On the Importance of Evaluating Foundational Time-Series Models One Step Ahead in an Investment Context

Peter Cotton

November 2026

Abstract

Popular benchmarks for foundational time-series models pool forecast errors over long horizons. Under frictionless log utility, an investor who can trade every interval in a liquid market does not use those horizons: the optimal position depends only on the conditional distribution of the next return, and the investor's expected log-wealth regret from trading on a forecast is bounded by the forecast's excess expected log loss, the relative entropy $D_{\mathrm{KL}}(P\|Q)$ of the truth P from the forecast Q. With transaction costs, in the quadratic model of Gârleanu and Pedersen, later horizons enter through the expected-return path with weights that shift toward the next interval as costs fall. The one-step predictive is therefore the object to evaluate for that investor, its excess expected log loss controls the investor's regret without ranking imperfect forecasts' decisions, and a score pooled over horizons does not establish superiority for that decision. Two examples mark the limits: a model better at one step can be worse at twelve while iterating its own dynamics, and a forecast closer to the truth in relative entropy can induce the worse position. We recommend reporting one step ahead on its own and including an iterated one-step model among multi-step baselines.

1 Setting

A trader observes a liquid security at intervals (a day, an hour, a tick) and can trade once per interval at the quoted price. There is no market impact and, until Section 2.3, no transaction cost. Write P_t for the price, $r_{t+1} = P_{t+1}/P_t - 1$ for the return over the next interval, $\mathcal{F}_t$ for the trader's information at t, and x_t for the fraction of wealth held from t to $t+1$. We call the interval between trades the *decision interval*.

Does a forecast of $r_{t+2}, r_{t+3}, \ldots$ change the choice of x_t? If it does not, a benchmark that scores such forecasts does not by itself establish performance for today's decision. The contribution is deliberately small. The sufficiency result is Mossin, Samuelson, Merton and Hakansson; the regret bound is Barron and Cover; the hedging demand is Merton; the cost-weighted aim portfolio is Gârleanu and Pedersen; the comparison of direct and iterated forecasts is Marcellino, Stock and Watson. The case for evaluating forecasts by the decision they serve is Granger and Pesaran (2000) and Pesaran and Skouras (2002). We line these up against the way multi-step forecasts of foundational time-series models are scored, keeping apart three questions: which distribution determines the optimal action, whether iterated or direct multi-step forecasts are better, and which loss evaluates a forecast for a stated decision.

2 Which distribution determines the position

2.1 Log utility

Wealth evolves as $W_{t+1} = W_t(1 + x_t r_{t+1})$ and the trader maximises $\mathbb{E}[\log W_T]$. Assume positive wealth, integrability of $\log(1 + x r_{t+1})$ on the feasible set, a feasible set $x_t \in \mathcal{X}_t(s_t)$ that does not depend on wealth and whose future members are unaffected by current choices (so no turnover constraint couples x_t to x_{t+1}), and an exogenous opportunity process: a state s_t adapted to $\mathcal{F}_t$ such that the conditional law of (r_{t+1}, s_{t+1}) given $\mathcal{F}_t$ depends on $\mathcal{F}_t$ only through s_t and is unaffected by the trader's choices. The dynamic programme is

$$V_t(W_t, s_t) = \max_{x_t} \mathbb{E}\big[\, V_{t+1}(W_{t+1}, s_{t+1}) \,\big|\, s_t \big]. \tag{1}$$

