International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(18s) 2026 pp. 1406-1430
https://doi.org/10.70917/ijcisim-2026-4990
Article

VOLATILITY FORECASTING FROM ECONOMETRICS TO ARTIFICIAL INTELLIGENCE: A FOUR-STAGE REVIEW OF THE EVIDENCE PIPELINE FROM FORECAST ACCURACY TO PORTFOLIO PERFORMANCE

Nikita I. Lysenok

Faculty of Economic Sciences, HSE University, Moscow, Russian Federation
https://orcid.org/0000-0003-4660-1410
nilysenok@hse.ru

Abstract: A volatility forecast becomes useful only after it has passed through four stages: a model produces it, a trading rule consumes it, a portfolio aggregates the result, and an investor evaluates that result against a utility function. Each stage has its own mature evaluation apparatus, and each is studied in isolation. This review traces the forecast through the whole pipeline and asks where the evidence connecting the stages actually exists. Drawing on 153 studies published between 1963 and 2026, covering both the international and the Russian-language literature, it makes four contributions. First, it introduces the four-stage pipeline itself as an organising framework for a literature that has never been surveyed across, rather than within, its stages. Second, it develops a classification of the channels through which a forecast reaches a trading decision — trade parameters, entry gating, position sizing and allocation. The first three channels and their comparative evaluation were introduced by the author in Lysenok (2026a); the framework is generalised here to four channels and applied as an organising instrument for the literature, which is shown to be structured one channel at a time, with no comparison under common conditions. Third, it maps the state of the evidence stage by stage and reports a quantitative asymmetry: 81 of the 153 studies evaluate forecasts by a statistical loss alone, 35 evaluate trading or portfolio outcomes without reference to forecast quality, and only 8 apply both criteria. The only direct comparison of channels under common conditions is the author's own, and its results — gains concentrated entirely in the channels that govern the commitment of capital, with the trade-parameter channel delivering no measurable economic value — indicate that the transitions this review maps are not formalities. Fourth, it systematises for an international readership a body of Russian-language work on volatility forecasting that has remained inaccessible behind a language barrier. The review concludes that statistical accuracy is a defensible proxy for economic value only conditionally on the channel through which the forecast reaches the allocation of capital, and sets out five research tasks that follow from the mapping.
Keywords: realized volatility; artificial intelligence; machine learning; deep learning; forecast evaluation; trading strategies; position sizing; economic value; systematic review; emerging markets

1. Introduction

Volatility is the most predictable second-order property of financial returns, and the effort devoted to forecasting it more accurately is correspondingly large. The methodological literature has produced a sequence of increasingly capable model classes, from the parametric conditional-variance families through the realized-volatility framework to gradient boosting and deep neural architectures, and it has developed a rigorous apparatus for deciding

which of them forecasts better. That apparatus is self-contained. It compares models by a loss function, tests the significance of differences in that loss, and stops.

The stopping point matters, because a forecast is not an end in itself. Before it can be worth anything, it must pass through a sequence of stages. A model produces the forecast. A trading rule consumes it and turns it into a decision about whether to hold a position and how large that position should be. Individual positions are aggregated into a portfolio. The portfolio is finally evaluated against the preferences of the investor who bears its risk. Each of these four stages has developed its own literature, its own benchmark practices and its own metrics, and each of those literatures is largely self-referential. Studies of forecasting accuracy rarely follow the forecast into a trading decision. Studies of trading strategies rarely ask how good the forecast inside them is. Studies of volatility-managed portfolios use a single mechanism of exposure scaling and apply it to passive factor portfolios. Studies of economic value apply the utility framework to covariance timing rather than to active systems.

The consequence is a set of transitions between stages on which the evidence is thin. Whether an improvement in statistical accuracy survives the passage into a trading decision is an empirical question that has been asked far less often than the question of which model is more accurate. Whether the mechanism of volatility management that works on passive factor portfolios also works inside a book of heterogeneous active strategies is another. Whether the resulting system pays for itself net of the cost of building it is a third. This review is organised around those transitions rather than around the model classes, which is what distinguishes it from existing surveys of the volatility forecasting literature.

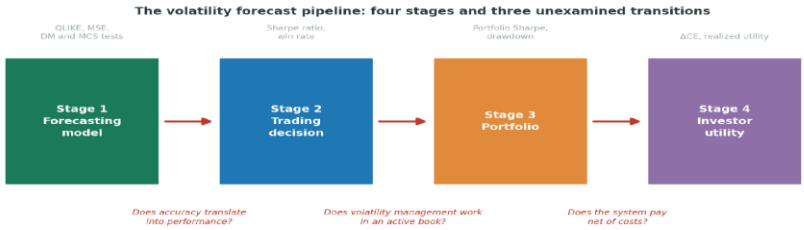

Figure 1. The volatility forecast pipeline: four stages, the metric conventionally used at each, and the three transitions on which evidence is scarce.

Four contributions follow. The first is the pipeline framework itself. Surveys of this literature exist within stages — most recently the comprehensive model-level review of Leushuis and Petkov (2026), which evaluates thirty-two forecasting specifications by goodness-of-fit criteria — but no survey has traversed the stages, and the four-stage representation of Figure 1, with its identification of the transitions as the objects requiring evidence, is introduced here. The second is a classification of the channels through which a volatility forecast can enter a trading decision. A forecast may scale the stop-loss and take-profit distances of a trade whose entry has already been determined; it may gate entry into a position; it may determine the size of that position; or it may set weights across strategies within a book. These are four distinct mechanisms with different sensitivities to forecast quality. The distinction between the first three, together with the first comparative evaluation of them on a single data set, was introduced by the author in Lysenok (2026a); the present review generalises the scheme to four channels by adding the allocation channel, adopts it as the organising instrument for Section 4, and uses it to show that the existing evidence is structured one channel at a time, with no study comparing them under common conditions. The same body of work provides what is, to the author's knowledge, the first introduction of a volatility forecast into a multi-strategy book of active strategies evaluated by both statistical and economic criteria, which is the intersection that Section 5 identifies as empty in the prior literature. The channel labels used here follow that earlier work, so that results reported under the original scheme remain directly comparable.

The third is a quantitative mapping of the evidence. Of the 153 studies surveyed here, 81 evaluate forecasts by a statistical loss alone, 35 evaluate trading or portfolio outcomes without any assessment of the quality of the forecast

that drives them, and only 8 apply both criteria within a single design. That asymmetry is the central finding of the systematisation, and it is what justifies organising a review around transitions rather than around methods.

The fourth is the systematisation of a Russian-language body of work on volatility forecasting. That literature has developed its own line of enquiry — on the role of oil prices and sanctions, on microstructure constraints affecting the choice of sampling frequency, on the applicability of HAR-type models to a concentrated emerging market — and it is effectively closed to an international readership by the language barrier. Section 7 organises it by the same four stages used elsewhere in the review.

The remainder of the paper is organised as follows. Section 2 describes the search protocol and the criteria for inclusion. Sections 3 to 6 traverse the four stages of the pipeline in turn. Section 7 covers emerging markets, with the Russian market as the principal case. Section 8 synthesises the evidence and identifies where the chain breaks. Section 9 concludes with five research tasks.

2. Review methodology

The search covered Scopus, Web of Science, RePEc and, for the Russian-language corpus, eLibrary. Queries combined terms for the object of forecasting (realized volatility, conditional variance, volatility forecasting) with terms for each stage of the pipeline (trading strategy, position sizing, volatility targeting, volatility timing, economic value, certainty equivalent, portfolio construction). The period covered is 1963 to 2026, beginning with the first documentation of non-normal return distributions and extending to the most recent work available at the time of writing.

Inclusion required that a study contribute evidence to at least one stage of the pipeline and report an evaluation against a benchmark, whether statistical or economic. Methodological contributions without an empirical component were included where they define the evaluation apparatus itself — loss functions, tests of predictive accuracy, bootstrap procedures for inference on performance ratios — since the argument of this review concerns precisely how evaluation is conducted. Purely descriptive market commentary, textbook expositions without original results and studies of implied volatility surfaces unconnected to forecasting were excluded.

The resulting corpus comprises 153 studies. Each was classified along five dimensions: the stage or stages of the pipeline it addresses, the channel of integration where applicable, the type of evaluation applied (statistical loss, economic outcome, or both), the market examined, and the sampling frequency of the data. Studies spanning several stages are counted in each, which is why stage-level counts sum to more than the number of studies. The classification is what makes the quantitative statements in Sections 8 and 9 possible; it is also, inevitably, a judgement, and borderline cases were resolved in favour of the narrower category.

Figure 2 shows the distribution of the corpus across stages and periods of publication. The concentration is pronounced and is itself the first result of the review.

Two limitations of the protocol should be stated. The corpus is weighted towards equity markets, since that is where the four stages have all been studied; the fixed-income and commodity literatures are represented only where they inform the general argument. And the classification of evaluation type is binary within each study, whereas some studies report an economic metric as a robustness check without treating it as a criterion. Such cases were classified by the criterion on which the study’s conclusions rest, which is a judgement that a different reviewer might make differently in perhaps a dozen instances. Neither limitation affects the direction of the asymmetry reported in Section 8, whose magnitude is too large to be sensitive to marginal reclassification.

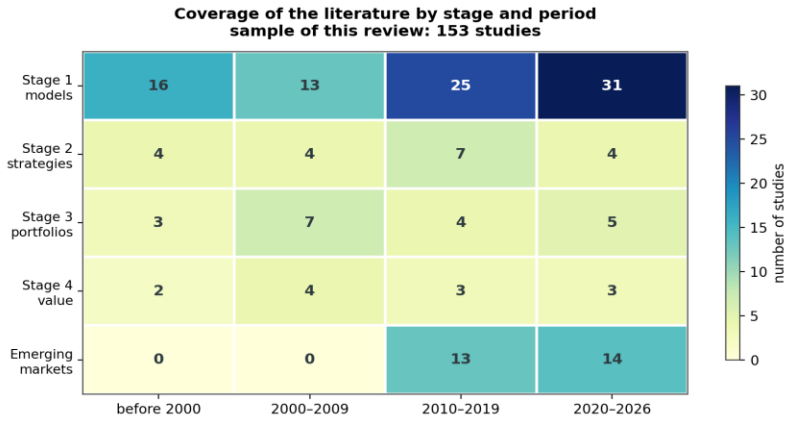

Figure 2. Coverage of the reviewed literature by pipeline stage and period of publication. A study spanning several stages is counted in each.

3. Stage one: forecasting models

3.1. What the object of forecasting requires of a model

The empirical properties that a volatility model must accommodate were established well before the models themselves. Returns are not normally distributed and their dispersion is not constant (Mandelbrot, 1963); volatility clusters, reverts to a long-run level, responds asymmetrically to the sign of returns, exhibits jumps and co-moves across assets (Cont, 2001). These properties constitute the specification problem, and the sequence of model families reviewed below can be read as successive attempts to represent more of them at acceptable cost.

