Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Forecasting accuracy to portfolio value: Model risk and regime dependence in an emerging equity market

Darko Vuković

vdarko@hotmail.rs

Geographical Institute “Jovan Cvijić” SASA

Stanislav Gushchin

Vega Institute Foundation

Boris Lodiagin

HSE University

Gleb Poleshkin

Vega Institute Foundation

Research Article

Keywords: factor investing, return prediction, machine learning, regime dependence, portfolio selection, emerging markets

Posted Date: September 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11010674/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Forecasting accuracy to portfolio value: Model risk and regime dependence in an emerging equity market

Abstract: Portfolio managers have uncertainty at two levels: the expected returns they estimate and the forecasting model on which those estimates depend. We address this problem using Moscow Exchange equities from October 2016 to September 2025 and compare eight linear, adaptive, nonlinear and sequence-based models in a common rolling out-of-sample framework. All forecasts use the same information set, the same reliability adjustment and the same constrained mean-variance allocation rule. CatBoost has the smallest average prediction error, but none of the models have a meaningful ability to rank cross-sectionally. Portfolio results are different from statistical rankings: PatchTST achieves competitive risk-adjusted performance, notwithstanding poor unconditional forecast accuracy. This inconsistency cannot be solved by model selection. The recent-model-performance selector performs lower than the fixed-model strategies. The regime-based rule produces the best nominal results but does not produce a statistically reliable advantage.

Keywords: factor investing; return prediction; machine learning; regime dependence; portfolio selection; emerging markets.

JEL classification: C53; G11; G12; G17

1. Introduction

Estimates of expected returns are essential to portfolio optimization, but they are never independent of the model that generates them. The problem is more general than classical parameter uncertainty. Investors must also decide which forecasting framework to believe, whether the recent model performance will persist, when changing market conditions warrant a different specification, and whether improved statistical accuracy translates into better portfolio outcomes. Such questions are of importantly relevance in settings where the association between characteristics of firms and future returns is unstable. Size, value and momentum effects have been widely documented but their associated premia vary across countries and over time (Fama and French 1992, 2015; Carhart 1997; Asness, Moskowitz and Pedersen 2013; Cakici, Fabozzi and Tan 2013). The key questions for portfolio management are how well the forecasting model adapts to shifting market conditions and how accurately expected returns can be estimated.

This question is especially relevant in emerging equity markets where the macro-economic and financial environment can change in relatively short estimation windows. Inflation, volatility,

liquidity and market drawdowns may affect the level of expected returns and the information contained in firm characteristics. The Moscow Exchange (MOEX) is a useful setting in which to study this issue. The Russian equity market went through a period of sustained inflationary pressure between 2016 and 2025; the disruption related to the COVID-19 pandemic and the geopolitical and monetary tightening cycle that began in 2022. These events produced very different investment conditions over a relatively short sample period. Therefore, a forecasting relation estimated over the whole period may be averaging over states in which the same signal has different economic meanings. In such circumstances changes in market conditions can influence return relationships, change the relative usefulness of competing forecasting models and ultimately influence portfolio allocations.

Two aspects of literature are especially relevant to this problem. The research on conditional asset-pricing recognizes that expected returns are based on information available to investors and that risk premia are conditional on economic conditions (Ferson and Harvey 1991, 1993; Ang and Bekaert 2002). Machine-learning research also takes up the same instability from another direction, acknowledging richer nonlinearities, interactions and temporal structures than standard linear models allow (Gu, Kelly, and Xiu 2020; Chen, Pelger, and Zhu 2024). These approaches have strengthened the evidence on return predictability but the implications for portfolio choice are still not fully settled. Although machine-learning studies tend to rank models by predictive loss, portfolio studies tend to focus on estimation error and often take the expected-return specification as given. While the literature on forecast selection and combination has considered model blending and switching, the actual forecasting architecture has received less attention as a possible source of portfolio risk in a changing market environment.

This distinction is important because statistical forecasting performance and investment value are not the same. Mean squared error (MSE) is a summary measure of forecast accuracy over the entire cross-section, but portfolio outcomes are in large part determined by the smaller subset of securities that rank near the top of the distribution of expected returns. The security selection, weight constraints, covariance estimates, and the interaction between the expected returns and the allocation rule are factors affecting the outcome for the portfolio. A model may be a poor performer in the average but still identify a subset of securities that are useful for portfolio formation. In contrast, a model with relatively low average forecast error might have little economic value if it does not order the most relevant securities correctly. The divergence

between statistical and economic performance has been noted before (Cenesizoglu and Timmermann 2012; Kelly, Malamud, and Zhou 2024) but is relevant here when multiple forecasting architectures are compared in the same regime-dependent investment environment.

Our study addresses three related questions in this context. First, does the performance of expected-return models vary across observable states of inflation and market stress? Second, do models that perform well in terms of forecast loss or cross-sectional ranking generate more powerful portfolios under a common allocation rule? Third, does selection of forecasting models based on recent out-of-sample performance or current market conditions improve the results of portfolios?

Our empirical examination employs monthly data from the MOEX spanning January 2014 to September 2025, with the forecasting panel commencing in October 2016 following the establishment of necessary lags, filters, and market-state variables (24,636 security-month observations for 290 securities). The information set combines cross-sectional size, value, momentum and illiquidity rankings with market volatility, high-inflation and drawdown-stress factors, as well as interactions to capture state-dependent relationships. We consider eight forecasting specifications: OLS, ridge, lasso, elastic net, rolling OLS, Kalman time-varying-parameter model, CatBoost and PatchTST. The models are different ways of responding to instability. Regularized regressions manage estimation noise through shrinkage, rolling OLS emphasizes recent information by eliminating older data, the Kalman framework permits gradual adjustments of coefficients, CatBoost detects nonlinear interactions, and PatchTST exploits temporal correlations across sequences of prior observations.

All models are compared under the same rolling out-of-sample protocol and the same portfolio construction framework to make the comparison economically meaningful. Forecasts undergo the same lagged reliability adjustment before being fed into a quarterly long-only mean-variance allocation. The candidate set at each rebalancing date contains 25 securities out of 290, with the highest expected returns after reliability adjustment. The individual weights are capped at 15%, and all strategies employ the same covariance estimate and optimization procedure. Holding these elements constant, the differences in portfolio performance are mainly due to the differences in the forecasts of expected returns and not due to the differences in portfolio construction.

The results show that no single model dominates all evaluation criteria. CatBoost records the lowest average mean squared error, but cross-sectional rank correlations remain weak across all specifications. Forecast errors also rise during high-inflation and drawdown-stress periods, indicating that the state of the market is important for assessing forecast risk. The portfolio results, however, do not reproduce the statistical ranking of the models. PatchTST performs relatively poorly according to unconditional forecast diagnostics but delivers risk-adjusted portfolio performance close to the regularized linear benchmark and a smaller maximum drawdown. This divergence indicates that average predictive accuracy alone does not reveal how useful a model will be once its forecasts are translated into actual positions.

The results of active model selection further qualify this. The selector based on recent out of sample forecast performance also underperforms the fixed model strategies, suggesting that recent statistical rankings are too unstable to be used to guide model selection. The regime-based rule achieves the highest nominal annualized return and Sharpe ratio, but its bootstrap confidence intervals largely overlap those of the unconditional benchmarks. Moreover, the rule was not specified in advance of the portfolio backtest, so the observed benefit of the rule should be considered exploratory, not as evidence of a validated switching strategy. The broader implication is that observable market regimes may be useful in identifying periods when forecasting risk is high, but these regimes do not necessarily offer an automatic reliable basis for switching between forecasting architectures.

We make three contributions to literature: First, it treats the forecasting architecture itself as an input to the portfolio and subject to model uncertainty. It broadens the traditional focus on estimation error in a given model to include uncertainty about which model generates expected returns. Secondly, in our study we assess forecast loss, cross-sectional ranking and realized portfolio performance in a single controlled design. Since the same information set, reliability adjustment, covariance estimation, security-selection rule, and portfolio constraints are used, the comparison shows directly that the statistical and economic rankings of forecasting models can differ materially. This is consistent with the view that a lower average forecast loss does not necessarily imply more useful signals for portfolio decisions (Cenesizoglu and Timmermann 2012; Kelly, Malamud, and Zhou 2024). Third, we study the extent to which active model selection can reduce this uncertainty. We find that conditioning the choice of model on recent performance or market regimes does not necessarily add portfolio value and may add another layer of estimation risk. In this sense, information on market conditions seems

more useful to identify periods when the model is less reliable, rather than to dictate which model investors should use.

2. Literature background

2.1 Conditional expected returns and market states

The empirical literature shows evidence that firm characteristics are related to expected stock returns. Size is a key factor in Banz (1981) and Fama and French (1992, 1993), and momentum is an important return pattern as shown by Jegadeesh and Titman (1993) and Carhart (1997). Fama and French (2015) develop the factor model for profitability and investment, and Asness, Moskowitz, and Pedersen (2013) find value and momentum in markets and across asset classes. The findings provide an economic rationale for adding size, value, and momentum as predictors, but much weaker support for the assumption that the related return premia are stable over time. In particular, the evidence from emerging markets shows large variation across countries and time periods (Cakici, Fabozzi, and Tan (2013); Hou, Karolyi, and Kho (2011)). For the Russian market, another source of variation is the institutional and macroeconomic situation, which is hard to capture by a single stable relationship (Goriaev and Zabotkin 2006). Moreover, momentum effects are contingent on previous market conditions and are subject to sharp reversals during times of distress and subsequent recovery (Cooper, Gutierrez, and Hameed 2004; Daniel and Moskowitz 2016). More generally, the presence of structural breaks and the business cycle conditions can have an impact on the strength of return predictability (Pástor and Stambaugh 2001; Henkel, Martin and Nardari 2011).

Conditional asset pricing incorporates this instability into the information set available to investors. As the economic environment and investment opportunities change, so do the expected returns and risk exposures (Merton 1973; Cochrane 2005). Ferson and Harvey (1991, 1993) show that macroeconomic information helps to explain variation in domestic and international risk premium. Related evidence that conditioning information matters for expected returns and cross-sectional pricing is provided by Harvey (1989) and Lettau and Ludvigson (2001). Regime models tackle the same problem in a discrete setting where parameters and portfolio decisions vary across distinct market states (Hamilton 1989; Ang and Bekaert 2002; Guidolin and Timmermann 2008). In this paper we follow this conditional viewpoint but define market states with observable information instead of latent regimes. Inflation, volatility and drawdown are treated as conditions known at the forecast date and against which the performance of competing models can be judged.

Each of these conditions has a straightforward economic interpretation. The connection between inflation and stock returns is horizon-dependent and relies on how inflation influences real activity, discount rates, and firms' cash flows (Fama and Schwert 1977; Fama 1981; Boudoukh and Richardson 1993; Bekaert and Wang 2010; Katz, Lustig, and Nielsen 2017). Volatility reflects changes in both uncertainty and hedging demand and has been linked to changes in the expected returns of stocks (Campbell and Hentschel 1992; Ang et al. 2006; Campbell et al. 2018). Returns are affected by liquidity through trading costs, price impact and systematic liquidity risk (Amihud and Mendelson 1986; Amihud 2002; Pástor and Stambaugh 2003; Acharya and Pedersen 2005). These effects are especially pronounced in emerging markets and during periods of funding stress (Bekaert, Harvey, and Lundblad 2007; Brunnermeier and Pedersen 2009). Drawdown adds information on accumulated downside stress not contained in volatility alone (Chekhlov, Uryasev, and Zabarankin 2005).

The adaptive markets hypothesis is a useful framework for relating these evolving conditions to model selection. If relationships between forecasts change with competition, learning, institutions and market structure, the same relationship need not be equally useful in every period (Lo 2004, 2017). If the relationships between returns change slowly, a stable linear specification may suffice. More adaptive models, such as those that adapt coefficients, capture nonlinear interactions or account for temporal dependence, may be more useful in less stable periods. However, more flexibility is not necessarily better, since the tradeoff between adaptation and estimation error can also depend on market conditions.

2.2 Forecasting flexibility, regularization, and model uncertainty

This tradeoff has become a central issue in the asset pricing literature of machine learning. Nonlinear methods can capture interactions among predictors that are hard to capture using traditional linear specifications (Gu, Kelly, and Xiu, 2020). At the same time, more flexible models provide greater room for estimation errors when the return signal is weak. Recent work document significant emphasis on restrictions, regularization and economic structure. Chen, Pelger, and Zhu (2024) incorporate no-arbitrage restrictions into deep-learning asset-pricing models, and Freyberger, Neuhierl, and Weber (2020) and Bryzgalova, Pelger, and Zhu (2025) use more structured approaches to the return cross-section. Similarly, Kelly, Malamud and Zhou (2024) demonstrate that model complexity can be useful in noisy-return prediction when paired with effective regularization.

The value is even greater when predictive signals are directly embedded in the allocation of capital across the portfolio. Avramov, Cheng, and Metzker (2023) show that the economic gains of machine learning are reduced once realistic portfolio constraints are considered. Kozak, Nagel, and Santosh (2018, 2020) demonstrate how shrinkage can be helpful for estimation when there are many weakly identified components of the return cross-section. Nagel (2021) emphasizes regularization as a central aspect of financial machine learning. More flexibility does not eliminate the need for regularization. Time variation in model parameters may increase estimation uncertainty when the available history is limited (Lewellen and Nagel 2006).

The models compared here offer different solutions to this problem. Ridge, lasso and elastic net retain the linear specification, but they reduce coefficient instability through shrinkage. Rolling OLS changes by only using more recent observations for the estimation. The coefficients are updated slowly as new data becomes available in the Kalman specification. Catboost allows for nonlinear interactions between current and lagged information. PatchTST adopts a different approach and learns from the ordered history of predictors through a transformer-based sequence structure (Nie et al. 2023). The comparison is not simply one of simple versus complex models. It relies on a number of assumptions about the persistence of return relationships, the rate at which they should be allowed to change, and the degree of estimation noise an investor should be willing to endure in return for added flexibility.

This difference is relevant for our empirical setting. The MOEX sample has some obvious changes in inflation and market stress, but a relatively short time dimension to estimate these. A rapidly adapting model may be more responsive to true structural change but may also be noisier. The relevant comparison is economic: how much better forecasts are as a consequence of this extra flexibility so that portfolio choices are better.

2.3 Forecasting accuracy, portfolio value, and model-selection risk

Forecast accuracy is only one measure of that comparison. The literature suggests that statistical predictability is not necessarily economically valuable. As Welch and Goyal (2008) show, many return predictors perform poorly out-of-sample. Rapach and Zhou (2013) also make a distinction between statistical forecast quality and economic value. The difference is highly relevant for cross-sectional portfolio formation. Metrics like MSE measure the overall predictive accuracy across all securities, whereas a long-only portfolio depends mostly on the

top predicted return stocks. Even if a model has relatively large errors elsewhere in the cross-section, it can still be useful for security selection. The other way round is also possible: low average forecast error does not guarantee that the most relevant securities for investment are ranked correctly. Cenesizoglu and Timmermann (2012) show that statistical improvements in return forecasts do not necessarily imply improved portfolio performance.

