

The Sign-Power Generalized Forecast Functional: A Copula-Theoretic Approach of Nonlinear Predictors

Dimitrios Thomakos*

July 18, 2026

Abstract

We study the one-step-ahead forecast functional $f_{t+1|t} = a + b \operatorname{sign}(x_{t-d}) |x_{t-d}|^g$, a four-parameter family that nests the random walk, the linear lag- d predictor, and a continuum of odd power-law transformations of a single lagged observation. Coined the *Sign-Power Generalized Forecast* (SPGF), this functional is motivated both by its inherent simplicity and by the potential failure of linearity for financial time series, volatility proxies, and macro-financial indicators exhibiting skewness, heavy tails, and asymmetric serial dependence. We derive the population-optimal parameters (a^*, b^*, d^*, g^*) under squared-error loss, absolute-error loss, and directional sign accuracy for a general strictly stationary first-order Markov process $\{x_t\}$ whose dependence structure is characterised, via the Darsow et al. (1992) $*$ -product, by a one-step copula C_1 and an invariant marginal F . The MSE optimum reduces to a generalised correlation-maximisation problem; the MAE optimum reduces to L_1 median regression with an implicit first-order condition; and the sign-accuracy optimum reduces to a one-dimensional threshold problem whose solution is a concordance functional of the $(d+1)$ -step copula $C_{d+1} = C_1^{*(d+1)}$, invariant in the power exponent g . All three optima are generically distinct, and we give exact conditions for their coincidence. A simulation-consistent estimator is derived for the empirically relevant case in which $C_1^{*(d+1)}$ has no closed form. We apply the framework to financial returns, derive three operational improvements over the baseline SPGF-SIGN rule—a multi-lag logistic model in sign space, a direct copula-conditional probability estimator, and a majority-vote ensemble—and evaluate five sign forecasters in a rolling-window framework reporting MSE, RMSE, MAE, MASE, directional accuracy, Diebold–Mariano and Pesaran–Timmermann tests, information ratio, and cumulative long-short P&L. The results strongly support the theoretical underpinnings and the practical efficacy of the proposed method.

Keywords: nonlinear forecasting; sign-power transform; copula; Darsow–Nguyen–Olsen product; concordance; directional accuracy; median regression; financial returns; Pesaran–Timmermann test; information ratio, simplicity.

JEL codes: C14, C22, C53, C58, G17.

*Professor and Director, Cybernetics and Artificial Intelligence Laboratory (CAIL), Department of Business Administration, National and Kapodistrian University of Athens, Athens, Greece. E-mail: dthomakos@ba.uoa.gr. This is part of AI-assisted research under the CAIL Quality Standard v1.0. The paper presents the theory only and no empirical results as of yet. Please consult my Python implementations either at the Laboratory’s website or my github repository.

1 Introduction

1.1 The potential failure of linearity in empirical forecasting

The standard workhorse of time-series forecasting is the linear, first-order autoregressive, predictor $f_{t+1|t} = a + b x_{t-d}$, with $d = 1$. Under joint Gaussianity the conditional expectation is linear and the minimum-mean-squared-error predictor coincides with the ordinary least squares projection (Whittle, 1983; Brockwell and Davis, 1991). Moreover, under Gaussianity the OLS predictor is simultaneously optimal under MSE, MAE, and every loss function that is a monotone function of the absolute error (Granger, 1969b). This triple optimality collapses as soon as the joint law of (x_{t+1}, x_{t-d}) departs from ellipticity, which it routinely does for the data types where forecasting matters most.

Financial log-returns exhibit four well-documented non-Gaussian features (Mandelbrot, 1963; Fama, 1965; Cont, 2001): heavy tails with tail-index estimates in $[3, 5]$ for major equity indices; volatility clustering and GARCH-type conditional heteroskedasticity (Engle, 1982; Bollerslev, 1986); negative skewness and the leverage effect (Black, 1976; Christie, 1982); and asymmetric tail dependence, whereby large negative shocks are more serially correlated than large positive shocks (Longin and Solnik, 2001; Ang and Bekaert, 2002). Macroeconomic indicators display comparable features, including regime-switching dynamics (Hamilton, 1989), threshold autoregression (Tong, 1990; Tsay, 1989), and sign-asymmetric persistence (Bekaert and Wu, 1998). In all these settings there exists some nonlinear predictor that dominates the linear rule under at least one standard loss function.

The literature on nonlinear time-series forecasting is vast. Early contributions include the threshold autoregressive (TAR) family of Tong (1990), smooth transition autoregressions of Teräsvirta (1994), neural-network predictors (Kaastra and Boyd, 1996; White, 2006), functional-coefficient models (Cai et al., 2000), additive nonparametric regressions (Fan and Yao, 2003), and Bayesian nonparametric approaches (Hjort et al., 2010). The common limitation of these approaches is that they typically optimize a single criterion, usually squared-error loss, and do not characterize in closed form how the optimal predictor changes when the loss function changes. The SPGF framework attempts an intervention in this gap by making loss-function sensitivity itself part of the object being studied rather than an afterthought of empirical comparison.

1.2 The sign-power forecast functional

We study the *Sign-Power Generalized Forecast* functional:

$$f_{t+1|t} = a + b \operatorname{sign}(x_{t-d}) |x_{t-d}|^g, \quad (a, b, d, g) \in \mathbb{R} \times \mathbb{R} \times \mathbb{Z}_{\geq 0} \times \mathbb{R}_{>0} \quad (1)$$

The key ingredient is the sign-power transform

$$z = \text{sign}(x)|x|^g \quad (2)$$

which is an odd, strictly monotone transformation preserving the sign of its argument. At $g = 1$ the SPGF reduces to the linear lag predictor. As $g \rightarrow 0^+$ it converges to a pure sign predictor, retaining directional information while discarding magnitude information. For $g \in (0, 1)$ the transform compresses large lagged observations and provides heavy-tail robustness; for $g > 1$ it amplifies large past moves relative to small ones and can capture momentum or leverage-type effects. The exponent therefore functions as a forecasting analogue of a power-transformation parameter: it alters predictive geometry rather than serving merely as a notational generalisation.

The family in (1) nests many predictors used in practice. The special cases $g = 1$, $a = 0$, $b = 1$ give the naive lag- d predictor; $g = 1$ with OLS (a, b) gives the linear AR(1)-style predictor; and $a = 0$, $b \in \{-1, 1\}$ gives a sign-based contrarian or momentum rule (Harvey et al., 1999; Jegadeesh and Titman, 1993). The SPGF therefore provides a unified parametric umbrella for asking which of these special cases is optimal under each loss function.

The sign-power transform $\text{sign}(x)|x|^g$ has appeared in the literature in two adjacent but distinct contexts, and the present paper is explicitly not a rediscovery of either. Boshnakov and Ravagli (2021) study the same function in the time-series statistics literature, showing that signed power transforms of ARCH-type processes remain ARCH-type and exploiting the transform's sign-preserving and moment-regularising properties to construct white noise tests when fourth moments are absent. Their interest is in the distributional properties of the *transformed outcome process*, not in using the transform as a *predictor ingredient*; no forecasting optimality, copula structure, or loss comparison is involved. On the loss-function side, Elliott et al. (2005) (EKT) define the two-parameter family $\mathcal{L}(e; \alpha, p) = [\alpha + (1 - 2\alpha)\mathbf{1}\{e < 0\}]\lvert e \rvert^p$, where p controls the power of the forecast error and α controls asymmetry. The SPGF exponent g is the analogous object on the *predictor* side: it controls how past magnitudes are reweighted before being used to form the forecast, whereas the EKT p controls how forecast errors are penalised after the forecast is formed. These are genuinely different parameters even when both equal one, and the SPGF framework makes no assumption about the shape of the loss function beyond the three specific criteria studied. Finally, the concordance functional $\kappa_{d+1} = \mathbb{E}[\text{sign}(x_{t+1}) \text{sign}(x_{t-d})]$ that emerges as the solution to the SIGN outer optimum (Theorem 5.4) is precisely the lag- $(d+1)$ zero-quantile autocorrelation measured by the quantilogram of Linton and Whang (2007) and its multivariate extension by Han et al. (2016). The quantilogram is a diagnostic tool for *measuring* directional dependence; the SPGF provides the optimisation-theoretic foundation explaining *why* that quantity is the right object to maximise within the sign-power predictor family.

1.3 The copula approach to serial dependence

The central theoretical tool is the copula representation of the joint law of (x_{t+1}, x_{t-d}) . By Sklar's theorem (Sklar, 1959), this joint law is characterised by the pair (F, C_{d+1}) , where F is the invariant marginal and C_{d+1} is a copula. For a first-order Markov process with one-step copula C_1 , Darsow et al. (1992) established that $C_{d+1} = C_1^{*(d+1)}$, the $(d+1)$ -fold composition under the $*$ -product

$$(A * B)(u, v) = \int_0^1 \partial_2 A(u, t) \partial_1 B(t, v) dt \quad (3)$$

This decomposition makes the SPGF theory tractable: F and C_1 can be estimated separately, and all multi-step dependence is generated by iterating C_1 through (3).