3.2. The parametric tradition

The autoregressive conditional heteroskedasticity framework of Engle (1982), generalised by Bollerslev (1986), models conditional variance as a function of past squared innovations and past variances, thereby capturing clustering and mean reversion within a compact parameterisation. Asymmetry was subsequently accommodated by the exponential specification of Nelson (1991) and by the threshold term of Glosten, Jagannathan and Runkle (1993); Engle and Ng (1993) formalised the diagnostic apparatus for detecting it. The family expanded rapidly, and Rossi (2010) provides a systematic account of its univariate branch.

The comparative evidence has not been kind to elaboration. Hansen and Lunde (2005) compared 330 specifications and found no consistent improvement over the simplest GARCH(1,1) in forecasting exchange rate and equity volatility. That result has proved durable, and it establishes an important discipline for the rest of the review: within a model family, added structure does not reliably buy accuracy, and claims to the contrary require testing against the simplest available member of the family rather than against a strawman.

A separate branch addresses persistence through fractional integration, reproducing the slow decay of the autocorrelation function at the cost of a parameter that is sensitive to sample length and sampling frequency; its forecasting advantage does not survive consistently out of sample. The lesson carries through the rest of the stage: long memory must be represented somehow, and the specifications that endure are those that represent it cheaply rather than exactly.

3.3. Realized volatility and the shift to intraday data

The decisive change came not from a better parameterisation but from a better measurement. Andersen and Bollerslev (1998) showed that the apparent poor performance of standard models was substantially an artefact of using squared daily returns as the evaluation target, and that a measure constructed from intraday returns changes the assessment. The formal development followed in Andersen, Bollerslev, Diebold and Labys (2001; 2003), which

established realized volatility as a consistent estimator of integrated variance and moved the object of forecasting from a latent to an observable quantity. The practical consequence is that volatility forecasting becomes a time-series problem on an observed series rather than a filtering problem on an unobserved one, which is what opened the field to the model classes discussed in Sections 3.6 and 3.7.

Sampling frequency is the parameter that governs the trade-off between the informativeness of the estimator and contamination by microstructure noise. The heterogeneous market hypothesis of Müller et al. (1997) provides the economic rationale for treating different sampling horizons as reflecting different classes of market participant, and it supplies the intuition on which the model of the following subsection rests.

The choice is not merely technical. Sampling too finely admits bid-ask bounce and discreteness effects that bias the estimator upward; sampling too coarsely discards information and raises its variance. The conventional five-minute interval on developed equity markets represents a compromise calibrated to the spread structure and turnover of those markets, and it is not automatically appropriate elsewhere. Section 7 returns to this point, since the transferability of the standard is one of the substantive questions that emerging-market applications raise. What matters for the argument here is that the sampling decision made at Stage one propagates through the entire pipeline: a forecast built on a noisier target will be evaluated as less accurate at Stage one and may or may not be less useful at Stage two, and no study in the corpus separates these two effects.

3.4. The heterogeneous autoregressive family

The model of Corsi (2009) represents long memory through an additive cascade of daily, weekly and monthly realized volatility components. Its efficiency is remarkable: three terms reproduce persistence that an autoregressive specification would require some twenty lags to approximate, and the resulting model remains linear and interpretable. It has become the standard benchmark, and the burden of proof in this literature now rests on any model that claims to beat it.

Its extensions follow the empirical properties enumerated in Section 3.1. Andersen, Bollerslev and Diebold (2007) add a jump component, separating the continuous part of quadratic variation using the bipower variation of Barndorff-Nielsen and Shephard (2004). Patton and Sheppard (2015) decompose by the sign of returns and find that negative jumps raise subsequent volatility substantially more than positive ones of comparable magnitude. Bollerslev, Patton and Quaedvlieg (2016) exploit the measurement error in the realized estimator itself, allowing the coefficients to depend on the precision of the estimate. Clements and Preve (2021) collate the estimation choices that materially affect the performance of the family in applied work. The wider family record confirms the pattern across settings: time-varying coefficients improve performance when regimes shift (Wang, Ma, Wei and Wu, 2016), frequency decompositions of the target add accuracy (Gong and Lin, 2019), the multivariate generalisation is robust to non-Gaussian features (Čech and Barunik, 2017), the specification is more stable than its fractionally integrated competitor across horizons and sectors (Izzeldin, Hassan, Pappas and Tsionas, 2019), it transfers to commodity markets (Degiannakis, Filis, Klein and Walther, 2022), and it scales to panel estimation across asset classes with documented economic gains (Bollerslev, Hood, Huss and Pedersen, 2018).

Further variants — logarithmic targets, multivariate cascades admitting cross-sectional information, time-varying coefficients — each improve fit at the cost of parameters, and the comparative evidence follows the pattern of Section 3.2: real but modest gains that are not uniform across markets or horizons.

3.5. Hybrids within the econometric tradition

Several specifications combine the realized measure with the conditional-variance framework rather than replacing one with the other. The realized GARCH of Hansen, Huang and Shek (2012) models returns and realized measures jointly; the HEAVY specification of Shephard and Sheppard (2010) is designed for rapid adaptation to shifts in the level of volatility; the mixed-frequency approach of Ghysels, Santa-Clara and Valkanov (2006) admits predictors sampled at different frequencies within a single regression. A parallel line uses the option-implied volatility as a predictor, following the finding of Christensen and Prabhala (1998) that it contains information beyond the

historical series; the systematic comparisons of Kambouroudis, McMillan and Tsakou (2016; 2021) establish that its inclusion as a regressor improves HAR-type forecasts consistently across markets, whereas no single augmented specification dominates in all of them.

3.6. Machine learning and deep architectures

Machine learning entered volatility forecasting through the same door as the rest of empirical asset pricing (Gu, Kelly and Xiu, 2020). Three classes have received sustained attention. Regularised regressions extend the linear framework by admitting many lags and shrinking uninformative coefficients; Audrino and Knaus (2016) show that a LASSO specification is comparable to HAR but does not reproduce its lag structure, and Ding, Kambouroudis and McMillan (2021) find it superior out of sample across eight international indices. Tree ensembles partition the predictor space without a prior functional form: Luong and Dokuchaev (2018) document gains from random forests, Christensen, Siggaard and Veliyev (2023) find that tree-based methods improve on the HAR family when a rich feature set is available, and Teller, Pigorsch and Pigorsch (2022) report that gradient boosting outperforms both HAR and recurrent networks at the one-day horizon.

Deep architectures form the third class. Recurrent networks with gating mechanisms (Hochreiter and Schmidhuber, 1997; Cho et al., 2014) were the first to be applied, with Bucci (2020) reporting gains over econometric alternatives on long samples. Convolutional designs followed, including the HARNet of Reisenhofer, Bayer and Hautsch (2022), which initialises a dilated convolutional network at the coefficients of the HAR model and thereby guarantees performance at least equal to its benchmark from the outset. Attention-based architectures constitute the current frontier: Ramos-Pérez, Alonso-González and Núñez-Velázquez (2021) and Frank (2023) report improvements from transformer variants, and graph-based designs extend the approach to the cross-section (Brini and Toscano, 2025). Moreno-Pino and Zohren (2024) apply dilated causal convolutions directly to high-frequency data, building on the time-series convolutional designs of Borovykh, Bohte and Oosterlee (2018), and Betz, Heil and Peter (2023) train convolutional networks on image representations of intraday returns. The applied record around these architectures is broad: recurrent networks retain accuracy in turbulent periods (Souto and Moradi, 2023), convolutional-recurrent hybrids outperform standalone learners (Ma, Hong and Song, 2024), boosting implementations dominate intraday prediction tasks (Wu and Wang, 2021), ensembles of boosting with networks improve on either component (Zhang, 2022), kernel methods capture nonlinearities without architectural complexity and remain stable across regime shifts (LeBaron, 2018; Şanlı, Balcılar and Özmen, 2025), and textual sentiment adds predictive content to the standard feature set (Audrino, Sigrist and Ballinari, 2020). The theoretical warrant for the flexible class as a whole is the universal approximation property (Hornik, Stinchcombe and White, 1989), and the representation-learning framework of Zerveas et al. (2021) supplies the general template that the transformer applications instantiate.

The aggregate verdict is not settled. Branco, Rubesam and Zevallos (2024), forecasting ten global indices, find no evidence that nonlinear machine learning beats linear models by a statistically significant margin. Audrino and Chassot (2025), working with 1,445 US equities and evaluating by QLIKE, mean squared error and a realized utility criterion, find that a carefully specified HAR model with an appropriate re-estimation scheme is not surpassed by machine learning alternatives given the same predictors. Their result is methodologically important beyond its substantive finding: it demonstrates that the fitting scheme, and not only the hypothesis class, determines the outcome of such comparisons.

An important qualification concerns the information set rather than the algorithm. The comparisons that favour the econometric benchmark are typically conducted with a restricted predictor set — lagged realized measures and an implied volatility index — which is precisely the information that the HAR cascade was designed to summarise efficiently. The comparisons that favour flexible learners typically supply a much larger predictor space including cross-sectional, macroeconomic and microstructure variables. It follows that the disagreement in the literature is partly about what is being compared: whether a flexible model can extract more from the same information, or whether it can exploit information that the parsimonious specification cannot represent at all. These are different claims with different implications, and studies rarely distinguish them explicitly.

The horizon dimension adds a further complication that is directly relevant to Stage two. The relative advantage of flexible learners is concentrated at short horizons, where the recent intraday history is most informative and where interactions among realized measures carry signal that an additive cascade cannot represent. As the horizon lengthens and the target is aggregated, the mean-reverting component dominates, the effective number of independent observations falls, and the variance of a flexible estimator rises while its marginal informational advantage declines. A trading system does not consume forecasts at an arbitrary horizon: it consumes them at the horizon of its rebalancing decision. The choice of model is therefore not separable from the design of the system that will use it, which is a first instance of the general point this review develops.

3.7. Model class and the structure of the data

The instability of the aggregate verdict has an explanation outside finance. Comparative studies of learning on tabular data consistently find that gradient-boosted trees remain competitive with, and frequently superior to, deep architectures (Shwartz-Ziv and Armon, 2022; Grinsztajn, Oyallon and Varoquaux, 2022), and surveys of deep learning on such problems reach compatible conclusions (Borisov et al., 2022). Three properties are advanced as the explanation, and all three apply to volatility panels: tree ensembles select features implicitly and are therefore robust to uninformative predictors; they represent non-smooth, piecewise-constant target functions efficiently; and they do not require the sample sizes that high-capacity sequential models need to estimate their parameters.

The relevance to this review is direct. A volatility panel constructed from a few hundred engineered predictors on a market with a few dozen liquid instruments is exactly the regime in which this literature predicts trees to win and sequence models to struggle. It follows that the label "machine learning" is a poor unit of comparison: the families it covers differ from one another by margins comparable to, and in some samples larger than, the margin separating either of them from the econometric benchmark.

