Does the Virtue of Complexity Travel? Lessons from Foreign Exchange Forecasting

Rehim Kılıç*

August 26, 2025

Abstract

This paper investigates whether the "virtue of complexity" in equity return prediction extends to exchange rate forecasting. Using nonlinear Ridge regressions with Random Fourier Features (Ridge–RFF), we evaluate predictive performance relative to linear regression and the robust random walk benchmark across three sets of economic fundamentals—traditional monetary, expanded monetary and non-monetary, and Taylor-rule predictors—under rolling training windows of 12, 60, and 120 months. Our results offer a cautionary perspective. Complexity delivers only modest, localized gains: in very small samples with rich predictor sets, Ridge–RFF can outperform linear regression. Yet these improvements never translate into systematic gains against the random walk. As training windows expand, Ridge–RFF quickly loses ground, and linear regression increasingly dominates, at times even outperforming the random walk under expanded fundamentals. Market-timing analyses reinforce these findings: while complexity-based strategies yield occasional short-sample gains, they are unstable and prone to underperformance, whereas simpler linear and random walk strategies provide more robust economic value. By incorporating formal forecast evaluation tests—including Clark–West and Diebold–Mariano—we show that apparent gains from complexity are fragile and rarely statistically significant. Overall, our evidence points to a limited virtue of complexity in FX forecasting: complexity can help under narrowly defined conditions, but parsimony and the random walk benchmark remain more reliable across samples, predictor sets, and economic evaluations.

JEL Classification: F41, C50, G11, G15.
Keywords: Exchange rate, predictability, Ridge, RFF, complexity, machine learning, nonlinearity, random walk.

*Federal Reserve Board, Washington, DC E-mail: rehim.kilic@frb.gov. The views presented in this paper are solely those of the author and do not represent those of the Board of Governors or any entities connected to the Federal Reserve System.

1 Introduction

Accurately forecasting exchange rates remains one of the most challenging and enduring puzzles in international finance, dating back to the seminal study by Meese and Rogoff (1983). Their influential finding—that structural models based on economic fundamentals consistently underperform a naive random walk in out-of-sample predictions—continues to be supported by subsequent empirical investigations (see, Rossi, 2013, fior an excellent review of the literature). Despite considerable theoretical advancements and methodological innovations, the so-called Meese–Rogoff puzzle persists, emphasizing the robustness of the random walk benchmark and questioning the predictive power of economic fundamentals.

Yet recent studies have demonstrated incremental progress by identifying specific conditions under which exchange rate predictability may indeed emerge. Empirical research highlights the relevance of employing richer predictor sets, accounting for structural policy regime changes, and utilizing methodological refinements such as adaptive modeling frameworks and machine learning approaches. In particular, the integration of monetary and non-monetary fundamentals, global risk measures, and Taylor-rule fundamentals has shown promise under certain contexts and horizons (see, Molodtsova and Papell, 2009; Zorzi et al., 2015; Pfahler, 2022; Engel and Wu, 2024; Filippou et al., 2025, among others).

Parallel advances in econometric modeling—especially driven by high-dimensional techniques and machine learning—have recently sparked renewed interest in exploiting nonlinear relationships in forecasting contexts. Notably, Kelly et al. (2024) have reinvigorated discussions around model complexity in equity return forecasting, introducing the concept of the *virtue of complexity* (VoC). They argue that high-dimensional predictive models employing nonlinear transformations, such as Ridge regression augmented with Random Fourier Features (RFF), can significantly outperform simpler linear models by capturing subtle predictive signals through appropriate regularization. These findings have triggered a nuanced debate. Critics like Nagel (2025) and Buncic (2025) warn that apparent improvements may reflect mechanical volatility-timing artifacts rather than genuine economic structure. In response, Kelly and Malamud (2025) enrich this narrative by distinguishing between nominal and effective complexity, exploring how implicit regularization and sample-dependent learning dynamics determine when complexity becomes beneficial—especially in small samples—and introducing the idea of ensemble complexity as a way to harness complexity more robustly.

Motivated by these recent methodological debates and the persistent exchange rate forecasting puzzle, this paper examines whether the documented *virtue of complexity* identified primarily in equity markets extends meaningfully to the distinct context of exchange rate predictability. Specifically, we address three central research questions: (1) Does complexity,

implemented via Ridge regressions augmented with Random Fourier Features, yield meaningful out-of-sample predictive advantages over simpler linear models and the robust random walk benchmark in the context of exchange rate forecasting? (2) How sensitive are the predictive benefits of complexity to the variation in the predictor set? (3) Do complex forecasting models provide superior economic value when evaluated through practical market-timing strategies, thereby offering tangible benefits to policymakers and practitioners?

We address these questions using three distinct sets of economic fundamentals commonly examined in the literature: (i) traditional monetary fundamentals (money supply growth, interest rate differentials, output growth, and inflation differentials), (ii) an expanded set that combines monetary and non-monetary fundamentals—including global risk indicators as proposed by Engel and Wu (2024), and (iii) Taylor-rule-based fundamentals that incorporate interest rate differentials and output gaps, following Molodtsova and Papell (2009). To ensure a comprehensive evaluation, our empirical design considers multiple training sample sizes (12, 60, and 120 months). Forecasting performance is assessed rigorously using established statistical metrics—out-of-sample R^2 and Mean Squared Prediction Error (MSPE)—but extends beyond the measures typically reported in recent VoC studies on equity returns. Specifically, we employ Clark–West (CW) tests to evaluate predictive accuracy relative to the random walk benchmark, and Diebold–Mariano (DM) tests to compare the forecast performance of high-complexity Ridge–RFF models against conventional linear regressions. These additional tests sharpen inference on statistical significance and robustness. Finally, we complement the statistical evaluation with an assessment of economic relevance through currency market-timing strategies, considering both single-currency portfolios and equal-weighted portfolios of eight currencies.

This paper makes three contributions. First, we provide one of the first rigorous tests of the *virtue of complexity* hypothesis in the context of exchange rate forecasting, extending a debate largely centered on equity returns. Second, our empirical analysis shows that while nonlinear Ridge–RFF models can deliver localized gains in very small samples with rich fundamentals, these advantages are fragile: as sample sizes grow, linear regressions often match or exceed their performance, and the random walk remains a remarkably robust benchmark. Third, by linking statistical forecast performance to market-timing portfolio outcomes, we demonstrate that complexity rarely translates into durable economic value. In doing so, we also extend the VoC literature methodologically: whereas recent equity-based studies have relied primarily on out-of-sample R^2 and Sharpe ratios, we incorporate Clark–West (CW) and Diebold–Mariano (DM) tests to assess forecast significance and relative accuracy, thereby providing sharper inference on the robustness of predictive gains. These findings highlight both the limits of

complexity in FX and the continued importance of parsimony, offering guidance to researchers, practitioners, and policymakers on when (and when not) complexity is worth pursuing.

The remainder of the paper proceeds as follows. Section 2 provides an review of the related literature and contributions our findings relative to the recent literature. Section 3 outlines the empirical methodology and data and construction of the exchange rate fundamentals used. Section 4 presents detailed empirical results and statistical analyses, while Section 5 discusses the economic implications derived from market-timing strategies under both single-currency and equal-weighted currency portfolios. Section 6 concludes and identifies promising avenues for future research with additional details on data in Appendix A.

2 Literature Review

Empirical exchange rate forecasting has long been challenged by the seminal finding of Meese and Rogoff (1983) that no structural model could consistently beat a naive random walk. This "random walk benchmark" — typically a no-change forecast — remains notoriously difficult to outperform in out-of-sample tests. Early and recent studies alike confirm that a random walk (especially without drift) often produces lower forecast errors than economic models. For example, Cheung et al. (2005) examined an expanded set of traditional models (including monetary and productivity-based specifications) and found that no model consistently outperformed a random walk in terms of mean squared error. An updated analysis by Cheung et al. (2017) reached a similar conclusion: even with new variants (e.g., real interest rate parity with shadow rates, Taylor-rule-based models), "the more recent models do not consistently outperform older ones" or the random walk, especially for major pairs like EUR/USD. They did note, however, a few instances of predictive success at longer horizons — for instance, models beating the random walk at the 5-year horizon occurred more frequently than at short horizons. This aligns with earlier hints in the literature that any predictability might emerge only at long horizons or under specific conditions, though such gains are modest and sample-dependent. Overall, the random walk without drift remains the toughest benchmark to beat in exchange rate forecasting, underscoring the persistence of the Meese–Rogoff puzzle.