Why a first-order Markov structure suffices. The SPGF forecast functional $f_{t+1|t} = a + b \operatorname{sign}(x_{t-d})|x_{t-d}|^g$ depends on the observed history only through the single variable x_{t-d} and the parameters (a, b, d, g) . Consequently, its optimality under any loss function depends exclusively on the joint law of the pair (x_{t+1}, x_{t-d}) , not on the full joint distribution of the entire process. We therefore place ourselves in the class of strictly stationary first-order Markov processes, for which the joint law of any pair (x_t, x_{t+k}) is completely determined by the one-step copula C_1 and the invariant marginal F via the DNO composition $C_k = C_1^{*k}$. This assumption is not an artificial restriction: it is the minimal dependence structure required to make the problem analytically tractable, and it contains as special cases all linear AR(1) processes, all nonlinear autoregressions of order one, and all copula-based Markov chains that are routinely used in financial and macroeconomic modelling. Any process whose dependence decays with the lag and is generated by a stationary Markov transition kernel fits this framework; higher-order dependence would only affect forecasts that use multiple lags, which are not considered in the single-lag SPGF family. Thus, the Markovian setting is exactly the natural domain for the theory developed here, and it is the setting in which the DNO $*$ -product provides a complete and rigorous characterisation of all multi-step joint laws required for the outer optimisation over (d, g) .

The copula approach to time-series modeling has developed rapidly since Patton (2004) introduced copula-based models for conditional dependence. Key contributions include the semiparametric Markov estimator of Chen and Fan (2006), the comprehensive review of Patton (2012), and extensions to high-dimensional settings such as Brechmann and Czado (2012). The algebraic properties of the Darsow–Nguyen–Olsen product have been studied by Ibragimov (2009), Longla (2012), and Durante and Sempi (2015); see also Joe (2014) for the broader dependence-modeling context. Copulas have been applied to financial return modeling by Jondeau and Rockinger (2006); Rodriguez (2007); Patton (2006), establishing empirically the non-Gaussian and asymmetric tail dependence that motivates the present framework. In the present paper that perspective

is pushed one step further: the copula is not only a device for estimation, but the object through which lag dependence and forecast optimality are linked analytically.

1.4 Directional forecasting

Predicting the sign of returns has attracted sustained interest in financial econometrics. Pesaran and Timmermann (1995) established that market timing ability is equivalent to sign-predictability, while Pesaran and Timmermann (1992) proposed the directional-accuracy test now standard in the literature. Leitch and Tanner (1991) showed that directional accuracy correlates more closely with profitability than MSE, motivating a separate treatment of the sign criterion. Sign prediction has been studied using logistic regression (Anatolyev and Gerko, 2005; Nyberg, 2011), probit models (Kauppi and Saikkonen, 2008), Markov-switching models (Hamilton, 1989; Filardo, 1994), neural networks (Kaastra and Boyd, 1996), and ensemble methods (Timmermann, 2006). Linton and Whang (2007) propose the quantilogram as a correlogram of sign hits for testing directional predictability without specifying an optimal predictor family. None of these approaches derives the optimal sign forecast within a parametric predictor family from first principles under a copula-based dependence model.

Forecast evaluation is conducted here using the Diebold–Mariano test (Diebold and Mariano, 1995; West, 1996) for comparative accuracy, the Pesaran–Timmermann test (Pesaran and Timmermann, 1992) for absolute directional predictability, and economic criteria including the information ratio (Sharpe, 1966; Grinold and Kahn, 1989) and the mean absolute scaled error (Hyndman and Koehler, 2006). The companion Python implementation reports all of these, providing a unified evaluation infrastructure.

1.5 Contributions and outline

This paper makes four contributions.

- (i) *Exact inner optima.* For fixed (d, g) we derive closed-form or implicitly defined optimal (a^*, b^*) under MSE, MAE, and SIGN for an arbitrary stationary first-order Markov process (section 3).
- (ii) *The Darsow et al. DNO outer optimum.* The outer optimisation over (d, g) reduces to integral functionals of $(F, C_1^{*(d+1)}, g)$ via the DNO $*$ -product (section 4).
- (iii) *g -invariance of the SIGN criterion.* The directional-accuracy optimum depends on the lag structure but not on the power exponent g ; under symmetry and sign-monotonicity it reduces to a concordance functional of C_{d+1} (section 5).
- (iv) *Financial applications and operational sign forecasters.* We characterise why financial returns are the natural domain for the general theory, derive three operational improvements over the baseline SPGF-SIGN estimator, and evaluate five sign

forecasters—AR1-Sign, SPGF-Sign, MultiLag-Logistic, CopulaProb, and Ensemble—in a rolling-window backtest with comprehensive statistical and economic evaluation (sections 7 and 8).

The rest of the paper is structured as follows: section 2 establishes notation; section 3 solves the inner problems; section 4 develops the DNO connection; section 5 establishes the structural features of the sign-accuracy problem; section 6 derives the simulation estimator; section 7 develops the financial application and the operational forecasters; section 8 describes the companion implementation and evaluation protocol. section 9 concludes.

2 The forecast functional and the three optimization problems

Let $\{x_t\}_{t \in \mathbb{Z}}$ be a strictly stationary, first-order Markov process (x_{t+1}, x_{t-d}) on $\mathbb{R}$ with invariant marginal CDF F , assumed continuous. The MSE criterion requires finite second moments of y and z ; the MAE and SIGN criteria require weaker conditions (first moments for MAE, none for SIGN), which are stated at the relevant points in the text. The copula-DNO framework and the simulation estimator do not require any moment assumptions beyond the existence of F and its quantile function. Having said this, for fixed $d \geq 0$ and $g > 0$, define:

$$z \equiv z_{t-d}(g) = \text{sign}(x_{t-d})|x_{t-d}|^g \quad (4)$$

and write $y \equiv x_{t+1}$ for the forecast target. The SPGF predictor is $y = a + bz$. Then the inner optimization problems solve:

Definition 2.1 (The three forecast-optimisation problems). For fixed (d, g) , define:

$$(a^*, b^*)_{\text{MSE}} = \arg \min_{a, b} \mathbb{E}[(y - a - bz)^2], \quad (5)$$

$$(a^*, b^*)_{\text{MAE}} = \arg \min_{a, b} \mathbb{E}|y - a - bz|, \quad (6)$$

$$(a^*, b^*)_{\text{SIGN}} = \arg \max_{a, b} P(\text{sign}(y) = \text{sign}(a + bz)) \quad (7)$$

The outer optimization problem is:

$$(d^*, g^*) = \arg \min_{d \geq 0, g > 0} \mathcal{L}((a^*, b^*)_{\bullet}(d, g); d, g) \quad (8)$$

where $\mathcal{L}$ is the minimised loss, or the negative of maximised accuracy in the SIGN case, and $\bullet \in \{\text{MSE}, \text{MAE}, \text{SIGN}\}$.¹

¹All expectations in this paper are taken with respect to the stationary joint distribution of the process. This is the correct population criterion for a fixed-parameter forecast rule, because the parameters (a, b, d, g)

In the paper we solve (5)–(7) for fixed (d, g) , express the resulting objectives as functionals of the joint law of (y, z) , and then reduce that joint law via the DNO $*$ -product to a functional of (F, C_1, d, g) .

3 Inner optimisation: exact first-order conditions

3.1 Squared-error loss

Theorem 3.1 (MSE inner solution). *For fixed (d, g) with $\text{Var}(z) > 0$, the unique minimiser of (5) is*

$$b_{\text{MSE}}^* = \frac{\text{Cov}(y, z)}{\text{Var}(z)}, \quad a_{\text{MSE}}^* = \mathbb{E}[y] - b_{\text{MSE}}^* \mathbb{E}[z] \quad (9)$$

and the minimised loss is

$$\mathcal{L}_{\text{MSE}}(d, g) = \text{Var}(y)(1 - \rho^2(y, z)), \quad \rho(y, z) = \frac{\text{Cov}(y, z)}{\sqrt{\text{Var}(y) \text{Var}(z)}} \quad (10)$$

Proof. The objective is strictly convex quadratic in (a, b) . Setting its gradient equal to zero yields the normal equations $\mathbb{E}[y - a - bz] = 0$ and $\mathbb{E}[z(y - a - bz)] = 0$, whose unique solution is (9). Substituting back gives (10). $\square$

Theorem 3.1 shows that the SPGF family preserves the familiar projection logic of linear forecasting while enlarging the choice set over which correlation is maximised. Once the lagged regressor is allowed to pass through a sign-preserving power map, the outer MSE problem becomes a joint problem of memory selection and magnitude reweighting. In non-Gaussian environments this reweighting is substantive rather than cosmetic, because the linear benchmark $g = 1$ need not be the correlation-maximising transformation.

