Research Inventy: International Journal of Engineering and Science
Vol.16, Issue 6 (June 2026), PP 62-70
Issn (e): 2278-4721, Issn (p):2319-6483, www.researchinventy.com

Transmission of cross-market information in Vietnam's steel industry: Economic foundations and portfolio management implications of the integrated learning framework for HPG stocks

Phan Dieu My
University of Da Nang, Vietnam University of Information and Communication Technology, Da Nang, Vietnam

ABSTRACT
This study investigates the economic mechanisms through which cross-market information transmits equity valuation in Vietnam's marginal steel sector, using Hoa Phat Group (HPG) as a representative case. Based on the Efficient Market Hypothesis (EMH), Modern Portfolio Theory (MPT), and multi-factor asset valuation theory, we identify three structural transmission channels – vertical supply chain linkages (upstream iron ore and HRC price shocks), horizontal competitive spillover effects (HSG and NKG peer steel companies), and intermediate streams finance (the driving force of the banking and brokerage sector) – and quantify their empirical implications through the Stacking Ensemble machine learning framework. Through 1,330 daily observations (April 2020 – November 2025) and a rigorously enforced Walk-Forward Validation protocol, the aggregate achieved MAPE = 1.57%, R^2 = 0.957, and Directional Accuracy = 51.5% on an out-of-sample test suite of 267 sessions. Analyses by PEST and Porter Five Forces show that HPG's price dynamics cannot be modeled in isolation: downstream construction demand proxies (CTD_Mom5, ranked 10th), iron ore positioning ratios (IronOre_ratio, ranked 12th), and brokerage momentum signals (SSI_Mom5, ranked 8th) outperforming all macroeconomic variables. The ± model's suitable prediction interval (499 VND, 76.4% empirical coverage) provides a data-driven basis for segmented Kelly position sizing and dynamic stop-loss correction, providing specific portfolio management implications for institutional investors navigating the market characterized by retail investor concentration high, information asymmetry and limited risk hedging tools.
Keywords: *Diffusion of cross-market information; Border markets; Vietnam's steel industry; Stacking complexes; Predict the right thing; Kelly Criteria; EMH; MPT.*

Date of Submission: 14-06-2026 Date of Acceptance: 27-06-2026

I. Introduction

Vietnam's stock market occupies a special place in the global financial landscape. As a marginal market undergoing rapid institutionalization – with a total listed market capitalization exceeding VND6 million billion by the end of 2025 and retail investor accounts exceeding VND9 million (SSC, 2025) – it exhibits characteristics that simultaneously invite and challenge conventional asset valuation assumptions. Individual investors account for about 85-90% of the daily trading value, amplifying herd behavior, momentum effects, and slow information absorption (Vo & Phan, 2019). These characteristics create a structural wedge between the theoretical standard of the Efficient Market Hypothesis and the observable microstructure of price formation, providing fertile ground for systematic, data-driven forecasting strategies.

In this market, Hoa Phat Group (code: HPG) is a preeminent industrial shareholder. The Group controls about 35% of the domestic construction steel market share and more than 40% of steel pipe production (Hoa Phat Group, 2025), while maintaining a market capitalization that is always in the top 5 on the Ho Chi Minh City Stock Exchange (HOSE). The average daily trading volume is more than 25 million shares and has exceeded 75 million in peak sessions (VCBS, 2024). HPG's price trajectory during the study period (April 2020 - November 2025) spans an extraordinary range - from the COVID-19 bottom of approximately VND 15,900 to the all-time high of VND 64,800 - creating a benchmark daily profit deviation of 1.85% compared to VN-Index's 1.2-1.5%, making HPG an ideal laboratory for testing multi-information integration models source.

The central thesis of this paper is that HPG's daily closing price cannot be fully modeled as a single stock process. The three structural economic channels embed cross-market signals into a company's equity price with systematic, exploitable relationships. Firstly, as a vertically integrated blast furnace manufacturer that imports 55-60% of production costs in the form of iron ore and coke, HPG faces a global commodity cycle with a transfer delay of one to three days (Chen et al., 2019). Secondly, as a market leader in centralized monopoly together with Hoa Sen Group (HSG) and Nam Kim Group (NKG), competitive dynamics, inventory cycles, and shared macro shocks create strong price co-movements (Hou & Moskowitz, 2005). Third, the dominant retail investor base is intermediated through brokerage firms that simultaneously manage margin lending – making the dynamics of the brokerage sector a leading indicator of leverage and risk appetite across the market (Barberis et al., 2005).

These channels are not merely descriptive; They have direct implications for model architecture. HPG's own time-series limited forecasting system systematically removes the information embedded in correlated securities, leaving predictable residual variance on the table. This paper builds on our previous Scopus indexed work on hybrid AI crypto forecasting (Ngo & Phan, in press) and extends it in three directions: (i) replacing on-chain blockchain signals with stock and commodity cross-market features that align with industrial equity; (ii) replacing the deep learning architecture with a Stacking Ensemble (GBR + RF + ElasticNet → Ridge) that is better calibrated for medium-sized datasets with interpretability requirements; and (iii) enriching the risk management layer with non-distributed Match Prediction intervals and position sizing based on Kelly criteria, based on modern portfolio management theory.

