta-analysis dapat menjadi tahap lanjutan apabila studi menyediakan effect size yang sebanding, misalnya korelasi antara ukuran kinerja, selisih rata-rata return yang telah distandardisasi, atau estimasi alpha dengan standard error yang dapat dikonversi secara sah.

Interpretasi Hasil dan Implikasi

Mengapa Ukuran Kinerja Memberikan Hasil Berbeda?

Perbedaan hasil muncul karena ukuran kinerja tidak mengukur konsep risiko yang sama. Sharpe menggunakan volatilitas total; Treynor menggunakan beta pasar; Jensen mengukur intercept model CAPM; Sortino menitikberatkan downside deviation; dan information ratio menilai active return relatif terhadap tracking error. Portofolio yang mempunyai risiko spesifik besar dapat terlihat kurang menarik menurut Sharpe tetapi relatif baik menurut Treynor jika beta pasarnya rendah. Sebaliknya, portofolio dengan tracking error tinggi dapat memiliki information ratio rendah meskipun return absolutnya positif.

Perbedaan juga muncul karena benchmark. Beta, alpha, active return, dan tracking error merupakan ukuran yang bergantung pada indeks pembanding. Benchmark yang tidak mewakili gaya investasi dapat menghasilkan kesimpulan yang keliru. Oleh karena itu, peneliti harus menjelaskan alasan pemilihan IHSG, LQ-45, IDX30, indeks sektoral, indeks syariah, atau benchmark lain.

Apakah Alpha Menunjukkan Kemampuan Manajer?

Alpha positif merupakan indikasi bahwa return melampaui prediksi model, bukan bukti kausal bahwa manajer memiliki kemampuan superior. Model dapat tidak lengkap, faktor risiko tertentu tidak dimasukkan, atau hasil dipengaruhi keberuntungan. Pengujian skill memerlukan model faktor yang sesuai, data bebas survivorship bias, pengendalian biaya, uji stabilitas lintas periode, dan validasi out-of-sample. Carhart (1997) memperlihatkan pentingnya kehati-hatian dalam menginterpretasikan persistensi.

Implikasi bagi Investor dan Manajer Investasi

Investor ritel sebaiknya tidak memilih produk hanya berdasarkan return satu tahun atau satu rasio. Evaluasi dapat dimulai dengan Sharpe dan Sortino untuk risiko total dan downside, kemudian dilengkapi alpha atau information ratio jika benchmark relevan tersedia. Investor perlu memeriksa konsistensi kinerja lintas periode, biaya, likuiditas, konsentrasi aset, dan kesesuaian profil risiko.

Manajer investasi perlu menyajikan kinerja secara transparan dengan benchmark yang tepat, menjelaskan risiko, biaya, perubahan strategi, dan periode pengukuran. Penggunaan beberapa ukuran mengurangi risiko komunikasi yang menonjolkan indikator yang paling menguntungkan saja.

Implikasi bagi Akuntansi dan Tata Kelola

Dalam perspektif Magister Akuntansi, evaluasi kinerja portofolio berkaitan dengan akuntabilitas pengelolaan aset, pengukuran hasil investasi, pengungkapan risiko, dan kualitas informasi bagi investor. Penyajian return tanpa risiko dan benchmark dapat memberi gambaran yang tidak lengkap. Pelaporan yang baik perlu membedakan return sebelum dan sesudah biaya, mengungkap metode pengukuran, serta konsisten dalam penggunaan benchmark dan periode.

Relevansi untuk Pasar Modal Indonesia

Pasar Indonesia memiliki karakteristik likuiditas, konsentrasi sektor, partisipasi investor ritel, serta variasi produk syariah dan konvensional yang perlu dipertimbangkan dalam desain penelitian. Penggunaan indeks pasar luas saja mungkin tidak cukup untuk menilai dana dengan mandat sektoral atau syariah. Selain itu, periode pandemi dan perubahan suku bunga dapat mengubah volatilitas serta korelasi antar-aset. Penelitian perlu membandingkan kondisi normal dan stres pasar tanpa memilih periode secara oportunistik.

RESEARCH GAP DAN AGENDA PENELITIAN

Tabel 6. Kesenjangan dan Agenda Penelitian

Research gap	Kondisi literatur	Agenda penelitian yang disarankan
1. Dominasi ukuran klasik	Banyak studi Indonesia masih mengandalkan Sharpe, Treynor, Jensen.	Bandingkan ukuran klasik dengan Sortino, M^2, information ratio, dan model multifaktor.
2. Ketergantungan benchmark	Kesimpulan dapat berubah saat benchmark diganti.	Uji beberapa benchmark yang sesuai dengan mandat portofolio dan lakukan robustness test.