Corollary 3.2 (Outer MSE problem). *Since $\text{Var}(y)$ is independent of (d, g) ,*

$$(d^*, g^*)_{\text{MSE}} = \arg \max_{d, g} \rho^2(x_{t+1}, \text{sign}(x_{t-d})|x_{t-d}|^g)$$

3.2 Absolute-error loss

Theorem 3.3 (L_1 inner solution). *Suppose that for every b in a neighbourhood of b^* , the distribution of $y - bz$ has a continuous density strictly positive at its median. Then $(a^*, b^*)_{\text{MAE}}$*

are constants that must perform well on average over all possible histories. Minimizing the unconditional risk is equivalent to minimizing the average conditional risk. Since the process is stationary, the joint distribution of (y, z) is independent of t , and the sample average of the loss converges to the unconditional risk.

satisfies

$$a^* = \text{Med}(y - b^*z), \quad (11)$$

$$0 = \mathbb{E}[z \text{sign}(y - a^* - b^*z)]. \quad (12)$$

Proof. For fixed b , the map $a \mapsto \mathbb{E}|y - bz - a|$ is convex with subgradient $-\mathbb{E}[\text{sign}(y - bz - a)]$ and is minimised at the median of $y - bz$, giving (11). Differentiating the profiled objective with respect to b yields (12) under the stated regularity conditions. $\square$

Remark 3.1 (Median regression interpretation). Equation (12) is a sign-orthogonality condition for the residual $r = y - a^* - b^*z$, the L_1 analogue of the OLS orthogonality condition. Jointly, (11)–(12) define the population median regression of y on z in the sense of Koenker and Bassett (1978); Koenker (2005).

Corollary 3.4 (Outer MAE problem). *The outer MAE optimisation over (d, g) requires minimising*

$$\mathcal{L}_{\text{MAE}}(d, g) = \mathbb{E} |y - a^*(d, g) - b^*(d, g)z_{t-d}(g)|$$

where (a^*, b^*) solve (11)–(12). Although no closed-form expression for $\mathcal{L}_{\text{MAE}}$ exists in general, its copula representation is

$$\mathcal{L}_{\text{MAE}}(d, g) = \int_{[0,1]^2} |F^{-1}(v) - a^*(d, g) - b^*(d, g) \text{sgn}(F^{-1}(u))| F^{-1}(u)^g | \, dC_{d+1}(u, v)$$

Corollary 3.5 (Coincidence of MSE and MAE solutions). *If the conditional law of y given z is symmetric about $a_{\text{MSE}}^* + b_{\text{MSE}}^*z$ for every z , then $(a^*, b^*)_{\text{MAE}} = (a^*, b^*)_{\text{MSE}}$.*

Proof. Under conditional symmetry, the conditional mean and conditional median coincide at the MSE solution, so (11)–(12) are satisfied by $(a^*, b^*)_{\text{MSE}}$. $\square$

Corollary 3.5 identifies exactly when the common intuition that MAE is just a more robust version of MSE is valid. When conditional symmetry fails, the mean and median target genuinely different population objects, and the SPGF parameters that optimize them need not agree. This separation is especially relevant in financial environments with asymmetric dependence and lower-tail concentration, where the divergence between magnitude-sensitive criteria is economically meaningful rather than statistically incidental.

3.3 Directional-accuracy loss

Theorem 3.6 (Reduction to a scalar threshold). *For fixed (d, g) and sign $s = \text{sign}(b) \in \{-1, 1\}$, problem (7) depends on (a, b) only through the threshold $\tau = -a/b$, via*

$$\text{sign}(a + bz) = s \text{sign}(z - \tau) \quad (13)$$

Writing

$$\Psi(\tau; d, g) = \mathbb{E}[\text{sign}(y) \text{sign}(z - \tau)] \quad (14)$$

we have

$$P(\text{sign}(y) = \text{sign}(a + bz)) = \frac{1}{2} + \frac{s}{2} \Psi(\tau; d, g) \quad (15)$$

Proof. Since $a + bz = b(z - \tau)$, the sign of $a + bz$ depends on (a, b) only through s and τ . The probability identity then follows from the indicator relation $\mathbf{1}\{uv = 1\} = \frac{1}{2}(1 + uv)$ for $u, v \in \{-1, 1\}$. $\square$

Theorem 3.7 (Threshold first-order condition). *If z has density $f_z > 0$ on its support and $\tau \mapsto \mathbb{E}[\text{sign}(y) \mid z = \tau]$ is continuous, then any interior maximiser τ^* of $\Psi(\cdot; d, g)$ satisfies*

$$\mathbb{E}[\text{sign}(y) \mid z = \tau^*] = 0 \quad (16)$$

Proof. Write $\Psi(\tau; d, g) = \int_{-\infty}^{\infty} m(u) f_z(u) \, du - \int_{-\infty}^{\tau} m(u) f_z(u) \, du$, where $m(u) = \mathbb{E}[\text{sign}(y) \mid z = u]$. Differentiating under the integral sign via Leibniz's rule gives

$$\frac{\partial \Psi}{\partial \tau}(\tau; d, g) = -m(\tau) f_z(\tau) - m(\tau) f_z(\tau) = -2 \mathbb{E}[\text{sign}(y) \mid z = \tau] f_z(\tau)$$

Since $f_z(\tau) > 0$ on the interior of the support of z , any interior stationary point τ^* must satisfy $\mathbb{E}[\text{sign}(y) \mid z = \tau^*] = 0$, which is (16). $\square$

Corollary 3.8 (One-parameter ray of optimal sign rules). *Fix τ^* solving (16) and let $s^* = \text{sign}(\Psi(\tau^*; d, g))$. Every pair (a, b) with $\text{sign}(b) = s^*$ and $a = -b\tau^*$ attains*

$$\text{Accuracy}^*(d, g) = \frac{1}{2} + \frac{1}{2} |\Psi(\tau^*; d, g)| \quad (17)$$

The sign criterion therefore does not identify (a, b) individually.

Remark 3.2 (Scope and identification). The non-identification in Theorem 3.8 is structural rather than pathological. The sign operator is invariant to positive rescaling, so the criterion is rank-based and identifies only a separating threshold and direction, not a unique linear coefficient pair. In that sense the SIGN problem is closer to a classification problem than to a cardinal level-forecasting problem: what matters is on which side of the threshold the predictor falls, not the numerical scale on which it is represented.

4 Outer optimization via the Darsow–Nguyen–Olsen $*$ -product

4.1 Sklar decomposition and one-step copula

By Sklar's theorem, the stationary joint distribution of (x_t, x_{t+1}) has the representation

$$H_1(x, x') = C_1(F(x), F(x'))$$

for a unique copula C_1 when F is continuous. The copula C_1 encodes the entire one-step serial dependence independently of the marginal behaviour.

4.2 The DNO product

Definition 4.1 (Darsow–Nguyen–Olsen $*$ -product). For copulas A and B , define

$$(A * B)(u, v) = \int_0^1 \partial_2 A(u, t) \partial_1 B(t, v) dt \quad (18)$$

Theorem 4.2 (Markov composition). *The $*$ -product is bilinear and associative. If $\{x_t\}$ is a first-order Markov process with one-step copula C_1 , then for every $k \geq 1$ the copula of (x_t, x_{t+k}) is*

$$C_k = C_1^{*k} \quad (19)$$

Remark 4.1 (Closed-form powers). Closed-form $*$ -powers are available essentially only for elliptical copulas. For many Archimedean and other non-elliptical families, C_1^{*k} leaves the original family for $k \geq 2$ and has no usable closed form. This is why the simulation estimator of section 6 is a theoretical necessity rather than merely a computational convenience. The more realistic the dependence departs from the Gaussian benchmark, the more central simulation becomes to the actual operational meaning of the theory.

4.3 The joint law of (y, z) as a pushforward

Fix (d, g) . By stationarity and the Markov property, (x_{t-d}, x_{t+1}) has copula $C_{d+1} = C_1^{*(d+1)}$ and common marginal F . Writing $u = F(x_{t-d})$ and $v = F(x_{t+1})$, with $(u, v) \sim C_{d+1}$, we have

$$x_{t-d} = F^{-1}(u), \quad y = F^{-1}(v), \quad z = \text{sign}(F^{-1}(u)) |F^{-1}(u)|^g \quad (20)$$

Thus every quantity entering the MSE, MAE, and SIGN objectives is an integral functional of (F, C_{d+1}, g) . In particular,

$$\mathbb{E}[y] = \int_0^1 F^{-1}(v) dv, \quad (21)$$

$$\mathbb{E}[z] = \int_0^1 \text{sign}(F^{-1}(u)) |F^{-1}(u)|^g du, \quad (22)$$

$$\text{Cov}(y, z) = \int_{[0,1]^2} F^{-1}(v) \text{sign}(F^{-1}(u)) |F^{-1}(u)|^g dC_{d+1}(u, v) - \mathbb{E}[y] \mathbb{E}[z] \quad (23)$$

This representation reduces the outer optimisation problem to a low-dimensional search over (d, g) once (F, C_1) are specified or estimated.