Proposition 1 (The next return's law is sufficient; Mossin 1968, Samuelson 1969, Merton 1969)**.** *Under log utility and the assumptions above, $V_t(W, s) = \log W + g_t(s)$ for some g_t, and*

$$x_t^* = \arg\max_x \mathbb{E}\big[\log(1 + x r_{t+1}) \,\big|\, s_t\big].$$

Proof. Induct backward from $V_T = \log W$. If $V_{t+1} = \log W + g_{t+1}(s)$ then

$$\mathbb{E}[V_{t+1}(W_{t+1}, s_{t+1}) \mid s_t] = \log W_t + \mathbb{E}[\log(1 + x_t r_{t+1}) \mid s_t] + \mathbb{E}[g_{t+1}(s_{t+1}) \mid s_t],$$

and because the opportunity process is exogenous only the middle term depends on x_t. Hence $g_t(s) = \max_x \mathbb{E}[\log(1 + x r_{t+1}) \mid s] + \mathbb{E}[g_{t+1}(s_{t+1}) \mid s]$. □

In English: the trader collects any predictability in later returns by trading when those intervals arrive, so today's position only has to be right about r_{t+1}. The state must be sufficient for the continuation value as well as for the next return's law. Holding the conditional law of r_{t+1} and the feasible set fixed, predictability of r_{t+12} from $\mathcal{F}_t$ changes g_t and leaves the argmax alone; a signal about r_{t+12} that also moves the conditional law of r_{t+1} enters through that law. Kelly sizing (Kelly, 1956) is the same statement with r a bet.

2.2 Other utilities

For $u(W) = W^{1-\gamma}/(1-\gamma)$ with $\gamma \neq 1$ the value function factorises as $V_t(W, s) = W^{1-\gamma} H_t(s)$. When the conditional investment opportunity set is deterministic, for example independent returns with the filtration they generate and deterministic feasible sets, H_{t+1} is constant and the problem is again one-period (Samuelson, 1969). Independence of the realised returns is not enough under a larger filtration: a signal that is informative about later returns in some states and not in others makes H_{t+1} vary with the state, and the position hedges against the less informed branch even though the returns themselves are independent. When returns are predictable, $H_{t+1}(s_{t+1})$ can covary with r_{t+1} and Merton (1971, 1973) obtain, in continuous time with one state variable,

$$x_t^* = \frac{\mu_t - r^f}{\gamma \sigma_t^2} \;+\; \frac{\rho\, \sigma_s}{\sigma_t} \, \frac{-J_{Ws}}{W J_{WW}},$$

where ρ is the correlation between the return and the state innovation and J is the value function. The second term is the hedging demand. It vanishes for $\gamma = 1$ and for $\rho = 0$, and it depends on how the opportunity set moves, through J_{Ws}.

Remark 1 (Frequent trading does not shrink the hedging demand)**.** *The decision interval and the remaining horizon $T-t$ are different quantities. Merton's formulation already allows continuous trading, and the hedging term is a function of the state dynamics with no factor in the trading interval. Campbell and Viceira (1999) find in calibrated models with persistent expected returns that the hedging demand may double average stock demand for investors with risk aversion above one, with a material loss from a myopic policy; Hakansson (1971) shows that with serially correlated returns only log utility is myopic in general. The term can also be large relative to a myopic position that is itself near zero. The exact statement of this section is for log utility; for other utilities the frictionless position depends on the law of motion of the opportunity set.*

2.3 Transaction costs

Add a cost of trading and keep the trader once per interval in a liquid market. We switch to a tractable quadratic model and to one security. Gârleanu and Pedersen (2013) study a trader holding n_t shares with trading cost $\frac{1}{2}\lambda\sigma^2(n_t-n_{t-1})^2$, mean-variance preferences with risk aversion γ over the excess price change ΔP_{t+1} (net of financing; equivalently take the risk-free rate to be zero), variance σ^2, discount rate ρ, and expected excess price changes driven by mean-reverting factors. Write $m_h = \mathbb{E}_t[\text{Markowitz}_{t+h}]$ with $\text{Markowitz}_{t+h} = \mathbb{E}_{t+h}[\Delta P_{t+h+1}]/(\gamma\sigma^2)$, so that forecast lead $j = h+1$. The optimal policy moves the holding a fixed fraction of the way toward an *aim portfolio*,

$$n_t = n_{t-1} + \frac{a}{\lambda}\,(\text{aim}_t - n_{t-1}), \qquad \text{aim}_t = \sum_{h\geq 0} z(1-z)^h\, m_h, \tag{2}$$

with

$$a = \frac{-\big(\gamma(1-\rho)+\lambda\rho\big) + \sqrt{\big(\gamma(1-\rho)+\lambda\rho\big)^2 + 4\gamma\lambda(1-\rho)^2}}{2(1-\rho)}, \qquad z = \frac{\gamma}{\gamma+a}.$$

The model's n_t is a share holding and our x_t a wealth fraction, so we quote (2) in its own units.