Nevertheless, numerous studies in the last decade have explored when and how economic fundamentals can help forecast exchange rates. Rossi (2013) provides a comprehensive survey of post-2000 findings, concluding that the answer to "Are exchange rates predictable?" is "It depends." Predictability appears strongest under specific choices of predictors, models, and horizons. Notably, Rossi finds that using certain economic fundamentals like Taylor rule differentials (which proxy relative monetary policy stances) or external imbalances (e.g., net foreign asset positions) can yield improved forecasts under the right conditions. In particular,

these fundamentals showed promise at shorter horizons when used in simple linear models with limited parameter estimation. This result echoes earlier work suggesting that imposing theoretical long-run relationships can aid forecasting: for example, monetary models that enforce long-run equilibrium (via error-correction terms) had some success at multi-year horizons in the 2000s.

A recent study by Engel and Wu (2024) further supports the perspective of circumstantial feature of predictability by showing significant improvements in the performance of standard exchange rate models since the early 2000s. Engel and Wu argue that these models, incorporating both monetary and non-monetary fundamentals alongside global risk and liquidity measures, perform substantially better in recent decades compared to the earlier periods (1970s–1990s). They attribute this enhanced predictability largely to improved and more credible monetary policy regimes, such as inflation targeting, which reduce the scope for self-fulfilling expectations and excessive volatility.

Similarly, studies have re-examined purchasing power parity (PPP) as a predictor. While PPP was historically critiqued for poor short-run performance, recent evidence suggests it can be useful over longer periods. Zorzi et al. (2015) show that a calibrated half-life PPP model (assuming the real exchange rate slowly mean-reverts to its equilibrium) forecasts real and nominal exchange rates better than a random walk at both short and long horizons. In fact, a series of papers find that a simple PPP-based forecast often significantly outperforms the random walk, provided one assumes realistic slow adjustment of exchange rates toward their PPP value. The intuition is that even if fundamental/value relationships are weak in the short run, the gradual pull of exchange rates toward relative price parity can provide exploitable forecasting power at medium to long horizons. By contrast, other traditional fundamentals have a more mixed record. Standard monetary models (which include money supplies, outputs, and interest rates) and uncovered interest parity (UIP, using interest differentials) generally fail to beat the random walk at short horizons. There is some evidence of monetary models improving forecast accuracy at longer horizons or in panel estimations, but results are not consistent across samples. For instance, Molodtsova and Papell (2009) found that incorporating Taylor rule fundamentals yields better-than-random-walk forecasts for certain currency pairs during the early 2000s, though such gains dissipate in other periods. Likewise, interest rate parity deviations alone have limited predictive power due to time-varying risk premia (the forward premium puzzle). Researchers have tried augmenting models with proxies for risk or liquidity conditions — e.g., global volatility indices or financial stress measures — to capture these time-varying premiums. Cheung et al. (2017) report that adding risk and liquidity factors can improve the in-sample fit of a sticky-price monetary model, but the out-of-sample predictive improvement remains still unimpressive. In summary, traditional

macro fundamentals by themselves have struggled to consistently forecast exchange rates better than a random walk, except in specific circumstances such as enforcing long-run PPP or when particular fundamental imbalances become extreme.

One reason for the inconsistent performance of fundamentals is structural instability — the idea that the relationship between exchange rates and fundamentals may change over time. Recent research has tackled this challenge through model uncertainty and regime-switching approaches. For example, Kouwenberg et al. (2017) develop a model selection framework that allows the set of relevant fundamentals to shift over time, aligning with theories that investors pay selective attention to different variables in different periods. They design an adaptive forecasting rule that at any given time picks the best-performing fundamentals-based model out of a broad menu of economic variables. This approach yields notable forecasting gains: out-of-sample tests show it significantly beats the random walk for 5 out of 10 major currencies. The selected fundamentals — and their weights — vary over time, suggesting that part of the exchange rate disconnect may be due to markets periodically rotating their focus (e.g., from interest differentials at one time to terms-of-trade or external deficits at another). Such findings are consistent with the "scapegoat" models of Bacchetta and van Wincoop (2004), which posit that agents might rationally latch onto different fundamentals in explaining currency movements when true underlying drivers are unobservable. By accounting for model uncertainty and allowing for time-varying parameter emphasis, the literature shows improved forecast accuracy and even economic value (e.g., the adaptive forecasts can inform profitable currency trading strategies). Other studies exploit panel data and factor models to pool information from multiple currencies. Engel and West (2012) extract common factors from a panel of bilateral dollar exchange rates and combine them with fundamentals; at long horizons (8 to 12 quarters) in more recent samples, these factor-based forecasts modestly outperform the random walk. Similarly, panel econometric techniques can increase predictive power by sharing information across countries: Ince and Kubler (2014) finds that using panel estimation with real-time data helps uncover predictability that is missed in single-currency, revised-data regressions. The general message is that incorporating more data (cross-sectional or temporal) and allowing for structural change can enhance the forecastability of exchange rates, albeit incrementally.

In the past several years, a wave of novel methodological developments — particularly those drawn from machine learning (ML) and big data — have been applied to exchange rate prediction. The motivation is that flexible, data-driven algorithms might detect complex nonlinear relationships or interactions in the data that elude traditional linear models. One strand of this research applies relatively simple ML tools to fundamentals-based forecasting. For example, Amat (2018) use regularization techniques (ridge regressions and an exponentially

weighted averaging of predictors) to forecast exchange rates with a range of economic fundamentals. They report slight improvements over OLS: while no large reduction in RMSE was found, their models were able to predict the correct direction of quarterly exchange rate changes a bit above 50% of the time for most major currency pairs — a small but non-trivial gain given that a random guess would be 50%. Building on that work, Pfahler (2022) explore more complex ML models like artificial neural networks (ANNs) and gradient-boosted trees (XGBoost) in a panel of 10 OECD currencies. Their findings indicate that nonlinear ML methods can indeed outperform the random walk under certain setups. In particular, when predicting the *direction* of change rather than exact amounts, the ML models showed significant predictive power. The XGBoost models beat the random walk's directional accuracy by a small margin (sometimes significantly so), and the ANN models by an even larger margin — often statistically significant at the 1% level. However, an interesting nuance emerged: these strong results depended on including "time dummy" variables in the model, suggesting that the algorithms were capturing some time-specific effects or regimes. In fact, an ANN using only time fixed effects (with no fundamentals) could predict well out-of-sample in many cases, raising the concern that ML might be picking up patterns such as trends or momentum rather than genuine economic relationships. Nonetheless, when the ANN was fed both time dummies and fundamental variables (especially those from a monetary model), its performance improved beyond using time dummies alone. This implies that the ML model was able to exploit interactions between economic fundamentals and time-specific factors — possibly capturing how the impact of fundamentals changes over different periods. In sum, machine learning methods have shown promise in extracting predictive signals from data that traditional approaches deemed unforecastable, but they also highlight the importance of handling structural change and of carefully interpreting what drives the forecast gains.

More advanced data-driven approaches have likewise been tested. Deep learning architectures — such as recurrent neural networks (LSTM), convolutional neural networks, and transformers — have been applied to exchange rate series with large sets of input features. Meng et al. (2024), for example, forecast the Chinese RMB/USD rate using 40 input features spanning macroeconomic indicators and market data. They find that a sophisticated transformer-based model achieved the best accuracy, outperforming simpler models on this task. Notably, their study emphasizes the role of economic fundamentals even in a data-rich, machine-learning context: using explainability techniques, they show that variables such as China–U.S. trade volumes and the exchange rates of major related currencies were among the most influential features for the model's predictions. In other words, the deep learning model's success still hinged on fundamental economic information, but the flexible ML framework was better able to capture the nonlinear and interactive effects of those fundamentals on

the exchange rate. Similarly, in a recent study by Filippou et al. (2025) demonstrates that short-horizon exchange rates can be predicted by combining economic fundamentals with machine learning methods. Using an ensemble of elastic net and deep neural networks applied to country-level and global variables, they show consistent outperformance over the random walk benchmark—especially during periods of financial stress. Their findings highlight the importance of modeling nonlinearities and state-dependent dynamics, offering both statistical and economic value in forecasting. These results generally align with a broader insight in the forecasting literature that more complex models can utilize large information sets more effectively. This paper is closely related to the recent literature on return predictability and the so-called "virtue of complexity (VoC)." While most existing studies, including Kelly et al. (2024), focus on equity return forecasting, our results provide an external test of these ideas in the distinct setting of exchange rate predictability. Kelly et al. (2024) argue that high-dimensional predictive models—implemented with random Fourier features and ridge regularization—can outperform simpler alternatives, especially in short-horizon equity return forecasts. They attribute this to "benign overfitting," where large models, when appropriately regularized, exploit weak but pervasive predictive signals.