This difference can be reinforced by the portfolio rule. The estimates of expected returns are used directly in the mean-variance optimization, resulting in portfolio weights that are sensitive to estimation and model uncertainty (Garlappi, Uppal, and Wang 2007). DeMiguel, Garlappi, and Uppal (2009) find that estimation error can make simple diversification a viable alternative to optimized diversification. While such constraints as long-only positions and weight caps may reduce this sensitivity, they cannot entirely remove it, particularly when the portfolio is constructed from a relatively small number of securities with high predicted returns. Such a condition motivates our use of a common allocation procedure for all forecasting models. Fixing the candidate universe, covariance estimate, reliability adjustment, and portfolio constraints allows the economic comparison to focus on differences in the expected-return signals.

A closely related source of uncertainty is the choice of model. Combinations of forecasts are robust because they diversify across imperfect models that are imperfect in different ways (Timmermann 2006; Rapach, Strauss, and Zhou 2010). A selector, in contrast, puts weight on the model that appears to be the most appropriate at a given time. This may be attractive when the model performance changes across regimes, but the decision itself has to be inferred from past observations. The short track record suggests that recent differences across models may be picking up noise rather than persistent changes in forecasting ability. Conditioning the comparison on specific market states makes the sample even smaller. Model selection does not eliminate model uncertainty but introduces an additional decision on which model should be trusted now.

This is where the strands of literature converge. In conditional asset pricing, expected returns are linked to changes in market conditions. There are various ways to respond to those changes, using modern forecasting techniques. However, portfolio research suggests that a statistically superior forecast is not necessarily a better investment signal. We do not know if the relative performance of competing forecasting architectures holds when their predictions are evaluated

under a common portfolio rule, or if changing models as market conditions change can improve upon fixed model strategies. We investigate these questions in a common rolling out of sample framework, where models are compared using forecast error, cross sectional ranking, and realized portfolio performance. Then we evaluate model selection based on recent performance and observable market conditions.

3. Data and methods

3.1 Data and investment universe

The analysis is based on a monthly panel of stocks traded on the MOEX from January 2014 to September 2025. We construct lags, rolling thresholds, volatility measures and market states from previous observations and the final forecasting panel spans from October 2016 to September 2025. A security-month is the unit of observation. Monthly returns are computed from the last valid trading price in each month, dividend adjusted prices where available. The monthly log return of security *i* is:

$$R_{i,t} = \ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right), \tag{1}$$

i denotes a security and t a month; $P_{i,t}$ is the last valid trading price of security i in month t, and $R_{i,t}$ its monthly log return, where the price term is the last valid monthly observation. Missing or non-positive prices are excluded. The risk-free rate is aligned to the monthly frequency, giving excess return:

$$R_{i,t}^{e} = R_{i,t} - R_{t}^{f}. \tag{2}$$

R_t^f is the monthly risk-free rate; $R_{i,t}^e$ is the excess return. Firm size is based on monthly market capitalization. Shares outstanding are aligned to their effective dates and carried forward between verified changes, giving:

$$\mathrm{MCAP}_{i,t} = P_{i,t} \times SO_{i,t}, \tag{3}$$

$SO_{i,t}$ is share-outstanding at the end of month t, carried forward between verified changes; $MCAP_{i,t}$ is market capitalization, the same quantity denoted $M_{i,t}$ in (25). Adjusted share count in Equation (3) is used for size classification and, where relevant, value-weighted aggregation. Quarterly book equity is mapped to the monthly panel and is used together with market

capitalization to construct the value characteristic. Liquidity is measured from daily trading data and introduced below with the remaining predictors.

The economic foundation follows the Fama-French-Carhart framework, with market, size, value, and momentum as the core return dimensions (Banz 1981; Fama and French 1992, 1993; Jegadeesh and Titman 1993; Carhart 1997). The market premium is:

$$\mathrm{MRP}_t = R_t^{\mathrm{mkt}} - R_t^f. \quad (4)$$

R_t^{mkt} is the return on the broad market index; MRP_t is the market premium. The size factor is constructed as the return spread between small- and large-capitalization portfolios:

$$\mathrm{SMB}_t = R_t^{\mathrm{Small}} - R_t^{\mathrm{Big}}. \quad (5)$$

R_t^{Small} and R_t^{Big} are the returns on portfolios of small- and large-capitalization securities; SMB_t is the size premium. The value factor is calculated as the return spread between high-value and low-value portfolios:

$$\mathrm{HML}_t = R_t^{\mathrm{High}} - R_t^{\mathrm{Low}}. \quad (6)$$

R_t^{High} and R_t^{Low} are the returns on portfolios with a high and a low value characteristic; HML_t is the value premium. The momentum factor is calculated as the return spread between winner and loser portfolios formed on prior stock performance:

$$\mathrm{MOM}_t = R_t^{\mathrm{Winner}} - R_t^{\mathrm{Loser}}. \quad (7)$$

R_t^{Winner} and R_t^{Loser} are the returns on portfolios formed on prior stock performance; MOM_t is the momentum premium. The baseline time-series model for individual securities is:

$$R_{i,t} - R_t^f = \alpha_i + \beta_{m,i}\mathrm{MRP}_t + \beta_{s,i}\mathrm{SMB}_t + \beta_{h,i}\mathrm{HML}_t + \beta_{mom,i}\mathrm{MOM}_t + \varepsilon_{i,t}. \quad (8)$$

α_i is the intercept of the regression for security i; $\beta_{m,i}, \beta_{s,i}, \beta_{h,i}, \beta_{mom,i}$ are its sensitivities to the market, size, value and momentum factors; $\varepsilon_{i,t}$ is the residual.

3.2 Market states and predictor set

The forecasting framework relates returns to observable market states. Instead of being added to the model as additional priced factors, inflation, volatility, and drawdown define the market environment in which the underlying return signals are assessed. In general terms the conditional factor relation can be written as:

$$R_{i,t}^{e} = \alpha_i + \sum_{k=1}^{K} \beta_{k,i}(S_t) F_{k,t} + \varepsilon_{i,t}. \quad (9)$$

$F_{k,t}$ is the value of factor k at time t, K the number of factors, and $\beta_{k,i}(S_t)$ the loading of security i on that factor, which depends on the market state S_t. All state variables are observed at the forecast date and therefore contain no future information. Inflation is constructed from the monthly consumer price index and expressed as an annualized continuously compounded rate:

$$\pi_t = 12 \times \ln\left(\frac{\mathrm{CPI}_t}{100}\right). \quad (10)$$

CPI_t is the monthly consumer price index; π_t is inflation expressed as an annualized continuously compounded rate. Its monthly change is:

$$\Delta\pi_t = \pi_t - \pi_{t-1}. \quad (11)$$

A backward-looking expectation is also formed entirely from information available before month t:

$$E_{t-1}(\pi_t) = \frac{1}{K}\sum_{j=1}^{K} \pi_{t-j}\,, \quad (12)$$

$E_{t-1}(\pi_t)$ is the backward-looking inflation expectation, the average over the K preceding months; this K is the length of the averaging window and is unrelated to the number of factors K in (9), which gives the inflation-surprise proxy:

$$\mathrm{Surprise}_t^{\pi} = \pi_t - E_{t-1}(\pi_t) \quad (13)$$

Our construction distinguishes current inflation from a simple backward-looking expectation available to investors prior to *t* realization (Fama and Schwert 1977). Inflation states are defined as rolling 60-month thresholds. The lower and upper quartiles are

$$Q^{\pi}_{0.20,t-1} = q_{0.20}(\pi_{t-60}, \dots, \pi_{t-1}), \tag{14}$$

$q_{0.20}(\cdot)$ and $q_{0.80}(\cdot)$ are the empirical quantiles at levels 0.20 and 0.80 over the past 60 months; $Q^{\pi}_{0.20,t-1}$ and $Q^{\pi}_{0.80,t-1}$ are the resulting thresholds, built only from data available up to month $t-1$.

$$Q^{\pi}_{0.80,t-1} = q_{0.80}(\pi_{t-60}, \dots, \pi_{t-1}) \tag{15}$$

The inflation state is classified as low, medium, or high according to:

$$\text{Regime}^{\pi}_{t} = \begin{cases} \text{Low}, & \pi_t \le Q^{\pi}_{0.20,t-1}, \\ \text{High}, & \pi_t \ge Q^{\pi}_{0.80,t-1}, \\ \text{Medium}, & \text{otherwise.} \end{cases} \tag{16}$$

The corresponding high-inflation indicator, used later as a market-wide predictor, is defined as:

$$S^{\text{infl}}_{t} = \mathbb{1}\{\text{Regime}^{\pi}_{t} = \text{High}\} \tag{17}$$

S^{infl}_{t} is a binary high-inflation indicator, equal to one in months classified in the upper regime; in the data and the code it is the predictor state_infl_high.

Market volatility is estimated from the broad market index return. Realized volatility is calculated over a fixed trailing window and annualized as:

$$\sigma^{\text{mkt}}_{t,L} = \sqrt{12} \times \text{sd}\left(R^{\text{mkt}}_{t-L+1}, \dots, R^{\text{mkt}}_{t}\right), \tag{18}$$

Where $\sigma_{t,L}^{\text{mkt}}$ is realized market volatility over the last L months, annualized by the factor $\sqrt{12}$; it enters the predictor vector as mkt_vol, and L is the lookback window. Following Andersen et al. (2003), realized return variation is used as the volatility proxy. Regimes are assigned by the same 60-month rolling-quantile rule:

$$Q_{0.20,t-1}^{\sigma} = q_{0.20}\left(\sigma_{t-60}^{\text{mkt}}, \dots, \sigma_{t-1}^{\text{mkt}}\right), \tag{19}$$

$$Q_{0.80,t-1}^{\sigma} = q_{0.80}\left(\sigma_{t-60}^{\text{mkt}}, \dots, \sigma_{t-1}^{\text{mkt}}\right). \tag{20}$$

The current value is classified as low, medium, or high relative to the trailing thresholds:

$$\text{Regime}_t^{\sigma} = \begin{cases} \text{Low}, & \sigma_t \le Q_{0.20,t-1}^{\sigma}, \\ \text{High}, & \sigma_t \ge Q_{0.80,t-1}^{\sigma}, \\ \text{Medium}, & \text{otherwise.} \end{cases} \tag{21}$$

Market stress is measured through drawdown of the broad market index. The drawdown in month t is calculated as the percentage decline of the current index level from its previous historical peak:

$$\text{DD}_t = \frac{P_t^{\text{mkt}}}{\max\limits_{s \le t} P_s^{\text{mkt}}} - 1. \tag{22}$$

Where P_t^{mkt} is the level of the market index in month t and $\max_{s \le t} P_s^{\text{mkt}}$ its previous historical peak; DD_t is the drawdown of the index from that peak.

Drawdown measures the cumulative loss from the previous market peak, a dimension of stress not captured by volatility alone (Chekhlov, Uryasev, and Zabarankin, 2005). The binary indicator is:

$$\text{Stress}_t^{\text{DD}} = \begin{cases} 1, & \text{DD}_t \le -0.10, \\ 0, & \text{DD}_t > -0.10. \end{cases} \tag{23}$$

$\text{Stress}_t^{\text{DD}}$ is a binary stress indicator, equal to one when the drawdown reaches at least 10%; in the data and the code it is the predictor state_dd_stress.

10% or more is defined as stress. High inflation and drawdown are included as indicator variables, but volatility enters the vector of forecast in a continuous fashion. The rule-based selector and the three-level volatility state for regime reporting. All expectations, thresholds, volatilities and historical highs are based on historical observations

The last forecasting panel includes 24,636 security-month observations for 290 securities of the Moscow Exchange during the period 31 October 2016 – 30 September 2025. The goal is total return over the next three months:

$$y_{i,t}^{(3)} = \prod_{h=1}^{3}(1 + r_{i,t+h}) - 1 \qquad (24)$$

Where $r_{i,t+h}$ is the monthly return of the security in the h-th month after date t; $y_{i,t}^{(3)}$ is the target variable, the cumulative return over the following three months, capped at the 1st and 99th percentiles of its distribution over the whole panel.

It is observed only when all three subsequent monthly returns are available. Size and value are defined as:

$$s_{i,t} = \ln(M_{i,t}), \qquad v_{i,t} = \ln\left(\frac{B_{i,t}}{M_{i,t}}\right) \qquad (25)$$

Where $s_{i,t}$ is size, the logarithm of market capitalization; $v_{i,t}$ is the value characteristic, the logarithm of the ratio of book equity to market capitalization. $M_{i,t}$ denotes market capitalization and $B_{i,t}$ book equity. Non-positive book-equity observations remain missing and are explicitly flagged. Monthly illiquidity follows an Amihud-type construction:

$$\ell_{i,t} = \frac{1}{D_{i,t}} \sum_{d=1}^{D_{i,t}} \frac{|R_{i,d}|}{V_{i,d}} \qquad (26)$$

Where $R_{i,d}$ is the daily return of the security on day d; $\ell_{i,t}$ is the Amihud-type illiquidity measure, the monthly average ratio of the absolute daily return to daily traded value. $D_{i,t}$ is the number of trading days and $V_{i,d}$ is daily traded value. Size, value, momentum, and illiquidity are transformed into centered within-month ranks. For a generic characteristic x:

$$C_{i,t}(x) = \frac{R_t(x_{i,t})}{N_t^x} - \frac{1}{2}, \tag{27}$$

Where x is any of the ranked characteristics (size, value, momentum, illiquidity); $R_t(x_{i,t})$ is the position of security i when all values observed at date t are ordered from smallest to largest; N_t^x is the number of securities with an observed value; $C_{i,t}(x)$ is the centered within-month rank.

The transformation preserves cross-sectional ordering while reducing sensitivity to differences in scale and extreme values. Two interactions allow characteristic effects to change with the observable states:

$$v_{i,t}^{infl} = v_{i,t}^{cs} S_t^{infl}, \qquad \ell_{i,t}^{stress} = \ell_{i,t}^{cs} S_t^{stress} \tag{28}$$

The superscript cs denotes the centered rank (27) of the corresponding characteristic; $v_{i,t}^{infl}$ and $\ell_{i,t}^{stress}$ are the interaction predictors value_cs_x_inflhigh and illiq_cs_x_stress, equal to the rank in months when the regime is active and to zero otherwise.

Table 1 presents the eleven predictors: four ranked characteristics, continuous volatility, two state indicators, two regime interactions, and two missingness indicators. Preprocessing parameters are estimated within the current training window and then held fixed for validation and test observations.

Table 1. *Predictive variables used in forecasting models*

Feature	Definition
size_cs	Centered cross-sectional rank of company size
value_cs	Centered cross-sectional rank of the value characteristic
mom_cs	Centered cross-sectional rank of the momentum signal
mkt_vol	Market volatility observed at the forecast date
illiq_cs	Centered cross-sectional rank of log-transformed illiquidity
state_infl_high	High-inflation regime indicator
state_dd_stress	Drawdown-stress regime indicator
value_cs_x_inflhigh	Interaction between value_cs and state_infl_high
illiq_cs_x_stress	Interaction between illiq_cs and state_dd_stress

Feature	Definition
value_missing	Indicator equal to one when the value characteristic is unavailable
illiq_missing	Indicator equal to one when the illiquidity characteristic is unavailable

Note. Stock-level ranks are formed within each month. Regime interactions use only state indicators available at the forecast date.

