Implied Volatility Surface Predictability: Evidence from TAIEX Index Options

Zong-Wei Yeh* Yen-Hsun Juan† Shih-Kuei Lin‡

August 22, 2026

*Department of Finance, National Chung Cheng University; zweyeh@ccu.edu.tw.

†Department of Money and Banking, National Chengchi University; 112352501@nccu.edu.tw.

‡Department of Money and Banking, National Chengchi University; square@nccu.edu.tw.

Implied Volatility Surface Predictability: Evidence from TAIEX Index Options

Abstract

We study which dimension of the implied volatility surface (IVS), the Level, the Slope, or the Curvature, is predictable out of sample and what information predicts it, using TAIEX index options over 213 out-of-sample months. We extract the three factors from raw option cross-sections following Collin-Dufresne et al. (2024) and predict their next-month levels from 101 predictors spanning six economic channels, using linear and nonlinear machine-learning models tuned by time-series cross-validation. Against a random-walk benchmark, predictability is concentrated in the Slope: the best specification attains an out-of-sample R^2_{RW} of $+16.2\%$, while the Level and the Curvature are not predictable. The predictive information spans credit, monetary, options-microstructure, and sentiment variables, and it survives the exclusion of every option-derived predictor. It is also economically valuable: a delta-hedged, conviction-weighted risk reversal on the Slope earns a net annualized Sharpe ratio of 1.08 after transaction costs.

Keywords: Implied volatility surface, machine learning, option strategies, TAIEX index options.
JEL codes: G13, G14, G17, C53.

1 Introduction

Market participants read the implied volatility surface (IVS) for information.[1] How the surface evolves is therefore central to derivative pricing, risk management, and portfolio allocation. A large literature predicts the level of implied volatility. How the surface as a whole evolves is much less well understood, and the exceptions are few (Gonçalves and Guidolin, 2006; Konstantinidi et al., 2008; Bernales and Guidolin, 2014). We ask three empirical questions: which dimension of the IVS (Level, Slope, or Curvature) is predictable out of sample; what information predicts it; and whether that predictability survives implementation as a tradable option position.

To fix what the shape characteristics of the IVS contain, we draw on the option profit-and-loss attribution framework of Carr and Wu (2020). Their approach is top-down: instead of specifying a dynamic model for the underlying and deriving option prices from it, they impose no-arbitrage directly on the option's profit and loss, which links the shape of the surface to the conditional moments that generate it without having to solve a dynamic model. Each shape characteristic thereby acquires an economic identity. The Level reflects the conditional variance of the underlying return. The Slope reflects the conditional covariance between the underlying return and changes in implied volatility; the leverage effect makes that covariance negative, tilting the surface into the familiar skew in which out-of-the-money puts are the more expensive. The Curvature reflects the volatility of implied volatility.

These identities sharpen the first question. To ask which dimension of the surface is predictable is to ask which of three conditional moments can be predicted from information available today. We take that question to TAIEX index options, traded on the Taiwan

[1]The surface has been used to measure risk-neutral expected variance (Britten-Jones and Neuberger, 2000; Jiang and Tian, 2005) and the compensation demanded for bearing uncertainty (Carr and Wu, 2009; Bollerslev et al., 2009), fears of disaster (Bates, 2000; Bollerslev and Todorov, 2011; Backus et al., 2011), the risk-neutral distribution and investor risk aversion (Rubinstein, 1994; Jackwerth, 2000), end-user demand pressure (Bollen and Whaley, 2004; Gârleanu et al., 2009), and the directional information in option order flow (Pan and Poteshman, 2006).

Futures Exchange. TAIEX index options were among the world's six most actively traded equity index options in 2018, and more than 200 million contracts changed hands in 2022.[2] The market has a concentrated domestic investor base in which retail and large-trader positioning can move the surface. Following Collin-Dufresne et al. (2024), we extract the three IVS factors by running daily cross-sectional regressions of implied volatility on an at-the-money (ATM) volatility-standardized moneyness metric, and take the coefficients on the last trading day of each month. We then assemble 101 predictors. Ninety-three of the ninety-five external variables fall into six economically motivated information channels: (i) underlying market characteristics, (ii) options microstructure, (iii) global and cross-market variables, (iv) macroeconomic and monetary conditions, (v) geopolitical risk, and (vi) Taiwan investor positioning. Six lagged IVS factors complete the set.

We predict next-month factor levels with eight machine-learning models. Five are linear: Ridge, Lasso, Elastic Net, principal-component regression, and partial least squares. Three are nonlinear: random forest, gradient-boosted regression trees, and XGBoost. To these we add three equal-weight forecast combinations. Each model's hyperparameters are chosen by five-fold time-series cross-validation inside every estimation window, so that no model is tuned on the data against which it is scored. All models are then compared against the random-walk benchmark, the standard in the IVS-prediction literature (Konstantinidi et al., 2008; Bernales and Guidolin, 2014), over 213 out-of-sample months with expanding estimation windows.

We report three main findings. The first is that predictability is concentrated in the Slope. Against the random-walk benchmark, the best specification attains an out-of-sample R^2_{RW} of $+16.2\%$ for the Slope (Clark–West $p < 0.01$), and seven of the eleven specifications we report are positive and significant at the 1% level. Neither the Level nor the Curvature is predictable: every specification is negative for the Level, and none is significant for the Curvature. The result is not the product of one episode. Splitting the

[2]Futures Industry Association, MarketVoice, "TAIFEX hits an all-time trading record in 2018" and "TAIFEX made tremendous strides in 2022."

213 evaluation months into halves, the Slope result holds in both and strengthens in the second, and the rolling 36-month R^2_{RW} stays above the benchmark for most of the sample. Partial least squares is the exception: it is strong early and weaker later, a pattern we do not have a clean explanation for.

The second concerns what information carries that predictability. Feature importance for the leading specification in each model family by out-of-sample R^2_{RW}, PLS, the random forest, and NLCombo, points to the same variables, and no single predictor accounts for more than eight percent of total importance. Beyond the lagged Slope itself, the call implied-volatility smirk and realized downside semivariance rank near the top for all three, the BBB corporate credit spread for PLS and the random forest, and University of Michigan consumer sentiment and the Taiwan policy and deposit rates for the random forest and NLCombo. Each has a natural economic reading. The credit spread measures corporate funding stress; the call smirk records demand pressure from options end-users; the domestic policy and deposit rates anchor the local risk-free curve against which the risk-neutral distribution is priced; downside semivariance measures realized left-tail risk; and consumer sentiment proxies for risk appetite. Options-market demand pressure is the one force among these already tied to the equity skew (Bollen and Whaley, 2004; Gârleanu et al., 2009), and the credit channel echoes the credit–index-option integration documented by Collin-Dufresne et al. (2024). Consistent with an economic rather than a mechanical reading, a predictor set that excludes every option-derived variable still predicts the Slope out of sample, so the signal is not merely the option market repeating its own information. We read the ordering of these information sets with more confidence than the contribution of any one channel, which the block-level estimates pin down only loosely.

The third finding is that this predictability is economically valuable. Directional timing on the Slope factor-mimicking portfolio earns a positive Sharpe ratio for every specification, reaching 0.86, with hit rates above 60%. That exercise is not investable, however, because the factor is not a traded asset. We therefore construct a risk reversal, a long

out-of-the-money put against a short out-of-the-money call, which is the option position whose value moves with the Slope. Each month we set the two strikes to scale with prevailing ATM implied volatility, size the position on the predicted change under a conditional mean–variance rule, and delta-hedge so that the P&L reflects the Slope rather than the direction of the index. Traded on the PLS predictions, the hedged strategy earns a net annualized Sharpe ratio of 1.08 ($t = 3.3$) after costs of 5 implied-volatility basis points, inside the 3 to 10 basis point spreads typical of liquid TAIEX options. Costs would have to rise several times over before the strategy stopped paying, and short-dated risk reversals perform best.

Related Literature and Contributions. Our work speaks to two literatures. The first studies the dynamics of the IVS. After Dumas et al. (1998) showed that a deterministic volatility function fitted in the cross-section does not survive out of sample, this literature has characterized the surface with principal components (Skiadopoulos et al., 2000; Cont and da Fonseca, 2002) and with parametric stochastic volatility models, which require multiple factors to fit the smirk and its term structure (Heston, 1993; Christoffersen et al., 2009; Andersen et al., 2015). Closest in spirit to our reading of the surface, Aït-Sahalia et al. (2021) invert its shape characteristics directly into volatility dynamics. We build on Collin-Dufresne et al. (2024), who extract Level, Slope, and Curvature from raw option cross-sections rather than from an interpolated grid, so that the three factors keep the same economic meaning in every month.

Our closest antecedent is Chen et al. (2023), who also study the level, slope, and curvature of implied volatility and also bring machine learning to bear on them. Three things separate the two papers. Their shape measures are three-point contrasts on a fixed grid of put deltas at a single 30-day maturity: the level is at-the-money implied volatility, the slope the out-of-the-money minus at-the-money difference, and the curvature a butterfly. Ours are least-squares coefficients fitted to every contract in the daily cross-section

at maturities from one week to one year, so that the Slope is a symmetric derivative at the money rather than a one-sided difference that also absorbs curvature. Their design is cross-sectional, predicting the shapes of 3,345 individual equity option surfaces from 94 firm fundamentals; ours is a time series, predicting one index surface from macro-financial and market state variables. And their estimator is a lasso-based combination of univariate predictions, a selection method, whereas the question we ask calls for comparing selection against pooling and linear against nonlinear specifications.

A longer line of work asks whether the surface is predictable without asking what predicts it. Gonçalves and Guidolin (2006) and Bernales and Guidolin (2014) model the daily coefficients of a deterministic volatility function with a vector autoregression, Chalamandaris and Tsekrekos (2010) and Kearney et al. (2019) do the same for currency and commodity option surfaces, and Konstantinidi et al. (2008) predict implied volatility indices. All find statistical predictability against a random walk, and all report delta-hedged profits that survive only moderate costs and selective trading rules. Two things separate us. Their second stage conditions on the surface's own fitted coefficients, so it can establish that the surface is predictable but not what predicts it, whereas our 101 predictors attribute the signal to identifiable economic forces and it survives the removal of every option-derived variable. And where they predict the surface as one object and trade vehicles exposed to the level of volatility, we evaluate the three dimensions separately and trade the one that turns out to be predictable, net of costs calibrated to observed spreads.

The second literature applies machine learning to asset pricing, where the standard setting is a large cross-section of assets and characteristics (Gu et al., 2020; Freyberger et al., 2020; Kozak et al., 2020; Chen et al., 2024; Bali et al., 2023), though the same methods have been taken to the term structure (Bianchi et al., 2021). When options enter, the surface is usually an input rather than a target: Kelly et al. (2023) feed IVS data into deep networks to predict returns. Where the surface has been the target, the information set has been the surface itself, as in the regression-tree predictions of Audrino and Colangelo (2010)

and the deep-learning predictions of Medvedev and Wang (2022). We ask instead what economic information predicts the surface.

The IVS suits these methods for two reasons. Its candidate predictors are numerous relative to a monthly sample and strongly correlated, since realized volatility, downside risk, credit spreads, and option demand all respond to common states, so methods that pool predictors are better matched to the design than methods that select among them (Zou and Hastie, 2005). And because the shape characteristics price conditional moments rather than realized quantities, the mapping from the economic state to the surface need not be linear. Machine learning has already been brought to volatility, where it improves on econometric benchmarks for realized volatility (Christensen et al., 2023; Zhang et al., 2024), but that work targets a single quantity, the level. The surface has three shape dimensions, and we evaluate all three under one protocol, so that the comparison across them reflects the data rather than differences in design.

Roadmap. Section 2 sets out the conceptual framework. Section 3 describes the data and factor extraction, and Section 4 the prediction methodology. Section 5 reports the out-of-sample results, Section 6 the economic analysis, and Section 7 concludes.