5 The sign-accuracy optimum and the g -invariance result

Theorem 5.1 (g -invariance of the sign criterion). *For every $g > 0$,*

$$\text{sign}(\text{sign}(x)|x|^g) = \text{sign}(x)$$

almost surely. Consequently the SIGN criterion depends on the lagged predictor only through the sign of x_{t-d} , not through the exponent g .

Proof. For $x \neq 0$, the transformed magnitude $|x|^g$ is strictly positive, so the sign of the transformed variable equals the sign of x . At $x = 0$ both sides are zero. $\square$

Corollary 5.2 (Zero threshold under symmetry and sign-monotonicity). *If F is symmetric about zero and $\mathbb{E}[\text{sign}(x_{t+1}) \mid x_{t-d} = \xi]$ is non-decreasing in ξ , then $\tau = 0$ satisfies (16) and globally maximises directional accuracy for every $g > 0$.*

Proof. Under symmetry, zero is the natural median threshold, and the monotonicity condition implies that the conditional sign expectation changes sign at zero and does so only once. $\square$

Remark 5.1 (What if sign-monotonicity fails?). Corollary 5.2 relies on the assumption that $\mathbb{E}[\text{sgn}(x_{t+1}) \mid x_{t-d} = \xi]$ is non-decreasing in ξ . This holds for many commonly used copula families (e.g., Gaussian, Student- t , Clayton with positive dependence) but may fail for processes with non-monotone dependence, such as those exhibiting periodic or cyclical patterns. In that case the optimal threshold τ^* may be non-zero and the optimal directional accuracy is given by the general formula (17) rather than the concordance identity (25). The simulation-based estimator of section 6 remains valid without any monotonicity assumption: it directly maximises the empirical analogue of $\Psi(\tau; d, g)$ over τ .

Corollary 5.3 (Zero threshold implies zero intercept). *Under the conditions of Corollary 5.2 (or more generally, whenever the optimal threshold is $\tau^* = 0$), the optimal intercept in the SPGF forecast is $a^* = 0$. The SPGF-SIGN predictor therefore simplifies to*

$$f_{t+1|t} = b^* \text{sign}(x_{t-d})|x_{t-d}|^g$$

and the sign of the forecast is simply the sign of the lagged value (up to a global sign factor). This property is a direct consequence of the symmetry of the copula and the marginal distribution around zero.

Theorem 5.4 (Concordance identity). *Under the conditions of Theorem 5.2, in particular with F symmetric about zero so that $F(0) = \frac{1}{2}$, the optimal directional accuracy depends on the lag d only through the concordance functional*

$$\kappa_{d+1} = \mathbb{E}[\text{sign}(x_{t+1}) \text{sign}(x_{t-d})] \quad (24)$$

and

$$\text{Accuracy}^*(d) = \frac{1}{2} + \frac{1}{2}\kappa_{d+1} \quad (25)$$

Moreover, κ_{d+1} admits the central-copula representation

$$\kappa_{d+1} = 4C_{d+1}\left(\frac{1}{2}, \frac{1}{2}\right) - 1 \quad (26)$$

which is Blomqvist's β evaluated at C_{d+1} .

Proof. Since $\tau^* = 0$ by Theorem 5.2, we have $\Psi(0; d, g) = \mathbb{E}[\text{sign}(y) \text{sign}(z)]$ for any $g > 0$. By the g -invariance theorem (Theorem 5.1), $\text{sign}(z) = \text{sign}(x_{t-d})$ a.s., giving $\Psi(0; d, g) = \kappa_{d+1}$. Substituting into (17) with $\tau^* = 0$ yields (25).

For (26), let $U = F(x_{t-d})$ and $V = F(x_{t+1})$, so $(U, V) \sim C_{d+1}$. Since F is symmetric about zero and continuous, $F(0) = \frac{1}{2}$, and therefore

$$\text{sign}(x_{t+1}) = \text{sign}(F(x_{t+1}) - \frac{1}{2}) = \text{sign}(V - \frac{1}{2}) \quad \text{a.s.}$$

and likewise $\text{sign}(x_{t-d}) = \text{sign}(U - \frac{1}{2})$ a.s. Hence

$$\kappa_{d+1} = \mathbb{E}[\text{sign}(V - \frac{1}{2}) \text{sign}(U - \frac{1}{2})] = 4C_{d+1}\left(\frac{1}{2}, \frac{1}{2}\right) - 1$$

where the last equality is the standard identity for Blomqvist's β (Durante and Sempi, 2015, Proposition 5.1.6). $\square$

Remark 5.2 (Structural separation). The SPGF parameters split naturally into two roles. The exponent g governs the magnitude-sensitive MSE and MAE criteria, whereas the lag structure governs the directional problem. In general, the MSE-, MAE-, and SIGN-optimal configurations need not coincide. This structural separation is one of the central implications of the paper: disagreements across forecast objectives can arise from the geometry of the predictor family itself, even before sampling uncertainty is taken into account.

6 Simulation-consistent estimation

For non-elliptical copulas, the multi-step copula $C_1^{*(d+1)}$ is typically unavailable in closed form. We therefore use simulation based on iterated Rosenblatt sampling. The purpose is not merely to approximate an intractable integral. The recursion generates exactly the Markovian dependence iterates that the outer optimisation requires, so the estimation strategy mirrors the population theory rather than bypassing it.

Theorem 6.1 (Conditional Rosenblatt sampling). *Let C_1 have conditional quantile function $Q(\xi | u)$, defined as the inverse in the second argument of the conditional copula distribution.*

Let $u_0 \sim U(0, 1)$ and for $i \geq 1$ draw $\xi_i \stackrel{iid}{\sim} U(0, 1)$ and set

$$u_i = Q(\xi_i \mid u_{i-1}) \quad (27)$$

Then $\{u_i\}$ is a stationary Markov chain on $[0, 1]$ with one-step copula C_1 , and (u_0, u_k) has copula C_1^{*k} for every $k \geq 1$.

Algorithm 6.2 (Simulation-consistent SPGF recovery). Given estimated F , estimated one-step copula C_1 , maximum lag D , a grid $\mathcal{G}$ for g , and Monte Carlo length N :

1. Simulate a single chain $\{u_i\}_{i=0}^{N+D}$ via (27) and set $x_i = F^{-1}(u_i)$.
2. For each lag $d \in \{0, \dots, D\}$ and exponent $g \in \mathcal{G}$, form pairs (y_n, z_n) with $y_n = x_{n+d+1}$ and $z_n = \text{sign}(x_n)|x_n|^g$.
3. For MSE, compute the OLS coefficients and associated loss.
4. For MAE, solve the median-regression first-order conditions numerically.
5. For SIGN, grid-search over thresholds for z_n and record the maximal directional accuracy.
6. Return the optimal configurations for MSE, MAE, and SIGN together with the associated parameters.

6.1 Practical implementation of the simulation estimator

The simulation estimator in Algorithm 6.2 is straightforward to apply, but three practical issues deserve attention.

Choice of Monte Carlo length N . The required N depends on the strength of serial dependence and the smoothness of the objective. For typical financial returns with weak autocorrelation, $N = 10^5$ gives negligible Monte Carlo error for the MSE and MAE objectives; for the SIGN criterion, which is rank-based, $N = 5 \cdot 10^4$ is usually sufficient. A simple diagnostic is to repeat the estimation with two independent chains and verify that the optimal (d^*, g^*) are stable.

Grid for g and d . For the power exponent g , a logarithmic grid over $[0.1, 3]$ (e.g., 10 points) captures the relevant curvature; finer grids can be used if computational budget permits. For the lag d , it is sensible to include all lags up to a maximum D beyond which the copula power $C_1^{*(d+1)}$ is numerically indistinguishable from independence. In practice, $D = 20$ is often adequate for daily financial data.

Reusing the simulated chain. The same simulated chain $\{x_i\}_{i=0}^{N+D}$ can be reused for every (d, g) pair because the pairs (y_n, z_n) are formed by shifting indices. This makes the grid search over (d, g) cheap once the chain is generated. The main computational cost is the Rosenblatt simulation from the fitted copula; this cost is incurred only once per rolling window.

Remark 6.1 (Consistency and efficiency). Under standard ergodicity conditions, ergodic averages from the simulated chain converge to the required population functionals. The same simulated chain can be reused across all (d, g) pairs, so the expensive step is simulation from the fitted copula rather than the outer grid search itself.

7 Application to financial return series

7.1 Why financial returns are the natural domain

Financial return series simultaneously exhibit four structural properties that activate every part of the presented SPGF theory. This is why the financial application is especially revealing rather than merely illustrative: each element of the SPGF setup acquires a direct empirical interpretation in return data.