In English: the expected excess price change at every lead enters today's trade, weighted by how long the position will be carried, and the weights depend on the cost only through λ/γ. Within this model, holding the other parameters fixed, lowering the cost shifts the weight distribution toward the nearer leads. Figure 1 plots the weights by lead. Their half-life is 0.28 leads at $\lambda/\gamma = 0.1$, 0.72 at 1, 2.2 at 10 and 6.9 at 100; the mass beyond lead twelve is 0.001 percent at $\lambda/\gamma = 1$, 2.3 percent at 10 and 30 percent at 100. The figure is a sensitivity analysis; we do not map a particular security and interval to a value of λ/γ.

Concentrated weights bound the absolute contribution of later leads and nothing more. Since $\text{aim}_t - m_0 = \sum_{h\geq 1} z(1-z)^h(m_h-m_0)$, a bound $\sup_{h\geq 1}|m_h-m_0| \leq C$ gives $|\text{aim}_t - m_0| \leq C(1-z)$. When m_0 is near zero the relative error need not be small: two predictive factors can cancel today and decay at different rates, leaving the whole aim to later leads. Other objectives need not admit fixed horizon weights at all; with proportional costs, constraints or nonlinear utility the policy can involve state-dependent trading regions.

The weights are a policy, and they do not convert into a per-horizon forecast score. If the policy is $x^* = c\sum_h w_h\mu_h$ and a forecaster supplies $\hat{\mu}_h = \mu_h + e_h$, the action error is $c\sum_h w_h e_h$, and for an objective quadratic in the action the regret is proportional to