3.3 Forecasting models and rolling estimation

The eight forecasting models are different ways of dealing with instability and not ordered from less complex to more complex specifications. OLS offers a stable linear benchmark. Ridge, lasso and elastic net regularize to curb estimation noise. Rolling OLS is more responsive because it uses a shorter window of history to estimate. The Kalman time varying parameter model allows the coefficients to evolve gradually with the arrival of new observations. CatBoost models nonlinear interactions between contemporaneous and lagged variables. PatchTST directly accounts for temporal dependence through ordered predictor histories. Common target, warm-up of eight quarters, forecast dates, and model classes: Summary Table 2.

Table 2. *Backtesting design and forecasting model classes*

Component	Specification
Target	Three-month forward cumulative return, capped once at the historical 1st/99th percentiles
Test frequency	Quarter-end rebalancing dates
Warm-up period	8 quarters
Validation window	2 quarters
Feature winsorization	1st and 99th percentiles, estimated on training data only
Kalman TVP grid	Process-noise $q \in \{10^{-5}, 10^{-4}, 10^{-3}\}$; initial covariance $P_0 = 10I$
Rolling OLS grid	4, 8, and 12 quarters
Linear models	OLS, ridge, lasso, elastic net
Adaptive linear models	Kalman time-varying-parameter regression; rolling OLS
Nonlinear tabular model	CatBoost with contemporaneous variables and one- to four-quarter lags
Sequential model	Global PatchTST-class model with eight-quarter sequences, patch length 4, and stride 2

Note. CatBoost is tuned over three parameter sets combining depth 4/6, learning rate 0.03/0.05, L2 leaf regularization 3/5/8, and 300/400 iterations. PatchTST is tuned over model dimension 64/96 and two or three layers; the batch size is 256, with 12 epochs and patience of 3.

For all models, the forecasting problem takes common form:

$$\hat{y}_{i,t}^{(m)} = f_m(X_{i,t}; \theta_{m,t}) \tag{29}$$

Where m indexes the eight compared models; $X_{i,t}$ is the vector of the eleven predictors of Table 1; $\theta_{m,t}$ are the parameters estimated only from data dated before t; f_m is the model-specific map and $\hat{y}_{i,t}^{(m)}$ the resulting forecast. Model parameters are estimated using only information dated before the test observation. At rebalancing date t, the expanding training sample is:

$$\Omega_t = \{(i,s): s < t,\ (i,s) \text{ is eligible}\}. \tag{30}$$

Where s is the date index of a training observation; Ω_t is the expanding training window, all security–date pairs dated strictly before the forecast date t.

Preprocessing is re-estimated within every training window. Missing observations are filled with the training sample, features are winsorized using training-only bounds, and scale-sensitive models receive the robustly standardized variable:

$$Z_{i,t,k} = \frac{\text{clip}(\text{fill}(X_{i,t,k}; \mu_{t,k}^0), L_{t,k}, U_{t,k}) - c_{t,k}}{d_{t,k}}. \tag{31}$$

Where k indexes the predictors; $\mu_{t,k}^0$ is the training-window median used to fill missing values; $L_{t,k}$ and $U_{t,k}$ are the 1st and 99th percentiles of the window, the winsorization bounds; $c_{t,k}$ is the median of the winsorized predictor and $d_{t,k}$ its interquartile range; $fill(\cdot; \mu_{t,k}^0)$ substitutes the median for a missing value and $\text{clip}(\cdot, L, U)$ truncates a value to the bounds; $Z_{i,t,k}$ is the resulting robustly standardized predictor.

CatBoost uses the same training-window imputation and winsorization but does not require centering or scaling. The last two pre-test cross-sections form the validation block. Hyperparameters are selected primarily through MSE, with MAE and Spearman correlation used as secondary criteria. The two error measures are:

$$MSE_m(t) = \frac{1}{N_t}\sum_{i\in A_t}\left(y_{i,t} - \hat{y}_{i,t}^m\right)^2, \qquad MAE_m(t) = \frac{1}{N_t}\sum_{i\in A_t}\left|y_{i,t} - \hat{y}_{i,t}^m\right|, \qquad (32)$$

Where A_t is the cross-section of securities with a forecast at date t and N_t their number; $MSE_m(t)$ and $MAE_m(t)$ are the mean squared and the mean absolute forecast error of model m in that quarter.

After model selection, the preferred specification is refitted using all information available before the forecast date. For the linear models, the common prediction equation is (the Kalman model):

$$\hat{y}_{i,t}^m = a_{m,t} + \sum_{k=1}^{11} b_{m,t,k}\, Z_{i,t,k}. \qquad (33)$$

Where $a_{m,t}$ is the intercept, the forecast for an average security; $b_{m,t,k}$ is the coefficient on predictor k, the change in the forecast when the standardized predictor rises by one unit; both are re-estimated at every date t.

Penalized models minimize residual loss together with a regularization term. Ridge is obtained through the quadratic component and searches $\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100\}$; lasso through the absolute-value penalty and elastic through their combination net and use $\lambda \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$, with elastic-net weights $\gamma \in \{0.2, 0.5, 0.8\}$. Rolling OLS estimates the unpenalized objective on the most recent $w \in \{4, 8, 12\}$ quarterly cross-sections:

$$\sum_{(i,s)\in\Omega_t} \varepsilon_{i,s}\,(\theta)^2 + \lambda\left[\gamma\sum_{k=1}^{11}|b_k| + \frac{1-\gamma}{2}\sum_{k=1}^{11} b_k^2\right]. \qquad (34)$$

Where $\varepsilon_{i,s}(\theta)$ is the residual of a candidate parameter set θ on a training observation and b_k its coefficients; λ is the penalty strength and γ the elastic-net weight between lasso ($\gamma = 1$) and ridge ($\gamma = 0$); w in rolling OLS is the number of most recent quarterly cross-sections used.

The Kalman specification retains a linear observation equation but allows the coefficient vector to evolve. Its state transition is:

$$\theta_{K,s} = \theta_{K,s-1} + \eta_s, \qquad \mathrm{Var}(\eta_s) = qI, \tag{35}$$

Where $\theta_{K,s}$ is the vector of Kalman coefficients in quarter s; η_s is the process noise added to them at every step; q is the variance of that drift: the larger it is, the faster the coefficients are allowed to change. The observation equation is:

$$y_s = Z_s\theta_{K,s} + u_s, \qquad \mathrm{Var}(u_s) = \sigma_s^2 I. \tag{36}$$

Where y_s is the column of realized returns of all securities in quarter s; Z_s is the matrix, whose row holds a one for the intercept and the standardized predictors of a security; u_s are the residuals and σ_s^2 their variance, re-estimated in every quarter.

Given posterior covariance P_{s-1}, the prediction and updating steps are:

$$R_s = P_{s-1} + qI, \qquad P_s = \left(R_s^{-1} + \frac{Z_s{}'Z_s}{\sigma_s^2}\right)^{-1}, \tag{37}$$

Where P_s is the posterior covariance of the coefficients once quarter s has been taken into account and R_s the prior covariance before it; the recursion starts from $P_0 = 10I$.

And:

$$\theta_{K,s} = P_s\left(R_s^{-1}\theta_{K,s-1} + \frac{Z_s{}'y_s}{\sigma_s^2}\right). \tag{38}$$

Forecasts at t use coefficients filtered only through $t-1$. The process-noise parameter q controls the speed of coefficient adaptation. CatBoost receives both current predictors and four quarterly lags. If $\tilde{X}_{i,t}$ denotes the imputed and winsorized vector, its input is:

$$W_{i,t} = \left(\tilde{X}'_{i,t}, \tilde{X}'_{i,t-1}, \ldots, \tilde{X}'_{i,t-4}\right)' \in \mathbb{R}^{55}, \qquad \hat{y}_{i,t}^{CB} = G_t\left(W_{i,t}; \eta_{CB,t}\right). \tag{39}$$

Where $\tilde{X}_{i,t}$ is the predictor vector after imputation and winsorization, without centering or scaling; $W_{i,t}$ is the CatBoost input, the current quarter and four lags stacked together, that is $5 \times 11 = 55$ numbers; G_t is the fitted tree ensemble and $\eta_{CB,t}$ its hyperparameters.

CatBoost evaluates three prespecified combinations of depth, learning rate, L^2 leaf regularization, and boosting iterations. The current vector and four quarterly lags make available nonlinear interactions and one-year changes in characteristics without a parametric transition equation.

PatchTST uses a longer ordered history. For security with eight consecutive quarterly observations:

$$S_{i,t} = \left[Z_{i,t-7}, \ldots, Z_{i,t}\right]' \in \mathbb{R}^{8\times 11}. \tag{40}$$

Where $S_{i,t}$ is the PatchTST input, the matrix of a security's predictor history over eight consecutive quarters: eight rows, one per quarter, and eleven columns, one per predictor. The sequence is divided into overlapping patches and passed through a transformer encoder. The full tuning grids for the regularized models, rolling OLS, Kalman TVP, CatBoost, and PatchTST can be reported in the Appendix, while the principal model settings remain summarized in Table 2.

3.4 Reliability-adjusted expected returns

A key feature of the portfolio stage is that raw forecasts are not treated as equally reliable. Each forecast is updated for the recent error history of the security-model pair prior to the optimization. If not enough historical data for a particular security is available, it defaults to the latest aggregate error of the model. So, forecasts that are associated with weaker recent performance are shrunk toward zero before they affect portfolio weights.

Let $\bar{\delta}_{i,t}^{m}$ denote the lagged rolling error measure used for security $i$and model m. Relative forecast instability, the reliability coefficient, and the resulting expected return are:

$$s_{i,t}^{m} = \frac{\bar{\delta}_{i,t}^{m}}{\text{median}_{j \in A_t} \bar{\delta}_{j,t}^{m}}, \qquad q_{i,t}^{m} = \frac{1}{1+0.5 s_{i,t}^{m}}, \qquad \mu_{i,t}^{m} = q_{i,t}^{m} \hat{y}_{i,t}^{m}. \tag{41}$$

Where $\bar{\delta}_{i,t}^{m}$ is the rolling mean absolute error of model m for security i over the four preceding quarters; $s_{i,t}^{m}$ is that error relative to the median error across securities at the date; $q_{i,t}^{m}$ is the reliability coefficient, smaller the worse the recent forecasts; 0.5 is the fixed damping constant; $\mu_{i,t}^{m}$ is the forecast after scaling.

The coefficient satisfies $q_{i,t}^{m} \in (0, 1]$, and the reliability adjustment can attenuate a forecast but cannot amplify it. When no lagged error information exists, the raw forecast is retained. Forecast quality is evaluated with MSE, MAE, and the within-date Spearman rank correlation:

$$\rho_{m,t}^{S} = corr\left(rank_t\left(\hat{y}_{i,t}^{(m)}\right), rank_t\left(y_{i,t}^{(3)}\right)\right) \tag{42}$$

For $rank_t(\cdot)$ is the within-date rank operator; $\rho_{m,t}^{S}$ is the Spearman rank correlation between the forecasts and the realized returns, showing whether the model orders securities correctly. In Equations (32) and (42), N_t is the number of forecasted securities, $y_{i,t}$ is the realized forward-quarter target, and $\text{rank}_t(\cdot)$ is the within-date rank operator. Notation remains fixed from this point: X is the constructed eleven-feature vector, Z its training-window imputed, winsorized, and robust-scaled version, ŷ the raw model forecast, μ the reliability-scaled expected return, and ω the portfolio-weight vector.

3.5 Portfolio construction and performance evaluation

All forecasts are converted into portfolios using the same investment rule. At every quarterly rebalancing date, the 25 securities with the highest reliability-adjusted expected returns form the candidate universe A_t (290 shares in dataset). The portfolio solves:

$$\widehat{\omega}_t = \underset{\omega}{\text{argmax}} \left\{ \omega' \mu_t - \frac{\gamma}{2} \omega' \Sigma_t \omega \right\}, \tag{43}$$

For ω is the vector of portfolio weights and $\widehat{\omega}_t$ its optimal value; μ_t is the vector of expected returns of the selected model; Σ_t is the covariance matrix estimated over the last 24 quarters; γ is the risk-aversion coefficient. Subject to:

$$\sum_{i\in A_t} \omega_{i,t} = 1, \qquad 0 \le \omega_{i,t} \le 0.15, \qquad 5 \le \left|\{i: \omega_{i,t} > 0\}\right| \le 25, \qquad (44)$$

The covariance matrix Σ_t is estimated from the previous 24 quarterly returns and the risk-aversion parameter is set to $\gamma = 5$. The 15% individual weight cap means that in practice a fully invested portfolio requires at least seven positive holdings. The number of securities in the realized portfolios varies from 7 to 12. The weights are held constant at the time of rebalancing and are applied to the actual returns for the next quarter in equation (45).

An important component of our empirical design is a common allocation rule. All strategies use the same universe size candidate, covariance estimator, long only constraint, weight cap, reliability adjustment and optimization procedure. The allocation process is the same in this portfolio comparison, and the performance differences are mainly due to the expected return signals. Fixed weights at date tare applied on realized returns thereafter:

$$R_{p,t+1}^{(s)} = \sum_{i=1}^{N_t} w_{i,t}^{(s)}\, r_{i,t+1} \qquad (45)$$

The superscript (s) indexes the strategy; $w_{i,t}^{(s)}$ is the weight of security i fixed at date t; $r_{i,t+1}$ is its realized return of the following period; $R_{p,t+1}^{(s)}$ is the realized portfolio return.

Portfolio performance is evaluated using return, volatility, downside and tail-risk measures. The principal risk-adjusted and drawdown statistics are:

$$Sharpe_s = \frac{\overline{R}_p^{(s)}}{\sigma_p^{(s)}}, \qquad MaxDD_s = \min_{\tau}\left(\frac{W_{s,\tau}}{\max_{u\le\tau} W_{s,u}} - 1\right) \qquad (46)$$

Where $\overline{R}_p^{(s)}$ is the mean return of strategy s and $\sigma_p^{(s)}$ its volatility, $Sharpe_s$ being their ratio; $W_{s,\tau}$ is the cumulative portfolio value of the strategy at time τ and $MaxDD_s$ its largest decline

from a previous peak. Annualized return and volatility, Sortino and Calmar ratios, and 5% VaR and CVaR are calculated from the same quarterly return series. Sampling uncertainty is assessed using 1,000 moving-block bootstrap replications with a four-quarter block length.

3.6 Model-selection strategies and diagnostic checks

The final component of the design considers the possible advantages of changing the forecasting model over time. We compare four strategies: an always-linear Ridge strategy, an always-transformer PatchTST strategy, a performance-based selector, and an exploratory regime-based selector. The regime rule maps market conditions observed in *t* to a forecasting model:

$$m_t^* = \begin{cases} \text{PatchTST}, & \text{if Stress}_t^{\text{DD}} = 1 \text{ or Regime}_t^{\sigma} = \text{High}, \\ \text{CatBoost}, & \text{if Regime}_t^{\pi} = \text{High}, \\ \text{Linear}, & \text{otherwise.} \end{cases} \tag{47}$$

Where m_t^* is the model assigned by the deterministic regime rule at date t; Linear here denotes the ridge regression used as the linear benchmark.