The article is structured as follows. Part 2 examines the economic and financial theories that underlie the cross-market transmission hypothesis. Part 3 describes the PEST and Porter Five Forces environment in HPG's investment environment. Part 4 details the data, feature engineering process, and model architecture. Part 5 presents the empirical results and explains them through the lens of financial theory. Part 6 develops portfolio management impacts. Season 7 ends with limitations and future directions.

II. Theoretical basis: Information transmission and asset valuation

2.1 Efficient market hypothesis and border market anomalies

Fama's (1970) efficient market hypothesis holds that the price of a security at any given time fully reflects all available information, excluding systematic excess profits from historical data (weak form) or publicly disclosed fundamentals (strong selling form). Empirical evidence from developed markets broadly favors at least weak form efficacy. However, there is a growing body of literature documenting persistent anomalies in border and emerging economies where institutional weaknesses facilitate predictable mispricing (Gupta et al., 2014; Hoang & Mateus, 2021).

Vietnam's stock market exhibits a number of characteristics that reduce the applicability of strong EMH. The concentration of retail investors cited above creates herd cascades that cause prices to react poorly then overreact to public information (De Bondt & Thaler, 1985). Settlement friction price capping rules and circuit breakers (daily ±7% of the band on the HOSE) create intraday momentum and a reversal microstructure effect that persists through the days (Vo & Phan, 2019). The information asymmetry between institutional and individual investors is still pronounced, with thin analyst coverage outside the VN30 index. Lo and MacKinlay's (1988) variance ratio experiments, when applied to VN30 components, eliminated random walking behavior on short horizons – a finding replicated for HPG's profit chain in our dataset through ADF statistics ($p < 0.01$ at log-returns).

These anomalies do not imply the irrationality of the market; rather, they reflect the friction of transaction costs and dissemination of information characteristic of border markets. It is in the gap between theoretical efficiency and observational microstructure that machine learning synthesis methods find their economic basis: they can detect and exploit nonlinear, multi-source dependencies that neither linear econometric models nor single-variable engineering systems can capture.

2.2 Modern portfolio theory and systemic risk framework

Markowitz's (1952) modern portfolio theory established that asset prices are governed not only by distinct characteristics but by their covariant structure in the portfolio universe. The systemic risk of the asset – the non-diversifiable component – is what rational investors value. For a capital-intensive, cyclical-sensitive industry like HPG, this system component is crucial: the stock's beta against the VN-Index has historically exceeded 1.3, amplifying both the market's bullish and downward momentum.

Multi-factor extension of CAPM (Merton, 1973; Sharpe, 1964) demonstrates that the risks of a heavy industry are multidimensional, including commodity input cost shocks, tightening credit markets, and downstream demand cycles. Our feature engineering clearly operates this theoretical decay. The Commodity_spread variable (HRC steel price minus double the iron ore price) captures the blast furnace gross processing profit margin – which directly represents the supply chain factor load. Variable the profitability of the banking sector Bank_Ret representative credit cost risk. The VNI_Ret5 variable absorbs the aggregate market factor. The one-to-one mapping between financial theory and designed features is a distinct methodological contribution of this paper.

2.3 Three channels of cross-market information transmission

We identify three structural channels through which information outside the HPG is temporarily held into its equity price with systematic delay. These channels follow the classification of Hertzel et al. (2008), Hou and Moskowitz (2005), and Barberis et al. (2005):

Channel 1 – Vertical supply chain links. Blast furnace steel production requires about 1.6-2.0 tons of iron ore per ton of finished steel (World Steel Association, 2023). The 62% Fe iron ore futures, which are priced on the Singapore Exchange (SGX), move first, with a lag of one to three days before the cost shock is fully reflected in HPG's share valuation (Chen et al., 2019). The Commodity_spread feature – which captures the net processing gross profit margin on raw material costs – provides a real-time representation of profits resulting in earnings disclosure in weeks. Symmetrical downstream construction activities: as major contractors such as Coteccons (CTD) increased their momentum for five sessions, this signaled an increase in on-site steel demand before HPG's order book data was made public.

Channel 2 – Parallel and horizontal co-motion effects. In the domestic market where HPG, HSG, and NKG collectively hold ~70% of the market share, a common macroshock (e.g., anti-dumping duties on imported HRCs or policy stimulus in the construction sector) affects all three companies simultaneously. However, the company's specific responses to general shocks – inventory positioning, hedging, customer mix – vary, creating a lead-lag cross-cutting latency. Barberis et al. (2005) noted that retail investors tend to classify similar-cap stocks or classify the industry into spirit portfolios, make correlated buy/sell decisions during news events, and generate price co-volatility beyond fundamental co-volatility. Steel_Peer_Ret (equally weighted average daily return of HSG and NKG) operates this channel.