Subsequent work has refined and challenged this view. Kelly and Malamud (2025) emphasize the distinction between nominal complexity ($c = P/T$, the parameter-to-sample ratio) and effective complexity, which accounts for shrinkage and implicit regularization. They show that nominal complexity is central to out-of-sample performance in equities. Our evidence suggests that, in FX, high nominal complexity does not translate into lasting effective gains—particularly against the random walk—highlighting domain-specific limits of the VoC framework.

Our findings also resonate with critical perspectives. Nagel (2025) argues that RFF-based forecasts in small samples mimic volatility-timed momentum strategies rather than uncovering genuine nonlinear structure, consistent with our observation that RFF advantages vanish quickly as sample size grows. Similarly, Buncic (2025) shows that methodological choices—such as omitting intercepts or imposing rigid aggregation of features—can exaggerate the apparent benefits of complexity. In our FX application, once sample sizes expand and nominal complexity declines in relative terms, the edge of complex models dissipates. Linear regression and, in many cases, the random walk not only regain ground but consistently dominate, both statistically and economically. This pattern accords with the theoretical perspective of **?**, who stress that the slope of the VoC curve is shaped by the complexity ratio: as T rises relative to P, the benefits of high nominal complexity attenuate.

Overall, our results confirm the conditional nature of complexity's value. Consistent with Kelly et al. (2024), Ridge–RFF can provide localized gains in small samples with rich

predictor sets, reflecting benign overfitting. Yet these advantages do not scale: as training windows expand, the predictive advantage of complexity diminishes or reverses, and simpler models—linear regression or even the random walk—become more robust. This pattern echoes critiques by Nagel (2025) and Buncic (2025), while also aligning with **?**, who show that in high-dimensional settings the marginal benefits of complexity erode once noise in features is considered. In the FX domain, complexity delivers at most localized, fragile improvements, while parsimony proves systematically more reliable.

We also find that under the expanded set of monetary and non-monetary fundamentals, linear regression consistently outperforms the random walk benchmark when estimated with medium and especially large training samples (notably at the 120-month window) in out-of-sample tests. This result is in line with Engel and Wu (2024), underscoring the importance of richer predictor information and longer historical samples for recovering exchange rate predictability. Strikingly, however, the Ridge–RFF model underperforms even in this potentially data-rich environment—failing to beat the random walk and performing worse than simple linear regression—suggesting that additional complexity does not translate into incremental forecasting gains when fundamentals are informative and sample size is sufficient.

Taken together, evidence in this paper together with some of the recent studies on VoC literature highlights the conditional nature of complexity's appeal. While complexity can provide localized benefits in equity forecasting and in very small-sample FX exercises, its scope is limited, fragile, and heavily dependent on benchmarks. Our results reinforce the importance of including both parsimonious linear models and the random walk in evaluation, and they underscore that, in FX, the virtue of complexity is sharply constrained by data conditions and economic structure.

3 Empirical Methodology and Data

This section describes the random Fourier features (RFF) non-linear machine learning framework of (Rahimi and Recht, 2007, 2008) that Kelly et al. (2024) utilize to build increasingly more complex machine learning models, together with the linear models that has been used in the empirical exchange rate literature based on economic fundamentals. Our empirical analysis assesses the predictive performance of complex machine learning models, simpler linear forecasting models and the naive random walk for exchange rate returns. We apply these models across different sets of exchange rate fundamentals and varying forecast horizons and sample sizes to provide a robust assessment of the so-called "virtue of complexity" in the context of exchange rate forecasting.

Following the exchange rate forecasting literature, we define the target variable as the log return of the bilateral exchange rate between the U.S. dollar and a foreign currency:

$$\Delta s_t = \log(S_t) - \log(S_t),$$

where S_t denotes the spot exchange rate (defined as the domestic currency price of a foreign currency) observed at time t. The log of the nominal exchange rate, denoted as s_t, is the U.S. dollar price of a foreign currency. Exchange rates used are the U.S. exchange rate versus the major currencies including Australian dollar (AUD), Canadian dollar (CAD), Swiss franc (CHF), the euro (EUR), U.K. pound sterling (GBP), Japanese yen (JPY), Norwegian krone (NOK), and Swedish krona (SEK). Details of the data and sources used in in the analysis are provided in Appendix A.

We evaluate both short-horizon ($h = 1$ month) and long-horizon ($h = 12$ months) forecasts. The predictor information set consists of the time t measurable $k \times 1$ vector G_t containing three different sets of fundamentals motivated by the large literature on the so called "exchange rate disconnect" (Rossi, 2013; Molodtsova and Papell, 2009; Engel and Wu, 2024).

1. Traditional monetary fundamentals include money supply growth, inflation, interest rate, and output growth differentials between the home (US) and foreign countries which gives 4×1 vector, $G_t = [\Delta(m_t - m_t^*), \Delta(i_t - i_t^*), \Delta(\pi_t - \pi_t^*), \Delta(y_t - y_t^*)]'$ with '$*$' denoting the foreign variable and Δ is the first difference operator and represents a one-month difference. The inflation over previous 12 months in the U.S. and foreign country defined as $\pi = p_t - p_{t-12}$ and $\pi_t^* = p_t^* - p_{t-12}^*$ where p_t and p_t^* are the log U.S. and foreign consumer price indexes. Money supply in the U.S. and foreign country are measured by M_t, and M_t^* and monthly logarithms are denoted by m_t and m_t^*. Therefore, $\Delta(m_t - m_t^*)$ gives the money supply growth differential between the U.S. and foreign country. The U.S. and foreign nominal interest rates are denoted by i_t and i_t^* and measured by 3-month U.S. and foreign government bond rates.The log output in the U.S. and foreign countries are denoted by y_t and y_t^* with output is captured by monthly industrial production index. The output growth differential between the U.S. and foreign country is then defined by $\Delta(y_t - y_t^*)$. The traditional fundamentals is motivated by the monetary model of exchange rate determination with sticky prices and the predictor set is more consistent with classical monetary models such as those tested by Meese and Rogoff (1983) and revisited in subsequent studies (Rossi, 2013).

2. Expanded monetary and non-monetary fundamentals includes variables introduced by Engel and Wu (2024). These fundamentals are home and foreign real interest rate changes (Δr_t, and Δr_t^*), home and foreign inflation rates, π_t, and π_t^*, one-month lagged

real exchange rate, q_{t-1}, the U.S. trade balance on goods and services divided by U.S. GDP, $\frac{TB}{GDP}$, and a measure of monthly change in global risk aversion, ΔRisk_t). This gives us a 7×1 vector, $G_t = \left[\Delta r_t,\ \Delta r_t^*, \pi_t,\ \pi_t^*,\ \Delta\text{Risk}_t,\ q_{t-1},\ \frac{TB_t}{GDP_t}\right]'$. The real interest rates are defined as $r_t = i_t - \pi_t$ and $r_t^* = i_t^* - \pi_t^*$ and the log real exchange rate as $q_t = s_t + p_t^* - p_t$.

3. Taylor-rule-based fundamentals include the U.S. and foreign nominal interest rates, i_t and i_t^*, inflation rates, π_t and π_t^*, output gaps, y_t^{gap} and y_t^{gap*}, and the real exchange rate, q_t. Taylor-rule fundamentals give a 7×1 vector, $G_t = [i_t,\ i_t^*, \pi_t,\ \pi_t^*,\ y_t^{gap},\ y_t^{gap*},\ q_t]'$ (see, Molodtsova and Papell, 2009).