2 A Conceptual Framework for the IVS Factors

This section provides a conceptual framework for interpreting the three IVS factors that we predict. Our goal is neither to estimate a structural option-pricing model nor to derive a formal no-arbitrage prediction equation. Instead, we use the option P&L attribution framework of Carr and Wu (2020) to clarify the economic content of the Level, Slope, and Curvature factors, linking each to a distinct component of the joint price–volatility environment. This mapping motivates our factor-by-factor prediction design, while leaving the relative predictability of the three factors as an empirical question to be settled in Section 5.

Consider a European call option with strike K, expiry date T, underlying price S_t, and BMS implied volatility I_t. Let $B(t, S_t, I_t; K, T)$ denote the BMS option value. Applying Itô's lemma, the instantaneous P&L can be decomposed as:

$$\mathrm{d}B = \underbrace{B_t\,\mathrm{d}t + B_S\,\mathrm{d}S_t + B_I\,\mathrm{d}I_t}_{\text{first-order}} + \underbrace{\tfrac{1}{2}B_{SS}(\mathrm{d}S_t)^2 + \tfrac{1}{2}B_{II}(\mathrm{d}I_t)^2 + B_{IS}\,\mathrm{d}S_t\,\mathrm{d}I_t}_{\text{second-order}}, \tag{1}$$

where B_t (theta), B_S (delta), B_I (vega), B_{SS} (gamma), B_{II} (volga), and B_{IS} (vanna) are partial derivatives of the pricing function. The first-order terms capture the option's sensitivity to time, price, and volatility; the second-order terms capture the convexity exposures.

Taking risk-neutral expectations of equation (1) replaces each stochastic increment by its conditional rate. Under zero financing costs the underlying is a martingale, so the delta term drops out; the expected rate of change in implied volatility is $\mu_t \equiv \mathbb{E}_t[\mathrm{d}I_t/I_t]/\mathrm{d}t$; and the three second-order terms become the conditional variance rate of the underlying return, the conditional variance rate of the implied volatility's percentage change, and the covariation rate between the two, which we denote σ_t^2, ω_t^2, and γ_t. Imposing zero expected excess return under continuous price and implied volatility movements, Carr and Wu (2020) then show that the option's theta loss must be exactly compensated by the expected gains from its gamma, vega, volga, and vanna exposures:

$$-B_t = B_I I_t \mu_t + \tfrac{1}{2}B_{SS}S_t^2\sigma_t^2 + \tfrac{1}{2}B_{II}I_t^2\omega_t^2 + B_{IS}I_tS_t\gamma_t, \tag{2}$$

where all four moments are risk-neutral conditional rates. Using the fact that the BMS theta, cash vega, cash vanna, and cash volga can each be written in terms of the cash gamma ($B_{SS}S_t^2$), Carr and Wu (2020) obtain an algebraic pricing relation for the fair implied volatility:

$$I_t^2 = \underbrace{2\tau\mu_t I_t^2 + \sigma_t^2}_{\text{level}} + \underbrace{2\gamma_t z_+ + \omega_t^2 z_+ z_-}_{\text{smile shape}}, \tag{3}$$

where $z_{\pm} = \ln(K/S_t) \pm \frac{1}{2} I_t^2 \tau$ are convexity-adjusted moneyness measures and $\tau = T - t$ is the time to maturity.

Equation (3) expresses the fair implied variance at any strike as the sum of a component that is constant in moneyness and a component that varies with moneyness, and it is this structure that gives economic meaning to the shape of the IVS. The constant component, the Level of the surface, depends only on the conditional return variance σ_t^2 and the expected drift of implied volatility μ_t, and pins down at-the-money implied volatility. The term linear in moneyness (the Slope, or skew) is governed by γ_t, the conditional covariance between the underlying return and changes in implied volatility; a negative γ_t, the hallmark of the leverage effect, tilts the surface so that out-of-the-money puts carry higher implied volatility than out-of-the-money calls. The term quadratic in moneyness (the Curvature, or smile) is governed by ω_t^2, the conditional variance of implied volatility changes (the volatility of volatility), and lifts both deep out-of-the-money puts and calls relative to at-the-money options. The three shape features of the surface therefore correspond to three economically distinct objects: the conditional variance of returns, the return–volatility covariance, and the volatility of volatility. The three enter equation (3) symmetrically, so the framework gives each shape feature an economic identity without implying that any one of them is easier to predict than the others. Which dimension is predictable is an empirical question, and we take it up in Section 5.

3 Data and Factor Extraction

3.1 TAIEX Options Data

Our primary data consist of daily settlement prices for TAIEX index options traded on the Taiwan Futures Exchange (TAIFEX). The TAIEX option, launched in December 2001, is a European-style contract on the Taiwan Stock Exchange Capitalization Weighted Stock Index and ranks among the most actively traded equity-index options in the Asia-Pacific

region. Our sample spans April 2002 to December 2024, providing 273 monthly observations.[3] The sample covers multiple market cycles including the 2008 global financial crisis, the 2011 European sovereign debt crisis, the 2015 Chinese market selloff, the 2020 COVID-19 pandemic, and the 2022 monetary tightening cycle.

For each trading day, we collect settlement implied volatilities for all available call and put options across strikes and expirations. We apply standard filters: we exclude options with fewer than 7 or more than 365 calendar days to expiration, options with moneyness (K/S) outside $[0.85, 1.15]$, and options whose settlement implied volatility falls outside $[0.01, 2.00]$. We also drop days on which the at-the-money volatility proxy of Section 3.2 is missing or below 0.01, and days with fewer than five surviving contracts; in this sample the first screen removes one day and the second is not binding. We use out-of-the-money puts for moneyness below 1.0 and out-of-the-money calls for moneyness at or above 1.0, following the standard practice of retaining only out-of-the-money contracts (Jiang and Tian, 2005; Carr and Wu, 2009); those studies place the cutoff at the forward rather than the spot price, a distinction that is immaterial at the short maturities that dominate our sample. After filtering, our sample contains 552,955 option-day observations across 5,591 trading days, with an average of 99 observations per day. Panel A of Table 1 reports descriptive statistics for the raw options data.

[Insert Table 1 here]

Figure 1 displays the average IVS across all months. The surface exhibits the characteristic volatility skew, with implied volatility higher for deep out-of-the-money puts and, to a lesser extent, for deep out-of-the-money calls. It also displays an upward-sloping term structure that reflects greater uncertainty at longer horizons.

[Insert Figure 1 here]

[3]We begin in April 2002 rather than at the December 2001 launch because the first 60 trading days list too few contracts to estimate the daily cross-sectional regression reliably: they average 12 options per day (maximum 19), against 99 thereafter, and the daily R^2 averages 0.42 against 0.75 over the retained sample.

3.2 Quadratic Regression Factor Extraction

A key advantage of our approach is that the quadratic regression specification of Collin-Dufresne et al. (2024) extracts IVS factors directly from the observed option cross-section on each trading day, without requiring prior interpolation onto a fixed grid. This avoids distortions introduced by surface smoothing and uses all available option observations. Two details differ from their implementation: we standardize $\log(K/S)$ rather than $\log(K/F)$, since a traded TAIEX forward is not available at every maturity, and we fit every filtered contract rather than one representative option per moneyness bucket. Collin-Dufresne et al. (2024) name only the intercept and the linear coefficient; the labels Level, Slope, and Curvature are ours, and follow the term-structure convention of Litterman and Scheinkman (1991) and Diebold and Li (2006). On each trading day, for all options with moneyness $K/S \in [0.85, 1.15]$ and maturity $\tau \in [7, 365]$ days, we define the moneyness measure standardized by ATM implied volatility and the square root of maturity,

$$m_i = \frac{\log(K_i/S)}{\sigma^{\text{ATM}}\sqrt{\tau_i/365}}, \tag{4}$$

where σ^{ATM} is the at-the-money implied volatility on that day, computed as the mean implied volatility of all options with moneyness $K/S \in [0.98, 1.02]$. This standardization normalizes log-moneyness by the prevailing level of implied volatility, so that the moneyness measure reflects economic distance from the money in units of expected price variation. We then estimate the cross-sectional regression

$$\text{IV}_i = a + b \cdot m_i + c \cdot m_i^2 + \varepsilon_i. \tag{5}$$

The three coefficients have direct economic interpretation. The intercept a captures the Level of the IVS, at-the-money implied volatility after controlling for moneyness effects. The linear coefficient b captures the Slope (or skew): negative values indicate that out-of-the-money puts are more expensive than out-of-the-money calls, reflecting demand

for downside protection. The quadratic coefficient c captures the Curvature (or smile): positive values indicate that both deep OTM puts and calls carry higher implied volatility than ATM options, reflecting the pricing of two-sided tail risk. Unlike statistical decompositions, this parametrization is free from rotation indeterminacy and produces factors with unambiguous, time-invariant economic meaning. By scaling log-moneyness by σ^{ATM}, much of the cross-sectional variation that would otherwise appear as Curvature is absorbed into the moneyness measure, so the Curvature factor captures residual smile shape after removing the scaling effect of ATM implied volatility.

The daily regressions achieve a mean R^2 of 0.751 (median 0.828) across 5,591 trading days, confirming that the quadratic specification captures the cross-sectional IVS shape well. Monthly factors are the last trading day's coefficients in each month, yielding 273 monthly observations. Panel B of Table 1 reports the factor moments: the mean Level $\bar{a} = 0.195$ indicates average ATM implied volatility of 19.5%, substantially higher than the typical 15–16% for S&P 500 options, consistent with the higher realized volatility of the TAIEX. The mean Slope $\bar{b} = -0.020$ is negative, confirming the typical equity skew, and is strongly left-skewed (-3.13); the Curvature is leptokurtic, reflecting occasional sharp smiles. The three factor levels are highly persistent, with first-order monthly autocorrelations of 0.88 (Level), 0.59 (Slope), and 0.50 (Curvature), consistent with the highly persistent volatility factor that parametric models require to fit the index smirk (Christoffersen et al., 2009). This persistence is what makes the random-walk benchmark a demanding standard for Level predictions (Gonçalves and Guidolin, 2006; Bernales and Guidolin, 2014).

Figure 2 plots the series: the Level closely tracks aggregate uncertainty, spiking in 2008 (reaching 0.519) and 2020, while the Slope becomes sharply more negative in the same episodes as investors bid up downside protection.

[Insert Figure 2 here]

3.3 Prediction Targets

Our prediction targets are the next-month levels of the three factors, a_{t+1}, b_{t+1}, and c_{t+1}. Predicting levels rather than changes is a deliberate design choice: it tests whether the economic state determines the equilibrium IVS shape, rather than merely predicting short-term mean reversion. We evaluate against the random-walk benchmark, the standard in the IVS-prediction literature (Gonçalves and Guidolin, 2006; Bernales and Guidolin, 2014) and in the closely related work on implied volatility indices (Konstantinidi et al., 2008), defined in Section 4.4.

4 Prediction Methodology

4.1 Information Channel Framework

A central feature of our design is the organization of predictors into economically motivated information channels. This allows us to assess not only aggregate predictability but also which types of information drive the predictability of the IVS, and it provides an economic rationale for the predictor set, mitigating concerns that the large number of variables reflects unprincipled data mining. We assemble 95 external predictor variables (93 organized into six economic channels plus two derived risk measures) and add six lagged IVS factor variables, for $95 + 6 = 101$ predictors in total. The channel counts below sum to the 95 external variables ($12 + 14 + 13 + 18 + 6 + 30 + 2 = 95$); the six lagged IVS variables enter every model alongside them.

Channel 1: Underlying Market (12 variables). Market return, realized volatility, realized skewness and kurtosis, maximum daily loss and gain, Amihud illiquidity, upside and downside semivariance, the high–low range, and momentum at 3- and 12-month horizons. The theoretical motivation is the leverage effect (Black, 1976; Christie, 1982): negative equity returns increase financial leverage and hence expected future volatility, which

transmits to the IVS through no-arbitrage linkages between the physical and risk-neutral measures. The resulting asymmetry in conditional volatility is well documented (Bekaert and Wu, 2000).

Channel 2: Options Microstructure (14 variables). Change in open interest, number of contracts traded, put–call open-interest ratio, and cross-sectional moments of the implied volatility distribution (mean, standard deviation, skewness, kurtosis, range, smirk, call smirk, butterfly, term slope), together with changes in call and put total open interest. These variables capture the demand pressure that shapes the implied volatility function (Bollen and Whaley, 2004; Gârleanu et al., 2009); option volume and open interest also carry informed-trading content (Pan and Poteshman, 2006), which may foreshadow IVS movements.