Channel 3 - Financial intermediary and liquidity transmission. SSI and VNDirect Securities brokerage firms are among the largest margin lenders in the country. Their share price dynamics (SSI_Mom5, VND_Mom5) reflect changes in the aggregate margin credit cycle, thereby boosting the capacity of retail investors holding leveraged long positions in capital stocks as large as HPG. When brokerage momentum is positive during the five-session window, it signals an expanded margin book and elevated risk appetite – conditions that precede cash flows into high-beta industries with a lag of one to two sessions (Lo & MacKinlay, 1988). Commercial banks VCB and BID play a similar role at the wholesale financing level, capturing the tightening or easing of corporate credit conditions that directly affect HPG's financing costs for iron ore import debt in USD.

The lead-lag relationship embedded in all three channels arises from two fundamental frictions: heterogeneous processing speeds between types of investors (institutional vs. retail) and limited attention that limits the speed at which correlated asset signals are subject to price arbitrage. Our experimental results confirm their incremental prediction content in addition to univariate technical signals.

III. HPG's Investment Environment: PEST and Porter Five Forces Analysis

3.1 Macro Environmental Factors (PEST)

Vietnam's political environment provides a structurally supportive context for construction steel demand through sustainable public infrastructure investment. The North-South expressway program, Long Thanh International Airport and urban housing policy incentives jointly maintain the basic steel absorption floor level. The machine learning model operates this channel through downstream contractor momentum (CTD_Mom5, Feature Importance Rating 10; VCG_Mom5), capture the on-site activity that has been carried out rather than the project processes that have been announced. Trade remedies – specifically anti-dumping duties on imported HRCs effective from 2024 (SSI, 2025) – directly protect HPG's domestic market share and pricing power, which is quantified in the model through variable Commodity_spread (Rank 15).

Economic factors represent the highest frequency and highest level of HPG's valuation. Spot iron ore prices and futures, which are denominated in USD and traded on the SGX, account for 55-60% of blast furnace production costs. Our IronOre_ratio variable (Rank 12) – calculated by dividing the spot price divided by its ten-session moving average – captures the deviation from the cost equilibrium, signaling impending margin compression or expansion before quarterly earnings are revealed. The USD/VND exchange rate opens up a second economic transmission channel: the devaluation of the dong increases both import costs and market losses on HPG's foreign currency debt. USD_Ret1 and banking sector variables Bank_Ret together represent this monetary policy and balance sheet risk.

Social and behavioral factors are particularly prominent in a market where individual investors dominate trading. The 2020-2021 wave of F0 investors – in which an estimated 1.5 million new retail accounts were opened – created a huge increase in HPG's liquidity, with daily volumes exceeding 75 million shares during peak sessions. This herd dynamics is empirically captured by the SSI_Mom5 variable (Rank 8, Critical Weight 0.0007), which records the highest cross-market signal weighting in the entire feature set, emphasizing the superiority of liquidity cycle proxies over macroeconomic variables in a retail-dominated frontier market.

Technological factors include the availability of real-time global commodity price feeds and the digitization of Hoa Phat's blast furnace operations, both of which accelerate the transmission of information into stock prices.

3.2 Industry Competitive Forces (Porter Five Forces)

The domestic steel industry exhibits concentrated monopoly characteristics. Three producers – HPG, HSG and NKG – control about 70% of construction steel and tubing production, creating a market structure where competitive responses (prices, volumes, discounts) are quick and publicly observable. Results The importance of the feature confirms that the cross-market returns of peers (Steel_Peer_Ret, Rank 16) carry statistically significant predictions, consistent with Hou and Moskowitz's (2005) price delay mechanism on correlated securities.

The bargaining power of the supplier focuses on the global mining monopolies (Vale, Rio Tinto, BHP) that set the benchmark price of iron ore. HPG has partially mitigated this risk through partial upstream integration and long-term consumption contracts, but has essentially kept the price of imported ore at bay. The IronOre_ratio and Commodity_spread variables capture the power of this upstream supplier in real-time. The bargaining power of customers mainly belongs to large contractors and real estate developers, which have seasonal purchase cycles and correlate with the disbursement of public infrastructure. CTD_Mom5's 10th ranking on the importance of the feature confirms the theory that downstream customer demand signals lead HPG's stock reaction by one to two sessions.