The rule is economically transparent. The sequence model is assigned for high volatility or drawdown stress periods, CatBoost for high inflation conditions outside those states, and Ridge for the remaining periods. Importantly, this mapping was not locked independently before the portfolio backtest. It is proposed as an exploratory specification, not as a validated investment rule.

The performance-based selector in contrast, considers the last eight quarters. We compute a measure of tail instability as the mean-to-median MSE ratio:

$$TR_m = \frac{\text{mean}_\tau MSE_m(\tau)}{\text{median}_\tau MSE_m(\tau)}. \tag{48}$$

For τ runs over the last eight quarters used in the calculation; TR_m is the ratio of mean to median quarterly MSE: the more it exceeds one, the larger the role of a few particularly bad quarters. After standardizing the available performance measures across models, the composite score is:

$$C_{m,t} = -0.5z_t(MSE_m) + 0.3z_t(\rho_m^S) - 0.2z_t(TR_m), \qquad m_t^C = \underset{m}{\text{argmax}} C_{m,t}. \quad (49)$$

For $z_t(\cdot)$ standardizes statistic across the models available at date t, measuring how much better or worse a model is than the average; $C_{m,t}$ is the composite score and m_t^C the model with the highest score.

The selector favors models with smaller recent forecast error and better cross-sectional ranking but penalizes unstable forecast loss. Otherwise, the full eight-quarter trailing window is used, but if there are at least three observations from the current joint inflation-volatility-drawdown state, the score is based on those comparable quarters. The score is available from all components before the selection decision.

Several diagnostics are added to the selector results. Forecast errors are decomposed by states of inflation and drawdown and portfolio differences are analyzed by moving-block bootstrap inference. The post 2020 matched analysis separates quarters in which the regime rule picks PatchTST from quarters in which it picks some other model. We examine portfolio concentration by the maximum and top five weights, the Herfindahl-Hirschman index, effective holdings, and security-level return contributions. Our approach also considers the robustness of any apparent performance improvements, including their reliance on a small number of securities or time periods. The deterministic regime rule was not pre-specified before the backtest, so in all cases, the results for its portfolio are considered exploratory.

4. Empirical results

4.1 Do more flexible models forecast returns better?

The out-of-sample results provide little support for the view that more flexible models always lead to better return forecasts. Table 3 shows that the best mean MSE is obtained by CatBoost, with 0.0666. Elastic net, lasso, ridge and OLS are very close, with only small differences. Kalman TVP, PatchTST and rolling OLS tend to have considerably larger average forecast errors. This overall ranking is confirmed by Figure 1 for all 8 models.

Table 3. *Out-of-sample forecasting performance by model*

Model	MSE mean	MSE median	MAE mean	MAE median	Spearman mean	Spearman median
CatBoost	0.0666	0.0451	0.1761	0.1519	-0.0165	-0.0006
Elastic net	0.0675	0.0465	0.1774	0.1649	-0.0064	-0.0385
Lasso	0.0676	0.0465	0.1762	0.1625	0.0019	0.0141
Ridge	0.0680	0.0489	0.1796	0.1512	0.0110	0.0121
OLS	0.0684	0.0490	0.1807	0.1487	0.0167	0.0111
Kalman TVP	0.0740	0.0507	0.1929	0.1617	-0.0065	-0.0007
PatchTST	0.0880	0.0466	0.2119	0.1581	-0.0392	-0.0457
Rolling OLS	0.0928	0.0512	0.2177	0.1610	0.0197	0.0313

Source: calculated by authors.

The cross-sectional evaluation does not show the forecast quality as clearly. Mean Spearman correlations are near zero for all models. Rolling OLS has the highest value of 0.0197, OLS has 0.0167 and ridge has 0.0110. CatBoost and PatchTST have negative average correlations -0.0165 and -0.0392 respectively. Figure 2 shows that the ranking ability is weak in general across models. CatBoost beats in MSE, but this does not translate into a stronger ability to identify relative stock performance within each cross-section. Figure 3 provides another view

by comparing mean and median MSE, which shows how unusually large forecast errors in some quarters affect each model.

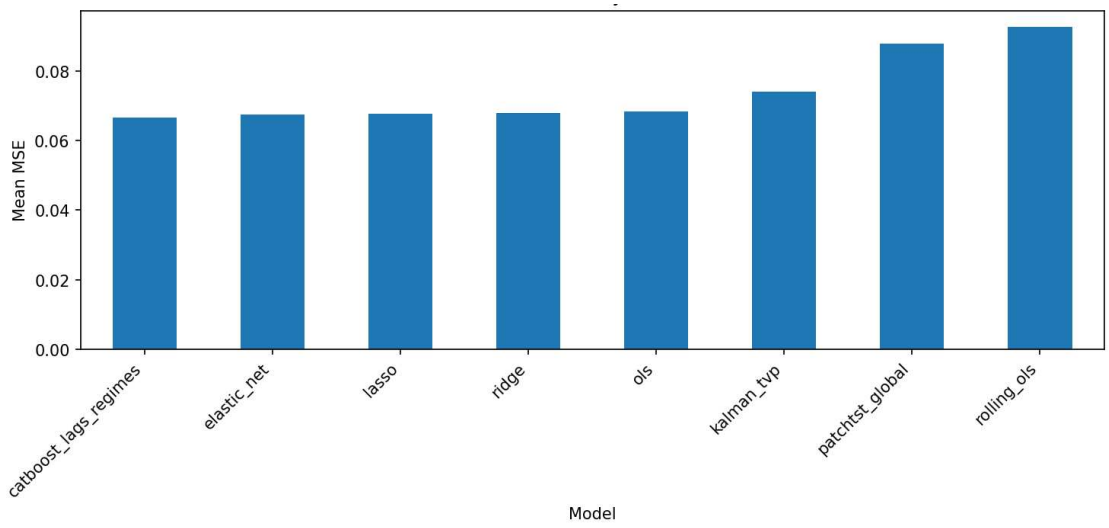


Figure 1. *Mean MSE by forecasting model. Source: compiled by the authors.*

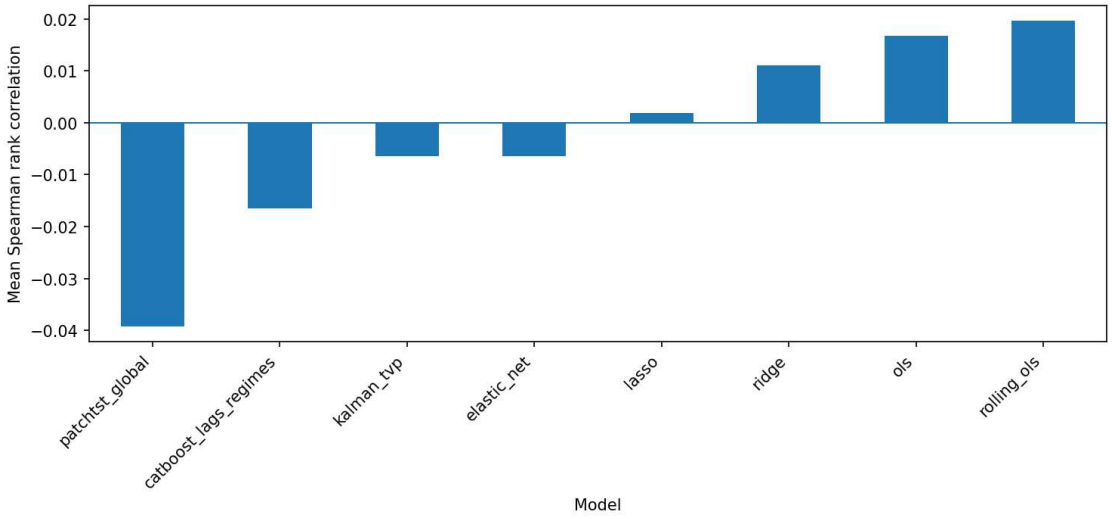


Figure 2. *Mean Spearman rank correlation by forecasting model. Source: compiled by the authors.*

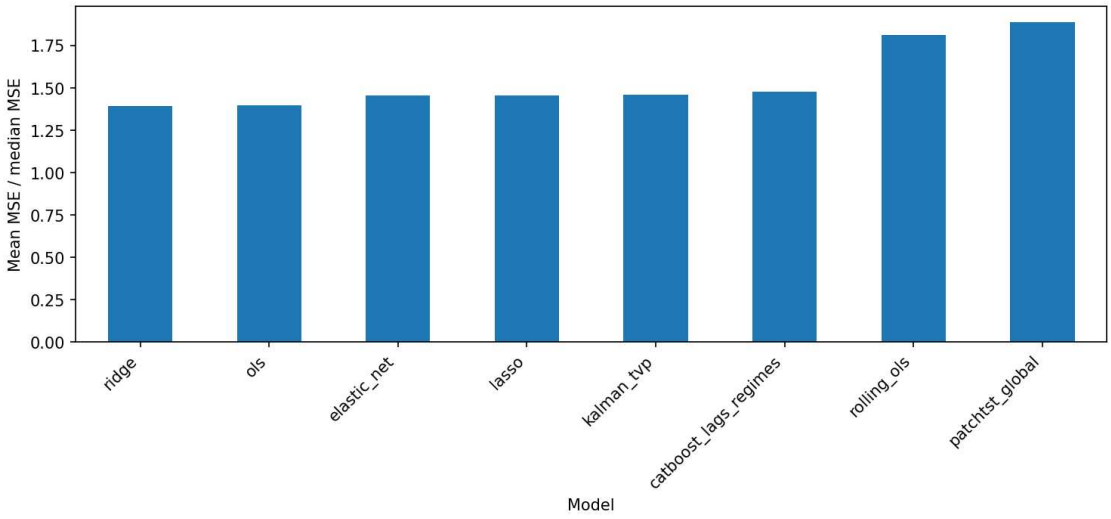


Figure 3. *Mean-to-median MSE ratio by forecasting model. Source: compiled by the authors.*

These differences are more evident with the market conditions. Table 4 shows that forecasting errors increase during periods of high inflation and drawdown stress, although the magnitude of the deterioration varies by model. OLS mean MSE increases from 0.0597 for outside high inflation to 0.0819 in high inflation, ridge increases from 0.0595 to 0.0813. CatBoost is more stable, with MSE increasing from 0.0617 to 0.0741. PatchTST shows a much worse drop from 0.0730 to 0.1112. We see a similar rise in forecasting loss during drawdown-stress periods.

Table 4. *Regime-conditioned forecasting performance*

Model	MSE, low inflation	MSE, high inflation	MSE, no stress	MSE, stress	Spearman, stress
Kalman TVP	0.0726	0.0763	0.0515	0.0886	0.0101
OLS	0.0597	0.0819	0.0510	0.0796	0.0447
Ridge	0.0595	0.0813	0.0506	0.0793	0.0406
Lasso	0.0607	0.0783	0.0485	0.0800	0.0348
Elastic net	0.0604	0.0785	0.0485	0.0798	0.0291
Rolling OLS	0.0990	0.0832	0.0589	0.1147	0.0558
CatBoost	0.0617	0.0741	0.0491	0.0778	-0.0301
PatchTST	0.0730	0.1112	0.0637	0.1037	-0.0202

Note. "Low inflation" denotes observations where the high-inflation indicator equals zero. "No stress" denotes observations where the drawdown-stress indicator equals zero. *Source: calculated by authors.*

The regime analysis shows a more general loss of forecasting power, rather than a systematic switch to another leading model. Quarterly MSE varies quite a bit over time as illustrated in Figure 4 . Figures 5 and 6 compare forecast errors across inflation and drawdown states. Appendix Table A1 provides the full regime decomposition. OLS and ridge maintain small

positive Spearman correlations during stress periods, at 0.0447 and 0.0406 respectively, but these are still economically modest and are associated with higher forecast errors.

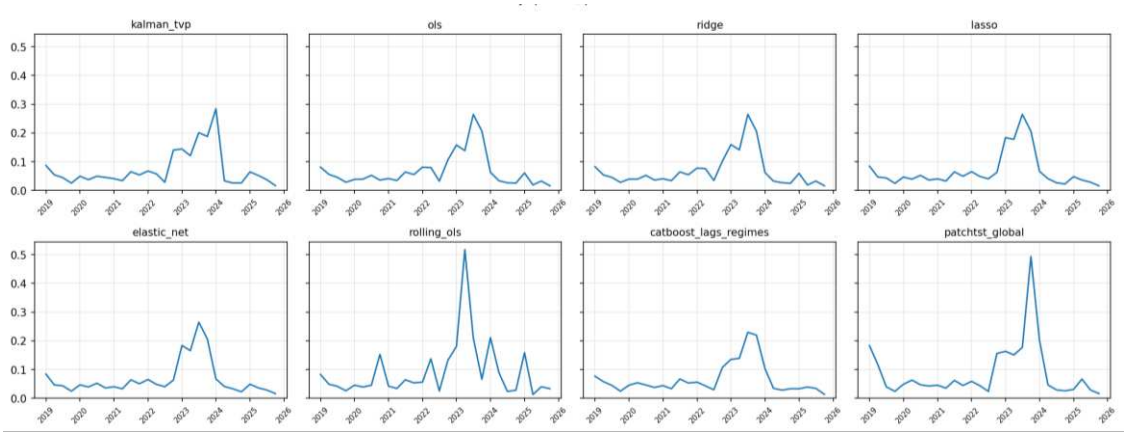


Figure 4. *Quarterly MSE by forecasting model. Source: compiled by the authors.*

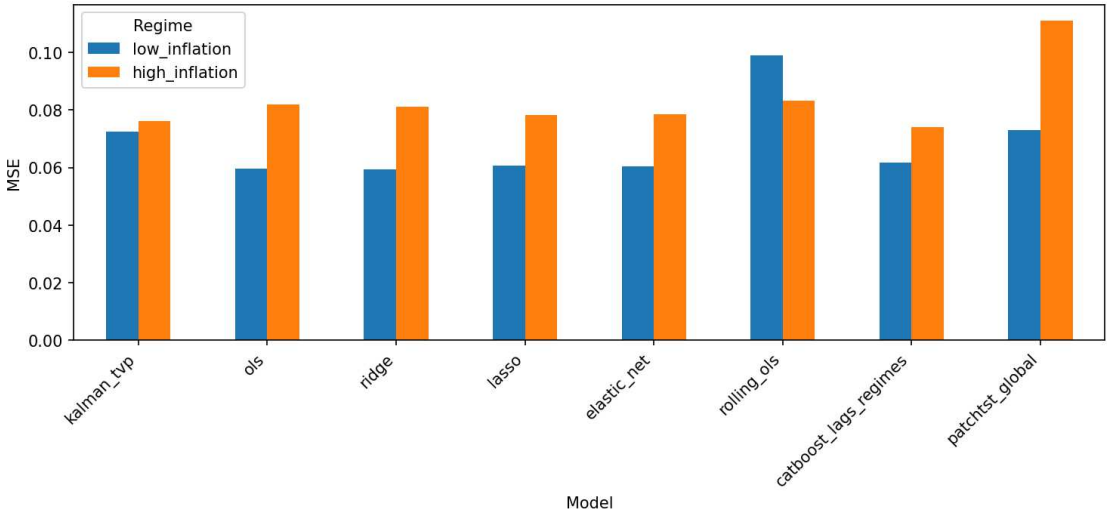


Figure 5. *MSE by inflation regime and forecasting model. Source: compiled by the authors.*