$$\Big(\sum_h w_h e_h\Big)^2 = \sum_{h,j} w_h w_j\, e_h e_j,$$

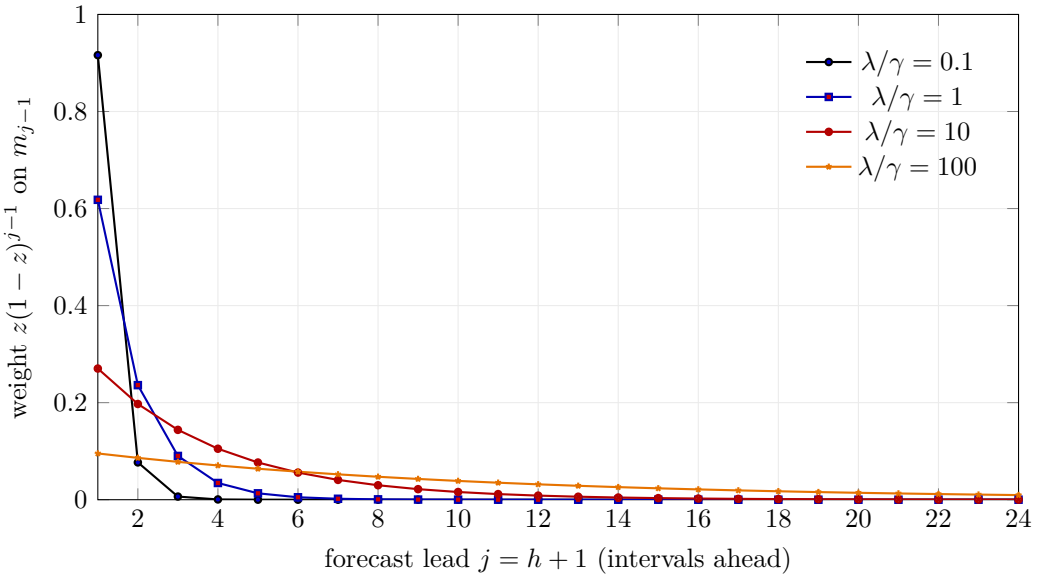

Figure 1: Weight of each forecast lead's expected excess price change in the aim portfolio of equation (2), per-interval discount rate $\rho = 0.0004$ (about ten percent a year at daily intervals). Lower cost puts the weight on the first lead; at $\lambda/\gamma = 1$ the first two leads carry 85 percent of it.

a sum of cross-horizon products in which errors cancel or reinforce. A decision loss has to be derived from the stated policy and objective; a weighted sum of separate horizon losses is a different object.

2.4 Excluded constraints and payoffs

A lock-up or a monthly mandate fixes a holding period, and the decision then depends on the cumulative return $R_{t,k} = \prod_{j=1}^{k}(1 + r_{t+j}) - 1$; a position too large to unwind in one interval imposes a liquidation schedule instead, which is a path problem. A futures expiry fixes neither; the position can be closed before it. Horizon risk limits and derivatives are compatible with trading once per interval and attach future-dependent valuation or constraints to today's choice; drawdown and barrier contracts depend on path functionals rather than on cumulative-return marginals. We exclude these constraints and payoffs.

3 Direct and iterated multi-step forecasts

3.1 What an iterated one-step model requires

The joint law of the next k returns given $\mathcal{F}_t$ factorises as

$$p(r_{t+1}, \ldots, r_{t+k} \mid \mathcal{F}_t) = \prod_{h=1}^{k} p\big(r_{t+h} \,\big|\, \mathcal{F}_t, r_{t+1}, \ldots, r_{t+h-1}\big).$$

A one-step model supplies the first factor at the current state. To supply the rest it must be a generative model with specified state dynamics: it must say how its state moves when a return is

appended, and it must model, or integrate out, any exogenous information that arrives between t and $t+h$. A one-step predictive at the present state and a complete transition model are different objects. A forecaster can score well when updated with realised observations and badly when iterated on its own simulated ones.

The chain rule also removes an apparent dichotomy. Any joint law admits an autoregressive factorisation, including a law fitted jointly across horizons; and separately fitted horizon marginals do not by themselves specify a joint law.

3.2 No ordering across horizons

A correct transition model gives correct forecasts at every horizon. Among imperfect models no ordering follows, with or without Markov structure in the truth.

Example 1 (Better at one step, worse at twelve)**.** *Let returns take the values $\pm a$ as the states of a two-state chain with transition matrix $M(p)$, p the probability of staying, so that* $\Pr(X_{t+h} = X_t) = \frac{1}{2}\big(1 + (2p-1)^h\big)$*. Let the truth be $M(0.9)$, model A be $M(0.8)$ and model B be $M(0.1)$, each iterating its own matrix. At one step A is far better: its excess expected log loss is about* 0.037 *against* 1.758 *for B. At twelve steps B is exactly right, because $M(0.1)^{12} = M(0.9)^{12}$.*

Whether a direct h-step forecast beats an iterated one-step model is therefore a question about estimation efficiency and misspecification. Marcellino et al. (2006) find iterated forecasts usually better on macroeconomic panels, more so at long horizons, and direct forecasts more robust when the one-step model is misspecified; Chevillon (2007) surveys the conditions on each side. It has to be measured.

3.3 Three targets

Three objects need separate names: a single interval's return r_{t+h}, the cumulative return $R_{t,h}$, and the path $(r_{t+1}, \dots, r_{t+h})$. A model that is better at r_{t+12} is not thereby better for a twelve-interval holding decision, which depends on $R_{t,12}$. Good cumulative-return marginals at several horizons do not establish that a model captures path dependence; different joint laws share those marginals and imply different drawdown risks.