Table 1. PEST and Porter Five Forces – Mapping to model features (Selected)

Element Categories	Economic Transmission Channel	Feature Variable	Importance Rating
Politics: Infrastructure Policy	Downstream construction steel demand signals	CTD_Mom5	10
Politics: Defensive-trade tariffs	HRC/Iron Ore Total Processing Margin	Commodity_spread	15
Economy: Iron ore cost shock	Upstream input cost deviation from the average value	IronOre_ratio	12
Economy: USD/VND exchange rate	Import costs and foreign exchange debt by market	USD_Ret1	Top 25
Society: Shepherding retail investors	Proxy Broker Margin Lending Capability	SSI_Mom5	8
Society: Peer Movement	Horizontal competitive pricing	Steel_Peer_Ret	16
Economics: Revenue Fundamentals	Basic anchor, anti-overfitting control	Revenue_Growth	14
Economy: Market System Risk	Market-beta proxies/risk appetite	VNI_Ret5	20

IV. Data and methodology

4.1 Data Source and Coverage

The empirical dataset covers 1,390 raw trading sessions from April 20, 2020 to November 7, 2025, covering a complete macro cycle: the COVID-19 liquidity crisis (2020), the F0 retail investor boom (2021), the contraction of monetary tightening (2022), and the sideways recovery period (2023-2025). After applying the scrolling window feature build (maximum lookback: 60 sessions), the top 60 NaN observations were removed, yielding a clean sample of 1,330 observations across 87 designed features derived from five types of sources: (i) HPG OHLCV (HOSE) data; (ii) closing price and volume for eight stocks on the cross market (HSG, NKG, VCB, BID, SSI, VND, CTD, VCG); (iii) the daily price of iron ore 62% Fe and the price of HRC futures contracts (SGX, LME via MXV); (iv) The interbank exchange rate of USD/VND (SBV); and (v) quarterly accounting rate for HPG (daily term).

4.2 Feature Engineering and Commodity_spread Innovations

The three feature lines that operate the three transmission channels are defined in Section 2.3. For each cross equity, we calculate two short-lag profit variables and five-session cumulative volume volatility. To address multi-linearity in industry groups, we also construct sector-average aggregate variables (Steel_Peer_Ret as equal-weighted averages of daily returns HSG and NKG; Bank_Ret is the equally weighted average of VCB and BID). Commodity features include one-session and five-session log profit proxies for iron ore and HRC, plus two positioning rate variables (price divided by the 10-session moving average). Commodity_spread feature is defined as:

$$Commodity\_spreadt = HRC\_Pricet - 2 \times IronOre\_Pricet$$

This expression reflects the economic logic of blast furnace production: about two tons of iron ore are consumed per ton of finished steel product equivalent to HRC (World Steel Association, 2023). As the spread widens, the gross profit margin widens, usually before adjusting for earnings to increase. When it compresses, the opposite dynamic unfolds. This feature provides the model with a daily, real-time frequency representation of profit information that would otherwise only be available with quarterly reporting delays – a new contribution compared to previous cross-stock forecasting studies in Vietnam (Do et al., 2022; Nguyen et al., 2020).

4.3 Model Architecture: Stacking Ensemble with Walk-Forward Validation

The forecasting system uses a two-level Stacking Ensemble (Wolpert, 1992). At Level 0, three heterogeneous base learners are trained in parallel on a 930-session training window: (i) Gradient Boosting Regressor (GBR, n_estimators=400, learning_rate=0.05, max_depth=5) – captures nonlinear interaction effects between cross-market features; (ii) Random Forest Regression (RF, n_estimators = 300, max_features = 0.7) – provides variance reduction via bootstrap aggregation; (iii) ElasticNet regression ($\alpha = 0.01$, l1_ratio = 0.5) – extracts linear relationships while applying L1/L2 regularization to control multi-linearity between highly correlated cross variables. At Level 1, a meta-learner Ridge (regularized L2, α=1.0) is trained on a super-feature matrix generated by the base learner's predictions on a 133-session validation window, the optimal weighting for their additional error structures.

Walk-Forward authentication divides 1,330 observations over time (70%/10%/20%) without any random disturbances, enforcing the causal arrow of time and eliminating forward-looking bias. The StandardScaler parameters are estimated exclusively on the training set and applied as freeze transformations to validation and test sets – a safeguard not found in the majority of Vietnam's domestic forecast studies (Pham & Nguyen, 2022). The out-of-sample evaluation was conducted once on a 267-session test set (January 2024 – November 2025).

4.4 Quantifying Uncertainty Through Proper Prediction

Benchmark score forecasting evaluation indicators (MAE, RMSE, MAPE, R^2) do not quantify the uncertainty per forecast – an important gap for risk management applications. We solve this problem using Inductive Fit Prediction (separation) (Lei et al., 2018; Vovk et al., 2005), a non-distributive method that provides valid range guarantees without the Gaussian error assumption. Non-conformity point $\alpha i = |yi\_val - \hat{y}i\_ensemble|$ computed on a validation set of 133 sessions, and the 90th quantum q_{90} extracted. Each prediction of the test set is then placed in parentheses as $[\hat{y}- q_{90}, \hat{y} + q_{90}]$. The experimental Q_{90} was 499 VND, which yielded a calibrated prediction interval of ±499 VND with an observed test set coverage rate of 76.4%. The gap between 90% nominal and 76.4% empirical coverage reflects the interchangeability of the financial time series over regime changes – a known limitation of static division fit, which we discuss further in Section 7.