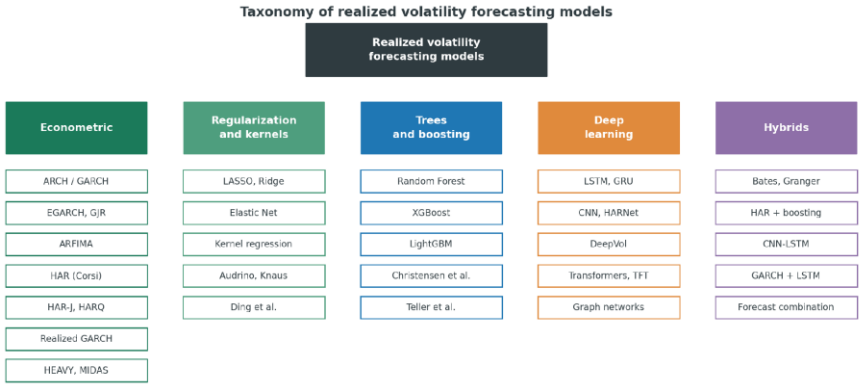

Figure 3. Taxonomy of realized volatility forecasting models by family, with representative specifications.

3.8. Combination

Bates and Granger (1969) established that a weighted combination of forecasts can dominate each of its constituents, and the subsequent literature has repeatedly found simple combinations difficult to improve upon (Genre et al., 2013). Applied across families rather than within one, the principle yields hybrids of econometric and machine learning forecasts, and the evidence assembled in Section 3.6 suggests why such combinations are attractive: the relative advantage of the two families varies with the forecast horizon and with the re-estimation scheme, so a weighted combination is a hedge against the instability that the comparative studies document.

The horizon dependence described above implies that a single set of combination weights is unlikely to be optimal across the horizon range, and the finding of Audrino and Chassot (2025) that the re-estimation scheme drives comparative outcomes applies with equal force to the combination layer. Empirical support on an emerging market is provided in Lysenok (2025), where a hybrid of the HAR-J specification with gradient boosting, its weights shifting

towards the econometric component as the horizon lengthens, improves on the benchmark significantly at every horizon under annual re-estimation.

3.9. The apparatus of comparison

Evaluation within this stage is methodologically mature. Patton (2011) established which loss functions preserve the true ranking of forecasts when the target is measured with error, identifying the quasi-likelihood and squared-error families as robust; the quasi-likelihood loss is additionally asymmetric in the direction that matters for risk management, penalising underestimation more heavily. Diebold and Mariano (1995) provide the test for the significance of differences in expected loss, and Hansen, Lunde and Nason (2011) extend the logic to sets of models. The multiplicity problem that arises from searching over many specifications is addressed by White (2000) and Romano and Wolf (2005), and its consequences for backtested strategies are set out by Bailey et al. (2014) and, in the context of factor discovery, by Harvey, Liu and Zhu (2016). Tashman (2000) collates the protocols for out-of-sample evaluation.

This apparatus answers one question completely and another not at all. It establishes which model forecasts more accurately, with a defensible notion of statistical significance. It says nothing about whether the difference is worth anything, and the remaining sections of this review are concerned with that question.

It is worth being precise about why the gap is not closed by the loss function itself. A robust loss preserves the ranking of forecasts with respect to the true conditional variance; it does not preserve the ranking of forecasts with respect to any particular decision that uses them. Two forecasts with identical expected loss can differ in the distribution of their errors — in whether they err in calm or turbulent states, in whether they overshoot or undershoot at the thresholds where a trading rule changes behaviour — and a decision rule that is sensitive to those states will not rank them equally. This is not a defect of the loss function but a limit on what a loss function can do, and it is the formal reason why the transitions in Figure 1 require separate evidence rather than following from Stage one by implication.

4. Stage two: integration into trading decisions

4.1. Systematic strategies and the inefficiencies they exploit

The strategies that consume volatility forecasts fall into a small number of families defined by the market inefficiency they exploit. Counter-trend rules rest on the mean reversion documented by Poterba and Summers (1988); trend-following rules on the momentum established by Jegadeesh and Titman (1993) in the cross-section and by Moskowitz, Ooi and Pedersen (2012) in the time series; relative-value rules on the convergence property analysed by Gatev, Goetzmann and Rouwenhorst (2006). The profitability of technical rules more generally has been examined since Brock, Lakonishok and LeBaron (1992), formalised by Lo, Mamaysky and Wang (2000), surveyed by Park and Irwin (2007), and shown by Menkhoff (2010) to be widely used by institutional managers, which establishes that the object of this stage is a live practice and not merely an academic construct.

The distinction between families matters for the argument because they have opposite exposures to volatility. Counter-trend rules profit when prices oscillate around a level and lose when a directional move persists, which makes them vulnerable in high-volatility states. Trend rules profit from persistence and are eroded by whipsaw, which makes them vulnerable in quiet, range-bound conditions. A volatility forecast is therefore not a uniformly favourable or unfavourable signal for a book of strategies: it is a signal whose interpretation depends on which strategy is being considered. This is the structural reason why a gating mechanism can add value that a uniform exposure adjustment cannot, and it is a feature of active books that the passive-portfolio literature of Section 5.3 has no occasion to encounter.

Risk management inside these strategies is itself a use of volatility information, though usually of realized rather than forecast volatility. Distances for stop-loss and take-profit levels are conventionally set as multiples of a range-based estimator, position sizes as a function of the same quantity. The substitution of a forecast for a backward-looking

estimate at these points is the narrowest form of integration and the one considered in Section 4.3; the broader forms considered in Sections 4.4 and 4.5 change what the strategy does rather than how it parameterises what it was going to do anyway.

4.2. A classification of integration channels

A volatility forecast can reach a trading decision through several mechanisms, and the literature has treated them as interchangeable applications of a single idea. They are not. The classification used here distinguishes four channels by what the forecast determines, which is also what governs the sensitivity of the outcome to forecast quality.

The scheme originates in Lysenok (2026a), which defined three integration channels — trade parameters, entry gating and position sizing, labelled B, C and D against a forecast-free baseline A — and evaluated them on a common data set, a common strategy universe and a common inference procedure. That study, to the author's knowledge, is the first to compare integration mechanisms rather than to examine one of them; the individual mechanisms had been introduced separately by Kaminski and Lo (2014), Fleming, Kirby and Ostdiek (2001) and Moreira and Muir (2017) respectively. The present review retains those labels so that results remain comparable across the two papers, and extends the scheme with a fourth channel, allocation, which operates at the portfolio rather than the position level and is discussed in Section 5.

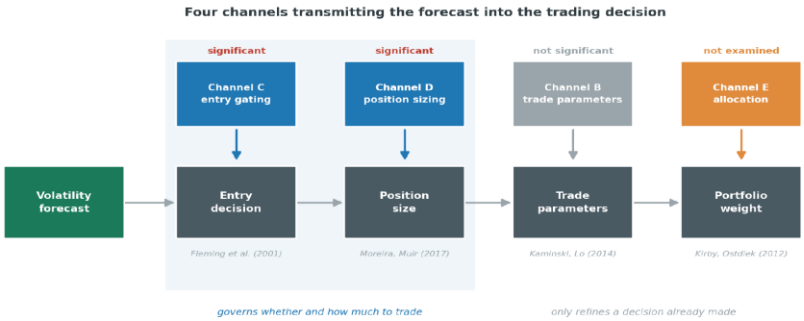

Figure 4. Four channels transmitting a volatility forecast into the trading decision. Shading marks the channels in which the forecast governs the commitment of capital.

Table 1. Classification of integration channels

Channel	Mechanism	What it determines	Principal studies	Established effect
B. Trade parameters	Scaling of stop-loss and take-profit distances	How an open position is managed	Kaminski and Lo (2014)	Modest and not reliably significant
C. Entry gating	Admission mask on the percentile of the forecast	Whether a position is opened at all	Fleming, Kirby and Ostdiek (2001)	Significant gains in investor utility
D. Position sizing	Exposure scaled inversely to the forecast	How much capital is committed	Moreira and Muir (2017); Barroso and Santa-Clara (2015)	Significant, primarily through risk reduction
E. Allocation	Weights across strategies or assets	How capital is distributed within a book	Kirby and Ostdiek (2012); Conrad, Kleen and Lönn (2025)	Established for passive portfolios, not for active books

Source: compiled by the author. Channels B, C and D follow the classification introduced in Lysenok (2026a); channel E is added in the present review.

The analytically important boundary runs between channel B on one side and channels C, D and E on the other. In B the forecast modifies the management of a position whose existence has already been settled by a directional signal. In the others it determines whether capital is committed and in what quantity. If forecast quality has economic value, there is no reason for these two groups to respond to it identically, and the evidence reviewed below indicates that they do not.

The formal intuition is that the channels differ in the curvature of the mapping from forecast to outcome. Where the forecast enters as a smooth multiplier on a quantity that is itself optimised — a stop distance, a take-profit level — the optimisation can absorb a systematic error in the forecast by shifting the multiplier, so the derivative of outcome with respect to forecast quality is small in the neighbourhood of the optimum. Where the forecast enters a threshold that determines a discrete action, no such absorption is available: an error that moves the forecast across the threshold changes the action entirely. Channels C and D differ in degree here — gating is a pure threshold, sizing is continuous but unbounded in its effect on capital at risk — but both belong to the class in which forecast error is not absorbed by parameter recalibration.

4.3. The trade-parameter channel

Kaminski and Lo (2014) provide the reference analysis of stop-loss rules, showing that they add value under momentum dynamics and subtract it under mean reversion, with a magnitude they characterise as modest. The theoretical reason is visible in the structure of the channel: a parameter optimisation can compensate for a less accurate forecast by recalibrating the response function, so the marginal contribution of forecast quality is attenuated by construction. This is the channel in which improvements in forecasting should be expected to matter least, and the empirical results assembled in Section 8 are consistent with that expectation.

4.4. The entry-gating channel

Gating uses the forecast to determine admissibility rather than magnitude. Fleming, Kirby and Ostdiek (2001) provide the canonical implementation in the volatility-timing framework, and the mechanism has since been applied through regime classification, where the forecast is discretised into states and different strategy types are admitted in different states. The economic logic is asymmetric with respect to the risk-free rate: capital withheld from the market earns the money-market rate, so the opportunity cost of a correct refusal to trade falls as that rate rises. This dependence has received almost no attention, for the straightforward reason that most of the evidence was generated in a low-rate environment. The one direct measurement is Lysenok (2026c): on the Russian market, as the policy rate rose from 9.9 to 21 per cent, the relative advantage of the forecast-informed system over its forecast-free counterpart widened from a factor of 1.7 to a factor of 4.7, with the base system’s Sharpe ratio collapsing from 2.99 to 0.40 while the gated system never fell below 1.87. The rate environment is not background here; it is an amplifier of the channel effect, and it defines the fourth research task of Section 9.

Implementations of the discretisation vary in ways that are not innocuous. A percentile rule applied to a rolling window adapts to the level of volatility but is sensitive to the window length; a fixed threshold is stable but drifts out of calibration as the unconditional level shifts; a latent-state model estimated jointly with the forecast introduces a second layer of estimation error. The comparative evidence on these choices is thin, and since the gating channel is the one in which the forecast has the largest measured effect, the sensitivity of results to the discretisation rule is a gap of practical importance rather than a technicality.