We begin with a standard linear regression framework, estimated using ordinary least squares (OLS), where exchange rate returns between t and $t+1$ are regressed on a set of macroeconomic fundamentals at date t:

$$\Delta s_{t+1} = G_t'\beta + \varepsilon_{t+1}. \tag{1}$$

This direct forecasting specification is also called ''single-equation, lagged fundamental model" and has been used extensively in the empirical exchange rate prediction literature (Rossi, 2013).[1]

If the relationship between exchange rate returns and predictors are not linear, following Kelly et al. (2024), one can assume that the true predictive model for exchange rate returns, Δs_t, follows the process:

$$\Delta s_{t+1} = f(G_t) + \varepsilon_{t+1}, \tag{2}$$

where G_t is "fixed set of predictive signals, and $f(.)$ a smooth function " (p. 460 Kelly et al., 2024). Although the set of predictors G_t may be known to the researcher, the prediction function $f(.)$ is unknown and "can be approximated with a sufficiently wide neural network (Kelly et al., 2024):

$$f(G_t) \approx \sum_{i=1}^{P} Z_{i,t}\beta_i \tag{3}$$

where $Z_{i,t} = g(\omega_i' G_t)$, P is the number of terms used in the approximation to $f(G_t)$, $g()$ is a predefined non-linear activation function with weight vector ω_i, and G_t is the vector of predictor variables under a given set of fundamentals. The approximating model for the true predictive model in (2) then takes the form:

[1]One advantage of this direct forecasting equation is that is less prone to endogeneity of fundamentals and hence, OLS can be used to estimate parameters.

$$\Delta s_t = \sum_{i=1}^{P} Z_{i,t}\beta_i + \tilde{\varepsilon}_t. \tag{4}$$

The approximation accuracy of $\sum_{i=1}^{P} Z_{i,t}\beta_i$ in (4) to the unknown (true) function $f(.)$ in (3) depends on the number of predictor terms (or features) P. With a fixed number of data points T used to train the model, one has to decide how large a P to use. A model is said to be simple, when $P << T$ (number of features is much less than the training sample size), and will have low variance due to a parsimonious parametrization, but it will only provide a crude approximation to the true function $f(.)$. A high-complexity model where $P >> T$ will have better approximating properties to $f()$, but may be poorly behaved and require regularization (or shrinkage estimation), which can increase bias. As in Kelly et al. (2024), complexity c is defined as the ratio of the number of features P to the number of training observations T, $c = P/T$. In a recent paper, Kelly and Malamud (2025) labels this as 'nominal complexity' in order to distinguish it from the notion of 'effective complexity' that is introduced by Nagel (2025) and formally described as the ratio of effective number of parameters to sample size where effective number of parameters accounts for shrinkage, bias, or implicit regularization.

Kelly et al. (2024) show theoretically that expected out-of-sample forecast accuracy and portfolio performance are strictly increasing in model complexity when appropriate shrinkage is applied. To verify their theoretical results, they utilize the RFF framework to be able to smoothly transition from a low-complexity model to a high complexity one. Specifically, the RFF methodology takes as input the $k \times 1$ vector G_t of predictor variables under a given set of fundamentals and converts the information contained in G_t into (a pair of) new signals $S_{i,t}$ defined as:

$$Z_{i,t} = \left[\sin(\gamma\omega_i' G_t)\ \cos(\gamma\omega_i' G_t)\right]', \tag{5}$$

where $\omega_i \sim i.i.d.N(0, I)$ is a $k \times 1$ random weight vector, I is the identify matrix of conformable size, and γ is a standard deviation parameter set to 2 in Kelly et al. (2024) that controls the variability in the Gaussian draws of ω_i. To generate P RFFs, one needs to generate $P/2$ weights ω_i and then evaluate the transformation in (5). This approach allows generating larger predictor sets based on a limited number of k predictors and thereby more *'complex'* models from the same underlying fundamentals and the associated information set by simply creating new random weight vectors ω_i and RFF pairs $Z_{i,t}$ from (5).

When $P >> T$, estimation of β_i is obtained by minimizing the Ridge objective function:

$$\widehat{\beta}_i(\lambda) = \arg\min_{\beta} \left[\sum_{t=1}^{T} \left(\Delta s_t - \sum_{i=1}^{P} Z_{i,t}\beta_i \right)^2 + \lambda \sum_{i=1}^{P} \beta_i^2 \right], \tag{6}$$

where λ is the regularization parameter (z in Kelly et al., 2024) and $\widehat{\beta}(\lambda)$ is the vector of regularized least squares estimates of $P \times 1$ vectors of stacked $\{\beta_i\}_{i=1}^{P}$. The regularization term λ penalizes the size of the coefficients, thereby mitigating overfitting when the number of predictors is large relative to the sample size. Following Kelly et al. (2024), we set $\lambda = 1000$.

Given the Ridge regression estimates of $\widehat{\beta}(\lambda)$, $h-$step ahead forecasts of Δs_t, given information up to time t, are computed as

$$\widehat{\Delta s_{t+h|t}} = \sum_{i=1}^{P} Z_{i,t+h}\widehat{\beta}_i(\lambda), \tag{7}$$

where following the exchange rate literature, we compute forecasts directly using observed actual rather than the forecasted values of fundamentals or in the case of Ridge with RFF generated predictors corresponding to date $t+h$ (see, Meese and Rogoff, 1983; Rossi, 2013, among others).[2]

To evaluate out-of-sample predictive performance, we use rolling window regressions with window lengths of 12, 60, and 120 months. At each step, the model is re-estimated using only the training window, and forecasts are generated for the subsequent 1-month ahead exchange rate return. Similar to Kelly et al. (2024), we consider training window sizes as small as $T = 12$ and since in our case, G_t is of dimensions $k = 4$ and 7 depending on the set of fundamentals used, we do not have Ridgeless regression as in Kelly et al. (2024), but note that in our case, linear regression corresponds to $\lambda = 0$ without the need for using Moore-Penrose pseudo-inverse to solve $\hat{\beta}$. We follow Kelly et al. (2024) and set $P = 12000$ and $\lambda = 1000$ in our assessment of the complexity.

We evaluate the forecasting performance of linear and high-complexity (nonlinear or Ridge-RFF) models against each other as well as random walk. Following Kelly et al. (2024), we do not include an intercept term in linear and Ridge-RFF models and compare them against a driftless random walk.[3] We use out-of-sample R^2, Mean Squared Prediction Error

[2]Despite predictors might be endogenous in such regressions, they are often used as they avoid the potential lack of predictability stemming from the poor forecasts of the fundamentals, (see, Rossi, 2013, for a review of studies that use actual values of fundamentals rather than their forecasts in the exchange rate prediction literature).

[3]Note that the terms 'linear' and 'OLS' are used interchangeably, as are 'nonlinear' and 'Ridge-RFF', throughout the paper.

(MSPE) ratios, Clark-West (CW) and Diebold-Mariano (DM) forecast comparison tests. We use CW test in comparing the high-complexity and linear regressions relative to the random walk benchmark as random walk is nested while use DM test in comparing Ridge-RFF against the linear benchmark as linear benchamrk not necessarily nested within the Ridge-RFF.

Similar to Kelly et al. (2024), we further evaluate the economic relevance of forecasts by implementing market-timing strategies that take positions in currency portfolios based on $1-$ month ahead predicted returns. At time t, the model-based strategy takes a long (short) position that is determined by the exchange rate return forecast under a model, yielding a portfolio return of $\widehat{\Delta s_{t+1|t}} \cdot \Delta s_{t+1}$, where $\widehat{\Delta s_{t+1|t}}$) denotes the forecast of the excess exchange rate return under a model (Ridge-RFF and linear) and Δs_{t+1} is the realized exchange rate return on date $t+1$. We construct market timing strategy under random walk in a similar fashion where positions are based on the the previous period's exchange rate return, i.e., $\Delta s_t \cdot \Delta s_{t+1}$. Portfolio returns under Ridge-RFF are averaged across 1000 simulations as in Kelly et al. (2024). We use both single-currency market-timing portfolios for each currency as well as equal-weighted eight-currency portfolios by aggregating across currencies at each point in time and evaluate performance using a range of out-of-sample metrics, including Sharpe ratios (SR), Information ratios (IR), skew, maximum loss and t-tests.

4 Out-of-sample Performance Statistics

In this section, we present and discuss out-of-sample performance of linear and nonlinear models relative to the random walk benchmark and to each other, using out-of-sample (OOS) R^2, Mean Squared Prediction Error (MSPE), and the Clark–West (CW) and Diebold–Mariano (DM) predictive accuracy tests across three predictor sets—traditional monetary fundamentals, the expanded monetary and nonmonetary set of Engel and Wu (2024) (we label this set by monetary+), and Taylor-rule fundamentals of Molodtsova and Papell (2009). Because the latter two sets include more predictors, they provide a natural setting to test whether greater nominal complexity improves performance through Ridge–RFF.