Channel 3: Global and Cross-Market (13 variables). The VIX level and derived measures (change, moving average, term slope, range, volatility), exchange-rate changes (USD/TWD, JPY/USD, EUR/USD, DXY, KRW/USD), the NASDAQ return, and the oil return. Global variables reflect cross-market volatility spillovers: the VIX serves as a global "fear gauge" that may lead IVS movements in smaller markets, while exchange-rate changes, and particularly JPY/USD as a barometer of global risk appetite, transmit risk-on and risk-off dynamics to the TAIEX options market.

Channel 4: Macroeconomic and Monetary (18 variables). U.S. Treasury yields and changes (10-year, 2-year, 3-month), term spreads, the Federal Funds rate, credit spreads (high-yield and BBB OAS), the global Economic Policy Uncertainty index (Davis, 2016), built as a GDP-weighted average of the national indices of Baker et al. (2016), U.S. inflation, and University of Michigan consumer sentiment. Credit spreads proxy for aggregate credit conditions and funding stress, while consumer sentiment captures investor mood that affects the demand for option protection. (We exclude two slow-moving U.S. real-activity series, industrial production and unemployment, whose mechanical link to a forward-looking risk-neutral skew is weak.)

Channel 5: Geopolitical Risk (6 variables). Global, Taiwan, China, and Korea GPR indices (Caldara and Iacoviello, 2022) and the changes in the global and Taiwan indices. Given Taiwan's unique geopolitical position, regional geopolitical risk may directly affect the demand for option protection: geopolitical events such as military tensions, diplomatic crises, and territorial disputes are precisely the low-probability, high-impact scenarios that drive demand for downside protection. The China and Korea indices capture regional spillover relevant to the TAIEX market.

Channel 6: Taiwan Investor Positioning (30 variables). Data from the Taiwan Economic Journal (TEJ): Taiwan macro indicators (CPI growth, policy and deposit rates, M2 growth, PMI, business-cycle sentiment, unemployment, exchange-rate change, export-order growth, and derived changes); margin trading (short-margin ratio and percentage changes in margin and short balances); futures large-trader open interest (total OI change, top-5/top-10 net positions, concentration spreads, and institutional and retail volume shares); the foreign-investor share of futures volume and net trading ratio; and options large-trader open interest (call/put top-5 net positions, put–call ratio, and net sentiment). This channel captures Taiwan-specific institutional positioning not reflected in global variables; the TAIEX option market features a concentrated domestic investor base in which large-trader and foreign-investor positioning can move the IVS.

We additionally include two **derived risk measures**: the variance risk premium (VRP = IV_Mean − RealizedVol) and the VIX–IV spread. The complete list of all 95 external variables with definitions is provided in Appendix A.

4.2 Predictor Variables

All 101 predictors are measured at the end of month t and used to predict $f_{k,t+1}$, the next month's factor level, where $f_k \in \{a, b, c\}$. The set comprises 95 external variables and six lagged IVS variables (lagged levels and first differences of all three factors), which capture the persistence and cross-factor dynamics inherent in the IVS. The expansion from

parsimonious predictor sets (fewer than 20 variables in prior work) is motivated by the hypothesis that the equilibrium IVS shape responds to a rich set of economic forces. The high predictor-to-observation ratio ($p/n \approx 101/60 = 1.68$ at the start of the out-of-sample period) creates a curse-of-dimensionality challenge: methods that select or over-shrink correlated predictors are expected to struggle, whereas methods that pool them (latent-component regressions and nonlinear ensembles) are better positioned to extract signal that is distributed across many weak, collinear variables. All predictors are standardized (zero mean, unit variance) within each expanding estimation window, ensuring that regularization penalties treat all variables symmetrically and that no look-ahead information contaminates the standardization.

4.3 Prediction Models

We employ eight machine-learning models drawn from two families, linear and nonlinear, plus three forecast combinations. The general prediction problem is

$$f_{k,t+1} = g(\mathbf{x}_t) + \varepsilon_{k,t+1}, \tag{6}$$

where $\mathbf{x}_t$ is the vector of predictors. Ordinary least squares is omitted: with $p \gg n$ it is infeasible in early windows and uninformative.

Linear models. The five linear models differ in how they discipline a predictor set larger than the sample. Three penalize the coefficients directly. **Ridge regression** adds an ℓ_2 penalty $\lambda\|\boldsymbol{\beta}\|_2^2$ to the OLS objective, shrinking all coefficients toward zero without setting any to zero. **Lasso** uses an ℓ_1 penalty $\lambda\|\boldsymbol{\beta}\|_1$ that induces sparsity by setting irrelevant coefficients exactly to zero. **Elastic Net** combines both:

$$\min_{\alpha,\boldsymbol{\beta}} \sum_s (f_{k,s+1} - \alpha - \boldsymbol{\beta}'\mathbf{x}_s)^2 + \lambda \left[\rho\|\boldsymbol{\beta}\|_1 + (1-\rho)\|\boldsymbol{\beta}\|_2^2\right], \tag{7}$$

where $\rho \in [0, 1]$ balances variable selection and coefficient stability.

The other two instead compress the predictors before fitting. **Principal-component regression (PCR)** projects the predictors onto their leading principal components and regresses the target on these orthogonal factors, mitigating multicollinearity. **Partial least squares (PLS)** constructs latent components that maximize covariance with the target, so that, unlike PCR, the dimension reduction is supervised. Both use all predictors through a small number of latent factors, making them well suited to a signal distributed across many collinear variables.

Nonlinear models. **Random forest** averages an ensemble of regression trees grown on bootstrap samples with random feature subsets. **Gradient-boosted regression trees (GBRT)** build shallow trees sequentially, each fitting the residuals of the previous ensemble under a Huber loss with shrinkage. **XGBoost** extends boosting with a regularized objective penalizing tree complexity. All three capture nonlinearities and interactions and tolerate collinear predictors.

Forecast combinations. Following Rapach et al. (2010), we construct three equal-weighted combinations: LinCombo, the average of the five linear models (Ridge, Lasso, Elastic Net, PCR, PLS); NLCombo, the average of the three nonlinear models (random forest, GBRT, XGBoost); and Combo, the average of all eight. The linear/nonlinear split tests whether combining within a family outperforms combining across both. Equal weighting is chosen for its robustness: simple averaging is difficult to beat out of sample when the individual predictions are themselves noisy (Rapach et al., 2010), and the refinements that improve on it, such as the combination elastic net of Rapach and Zhou (2020), require estimating additional weights from the same short sample.

4.4 Hyperparameter Selection and Out-of-Sample Evaluation

We employ an expanding (recursive) window scheme. The initial training window spans 60 monthly observations (2002:04–2007:03), yielding the first out-of-sample prediction for April 2007. At each subsequent month the window expands by one observation, all models are re-estimated including hyperparameter selection, and the next prediction is generated, producing $T_{\text{OOS}} = 213$ out-of-sample predictions (2007:04–2024:12). To prevent look-ahead bias, the training set at each prediction date uses only data through month $t - 1$.

Hyperparameter selection follows a uniform validation scheme applied to all models. Within each estimation window, we tune every model's hyperparameters by 5-fold time-series cross-validation (minimizing the mean validation mean-squared error across folds), and the selected configuration is refit on the full window before predicting the next month. The hyperparameter grids searched for each model are reported in Appendix Table A2. Targets are standardized within each window. This single, uniform scheme, applied in place of model-specific procedures, ensures that the comparison across model families is on an equal footing.[4]

We evaluate prediction accuracy against the random-walk (RW) benchmark, which predicts the next month's factor as the current level plus the expanding-window mean of past changes:[5]

$$\hat{f}^{\text{RW}}_{k,t+1} = f_{k,t} + \frac{1}{t}\sum_{s=1}^{t} \Delta f_{k,s}, \qquad \Delta f_{k,s} = f_{k,s} - f_{k,s-1}. \tag{8}$$

Because the IVS factors are highly persistent (Section 3.2), the random walk is a hard and standard benchmark in the IVS-prediction literature (Gonçalves and Guidolin, 2006; Bernales and Guidolin, 2014; Konstantinidi et al., 2008); the average drift is near zero, so (8)

[4] We use k-fold time-series cross-validation rather than a single chronological holdout: in a sample of this size, a single held-out validation block is noisy and selects unstable hyperparameters for the linear models, while k-fold cross-validation uses the data more efficiently and is the small-sample-robust choice.

[5] In the implementation the expanding mean at month t runs one period further than equation (8) specifies, including the change realized between t and $t+1$, which is the change being predicted. This makes the benchmark marginally sharper than stated, so the R^2_{RW} we report are, if anything, conservative; the effect is $O(1/t)$ and negligible over 213 months.

is effectively the no-change prediction. Our primary metric is

$$R^2_{\mathrm{RW}} = 1 - \frac{\sum_{t=1}^{T_{\mathrm{OOS}}} (f_{k,t} - \hat{f}_{k,t})^2}{\sum_{t=1}^{T_{\mathrm{OOS}}} (f_{k,t} - \hat{f}^{\mathrm{RW}}_{k,t})^2}. \tag{9}$$

A positive R^2_{RW} indicates improvement upon the random walk. The statistic follows the out-of-sample R^2 convention of the return-predictability literature (Campbell and Thompson, 2008; Welch and Goyal, 2008), with the random walk replacing the prevailing mean as the benchmark. We test predictive accuracy with the Clark and West (2007) one-sided test using Newey and West (1987) HAC standard errors. The test is derived for nested models; the random walk is nested in our linear specifications, which include the lagged factor, but not literally in the tree ensembles, so for those the statistic should be read as an approximate guide.

5 Empirical Results

5.1 Out-of-Sample Predictability

Table 2 presents the out-of-sample R^2_{RW} for all eight individual models and three combinations across the three factor levels, over 213 out-of-sample months.

[Insert Table 2 here]

The results reveal a clear hierarchy across IVS dimensions and across models. First, the Slope, which Section 2 identifies with the conditional return–volatility covariance γ_t that the leverage effect makes negative, is the predictable dimension. The nonlinear combination (NLCombo) achieves the highest $R^2_{\mathrm{RW}} = +16.2\%$ (CW $p < 0.01$), a 16% reduction in mean-squared prediction error relative to the random walk. Partial least squares attains $+14.6\%$, random forest $+14.0\%$, XGBoost $+10.4\%$, gradient boosting $+10.1\%$, the overall combination $+9.9\%$, and principal-component regression $+7.3\%$, all significant at the 1%

level. Every nonlinear model is significantly positive, and so are PCR and PLS.

Second, and equally informative, the remaining three linear models fail: Lasso (-42.9%), Ridge (-30.9%), and Elastic Net (-12.3%) are all negative, as is their combination LinCombo (-2.7%). The ordering does not sort on linearity. PLS, a linear model, matches the random forest ($+14.6\%$ versus $+14.0\%$), while Ridge and Lasso, equally linear, are the two worst specifications in the table. The specifications that succeed work with a small number of derived components or splits rather than with coefficients on the individual predictors. We do not press the comparison further, since it rests on the ordering of eleven specifications for a single factor in a single market. Sections 5.3 and 5.4 examine the structure of the signal directly.

Third, neither the Level nor the Curvature factor is robustly predictable. Every specification is negative for the Level, which captures aggregate implied volatility driven by uncertainty shocks that arrive unpredictably and are already well captured by the persistent random-walk benchmark. For Curvature, only a few nonlinear models are marginally positive (random forest $+5.8\%$, gradient boosting $+6.5\%$) and none is significant at the 10% level; the ATM-IV standardization absorbs much of the smile dynamics, leaving residual Curvature with little structure for economic fundamentals to exploit.