A related observation concerns exposure. A gating rule reduces the fraction of time a system is in the market, and any reduction in exposure mechanically reduces both return and risk. The diagnostic that separates a genuine forecast-driven filter from a mechanical de-risking is whether return is preserved while exposure falls: if a system trades materially less often without losing return, the entries it declined were disproportionately unprofitable, which a random or calendar-based reduction would not achieve. This diagnostic is straightforward to compute and is reported only occasionally.

4.5. The position-sizing channel

Scaling exposure inversely to a volatility estimate is the most extensively studied mechanism. Hallerbach (2012) provides the formal argument for its optimality over time; Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016) demonstrate that it removes the crash risk of momentum strategies; Baltas and Kosowski (2019) show that the choice of volatility estimator materially affects the result, which links this stage back to Stage one. Clements and Vatsa (2025) decompose the forecast by frequency and find that the low-frequency component produces the better portfolio outcome, concluding that forecast evaluation should be aligned with the investment objective rather than conducted independently of it — a conclusion that anticipates the argument of this review.

4.6. What has not been compared

Each channel has been studied; the channels have not been compared. No study in the corpus evaluates all mechanisms under a common data set, a common strategy universe and a common inference procedure, which means that the practically decisive question — given a fixed budget for model development, where should a forecast be inserted — has not been posed in a form that admits an answer.

Part of the reason is methodological cost. A clean comparison requires that everything other than the channel be held constant, which is straightforward for gating, where strategy parameters can be carried over unchanged and only the input forecast varies, but not for sizing or parameter scaling, where the parameters interact with the forecast and must be re-optimised for each configuration. The asymmetry in cost has shaped the literature: the channel that is cheapest to test in isolation is not the channel that most studies test, and the configurations that would make the comparison informative are the ones that require the most computation.

The exception, and the only direct evidence in the corpus, is Lysenok (2026a), which evaluates three channels against a forecast-free baseline on a fixed universe of seventeen liquid Moscow Exchange equities, a fixed book of six strategies spanning counter-trend, trend and range families, and a common bootstrap inference procedure, over a walk-forward control period from 2022 to 2026. The results separate the channels sharply. Entry gating and position sizing raise the Sharpe ratio of the multi-strategy portfolio by +1.19 and +1.22 respectively ($p < 0.001$), delivering certainty-equivalent gains of up to +2.07 per cent per annum for a conservative investor; the trade-parameter channel yields +0.15 of Sharpe ratio and no certainty-equivalent gain distinguishable from zero at any level of risk aversion. A companion study (Lysenok, 2026b) reverses the experiment, holding the channels fixed and varying the forecasting model: a 28 per cent improvement in QLIKE is worth +0.79 to +0.85 of Sharpe ratio when routed through gating or sizing, and +0.04, indistinguishable from zero, when routed through trade parameters. Read together, the two designs establish the same boundary from both directions — varying the channel with the model fixed, and varying the model with the channel fixed — which is the strongest form the evidence can currently take.

The strongest indirect evidence that the question matters comes from Wade (2026), who evaluates graph neural network forecasts on 465 US equities and finds that the model with the lowest forecast error, the model with the highest cross-sectional ranking accuracy and the model with the highest portfolio Sharpe ratio are three different models. Forecast accuracy, ranking quality and portfolio performance are related but not interchangeable objectives, and the ordering of models is not preserved across them.

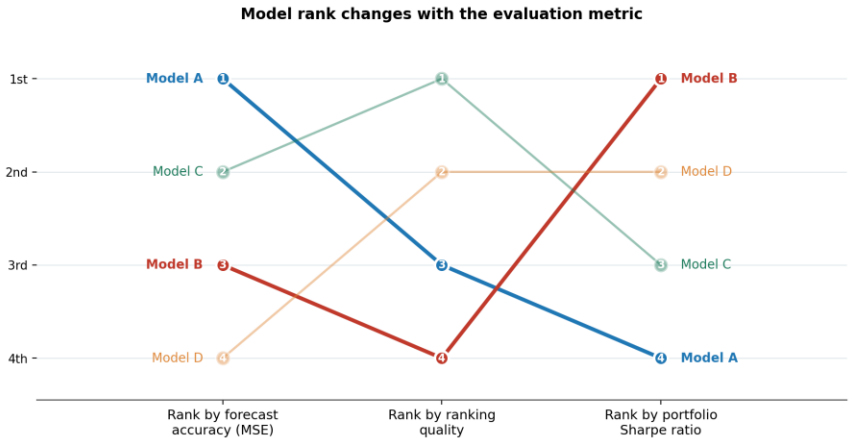

Figure 5. Ranking of the same models under three evaluation metrics. Crossing lines indicate that the ordering is not preserved. Model labels are generic; the figure conveys the character of the divergence rather than specific values.

4.7. The scope of the direct evidence

It should be stated plainly that the direct comparison of channels rests on a single market, and the review does not claim more for it than that. Two considerations govern how the limitation reads. First, the conditions of the test were adverse rather than favourable: a control period in which the market index lost 28 per cent, a policy rate reaching 21 per cent that priced every act of capital commitment against a demanding alternative, and a structural break of unusual magnitude in 2022. A channel effect that emerges under those conditions is unlikely to be an artefact of a benign sample. Second, the design is inexpensive to replicate, since the gating channel requires only the substitution of the input forecast with all strategy parameters held fixed; replication across markets is the first task of the research agenda in Section 9, and until it accumulates, the channel boundary should be treated as a well-evidenced hypothesis rather than a settled fact.

5. Stage three: aggregation into portfolios

5.1. Weighting schemes and the difficulty of beating equal weights

The mean-variance framework of Markowitz (1952) provides the reference solution, and the persistent finding of the applied literature is that it is difficult to implement profitably. DeMiguel, Garlappi and Uppal (2009) show that equal weighting is not outperformed by optimisation across a wide range of data sets, because estimation error in the inputs offsets the theoretical advantage of the optimum. Kirby and Ostdiek (2012) qualify the conclusion by demonstrating that simple active schemes, including inverse-volatility weighting, do outperform naive diversification once turnover is controlled, which is the point at which volatility forecasting enters portfolio construction.

The schemes that occupy the middle ground between equal weighting and full optimisation share a common feature: they use second moments and discard expected returns. Inverse-volatility weighting assigns capital in proportion to the reciprocal of forecast volatility; risk parity equalises risk contributions across components; minimum-variance solves the optimisation with the mean vector suppressed. All three are direct consumers of volatility forecasts, and all three are motivated by the same finding — that errors in expected returns dominate errors in second moments — which makes them the portfolio-level counterpart of the channel classification developed in Section 4.2. Where a forecast determines weights rather than a binary decision, the effect of forecast error is smoothed across components, which is why the allocation channel occupies an intermediate position in the taxonomy.

5.2. Estimation error as the binding constraint

Chopra and Ziemba (1993) establish the relative importance of errors in expected returns, variances and covariances for portfolio choice, and the ordering they document explains why volatility forecasting is a more tractable route to improvement than return forecasting. Jagannathan and Ma (2003) show that weight constraints function as an implicit regularisation of the covariance estimate, and Ledoit and Wolf (2004) provide the explicit shrinkage estimator for the high-dimensional case. This line establishes the conditions under which better second-moment estimates translate into better portfolios, and it is the closest the literature comes to addressing the third transition of Figure 1 directly.

5.3. Volatility-managed portfolios

The empirical foundation of the low-risk approach is the anomaly documented by Ang, Hodrick, Xing and Zhang (2006), the inverse relationship between volatility and risk-adjusted return that makes second-moment forecasts directly monetisable. Moreira and Muir (2017) show that scaling factor exposures by the inverse of recent realized variance earns significant alpha for the major equity factors, a result that has generated both extension and dispute. Cederburg, O'Doherty, Wang and Yan (2020) evaluate 103 strategies and find that volatility-managed versions rarely outperform their unmanaged counterparts out of sample once the scaling rule must be estimated in real time. Harvey, Hoyle, Korgaonkar, Rattray, Sargaison and Van Hemert (2018) reach a compatible but differently framed conclusion: the principal and reliable effect of volatility targeting across asset classes is the stabilisation of realized risk rather than the enhancement of returns.

These findings are not as contradictory as they appear once the evaluation criterion is made explicit. A mechanism that reduces variance without reducing mean return raises risk-adjusted performance whether or not it generates alpha in a factor regression, and the disagreement in the literature is substantially a disagreement about which of these is the relevant measure. Conrad, Kleen and Lönn (2025) provide recent evidence on the constructive side, showing that using forecasts rather than retrospective risk measures to sort and weight a large cross-section delivers significant economic gains net of transaction costs.

5.4. Multi-strategy books

A parallel literature studies portfolios of strategies rather than of assets. Fung and Hsieh (1997) characterise the return profiles of dynamic trading strategies; Agarwal and Kale (2007) compare multi-strategy structures with funds of funds; Newton, Platanakis, Stafylas, Sutcliffe and Ye (2021) examine diversification across strategy types. The institutional context is set out by Stulz (2007) and the practice by Pedersen (2015). What this literature does not do is introduce a volatility forecast: the diversification benefit is analysed on the basis of realized correlations among strategy returns, not as a function of the forecast that drives the strategies.

5.5. The empty intersection

The two lines of Sections 5.3 and 5.4 do not meet. Volatility management is established on passive factor portfolios and asset-class overlays; multi-strategy books are studied without reference to forecast quality. Whether the mechanism that works on a factor portfolio also works inside a book of heterogeneous active strategies with discrete entry signals is, on the evidence assembled here, an open question, and it is the second of the three transitions identified in Figure 1.

The first study to occupy the intersection is, to the author's knowledge, Lysenok (2026a), which introduces a volatility forecast into a multi-strategy book through the three-level aggregation of instruments into strategy sleeves and sleeves into a portfolio, and evaluates the result by both statistical and economic criteria. Two of its findings bear directly on the questions this section has raised. The diversification benefit of the multi-strategy structure survives the introduction of the forecast — the average pairwise correlation among the six strategy sleeves is 0.09, with counter-trend and trend sleeves negatively correlated — so the forecast does not homogenise the book, which was not obvious in advance given that all sleeves respond to the same volatility signal. And the improvement operates through the composition of trading rather than through uniform de-risking: under the gating configuration the book is in the market

less than half as often while its return is preserved, which is the diagnostic, introduced in Section 4.4, that separates a forecast-driven filter from a mechanical reduction of exposure. Whether these properties transfer to books with different strategy mixes and to other markets is unexamined, and the claim of novelty here is correspondingly narrow: not the multi-strategy structure, which Section 5.4 attributes to its authors, and not volatility management, which Section 5.3 attributes to its own, but their combination under joint statistical and economic evaluation.

There is a structural reason to expect the transfer to be imperfect. A factor portfolio is continuously invested, so scaling its exposure is a smooth operation on a position that always exists. An active book is invested intermittently, and its strategies have opposite volatility exposures as described in Section 4.1, so scaling the book uniformly conflates two distinct operations: adjusting the size of positions that would be taken anyway, and altering which positions are taken. The aggregation level at which the scaling is applied — individual instrument, strategy sleeve, or the book as a whole — determines which of these dominates, and the literature contains no systematic treatment of that choice.