4 Forecast quality and decision quality

Proposition 1 names the distribution that determines the optimal action. Two further facts fix what a score of that distribution does and does not tell the trader. Among misspecified forecasts the ranking depends on the loss used to make it (Patton, 2020); the two facts make that concrete for this trader.

Proposition 2 (Relative entropy bounds the log-utility regret; Barron and Cover 1988)**.** *Let P be the true conditional law of r and Q a forecast with $P \ll Q$ and $D_{\mathrm{KL}}(P\|Q) < \infty$, with densities p and q taken relative to a common dominating measure; let the feasible set of fractions be convex and common to both, with $1 + xr > 0$ on it; and let the expectations below be finite. With $x_P \in \arg\max_x \mathbb{E}_P \log(1+xr)$ and $x_Q \in \arg\max_x \mathbb{E}_Q \log(1+xr)$,*

$$0 \le \mathbb{E}_P\left[\log \frac{1 + x_P r}{1 + x_Q r}\right] \le D_{\mathrm{KL}}(P \,\|\, Q).$$

Proof. Optimality of x_Q under Q over a convex set gives the first-order condition $\mathbb{E}_Q\big[(1 + xr)/(1 + x_Q r)\big] \le 1$ for every feasible x, in particular for x_P. Then

$$\begin{aligned}\mathbb{E}_P\left[\log \frac{1 + x_P r}{1 + x_Q r}\right] &= D_{\mathrm{KL}}(P\|Q) + \mathbb{E}_P\left[\log\Big(\frac{q}{p}\,\frac{1 + x_P r}{1 + x_Q r}\Big)\right] \\ &\le D_{\mathrm{KL}}(P\|Q) + \log \mathbb{E}_Q\left[\frac{1 + x_P r}{1 + x_Q r}\right] \;\le\; D_{\mathrm{KL}}(P\|Q),\end{aligned}$$

by Jensen's inequality and the first-order condition. The lower bound is optimality of x_P under P. □

In English: the expected log-wealth lost over one interval by trading on a forecast is at most the forecast's excess expected log loss under the true law, $L_P(Q) - L_P(P) = D_{\mathrm{KL}}(P\|Q)$ with $L_P(Q) = -\mathbb{E}_P \log q(r)$. The raw expected loss $L_P(Q)$ does not bound anything: it can be negative for a density, it depends on the units of r, and $L_P(P)$ is unknown. A smaller excess loss tightens the bound; an observed benchmark difference estimates the improvement with sampling error. Over many intervals in the same frictionless exogenous setting the regret in terminal log wealth is bounded by the sum of the expected conditional divergences, $\mathbb{E}\log(W_T^*/W_T^Q) \le \sum_{t<T} \mathbb{E}[D_{\mathrm{KL}}(P_t\|Q_t)]$. None of this ranks decisions.

Example 2 (The better forecast induces the worse position)**.** *Let returns take the values* $\{-0.2, 0, 0.2\}$ *with truth* $(0.20, 0.50, 0.30)$*, forecast* $A = (0.21, 0.50, 0.29)$ *and forecast* $B = (0.32, 0.20, 0.48)$*. The log-optimal fraction is* $x^*(q) = \big(q(0.2) - q(-0.2)\big)/\big(0.2\,[q(0.2) + q(-0.2)]\big)$*, which equals* 1.00 *under the truth,* 0.80 *under* A *and* 1.00 *under* B*, while* $D_{\mathrm{KL}}(P\|A) \approx 0.0004$ *and* $D_{\mathrm{KL}}(P\|B) \approx 0.223$*. Proposition 2 bounds* A*'s regret by* 0.0004124*, which* A *attains, and* B*'s by* 0.223*;* B*'s regret is zero.*