- (i) *Heavy-tailed marginals.* The invariant marginal of daily and weekly returns is substantially heavier-tailed than Gaussian (Mandelbrot, 1963; Cont, 2001). This makes the power transform genuinely informative for magnitude-sensitive criteria.
- (ii) *Asymmetric tail dependence and leverage.* Large negative shocks are more likely to be followed by further negative shocks than large positive shocks are to be followed by further positive shocks (Black, 1976; Christie, 1982; Longin and Solnik, 2001; Ang and Bekaert, 2002). This is exactly the type of serial asymmetry that makes copula modeling useful and Gaussian approximations inadequate.
- (iii) *Non-trivial but small directional concordance.* Daily equity return predictability is typically weak in magnitude but can still be statistically and economically meaningful when evaluated on long samples (Pesaran and Timmermann, 1995; Leitch and Tanner, 1991).
- (iv) *Horizon-dependent dependence decay.* The dependence structure between x_t and x_{t+k} decays with k in a way governed by the underlying Markov operator, making lag selection empirically substantive rather than cosmetic.

These features do not merely motivate the SPGF framework in the abstract; they determine the form of practically useful sign forecasters. Heavy tails make the magnitude transform relevant for MSE and MAE. Asymmetric dependence motivates direct

conditional-probability modeling. Non-trivial short-run concordance motivates directional forecasting itself. Horizon-dependent decay makes lag selection and multi-lag aggregation operationally relevant.

7.2 Limitations of the baseline SPGF-SIGN estimator

The baseline SPGF-SIGN rule, while theoretically grounded, has four practical limitations in financial applications.

- (i) *Simulation variance.* The concordance and threshold are estimated from a Rosenblatt chain rather than directly from the observed return sample, so Monte Carlo noise can matter when the simulation length is too small.
- (ii) *Single-lag information loss.* A single-lag sign rule discards directional information available at neighbouring lags when several short lags are jointly informative.
- (iii) *Binary-versus-probability information loss.* A thresholded sign rule produces only a binary decision, whereas the full conditional probability of a positive return contains more information and can support confidence-weighted decisions.
- (iv) *g -parameter inheritance.* Because the sign criterion is g -invariant, a baseline implementation often inherits a magnitude scale from an MSE-oriented step, even though that scale is not intrinsic to the directional problem.

These limitations motivate three operational improvements over the baseline SPGF-SIGN estimator: a multi-lag sign-space logistic model, a direct copula-conditional probability estimator, and an ensemble that combines several directional signals. We propose and utilize these extensions next.

7.3 From theory to practice: approximations and extensions

The baseline SPGF-SIGN rule derived in Section 5 is the population-optimal single-lag sign forecaster under the concordance identity. Its practical limitations motivate three operational refinements. These refinements are not arbitrary heuristics; they are natural approximations that either (i) relax the single-lag restriction by aggregating information over several lags (MultiLag-Logistic), (ii) exploit the full conditional probability rather than just its sign (CopulaProb), or (iii) combine complementary signals to reduce variance (Ensemble). The MultiLag-Logistic model is a direct consequence of the g -invariance theorem: because the sign criterion depends only on lagged signs, a flexible logistic model in sign space is a parsimonious way to capture multi-lag concordance. The CopulaProb forecaster is an empirical implementation of the conditional probability (29), which contains more information than the binary sign. The Ensemble combines these approximations to guard against misspecification of any single component. Thus,

while each of these forecasters departs from the exact SPGF-SIGN rule, they all remain rooted in the same copula-based theory and are designed to address the specific weaknesses of the single-lag threshold rule. We present these in turn and also discuss some additional implementations that are of practical interest.

7.4 Multi-lag logistic regression in sign space

Because the sign criterion depends on (x_{t-d}, x_{t+1}) only through their signs, a natural extension works directly in sign space. Define:

$$s_t = \text{sign}(x_t) \in \{-1, 1\}$$

almost surely under continuity of the marginal law and then define:

Definition 7.1 (Multi-lag sign model). For maximum lag D , define the logistic sign model

$$P(s_{t+1} = 1 \mid s_t, s_{t-1}, \dots, s_{t-D}) = \Lambda\left(\beta_0 + \sum_{d=0}^D \beta_{d+1} s_{t-d}\right), \quad \Lambda(u) = \frac{1}{1 + e^{-u}} \quad (28)$$

Equivalently, with $y_t = \mathbf{1}\{x_{t+1} > 0\}$, this is a Bernoulli regression driven solely by lagged signs.

The model in (28) should be viewed as an operational sign-space extension motivated by the g -invariance theorem, not as a claim that logistic regression is the unique population-optimal rule under the original SPGF criterion. Its value lies in aggregating weak directional signals across several nearby lags in a parsimonious and interpretable form.

Remark 7.1 (Ridge regularisation). In rolling windows, especially at lower frequencies or under short samples, the sign design matrix may be ill-conditioned or close to separation. Ridge regularisation is therefore used as an implementation stabiliser rather than as a theoretical ingredient of the population model.

7.5 Testing the zero-threshold hypothesis

The condition $\tau^* = 0$ — equivalently $a^* = 0$ in the SPGF-SIGN predictor — is a strong structural implication of the symmetry assumptions in Corollary 5.2. It is therefore of practical interest to test whether the data support this restriction. A direct test on the threshold τ^* estimated from Algorithm 6.2 is complicated by the non-smoothness of the sign objective and the simulation noise. However, the sign-space logistic model of Definition 7.1 provides a straightforward alternative.

Consider the single-lag sign model

$$\log \frac{P(s_{t+1} = +1 \mid s_t = s)}{P(s_{t+1} = -1 \mid s_t = s)} = \beta_0 + \beta_1 s, \quad s \in \{-1, +1\}$$

The threshold $\tau^* = 0$ in the original SPGF-SIGN rule corresponds exactly to the condition that the forecast sign is independent of the magnitude of z , i.e., it depends only on $\text{sign}(z) = \text{sign}(x_t)$. Under the symmetry conditions of Corollary 5.2, the conditional probability of a positive return given a positive lag equals that given a negative lag:

$$P(x_{t+1} > 0 \mid x_t > 0) = P(x_{t+1} > 0 \mid x_t < 0) = \frac{1}{2}$$

which implies $\beta_0 = 0$ in the logistic model. Conversely, if $\beta_0 = 0$, then the logistic model produces a sign forecast that is $\text{sign}(\beta_1 s_t)$, which is a pure sign rule with threshold zero. Thus, the null hypothesis

$$H_0 : \tau^* = 0 \iff a^* = 0$$

is equivalent (under the logistic model) to

$$H_0 : \beta_0 = 0$$

A Wald test for H_0 is obtained from the maximum likelihood estimator $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ of the logistic regression. The asymptotic covariance matrix is the inverse of the Fisher information:

$$\widehat{\text{Var}}(\hat{\beta}) = \left(\mathbf{S}^\top \hat{\mathbf{W}} \mathbf{S} \right)^{-1}$$

where $\mathbf{S}$ is the $T \times 2$ design matrix with rows $(1, s_t)$ and $\hat{\mathbf{W}} = \text{diag}(\hat{p}_t(1 - \hat{p}_t))$ with $\hat{p}_t = \sigma(\hat{\beta}_0 + \hat{\beta}_1 s_t)$. The Wald statistic

$$W = \frac{\hat{\beta}_0^2}{\widehat{\text{Var}}(\hat{\beta}_0)} \xrightarrow{d} \chi_1^2 \quad \text{under } H_0$$

A p -value below a chosen significance level (e.g., 0.05) provides evidence against the zero-threshold hypothesis, indicating that the optimal SIGN threshold is not zero and that the symmetric conditions of Corollary 5.2 are violated.

For the multi-lag model (Definition 7.1), the same logic applies: the null that the optimal threshold is zero for every lag is equivalent to $\beta_0 = 0$ (intercept zero) in the joint logistic regression of s_{t+1} on $(s_t, \dots, s_{t-D})$. The Wald test for $\beta_0 = 0$ then tests the joint symmetry of the entire sign process. In practice, one may also test the joint hypothesis $\beta_0 = \beta_1 = \dots = \beta_D = 0$ to assess whether any sign predictability exists, but the isolated test on β_0 specifically addresses the zero-threshold condition.

Remark 7.2 (Interpretation of a rejected null). If the test rejects $H_0 : \beta_0 = 0$, then the optimal sign forecast under the SPGF-SIGN criterion has a non-zero threshold $\tau^* \neq 0$, and hence $a^* \neq 0$. This implies that the forecast sign is not simply the sign of the lagged value; it depends on the magnitude of the lagged return through the threshold. Such a finding is consistent with asymmetric dependence or non-zero mean returns. In that

case, the practitioner must estimate the full threshold τ^* (e.g., via Algorithm 6.2) rather than assuming a symmetric rule. The test therefore provides a diagnostic for whether the simplifying zero-intercept rule is justified by the data.

7.6 Direct copula-conditional probability estimation

The threshold first-order condition says that the optimal SIGN threshold is located where the conditional probability of a positive return equals one half. A natural refinement is therefore to estimate the full conditional probability

$$P(x_{t+1} > 0 \mid x_t = x),$$

which contains strictly more information than the thresholded sign forecast alone.