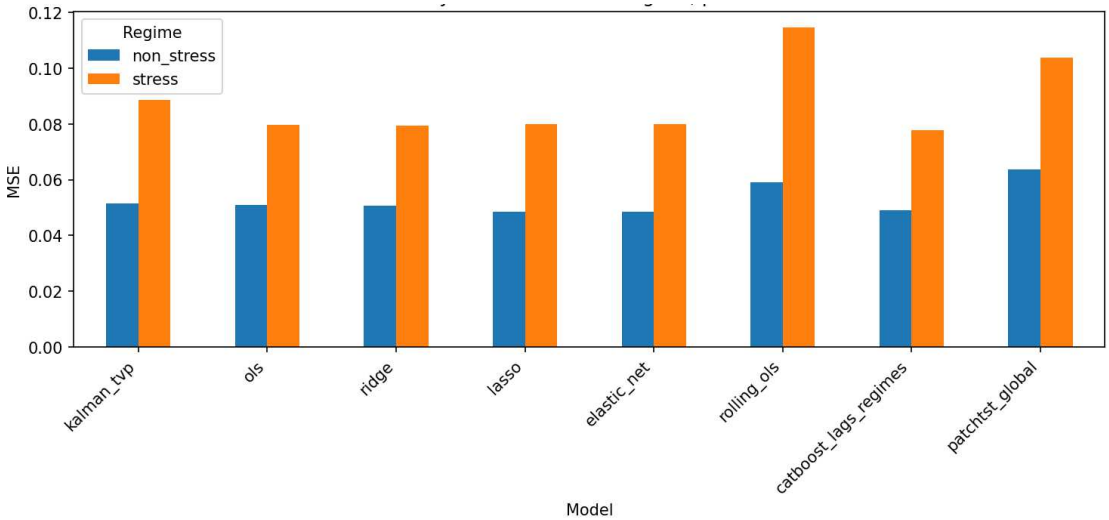


Figure 6. *MSE by drawdown-stress regime and forecasting model. Source: compiled by the authors.*

Feature attribution provides further insight into the sources of these differences between models. In the OLS, the penalized regressions, rolling OLS and Kalman TVP, momentum is the main contributor. As shown in Table 5, CatBoost and PatchTST put more weight on lagged market volatility, and lagged momentum is also one of their top predictors. We find that the more flexible models rely more on past market conditions, whereas the linear specifications rely more on contemporaneous cross-sectional characteristics. Appendix Table A2 reports the top five complete attribution rankings.

Table 5. *Leading forecasting drivers by model*

Model	Rank 1	Rank 2	Rank 3
Kalman TVP	mom_cs (0.0672)	mkt_vol (0.0408)	state_infl_high (0.0187)
OLS	mom_cs (0.0535)	state_infl_high (0.0211)	mkt_vol (0.0189)
Ridge	mom_cs (0.0522)	mkt_vol (0.0187)	state_infl_high (0.0181)
Lasso	mom_cs (0.0261)	mkt_vol (0.0089)	state_dd_stress (0.0066)
Elastic net	mom_cs (0.0260)	mkt_vol (0.0103)	state_infl_high (0.0072)
Rolling OLS	mom_cs (0.1072)	mkt_vol (0.0744)	state_dd_stress (0.0632)
CatBoost	mkt_vol_lag2 (0.0256)	mom_cs (0.0184)	mkt_vol_lag4 (0.0098)
PatchTST	mkt_vol_lag3 (0.0221)	mkt_vol_lag5 (0.0202)	mom_cs_lag4 (0.0182)

Note. Values in parentheses mean absolute attribution values. The full top five ranking is reported in Appendix Table A2. Source: calculated by authors.

The Kalman diagnostics provide a further perspective on the evolution of the model. The estimated coefficients over time are reported in Figure 7 and the frequency of each process noise parameter is reported in Figure 8. The estimated coefficients vary substantially through the sample. This indicates that modifying the specification changes the return relationships as new information becomes available. Despite this flexibility, out of sample forecast errors are larger than those of the more stable regularized models.

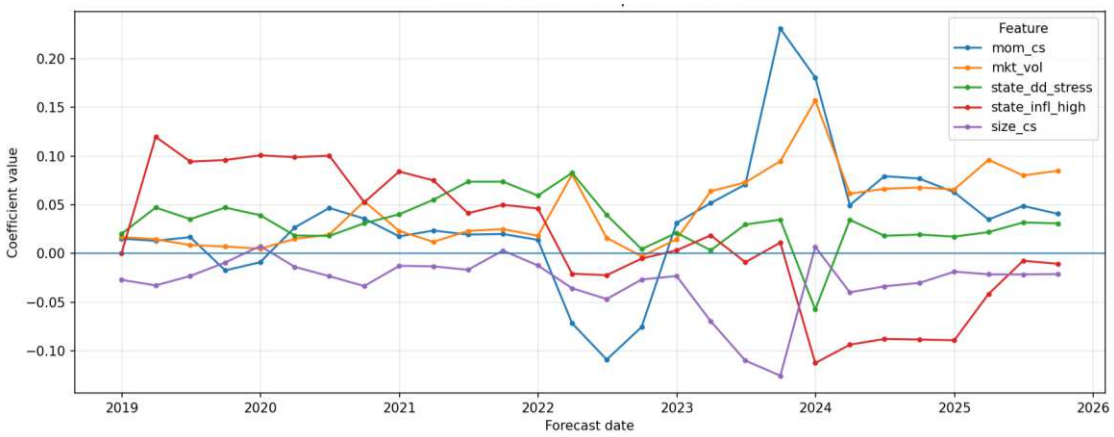


Figure 7. *Kalman TVP coefficient paths over time. Source: compiled by the authors.*

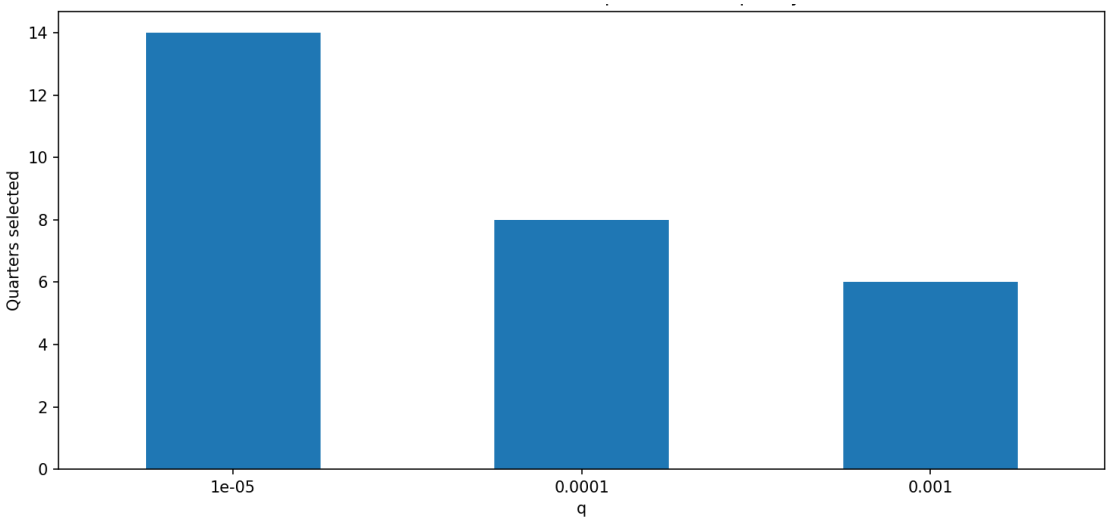


Figure 8. *Frequency of selected Kalman TVP state-noise parameters. Source: compiled by the authors.*

4.2 Do statistical winners become portfolio winners?

The portfolio results show a different model ranking. Table 6 compares the fixed linear ridge strategy, the fixed transformer PatchTST strategy, the performance based composite selector and the exploratory regime-based selector. We can see the contrast with the forecasting results in Table 3. In terms of average forecast error, CatBoost is the best, while PatchTST is one of the worst models for both mean MSE and Spearman correlation. However, the PatchTST portfolio is still economically competitive, indicating that the statistical forecast rankings do not directly translate into portfolio performance.

Table 6. *Portfolio strategy performance*

Strategy	N	Mean return	Annualized return	Annualized vol.	Sharpe	Sortino	Max drawdown	Calmar	VaR 5%	CVaR 5%
Always linear	28	0.0733	0.2785	0.3141	0.8869	4.3450	-0.2450	1.1370	-0.1003	-0.1136
Always transformer	28	0.0654	0.2501	0.2736	0.9141	3.7910	-0.1935	1.2920	-0.0879	-0.0916
Composite selector	28	0.0581	0.2115	0.2845	0.7434	2.3960	-0.3181	0.6649	-0.1267	-0.1529
Rule-based selector	28	0.0837	0.3310	0.3061	1.0820	4.3610	-0.1935	1.7100	-0.0878	-0.0981

Source: calculated by authors.

The fixed linear strategy achieved an annualized return of 27.85%, annualized volatility of 31.41%, and a Sharpe ratio of 0.8869. The fixed transformer portfolio has a slightly lower annualized return of 25.01%, but a lower volatility of 27.36% resulting in a slightly higher Sharpe ratio of 0.9141. Its maximum drawdown is also lower, -19.35% compared to -24.50% for the linear strategy. PatchTST exhibits competitive portfolio-level performance despite relatively weak unconditional forecasting performance.

This difference is important as MSE and portfolio performance measure different aspects of the forecasts. MSE is calculated over the entire set of forecasted securities, including many stocks that never make it into the portfolio. The investment rule, however, only retains the 25 securities with the highest reliability-adjusted expected returns and then finds their weights under common only long and concentration constraints. A model may do badly over much of the cross-section but still be able to pick out a useful group of securities at the top end of the forecast distribution. However, a lower average forecast error is no guaranty that the stocks most relevant for portfolio formation are ranked correctly. Figure 9 shows the accumulation of these differences with time. The strategies do not diverge gradually over the evaluation period with much of the divergence occurring later in the sample. Their Sharpe ratios are compared in Figure 10. Maximum drawdowns are reported in Figure 11. Also shown in Figure 12 are the

annualized portfolio returns in each state of inflation. Our results confirm that the statistical ranking of forecasting models does not translate into the ranking of the portfolios generated from their forecasts.

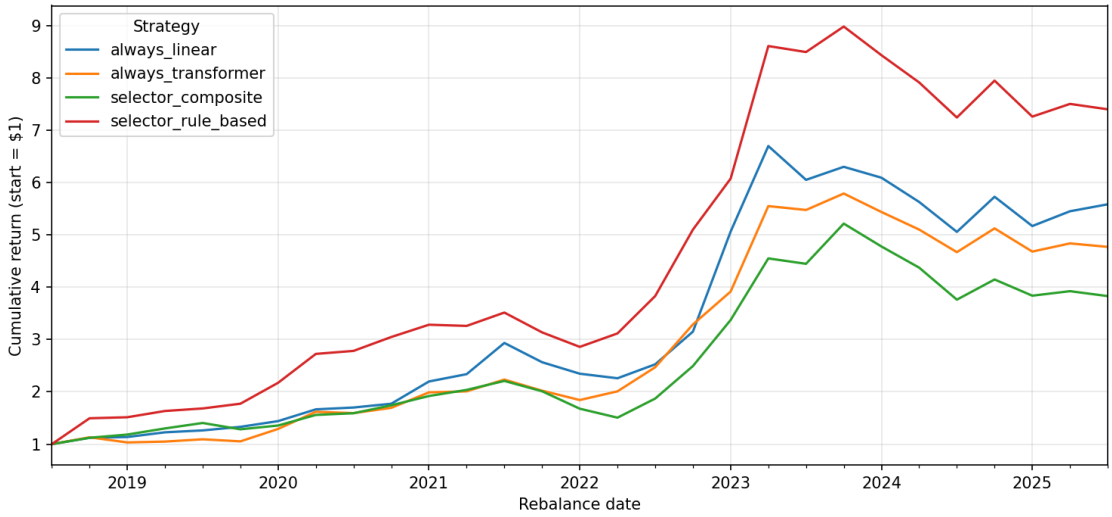


Figure 9. *Cumulative portfolio wealth by strategy. Source: compiled by the authors.*

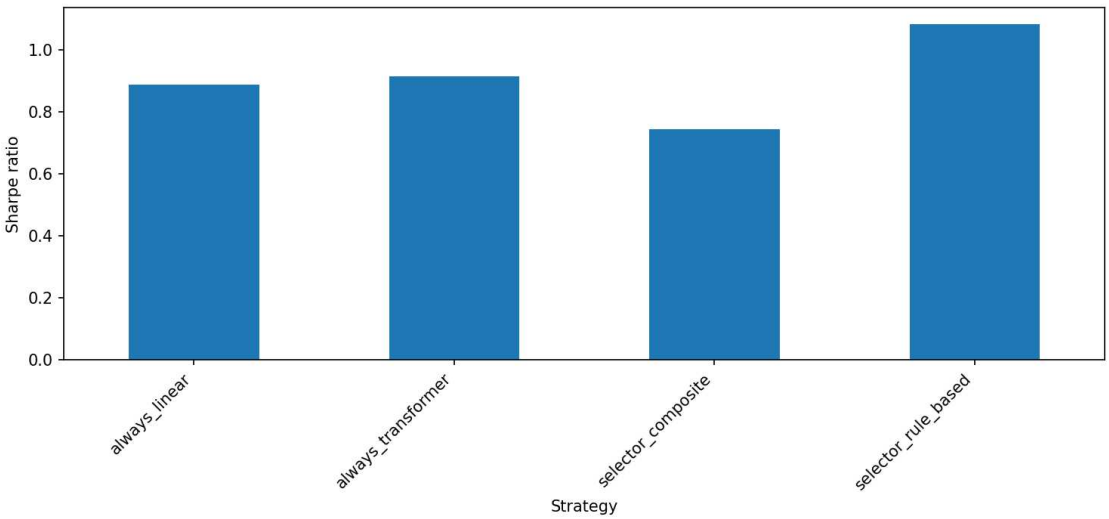


Figure 10. *Sharpe ratio by portfolio strategy. Source: compiled by the authors.*