Each table corresponds to a *single* training-window length and contains three panels—A: Traditional, B: Monetary+, and C: Taylor rule—so that all predictor sets are compared side-by-side for the same T. Within each panel, columns are grouped into three comparisons: *Linear vs. Random Walk* (reporting R^2_{Lin} and CW against RW), *Ridge-RFF vs. Random Walk* (reporting R^2_{RFF} and CW against RW), and *Ridge-RFF vs. Linear* (reporting the MSPE ratio $MSPE_r = \frac{MSPE_{RFF}}{MSPE_{Lin}}$ and the DM test).[4] Positive and statistically significant

[4]Equivalently, $MSPE_r = \frac{1-R^2 RFF}{1-R^2 Lin}$ allows back-out of model MSPEs from the reported OOS R^2.

CW (DM) values favor the first model named in the column header; negative and significant values favor the second. We employ DM because OLS need not be nested within Ridge–RFF.

Table 1 reports the 1-month ahead forecasting results under traditional, monetary+, and Taylor rule fundamentals in Panels A, B, and C, respectively for the 12-month rolling window. In Panel A with traditional fundamentals where the number of predictors $k = 4$, neither linear regression nor Ridge-RFF succeeds in improving upon the random walk. Out-of-sample R^2—values are uniformly negative across currencies—especially so for the linear model—while Clark–West tests rarely reject the null of equal predictive accuracy (only CAD for the linear model, and CAD and JPY for Ridge-RFF). Across all currencies, high-complexity RFF based model achieves substantially lower MSPE ratios (0.551–0.747) and statistically significant Diebold–Mariano test rejections in its favor against the linear regression.

Results in Panels B and C extend the analysis to the broader predictor sets—monetary and non-monetary fundamentals ($k = 7$) and Taylor-rule fundamentals ($k = 7$). The linear model performs very poorly in these settings, generating severely negative out-of-sample R^2 values for all currencies as the number of parameters estimated increases in rolling regressions. It consistently fails to outperform either the random walk or Ridge-RFF, and under monetary+ fundamentals the Clark–West test even turns significantly negative at the 10% level for CAD, NOK, and SEK, formally favoring the random walk over the linear regression. In contrast, Ridge-RFF continues to dominate the linear model across most currencies, producing near-zero MSPE ratios and large statistically significant (at 1% level) positive Diebold–Mariano statistics. Taken together, these findings confirm the pattern observed with traditional fundamentals: at high nominal complexity, Ridge-RFF reliably improves upon linear regression but does not deliver systematic gains against the random walk benchmark across all currencies. These gains reflect not just nominal complexity, but also the model's ability to exploit "effective complexity," as Ridge regularization ($\lambda = 1000$) selects only a subset of relevant RFFs at each estimation. In sum, when T is small and complexity is at its peak, Ridge-RFF shows advantages over linear regression, but these do not translate into systematic gains against the random walk benchmark.

With the increase in the rolling window size to 60 months, reported in Panels A–C of Table 2, the linear model begins to show improved performance. Out-of-sample R^2 values remain negative, but the magnitudes are substantially smaller, and in several cases the linear model outperforms the random walk, as indicated by positive and statistically significant CW test statistics. For instance, under the Taylor rule fundamentals the linear model outperforms the random walk for three of the eight currencies (GBP, JPY, and NOK) at the 10% or 5% levels. Similarly, the linear model beats the random walk for CAD under traditional

Table 1: Out-of-sample forecast comparisons statistics for 12-month rolling-window across traditional, monetary+, and Taylor rule fundamentals

	Linear vs. Random Walk		Nonlinear vs. Random Walk		Nonlinear vs. Linear	
	R^2_{Lin}	CW	R^2_{Nlin}	CW	$MSPE_r$	DM
Panel A: Traditional fundamentals						
AUD	−1.274	0.491	−0.252	−0.003	0.551	3.270***
CAD	−0.726	1.767**	−0.104	2.027**	0.640	5.148***
CHE	−1.007	0.678	−0.244	0.245	0.620	3.752***
EUR	−1.156	−0.188	−0.208	−0.417	0.560	2.674***
GBP	−0.754	−0.886	−0.251	0.522	0.713	4.585***
JPY	−0.990	1.084	−0.153	2.670***	0.579	3.442***
NOK	−0.636	−0.506	−0.222	0.519	0.747	2.323**
SEK	−0.771	−0.428	−0.209	0.113	0.682	4.059***
Panel B: Monetary and non-monetary fundamentals						
AUD	−4.579	0.133	−0.153	−0.942	0.207	5.187***
CAD	−4.949	−1.400*	−0.147	0.226	0.193	4.649***
CHE	−4.395	−0.403	−0.128	−0.186	0.209	5.527***
EUR	−3.809	0.430	−0.104	1.017	0.230	6.372***
GBP	−3.487	−1.275	−0.099	−0.153	0.245	7.905***
JPY	−17.755	−0.701	−0.163	−0.341	0.062	1.773**
NOK	−151.536	−1.309*	−0.147	−0.053	0.008	1.180
SEK	−6.629	−1.284*	−0.123	0.041	0.147	4.823***
Panel C: Taylor Rule fundamentals						
AUD	−10.167	0.590	−0.224	1.069	0.110	4.825***
CAD	−6.256	−1.057	−0.392	−0.646	0.192	7.028***
CHE	−5.445	−0.251	−0.300	1.326*	0.202	5.879***
EUR	−4.604	−0.102	−0.356	0.651	0.242	6.472***
GBP	−4.490	0.953	−0.266	0.131	0.231	6.909***
JPY	−28.664	1.676**	−0.313	1.787**	0.044	1.267
NOK	−5.868	0.409	−0.366	−1.026	0.199	8.071***
SEK	−6.412	0.900	−0.365	−0.035	0.184	5.913***

This table reports 1-month ahead out-of-sample forecast accuracy performance comparison statistics between the linear (Lin) and random walk (RW), Ridge-RFF (RFF) and random walk, and Ridge-RFF and linear (see, the main text for details). The forecast target is the $1-$month ahead forecast of log USD exchange rate return relative to given currency. Linear vs. Random Walk and Ridge-RFF vs. Random walk panels report oos R^2 values (i.e., R^2_{Lin}, and R^2_{RFF}, respectively), and Clark-West (CW) statistic for equal predictive accuracy against RW. The columns for Ridge-RFF vs. Linear report Mean Squared Prediction Error (MSPE) ratios between RFF and Lin, and Diebold-Mariano (DM) test for equal forecast accuracy between these models. *, **, and *** for CW and DM tests indicate rejection of the null of equal predictive accuracy in favor of the first model (i.e., Lin against the RW and RFF against the RW under CW test, and RFF against the Lin under DM test) if the reported test is positive and in favor of the second model if negative.

fundamentals (Panel A) and for GBP, NOK, and SEK under the monetary+ fundamentals at 10% or 1% significance levels (Panel B).

At the same time, Ridge-RFF performance weakens relative to both the linear model and the random walk. While OOS R^2 values improve modestly, MSPE ratios often reach or exceed unity, indicating that Ridge-RFF offers no efficiency gains over linear regression. For example, MSPE ratios exceed one for all currencies except for AUD and EUR, with DM tests generally failing to reject equal predictive accuracy under the monetary fundamentals (Panel A). In several cases, the balance even tips in favor of the linear model: the DM test rejects in favor of OLS for CHE under traditional fundamentals and for EUR, GBP, NOK, and SEK under Taylor rule fundamentals, with negative and statistically significant DM values at the 10% or 5% levels. Moreover, Ridge-RFF begins to underperform the random walk, with CW tests not only negative but also statistically favoring the benchmark for CHE and NOK under traditional fundamentals and for AUD and EUR under monetary+ fundamentals. In the case of JPY under Taylor rule fundamentals, both linear regression and Ridge–RFF outperform the random walk, as confirmed by statistically significant CW test rejections, while the DM test indicates no significant difference between the two models, underscoring that they remain broadly competitive with each other.

With the increase in the training window to 120 months, reported in Panels A–C of Table 3, the relative performance of linear and Ridge-RFF models shifts decisively in favor of parsimony. First, the linear model shows clear signs of improvement. Out-of-sample R^2 values, while often still slightly negative, are closer to zero and accompanied by several cases of statistically significant CW test rejections in favor of the linear model against the random walk. Under traditional fundamentals (Panel A), CAD and CHE exhibit positive CW test statistics at the 1% and 5% levels, respectively, while under Taylor rule fundamentals (Panel C), where the linear model achieves statistically significant gains for AUD and EUR at the 5% level. Similar to Engel and Wu (2024), the strongest results emerge under monetary+ fundamentals (Panel B), the linear model outperforms the random walk for five out of eight currencies including AUD, EUR, GBP, NOK, and SEK, all at conventional significance levels.