Remark 7.3 (Conditional copula notation). For a copula C with continuous partial derivative, write

$$h(v \mid u) := \frac{\partial C(u, v)}{\partial u}$$

for the conditional distribution function of V given $U = u$, so that $h(v \mid u) = P(V \leq v \mid U = u)$ whenever $(U, V) \sim C$. This is the standard conditional copula function used in Joe (1997) and Patton (2012).

Proposition 7.2 (Direct copula-conditional probability). *Let F be the invariant marginal and C_1 the one-step copula of a stationary process. Write $u = F(x_t)$ and $u_0 = F(0)$, and let $h(\cdot \mid u)$ be the conditional copula function defined in Remark 7.3. Then*

$$P(x_{t+1} > 0 \mid x_t = x) = 1 - h(u_0 \mid u) \tag{29}$$

where $h(u_0 \mid u) = \partial_1 C_1(u, u_0)$ in the notation of Joe (1997). Thresholding this probability at $1/2$ yields the associated binary sign forecast.

Proof. By Sklar's theorem, with $U = F(x_t)$ and $V = F(x_{t+1})$ jointly distributed as C_1 ,

$$P(x_{t+1} \leq 0 \mid x_t = x) = P(V \leq F(0) \mid U = u) = P(V \leq u_0 \mid U = u) = h(u_0 \mid u)$$

Taking complements gives (29). □

Operationally this construction has both a parametric and a nonparametric version. In the parametric version, $\partial_1 C_1$ is evaluated from a fitted family such as Gaussian or Student- t . In the nonparametric version used for robustness in the companion code, the conditional probability is approximated by an empirical k -nearest-neighbour conditional copula fraction. The former is analytically transparent; the latter reduces sensitivity to copula misspecification.

7.7 Ensemble combination

The four component sign forecasters—AR1-Sign, SPGF-Sign, MultiLag-Logistic, and CopulaProb—exploit complementary aspects of the return process. AR1-Sign is the linear benchmark. SPGF-Sign is the direct implementation of the paper’s baseline concordance theory. MultiLag-Logistic aggregates sign information across several lags. CopulaProb uses the fitted one-step conditional sign probability. Their errors need not be perfectly aligned, so a simple ensemble can improve stability.

Proposition 7.3 (Operational ensemble forecast and accuracy bound). *Let*

$$s_{t+1|t}^{\text{AR1}}, \quad s_{t+1|t}^{\text{SPGF}}, \quad s_{t+1|t}^{\text{MLL}}, \quad s_{t+1|t}^{\text{CP}} \in \{-1, 1\}$$

be the four component binary forecasts, with associated probabilities

$$p_{t+1|t}^{\text{AR1}}, \quad p_{t+1|t}^{\text{SPGF}}, \quad p_{t+1|t}^{\text{MLL}}, \quad p_{t+1|t}^{\text{CP}} \in [0, 1].$$

Define the ensemble sign forecast by majority vote over the four binary forecasts. In the event of a 2–2 tie, set the ensemble sign equal to the CopulaProb sign forecast. Define the ensemble probability forecast by equal-weight averaging:

$$p_{t+1|t}^{\text{Ens}} = \frac{1}{4} \left(p_{t+1|t}^{\text{AR1}} + p_{t+1|t}^{\text{SPGF}} + p_{t+1|t}^{\text{MLL}} + p_{t+1|t}^{\text{CP}} \right). \quad (30)$$

Moreover, let $p_k = P(\hat{s}_{t+1}^{(k)} = s_{t+1}) > \frac{1}{2}$ be the hit rate of the k -th component, $\bar{p} = \frac{1}{K} \sum_k p_k$ the mean hit rate, and $\bar{\rho}$ the mean pairwise error correlation across the $K = 4$ components. Then the variance of the equal-weight average probability score satisfies

$$\text{Var}(p_{t+1|t}^{\text{Ens}}) = \frac{1 + (K - 1)\bar{\rho}}{K} \bar{\sigma}^2, \quad (31)$$

where $\bar{\sigma}^2$ is the mean component probability-score variance. Consequently, for $\bar{\rho} < 1$, the ensemble reduces forecast variance relative to any single component by a factor of $[1 + (K - 1)\bar{\rho}]/K < 1$, with the reduction increasing as $\bar{\rho}$ decreases (Timmermann, 2006; Bates and Granger, 1969).

The tie-breaking rule gives preference to CopulaProb because it uses the richest one-step conditional object among the four component methods, namely the fitted copula-based conditional event probability. Equal weighting is retained to avoid introducing an additional estimation layer for combination weights in short rolling samples. This choice is also consistent with the forecast-combination literature, where equal weighting often performs well because it exchanges a small amount of theoretical efficiency for a meaningful reduction in estimation error.

7.8 Economic evaluation criteria

Because the Python financial module produces both directional forecasts and probability-based decision scores, statistical accuracy alone is not sufficient for evaluation. We therefore report level-domain criteria (MSE, RMSE, MAE, MASE), directional criteria (HitRate, Pesaran–Timmermann, sign-based Diebold–Mariano comparisons), and economic criteria (information ratio and cumulative long-short P&L).

Pesaran–Timmermann test

The Pesaran–Timmermann statistic tests whether forecast signs and realised signs are more aligned than would be expected under independence (Pesaran and Timmermann, 1992). Let $\hat{p}_y$ be the empirical frequency of positive outcomes, $\hat{p}_f$ the empirical frequency of positive forecasts, and $\hat{p}_c = \hat{p}_y \hat{p}_f + (1 - \hat{p}_y)(1 - \hat{p}_f)$ the hit rate expected under independence. Then

$$PT = \frac{\hat{p} - \hat{p}_c}{\sqrt{\widehat{\text{Var}}(\hat{p}) - \widehat{\text{Var}}(\hat{p}_c)}} \quad (32)$$

is asymptotically standard normal under the null (Pesaran and Timmermann, 1992), where

$$\widehat{\text{Var}}(\hat{p}) = \frac{\hat{p}_c(1 - \hat{p}_c)}{T}, \quad (33)$$

$$\widehat{\text{Var}}(\hat{p}_c) = \frac{(2\hat{p}_y - 1)^2 \hat{p}_f(1 - \hat{p}_f) + (2\hat{p}_f - 1)^2 \hat{p}_y(1 - \hat{p}_y)}{T} \quad (34)$$

with T the number of out-of-sample forecast periods. These expressions assume serially independent forecast errors; under dependence, replace T with the long-run variance estimated by a HAC procedure (West, 1996).

Information ratio and cumulative long-short P&L

Given sign forecast $\hat{s}_{t+1|t} \in \{-1, 1\}$ and realised return x_{t+1} , define the one-period strategy return by

$$r_{t+1}^S = \hat{s}_{t+1|t} x_{t+1}. \quad (35)$$

The information ratio is

$$IR = \frac{\mathbb{E}[r_{t+1}^S]}{\sqrt{\text{Var}(r_{t+1}^S)}}, \quad (36)$$

or its sample analogue in the rolling backtest. The cumulative gross return is

$$\text{PnL}(T) = \prod_{t=1}^T (1 + r_t^S). \quad (37)$$

These criteria are economically meaningful because they depend not only on directional accuracy but also on the magnitude distribution of correct and incorrect forecasts.

Mean absolute scaled error

To retain one scale-free level-domain measure alongside the directional metrics, we report the MASE

$$\text{MASE} = \frac{\text{out-of-sample MAE}}{\text{initial-window naive one-step MAE}}, \quad (38)$$

where the denominator is computed once from the initial training window. In the financial sign module this should be viewed as an auxiliary comparability diagnostic rather than the primary target of the sign forecasters.

8 Companion implementation and evaluation protocol

The complete financial-forecasting framework is implemented in the companion file `SPGF_financial.py`. The implementation mirrors the theoretical construction while providing a bias-free rolling-window comparison of five operational sign forecasters.

The module includes an empirical marginal model, Gaussian and Student- t copula classes, an empirical conditional copula device, and five evaluated sign forecasters:

- (i) AR1-Sign, the linear sign baseline;
- (ii) SPGF-Sign, the paper’s original single-lag copula-concordance rule;
- (iii) MultiLag-Logistic, the sign-space logistic extension;
- (iv) CopulaProb, the direct copula-conditional probability forecaster;
- (v) Ensemble, the majority-vote combination of the previous four methods with CopulaProb tie-breaking and equal-weight probability averaging.

The rolling evaluation is strictly expanding-window. Each forecast at time t is produced using only observations up to $t - 1$. Lightweight models are refit every step, whereas the more expensive copula-based models may be refit every k steps for computational efficiency. The evaluation harness reports MSE, RMSE, MAE, MASE, HitRate, Diebold–Mariano comparisons under MSE, MAE, and sign-hit loss, together with the Pesaran–Timmermann statistic, information ratio, and cumulative long-short P&L for all five forecasters. A forward-looking-bias audit verifies that all training windows are expanding and use only historical data, that the MASE denominator is computed once from the initial training window, and that forecasting contexts are constructed without future information. The implementation also supports both real-data acquisition and calibrated synthetic data generation, allowing stress tests and robustness checks even

in environments without live network access. This matters for the paper’s methodological contribution because it allows the same code base to be used for both empirical comparison and controlled theory-consistent simulation exercises.