The bound is specific to logarithmic scoring under these conditions and says nothing about other proper scores. The same distinction applies to a model B that is the better forecaster at twelve steps while A is better at one. Let $\mathcal{G}_t \subseteq \mathcal{F}_t$ be the information model A uses. B's output is $\mathcal{F}_t$-measurable, so it adds nothing to the oracle law $p(r_{t+1} \mid \mathcal{F}_t)$. It carries no information beyond $\mathcal{G}_t$ when the law of r_{t+1} given $\mathcal{G}_t$ and B's output equals the law given $\mathcal{G}_t$ alone, which is a statement about an investor who already holds the exact $\mathcal{G}_t$-conditional law; an imperfect implementation of that law can still be improved by B's output. Mixture weights chosen by predictive log score are likewise a forecasting choice; a trader with a stated objective would in general choose others.

5 Testing a one-step model against a multi-step one

Three constructions use a one-step model A with a multi-step model B and can be tested for improvement without retraining either. Each is evaluated at the information set it uses.

A portfolio per horizon. At each h, mix the two predictive marginals with weights that follow realised score at that horizon. This needs A's marginal at every horizon, so A must be iterated as a generative model (Section 3.1) or supplied with separate horizon forecasts; we iterate. The weights report where each model is better. The mixed marginals do not specify a joint path law, so the construction is evaluated on marginals unless a dependence structure is added.

Marginal replacement. If B emits sample paths from a joint law p_B with first-step marginal $p_{B,1}$, give each path the importance weight $p_A(x_1)/p_{B,1}(x_1)$. The weighted law is

$$q(x_{1:H}) = p_B(x_{1:H})\,\frac{p_A(x_1)}{p_{B,1}(x_1)} = p_A(x_1)\,p_B(x_{2:H} \mid x_1),$$

A's first-step marginal with B's conditional law of the rest of the path given the first observation. It preserves that conditional law and need not preserve B's copula or unconditional correlations. Under the true law P, when the expectations are well defined and finite,

$$\mathbb{E}_P\big[\log q(X_{1:H}) - \log p_B(X_{1:H})\big] = \mathbb{E}_P\big[\log p_A(X_1) - \log p_{B,1}(X_1)\big],$$

so the path log score improves by exactly the first-step improvement, when there is one. The weighted law is the information projection of p_B onto the laws with first marginal p_A (Csiszár, 1975), and the identity is its Pythagorean equality; Krüger et al. (2017) tilt path forecasts toward external nowcasts in the same spirit, with moment targets. The identity says nothing about later marginals, cumulative-return scores or utility. The construction needs $p_{B,1}$ to be positive wherever p_A is, evaluable first-step densities, and control of weight degeneracy; it is made at time t.

Revision by the one-step innovation. A stale h-step forecast made at t can be revised, once r_{t+1} is seen, by the regression of its error on the surprise $r_{t+1} - \mathbb{E}_t^A[r_{t+1}]$. The revised forecast is an object of time $t+1$ and is compared with competitors that hold the same information at $t+1$ for the same target date. Rerunning B at $t+1$ makes the revision redundant only if the rerun absorbs the new observation adequately.

The experiment is then: B alone, A iterated alone, and each construction, at identical origins and information sets, scored on r_{t+h} per horizon, with cumulative-return and path-functional scores for the constructions that supply a joint law.

6 Benchmarks of foundational time-series models

GIFT-Eval (Aksu et al., 2024) evaluates on non-overlapping rolling windows in the test portion of each series, one window in some configurations and several in others; averages a quantile loss over nine levels and over every position of a window that spans several seasonal periods; normalises each metric against a seasonal naive per configuration; and reports average ranks, with the public leaderboard aggregating the normalised metrics across configurations by geometric mean. The suites surveyed in Cotton (2026) share the pooling of positions within a window.

For the once-per-interval trader in a liquid market four statements hold.

1. One step ahead at the decision interval should be reported on its own. For the log-utility trader the one-step predictive is the only input to the decision, and by Proposition 2 its excess expected log loss bounds the trader's regret. The decision interval belongs to the trader; a daily dataset scored at $h = 1$ speaks to a trader who trades daily.