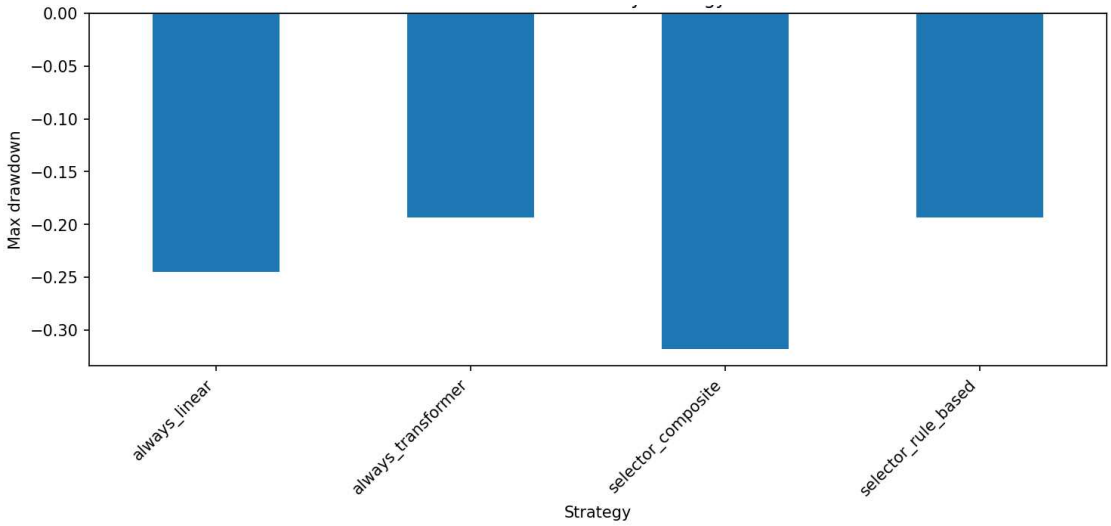


Figure 11. *Maximum drawdown by portfolio strategy. Source: compiled by the authors.*

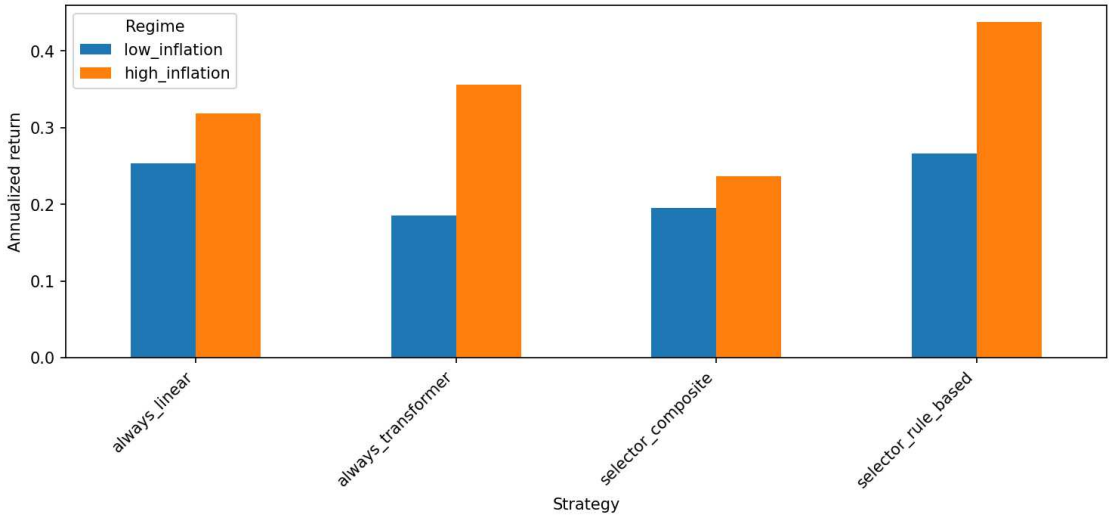


Figure 12. *Annualized return by inflation regime and portfolio strategy. Source: compiled by the authors.*

This finding must be taken with caution. Competitive portfolio performance does not mean PatchTST is a better return forecasting model. It still has a negative unconditional rank correlation and a sharply increasing MSE in more difficult market states. The evidence points to a difference between statistical forecast accuracy and usefulness in an economic sense. Portfolio construction may reward a relatively small fraction of the predicted cross-section that has little impact on an average loss measure. The latter also accounts for the fact that the comparison of the portfolios remains informative, despite the generally weak rank correlations. All strategies have the same reliability adjustment, candidate-universe rule, covariance estimator, optimizer and portfolio constraints. The change in model ranking is not attributable

to differences in portfolio construction. This happens when we filter different expected return signals through the same investment framework.

4.3 Does model switching add portfolio value?

The selector analysis directly tests if shifts in relative model performance lead to better portfolio performance. The regime-based rule seems successful at first. Besides, it presents the highest nominal annualized return of 33.10% and the highest Sharpe ratio of 1.0820 as shown in Table 6. It has a max drawdown of -19.35% which is the same as the fixed transformer strategy. The composite selector does much worse, with an annualized return of 21.15%, a Sharpe ratio of 0.7434 and a maximum drawdown of -31.81%.

These results are achieved with very different choices of models for the two selectors as shown in Table 7 and Figure 13. The regime-rule chooses PatchTST in 17 out of the 28 quarters, CatBoost in six and Ridge in five. The composite selector is rather more diluted and draws from seven models. The most common method is Kalman TVP (eight quarters), and then CatBoost and elastic net (six quarters each). PatchTST is not chosen. The recent MSE, Spearman correlation and loss instability drive the composite selector away from the model favoring the stronger fixed-model portfolio.

Table 7. *Model usage by selector strategies*

Strategy	CatBoost	Elastic net	Kalman TVP	Lasso	OLS	PatchTST	Ridge	Rolling OLS
Composite selector	6	6	8	3	1	0	1	3
Rule-based selector	6	0	0	0	0	17	5	0

Source: calculated by authors.

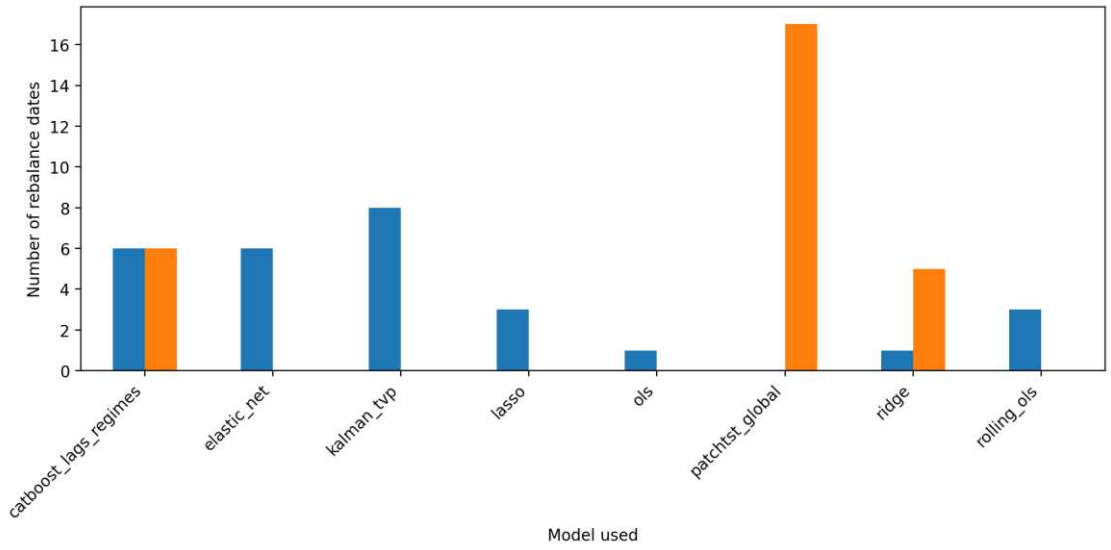


Figure 13. *Model use by selector strategy. Source: compiled by the authors.*

This is an important negative result. A model that has been doing well recently on statistical measures is not a good guide as to which model will generate the best follow-up portfolio. With only eight quarters in the evaluation window, changes in model rankings may reflect noise as well as true changes in forecasting ability. This further reduces the available history by conditioning the comparison on similar market regimes. Thus, in this sample, the additional adaptation introduced by the composite selector does not reduce model uncertainty.

The apparent full-sample superiority of the regime-based rule also loses some of its attractiveness once its origin is examined. In Appendix Table A3, we differentiate the quarters after 2020 in which the selector persists with PatchTST and the quarters in which it switches to a different model. Both strategies agree in 15 out of 19 quarters and differ in only four quarters. The selector has a mean quarterly return of 0.0568 over the entire post-2020 period, and the always-transformer strategy has a mean quarterly return of 0.0652. More importantly, the quarters in which the selector diverges from PatchTST do not lead to better subsequent performance. The full sample benefit of the regime rule cannot therefore be attributed to consistently successful switching.

This is supported by the bootstrap results. Table 8 Moving-block resampling: Comparison of the regime-based rule and the two fixed-model benchmarks. The median differences in annualized return, Sharpe ratio and Calmar ratio generally favor the rule-based strategy but the confidence intervals for these metrics overlap substantially. The overlap ratio for Sharpe ratio

is 0.9443 in relation to the always-linear strategy and 0.9563 in relation to the transformer strategy. The overlap with the transformer benchmark is 0.9795 for the annualized return. Therefore, the evidence is in favor of a nominal performance advantage, but not of a statistically robust or persistent dominance.

Table 8. *Bootstrap interval-overlap diagnostics for the rule-based selector*

Metric	Benchmark	Median difference	CI overlap ratio	Heavy overlap
mean_return	Always linear	0.0054	1.0000	True
annualized_return	Always linear	0.0374	1.0000	True
sharpe	Always linear	0.2187	0.9443	True
calmar	Always linear	0.3708	0.9882	True
cvar_5_return	Always linear	0.0129	0.4852	False
mean_return	Always transformer	0.0085	0.9876	True
annualized_return	Always transformer	0.0402	0.9795	True
sharpe	Always transformer	0.1267	0.9563	True
calmar	Always transformer	0.1561	0.9896	True
cvar_5_return	Always transformer	-0.0064	1.0000	True

Note. Median difference is defined as the rule-based selector minus the benchmark. A high overlap ratio indicates that the bootstrap intervals strongly overlap. *Source: calculated by authors.*

Portfolio concentration is another reason for caution. The weights of individual positions often hit the 15% cap in regime quarters where the rule for the regime selects PatchTST, and the five largest positions make up about three quarters of the total portfolio weight. Full concentration diagnostics are available in Appendix Table A4. The strong portfolio performance may be consistent with useful information in the upper tail of the forecast distribution, but it remains sensitive to the performance of a relatively small number of securities.

At the portfolio level the diagnostics point to the same problem. Figure 14 plots the number of holdings at each rebalance and highlights the limited breadth of the strategies. The associated Herfindahl concentration index is reported in Figure 15 and Figure 16 lists the five securities with the largest cumulative contributions to realized portfolio returns.

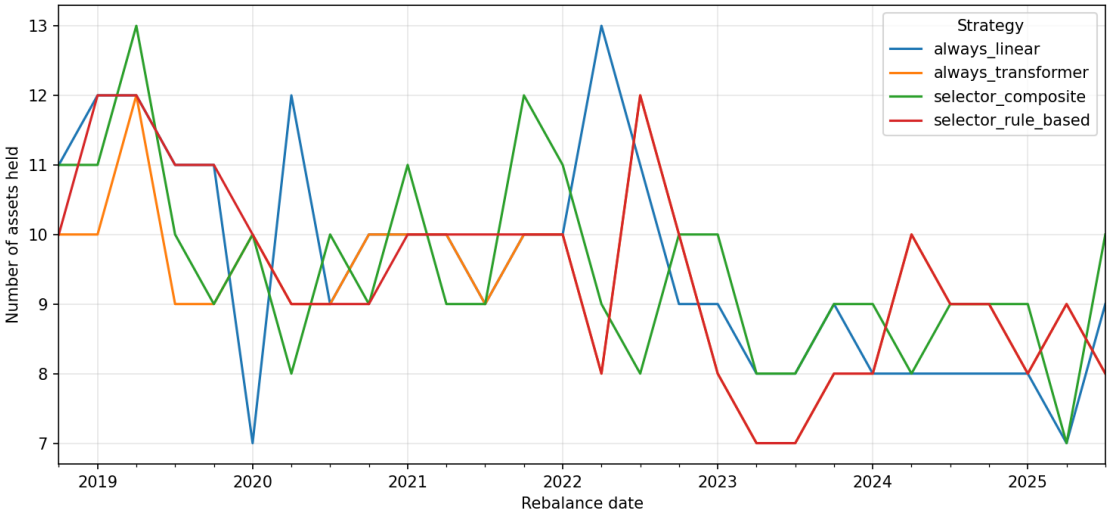


Figure 14. *Number of portfolio holdings by strategy. Source: compiled by the authors.*

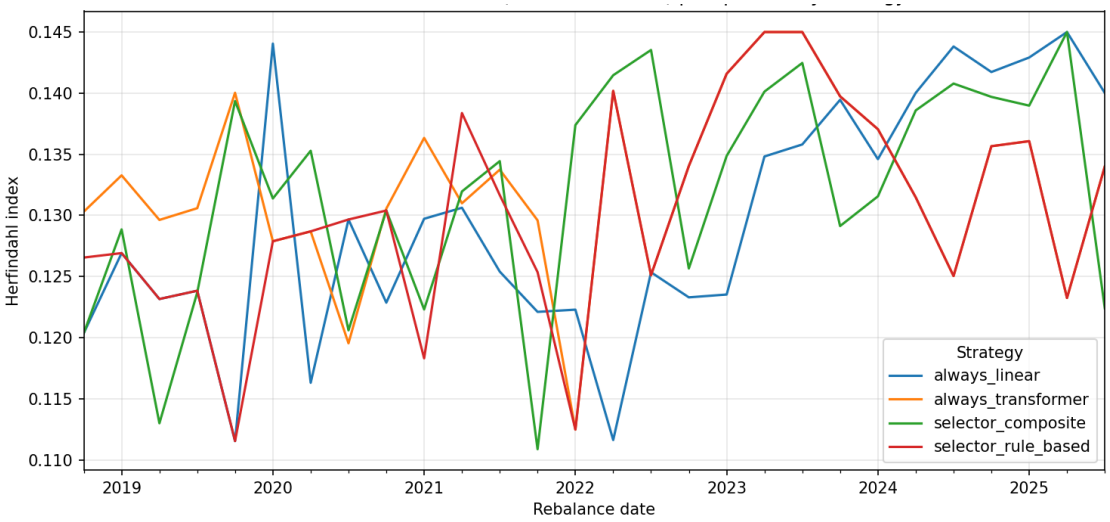


Figure 15. *Portfolio concentration (Herfindahl index) by strategy. Source: compiled by the authors.*