Research gap	Kondisi literatur	Agenda penelitian yang disarankan
3. Risiko downside belum merata	Standar deviasi dan beta masih dominan.	Gunakan downside deviation, maximum drawdown, expected shortfall, dan Ulcer Index.
4. Periode dan siklus pasar	Banyak studi memakai periode relatif pendek atau satu rezim pasar.	Gunakan data panjang dan bandingkan bull, bear, pandemi, serta periode pemulihan.
5. Bias dan biaya	Tidak semua studi melaporkan survivorship bias, biaya transaksi, dan turnover.	Gunakan data survivorship-bias-free serta return neto biaya.
6. Skill dan persistensi	Persistensi return sering dipakai sebagai proksi skill.	Pisahkan persistensi return, alpha, dan peringkat; lakukan uji out-of-sample.
7. Konteks syariah	Studi komparatif masih terbatas dan sampel sering sempit.	Bandingkan portofolio syariah dan konvensional dengan kontrol sektor, ukuran, dan faktor risiko.
8. Meta-analysis	Heterogenitas metode dan pelaporan statistik menghambat pooling.	Bangun dataset effect size dan standard error yang harmonis sebelum meta-analysis.

Usulan Model Penelitian Lanjutan

Salah satu rancangan penelitian yang dapat dikembangkan adalah evaluasi kinerja reksa dana saham Indonesia menggunakan pendekatan multi-metrik dan multifaktor. Variabel dependen dapat berupa Jensen alpha atau ukuran risk-adjusted return; variabel penjelas dapat mencakup expense ratio, turnover, ukuran dana (AUM), usia dana, arus dana, konsentrasi portofolio, dan kondisi pasar. Analisis dapat menggunakan panel data dengan fixed effects atau random effects sesuai hasil pengujian, serta robust standard errors.

Untuk menguji ketahanan, penelitian dapat membandingkan CAPM, Fama–French tiga faktor, dan Carhart empat faktor; membedakan return gross dan net; menggunakan IHSG dan benchmark kategori; serta memisahkan subperiode sebelum, selama, dan setelah pandemi. Jika menilai persistensi, gunakan ranking portofolio dan estimasi alpha out-of-sample, bukan hanya korelasi return masa lalu dan masa depan.

Keterbatasan Review

Draf ini belum melaporkan angka PRISMA final karena proses pencarian sistematis multi-database, deduplikasi, dan penilaian full text belum dilakukan secara lengkap. Beberapa artikel teridentifikasi melalui metadata dan abstrak; status indeksasi jurnal perlu diverifikasi pada tahun publikasi. Selain itu, perbedaan konteks dan statistik antarpenelitian membatasi kemungkinan penggabungan kuantitatif. Keterbatasan tersebut harus dinyatakan secara terbuka pada naskah final dan tidak boleh ditutupi dengan angka seleksi yang tidak terdokumentasi.

KESIMPULAN

Penilaian kinerja portofolio merupakan proses evaluasi return dengan mempertimbangkan risiko, benchmark, diversifikasi, biaya, dan horizon investasi. Sharpe, Treynor, dan Jensen tetap

menjadi ukuran dasar, tetapi masing-masing memiliki asumsi dan keterbatasan. M^2, Sortino, information ratio, dan model multifaktor memperluas analisis sesuai kebutuhan evaluasi. Literatur internasional menegaskan bahwa persistensi kinerja tidak selalu mencerminkan skill manajer karena dapat dijelaskan oleh eksposur faktor dan biaya. Bukti Indonesia menunjukkan bahwa peringkat dan kesimpulan kinerja dapat berubah menurut metode, benchmark, jenis aset, dan periode pengamatan.

Dengan demikian, evaluasi yang memadai sebaiknya menggunakan beberapa ukuran yang saling melengkapi, benchmark yang sesuai dengan mandat portofolio, pengendalian biaya dan bias, serta pengujian robustness. Research gap utama terletak pada perlunya penggunaan model multifaktor yang lebih luas, pengukuran risiko downside, data jangka panjang lintas siklus pasar, transparansi biaya, dan pengembangan basis data yang memungkinkan meta-analysis. Kajian ini adalah pemetaan awal berbasis literatur yang teridentifikasi; klaim SLR final sesuai PRISMA 2020 memerlukan penyelesaian pencarian database, screening full text, dan pelaporan angka flow diagram.