V. Experimental results

5.1 The Importance of the Cross-Market Feature: Evidence for Three Channels

Table 2 presents a ranking of the importance of the Gradient Boosting feature for the top 20 variables. The results provide strong quantitative support for the three-channel transmission hypothesis given in Section 2.3. The dominance of MA_5 (Importance = 0.9712) reflects the most clearly documented correlation in Vietnam's border equity returns (Lo & MacKinlay, 1988; Vo & Phan, 2019) – a stylized reality of the border market microstructure, where the dissemination of information takes place gradually and limited attention limits immediate price corrections.

Among the cross market variables, the financial intermediary channel (Channel 3) produces the highest ranking signal: SSI_Mom5 ranks 8th with an important weighting of 0.0007 – the most informative external feature in the system. This result is economically visual: SSI's five-session accumulation momentum is a high-frequency, continuously updated indicator of retail-credit margin expansion, leading capital inflows into high-beta large-cap stocks like HPG for one to two sessions. VND_Mom5 (Ranked 19) provides authentic secondary broker signals.

The supply chain channel (Channel 1) produces the second strongest crossover market block: CTD_Mom5 (Rank 10, 0.0005) and IronOre_ratio (Rank 12, 0.0004) rank above all macroeconomic variables. CTD's weekly dynamics record on-site construction activity that was carried out with specific demand signals prior to HPG's order confirmation; IronOre_ratio captures deviations from the cost balance, which, when increased, signals potential profit compression before the quarterly cost is announced. Proprietary Commodity_spread variable (Rank 15) – HRC minus twice the price of iron ore – confirms the interpretation of the gross processing margin: wider spreads predict higher profits, and the model assigns this feature a Steel_Peer_Ret in line with the important weighting (Rank 16), turn the channel into horizontal competition.

Table 2. Feature Importance – Top 20 Variables (GBR Level 0 Learners)

Grade	Feature Variable	Categories/Channels	Importance
1	MA_5	Technique – short-term momentum anchor	0.9712
2	MA_10	Technical – 10-session trend	0.0117
3	MA_ratio_60	Technical position ratio – quarterly	0.0015
4	MA_ratio_5	Technical – overbought/oversold proxies	0.0013
5	BB_position	Technique - Bollinger Band Placement	0.0010
6	HPG_Return_3d	Internal - 3 Days Late Returns	0.0009
7	MA_20	Technical - monthly trends	0.0009
8	SSI_Mom5	Channel 3 - broker/margin liquidity	0.0007

9	HPG_Return_2d	Internal – 2 Days Late Declaration	0.0005
10	CTD_Mom5	Channel 1 – downstream demand signal	0.0005
11	MA_ratio_20	Technical position ratio – monthly	0.0004
12	IronOre_ratio	Channel 1 – upstream cost bias	0.0004
13	Vol_20d	Technique – alternating volatility	0.0003
14	Revenue_Growth	Basic – income anchor	0.0003
15	Commodity_spread	Channel 1 – blast furnace edge	0.0003
16	Steel_Peer_Ret	Channel 2 - peer-to-peer spread	0.0003
17	MACD_signal	Technique – momentum smoothing	0.0003
18	MA_ratio_10	Technical – bi-weekly placement	0.0003
19	VND_Mom5	Channel 3 – secondary broker signals	0.0003
20	VNI_Ret5	Macro-market system risk	0.0003

5.2 Out-of-sample forecast performance

Table 3 presents the comparative performance of the five model architectures across a 267-session test suite. The Stacking Ensemble achieves MAPE = 1.57%, R^2 = 0.957, and Directional Accuracy (DA) = 51.50%. ElasticNet independently recorded the lowest absolute error indices (MAE = 344.30 VND, MAPE = 1.43%) and the highest DA (52.26%) – a model that is consistent with the linear characteristics of the 2024-2025 horizontal consolidation period, in which regular linear models benefit from a reduction in nonlinear noise compared to tree-based learners (see Section 5.3 for explanation). The Naive Baseline achieves superior MAE (298.72 VND) and R^2 (0.973) by exploiting stochastic walking inertia, but drops to DA=5.64%, confirming that error magnitude and directional accuracy are economically separated: a model that always predicts tomorrow's price equal to today's price will achieve a low MAE in a slow-moving market while supply zero guidance grant.

Table 3. Out-of-sample performance on a 267-session test suite

Model	MAE (VND)	RMSE (VND)	MAPE (%)	R^2	DA (%)
Slope Strength (GBR)	458.75	636.50	1.94	0.945	45.11
Random Forest (RF)	461.71	650.21	1.95	0.942	47.74
Elastic Network	344.30	502.77	1.43	0.966	52.26
Stacking Ensemble	386.81	563.41	1.57	0.957	51.50
Innocent baseline	298.72	446.65	1.24	0.973	5.64

5.3 Structural Strength of Ensemble vs. ElasticNet

ElasticNet's apparent superior performance on test indicators requires careful economic interpretation rather than a simple conclusion that linear models prevail. The testing period (2024-2025) coincides with the time when HPG prices accumulate in a narrow range around VND 23,000 - 28,000, characterized by low volatility, reduced speculative capital flows and relatively stable commodity spreads. In this 'static linear' mode, a simple regularized linear model faces minimal nonlinear noise and naturally records lower absolute errors.