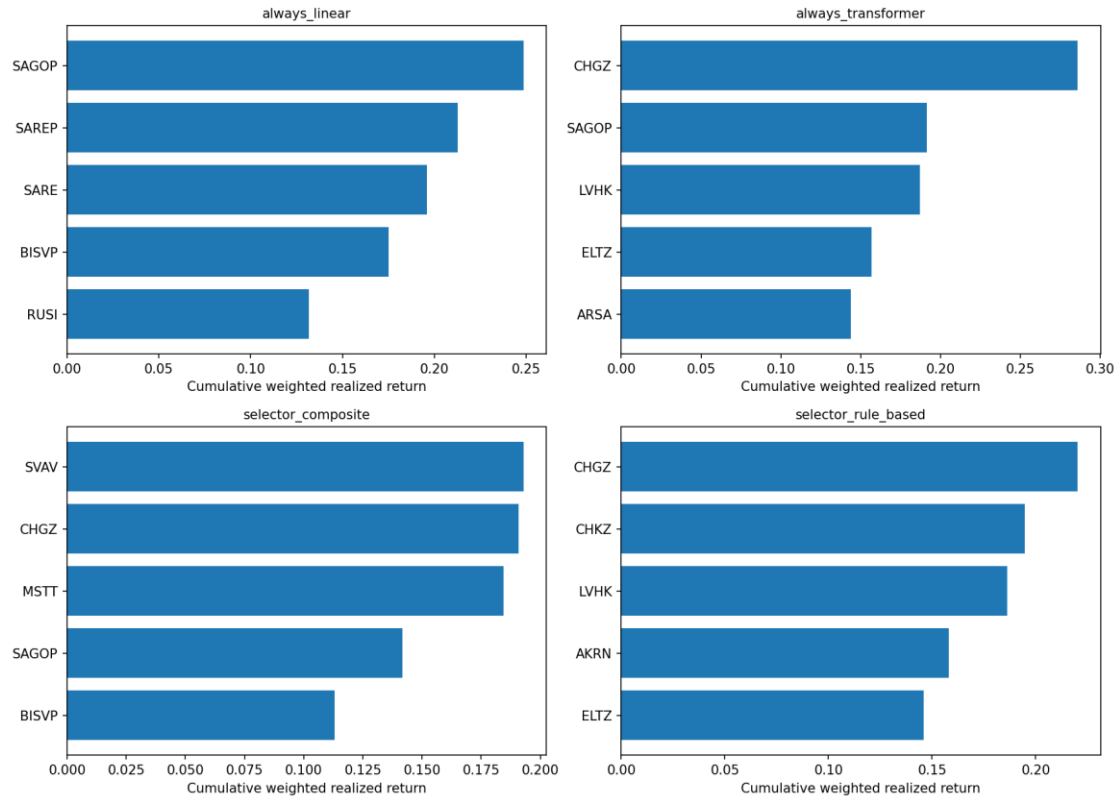


Figure 16. *Top five contributors to total realized return by strategy. Source: compiled by the authors.*

However, there is no evidence from the selector data to support the view that active switching reliably enhances portfolio performance. It is obvious that the most recent performance selector is worse than the fixed-model choices. The regime-based rule gives appealing full-sample results, but a lot of its later performance is due to times when it is simply following PatchTST. The quarters after 2020, where the rule deviates from PatchTST do not result in better performance, and the bootstrap intervals are largely overlapping with those of the fixed model benchmarks. Because regime mapping was not established ex ante, the portfolio results are considered exploratory and do not validate the switching strategy.

The main point of this section is not that one model is identified as superior, but rather the general pattern that is observed across the forecasting and portfolio results. More flexible models do not always lead to better return forecast, statistical rankings do not translate directly to portfolio rankings, and neither recent-performance nor regime-based selection eliminate model uncertainty. The quality of the forecast in portfolio management should be assessed not only in terms of statistical loss, but also in terms of the quality of the resulting signals after security selection, reliability adjustment and portfolio optimization.

5. Discussion

5.1 Forecast-model choice as portfolio model risk

The results suggest that not only parameter uncertainty, which is more familiar in portfolio choice, but also model uncertainty should be considered. The portfolio manager must guess expected returns, within a given specification. But there is a prior question: is that specification appropriate for the market environment in which it is being used? The distinction may be less important in relatively stable circumstances. But if inflation, volatility, liquidity and market stress change materially, uncertainty about the forecasting model itself is part of the investment problem.

This interpretation is consistent with recent research showing that flexible return models can extract economically useful information if proper complexity control is applied. Kelly, Malamud, and Zhou (2024) find that more highly parameterized models can outperform less parameterized specifications if regularization is done well. Chen, Pelger, and Zhu (2024) demonstrate that deep-learning models can take advantage of conditioning information and time variation when the economic structure is embedded in the model. Our results do not contradict these findings but qualify them in a smaller and less stable emerging market context. CatBoost has the lowest mean forecast error, but the additional flexibility is not always rewarded. Kalman TVP, rolling OLS, and PatchTST do not beat regularized linear models on standard forecasting metrics. Complexity can be helpful, but it is not necessarily better.

Our contribution is to apply this argument directly to portfolio choice. The models discussed here reflect different assumptions about persistence and adaptation. Shrinkage stabilizes coefficient estimates. Rolling estimation puts more emphasis on recent observations. Kalman specification allows for smooth evolution of coefficients. CatBoost captures nonlinear interactions. PatchTST emphasizes more on temporal structure. Portfolio choice depends not only on the expected return estimates but also on the information processing rule through which they are generated. The relative performance of these rules' changes over time, and the forecasting model choice is a source of portfolio model risk.

This view is closely related to recent work by Gehrig, Sögner, and Westerkamp (2025) who motivate shrinkage as a response to parameter and model uncertainty in portfolio choice and show that unconstrained optimal policies can be unstable out of sample. Their analysis concerns the uncertainty of portfolio policies. Our results suggest an associated source of uncertainty

earlier in the investment process: the model used to generate expected-return inputs may be unstable itself. This is partially explained by the reliability adjustment, which shrinks forecasts toward zero if they have performed poorly in the recent past for a given asset and model. But it doesn't answer the separate question of how appropriate the forecasting architecture is to the current market conditions.

This identification is easier in observable regimes, but it is not solved by them. Forecast errors are almost uniformly increased in high inflation and drawdown stress periods for all models. CatBoost drops less than several linear models in high inflation. PatchTST drops more sharply in terms of MSE. But once these forecasts are translated into positions the ranking of the portfolio changes again. The evidence does not indicate that one architecture is intrinsically better suited for stressed markets. Adverse market conditions seem rather to create uncertainty about expected returns and the models used to estimate these. This interpretation is consistent with the conditional asset pricing and adaptive market hypothesis as the change in the market environment can influence the return relationships and the relative reliability of competing forecasting models.

Further support for this interpretation is provided by the model selection results. The composite selector does not improve portfolio performance in trying to identify the currently best forecasting model but uses only information available at the decision date. Model selection introduces a second estimation problem. One must estimate expected returns and, from a brief performance history, infer the relative quality of the models that generated those estimates. Uncertainty can only be reduced by model switching, if the switching rule is accurately enough estimated.

5.2 Why does forecasting accuracy and portfolio value diverge?

Our most significant empirical result is the difference between the statistical and economic ranking of the models. CatBoost has the lowest average MSE, but PatchTST has relatively poor results on both MSE and unconditional cross sectional rank correlation. Nevertheless, the PatchTST portfolio is superior to the regularized linear benchmark in terms of higher Sharpe ratio and lower maximum drawdown. We find a model can be poor in terms of standard forecast diagnostics and be useful after its signals are mapped into portfolio positions. This is because MSE and portfolio performance measure different aspects of the forecasting problem. The MSE weights each forecast in the cross section. The portfolio does not. First, it applies reliability

adjustment and retains only 25 securities with the highest adjusted expected returns. Then, it applies a constrained mean-variance weighting rule. MSE is affected by forecast errors for securities that never enter the candidate set but not realized portfolio performance directly. The portfolio value depends on a smaller part of the forecasting distribution, and on the interaction of expected returns with covariances and portfolio constraints.

This distinction is increasingly stressed in literature. Avramov, Cheng, and Metzker (2023) show that realistic investment constraints and trading costs may alter the economic value of machine-learning signals. A similar conclusion is reached by Rigamonti (2024), who finds that predictive gains do not necessarily translate into strategies that outperform simpler alternatives once implementation considerations are considered. Our results point in the same direction: forecast performance should be evaluated in relation to the portfolio decision that it is meant to support. The results are also in line with the arguments developed by Kelly, Malamud and Zhou (2024). They find that the model complexity could improve prediction and portfolio performance when the regularization is effective. Our evidence offers a different view. The value of complexity may depend on the context of its judgment. If portfolio formation is mainly based on a relatively small part of the cross-section, a flexible model may not dominate the average forecast loss and be economically competitive.

A similar conclusion is reached in Bryzgalova, Pelger and Zhu (2025) who show that structured portfolios based on firm characteristics can outperform a number of alternatives constructed from machine learning predictions. Their results highlight the role of both predictive flexibility and the transformation of information into investable positions for economic value. Our study reaches a similar conclusion with a different design: even with the portfolio rule fixed, statistical superiority at the forecasting stage does not imply superior portfolio performance.

Concentration diagnostics should be handled with care. The PatchTST related portfolios usually have at least one security at the 15% cap, and the five largest positions account for a large portion of the portfolio weight and realized contribution. The evidence suggests that useful information could be concentrated among securities at the top of the reliability adjusted ranking. But that is not to say that PatchTST always gives a better picture of what is happening with these stocks. The portfolio sample only has 28 quarters so a few securities with strong results could produce the same pattern.

Forecasts also show up as realized returns when considering portfolio constraints and reliability adjustment. The 15% weight cap constrains the impact of extreme expected return signals, and reliability scaling mitigates the impact of forecasts that have exhibited poor performance for a given security and model. The covariance matrix can also down-weight a high predicted return stock if it contributes disproportionately to the risk of the portfolio. Portfolio formation is more than ranking by predicted return. Expected returns interact with forecast reliability, covariances and portfolio constraints to determine final positions.

Control mapping is also an important feature of our investigation. All forecasting models share the same reliability adjustment, top 25 candidate set, covariance estimator, long-only restriction, 15% cap, and mean-variance objective. The differences in portfolio outcomes can be more easily attributed to the expected return signals themselves. In this manner, our study separates uncertainty related to the forecasting model from differences in portfolio construction and offers insights into the impact of model choice on realized investment performance.

5.3 Implications for portfolio managers

For portfolio management, the first implication is that the choice of return model should not be based on MSE alone. CatBoost has the best average squared error performance, but this does not directly translate into portfolio results. In contrast, PatchTST looks unattractive if evaluation is stopped at unconditional forecast statistics, but its portfolio remains competitive. The forecasting loss is useful to evaluate predictive accuracy, but it does not answer the same question as portfolio performance.

A second implication is that forecasting models should be tested through the portfolio rule in which they will be used. Security selection, portfolio constraints, estimation of covariance, risk aversion, reliability adjustment and turnover and trading costs affect the translation of a forecast into an investment position. For example, Huang et al. (2025) question the classical distinction between prediction and optimization and develop a framework that more directly targets the investment objective. We employ a two-stage forecast and optimization framework. This allows us to disentangle the sources of portfolio performance. Forecasting models should be evaluated in terms of the investment decision they are supposed to support.

The third implication has to do with model switching. Recent model rankings should be considered noisy signals rather than strong evidence of persistent dominance. The composite

selector shows this situation. The selector combines recent MSE, rank correlation and tail loss using only information available at the time but still underperforms both fixed model benchmarks. The selector only has eight quarters of history and needs to find real changes in prediction ability from normal sampling variation. The effective sample is further reduced by conditioning on the current joint state of the market. Short term performance can lead to frequent model changes that add further estimation noise, rather than reduce uncertainty.

This may explain why model averaging and shrinkage are still hard standards to beat. A selector assigns to the model that looks strongest at a given point in time, while forecasting combinations spread risk over a number of imperfect models. A choice that is only worth it when relative model performance stands. This interpretation is consistent with the forecast combination literature (Timmermann 2006; Rapach, Strauss, and Zhou 2010) and with the robustness and shrinkage literature in portfolio choice. For example, Gehrig, Sögner, and Westerkamp (2025) observe that reducing an estimated portfolio policy toward a more stable benchmark can lead to out of sample performance.

A fourth implication is about the role of regime information. Market states may be more useful for identifying periods of high model risk than for selecting specific forecasting architecture. Higher inflation and drawdown stress are associated with larger forecast errors for most models. Such information helps portfolio managers identify periods during which they should be more cautious in estimating expected returns. The response could be stronger contraction, tighter position limits, or less active exposure. Nor does our evidence support a mapping from each market regime to a single forecasting model.

The deterministic regime rule illustrates this distinction. It has the highest full sample annualized return and Sharpe ratio. However, the post-2020 analysis demonstrates that this edge is not due to removing PatchTST. Its bootstrap intervals also overlap considerably with those of both fixed model benchmarks. Our results indicate a possible relation between market states and model performance that needs to be confirmed independently, as the rule was not pre-specified before the portfolio backtest. These are not dependable trading rule.

The distinction is practically important, too. A high volatility state should not be automatically interpreted as a signal to adopt one specific forecasting model. Instead, it could suggest that the expected return estimates of all models are less reliable. A portfolio manager may react by

having a stronger forecast shrinkage, tighter exposure limits, more model averaging or a higher threshold for changing the forecasting architecture.

The theoretical contribution is not a new factor or a claim of superiority of one machine learning architecture over another. The contribution is to consider forecast model choice as a component of portfolio model risk. Conditional asset pricing shows that expected returns may depend on the economic environment. Machine Learning research shows that different architectures learn different forms of non-linearity and time variation. Portfolio theory has long stressed the consequences of estimation error. Our evidence ties these ideas together. Changes in the return environment led to uncertainty in the estimated returns of a model and the choice of model to generate them. Another estimation problem is the choice of models and may thus lead to greater uncertainty in the portfolio.

The empirical contribution is the investigation of this issue within a common forecast to portfolio framework. Recent work documents the promise of high dimensional return prediction (Chen, Pelger, and Zhu 2024; Kelly, Malamud, and Zhou 2024) and the value of more structured and interpretable cross-sectional representations (Bryzgalova, Pelger, and Zhu 2025). We relate these two views. More flexible models may enhance some aspects of forecasting but the value of such enhancements for investment depends on which part of the forecast distribution is relevant for portfolio formation and how the signals generated from the forecast are converted into positions.

Several limitations circumscribe the scope of the results. The portfolio analysis is restricted to 28 quarters, which reduces the number of independent inflation and stress episodes. Our study is also limited to a single emerging equity market, and changes in the composition of the Moscow Exchange may increase the importance of individual securities. Portfolio concentration also restricts further inference of general cross sectional predictive ability. The fixed cap on the full-panel forecast target should be compared to a version estimated separately within each training window. The economic ranking can change with alternative covariance estimators and portfolio objectives even if the forecasts are the same.

These constraints offer clear avenues for future research. The rule of regime should be pre-specified and tested on an independent market or a truly untouched holdout period. The analysis can be extended to other emerging markets to help distinguish features specific to the Moscow

Exchange and a more general relation between macroeconomic instability and forecast model risk. Our evidence at this point suggests that model selection is part of the portfolio decision itself and that neither lower statistical error nor more frequent adaptation is enough to demonstrate superior investment value.

6. Conclusion

Our paper investigates the impact of forecast flexibility and model switching on portfolio decisions in an emerging market with evolving economic conditions. We compare eight forecasting models on the Moscow Exchange with the same information set, rolling out of sample procedure, reliability adjustment and constrained mean variance allocation rule. Fixing the portfolio construction process we can attribute the differences in performance mainly to the expected return forecasts.