5.2 Subsample Stability

A concern with any out-of-sample exercise is whether the aggregate R^2_{RW} masks temporal instability. We split the 213-month evaluation period into two approximately equal halves and compare R^2_{RW} for the Slope across subsamples. The Slope result is stable and, for the nonlinear models, improving: random forest delivers $+10.6\%$ in the first half and $+23.5\%$ in the second, and the nonlinear combination $+15.4\%$ and $+18.3\%$; partial least squares is strong in the first half ($+17.7\%$) and weaker in the second ($+5.9\%$), while the overall combination improves from $+6.8\%$ to $+18.6\%$. The second-half improvement of the nonlinear models suggests that the accumulating training sample strengthens estimation and

that the predictive relationship between economic fundamentals and the volatility skew has become more reliable in recent years. Figure 3 plots the rolling 36-month R^2_{RW} for all three factors and confirms that the Slope stays predominantly above the benchmark for the models that pool predictors, while Level and Curvature sit mostly below.

[Insert Figure 3 here]

5.3 Where Does Predictability Come From?

To locate the source of Slope predictability, Table 3 predicts the Slope from each information block separately, which lets us ask whether the result is merely skew mean-reversion, a natural concern since the surface's own recent history is itself among the predictors. The lagged IVS factors are indeed informative on their own: the six lags deliver $R^2_{\text{RW}} = +8.4\%$ for partial least squares and $+6.6\%$ for random forest, confirming that factor persistence and cross-factor dynamics carry genuine predictive content. The predictability is not, however, only autoregressive. External information that excludes all option and IVS variables, that is, pure macro-financial and market conditions carrying no information from the option surface, still predicts the Slope at $R^2_{\text{RW}} = +10.2\%$ for random forest (and $+6.2\%$ for PLS), establishing that economic fundamentals carry predictive content independent of the surface's own history.

[Insert Table 3 here]

These two sources are complementary rather than redundant: adding the external information to the lagged factors raises the full-model accuracy to $+14.0\%$ (random forest) and $+14.6\%$ (PLS), a gain of six percentage points for PLS and seven for the random forest over the lags-only prediction. The predictability is therefore concentrated neither in the surface's own persistence nor in any single external channel. Taken individually, only two external blocks are positive for the random forest, option microstructure at $+9.8\%$ and the macro/monetary channel at $+3.3\%$, and no single block approaches the full-model level;

the remaining blocks are individually weak or negative, as is expected when a handful of noisy, collinear predictors are asked to beat a demanding random-walk benchmark on their own. What is robust is not the sign of any one block but the fact that both partial information sets fall well short of the full set on their own, and that both models converge to roughly $+14\%$ once the full predictor set is used. The ordering of the two partial sets is not itself stable: external information dominates the lags for the random forest but not for PLS.

5.4 Feature Importance: Distributed Multi-Channel Predictability

We examine feature importance from the leading specification in each model family by out-of-sample R^2_{RW}. The three embody complementary functional forms: partial-least-squares importance (a linear, supervised view), random-forest impurity importance (a nonlinear view), and NLCombo (the equal-weight average of the three tree models), each normalized to a percentage of total importance. Table 4 reports the leading predictors for the Slope.

[Insert Table 4 here]

Beyond the lagged Slope, which captures mean-reverting skew dynamics, the leading random-forest predictors span four channels: options microstructure (the call implied-volatility smirk, 4.7%), macroeconomic and monetary conditions (University of Michigan consumer sentiment, 4.7%, and the BBB corporate credit spread, 2.7%), Taiwan variables (the policy rate, 3.7%, and the deposit rate, 3.5%), and the underlying market (realized downside semivariance and 12-month momentum, 2.9% each). PLS ranks the same variables highly and spreads importance still more evenly, with no variable above 2.6%. NLCombo places the call smirk first at 7.5%, an ordering it inherits from gradient boosting, which puts 14.9% of its weight on that one variable. That the same predictors recur across a linear, a nonlinear, and a combination specification suggests the pattern is not an artifact of any single estimator. Concentration nonetheless differs sharply across methods:

the five largest weights account for 46% of total importance for Lasso and 43% for Elastic Net, against 11 to 13% for Ridge, PCR, and PLS. Figure 4 shows the ranking across all eight models.

Each of these variables has a natural economic reading. The credit spread captures corporate funding stress, the call smirk demand pressure in the options market, the domestic policy and deposit rates the local risk-free curve against which the risk-neutral distribution is priced, and consumer sentiment broad risk appetite. Of the four, options-market demand pressure is the one with a long-standing link to the equity skew (Bollen and Whaley, 2004; Gârleanu et al., 2009), and the credit channel has a counterpart in the integration of credit and index-option markets documented by Collin-Dufresne et al. (2024). We are not aware of prior work isolating the monetary and sentiment variables in a predictive design for the risk-neutral skew.

[Insert Figure 4 here]

6 Economic Analysis

Having established the statistical predictability of the Slope, we investigate whether it translates into economic value. We use directional factor timing as a prediction-quality diagnostic and implementable risk reversals as the economic test, and we examine sensitivity to transaction costs and maturity.

6.1 Directional Factor Timing

We first assess the information content of the Slope predictions through a directional timing strategy. At the end of month t, each model generates a prediction $\hat{b}_{t+1}$. The predicted direction of the Slope change is $\text{sign}(\hat{b}_{t+1} - b_t)$, and the strategy earns $r_{t+1} = \text{sign}(\hat{b}_{t+1} - b_t) \times (b_{t+1} - b_t)$ on the Slope factor-mimicking portfolio. Because the factor is not a traded asset, we interpret this as a prediction-quality statistic, the economic informa-

tion value of the directional prediction, rather than an investable return. Table 5 reports the annualized performance for all eleven specifications. Every model earns a positive Sharpe ratio and a hit rate above 60%: XGBoost leads (Sharpe 0.86, Sortino 1.24, hit rate 64%), followed by random forest (0.77, hit rate 68%) and the overall combination (0.76, hit rate 67%); even the weakest specifications (Ridge, PLS) exceed 0.45. The Sortino ratios, which penalize only downside variation, generally exceed the Sharpe ratios, indicating that the strategy's volatility is not dominated by losses. Figure 5 displays the cumulative directional returns. The uniformly positive directional performance indicates that the Slope predictions carry economically meaningful information across the entire model spectrum.

[Insert Table 5 and Figure 5 here]

6.2 Risk Reversal Strategies

A risk reversal is the tradable counterpart of the Slope. Long an out-of-the-money put against a short out-of-the-money call, it gains precisely when out-of-the-money puts richen relative to out-of-the-money calls, which is what a fall in the Slope means. The obvious alternatives do not have this property: a straddle is exposed to the Level and a butterfly to the Curvature, so neither would test the prediction we actually produce. And unlike the factor-mimicking portfolio of Section 6.1, a risk reversal is a position an investor can hold, at observable prices and with observable spreads. It is therefore the vehicle through which the information in our Slope predictions can be turned into a return.

Three choices keep the traded position aligned with the object being predicted. First, because our factors are extracted from moneyness standardized by ATM implied volatility, a strike set at a fixed percentage distance would drift in factor units as volatility changes; we therefore reset both strikes each month to a fixed distance in standardized moneyness, the units the factor is measured in. Second, a long put against a short call is a short position

in the underlying, so the unhedged P&L would be dominated by the index return rather than by the change in the skew; we delta-hedge the two legs and report the hedged P&L throughout. Third, the position is sized on the strength of the prediction rather than on its sign alone, which uses the magnitude of the predicted change and the time variation in its dispersion. Appendix B sets out the construction in full.

Performance. Table 6 reports the delta-hedged, conviction-weighted risk reversals after 5 implied-volatility-basis-point transaction costs. Partial least squares attains a net Sharpe ratio of 1.08 ($t = 3.3$, Sortino 3.22), followed by LinCombo (0.95), PCR (0.93), and the overall combination (0.84); the nonlinear models are positive but lower (0.5–0.65). The Sortino ratios exceed the Sharpe ratios for every model, confirming that the strategy's risk is not concentrated on the downside. That the latent-component linear models and their combination lead the tradable strategy, whereas the nonlinear combination leads the statistical R^2_{RW}, indicates that statistical predictability and tradability, though related, are not identical: the risk reversal rewards the directional and conviction content of the prediction, which the latent-component models capture well. Figure 6 plots the cumulative hedged P&L.

[Insert Table 6 and Figure 6 here]

6.3 Transaction Cost Sensitivity and Maturity Robustness

Transaction costs are measured in implied-volatility basis points (IV bp), which translate into index-point costs through the option's vega. For a TAIEX ATM option at 60-day maturity, one IV basis point corresponds to roughly 0.2 index points, and the round-trip charge on both legs of the risk reversal averages about 0.4 index points per IV bp; our baseline 5 IV bp cost is conservative relative to the typical 3–10 IV bp bid–ask spreads of liquid TAIEX options. Panel A of Figure 7 displays the risk-reversal Sharpe ratio as a function of transaction costs for the leading specification in each model family by out-of-sample R^2_{RW}.

At 5 IV bp it is 1.08 for PLS, 0.65 for random forest, and 0.63 for NLCombo; at 10 IV bp, 0.99, 0.55, and 0.53. The break-even cost is near 54 IV bp for PLS and in the mid-thirties for the two nonlinear specifications, in every case several times the spreads that liquid TAIEX options actually trade at. Panel B examines maturity robustness by constructing risk reversals at 30, 60, 90, 120, and 180 days. Short-dated risk reversals perform best for all three specifications, with 30-day Sharpe ratios of 1.09 (PLS), 0.72 (random forest), and 0.74 (NLCombo), consistent with the information-incorporation hypothesis that near-term options respond more rapidly to predicted Slope changes. The decay is monotone for PLS but not for the two tree-based specifications, which recover at 180 days after a trough at 120; we do not read the profile beyond the short end as informative. Figure 8 plots realized against predicted monthly changes in all three factors; the Slope row shows the tightest alignment with the 45-degree line for every model.

[Insert Figures 7 and 8 here]

7 Conclusion

Predictability in the IVS is concentrated in its asymmetry. The Level and the Curvature move largely unpredictably from month to month. The Slope, which prices the covariance between returns and volatility, responds to the economic state in a way that survives a demanding out-of-sample benchmark and an implementable option position.

Three implications follow. The surface should not be treated as one object to be predicted: its dimensions carry different economic content and different degrees of predictability, so evaluating them together conceals the only result there is to find. The information that moves the skew is not confined to the option market, since a predictor set that excludes every option-derived variable retains predictive content on its own, which locates the equilibrium skew in the broader macro-financial state. And statistical predictability is not tradability: a prediction acquires economic value only when the

position taken on it is exposed to the quantity actually predicted, which in options is a design problem rather than an implementation detail. How far these conclusions travel we cannot say, at higher frequencies or in markets with different investor bases.

References

Aït-Sahalia, Y., Li, C., and Li, C. X. (2021). Implied stochastic volatility models. *Review of Financial Studies*, 34(1):394–450.

Andersen, T. G., Fusari, N., and Todorov, V. (2015). Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83(3):1081–1145.

Audrino, F. and Colangelo, D. (2010). Semi-parametric forecasts of the implied volatility surface using regression trees. *Statistics and Computing*, 20(4):421–434.

Backus, D., Chernov, M., and Martin, I. (2011). Disasters implied by equity index options. *Journal of Finance*, 66(6):1969–2012.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4):1593–1636.

Bali, T. G., Beckmeyer, H., Mörke, M., and Weigert, F. (2023). Option return predictability with machine learning and big data. *Review of Financial Studies*, 36(9):3548–3602.

Bates, D. S. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, 94(1–2):181–238.

Bekaert, G. and Wu, G. (2000). Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, 13(1):1–42.

Bernales, A. and Guidolin, M. (2014). Can we forecast the implied volatility surface dynamics of equity options? predictability and economic value tests. *Journal of Banking & Finance*, 46:326–342.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). Bond risk premiums with machine learning. *Review of Financial Studies*, 34(2):1046–1089.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, pages 177–181, Washington, DC. American Statistical Association.

Bollen, N. P. B. and Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance*, 59(2):711–753.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11):4463–4492.

Bollerslev, T. and Todorov, V. (2011). Tails, fears, and risk premia. *Journal of Finance*, 66(6):2165–2211.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671.

Britten-Jones, M. and Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *Journal of Finance*, 55(2):839–866.

Caldara, D. and Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, 112(4):1194–1225.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531.

Carr, P. and Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 22(3):1311–1341.