6. Stage four: economic value to the investor

6.1. From a statistical metric to a utility function

A difference in a loss function does not establish that anyone would pay for the better forecast. The framework that addresses this defines the value of a forecast as the fee that would make an investor with specified preferences indifferent between the forecast-based portfolio and a benchmark. The formulation in terms of the certainty equivalent of a mean-variance investor is due to Fleming, Kirby and Ostdiek (2001), and its principal virtue is that it aggregates changes in mean and in variance into a single number on the scale of return, which is the scale on which the decision to invest in a forecasting system is actually taken.

The construction requires a coefficient of risk aversion, which is not observable, and results are conventionally reported across a range of values. The behaviour of the estimate across that range is itself informative: if the increment in certainty equivalent rises with risk aversion, the gain is being generated predominantly through variance reduction; if it falls, through higher mean return. This diagnostic is available at no additional cost and distinguishes the two economic mechanisms that a difference in Sharpe ratios conflates, yet it is reported in only a minority of the studies that compute the metric at all.

6.2. Estimates and their range

Fleming, Kirby and Ostdiek (2001; 2003) estimate that a risk-averse investor would pay between 50 and 200 basis points per year for the improvement in covariance estimates obtained by moving from daily to intraday data. Marquering and Verbeek (2004) extend the framework to the joint prediction of returns and volatility. Bollerslev, Patton and Quaedvlieg (2018) formulate a realized utility criterion in which the investor targets a constant level of risk and the quality of the volatility forecast determines how closely that target is met; this criterion has since been adopted as a secondary metric in several forecasting studies, including Zhang, Zhang, Cucuringu and Qian (2024) and Audrino and Chassot (2025), and, in the panel setting of Bollerslev, Hood, Huss and Pedersen (2018), it demonstrates significant economic gains from exploiting volatility similarities across asset classes. It represents the most direct existing bridge between Stage one and Stage four.

6.3. Costs, thresholds and the scale of operation

Economic value net of implementation cost is a different quantity from economic value gross of it, and the distinction determines who can use a forecasting system. Transaction costs are routinely incorporated; the fixed costs of building and maintaining the system — data acquisition, infrastructure, personnel — are not. Since these costs are largely independent of the capital deployed against them, they imply a threshold below which a forecasting system cannot pay for itself regardless of its statistical quality. The threshold is rarely computed. The one explicit calculation in the corpus is again the author's: in Lysenok (2026a) the annual cost of operating the forecasting system, estimated from a survey of market practitioners, is set against the certainty-equivalent gain to yield a break-even capital of approximately 160 million roubles, a level available, on central bank data, to fewer than three per cent of active

investors on that market. The number itself is market-specific; the exercise is not, and its absence elsewhere leaves the question of who can afford a forecasting system outside the scope of a literature that is nominally about economic value.

6.4. Inference on performance measures

The Sharpe ratio (Sharpe, 1994) remains the standard summary of risk-adjusted performance, and inference on differences between Sharpe ratios requires care because returns are neither independent nor normally distributed. Ledoit and Wolf (2008) provide a robust procedure based on the stationary bootstrap of Politis and Romano (1994), and its use has become the expected standard in this literature.

Alternative summaries address the asymmetry that the Sharpe ratio ignores. Downside-deviation measures penalise only adverse dispersion, which matches the preferences of most investors more closely than a symmetric variance; drawdown-based measures relate return to the worst peak-to-trough loss and correspond to the constraint that most institutional mandates actually impose. Neither displaces the Sharpe ratio as the reporting convention, and the practical consequence is that a mechanism whose principal effect is on the shape of the loss distribution rather than on its variance will be undervalued by the standard metric. Since volatility management is precisely such a mechanism, this is not a neutral choice of convention.

6.5. What has been measured and what has not

The apparatus of Stage four has been applied predominantly to covariance timing in asset allocation problems and to factor portfolios. Its application to books of active strategies with discrete entry decisions is rare, and its application to markets with a materially non-zero risk-free rate is rarer still. Since the arithmetic of the certainty equivalent depends directly on the return available to capital that is not deployed, this is not a minor gap: the entire quantitative literature on the economic value of volatility timing was generated under conditions in which withholding capital was nearly costless in nominal terms, and the extent to which its conclusions transfer to a high-rate environment has not been established.

7. Emerging markets and the Russian case

Evidence on all four stages is concentrated in developed markets. The emerging-market literature is dominated by Stage one, and the Russian corpus, which is substantial and largely inaccessible in English, illustrates the pattern clearly.

The early work applied the conditional-variance framework to daily index data and established the macroeconomic drivers, with oil prices, the exchange rate and the policy rate identified as significant (Fedorova and Nazarova, 2010; Fedorova and Buzlov, 2013). Asaturov and Teplova (2014) documented volatility spillovers from global markets using multivariate specifications, which is the empirical basis for including global assets in the predictor set of any model built on this market. The decisive methodological shift came with Aganin (2017), who compared 88 GARCH specifications, 10 HAR-RV modifications and 4 ARFIMA models on ten Russian assets using five-minute data and found HAR-RV specifications entering the model confidence set, thereby transferring to this market the finding established internationally in Section 3.4. Aganin (2020) and Aganin and Peresetsky (2018) linked index and currency volatility to oil prices and sanctions.

A parallel line places the market in an international context. Peresetsky (2014) and Korhonen and Peresetsky (2016) identify the contribution of global and political shocks; Pogorelova and Peresetsky (2020) extract a global stochastic trend from non-synchronous volatility observations; Manevich, Peresetsky and Pogorelova (2022) and Aganin, Manevich, Peresetsky and Pogorelova (2023) extend the comparison to cryptocurrency markets. Alternative data sources have been examined by Bazhenov and Fantazzini (2019), who combine search-query data with implied volatility, and by Sidorov, Date and Balash (2013), who incorporate news analytics into the conditional variance specification.

Machine learning has entered this literature recently and unevenly. Patlasov and Garafutdinov (2024) apply recurrent architectures to index volatility; Patlasov (2025) develops hybrid deep learning approaches for exchange-traded funds; Glebova and Kovaleva (2024) examine forecastability under the sanctions regime. These studies work predominantly with index or fund data at daily frequency rather than with individual equities at intraday frequency, which is a material restriction given the evidence of Section 3.3. The microstructure reason is documented by Gurov (2023): spreads outside the most liquid names and measurable price impact within the index constrain the sampling frequency at which a realized estimator remains informative, so the five-minute standard of developed markets is not directly transferable.

Stages two to four are represented far more thinly. Asaturov and Teplova (2014) and Lakshina (2016) address hedging with explicit risk preferences; Teplova and Mikova (2014) examine the determinants of momentum strategies. Systematic evaluation of how volatility forecasts translate into trading and portfolio outcomes on this market has appeared only recently, in a sequence of studies by the author: the forecasting model and its accuracy (Lysenok, 2025), the comparison of integration channels (Lysenok, 2026a), the model-family comparison under a fixed channel structure (Lysenok, 2026b) and the dependence of the effect on the policy rate (Lysenok, 2026c). The gap is therefore not merely one of coverage but of stage: an emerging market with a high policy rate is precisely the environment in which the arithmetic of Section 6.5 differs most from the developed-market baseline, and it is the environment least studied at the stages where that arithmetic operates.

Table 3. The Russian-language corpus by pipeline stage, data frequency and type of evaluation

Line of enquiry	Stage	Data frequency	Evaluation	Principal studies
Parametric volatility models	1	Daily	Statistical	Fedorova and Nazarova (2010); Fedorova and Buzlov (2013)
Realized volatility and HAR	1	Intraday	Statistical	Aganin (2017; 2020); Aganin and Peresetsky (2018)
Global factors and spillovers	1	Daily	Statistical	Peresetsky (2014); Pogorelova and Peresetsky (2020)
Machine learning and neural networks	1	Daily	Statistical	Patlasov and Garafutdinov (2024); Patlasov (2025); Glebova and Kovaleva (2024)
Alternative data sources	1	Daily	Statistical	Bazhenov and Fantazzini (2019); Sidorov, Date and Balash (2013)
Microstructure and liquidity	1	Intraday	—	Gurov (2023)
Strategies and hedging	2	Daily	Economic	Asaturov and Teplova (2014); Teplova and Mikova (2014); Lakshina (2016)

Forecast integration into active books	**2–4**	**Intraday**	**Both**	**Lysenok (2025; 2026a; 2026b; 2026c)**

Source: compiled by the author.

The table makes the structure of the gap explicit. Six of the eight lines of enquiry sit entirely at Stage one; five of them work at daily frequency, which is the restriction that Section 3.3 identifies as consequential; and only one line applies both statistical and economic criteria. The pattern is not specific to this market — it reproduces at smaller scale the asymmetry documented for the corpus as a whole in Section 8 — but it is more pronounced here, and it coincides with the market conditions under which the unexamined questions would be most informative to study.

8. Synthesis: where the chain breaks

Table 2 consolidates the review. For each stage it records what has been established, the metric by which it is established, and the question that remains open.

Table 2. State of the evidence by pipeline stage

Stage	What is established	Evaluation metric	Remaining gap	Studies
1. Forecasting models	Model hierarchy known; HAR remains a strong benchmark; the machine learning advantage is not uniform	QLIKE, MSE, Diebold–Mariano and MCS tests	Evaluation does not extend beyond the loss function	**85**
2. Trading decisions	Each integration channel examined separately	Sharpe ratio, win rate	Channels never compared under common conditions	**19**
3. Portfolios	Volatility management established on passive factor portfolios	Portfolio Sharpe, maximum drawdown	Active multi-strategy books not examined	**20**
4. Investor utility	Apparatus mature; gains of 50–200 b.p. for developed markets	ΔCE, realized utility, cost thresholds	Applied to covariance timing, not to active books	**12**
Emerging markets	Forecasting models adapted; economic evaluation largely absent	Predominantly statistical metrics	High risk-free-rate regime unexamined	**27**
TOTAL	**Of 153 studies, 8 connect statistical forecast quality to an economic outcome within one design**	—	—	**153**

Source: compiled by the author. A study spanning several stages is counted in each, so stage counts sum to more than the total.

8.1. The quantitative asymmetry

The distribution of evaluation practice across the corpus is shown in Figure 6. Eighty-one studies evaluate forecasts by a statistical loss alone. Thirty-five evaluate a trading or portfolio outcome without any assessment of the quality of the forecast driving it. Eight apply both criteria within a single design. The remaining studies are methodological or definitional and have no empirical evaluation to classify. This is the principal quantitative result of the systematisation, and it holds across the full period covered.