The Stacking Ensemble's 42.5 VND incremental MAE disadvantage compared to ElasticNet represents a deliberate structural trade-off: the set maintains nonlinear base learning (GBR, RF) calibrated to handle mode switches – margin compression due to commodity shocks, sudden changes in policy rates, F0 investment wave - at the cost of some accuracy in periods of atypical low volatility. From the perspective of a multi-period investment horizon, a pattern that performs worse than 42.5 VND/session in a sideways market but avoids oriented failure during periods of volatility yields superior risk-adjusted returns, in line with the slightly higher DA of the set (51.50% vs. 52.26% of ElasticNet) shows only lagging behind after being modest in this indicator, while achieving greater resilience to structural faults. The operational determinant rule we propose in Section 6 clearly explains this cyclical trade-off through mode conditional model selection.

VI. Meaning of portfolio management for border market investors

6.1 Suitable Prediction Interval as Dynamic Stop-Loss Anchor

The matching prediction range of ±499 VND - corresponding to about 1.9% of the average price in HPG's test period - translates directly into a principled stop-loss frame. Instead of imposing stop-outs based on arbitrary percentages (a common retail practice), institutional portfolio managers can draw stop-loss thresholds from the empirical error distribution of a validated pattern. We recommend a cautious rate factor of $1.5 \times q_{90} \approx$ 749 VND, which provides a daily worst-case loss buffer that includes tail errors observed in the validation set.

For a long position of N shares opened at P_entry price, S_stop trigger stop loss and risk capital (CaR) per session is:

$$S\_stop = P\_entry - 1.5 \times q_{90} = P\_entry - 749\ VND$$

$$CaR\ per\ session = N \times 749\ VND$$

This experimental calibration offers two advantages over the heuristic method. First, it is based on the model's actual generalized error distribution, not a Gaussian form or hypothetical parameter – making it robust with the fat-tail return distribution documented in Table 2 (daily return kurtosis of HPG = 5.23). Second, the period can be dynamically updated as the model is periodically retrained on an extended data window, allowing the stop loss to adapt to changing volatility modes instead of maintaining a fixed percentage.

An empirical coverage ratio of 76.4% (compared to 90% nominal) reflects the non-interchangeability of HPG's profit chain across market modes – a known limitation of Static Split Conformity Prediction when applied to non-fixed financial data (Gibbs & Candès, 2021). Moving on to Adaptive Rolling Fit Prediction (rolling q_{90} in W = the last 50 sessions) is a future natural upgrade that will bring the experimental range closer to the nominal goal.

6.2 Fractional Kelly Criteria for Position Size

The Kelly criterion provides the theoretically optimal f* capital portion for allocation to binary bets with a win probability p and a win/loss ratio of b (Kelly, 1956, as operated in Thorp, 2008):

$$f^* = p - (1 - p) / b$$

Applying DA = 51.5% of the aggregation as p and considering the average daily HPG win/loss ratio (estimated from the tester) as b ≈ 1.0 (daily return approximately symmetrical approximation of about 0 for an average reverted sideways market), the full Kelly fraction would be f* = 0.015 – i.e. 1.5% of capital per trade. This is already a fractional Kelly allocation, consistent with the academic consensus that half-Kelly or quarter-Kelly strategies are preferred in practice to reduce the volatility of outcomes (MacLean et al., 2010).

We recommend a halved Kelly adjustment due to the limitations that limit the reliability of the model: allocate 0.75% of the portfolio capital to each HPG-oriented signal. This fraction is adjusted on the signal filter: trades are taken only when the Match Prediction interval of the set does not coincide with the previous day's close – i.e. when $|\hat{y} - y_{t-1}| > q_{90}$. This confidence-based filtering mechanism reduces the expected frequency of transactions while improving the average quality of the signals executed, a technique that has been validated in our previous cryptocurrency research (Ngo & Phan, print press) and supported in the quantitative trading literature (Thorp, 2008).

6.3 Scenario-based dynamic tilt via Monte Carlo

In addition to daily directional signals, the system generates long-term scenario forecasts through iterative Synthetic Predictions combined with Monte Carlo Brown Motion Geometry (GBM) simulations (1,000 paths, 252 trading days). The three macro scenarios – BULL (P90 line, 12-month moving average: 31,366 VND, +20.2%), BASE (P50 line, 25,976 VND, -0.5%), BEAR (P10 line, 21,518 VND, -17.6%) – provide a probability-risk cone that institutional investors can use to correct the strategic asset allocation bias.