Carr, P. and Wu, L. (2020). Option profit and loss attribution and pricing: A new framework. *Journal of Finance*, 75(4):2271–2316.

Chalamandaris, G. and Tsekrekos, A. E. (2010). Predictable dynamics in implied volatility surfaces from OTC currency options. *Journal of Banking & Finance*, 34(6):1175–1188.

Chen, D., Guo, B., and Zhou, G. (2023). Firm fundamentals and the cross-section of implied volatility shapes. *Journal of Financial Markets*, 63:100771.

Chen, L., Pelger, M., and Zhu, J. (2024). Deep learning in asset pricing. *Management Science*, 70(2):714–750.

Christensen, K., Siggaard, M., and Veliyev, B. (2023). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, 21(5):1680–1727.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4):407–432.

Christoffersen, P., Heston, S., and Jacobs, K. (2009). The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science*, 55(12):1914–1932.

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.

Collin-Dufresne, P., Junge, B., and Trolle, A. B. (2024). How integrated are credit and equity markets? Evidence from index options. *Journal of Finance*, 79(2):949–992.

Cont, R. and da Fonseca, J. (2002). Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1):45–60.

Davis, S. J. (2016). An index of global economic policy uncertainty. Working Paper 22740, National Bureau of Economic Research, Cambridge, MA.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.

Dumas, B., Fleming, J., and Whaley, R. E. (1998). Implied volatility functions: Empirical tests. *Journal of Finance*, 53(6):2059–2106.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56(1):329–352.

Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*, 33(5):2326–2377.

Gârleanu, N., Pedersen, L. H., and Poteshman, A. M. (2009). Demand-based option pricing. *Review of Financial Studies*, 22(10):4259–4299.

Gonçalves, S. and Guidolin, M. (2006). Predictable dynamics in the S&P 500 index options implied volatility surface. *Journal of Business*, 79(3):1591–1635.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies*, 13(2):433–451.

Jiang, G. J. and Tian, Y. S. (2005). The model-free implied volatility and its information content. *Review of Financial Studies*, 18(4):1305–1342.

Kearney, F., Shang, H. L., and Sheenan, L. (2019). Implied volatility surface predictability: The case of commodity markets. *Journal of Banking & Finance*, 108:105657.

Kelly, B. T., Kuznetsov, B., Malamud, S., and Xu, T. A. (2023). Deep learning from implied volatility surfaces. Research Paper 23-60, Swiss Finance Institute.

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

Konstantinidi, E., Skiadopoulos, G., and Tzagkaraki, E. (2008). Can the evolution of implied volatility be forecasted? evidence from european and us implied volatility indices. *Journal of Banking & Finance*, 32(11):2401–2411.

Kozak, S., Nagel, S., and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292.

Litterman, R. and Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, 1(1):54–61.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Medvedev, N. and Wang, Z. (2022). Multistep forecast of the implied volatility surface using deep learning. *Journal of Futures Markets*, 42(4):645–667.

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

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2):228–250.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Pan, J. and Poteshman, A. M. (2006). The information in option volume for future stock prices. *Review of Financial Studies*, 19(3):871–908.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23(2):821–862.

Rapach, D. E. and Zhou, G. (2020). Time-series and cross-sectional stock return forecasting: New machine learning methods. In Jurczenko, E., editor, *Machine Learning for Asset Management: New Developments and Financial Applications*, pages 1–33. Wiley, Hoboken, NJ.

Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance*, 49(3):771–818.

Skiadopoulos, G., Hodges, S., and Clewlow, L. (2000). The dynamics of the S&P 500 implied volatility surface. *Review of Derivatives Research*, 3(3):263–282.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

Zhang, C., Zhang, Y., Cucuringu, M., and Qian, Z. (2024). Volatility forecasting with machine learning and intraday commonality. *Journal of Financial Econometrics*, 22(2):492–530.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2):301–320.

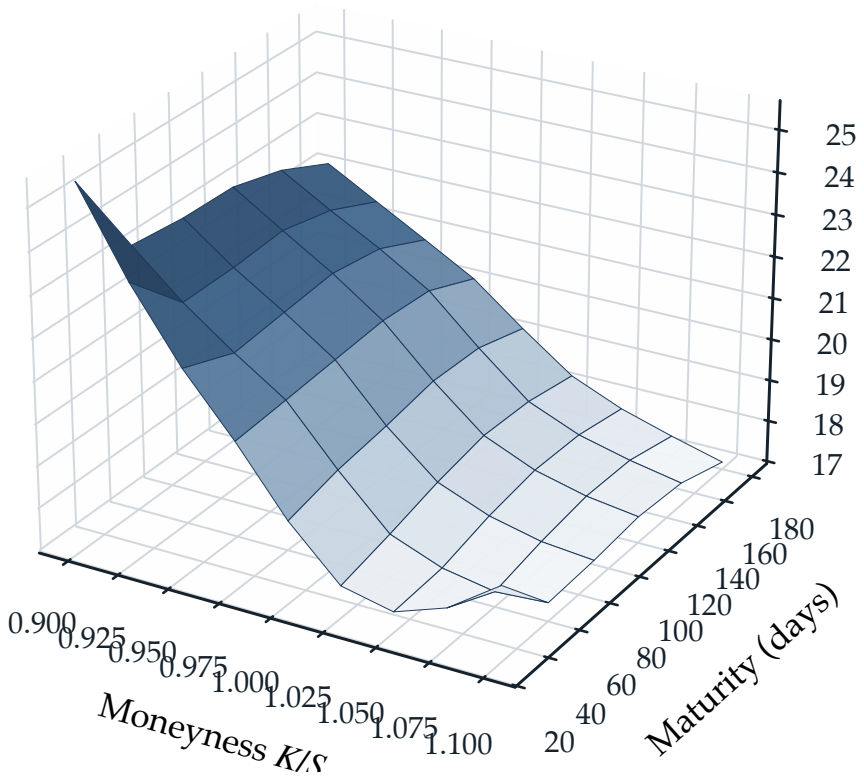


Figure 1: The average TAIEX IVS. This figure plots the average Black–Scholes settlement IVS of TAIEX index options over the 2002–2024 sample. For each month the observed option implied volatilities are interpolated onto a fixed grid of moneyness (K/S) and time to maturity (calendar days), and the figure displays the average of these monthly grids across all 273 months; implied volatility is in percentage points. The surface is constructed directly from observed option-level implied volatilities and is independent of the quadratic-regression factor extraction of Section 3.2.



Figure 2: Time series of the IVS factors. This figure plots the monthly Level (a), Slope (b), and Curvature (c) factors, each obtained as the last-trading-day coefficient of the daily cross-sectional quadratic regression IV $= a + b\,m + c\,m^2$, where $m = \log(K/S)/(\sigma^{\text{ATM}}\sqrt{\tau/365})$ is ATM-IV-standardized log-moneyness (equation 5). The Level equals at-the-money implied volatility; a more negative Slope indicates a steeper equity skew; a larger Curvature indicates a more pronounced smile. Shaded bands mark the 2008 global financial crisis and the 2020 pandemic.

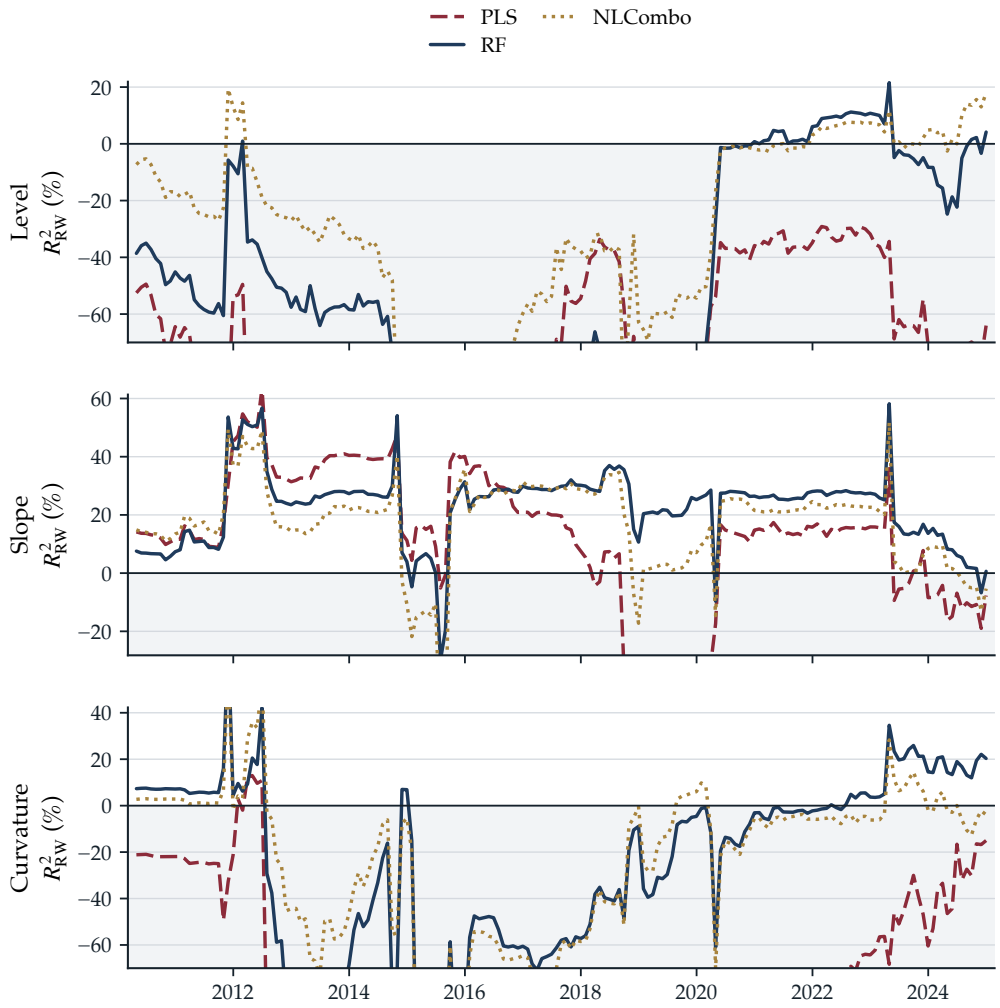


Figure 3: Rolling out-of-sample predictive accuracy by factor. This figure plots the rolling 36-month out-of-sample R^2_{RW} (equation 9) for each factor level, computed within each trailing 36-month window against the random-walk benchmark, for the leading specification in each model family by out-of-sample R^2_{RW}: partial least squares (linear), random forest (nonlinear), and NLCombo (combination). Values above zero (unshaded) indicate that the model outperforms the random walk over that window; values below zero (shaded) indicate underperformance. The vertical axis is truncated for readability.