By contrast, Ridge-RFF performance deteriorates further. Out-of-sample R^2 values are consistently more negative than those of the linear model, and CW tests rarely provide evidence of predictive gains relative to the random walk. Moreover, MSPE ratios are uniformly above one, and DM tests overwhelmingly reject the null of equal predictive accuracy in favor of the linear model under Monetary+ and Taylor rule fundamentals. DM tests are not only negative but also highly significant (often at the 1% level) across virtually all predictor sets and currencies. Even in cases where Ridge-RFF attains positive CW values (e.g., JPY and

Table 2: Out-of-sample forecast comparisons statistics for 60-month rolling-window across traditional, monetary+, and Taylor rule fundamentals

	Linear vs. Random Walk		Nonlinear vs. Random Walk		Nonlinear vs. Linear	
	R^2_{Lin}	CW	R^2_{Nlin}	CW	$MSPE_r$	DM
Panel A: Traditional fundamentals						
AUD	−0.155	0.964	−0.116	−0.860	0.966	0.603
CAD	−0.060	2.294**	−0.065	−0.164	1.005	−0.124
CHE	−0.097	0.718	−0.182	−2.488***	1.078	−1.338*
EUR	−0.077	−0.276	−0.076	−0.886	0.999	0.018
GBP	−0.056	0.101	−0.079	0.605	1.021	−0.578
JPY	−0.109	0.312	−0.137	−0.393	1.025	−0.583
NOK	−0.093	−0.332	−0.204	−1.290*	1.101	−1.129
SEK	−0.097	−1.062	−0.100	−0.211	1.003	−0.086
Panel B: Monetary and non-monetary fundamentals						
AUD	−0.353	0.854	−0.238	−1.635*	0.915	0.834
CAD	−0.280	−0.620	−0.242	−0.277	0.970	0.412
CHE	−0.266	−1.007	−0.210	−1.178	0.956	0.601
EUR	−0.227	0.216	−0.338	−1.327*	1.091	−0.802
GBP	−0.136	2.420***	−0.118	0.642	0.984	0.274
JPY	−0.246	0.520	−0.208	−0.208	0.970	0.531
NOK	−0.224	1.341*	−0.206	−0.966	0.986	0.202
SEK	−0.198	1.386*	−0.205	−0.896	1.005	−0.095
Panel C: Taylor Rule fundamentals						
AUD	−0.234	0.721	−0.224	1.134	0.992	0.152
CAD	−0.292	0.429	−0.395	−0.754	1.080	−0.917
CHE	−0.213	0.325	−0.301	1.263	1.073	−1.023
EUR	−0.221	1.202	−0.362	−0.116	1.116	−1.294*
GBP	−0.218	1.849**	−0.308	−0.448	1.074	−1.382*
JPY	−0.282	1.420*	−0.287	1.874**	1.003	−0.053
NOK	−0.221	1.783**	−0.338	−0.552	1.096	−1.888**
SEK	−0.258	0.773	−0.353	0.068	1.075	−1.366*

See, Table 1.

Table 3: Out-of-sample forecast comparisons statistics for 120-month rolling-window across traditional, monetary+, and Taylor rule fundamentals

	Linear vs. Random Walk		Nonlinear vs. Random Walk		Nonlinear vs. Linear	
	R^2_{Lin}	CW	R^2_{Nlin}	CW	$MSPE_r$	DM
Panel A: Traditional fundamentals						
AUD	−0.067	0.257	−0.056	−0.792	0.989	0.484
CAD	−0.010	2.471***	−0.019	0.845	1.009	−0.410
CHE	−0.019	1.654**	−0.104	−0.957	1.083	−1.130
EUR	−0.017	0.539	−0.038	−0.396	1.020	−0.592
GBP	−0.024	0.409	−0.021	0.746	0.997	0.157
JPY	−0.060	−0.204	−0.059	−0.498	0.999	0.030
NOK	0.016	1.096	−0.157	−0.371	1.176	−1.388*
SEK	−0.057	−1.300*	−0.087	−0.811	1.028	−0.795
Panel B: Monetary and non-monetary fundamentals						
AUD	−0.081	1.754**	−0.244	−1.189	1.151	−2.465***
CAD	−0.073	0.853	−0.250	−0.470	1.165	−2.853***
CHE	−0.095	0.193	−0.238	−0.390	1.131	−1.874**
EUR	0.002	2.878***	−0.449	−1.007	1.451	−3.337***
GBP	−0.024	2.682***	−0.140	0.884	1.113	−2.217**
JPY	−0.103	0.466	−0.233	0.220	1.118	−1.934**
NOK	−0.088	2.007**	−0.187	0.210	1.091	−1.377*
SEK	−0.069	1.856**	−0.212	−0.366	1.134	−2.540***
Panel C: Taylor Rule fundamentals						
AUD	−0.046	2.043**	−0.235	1.103	1.181	−3.301***
CAD	−0.062	1.145	−0.414	−0.781	1.332	−3.193***
CHE	−0.105	0.448	−0.314	1.446*	1.189	−3.128***
EUR	−0.092	2.139**	−0.488	−0.882	1.363	−3.147***
GBP	−0.126	0.786	−0.335	−0.565	1.186	−3.407***
JPY	−0.124	0.258	−0.305	1.779**	1.161	−2.033**
NOK	−0.098	0.972	−0.343	−0.558	1.223	−3.952***
SEK	−0.126	−0.040	−0.368	0.068	1.215	−3.380***

See, Table 1.

CHE under Taylor rule fundamentals), the corresponding DM tests indicate that these modest gains are eclipsed by the stronger and more robust performance of the linear model.

Taken together, the 120-month results highlight a decisive reversal: Ridge-RFF loses the edge it displayed in very small samples, while the linear model steadily gains ground, occasionally outperforming the random walk benchmark and dominating Ridge-RFF in direct comparisons.

The results across the three training windows (12, 60, and 120 months) reveal a coherent narrative about the role of complexity in FX predictability: At short horizons ($T = 12$), nominal complexity is at its peak ($p = P/T$ very large), and Ridge-RFF demonstrates tentative advantages over linear regression. It produces lower MSPE ratios and statistically significant DM rejections in its favor across several currencies. Yet, these gains are relative: Ridge-RFF still fails to systematically outperform the random walk benchmark, which remains a formidable hurdle and a benchmark that has not been explored by the recent studies exploring the 'virtue of complexity' in equity return prediction.

At intermediate horizons ($T = 60$), the picture becomes more balanced. The linear model begins to close the gap, with smaller negative R^2 values and occasional statistically significant outperformance of the random walk. Ridge-RFF's edge diminishes: MSPE ratios hover around one, DM tests often fail to reject, and in some cases significantly favor the linear model. Ridge-RFF also begins to underperform the random walk in selected currencies. This marks a turning point, where the "virtue of complexity" weakens and parsimony regains traction.

At long horizons ($T = 120$), the balance tips decisively in favor of parsimony. The linear model consistently outperforms Ridge-RFF, both in terms of MSPE ratios and DM tests, with multiple cases of statistical significance, especially under Monetary+ and Taylor rule fundamentals. It also delivers meaningful gains against the random walk under several predictor sets. Ridge-RFF, by contrast, fails to match this performance, producing uniformly inferior outcomes relative to linear regression and often underperforming the random walk.

Overall, the evidence suggests that the benefits of complexity in FX return prediction are highly sensitive to sample size. In very small samples, consistent with the notion of benign overfitting (Kelly et al., 2024), Ridge-RFF can exploit nominal complexity to improve upon linear regression, though never consistently against the random walk. As the training window expands, however, effective complexity declines, regularization reduces the marginal contribution of RFFs, and linear regression reasserts itself as the stronger predictor—at times even outperforming the random walk.