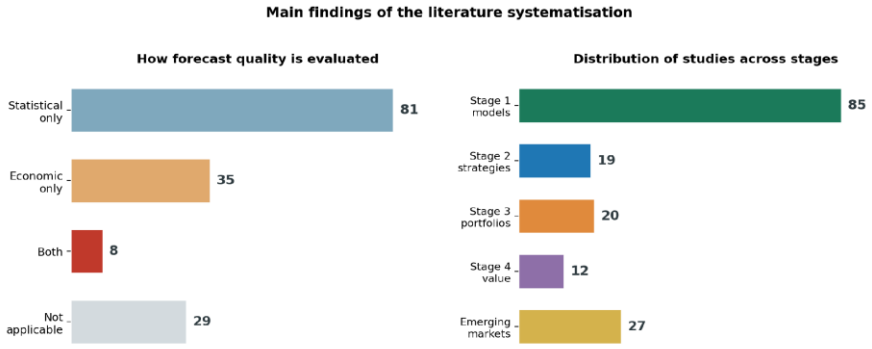

Figure 6. Distribution of the 153 reviewed studies by type of evaluation and by pipeline stage.

Figure 7 renders the same classification as a flow from stage to evaluation criterion, and it makes the structure of the field visible in a way the marginal counts cannot: the mass of Stage one flows almost undivided into statistical evaluation, the strategy and portfolio stages flow into economic evaluation, and the ribbons connecting any stage to joint evaluation are the thinnest in the diagram. The field is not merely under-supplied with joint evaluation; it is organised so that the two criteria are applied by different literatures to different objects.

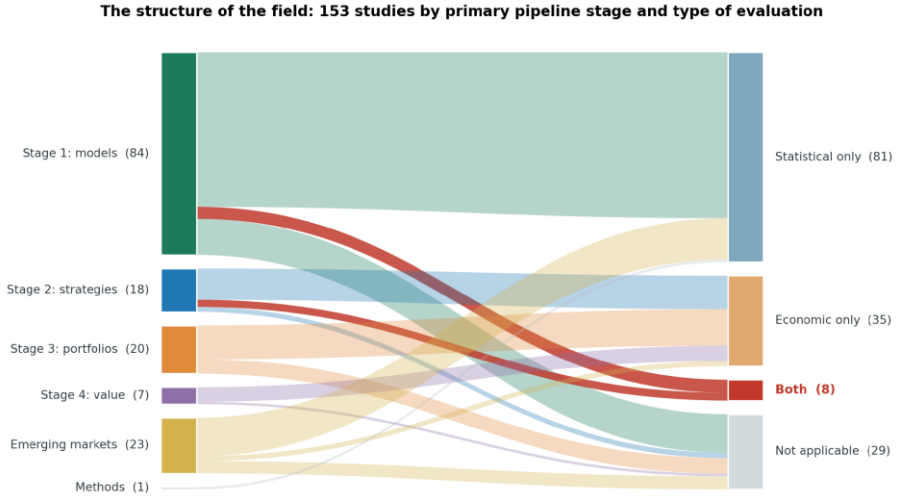

Figure 7. The structure of the field: 153 studies routed from their primary pipeline stage to the type of evaluation they apply. Ribbon width is proportional to the number of studies; the flows into joint evaluation are highlighted.

8.2. Three apparent contradictions

Three disagreements in the literature dissolve once the stage and the criterion are made explicit. The first is between Moreira and Muir (2017), who find alpha from volatility management, and Cederburg et al. (2020), who find it absent out of sample: the disagreement concerns the estimation of the scaling rule in real time rather than the mechanism itself, and the qualification of Harvey et al. (2018) — that the reliable effect is risk stabilisation rather than

return enhancement — reconciles them. The second is between studies reporting machine learning gains and Branco et al. (2024) and Audrino and Chassot (2025), who do not: Section 3.7 supplies the resolution, in that the aggregate label conflates families with different inductive biases, and Audrino and Chassot add that the fitting scheme of the benchmark is itself decisive. The third is the divergence documented by Wade (2026) between rankings under different metrics, which is not a contradiction at all but the direct expression of the review's central claim, and which the paired designs of Lysenok (2026a; 2026b) resolve into a mechanism: rankings diverge because the channel, not the model, determines whether accuracy is convertible.

8.3. What the mapping implies

The implication for practice inverts the usual research sequence. The first question is not which model forecasts best but whether the system that will consume the forecast contains a channel through which forecast quality can propagate to the allocation of capital. Where no such channel exists, improvements in the model will not reach the portfolio, and the appropriate investment is in the architecture of the system. Where such a channel exists, the choice of model becomes consequential, and the evidence assembled here indicates that it should be indexed to the horizon at which the system consumes the forecast and that combination across families is more robust than replacement of one family by another.

The channel classification is offered as the instrument that makes such a comparison expressible. Its value is not that it enumerates mechanisms already known individually, but that it fixes a common vocabulary in which results obtained under different mechanisms can be placed side by side, and identifies the property — whether the forecast governs the commitment of capital — along which they separate.

8.4. What the mapping cannot establish

Three limits on the inference should be stated. First, the counts reported here describe the literature, not the world: that few studies connect statistical and economic evaluation is evidence about research practice, and it would be a fallacy to read it as evidence that the connection is weak. The direction of the substantive relationship is what Sections 4 and 6 assess, and the evidence there is suggestive rather than conclusive.

Second, publication practice plausibly biases the observed distribution. A study that finds no economic value in a statistically superior forecast is harder to place than one that finds value, and the seven studies applying both criteria are unlikely to be a random sample of the studies that computed both. The direction of that bias works against the review's central claim, which strengthens rather than weakens it, but the magnitude cannot be estimated from the corpus.

Third, the classification into channels is a construct imposed on a literature that did not use it. Studies were assigned to channels by what their mechanism does, not by how the authors described it, and a small number of designs combine mechanisms in ways that resist a single assignment. The taxonomy is offered as an instrument for organising future comparisons rather than as a description of how the field has understood itself.

9. Conclusion

This review has traced the volatility forecast through the four stages that separate a model from an investor and has mapped the evidence available at each. The mapping supports three conclusions. The evaluation apparatus of the forecasting literature is mature and self-contained, answering completely the question of which model is more accurate and not at all the question of whether the difference is worth anything. The mechanisms by which a forecast enters a trading decision are distinct in their sensitivity to forecast quality, and they have been studied one at a time rather than compared. And the quantitative asymmetry in evaluation practice — 81 studies statistical, 35 economic, 8 both — indicates that the transition between statistical and economic assessment is not a minor gap in an otherwise complete literature but its principal structural feature.

Five research tasks follow, each stated so that the measurement required is explicit.

First, the relationship between forecast accuracy and economic outcome should be estimated across a full range of models rather than inferred from a comparison of two. The relevant question is whether the relationship is monotone, and the entry-gating channel offers the cheapest route to answering it, since strategy parameters can be held fixed while only the input forecast varies — a property demonstrated in practice in Lysenok (2026b), where the substitution required no re-optimisation.

Second, the field needs a common protocol for economic evaluation. Three incompatible metrics are currently in use — the increment in certainty equivalent, the realized utility criterion, and the difference in Sharpe ratios — and results reported under one are not comparable with results reported under another, although all three derive from the same utility formulation.

Third, sampling frequency should be treated as an experimental variable rather than a design constraint. Running an identical pipeline on intraday, hourly and daily targets would identify the point at which economic value is lost, and would establish whether the conclusions of the intraday literature transfer to markets where intraday data are unavailable or prohibitively expensive.

Fourth, the dependence of the results on the risk-free rate should be quantified. The arithmetic of withholding capital changes materially when the money-market alternative yields a substantial return, and the recent experience of emerging markets operating at policy rates above fifteen per cent provides a natural experiment that the developed-market literature cannot supply.

Fifth, robustness to regime breaks should be measured in economic rather than statistical terms. The speed at which model classes degrade after a structural break and the speed at which they recover under re-estimation are known in terms of forecast loss for a few cases; the corresponding question in terms of portfolio outcome has not been posed.

The unifying point is that forecast evaluation should be aligned with the decision the forecast is meant to inform. Statistical accuracy remains a valid proxy for economic value, but only conditionally, and the condition is the existence of a channel through which accuracy can reach the allocation of capital.

10. Declarations

Funding. This research received no external funding.

Conflicts of interest. The author declares no conflict of interest. The empirical studies whose results are cited in Sections 4 to 7 (Lysenok, 2025; 2026a; 2026b; 2026c) are the author's own prior publications; they are identified as such wherever cited, and no result from them is reported here for the first time.

Data availability. The classification of the 153 reviewed studies underlying Figures 2, 6 and 7 and Tables 2 and 3 is available from the author on request.

Use of artificial intelligence. AI-based tools were used in the preparation of the manuscript for literature search assistance, figure generation and language editing. All classifications, judgements and conclusions are the author's own, and all sources were verified by the author.

REFERENCES

1. Aganin A.D. (2017). Comparison of GARCH and HAR-RV models for forecasting realized volatility in the Russian market. Applied Econometrics, 48, 63-84. (In Russian)
2. Aganin A.D. (2020). Volatility of the Russian stock index: oil and sanctions. Voprosy Ekonomiki, 2, 86-100. (In Russian)
3. Aganin A.D., Manevich V.A., Peresetsky A.A., Pogorelova P.V. (2023). Comparison of volatility forecasting models for cryptocurrencies and the stock market. HSE Economic Journal, 27(1), 9-32. (In Russian)
4. Aganin A.D., Peresetsky A.A. (2018). Volatility of the rouble exchange rate: oil and sanctions. Applied Econometrics, 52, 5-21. (In Russian)
5. Agarwal, Vikas & Kale, Jayant R. (2007). On the Relative Performance of Multi-Strategy and Funds of Hedge Funds. CFR Working Papers 07-11, University of Cologne, Centre for Financial Research (CFR).