A real-world implementation framework maps scenario probability to portfolio weighting adjustments. According to the BASE (highest probability) scenario, characterized by stable domestic steel demand and restrained global commodity prices), the recommended HPG allocation reflects the level of exposure to the benchmark index. In BULL conditions (triggered by the positive Commodity_spread expansion sustained for 20+ sessions plus strong CTD/VCG downstream momentum), the Kelly segment weight is expanded to 1.5× the base allocation. In a BEAR condition (triggered by a persistent negative IronOre_ratio deviation plus credit tightening signals in the Bank_Ret), the position is reduced to a base of 0.5× or completely replaced by short-term protection through a hedged call option – a multi-instrument extension that is beyond the scope of this article but follows in a natural from the risk cone framework.

6.4 Technological inequality and implications for market access

The system described in this paper requires access to real-time commodity price feeds (SGX iron ore futures, LME HRC), eight cross-stock data streams, Python-based synthetic inference workflows, and the ability to retrain the model periodically. These requirements are within the technical capabilities of institutional asset managers, fund operators and quantitative hedge funds – but are still inaccessible to the retail F0 investor segment, which accounts for more than 85% of Vietnam's daily trading volume. This technological disparity has regulatory implications and market microstructures that deserve explicit acknowledgement.

If aggregate-based systems with multi-market information integration achieve systematically higher directional accuracy than publicly available technical analysis tools, they create information asymmetry that benefits institutional participants and reinforces the concentration of alpha generation in a minority complex technology. This dynamic is not unique to Vietnam – it mirrors debates in developed markets about high-frequency trading and algorithmic advantage – but is likely to be more consequential in a marginal market where retail investors' financial literacy and sophistication of risk management remain limited. We mark this as an area that needs to be monitored by regulators such as the State Securities Commission (SSC) as the algorithmic trading infrastructure develops in Vietnam.

VII. Conclusions and limitations

This article has hypothesized the transmission of information across the market for Vietnam's border steel industry through three contributions. First, we laid the technical foundation of the feature of an integrated forecasting system in established financial theory – EMH, MPT, CAPM, multi-factor and supply chain economics – that maps each theoretical channel to a set of specifically designed variables. Second, we have empirically confirmed that the three defined transmission channels (supply chain, horizontal competition, and financial intermediary) produce statistically significant incremental predictive content in addition to univariate technical signals, with brokerage momentum (SSI_Mom5), The need for downstream construction (CTD_Mom5) and iron ore positioning (IronOre_ratio) is ranked among the top 12 characteristics in the GBR importance hierarchy. Third, we translated the model's outputs into specific portfolio management protocols – stop-losses derived from Matching Predictions, fractional Kelly position sizing, and dynamic tilts based on the Monte Carlo scenario – based on empirical error distributions rather than arbitrary guesses.

The study has some limitations. The matching prediction intervals provide empirical coverage of 76.4% against the 90% nominal target, reflecting impermanence in the 2020-2025 cycle. Transitioning to adaptive roller conformalization (Gibbs & Candès, 2021) will address this gap. The feature engineering process is tailor-made for HPG's economic environment: cross-variables (CTD, SSI, IronOre) are economically driven by HPG's supply chain structure and cannot be transferred to other industries without restructuring based on industry-specific analysis. The 267-session test phase coincided with an unusually low volatility sideways mode, partly in favor of the more linear ElasticNet than Ensemble; Vertical evaluation across multiple market cycles will reinforce claims of generalizability. Finally, the absence of intraday microstructure data and investor sentiment signals (from social media, news) represents a known information boundary that future research can address through the integration of NLP-based sentiment indicators and high-frequency feature building.

Despite these limitations, the results demonstrate that systematic exploitation of multi-market information transmission channels in marginal stock markets can produce modest but statistically significant directional advantages (DA = 51.5% vs. 5.64% mediocre for the Naive Baseline), and these advantages can be operated into disciplined, quantitative portfolio management protocols that are consistent with the risk tolerance of institutionalized investors in Vietnam institutionalizing the stock market.

References

[1] Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. Journal of Financial Economics, 75(2), 283–317. https://doi.org/10.1016/j.jfineco.2004.04.003

[2] Chen, T., Huang, Y., & Wang, S. (2019). Commodity price shocks and stock market reactions: Evidence from the Chinese steel sector. Resources Policy, 61, 412–423. https://doi.org/10.1016/j.resourpol.2019.02.015

[3] De Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact? The Journal of Finance, 40(3), 793–805. https://doi.org/10.1111/j.1540-6261.1985.tb05004.x

[4] Diebold, F. X., & Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. International Journal of Forecasting, 28(1), 57–66. https://doi.org/10.1016/j.ijforecast.2011.02.006

[5] Dietterich, T. G. (2000). Ensemble methods in machine learning. Multiple Classifier Systems, 1–15. https://doi.org/10.1007/3-540-45014-9_1

[6] Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Machine Learning in Finance: From Theory to Practice. Springer. https://doi.org/10.1007/978-1-4939-9933-0

[7] Do, H. M., Nguyen, T. T., & Le, V. H. (2022). Stock price forecasting using machine learning: Evidence from an emerging market. Journal of Asset Management, 23(4), 312–325.