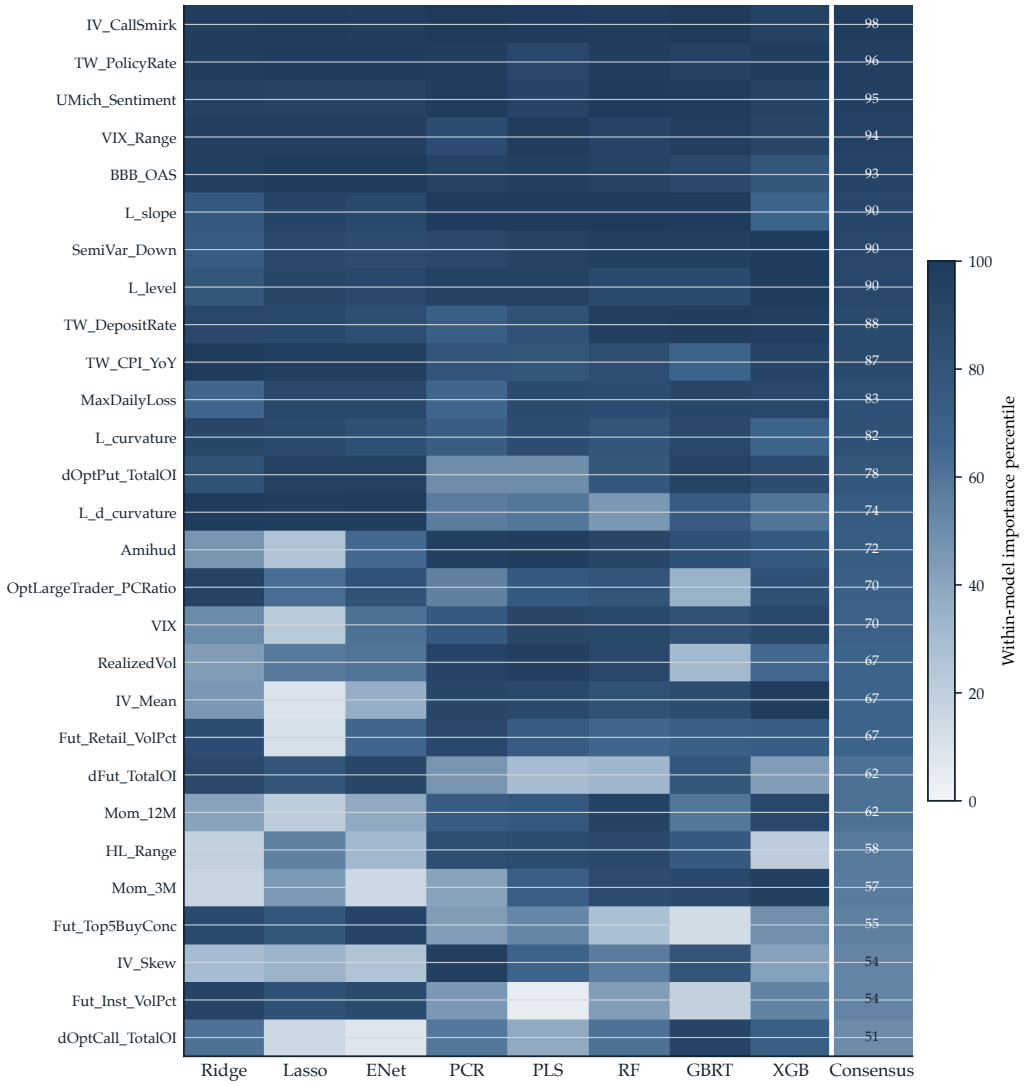


Figure 4: Feature importance across models for the Slope factor. This heatmap reports feature importance for the Slope across all eight individual models. Each column is a model and each row a predictor; a cell gives the predictor's importance percentile within that model (darker is more important), so importances are comparable across models with differing natural scales. Importance is impurity importance for the nonlinear models (RF, GBRT, XGB), the average absolute standardized coefficient $|\hat{\beta}|$ for the penalized linear models and PLS, and |PCA loadings $\times$ coefficient| for PCR. The "Consensus" column is the cross-model average percentile; rows are the union of each model's top predictors, ordered by consensus. Variable definitions are in Appendix A.

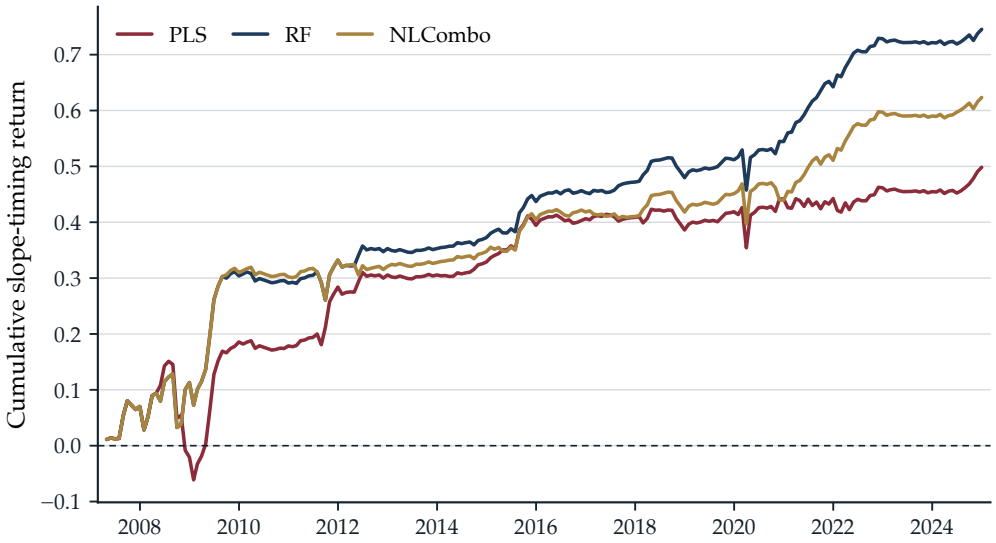


Figure 5: Cumulative performance of the directional Slope-timing strategy. This figure plots the cumulative return of the directional Slope-timing strategy over the 213 out-of-sample months (2007:04–2024:12). Each month the strategy takes a unit position in the Slope factor-mimicking portfolio in the predicted direction, $\text{sign}(\hat{b}_{t+1} - b_t)$, earning $\text{sign}(\hat{b}_{t+1} - b_t) \times (b_{t+1} - b_t)$ in Slope-factor units. Because the Slope factor is not a traded asset, the strategy is a prediction-quality diagnostic rather than an investable return. Models are the leading specification in each model family by out-of-sample R^2_{RW}: partial least squares (linear), random forest (nonlinear), and NLCombo (combination).

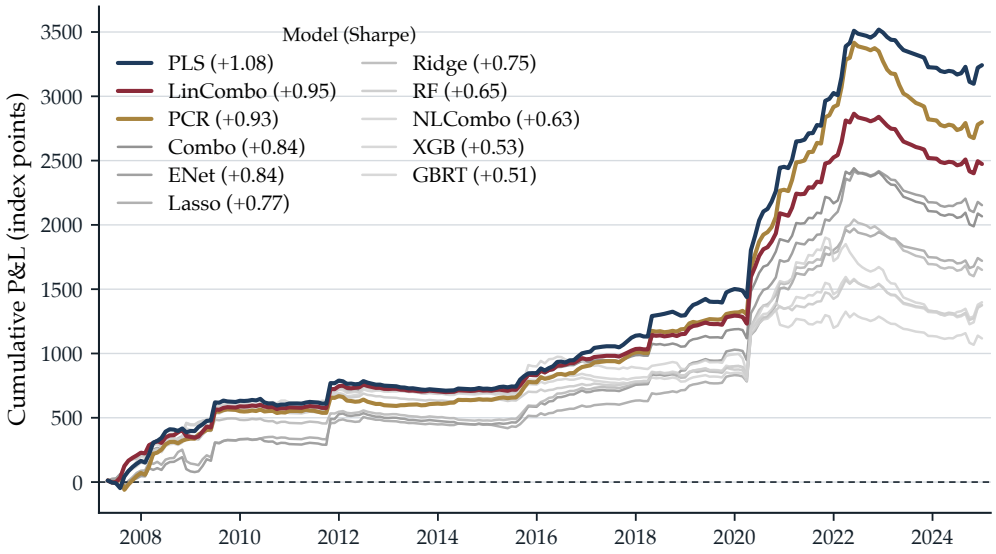


Figure 6: Cumulative P&L of the risk-reversal strategy. This figure plots the cumulative delta-hedged P&L of the conviction-weighted risk reversal for each model over the 213 out-of-sample months. The risk reversal is long one out-of-the-money put and short one out-of-the-money call at adaptive strikes $K = S\exp(\pm\,\sigma^{\mathrm{ATM}}\sqrt{\tau/365}\,\delta)$ with $\delta = 1$ and $\tau = 60$ days (equation 10), delta-hedged so the P&L reflects the volatility skew rather than the direction of the underlying, sized by the conviction rule of equation (11), and reported net of 5 IV bp transaction costs. P&L is in TAIEX index points. The three highest-Sharpe models are drawn in colour and the remainder in grey; the legend reports annualized Sharpe ratios.

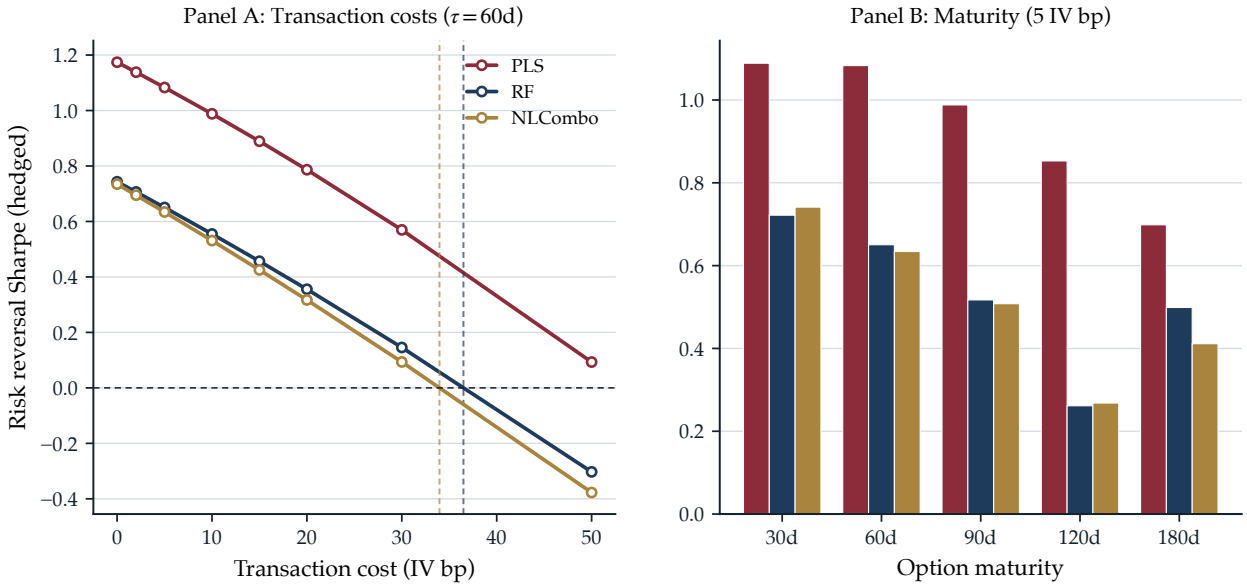


Figure 7: Robustness of the risk reversal to trading assumptions. This figure reports the annualized Sharpe ratio of the delta-hedged, conviction-weighted risk reversal for the leading specification in each model family by out-of-sample R^2_{RW}: partial least squares (linear), random forest (nonlinear), and NLCombo (combination). Panel A varies the round-trip transaction cost from 0 to 50 implied-volatility basis points (IV bp) at $\tau = 60$ days, with costs charged on each option leg in proportion to its vega; the dashed vertical lines mark the break-even cost for the random forest and NLCombo, while the PLS break-even, at approximately 54 IV bp by linear extrapolation of the 30–50 IV bp segment, lies beyond the plotted range. Panel B varies the option maturity from 30 to 180 days at a cost of 5 IV bp.