There is no dominant model for all evaluation criteria. Catboost has the lowest average MSE, but cross-sectional rank correlations are weak for all models. When forecasts are transformed to portfolios, the ranking changes. Although the unconditional forecast metrics indicate that the PatchTST has a relatively weak performance, the risk adjusted portfolio performance is close to the regularized linear benchmark and has a smaller maximum drawdown. Thus, statistical forecast accuracy and portfolio value are two different ways to measure the performance of the model.

This difference does not go away by changing models over time. The selector based on recent forecast performance is worse than the fixed model strategies. The regime rule has the highest nominal return and Sharpe ratio. However, the bootstrap intervals for the regime rule overlap substantially with the fixed benchmarks. Neither do its post-2020 switches result in consistent improvements. As the rule was not specified before the portfolio backtest these results should be viewed as exploratory and not validation of an investment strategy.

A more general implication is that portfolio managers are uncertain on two levels. The first concerns estimates of expected return in each model. The second is about the choice of the forecasting model itself. Observable market regimes can identify periods when forecasts are more unreliable, but do not give a clear option for which forecasting architecture to choose. Model choice is an additional element of portfolio model risk.

References

Acharya, V. V., & Pedersen, L. H. (2005). Asset pricing with liquidity risk. Journal of Financial Economics, 77(2), 375–410. https://doi.org/10.1016/j.jfineco.2004.06.007

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets, 5(1), 31–56. https://doi.org/10.1016/S1386-4181(01)00024-6

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. Journal of Financial Economics, 17(2), 223–249. https://doi.org/10.1016/0304-405X(86)90065-6

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica, 71(2), 579-625. https://doi.org/10.1111/1468-0262.00418

Ang, A., & Bekaert, G. (2002). International asset allocation with regime shifts. Review of Financial Studies, 15(4), 1137–1187. https://doi.org/10.1093/rfs/15.4.1137

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. https://doi.org/10.1111/j.1540-6261.2006.00836.x

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. Journal of Finance, 68(3), 929–985.

Avramov, D., Cheng, S., & Metzker, L. (2023). Machine learning vs. economic restrictions: Evidence from stock return predictability. Management Science, 69(5), 2587–2619.

Banz, R. W. (1981). The relationship between return and market value of common stocks. Journal of Financial Economics, 9(1), 3-18. https://doi.org/10.1016/0304-405X(81)90018-0

Bekaert, G., & Wang, X. (2010). Inflation risk and the inflation risk premium. Economic Policy, 25(64), 755–806.

Bekaert, G., Harvey, C. R., & Lundblad, C. (2007). Liquidity and expected returns: Lessons from emerging markets. The Review of Financial Studies, 20(6), 1783–1831. https://doi.org/10.1093/rfs/hhm030

Boudoukh, J., & Richardson, M. (1993). Stock returns and inflation: A long-horizon perspective. American Economic Review, 83(5), 1346–1355.

Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. The Review of Financial Studies, 22(6), 2201–2238. https://doi.org/10.1093/rfs/hhn098

Bryzgalova, S., Pelger, M., & Zhu, J. (2025). Forest through the trees: Building cross-sections of stock returns. The Journal of Finance, 80(5), 2447–2506. https://doi.org/10.1111/jofi.13477

Cakici, N., Fabozzi, F. J., & Tan, S. (2013). Size, value, and momentum in emerging market stock returns. Emerging Markets Review, 16, 46–65.

Campbell, J. Y., & Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. Journal of Financial Economics, 31(3), 281–318. https://doi.org/10.1016/0304-405X(92)90037-X

Campbell, J. Y., Giglio, S., Polk, C., & Turley, R. (2018). An intertemporal CAPM with stochastic volatility. Journal of Financial Economics, 128(2), 207–233.

Carhart, M. M. (1997). On persistence in mutual fund performance. The Journal of Finance, 52(1), 57-82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Cenesizoglu, T., & Timmermann, A. (2012). Do return prediction models add economic value? Journal of Banking and Finance, 36(11), 2974–2987.

Chekhlov, A., Uryasev, S., & Zabarankin, M. (2005). Drawdown measure in portfolio optimization. International Journal of Theoretical and Applied Finance, 8(1), 13-58. https://doi.org/10.1142/S0219024905002767

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

Cochrane, J. H. (2005). Asset pricing (Rev. ed.). Princeton University Press.

Cooper, M. J., Gutierrez, R. C., Jr., & Hameed, A. (2004). Market states and momentum. Journal of Finance, 59(3), 1345–1365.

Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. Journal of Financial Economics, 122(2), 221–247.

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

Fama, E. F. (1981). Stock returns, real activity, inflation, and money. American Economic Review, 71(4), 545–565.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. The Journal of Finance, 47(2), 427-465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3-56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics, 116(1), 1–22.

Fama, E. F., & Schwert, G. W. (1977). Asset returns and inflation. Journal of Financial Economics, 5(2), 115-146. https://doi.org/10.1016/0304-405X(77)90014-9

Ferson, W. E., & Harvey, C. R. (1991). The variation of economic risk premiums. Journal of Political Economy, 99(2), 385–415. https://doi.org/10.1086/261755

Ferson, W. E., & Harvey, C. R. (1993). The risk and predictability of international equity returns. Review of Financial Studies, 6(3), 527–566. https://doi.org/10.1093/rfs/6.3.527

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

Garlappi, L., Uppal, R., & Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. Review of Financial Studies, 20(1), 41–81.

Gehrig, T., Sögner, L., & Westerkamp, A. (2025). Extending the demand system approach to asset pricing. Financial Markets and Portfolio Management, 39(1), 133–166.

Goriaev, A., & Zabotkin, A. (2006). Risks of investing in the Russian stock market: Lessons of the first decade. Emerging Markets Review, 7(4), 380–397.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. Review of Financial Studies, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Guidolin, M., & Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. Review of Financial Studies, 21(2), 889–935. https://doi.org/10.1093/rfs/hhn006

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, 57(2), 357–384. https://doi.org/10.2307/1912559

Harvey, C. R. (1989). Time-varying conditional covariances in tests of asset pricing models. Journal of Financial Economics, 24(2), 289–317. https://doi.org/10.1016/0304-405X(89)90049-4

Henkel, S. J., Martin, J. S., & Nardari, F. (2011). Time-varying short-horizon predictability. Journal of Financial Economics, 99(3), 560–580.

Hou, K., Karolyi, G. A., & Kho, B. C. (2011). What factors drive global stock returns? Review of Financial Studies, 24(8), 2527–2574.

Huang, X., Newton, D. P., Platanakis, E., & Sutcliffe, C. (2025). Single-stage portfolio optimization with automated machine learning for M6. International Journal of Forecasting, 41(4), 1450–1460.

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. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Katz, M., Lustig, H., & Nielsen, L. (2017). Are stocks real assets? Sticky discount rates in stock markets. Review of Financial Studies, 30(2), 539–587.

Kelly, B. T., Malamud, S., & Zhou, K. (2024). The virtue of complexity in return prediction. Journal of Finance, 79(1), 459–503.

Kozak, S., Nagel, S., & Santosh, S. (2018). Interpreting factor models. The Journal of Finance, 73(3), 1183–1223. https://doi.org/10.1111/jofi.12612

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

Lettau, M., & Ludvigson, S. (2001). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. Journal of Political Economy, 109(6), 1238–1287. https://doi.org/10.1086/323282

Lewellen, J., & Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics, 82(2), 289–314.

Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. The Journal of Portfolio Management, 30(5), 15–29. https://doi.org/10.3905/jpm.2004.442611

Lo, A. W. (2017). Adaptive markets: Financial evolution at the speed of thought. Princeton University Press.

Merton, R. C. (1973). An intertemporal capital asset pricing model. Econometrica, 41(5), 867–887. https://doi.org/10.2307/1913811

Nagel, S. (2021). Machine learning in asset pricing. Princeton University Press.

Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. In Proceedings of the Eleventh International Conference on Learning Representations.

Pástor, L., & Stambaugh, R. F. (2001). The equity premium and structural breaks. Journal of Finance, 56(4), 1207–1239.

Pástor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. Journal of Political Economy, 111(3), 642–685. https://doi.org/10.1086/374184

Rapach, D. E., & Zhou, G. (2013). Forecasting stock returns. In G. Elliott & A. Timmermann (Eds.), Handbook of Economic Forecasting (Vol. 2A, pp. 328–383). Elsevier.

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

Rigamonti, A. (2024). Can machine learning make technical analysis work? Financial Markets and Portfolio Management, 38(3), 399–412.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), Handbook of Economic Forecasting (Vol. 1, pp. 135–196). Elsevier.

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

Appendix

Appendix Table A1. *Forecasting performance by inflation and drawdown regime*

Model	Regime	State	MSE	MAE	Spearman
Kalman TVP	high inflation	0	0.0726	0.1897	-0.0346
Kalman TVP	high inflation	1	0.0763	0.1978	0.0370
Kalman TVP	drawdown stress	0	0.0515	0.1577	-0.0321
Kalman TVP	drawdown stress	1	0.0886	0.2157	0.0101
OLS	high inflation	0	0.0597	0.1613	0.0216
OLS	high inflation	1	0.0819	0.2106	0.0092
OLS	drawdown stress	0	0.0510	0.1552	-0.0267
OLS	drawdown stress	1	0.0796	0.1971	0.0447
Ridge	high inflation	0	0.0595	0.1608	0.0130
Ridge	high inflation	1	0.0813	0.2085	0.0079
Ridge	drawdown stress	0	0.0506	0.1535	-0.0347
Ridge	drawdown stress	1	0.0793	0.1964	0.0406
Lasso	high inflation	0	0.0607	0.1637	0.0401
Lasso	high inflation	1	0.0783	0.1955	-0.1002

Model	Regime	State	MSE	MAE	Spearman
Lasso	drawdown stress	0	0.0485	0.1465	-0.3276
Lasso	drawdown stress	1	0.0800	0.1954	0.0348
Elastic net	high inflation	0	0.0604	0.1646	0.0340
Elastic net	high inflation	1	0.0785	0.1971	-0.1007
Elastic net	drawdown stress	0	0.0485	0.1479	-0.3267
Elastic net	drawdown stress	1	0.0798	0.1965	0.0291
Rolling OLS	high inflation	0	0.0990	0.2201	0.0270
Rolling OLS	high inflation	1	0.0832	0.2138	0.0084
Rolling OLS	drawdown stress	0	0.0589	0.1720	-0.0362
Rolling OLS	drawdown stress	1	0.1147	0.2472	0.0558
CatBoost	high inflation	0	0.0617	0.1707	-0.0181
CatBoost	high inflation	1	0.0741	0.1844	-0.0140
CatBoost	drawdown stress	0	0.0491	0.1460	0.0045
CatBoost	drawdown stress	1	0.0778	0.1956	-0.0301
PatchTST	high inflation	0	0.0730	0.2021	-0.0528
PatchTST	high inflation	1	0.1112	0.2269	-0.0182
PatchTST	drawdown stress	0	0.0637	0.1807	-0.0685
PatchTST	drawdown stress	1	0.1037	0.2321	-0.0202

Source: calculated by authors.

Appendix Table A2. *Top five mean absolute attribution values by model*

Model	Rank	Feature	Mean absolute attribution
CatBoost	1	mkt_vol_lag2	0.0256
CatBoost	2	mom_cs	0.0184
CatBoost	3	mkt_vol_lag4	0.0098
CatBoost	4	mkt_vol	0.0089
CatBoost	5	mom_cs_lag1	0.0089
Elastic net	1	mom_cs	0.0260
Elastic net	2	mkt_vol	0.0103
Elastic net	3	state_infl_high	0.0072
Elastic net	4	state_dd_stress	0.0066
Elastic net	5	illiq_cs_x_stress	0.0048
Kalman TVP	1	mom_cs	0.0672
Kalman TVP	2	mkt_vol	0.0408

Model	Rank	Feature	Mean absolute attribution
Kalman TVP	3	state_infl_high	0.0187
Kalman TVP	4	size_cs	0.0152
Kalman TVP	5	state_dd_stress	0.0150
OLS	1	mom_cs	0.0535
OLS	2	state_infl_high	0.0211
OLS	3	mkt_vol	0.0189
OLS	4	state_dd_stress	0.0186
OLS	5	size_cs	0.0104
PatchTST	1	mkt_vol_lag3	0.0221
PatchTST	2	mkt_vol_lag5	0.0202
PatchTST	3	mom_cs_lag4	0.0182
PatchTST	4	mom_cs_lag2	0.0178
PatchTST	5	mkt_vol_lag2	0.0157
Ridge	1	mom_cs	0.0522
Ridge	2	mkt_vol	0.0187
Ridge	3	state_infl_high	0.0181
Ridge	4	state_dd_stress	0.0173
Ridge	5	size_cs	0.0097
Rolling OLS	1	mom_cs	0.1072
Rolling OLS	2	mkt_vol	0.0744
Rolling OLS	3	state_dd_stress	0.0632
Rolling OLS	4	state_infl_high	0.0344
Rolling OLS	5	size_cs	0.0137

Source: calculated by authors.

Appendix Table A3. *Rule-based selector versus transformer benchmark after 2020*

Sample	N quarters	Identical to transformer	Switch away	Mean selector return	Mean transformer return	Mean difference	Median difference	Hit rate	Cumulative selector return	Cumulative transformer return	Cumulative difference
All post-2020	19	15	4	0.0568	0.0652	-0.0085	0.0000	0.0000	1.4280	1.8170	-0.3888

Sample	N quarters	Identical to transformer	Switch away	Mean selector return	Mean transformer return	Mean difference	Median difference	Hit rate	Cumulative selector return	Cumulative transformer return	Cumulative difference
Matches transformer	15	15	0	0.0692	0.0692	0.0000	0.0000	0.0000	1.3600	1.3600	0.0000
Switches away	4	0	4	0.0102	0.0504	-0.0402	-0.0252	0.0000	0.0291	0.1939	-0.1648

Source: calculated by authors.

Appendix Table A4. *PatchTST selected portfolio concentration diagnostics*

Forecast date	N assets	Portfolio return	Max weight	Top 5 weight	HHI	Effective N	Top contributor	Top contribution
2022-03-31	10	-0.0884	0.15	0.6491	0.1125	8.890	NLMK	-0.0336
2022-06-30	8	0.0900	0.15	0.7500	0.1402	7.133	ELTZ	0.1149
2022-09-30	12	0.2293	0.15	0.7213	0.1251	7.993	ELTZ	0.0486
2022-12-31	10	0.3314	0.15	0.7480	0.1340	7.460	KRKOP	0.1020
2023-06-30	7	0.4173	0.15	0.7500	0.1450	6.897	LVHK	0.1929
2024-03-31	8	-0.0610	0.15	0.7500	0.1371	7.296	RUSI	-0.0348
2024-09-30	9	-0.0847	0.15	0.7049	0.1250	7.997	RUSI	-0.0441
2025-06-30	9	0.0334	0.15	0.6969	0.1232	8.114	KRKOP	0.0131
2025-09-30	8	-0.0137	0.15	0.7500	0.1340	7.462	KRKN	0.0198

Note. The full concentration table contains all PatchTST selected quarters; this appendix reports the periods needed to reproduce the concentration interpretation in the text. *Source: calculated by authors.*