9 Conclusion

We have developed a general-process theory of the four-parameter Sign-Power Generalized Forecast functional under squared-error, absolute-error, and directional-accuracy loss. For fixed parameters (d, g) , the inner optimization problems reduce respectively to linear regression, median regression, and a threshold sign problem. The outer optimization problems reduce, through the Darsow–Nguyen–Olsen product applied to a single one-step copula, to functionals of the pair (F, C_1, d, g) . The directional-accuracy problem exhibits a structural asymmetry relative to the magnitude-sensitive criteria: it depends on lag structure but is invariant to the power exponent. The financial application shows why return forecasting is a natural empirical domain for the SPGF framework. Heavy tails make the magnitude transform relevant, asymmetric dependence motivates copula modeling, weak but persistent concordance motivates sign prediction, and horizon-dependent decay motivates lag selection and multi-lag aggregation. On this foundation we derive three operational improvements over the baseline SPGF-SIGN estimator—MultiLag-Logistic, CopulaProb, and Ensemble—and evaluate, together with AR1-Sign and SPGF-Sign, a five-method directional forecasting suite.

The companion Python implementation provides a reproducible platform for simulation and rolling-window evaluation. Its role is not merely computational: it operationalises the paper’s central message that forecast optimality depends on the loss function, and that a single parametric family can expose this dependence transparently. The empirical results and robustness analysis generated from this framework can therefore be read not only as a horse race among forecasting rules, but as an applied map of the theoretical separation between MSE-, MAE-, and SIGN-optimal prediction.

A Waldian perspective on loss selection. The theoretical development of the paper is inherently decision-theoretic: for a fixed loss function, the SPGF framework delivers the optimal action (a^*, b^*, d^*, g^*) in the Waldian sense, Wald (1950). However, the central finding that the MSE, MAE, and SIGN optima are generically distinct (Remark 5.2) raises a higher-order decision problem: which loss should the practitioner choose? This is not a modelling ambiguity but a fundamental reflection of the fact that different economic objectives imply different risk functions. The diagnostic tests of Appendix A.1—in particular the tests for sign-monotonicity and the equality of MSE and MAE slopes—provide data-driven evidence to guide this meta-decision. Rejection of sign monotonicity, for example, indicates that the simplifying zero-threshold rule is invalid, favouring the full SIGN threshold search; rejection of slope equality indicates that conditional

asymmetry is present, so MSE and MAE should not be treated as interchangeable. In practice, one may adopt a minimax regret criterion over the three losses, or form a convex combination $L_\lambda = \lambda_1 \text{MSE} + \lambda_2 \text{MAE} + \lambda_3 (1 - \text{HR})$ to reflect preferences, both of which are natural extensions of the present framework and are indeed supported by the companion code for anyone wishing to perform further experimentation.

References

- Anatolyev, S., and Gerko, A. (2005). A trading approach to testing for predictability. *Journal of Business & Economic Statistics*, 23(4), 455–461.
- Ang, A., and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15(4), 1137–1187.
- Bates, J. M., and Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, 20(4), 451–468.
- Bekaert, G., and Wu, G. (1998). Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, 13(1), 1–42.
- Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association*, 177–181.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.
- Boshnakov, G. N., and Ravagli, D. (2021). A signed power transformation with application to white noise testing. *arXiv preprint arXiv:2109.00280*.
- Brechmann, E. C., and Czado, C. (2012). Copula-based dependence modeling. *Journal of Multivariate Analysis*, 110, 1–3.
- Brockwell, P. J., and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer.
- Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451), 941–956.
- Chen, X., and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2), 307–335.
- Christie, A. A. (1982). The stochastic behavior of common stock variances. *Journal of Financial Economics*, 10(4), 407–432.
- Clark, T. E., and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291–311.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.
- Darsow, W. F., Nguyen, B., and Olsen, E. T. (1992). Copulas and Markov processes. *Illinois Journal of Mathematics*, 36(4), 600–642.

- Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Durante, F., and Sempi, C. (2015). Principles of copula theory. *CRC Press*.
- Elliott, G., Komunjer, I., and Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies*, 72(4), 1107–1125.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity. *Econometrica*, 50(4), 987–1007.
- Fan, J., and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.
- Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business*, 38(1), 34–105.
- Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics*, 12(3), 299–308.
- Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operational Research Quarterly*, 20(2), 199–207.
- Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operational Research Quarterly*, 20(2), 199–207.
- Granger, C. W. J., and Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19(7), 537–560.
- Grinold, R. C., and Kahn, R. N. (1989). *Active Portfolio Management*. McGraw-Hill.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384.
- Han, H., Linton, O., Oka, T., and Whang, Y.-J. (2016). The cross-quantilogram: measuring quantile dependence and testing directional predictability between time series. *Journal of Econometrics*, 193(1), 251–270.
- Harvey, C. R., Siddique, A., and others. (1999). Autoregressive patterns in asset returns. *Working paper*.
- Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambridge University Press.
- Hyndman, R. J., and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Ibragimov, R. (2009). Copula-based characterizations for higher-order Markov processes. *Econometric Theory*, 25(3), 819–846.

- Jegadeesh, N., and Titman, S. (1993). Returns to buying winners and selling losers. *Journal of Finance*, 48(1), 65–91.
- Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, 41(1/2), 133–145.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall.
- Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall/CRC.
- Jondeau, E., and Rockinger, M. (2006). The copula-GARCH model of conditional dependencies. *Journal of International Money and Finance*, 25(5), 827–853.
- Kaastra, I., and Boyd, M. (1996). Designing a neural network for forecasting financial time series. *Neurocomputing*, 10(3), 215–236.
- Kauppi, H., and Saikkonen, P. (2008). Predicting U.S. recessions with dynamic binary response models. *Review of Economics and Statistics*, 90(4), 777–791.
- Koenker, R., and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.
- Leitch, G., and Tanner, J. E. (1991). Economic forecast evaluation. *International Journal of Forecasting*, 7(4), 579–584.
- Linton, O., and Whang, Y.-J. (2007). The quantilogram: with an application to evaluating directional predictability. *Journal of Econometrics*, 141(1), 250–282.
- Longin, F., and Solnik, B. (2001). Extreme correlation of international equity markets. *Journal of Finance*, 56(2), 649–676.
- Longla, M. (2012). On the dependence structure of copula-based Markov chains. *Journal of Multivariate Analysis*, 104(1), 1–14.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(4), 394–419.
- Nitzan, S., and Paroush, J. (1982). Optimal decision rules in uncertain dichotomous choice situations. *International Economic Review*, 23(2), 289–297.
- Nyberg, H. (2011). Forecasting the direction of the stock market with dynamic binary models. *International Journal of Forecasting*, 27(2), 561–578.
- Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, 2(1), 130–168.
- Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2), 527–556.

- Patton, A. J. (2007). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256.
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256.
- Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110, 4–18.
- Pesaran, M. H., and Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4), 461–465.
- Pesaran, M. H., and Timmermann, A. (1995). Predictability of stock returns: robustness and economic significance. *Journal of Finance*, 50(4), 1201–1228.
- Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.
- Rodriguez, J. C. (2007). Measuring financial contagion: a copula approach. *Journal of Empirical Finance*, 14(3), 401–423.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23(3), 470–472.
- Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, 39(1), 119–138.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15(1), 72–101.
- Ter''asvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425), 208–218.
- Timmermann, A. (2006). Forecast combinations. In *Handbook of Economic Forecasting*, Vol. 1, 135–196.
- Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press.
- Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84(405), 231–240.
- Wald, A. (1950). *Statistical Decision Functions*. John Wiley & Sons, New York.
- West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5), 1067–1084.

Whittle, P. (1983). *Prediction and Regulation by Linear Least-Square Methods*. University of Minnesota Press.

White, H. (2006). *Approximate nonlinear forecasting methods*. Working notes.

Yeo, I. K., and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4), 954–959.

A Testable Hypotheses and Diagnostic Tests

This appendix presents four testable hypotheses that follow directly from the theoretical results in the main paper. Each test is stated with its mathematical foundation, required estimators, and sampling distribution. They are designed for implementation within the rolling-window backtest framework described in Section 8.

A.1 DNO concordance decay test

A.1.1 Theoretical background

The Darsow–Nguyen–Olsen composition theorem (Theorem 4.2) implies that the concordance

$$\kappa_{d+1} = \mathbb{E}[\text{sgn}(x_{t+1}) \text{sgn}(x_{t-d})]$$

is a deterministic function of the one-step copula C_1 :

$$\kappa_{d+1} = \Gamma_{d+1}(C_1), \quad \Gamma_{d+1}(C_1) := \int_{[0,1]^2} (2\mathbf{1}\{u > \tfrac{1}{2}\} - 1)(2\mathbf{1}\{v > \tfrac{1}{2}\} - 1) \, dC_1^{*(d+1)}(u, v), \quad (\text{A.1})$$

For a Gaussian copula with parameter ρ , $\Gamma_{d+1}(\rho) = \frac{2}{\pi} \arcsin(\rho^{d+1})$. In general, the DNO product defines a specific decay profile that can be tested against the observed sample concordances.