6. Agarwal, Vikas and Naik, Narayan Y., Risks and Portfolio Decisions Involving Hedge Funds.
7. Andersen T.G., Bollerslev T. (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. International Economic Review, 39(4), 885-905.
8. Andersen T.G., Bollerslev T. (1998). Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies. The Journal of Finance, 53(1), 219-265.
9. Andersen T.G., Bollerslev T., Diebold F.X. (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. The Review of Economics and Statistics, 89(4), 701-720.
10. Andersen T.G., Bollerslev T., Diebold F.X., Labys P. (2001). The Distribution of Realized Exchange Rate Volatility. Journal of the American Statistical Association, 96(453), 42-55.
11. Andersen T.G., Bollerslev T., Diebold F.X., Labys P. (2003). Modeling and Forecasting Realized Volatility. Econometrica, 71(2), 579-625.
12. Ang A., Hodrick R.J., Xing Y., Zhang X. (2006). The Cross-Section of Volatility and Expected Returns. The Journal of Finance, 61(1), 259-299.
13. Asaturov K.G., Teplova T.V. (2014). Construction of hedging coefficients for highly liquid stocks of the Russian market based on GARCH-class models. Economics and Mathematical Methods, 50(1), 37-54. (In Russian)
14. Audrino F., Chassot J. (2025). HARd to Beat: The Overlooked Impact of Rolling Windows in the Era of Machine Learning. International Journal of Forecasting.
15. Audrino F., Knaus S.D. (2016). Lassoing the HAR Model: A Model Selection Perspective on Realized Volatility Dynamics. Econometric Reviews, 35(8-10), 1485-1521.
16. Audrino F., Sigrist F., Ballinari D. (2020). The Impact of Sentiment and Attention Measures on Stock Market Volatility. International Journal of Forecasting, 36(2), 334-357.
17. Bailey D.H., Borwein J.M., López de Prado M., Zhu Q.J. (2014). Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance. Notices of the AMS, 61(5), 458-471.
18. Baltas, Nick and Baltas, Nick and Kosowski, Robert, Demystifying Time-Series Momentum Strategies: Volatility Estimators, Trading Rules and Pairwise Correlations (September 8, 2019).
19. Barndorff-Nielsen O.E., Shephard N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. Journal of Financial Econometrics, 2(1), 1-37.
20. Barroso P., Santa-Clara P. (2015). Momentum Has Its Moments. Journal of Financial Economics, 116(1), 111-120.
21. Bates, J.M. & Granger, C.W.J. (1969). The Combination of Forecasts. Journal of the Operational Research Society, 20(4), 451-468.
22. Bazhenov T.I., Fantazzini D. (2019). Forecasting realized volatility of Russian stocks using Google Trends and implied volatility. Russian Journal of Industrial Economics, 12(1), 79-88. (In Russian)
23. Berzon N.I., Lysenok N.I. (2021). Assessing the investment attractiveness of BRICS stock markets. Finance and Business, 17(4), 18-31. (In Russian)
24. Berzon N.I., Teplova T.V. (eds.) (2013). Innovations in Financial Markets. Moscow: HSE Publishing House. (In Russian)
25. Betz N., Heil T., Peter F. (2023). The Best of Both Worlds? Predicting Realized Volatility Based on Convolutional Neural Nets and Coloured Intraday Return Images. Working paper.
26. Black F., Scholes M. (1973). The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81(3), 637-654.
27. Bollerslev T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31(3), 307-327.
28. Bollerslev T., Hood B., Huss J., Pedersen L.H. (2018). Risk Everywhere: Modeling and Managing Volatility. The Review of Financial Studies, 31(7), 2729-2773.
29. Bollerslev T., Patton A.J. & Quaedvlieg R. (2018). Modeling and Forecasting (Un)Reliable Realized Covariances for More Reliable Financial Decisions. Journal of Econometrics, 207(1), 71-91.
30. Bollerslev T., Patton A.J., Quaedvlieg R. (2016). Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting. Journal of Econometrics, 192(1), 1-18.
31. Borisov V., Leemann T., Seßler K., Haug J., Pawelczyk M., Kasneci G. (2022). Deep Neural Networks and Tabular Data: A Survey. IEEE Transactions on Neural Networks and Learning Systems.
32. Borovykh A., Bohte S., Oosterlee C.W. (2018). Dilated Convolutional Neural Networks for Time Series Forecasting. Journal of Computational Finance.
33. Boudt K., Kleen O., Sjørup E. (2022). Analyzing Intraday Financial Data in R: The highfrequency Package. Journal of Statistical Software, 104(8), 1-36.
34. Branco R.R., Rubesam A., Zevallos M. (2024). Forecasting Realized Volatility: Does Anything Beat Linear Models? Journal of Empirical Finance, 78, 101524.
35. Brini A., Toscano G. (2025). SpotV2Net: Multivariate Intraday Spot Volatility Forecasting via Vol-of-Vol-Informed Graph Attention Networks. International Journal of Forecasting, 41(3), 1093-1111.
36. Brock W., Lakonishok J., LeBaron B. (1992). Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. The Journal of Finance, 47(5), 1731-1764.
37. Bucci A. (2020). Realized Volatility Forecasting with Neural Networks. Journal of Financial Econometrics, 18(3), 502-531.

38. Campbell J.Y. & Viceira L.M. (2002). Strategic Asset Allocation: Portfolio Choice for Long-Term Investors. Oxford University Press.
39. Čech F., Barunik J. (2017). On the Modelling and Forecasting of Multivariate Realized Volatility: Generalized Heterogeneous Autoregressive (GHAR) Model. Journal of Forecasting, 36(2), 181-206.
40. Cederburg S., O'Doherty M.S., Wang F., Yan X.S. (2020). On the Performance of Volatility-Managed Portfolios. Journal of Financial Economics, 138(1), 95-117.
41. Chen T., Guestrin C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 785-794.
42. Cho K., van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. Proceedings of EMNLP 2014, 1724-1734.
43. Chopra V.K., Ziemba W.T. (1993). The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. The Journal of Portfolio Management, 19(2), 6-11.
44. Christensen B.J., Prabhala N.R. (1998). The relation between implied and realized volatility. Journal of Financial Economics, 50(2), 125-150.
45. Christensen K., Siggaard M., Veliyev B. (2023). A Machine Learning Approach to Volatility Forecasting. Journal of Financial Econometrics, 21(5), 1680-1727.
46. Clements A., Preve D.P. (2021). A Practical Guide to Harnessing the HAR Volatility Model. Journal of Banking & Finance, 133, 106285.
47. Clements A., Vatsa P. (2025). Forecasting Realized Volatility Using HAR Models and Wavelet Decomposition: A Volatility-Timing Perspective. Working paper.
48. Conrad C., Kleen O., Lönn R. (2025). Volatility Forecasting for Low-Volatility Investing. International Journal of Forecasting, 570-586.
49. Cont R. (2001). Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. Quantitative Finance, 1(2), 223-236.
50. Corsi F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. Journal of Financial Econometrics, 7(2), 174-196.
51. Corsi, Fulvio and Kretschmer, Uta and Mittnik, Stefan and Pigorsch, Christian, The Volatility of Realized Volatility (November 28, 2005).
52. Daniel K., Moskowitz T.J. (2016). Momentum Crashes. Journal of Financial Economics, 122(2), 221-247.
53. Degiannakis S., Filis G., Klein T., Walther T. (2022). Forecasting Realized Volatility of Agricultural Commodities. International Journal of Forecasting, 38(1), 74-96.
54. DeMiguel V., Garlappi L., Uppal R. (2009). Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? The Review of Financial Studies, 22(5), 1915-1953.
55. Diebold F.X., Mariano R.S. (1995). Comparing Predictive Accuracy. The Journal of Business & Economic Statistics, 13(3), 253-263.
56. Ding Y., Kambouroudis D., McMillan D.G. (2021). Forecasting Realised Volatility: Does the LASSO Approach Outperform HAR? Journal of International Financial Markets, Institutions and Money, 74, 101386.
57. Engle R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987-1007.
58. Engle R.F., Ng V.K. (1993). Measuring and Testing the Impact of News on Volatility. The Journal of Finance, 48(5), 1749-1778.
59. Fama E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), 383-417.
60. Fedorova E.A., Buzlov D.A. (2013). Forecasting the Russian stock market using GARCH modelling. Financial Analytics: Science and Experience, 16. (In Russian)
61. Fedorova E.A., Nazarova Yu.N. (2010). Identifying factors influencing stock market volatility using a cointegration approach. Economic Analysis: Theory and Practice, 3, 17-24. (In Russian)
62. Fedorova E.A., Pankratov K.A. (2011). Modelling stock market volatility during crisis periods. Audit and Financial Analysis. (In Russian)
63. Fleming J., Kirby C., Ostdiek B. (2001). The Economic Value of Volatility Timing. The Journal of Finance, 56(1), 329-352.
64. Fleming, J., Kirby, C. & Ostdiek, B. (2003). The Economic Value of Volatility Timing Using "Realized" Volatility. Journal of Financial Economics, 67(3), 473-509.
65. Frank J. (2023). Forecasting Realized Volatility in Turbulent Times Using Temporal Fusion Transformers. FAU Discussion Papers in Economics.
66. Fung, William & Hsieh, David A. (1997). Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds. The Review of Financial Studies, 10(2), 275-302.
67. Gatev E., Goetzmann W.N., Rouwenhorst K.G. (2006). Pairs Trading: Performance of a Relative-Value Arbitrage Rule. The Review of Financial Studies, 19(3), 797-827.
68. Genre, V., Kenny, G., Meyler, A. & Timmermann, A. (2013). Combining Expert Forecasts: Can Anything Beat the Simple Average? International Journal of Forecasting, 29(1), 108-121.

69. Ghysels, E., Santa-Clara, P. & Valkanov, R. (2006). Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies. Journal of Econometrics, 131(1-2), 59-95.
70. Glebova A.G., Kovaleva A.A. (2024). Forecasting the volatility of the Russian stock market in the context of international economic sanctions. Finance: Theory and Practice, 28(1), 20-29. (In Russian)
71. Glosten L.R., Jagannathan R., Runkle D.E. (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. The Journal of Finance, 48(5), 1779-1801.
72. Gong X., Lin B. (2019). Modeling Stock Market Volatility Using New HAR-Type Models. Physica A, 516, 194-211.
73. Grinsztajn L., Oyallon E., Varoquaux G. (2022). Why Do Tree-Based Models Still Outperform Deep Learning on Typical Tabular Data? Advances in Neural Information Processing Systems, 35.
74. Gu S., Kelly B., Xiu D. (2020). Empirical Asset Pricing via Machine Learning. The Review of Financial Studies, 33(5), 2223-2273.
75. Gurov S.V. (2023). Illiquidity effects on the Russian stock market. HSE Economic Journal, 27(1). (In Russian)
76. Hallerbach W.G. (2012). A Proof of the Optimality of Volatility Weighting Over Time. Journal of Investment Strategies, 1(4), 87-99.
77. Hansen P.R., Huang Z., Shek H.H. (2012). Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility. Journal of Applied Econometrics, 27(6), 877-906.
78. Hansen P.R., Lunde A. (2005). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? The Journal of Applied Econometrics, 20(7), 873-889.
79. Hansen P.R., Lunde A., Nason J. (2011). The Model Confidence Set. Econometrica, 79(2), 453-497.
80. Harvey C.R., Hoyle E., Korgaonkar R., Rattray S., Sargaison M., Van Hemert O. (2018). The Impact of Volatility Targeting. The Journal of Portfolio Management, 45(1), 14-33.
81. Harvey C.R., Liu Y., Zhu H. (2016). …and the Cross-Section of Expected Returns. The Review of Financial Studies, 29(1), 5-68.
82. Hochreiter S., Schmidhuber J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.
83. Hornik K., Stinchcombe M., White H. (1989). Multilayer Feedforward Networks Are Universal Approximators. Neural Networks, 2(5), 359-366.
84. Izzeldin M., Hassan M.K., Pappas V., Tsionas M. (2019). Forecasting Realised Volatility Using ARFIMA and HAR Models. Quantitative Finance, 19(10), 1627-1638.
85. Jagannathan R., Ma T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. The Journal of Finance, 58(4), 1651-1683.
86. Jegadeesh N., Titman S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. The Journal of Finance, 48(1), 65-91.
87. Kambouroudis D.S., McMillan D.G., Tsakou K. (2016). Forecasting Stock Return Volatility: A Comparison of GARCH, Implied Volatility, and Realized Volatility Models. Journal of Futures Markets, 36(12), 1127-1163.
88. Kambouroudis D.S., McMillan D.G., Tsakou K. (2021). Forecasting Realized Volatility: The Role of Implied Volatility, Leverage Effect, Overnight Returns, and Volatility of Realized Volatility. Journal of Futures Markets, 41(10), 1618-1639.
89. Kaminski K.M., Lo A.W. (2014). When Do Stop-Loss Rules Stop Losses? Journal of Financial Markets, 18, 234-254.
90. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and, Liu, T.-Y. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, December 2017, 3149-3157.
91. Kim, H.Y. and Won, C.H. (2018) Forecasting the Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple GARCH-Type Models. Expert Systems with Applications, 103, 25-37.
92. Kirby C., Ostdiek B. (2012). It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naïve Diversification. Journal of Financial and Quantitative Analysis, 47(2), 437-467.
93. Korhonen I., Peresetsky A. (2016). What Influences Stock Market Behavior in Russia and Other Emerging Countries? Emerging Markets Finance and Trade, 52(5), 1210–1225.
94. Lakshina V.V. (2016). Dynamic hedging accounting for the degree of risk aversion. HSE Economic Journal, 1. (In Russian)
95. LeBaron B. (2018). Forecasting Realized Volatility with Kernel Ridge Regression. Working paper.
96. Ledoit O., Wolf M. (2004). A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. Journal of Multivariate Analysis, 88(2), 365-411.
97. Ledoit O., Wolf M. (2008). Robust Performance Hypothesis Testing with the Sharpe Ratio. Journal of Empirical Finance, 15(5), 850-859.
98. Leushuis R.M., Petkov N. (2026). Advances in Forecasting Realized Volatility: A Review of Methodologies. Financial Innovation, 12, Article 14.
99. Lo A.W., Mamaysky H., Wang J. (2000). Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. The Journal of Finance, 55(4), 1705-1765.
100. Luong C., Dokuchaev N. (2018). Forecasting of Realised Volatility with the Random Forests Algorithm. Journal of Risk and Financial Management, 11(4), 61.
101. Lysenok N.I. (2025). Application of machine learning for volatility forecasting and trading strategy enhancement on the Russian stock market. Fundamental and Applied Mathematics, 25(4), 90-107. (In Russian)