2. An equal-weight average over horizons has economic readings. An unnormalised sum of pinball losses is the objective of a collection of quantile decisions with additive asymmetric costs at different dates, and a uniformly distributed compulsory liquidation date gives an equal-weight average of horizon utilities. Neither reading is this trader's, so a pooled headline does not establish superiority for this trader.

3. The multi-step baseline should include an iterated one-step model with specified state dynamics, compared at identical origins and information on r_{t+h}, and on $R_{t,h}$ and path functionals where the models supply a joint law. Section 3.2 is the reason the comparison has to be run.

4. Where a decision is named, its regret belongs next to the score. Proposition 2 bounds it for the log-utility trader; Section 2.3 derives policy weights and shows why they do not convert into a per-horizon score.

This is why the `skaters` package scores `laplace(k=1)` and why its comparison with foundational time-series models on https://foundationalprediction.com is made one step ahead and never pooled across horizons. The package's twelve-step limit and its decimated multi-scale fans are engineering choices documented in the package; nothing here derives them. We leave to future work the experiment of Section 5 on the canonical panel, with predictive scores next to the regret of a stated trading rule.

Disclosure

Drafting was assisted by Anthropic's Claude. The two examples, the marginal-replacement identity and the pointer to the Barron–Cover bound came from reviews produced by OpenAI's ChatGPT. The author directed and checked the work.

References

Taha Aksu, Gerald Woo, Juncheng Liu, Xu Liu, Chenghao Liu, Silvio Savarese, Caiming Xiong, and Doyen Sahoo. GIFT-Eval: a benchmark for general time series forecasting model evaluation. arXiv:2410.10393, 2024.

Andrew R. Barron and Thomas M. Cover. A bound on the financial value of information. *IEEE Transactions on Information Theory*, 34(5):1097–1100, 1988.

John Y. Campbell and Luis M. Viceira. Consumption and portfolio decisions when expected returns are time varying. *Quarterly Journal of Economics*, 114(2):433–495, 1999.

Guillaume Chevillon. Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21(4):746–785, 2007. doi: 10.1111/j.1467-6419.2007.00518.x.

Peter Cotton. A stronger univariate benchmark for foundational time-series model evaluation. SSRN, https://skaters.microprediction.org/papers.html, 2026.

Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3(1):146–158, 1975.

Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic trading with predictable returns and transaction costs. *Journal of Finance*, 68(6):2309–2340, 2013.

Clive W. J. Granger and M. Hashem Pesaran. Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19(7):537–560, 2000.

Nils H. Hakansson. On optimal myopic portfolio policies, with and without serial correlation of yields. *Journal of Business*, 44(3):324–334, 1971.

John L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35(4): 917–926, 1956.

Fabian Krüger, Todd E. Clark, and Francesco Ravazzolo. Using entropic tilting to combine BVAR forecasts with external nowcasts. *Journal of Business & Economic Statistics*, 35(3): 470–485, 2017.

Massimiliano Marcellino, James H. Stock, and Mark W. Watson. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135(1–2):499–526, 2006.

Robert C. Merton. Lifetime portfolio selection under uncertainty: the continuous-time case. *Review of Economics and Statistics*, 51(3):247–257, 1969.

Robert C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413, 1971.

Robert C. Merton. An intertemporal capital asset pricing model. *Econometrica*, 41(5):867–887, 1973.

Jan Mossin. Optimal multiperiod portfolio policies. *Journal of Business*, 41(2):215–229, 1968.

Andrew J. Patton. Comparing possibly misspecified forecasts. *Journal of Business & Economic Statistics*, 38(4):796–809, 2020.

M. Hashem Pesaran and Spyros Skouras. Decision-based methods for forecast evaluation. In Michael P. Clements and David F. Hendry, editors, *A Companion to Economic Forecasting*, pages 241–267. Blackwell, 2002.

Paul A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, 51(3):239–246, 1969.