DAFTAR PUSTAKA

Abidah, A. N., Pratiwi, P., Oftafiana, T., & Aswad, M. (2024). Comparison of stock portfolio performance of conventional banks and Islamic banks using Sharpe ratio, Treynor ratio, and Jensen ratio (2021–2023). *Jurnal Ilmu Ekonomi Terapan, 9(2)*. https://doi.org/10.20473/jiet.v9i2.59272

Ali, M. A., Aqil, M., Alam, S. H., & Zaman, S. (2021). Evaluation of risk adjusted performance of mutual funds in an emerging market. *International Journal of Finance & Economics*. https://doi.org/10.1002/ijfe.2486

Andreas, A., & Basana, S. R. (2021). Analysis of the performance of Indonesian mutual stock funds using Sharpe, Treynor, Jensen and M2 method period 2010–2019. *International Journal of Financial and Investment Studies, 2(1), 1–9*. https://doi.org/10.9744/ijfis.2.1.1-9

Azizah, F., Handayani, S. S., & Rizaldy, M. (2024). Comparative analysis of Indonesia and Malaysia Sharia stock index performance using Sharpe, Treynor, and Jensen methods. *International Journal of Islamic Business and Economics, 8(1)*. https://doi.org/10.28918/ijibec.v8i1.7188

Borges, D. M. (2021). Evaluation of investment funds through different performance measures. *Revista Gestão da Produção Operações e Sistemas, 16(4)*. https://doi.org/10.15675/gepros.v16i4.2787

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance, 52(1), 57–82*. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Elton, E. J., & Gruber, M. J. (2020). A review of the performance measurement of long-term mutual funds. *Financial Analysts Journal, 76(3), 22–37*. https://doi.org/10.1080/0015198X.2020.1738126

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33(1), 3–56*. https://doi.org/10.1016/0304-405X(93)90023-5

Hilmy, A., & Shafira, L. V. (2022). IDX BUMN20 performance measurement with Sharpe, Treynor, and Sortino. *JEM17: Jurnal Ekonomi Manajemen, 7(1)*. https://doi.org/10.30996/jem17.v7i1.6510

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance, 23(2), 389–416*. https://doi.org/10.1111/j.1540-6261.1968.tb00815.x

Kiymaz, H., & Simsek, K. D. (2017). The performance of US-based emerging market mutual funds. *Journal of Capital Markets Studies, 1(1), 58–73*. https://doi.org/10.1108/JCMS-10-2017-003

Leković, M. (2017). Mutual funds portfolio performance evaluation models: Sharpe, Treynor and Jensen index. *Bankarstvo, 46(4), 108–133*. https://doi.org/10.5937/bankarstvo17041081

Malhotra, D., Mooney, T. J., Poteau, R. R., & Russel, P. (2023). Assessing the performance and risk-adjusted returns of financial mutual funds. *International Journal of Financial Studies, 11(4), 136*. https://doi.org/10.3390/ijfs11040136

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7(1), 77–91*. https://doi.org/10.2307/2975974

Modigliani, F., & Modigliani, L. (1997). Risk-adjusted performance. *The Journal of Portfolio Management, 23(2), 45–54*. https://doi.org/10.3905/jpm.1997.409608

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372, n71*. https://doi.org/10.1136/bmj.n71

Pangestuti, I. R. D., Wahyudi, S., & Robiyanto. (2017). Performance evaluation of equity mutual funds in Indonesia. *Jurnal Keuangan dan Perbankan, 21(4), 527–542*. https://doi.org/10.26905/jkdp.v21i4.1503

Qur'anitasari, Q., Nuzula, N. F., & Darmawan, A. (2019). Critical analysis of Sharpe, Treynor and Jensen methods in analyzing stock portfolio performance: LQ-45 stock studies. *Asia Pacific Management and Business Application, 8(2), 89–104*. https://doi.org/10.21776/ub.apmba.2019.008.02.2

Riandini, A. S., & Risman, A. (2022). Evaluation of equity mutual funds portfolio performance in Indonesia using Sharpe, Treynor and Jensen methods. *The EUrASEANs: Journal on Global Socio-Economic Dynamics, 3(34), 34–40*. https://doi.org/10.35678/2539-5645.3(34).2022.34-40

Sari, I. P., & Risman, A. (2021). Assessment of mutual fund performance using the calculation methods of Sharpe ratio, Treynor ratio, Jensen alpha, and M-square (Study on Indonesia mutual fund for the 2016–2018). *Sumerianz Journal of Business Management and Marketing, 4(4), 138–145*. https://doi.org/10.47752/sjbmm.44.138.145

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19(3), 425–442*. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business, 39(1), 119–138*. https://doi.org/10.1086/294846

Treynor, J. L., & Mazuy, K. K. (1966). Can mutual funds outguess the market? Harvard Business Review, 44(4), 131–136.

Yuliana, R., Siwiyanti, L., Purnomo, B. S., & Purnamasari, I. (2024). Determinants of investment portfolio performance: A systematic literature review. *The Eastasouth Management and Business, 2(2)*. https://doi.org/10.58812/esmb.v2i02.184