102. Lysenok N.I. (2026a). Effectiveness of volatility forecasts in active trading strategies of institutional investors on the Russian equity market. Fundamental and Applied Mathematics, 26(3), 33-42. (In Russian)
103. Lysenok N.I. (2026b). Which Model Families Pay? Econometric, Gradient Boosting and Recurrent Neural Network Volatility Forecasts in Active Trading Strategies on the Russian Stock Market. International Journal of Computer Information Systems and Industrial Management Applications, 18(17s), 533-549.
104. Lysenok N.I. (2026c). When Does Volatility-Informed Position Sizing Matter Most? Policy-Rate Dependence and Diversification Arithmetic of Multi-Strategy Trading in MOEX Index Stocks. Enterprise Development and Microfinance, 36(4s), 81-92.
105. Lysenok N.I., Berzon N.I., Rechmedina S.A. (2024). Factor investing: can stock price changes now be explained? Finance and Business, 20(4), 42-56. (In Russian)
106. Lysenok N.I., Mamochkin A.I. (2024). Potential for the application of financial structured products within BRICS. World Economics, 7. (In Russian)
107. Ma W., Hong Y., Song Y. (2024). On Stock Volatility Forecasting under Mixed-Frequency Data Based on Hybrid RR-MIDAS and CNN-LSTM Models. Mathematics, 12(10), 1538.
108. Mandelbrot B. (1963). The Variation of Certain Speculative Prices. The Journal of Business, 36(4), 394-419.
109. Manevich V.A., Peresetsky A.A., Pogorelova P.V. (2022). Stock market volatility and cryptocurrency volatility. Applied Econometrics, 65, 65-76. (In Russian)
110. Markowitz H. (1952). Portfolio Selection. Journal of Finance, 7(1), 77-91.
111. Marquering, W.A. & Verbeek, M. (2004). The Economic Value of Predicting Stock Index Returns and Volatility. Journal of Financial and Quantitative Analysis, 39(2), 407-429.
112. Menkhoff L. (2010). The Use of Technical Analysis by Fund Managers: International Evidence. Journal of Banking & Finance, 34(11), 2573-2586.
113. Moreira A., Muir T. (2017). Volatility-Managed Portfolios. The Journal of Finance, 72(4), 1611-1644.
114. Moreno-Pino F., Zohren S. (2024). DeepVol: Volatility Forecasting from High-Frequency Data with Dilated Causal Convolutions. Quantitative Finance, 24(8), 1105-1127.
115. Moskowitz T.J., Ooi Y.H., Pedersen L.H. (2012). Time Series Momentum. The Journal of Financial Economics, 104(2), 228-250.
116. Müller, U.A., Dacorogna, M.M., Davé, R.D., Olsen, R.B., Pictet, O.V. & von Weizsäcker, J.E. (1997). Volatilities of Different Time Resolutions — Analyzing the Dynamics of Market Components. Journal of Empirical Finance, 4(2-3), 213-239.
117. Nelson D. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica, 59(2), 347-370.
118. Newton, D., Platanakis, E., Stafylas, D., Sutcliffe, C. & Ye, X. (2021). Hedge Fund Strategies, Performance & Diversification: A Portfolio Theory & Stochastic Discount Factor Approach. British Accounting Review, 53(5), Article 101000.
119. Park C.-H., Irwin S.H. (2007). What Do We Know About the Profitability of Technical Analysis? Journal of Economic Surveys, 21(4), 786-826.
120. Patlasov D.A. (2025). Hybrid approaches to forecasting realized ETF volatility: deep learning and the recovery theorem. HSE Economic Journal, 29(1), 103-131. (In Russian)
121. Patlasov D.A., Garafutdinov R.V. (2024). Application of LSTM neural networks to modelling stock market volatility. Perm University Herald. Economy, 19(1), 41-51. (In Russian)
122. Patton A.J. (2011). Volatility Forecast Comparison Using Imperfect Volatility Proxies. Journal of Econometrics, 160(1), 246-256.
123. Patton A.J., Sheppard K. (2015). Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility. The Review of Economics and Statistics, 97(3), 683-697.
124. Pedersen, Lasse Heje (2015). Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined. Princeton University Press.
125. Peresetsky A.A. (2014). What Drives the Russian Stock Market: World Market and Political Shocks. International Journal of Computational Economics and Econometrics, 4(1/2), 82–95.
126. Peresetsky A.A., Yakubov R.I. (2017). Autocorrelation in an Unobservable Global Trend: Does It Help to Forecast Market Returns? International Journal of Computational Economics and Econometrics, 7(1/2), 152–169.
127. Pogorelova P.V., Peresetsky A.A. (2020). Isolating a global stochastic trend from non-synchronous observations of financial index volatility. Applied Econometrics, 57, 53-71. (In Russian)
128. Poon S.-H., Granger C. W. J. (2003). Forecasting Volatility in Financial Markets: A Review. Journal of Economic Literature, 41(2), 478-539.
129. Poterba J.M., Summers L.H. (1988). Mean Reversion in Stock Prices: Evidence and Implications. The Journal of Financial Economics, 22(1), 27-59.
130. Ramos-Pérez E., Alonso-González P.J., Núñez-Velázquez J.J. (2021). Multi-Transformer: A New Neural Network-Based Architecture for Forecasting S&P Volatility. Mathematics, 9(15), 1794.
131. Reisenhofer R., Bayer X., Hautsch N. (2022). HARNet: A Convolutional Neural Network for Realized Volatility Forecasting. arXiv:2205.07719.
132. Romano J.P., Wolf M. (2005). Stepwise Multiple Testing as Formalized Data Snooping. Econometrica, 73(4), 1237-1282.

133. Rossi E. (2010). Univariate GARCH models: a survey. Quantile, 8, 1-67. (In Russian)
134. Şanlı S., Balcılar M., Özmen M. (2025). Predicting the Volatility of Bitcoin Returns Based on Kernel Regression. Annals of Operations Research, 352(3), 505-542.
135. Sharpe W.F. (1994). The Sharpe Ratio. The Journal of Portfolio Management, 21(1), 49-58.
136. Shephard, N. & Sheppard, K. (2010). Realising the Future: Forecasting with High-Frequency-Based Volatility (HEAVY) Models. Journal of Applied Econometrics, 25(2), 197-231.
137. Shwartz-Ziv R., Armon A. (2022). Tabular Data: Deep Learning Is Not All You Need. Information Fusion, 81, 84-90.
138. Sidorov S.P., Date P., Balash V.A. (2013). Using news analytics data in GARCH models. Applied Econometrics, 29(1). (In Russian)
139. Souto H.G., Moradi A. (2023). Forecasting Realized Volatility through Financial Turbulence and Neural Networks. Economics and Business Review, 9(2), 133-159.
140. Stulz, René M. (2007). Hedge Funds: Past, Present, and Future. Journal of Economic Perspectives, 21(2), 175-194.
141. Tashman L.J. (2000). Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review. International Journal of Forecasting, 16(4), 437-450.
142. Teller A., Pigorsch U., Pigorsch C. (2022). Short-to Long-Term Realized Volatility Forecasting Using Extreme Gradient Boosting. Working paper.
143. Teplova T.V., Mikova E.S. (2014). Issuer size, trading activity and stock liquidity as determinants of momentum portfolio strategies. Vestnik NSU. Socio-Economic Sciences, 14(2), 14-23. (In Russian)
144. Teplova T.V., Sokolova T.V., Faizulin M.S., Kurkin A.V. (2022). Investor Sentiment and Anomalies in the Behaviour of Market Characteristics of Investment Assets. Moscow: INFRA-M. (In Russian)
145. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30.
146. Wade R. (2026). Do Better Volatility Forecasts Lead to Better Portfolios? Evidence from Graph Neural Networks. Working paper.
147. Wang Y., Ma F., Wei Y., Wu C. (2016). Forecasting Realized Volatility in a Changing World: A Dynamic Model Averaging Approach. Journal of Banking & Finance, 64, 136-149.
148. White H. (2000). A Reality Check for Data Snooping. Econometrica, 68(5), 1097-1126.
149. Wilder, J.W. (1978). New Concepts in Technical Trading Systems. Trend Research.
150. Wu Y., Wang Q. (2021). LightGBM Based Optiver Realized Volatility Prediction. Proceedings of IEEE CSAIEE, 227-230.
151. Zerveas G., Jayaraman S., Patel D., Bhamidipaty A., Eickhoff C. (2021). A Transformer-Based Framework for Multivariate Time Series Representation Learning. Proceedings of the 27th ACM SIGKDD Conference, 2114-2124.
152. Zhang C., Zhang Y., Cucuringu M., Qian Z. (2024). Volatility Forecasting with Machine Learning and Intraday Commonality. Journal of Financial Econometrics, 22(2), 492-530.
153. Zhang X. (2022). A Model Combining LightGBM and Neural Network for High-Frequency Realized Volatility Forecasting. Proceedings of ICFIED 2022, 2906-2912.