5 Market timing strategy performance

In this section, we evaluate model performance through their ability to generate market-timing portfolios at both the single-currency and equal-weighted levels, following the approach of Kelly et al. (2024). For the linear and Ridge–RFF models, positions are based on 1-month-ahead exchange rate return forecasts constructed under three sets of fundamentals—Traditional, Monetary+, and Taylor rule—using rolling training windows of 12, 60, and 120 months. Portfolio returns are generated by taking positions proportional to the model's forecast: when the forecast is positive, the strategy takes a long position; when negative, it reduces or reverses exposure. The same trading logic applies uniformly across all models, including the random walk benchmark, where the one-month lagged return dictates position-taking (see Section 3).

In addition to single-currency strategies, we also examine an equal-weighted market-timing portfolio, in which currency positions are aggregated into an equally weighted basket that is rebalanced monthly. Analyzing both types of strategies allows us to distinguish between model performance at the individual currency level and the aggregate portfolio level, thereby capturing differences between idiosyncratic predictability and systematic gains across currencies. In the following we first present and discuss results for the single currency strategies (5.1) and then the results under equal-weighted portfolio (5.2).

5.1 Single-currency portfolios

Table 4 reports results under the random walk–based market timing strategy while Tables 5, 6, and 7 under Ridge-RFF and linear regression-based strategies across traditional, monetary+, and Taylor-Rule fundamentals in columns A-C for $T = 12,\ 60,$ and 120, respectively.

Because random-walk-based strategy relies solely on lagged exchange rate returns, results in Table 4 remain broadly similar across training windows, with differences largely reflecting variation in the out-of-sample evaluation periods. Overall, the RW strategy generates modestly positive Sharpe ratios and information ratios relative to the market across most currencies, though statistical significance is generally weak. The standout case is the Japanese yen (JPY), which delivers consistently higher Sharpe ratios (ranging from 0.257 to 0.288) with t-statistics around 1.8, achieving statistical significance at the 10 percent level in all windows. Its information ratios relative to the market are also positive and significant, making JPY the clearest case where lagged return–based timing has predictive value. By contrast, CAD, EUR, and NOK tend to produce small or negative Sharpe ratios across all windows, indicating little or no timing ability. Other currencies (AUD, GBP, SEK, and CHF) display positive but statistically insignificant Sharpe ratios. Distributional properties are heterogeneous: GBP

Table 4: Market timing performance metrics: Random walk-based timing stratgey

	SR	t	IR v. Mkt	t	Skew	Max Loss
Panel A: Training window size $T = 12$						
AUD	0.064	0.450	0.040	0.281	0.327	10.552
CAD	−0.174	−1.214	−0.167	−1.166	−1.458	9.770
CHE	0.115	0.763	0.101	0.666	−1.778	9.953
EUR	0.058	0.289	0.056	0.275	−2.607	9.430
GBP	0.183	1.278	0.172	1.197	4.129	4.412
JPY	0.257	1.791*	0.236	1.646*	−0.259	8.201
NOK	−0.034	−0.234	−0.040	−0.276	0.151	7.377
SEK	0.113	0.787	0.075	0.522	2.895	7.057
Panel B: Training window size $T = 60$						
AUD	0.145	0.970	0.128	0.855	3.035	5.351
CAD	−0.173	−1.154	−0.168	−1.123	−1.396	9.461
CHE	0.053	0.335	0.030	0.189	−2.084	9.896
EUR	−0.063	−0.288	−0.065	−0.293	−3.583	9.798
GBP	0.185	1.240	0.169	1.129	4.842	4.471
JPY	0.268	1.793*	0.257	1.716*	0.219	8.650
NOK	0.009	0.060	−0.008	−0.051	0.905	4.878
SEK	0.171	1.143	0.122	0.815	3.174	6.961
Panel C: Training window size $T = 120$						
AUD	0.119	0.748	0.112	0.702	2.936	5.076
CAD	−0.167	−1.053	−0.166	−1.047	−1.334	8.988
CHE	0.047	0.278	0.036	0.215	−3.011	10.679
EUR	−0.178	−0.706	−0.201	−0.792	−0.745	4.313
GBP	0.170	1.074	0.176	1.111	5.344	4.434
JPY	0.288	1.814*	0.271	1.704*	−0.229	8.979
NOK	−0.045	−0.282	−0.048	−0.302	0.938	4.659
SEK	0.153	0.965	0.148	0.932	3.161	6.675

This table reports one-month-ahead out-of-sample market-timing strategy performance metrics for single-currency portfolios under the random walk (RW) benchmark. Reported statistics include the Sharpe ratio (SR) of each strategy with associated t-statistics for mean returns, the Information Ratio relative to the market (IR v. Mkt) with corresponding t-statistics, as well as skewness of returns (Skew) and the maximum loss (Max Loss) in standard deviation units. Statistical significance of mean returns and information ratios is denoted by *, **, and *** at the 10%, 5%, and 1% levels, respectively.

and SEK exhibit large positive skew, while CHF and EUR exhibit persistent negative skew. Drawdown risks remain material across all currencies, with maximum losses often between 7 and 10 standard deviations, underscoring the volatility inherent in return-based timing even for this simple benchmark. Notably, GBP combines reasonably strong Sharpe ratios with relatively low drawdowns (around 4.4), while CAD and CHF combine weak or negative Sharpe ratios with among the largest drawdowns, highlighting variation in risk-return trade-offs across currencies.

Turning to fundamentals-based models, Table 5 reports market-timing performance for individual currencies under the traditional monetary fundamentals. At the shortest window ($T = 12$), Ridge–RFF shows clear pockets of strength but no broad dominance. In particular, CAD and JPY stand out for RFF with economically meaningful and statistically significant Sharpe ratios (CAD: $SR = 0.306$, $t = 2.114$; JPY: $SR = 0.373$, $t = 2.580$), supported by significant information ratios against both the market and the random walk. Crucially, "IR

v. LIN" confirms incremental value for RFF in these two currencies (CAD: $t = 1.753$, JPY: $t = 2.414$). Elsewhere, performance is mixed: RFF's SR is near zero for AUD and CHE, negative for EUR, and small for GBP/NOK/SEK; the linear model's gains are limited to CAD (modestly significant) and JPY (positive but not significant). Drawdowns at $T = 12$ are nontrivial across both models—for example, AUD exhibits large maximum losses (about 11 s.d.) for both LIN and RFF—whereas CAD shows comparatively mild losses (about 4.4–4.5 s.d.), aligning with its stronger risk-adjusted performance. Skewness varies widely: LIN returns are strongly negatively skewed for GBP and EUR, while RFF skewness is more benign for JPY and mixed elsewhere.

With the 60-month window ($T = 60$), the balance tilts toward parsimony. Linear timing improves—CAD is now solidly significant (LIN: $SR = 0.345$, $t = 2.279$)—while RFF deteriorates sharply across most currencies. RFF delivers negative and often statistically significant Sharpe ratios for CHE ($SR = -0.426$, $t = -2.672$) and NOK ($SR = -0.391$, $t = -1.315$, with "IR v. Mkt" $t = -2.506$), and remains weak or marginal elsewhere (e.g., AUD, EUR, JPY, SEK). The new "IR v. LIN" column makes the underperformance explicit: RFF is significantly worse than LIN for CHE ($t = -2.587$), and negative (though not always significant) for several others. Importantly, tail risk rises under complexity: RFF maximum losses are elevated—e.g., CAD (about 15 s.d.), CHE (about 10.7 s.d.), AUD (about 10.1 s.d.)—and generally exceed those of LIN, underscoring fragility when the sample expands and nominal complexity ($c = P/T$) falls. LIN drawdowns are materially lower (typically 4–10 s.d.) and more in line with its steadier SR profile.

At the longest window ($T = 120$), the reversal is complete. Linear timing is most effective in CAD (LIN: $SR = 0.401$, $t = 2.494$), with additional positive though weaker SRs for CHE, EUR, and NOK; SEK and JPY are the notable laggards for LIN (negative or near zero). By contrast, RFF is broadly weak: SRs are negative for AUD, CHE, EUR, JPY, NOK, and SEK, and only small and insignificant for CAD and GBP. The "IR v. LIN" metric rarely favors RFF (mostly negative or near zero), indicating that any incremental value of complexity has evaporated in data-richer settings. Risk metrics reinforce this: RFF exhibits some of the largest drawdowns in the table—e.g., CHE (about 12.6 s.d.)—and more negative skew (e.g., CHE, SEK), whereas LIN's tail risk is comparatively contained for the few currencies where it performs best (e.g., CAD with Max Loss around 4.9 s.d.).