A.1.2 Hypothesis

For a fixed maximum lag D , test

$$H_0^{(\text{DNO})} : \quad \kappa_{d+1} = \Gamma_{d+1}(C_1), \quad d = 0, \dots, D-1,$$

against the alternative that at least one lag deviates.

A.1.3 Estimators

1. Fit $\hat{C}_1$ from the training sample using the semiparametric method (Section 4.1).
2. Simulate a Rosenblatt chain of length N from $\hat{C}_1$ (Algorithm 6.2) and compute the simulated concordances:

$$\tilde{\kappa}_{d+1} = \frac{1}{N-d-1} \sum_{n=1}^{N-d-1} \text{sgn}(x_{n+d+1}) \text{sgn}(x_n). \quad (\text{A.2})$$

3. From the observed sample, compute the empirical concordances:

$$\hat{\kappa}_{d+1} = \frac{1}{T-d-1} \sum_{t=d+1}^{T-1} \text{sgn}(x_{t+1}) \text{sgn}(x_{t-d}). \quad (\text{A.3})$$

A.1.4 Test statistic

Let

$$\mathbf{e}_D = (\hat{\kappa}_1 - \tilde{\kappa}_1, \hat{\kappa}_2 - \tilde{\kappa}_2, \dots, \hat{\kappa}_D - \tilde{\kappa}_D)^\top.$$

Under H_0 , $\mathbf{e}_D$ is asymptotically normal with mean zero. Its covariance can be estimated by a block bootstrap or a HAC estimator (Newey–West). The Wald statistic

$$W_{\text{DNO}} = \mathbf{e}_D^\top \hat{\Sigma}^{-1} \mathbf{e}_D \xrightarrow{d} \chi_D^2 \quad (\text{A.4})$$

provides a joint test. Rejection indicates misspecification of the first-order Markov assumption or the chosen copula family.

A.2 Upper bound on single-lag directional accuracy

A.2.1 Theoretical background

Theorem 5.5 establishes that, under the symmetry and sign-monotonicity conditions, the maximal achievable single-lag directional accuracy is

$$\text{Accuracy}^*(d) = \frac{1}{2} + \frac{1}{2} |\kappa_{d+1}|. \quad (\text{A.5})$$

Any forecaster using only a single lag d cannot exceed this bound.

A.2.2 Hypothesis

For a given single-lag forecaster (e.g., SPGF-SIGN) with empirical hit rate $\widehat{\text{HR}}_d$, test

$$H_0^{(\text{Bound})} : \widehat{\text{HR}}_d \leq \frac{1}{2} + \frac{1}{2} |\kappa_{d+1}|,$$

against the one-sided alternative that the forecaster exceeds the bound.

A.2.3 Estimators

- $\widehat{\text{HR}}_d$: out-of-sample hit rate of the single-lag SPGF-SIGN rule (or any competitor) at lag d .
- $\hat{\kappa}_{d+1}$: empirical concordance from the observed sample (or the DNO-implied value). We recommend using the empirical concordance for a direct data-driven comparison.

A.2.4 Test statistic

Define

$$\Delta_d = \widehat{\text{HR}}_d - \left(\frac{1}{2} + \frac{1}{2} |\hat{\kappa}_{d+1}| \right).$$

Under H_0 , $\Delta_d \leq 0$. A one-sided test uses

$$Z_d = \frac{\Delta_d}{\widehat{\text{se}}(\Delta_d)} \xrightarrow{d} N(0, 1), \quad (\text{A.6})$$

where $\widehat{\text{se}}(\Delta_d)$ is obtained via a block bootstrap that resamples pairs (x_{t+1}, x_{t-d}) and re-computes both $\widehat{\text{HR}}_d$ and $\hat{\kappa}_{d+1}$. Rejection ($Z_d > z_\alpha$) indicates that the forecaster exploits information beyond the single-lag symmetry bound, which may occur if multiple lags are used or if the symmetry/monotonicity assumptions fail.

A.3 Test for equality of MSE and MAE slope parameters

A.3.1 Theoretical background

Corollary 3.6 states that $b_{\text{MSE}}^* = b_{\text{MAE}}^*$ if and only if the conditional distribution of y given z is symmetric about the linear predictor. Thus a test of slope equality is a direct test of conditional asymmetry.

A.3.2 Hypothesis

For a fixed transformation $z = \text{sgn}(x_{t-d})|x_{t-d}|^g$, test

$$H_0^{(\text{Asym})} : b_{\text{MSE}}^* = b_{\text{MAE}}^*.$$

A.3.3 Estimators

In each rolling window:

- $\hat{b}_{\text{MSE}}$: OLS slope from the regression of y on z .
- $\hat{b}_{\text{MAE}}$: slope from the median (quantile 0.5) regression of y on z , solved via the first-order conditions (Theorem 3.3).

A.3.4 Test statistic (Hausman-type)

Let $\hat{\theta} = (\hat{b}_{\text{MSE}}, \hat{b}_{\text{MAE}})^\top$. Under the null, both estimators are consistent for the same population slope, but OLS is efficient under conditional symmetry. Define

$$\hat{q} = \hat{b}_{\text{MSE}} - \hat{b}_{\text{MAE}}.$$

Its asymptotic variance is

$$\widehat{\text{Var}}(\hat{q}) = \widehat{\text{Var}}(\hat{b}_{\text{MSE}}) + \widehat{\text{Var}}(\hat{b}_{\text{MAE}}) - 2\widehat{\text{Cov}}(\hat{b}_{\text{MSE}}, \hat{b}_{\text{MAE}}),$$

where the variances and covariances are obtained from the joint asymptotic distribution of OLS and quantile regression estimators (using the influence function approach; see Koenker (2005)). The Hausman statistic is

$$W_{\text{Hausman}} = \frac{\hat{q}^2}{\widehat{\text{Var}}(\hat{q})} \xrightarrow{d} \chi_1^2. \quad (\text{A.7})$$

A large value provides evidence of conditional asymmetry.

A.4 Test for sign-monotonicity

A.4.1 Theoretical background

Corollary 5.2 requires that

$$m(\xi) = \mathbb{E}[\text{sgn}(x_{t+1}) \mid x_{t-d} = \xi] \quad (\text{A.8})$$

is non-decreasing in ξ . If this fails, the optimal threshold τ^* may be non-zero, and the simplified zero-intercept rule is not justified.

A.4.2 Hypothesis

Test

$$H_0^{(\text{Mon})} : m(\xi_1) \leq m(\xi_2) \quad \text{for all } \xi_1 < \xi_2,$$

against the alternative that $m(\xi)$ is not non-decreasing.

A.4.3 Non-parametric test

1. Partition the observed x_{t-d} into K quantile bins (e.g., $K = \lfloor T^{1/3} \rfloor$). Let the bin midpoints be $\xi_1, \dots, \xi_K$ and let $\bar{m}_j$ be the average of $\text{sgn}(x_{t+1})$ within bin j .
2. Use the Jonckheere–Terpstra test for ordered alternatives (Jonckheere, 1954). Let $Y_{rt}^{(i)} = \text{sign}(x_{t+1})$ for observation r in bin i , with n_i observations per bin. The test statistic is

$$J = \sum_{1 \leq i < j \leq K} \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} \mathbf{1}\{Y_r^{(i)} \leq Y_s^{(j)}\}, \quad (\text{A.9})$$

which counts all pairwise comparisons across bins that are consistent with the ordered alternative (non-decreasing $m(\xi)$). Under the null of no trend, J is asymp-

totically normal with mean $\mu_J = \frac{1}{4}(N^2 - \sum_i n_i^2)$ and variance obtainable from Jonckheere (1954), where $N = \sum_i n_i$.

A.4.4 Interpretation

Rejection of $H_0^{(\text{Mon})}$ implies that the sign-monotonicity condition fails. In that case, the practitioner should estimate the full threshold τ^* via Algorithm 6.2 rather than assuming $\tau = 0$, and consider that the forecast sign may depend on the magnitude of the lagged value in a non-monotone way.

A.5 Summary of implementation

Table 1: Summary of diagnostic tests

Test	Required data	Primary statistic	Distribution
DNO decay	One-step copula fit, Rosenblatt chain	W_{DNO} (A.4)	χ_D^2
Hit-rate bound	Out-of-sample hit rate, emp. concordance	Z_d (A.6)	$N(0, 1)$
MSE vs MAE	OLS and median regression slopes	W_{Hausman} (A.7)	χ_1^2
Sign-monotonicity	Binned conditional sign averages	J (A.9)	Normal / bootstrap

All tests can be performed in a rolling-window fashion to assess the stability of the underlying assumptions over time. The companion code infrastructure (Section 8) already supports the necessary estimators; only the additional test statistics need to be implemented.