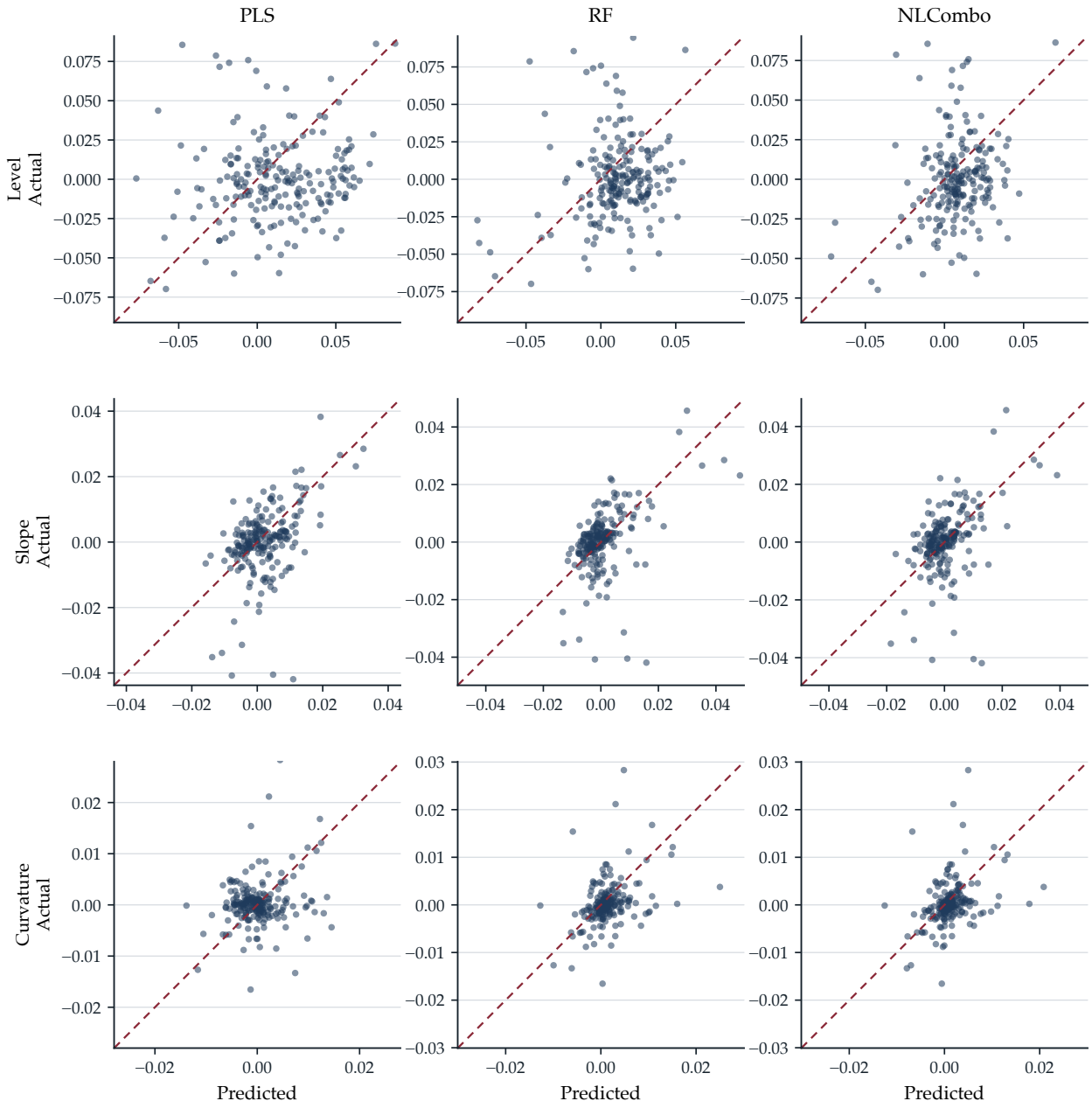


Figure 8: Realized versus predicted factor changes. This figure plots realized against predicted monthly factor changes for the Level, Slope, and Curvature, one row per factor and one column per model, pooled over the 213 out-of-sample months; the 45-degree line denotes perfect prediction. Models are the leading specification in each model family by out-of-sample R^2_{RW}: partial least squares (linear), random forest (nonlinear), and NLCombo (combination).

Table 1: Summary statistics

Notes. This table reports summary statistics for the options data and the extracted IVS factors. Panel A describes the 552,955 filtered option–day observations (5,591 trading days, April 2002–December 2024) used to estimate the daily cross-sectional regressions: Implied volatility is the Black–Scholes settlement implied volatility; Moneyness is the strike-to-spot ratio K/S; Maturity is calendar days to expiration; and Underlying is the TAIEX spot index. Panel B describes the 273 monthly factor levels (a, b, c), obtained as the last-trading-day coefficients of the daily regression IV $= a + bm + cm^2$ with ATM-IV-standardized moneyness $m = \log(K/S)/(\sigma^{\text{ATM}}\sqrt{\tau/365})$ (equation 5). Skew. and Kurt. are the third and fourth standardized moments (kurtosis in excess of the normal value of 3).

	Mean	Std. Dev.	Min	Median	Max	Skew.	Kurt.
Panel A: Filtered options data (552,955 option–days)							
Implied volatility	0.198	0.071	0.065	0.183	1.289	1.42	3.54
Moneyness (K/S)	0.995	0.081	0.850	0.994	1.150	0.05	−1.05
Maturity (days)	77.8	66.8	7	57	279	1.25	0.59
Underlying index	11,911	5,001	3,794	10,273	24,440	0.71	−0.58
Panel B: Monthly IVS factor levels (273 months)							
a (Level)	0.195	0.072	0.085	0.177	0.519	1.43	2.36
b (Slope)	−0.020	0.017	−0.127	−0.016	0.005	−3.13	13.94
c (Curvature)	0.007	0.014	−0.024	0.004	0.158	8.23	83.10

Table 2: Out-of-sample R^2_{RW} (%) for monthly IVS factor levels

Notes. This table reports the out-of-sample $R^2_{\text{RW}} = 1 - \text{MSE}_{\text{model}}/\text{MSE}_{\text{RW}}$ (equation 9) for each model's prediction of the next-month level of the Level, Slope, and Curvature factors, over 213 out-of-sample months (2007:04–2024:12). A positive value means the model's mean-squared prediction error is smaller than that of the random-walk benchmark (equation 8); a value of $+16\%$ means a 16% reduction in MSE relative to the random walk. Models are estimated on an expanding window with all hyperparameters chosen by 5-fold time-series cross-validation (Appendix Table A2). Statistical significance is from the one-sided Clark and West (2007) test with Newey and West (1987) HAC standard errors: ***, **, * denote $p < 0.01, 0.05, 0.10$; stars are shown only for $R^2_{\text{RW}} > 0$. The Clark–West statistic tests the null that the random walk is the true model against the alternative that the larger model has lower population mean-squared error; because it corrects the sample loss differential for the noise introduced by estimating parameters whose true values are zero, it can reject even when $R^2_{\text{RW}} < 0$. Restricting stars to $R^2_{\text{RW}} > 0$ therefore prevents significance from being read as in-sample improvement where there is none.

Model	Level (a)	Slope (b)	Curvature (c)
Linear models			
Ridge	-67.98	-30.91	-29.37
Lasso	-21.04	-42.94	-39.61
Elastic Net	-46.57	-12.29	-50.20
PCR	-55.01	$+7.32^{***}$	-7.87
PLS	-80.65	$+14.60^{***}$	-26.48
Nonlinear models			
Random Forest	-33.80	$+13.97^{***}$	$+5.81$
Gradient Boosting	-15.58	$+10.07^{***}$	$+6.46$
XGBoost	-15.75	$+10.37^{***}$	-27.56
Combinations			
LinCombo	-36.29	-2.70	-23.48
NLCombo	-13.06	$\mathbf{+16.16}^{***}$	$+1.60$
Combo	-15.67	$+9.90^{***}$	-8.81

Table 3: Predictor-block prediction of the Slope factor (R^2_{RW}, %)

Notes. This table reports the out-of-sample R^2_{RW} (equation 9) for predictions of the Slope factor level that use only the predictors in a single information block, for random forest (RF) and partial least squares (PLS), over 213 out-of-sample months. "# vars" is the number of predictors in the block; block membership follows Section 4.1 and Appendix A. "Lagged IVS only" uses the six lagged factor variables; "All ex-option-derived" uses every external variable except the options microstructure and option-derived measures; "Full" is the 101-predictor model. Significance stars are omitted for the reason given in Table 2: the Clark–West statistic can be significant even when $R^2_{\text{RW}} < 0$.

Block	# vars	RF	PLS
Lagged IVS only	6	+6.61	+8.39
Option microstructure	14	+9.84	−1.67
Underlying market	12	−1.75	+0.73
Global / cross-market	13	−5.70	−17.68
Macro / monetary	18	+3.31	−11.06
Geopolitical	6	−25.10	−27.85
TW investor positioning	30	−3.37	−17.54
All ex-option-derived	79	+10.20	+6.19
Full (all)	101	+13.97	+14.60

Table 4: Feature importance for the Slope factor (top 8)

Notes. This table reports the eight most important predictors of the Slope for the leading specification in each model family by out-of-sample R^2_{RW}: partial least squares (linear), random forest (nonlinear), and NLCombo (combination). Importance is the average absolute standardized coefficient $|\hat{\beta}|$ for PLS and impurity importance for the tree models; NLCombo importance is the equal-weight average of the normalized importances of its three constituents. For each model it is averaged over the 213 out-of-sample estimation windows and expressed as a percentage of that model's total importance, so the columns are internally comparable but the absolute scales differ across methods. The economic category of each predictor is given in parentheses; these are descriptive labels rather than the six information channels of Section 4.1, since the Taiwan policy and deposit rates belong to the Taiwan channel while the credit spread and consumer sentiment both sit in the macroeconomic and monetary channel; full variable definitions are in Appendix A.

PLS (linear)		Random Forest (nonlinear)		NLCombo (combination)	
Variable	%	Variable	%	Variable	%
Lagged Slope (IVS)	2.53	Lagged Slope (IVS)	5.87	Call smirk (option)	7.53
Call smirk (option)	2.30	UMich sentiment (macro)	4.74	Lagged Slope (IVS)	4.46
VIX range (global)	2.29	Call smirk (option)	4.69	UMich sentiment (macro)	4.35
Amihud illiq. (und.)	2.28	TW policy rate (mon.)	3.70	TW deposit rate (mon.)	3.58
Realized vol. (und.)	2.23	TW deposit rate (mon.)	3.48	Downside semivar. (und.)	3.51
BBB OAS (credit)	2.21	Downside semivar. (und.)	2.89	TW policy rate (mon.)	3.38
Lagged Level (IVS)	2.20	Momentum 12m (und.)	2.86	Lagged Level (IVS)	3.20
Downside semivar. (und.)	2.18	BBB OAS (credit)	2.70	VIX range (global)	2.68

Table 5: Directional Slope-timing performance (annualized)

Notes. This table reports the performance of the directional Slope-timing strategy over 213 out-of-sample months. Each month the strategy takes a unit position in the Slope factor-mimicking portfolio in the predicted direction $\text{sign}(\hat{b}_{t+1} - b_t)$, earning $\text{sign}(\hat{b}_{t+1} - b_t) \times (b_{t+1} - b_t)$; because the Slope factor is not a traded asset, this is a prediction-quality diagnostic rather than an investable return. Mean and Std are in factor units $\times 10^{-2}$ and annualized; Sharpe and Sortino are annualized ratios (the Sortino ratio replaces total volatility with downside deviation); Hit is the share of months in which the predicted direction is correct. The highest Sharpe ratio is in bold.

Model	Mean	Std	Sharpe	Sortino	Hit %
Linear models					
Ridge	2.49	5.52	0.45	0.50	64.3
Lasso	2.94	5.51	0.53	0.58	64.8
Elastic Net	3.38	5.48	0.62	0.64	66.7
PCR	3.18	5.49	0.58	0.61	63.8
PLS	2.81	5.51	0.51	0.52	62.0
Nonlinear models					
Random Forest	4.20	5.44	0.77	0.78	67.6
Gradient Boosting	3.29	5.49	0.60	0.64	63.8
XGBoost	4.66	5.40	**0.86**	1.24	64.3
Combinations					
LinCombo	3.73	5.47	0.68	0.75	65.7
NLCombo	3.51	5.48	0.64	0.66	66.2
Combo	4.16	5.44	0.76	0.93	67.1

Table 6: Risk-reversal strategy performance (delta-hedged, conviction-weighted)

Notes. This table reports the performance of the implementable risk-reversal strategy over 213 out-of-sample months: long one out-of-the-money put and short one out-of-the-money call at adaptive strikes $K = S\exp(\pm\,\sigma^{\text{ATM}}\sqrt{\tau/365}\,\delta)$ with $\delta = 1$ and $\tau = 60$ days (equation 10), delta-hedged so that the P&L isolates the volatility skew, sized by the conviction rule of equation (11), and net of 5 IV bp transaction costs. Mean and Std are the annualized P&L moments in TAIEX index points; Sharpe and Sortino are annualized, scale-invariant ratios (Sortino uses downside deviation); Hit is the share of profitable months. The highest Sharpe ratio is in bold.