[8] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

[9] Forbes, K. J., & Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements. The Journal of Finance, 57(5), 2223–2261. https://doi.org/10.1111/0022-1082.00494

[10] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[11] Gibbs, I., & Candès, E. J. (2021). Adaptive conformal prediction for time series. Advances in Neural Information Processing Systems, 34, 16607–16619.

[12] Gupta, R., Hammoudeh, S., Modise, M. P., & Nguyen, D. K. (2014). Can economic uncertainty, financial stress and consumer sentiments predict US equity premium? Journal of International Financial Markets, Institutions and Money, 33, 367–378.

[13] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

[14] Hertzel, M. G., Li, Z., Officer, M. S., & Rodgers, K. J. (2008). Inter-firm linkages and the wealth effects of financial distress along the supply chain. Journal of Financial Economics, 87(2), 374–387. https://doi.org/10.1016/j.jfineco.2007.01.005

[15] Hoang, T. H. V., & Mateus, C. (2021). Return, volatility and trading volume spillovers between Chinese and frontier equity markets. Journal of International Financial Markets, Institutions and Money, 74, 101392.

[16] Hoa Phat Group. (2025). Annual Report 2024. Corporate Disclosure, HOSE.

[17] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

[18] HOSE. (2025). Market Capitalisation and Liquidity Statistics. https://www.hsx.vn

[19] Hou, K., & Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. Review of Financial Studies, 18(3), 981–1020. https://doi.org/10.1093/rfs/hhi023

[20] Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Research, 259(2), 689–702. https://doi.org/10.1016/j.ejor.2016.10.031

[21] Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018). Distribution-free predictive inference for regression. Journal of the American Statistical Association, 113(523), 1094–1111. https://doi.org/10.1080/01621459.2017.1307116
[22] Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. Review of Financial Studies, 1(1), 41–66. https://doi.org/10.1093/rfs/1.1.41
[23] MacLean, L. C., Thorp, E. O., & Ziemba, W. T. (2010). Good and bad properties of the Kelly criterion. Wilmott Magazine, 45, 87–91.
[24] Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x
[25] Merton, R. C. (1973). Theory of rational option pricing. The Bell Journal of Economics, 4(1), 141–183. https://doi.org/10.2307/3003143
[26] Mohanty, S., Nandha, M., & Bansal, A. (2018). The impact of iron ore and steel price dynamics on industrial equity returns. Resources Policy, 59, 412–423. https://doi.org/10.1016/j.resourpol.2018.08.015
[27] MXV. (2025). Commodity Market Report 2020–2025: Iron Ore and HRC Price Dynamics. https://mxv.com.vn
[28] Ngo, H. Q., & Phan, D. M. (in press). Hybrid AI cryptocurrency price forecasting system with intelligent trading signals: Application to Chainlink. Mathematical Modelling of Engineering Problems.
[29] Nguyen, T. T., Tran, Q. H., & Pham, N. M. (2020). Single-stock versus multi-source data integration in financial forecasting. Expert Systems with Applications, 145, Article 113112.
[30] Pham, T. T. H., & Nguyen, V. A. (2022). Ứng dụng thuật toán Random Forest trong dự báo xu hướng giá cổ phiếu ngành Thép: Trường hợp thực nghiệm mã HPG. Tạp chí Chuyển đổi số Tài chính, 5(2), 23–34.
[31] Phan, D. M. (2026). Applying Ensemble Learning and Cross-Stock Information for Stock Price Forecasting: A Walk-Forward Validation Approach - Case Study: Hoa Phat Group (HPG) [Bachelor's Thesis, VKU].
[32] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 19(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x
[33] SSI. (2025). Market Strategy and Sector Analysis Report 2020–2025. https://www.ssi.com.vn
[34] UBCKNN. (2025). Annual Report: Vietnam Capital Market Development. Ministry of Finance. https://ssc.gov.vn
[35] VCBS. (2024). Strategic Investment and Company Analysis Report. Vietcombank Securities.
[36] Vo, X., & Phan, B. S. (2019). Market anomalies and technical trading strategies in emerging equity markets: The case of Vietnam. Quarterly Review of Economics and Finance, 74, 215–224. https://doi.org/10.1016/j.qref.2019.05.004
[37] Vovk, V., Gammerman, A., & Shafer, G. (2005). Algorithmic Learning in a Random World. Springer. https://doi.org/10.1007/b106715
[38] Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1
[39] World Steel Association. (2023). World Steel in Figures 2023. Worldsteel Committee on Economic Studies.
[40] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B, 67(2), 301–320. https://doi.org/10.1111/j.1467-986X.2005.00503.x