Model	Mean	Std	Sharpe	Sortino	Hit %
Linear models					
Ridge	93	124	0.75	2.24	50.7
Lasso	97	125	0.77	1.82	53.5
Elastic Net	121	145	0.84	1.95	54.0
PCR	158	170	0.93	1.97	61.5
PLS	183	169	**1.08**	3.22	56.3
Nonlinear models					
Random Forest	77	119	0.65	2.12	48.4
Gradient Boosting	63	124	0.51	1.06	47.9
XGBoost	78	147	0.53	0.94	48.8
Combinations					
LinCombo	139	146	0.95	2.70	54.0
NLCombo	79	124	0.63	1.61	48.4
Combo	116	138	0.84	2.30	52.1

A Variable Definitions and Model Hyperparameters

Table A1 provides complete definitions for the 95 external predictors, organized by the six information channels of Section 4.1, together with the six lagged IVS variables that complete the 101-predictor set. Table A2 reports the model hyperparameter grids; all are selected by 5-fold time-series cross-validation within each expanding window.

Table A1: Variable definitions

Variable	Description
Channel 1: Underlying Market (12)	
MktReturn	Monthly TAIEX index return
RealizedVol	Annualized standard deviation of daily TAIEX returns within the month
RealizedSkew	Skewness of daily TAIEX returns within the month
RealizedKurt	Excess kurtosis of daily TAIEX returns within the month
MaxDailyLoss	Largest single-day loss within the month (absolute value)
MaxDailyGain	Largest single-day gain within the month
Amihud	Amihud (2002) illiquidity ratio (average |return|/volume)
SemiVar_Down	Downside semivariance of daily returns (negative returns only)
SemiVar_Up	Upside semivariance of daily returns (positive returns only)
HL_Range	Average daily high−low range divided by closing price
Mom_3M, Mom_12M	TAIEX cumulative return over the past 3 and 12 months

continued on next page

Table A1, continued

Variable	Description
Channel 2: Options Microstructure (14)	
Opt_dOI	Monthly change in total option open interest
dOpt_NumContracts	Monthly percentage change in the number of contracts traded
PC_Ratio_OI	Put-to-call open-interest ratio
IV_Mean, IV_Std, IV_Skew, IV_Kurt	Cross-sectional mean, standard deviation, skewness, and kurtosis of implied volatility across options
IV_Range	Maximum minus minimum implied volatility across options
IV_Smirk	Out-of-the-money put implied volatility minus at-the-money implied volatility (30-day)
IV_CallSmirk	Out-of-the-money call implied volatility minus at-the-money implied volatility (30-day)
IV_Butterfly	Average out-of-the-money put/call implied volatility minus at-the-money implied volatility

continued on next page

Table A1, continued

Variable	Description
IV_TermSlope	180-day minus 30-day at-the-money implied volatility
dOptCall_TotalOI, dOptPut_TotalOI	Monthly percentage change in call and put total open interest
Channel 3: Global and Cross-Market (13)	
VIX, dVIX	CBOE VIX level and its monthly change
VIX_MA3	Three-month moving average of the VIX
VIX_TermSlope	VIX term-structure slope (longer- minus shorter-dated VIX)
VIX_Range	Intra-month high minus low of the VIX
VIX_Vol	Standard deviation of daily VIX changes within the month
dUSDTWD, dJPYUSD, dEURUSD, dDXY, dKRWUSD	Monthly changes in the USD/TWD, JPY/USD, EUR/USD, U.S. dollar index, and KRW/USD exchange rates
NASDAQ_Ret	Monthly NASDAQ Composite return

continued on next page

Table A1, continued

Variable	Description
OIL_Ret	Monthly WTI crude-oil return
Channel 4: Macroeconomic and Monetary (18)	
US10Y, dUS10Y; US2Y, dUS2Y; dUS3M	U.S. 10-year and 2-year Treasury yields and their changes; 3-month T-bill yield change
TermSpread_10_2, TermSpread_10_3M, dTermSpread	10-year minus 2-year and 10-year minus 3-month yield spreads, and the change in the former
FedFunds, dFedFunds	Federal Funds target rate and its monthly change
HY_OAS, dHY_OAS; BBB_OAS, dBBB_OAS	High-yield and BBB corporate option-adjusted credit spreads and their monthly changes
EPU_Global	Global Economic Policy Uncertainty index (Davis, 2016)
US_Inflation_YoY	U.S. CPI inflation, year-over-year

continued on next page

Table A1, continued

Variable	Description
UMich_Sentiment, dSentiment	University of Michigan consumer-sentiment index and its monthly change
Channel 5: Geopolitical Risk (6)	
GPR_Global, dGPR_Global	Global geopolitical-risk index (Caldara and Iacoviello, 2022) and its monthly change
GPR_Taiwan, dGPR_Taiwan	Taiwan-specific geopolitical-risk index and its monthly change
GPR_China, GPR_Korea	China- and Korea-specific geopolitical-risk indices
Channel 6: Taiwan Investor Positioning (30); source: Taiwan Economic Journal (TEJ)	
TW_CPI_YoY, dTW_CPI	Taiwan CPI year-over-year growth and its change
TW_PolicyRate, dTW_PolicyRate	Central Bank of Taiwan rediscount (policy) rate and its change
TW_DepositRate	One-year time-deposit rate

continued on next page

Table A1, continued

Variable	Description
TW_M2_YoY	M2 money-supply growth (year-over-year)
TW_PMI	Taiwan manufacturing Purchasing Managers' Index
TW_Sentiment, dTW_Sentiment	Taiwan business-cycle indicator and its change
TW_Unemployment	Taiwan unemployment rate
dTW_USDTWD	Monthly percentage change in the USD/TWD exchange rate
TW_ExportOrders_YoY	Taiwan export-order growth (year-over-year)
ShortMarginRatio	Aggregate short-balance-to-margin-balance ratio
dMarginBalance_Pct, dShortBalance_Pct	Monthly percentage changes in aggregate margin and short balances
dFut_TotalOI	Monthly percentage change in total futures open interest
Fut_Top5NetOI, Fut_Top10NetOI	Net position (buy minus sell) of the top-5 and top-10 futures traders

continued on next page

Table A1, continued

Variable	Description
Fut_Top5BuyConc, Fut_Top5SellConc, Fut_ConcSpread	Top-5 buy-side and sell-side concentration and their spread
Fut_Inst_VolPct, Fut_Retail_VolPct	Institutional and retail shares of futures trading volume
Fut_Foreign_VolPct	Foreign-investor (TEJ category A4) share of futures volume
Fut_Foreign_NetRatio	Foreign-investor net trading ratio, (buy−sell)/(buy+sell)
OptCall_Top5Net, OptPut_Top5Net	Net call and put positions of the top-5 option traders
OptLargeTrader_PCRatio, dOptLarge-Trader_PCRatio	Large-trader put-call open-interest ratio and its change
OptLargeTrader_Sentiment	Net call minus net put positions of the top-5 option traders

continued on next page

Table A1, continued

Variable	Description
Derived risk measures (2)	
VRP	Variance risk premium: cross-sectional mean implied volatility minus realized volatility
VIX_IV_Spread	VIX minus the TAIEX average implied volatility
Lagged IVS factors (6); used in every model	
L_level, L_slope, L_curvature	One-month lags of the Level, Slope, and Curvature factor levels
L_d_level, L_d_slope, L_d_curvature	One-month lags of the monthly changes in the three factors

Table A2: Model hyperparameter grids (5-fold time-series CV)

Model	Hyperparameter grid
Ridge	$\lambda \in \{10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$
Lasso	$\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$
Elastic Net	$\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$, $\rho \in \{0.2, 0.5, 0.8\}$
PCR	number of components $\in \{1, 3, 5, 10, 15, 20\}$
PLS	number of components $\in \{1, 3, 5, 10, 15\}$
Random Forest	300 trees; max_depth $\in \{2, 3, 5\}$; min_leaf 5; max_features = $\sqrt{p}$
Gradient Boosting	300 trees, Huber loss, subsample 0.5; max_depth $\in \{1, 2\}$, lr $\in \{0.01, 0.1\}$
XGBoost	300 trees, subsample/colsample 0.5, ℓ_1/ℓ_2 reg.; max_depth $\in \{1, 2\}$, lr $\in \{0.01, 0.1\}$

B Construction of the Risk-Reversal Strategy

This appendix details the trading vehicle used in Section 6.2.

Adaptive strikes. Our factors are extracted from implied volatility regressed on moneyness standardized by ATM implied volatility and the square root of maturity (equation 4). A fixed strike distance in percentage terms would therefore drift in factor units as volatility changes: the same 5% out-of-the-money strike sits far in the tail in calm markets and close to the money in turbulent ones. To keep the position at a constant distance in the units the factor is measured in, we reset both strikes each month to

$$K_{\text{put}} = S \cdot \exp\left(-\sigma^{\text{ATM}}\sqrt{\tau/365} \cdot \delta\right), \quad K_{\text{call}} = S \cdot \exp\left(+\sigma^{\text{ATM}}\sqrt{\tau/365} \cdot \delta\right), \tag{10}$$

with $\delta = 1$, one unit of standardized moneyness on each side, and a baseline maturity of $\tau = 60$ days. Implied volatilities at these strikes are read off the daily surface by interpolation, with strikes outside the observed grid clipped to its edge. Positions are held one month and rolled.

Delta hedging. We hedge the net Black–Scholes delta of the two legs at initiation, using the delta implied by each leg's own interpolated implied volatility, and report the hedged P&L throughout. The unhedged P&L is computed as well and is uniformly noisier.

Conviction weighting. A rule that takes a unit position based only on the sign of the prediction discards both the magnitude of the predicted change and the time variation in risk. For an investor with relative risk aversion γ allocating to a single risky strategy whose conditional expected return and variance are μ_t and σ_t^2, the optimal weight is $w_t^* = \mu_t/(\gamma\,\sigma_t^2)$ (Markowitz, 1952; Merton, 1969); the log-utility, or Kelly, rule is the special case $\gamma = 1$ (Kelly, 1956). Because the risk reversal is constructed to carry approximately unit exposure to the Slope, its conditional expected payoff is approximately proportional to

the predicted change, $\mu_t \approx \kappa\, s_t$ with $s_t \equiv \hat{b}_{t+1} - b_t$ and $\kappa < 0$, since the position gains when the Slope becomes more negative. Hence $w_t^* \propto -s_t/\sigma_t^2$. We implement a robust, leverage-constrained version,

$$w_t = -\operatorname{sign}(s_t) \cdot \min\Big(\max\left(|s_t|/\hat{\sigma}_{s,t},\ \underline{w}\right),\ \overline{w}\Big), \qquad \hat{\sigma}_{s,t} = \operatorname{std}\left(\{|s_\tau| : \tau < t\}\right), \tag{11}$$

with bounds $(\underline{w}, \overline{w}) = (0.3, 3)$, and form the strategy return $r_{t+1} = w_t\, R^{\mathrm{RR}}_{t+1}$, where R^{RR}_{t+1} is the delta-hedged risk-reversal P&L. The first six out-of-sample months carry unit weight, since $\hat{\sigma}_{s,t}$ is not yet estimable.

Two design choices deserve comment. Standardizing the prediction by its own expanding-window dispersion converts it into a z-score, so the position responds to how unusual the current signal is rather than to its raw size; this applies to the dispersion of the signal the same risk-scaling logic that volatility timing (Fleming et al., 2001) and time-series momentum (Moskowitz et al., 2012) apply to the dispersion of returns, and it uses only past data. The bounds implement fractional-Kelly shrinkage: unconstrained mean–variance weights are highly sensitive to estimation error in $\hat{\mu}$ and $\hat{\sigma}$ (Britten-Jones, 1999; DeMiguel et al., 2009), so capping the weight limits the damage from a mis-estimated signal. We stress that this sizing is applied to the option P&L, a genuine traded return for which the mean–variance and Kelly justifications are valid, and not to the factor change itself, which is not a return.

Transaction costs. Costs are charged on each leg in proportion to its Black–Scholes vega, in implied-volatility basis points, at both entry and exit. The baseline of 5 IV bp is conservative against the 3–10 IV bp bid–ask spreads of liquid TAIEX options; Section 6.3 varies the charge from 0 to 50 IV bp.