In sum, under traditional fundamentals, Ridge–RFF's advantages are localized to very small samples and a couple of currencies (CAD, JPY) where "IR v. LIN" confirms incremental content. As the training window grows from 12 to 60 and 120 months, linear timing becomes relatively stronger, whereas RFF performance turns persistently negative with higher drawdowns. The evolution across T is consistent with a limited, short-sample "virtue

Table 5: Market timing performance metrics for single currency portfolios under traditional fundamentals

		SR	t	IR v. Mkt	t	IR v. RW	t	IR v. LIN	t	Skew	Max Loss
Panel A: Training window size $T = 12$											
LIN	AUD	0.068	0.471	0.074	0.511	0.055	0.382			−3.193	11.082
	CAD	0.260	1.797*	0.264	1.822*	0.271	1.867*			2.023	4.466
	CHE	0.107	0.703	0.142	0.932	0.134	0.883			8.809	4.010
	EUR	−0.036	−0.175	−0.044	−0.215	−0.049	−0.239			−2.432	8.989
	GBP	−0.128	−0.884	−0.139	−0.957	−0.156	−1.074			−2.460	11.233
	JPY	0.155	1.071	0.144	0.991	0.120	0.827			2.870	7.004
	NOK	−0.129	−0.506	−0.163	−0.632	−0.200	−0.772			−1.448	6.268
	SEK	−0.081	−0.407	−0.091	−0.457	−0.082	−0.413			−0.746	5.535
RFF	AUD	0.000	−0.003	−0.031	−0.218	−0.011	−0.076	−0.011	−0.076	−2.090	11.078
	CAD	0.306	2.114**	0.294	2.024**	0.338	2.328**	0.255	1.753*	2.297	4.383
	CHE	0.036	0.234	0.008	0.054	0.001	0.007	0.046	0.299	−1.568	9.409
	EUR	−0.080	−0.397	−0.082	−0.405	−0.130	−0.637	−0.073	−0.357	−1.843	7.299
	GBP	0.075	0.519	0.060	0.415	0.049	0.339	0.131	0.902	1.319	5.688
	JPY	0.373	2.580***	0.354	2.442**	0.320	2.201**	0.350	2.414**	0.519	5.534
	NOK	0.123	0.484	0.053	0.208	0.154	0.595	0.193	0.750	−0.797	4.821
	SEK	0.022	0.110	0.017	0.086	0.022	0.108	0.090	0.452	−0.528	6.589
Panel B: Training window size $T = 60$											
LIN	AUD	0.135	0.896	0.152	1.008	0.156	1.033			0.508	10.527
	CAD	0.345	2.279**	0.344	2.273**	0.346	2.279**			1.575	5.152
	CHE	0.105	0.655	0.152	0.948	0.103	0.644			2.603	6.384
	EUR	−0.058	−0.260	−0.034	−0.151	−0.049	−0.221			−0.307	5.286
	GBP	0.015	0.102	−0.004	−0.026	−0.065	−0.431			3.644	5.603
	JPY	0.048	0.321	0.056	0.369	0.027	0.178			1.848	5.815
	NOK	−0.107	−0.360	−0.199	−0.660	−0.134	−0.448			0.496	4.824
	SEK	−0.242	−1.118	−0.230	−1.059	−0.256	−1.180			−2.862	8.271
RFF	AUD	−0.128	−0.850	−0.142	−0.943	−0.131	−0.868	−0.190	−1.261	−1.391	10.056
	CAD	−0.025	−0.163	−0.007	−0.049	−0.018	−0.116	−0.034	−0.221	−6.397	15.032
	CHE	−0.426	−2.672***	−0.491	−3.063***	−0.448	−2.802***	−0.414	−2.587***	−5.771	10.701
	EUR	−0.193	−0.871	−0.189	−0.850	−0.183	−0.823	−0.186	−0.834	−1.705	6.726
	GBP	0.096	0.635	0.062	0.408	−0.013	−0.089	0.106	0.700	7.461	6.552
	JPY	−0.060	−0.398	−0.073	−0.484	−0.093	−0.611	−0.068	−0.448	−1.569	7.802
	NOK	−0.391	−1.315	−0.757	−2.506**	−0.379	−1.262	−0.386	−1.291	−1.856	6.736
	SEK	−0.047	−0.216	−0.040	−0.184	−0.055	−0.252	0.123	0.566	−4.107	8.928
Panel C: Training window size $T = 120$											
LIN	AUD	0.039	0.244	0.042	0.261	0.044	0.277			1.166	7.568
	CAD	0.401	2.494**	0.401	2.490**	0.412	2.553**			1.417	4.938
	CHE	0.269	1.574	0.318	1.859*	0.274	1.600			6.779	6.753
	EUR	0.132	0.517	0.142	0.554	0.140	0.545			0.349	4.012
	GBP	0.064	0.399	0.068	0.422	−0.006	−0.037			2.307	5.923
	JPY	−0.033	−0.206	−0.017	−0.106	−0.049	−0.303			−3.012	11.045
	NOK	0.480	1.209	0.427	1.053	0.458	1.132			0.539	3.848
	SEK	−0.352	−1.423	−0.266	−1.068	−0.394	−1.585			−3.696	8.594
RFF	AUD	−0.119	−0.748	−0.119	−0.745	−0.125	−0.779	−0.180	−1.127	3.425	5.032
	CAD	0.130	0.807	0.136	0.846	0.120	0.742	0.026	0.161	−0.941	9.796
	CHE	−0.183	−1.075	−0.223	−1.301	−0.213	−1.245	−0.069	−0.404	−6.962	12.551
	EUR	−0.104	−0.408	−0.084	−0.326	−0.146	−0.568	−0.140	−0.543	0.551	4.100
	GBP	0.122	0.762	0.129	0.803	0.050	0.311	0.106	0.658	6.390	5.040
	JPY	−0.081	−0.506	−0.136	−0.848	−0.123	−0.765	−0.078	−0.485	−0.208	6.581
	NOK	−0.152	−0.383	−0.486	−1.199	−0.155	−0.384	−0.220	−0.541	−1.317	5.501
	SEK	−0.199	−0.804	−0.191	−0.767	−0.200	−0.805	0.033	0.132	−3.181	7.583

This table reports one-month-ahead out-of-sample market-timing strategy performance metrics for single-currency portfolios under the traditional monetary fundamentals. Results are presented for linear regression (LIN) and Ridge–RFF (RFF) models, evaluated relative to the market, the random walk (RW), and to each other. Reported statistics include the Sharpe ratio (SR) of each strategy with associated t-statistics for mean returns; Information Ratios (IR) relative to the market, LIN relative to RW, and RFF relative to both RW and LIN, each with corresponding t-statistics; as well as higher-moment and downside risk measures: skewness of returns (Skew) and the maximum loss (Max Loss) in standard deviation units. Statistical significance of mean returns and information ratios is denoted by *, **, and *** for the 10%, 5%, and 1% levels, respectively.

of complexity" that does not scale once more data become available and parsimony can capitalize on signal more reliably.

Table 6 evaluates timing strategies under the expanded set of monetary and non-monetary fundamentals. The results confirm that predictor richness does not translate into reliable benefits from added model complexity.

At the shortest horizon ($T = 12$), both models perform weakly overall. Linear regression delivers mostly negative Sharpe ratios, with particularly poor results for CAD, GBP, NOK, and SEK, each with negative and often marginally significant t-statistics. Large maximum losses—for instance, NOK suffers a drawdown of over 23 s.d.—highlight instability in LIN despite its simplicity. Ridge–RFF offers only isolated and statistically weak improvements, most notably for EUR ($SR = 0.216$, $t = 1.075$), while for most currencies its performance remains flat or negative. Importantly, the "IR v. LIN" entries are uniformly weak, suggesting that RFF does not extract incremental value from the richer predictor set even in small samples.

With medium-sized samples ($T = 60$), linear regression stabilizes. GBP is a clear standout: LIN achieves a Sharpe ratio of 0.404 ($t = 2.703$), significant at the 1% level, with supporting information ratios against the market and the random walk. Modest though less significant improvements are also observed for NOK and SEK. In contrast, Ridge–RFF systematically deteriorates. Across most currencies (e.g., AUD, CHE, EUR, NOK, SEK), RFF produces negative and often significant Sharpe ratios, coupled with elevated drawdowns. Notably, for EUR, RFF collapses to $SR = -0.284$ with "IR v. Mkt" $t = -1.337$, underscoring its fragility once longer training samples reduce nominal complexity. Tail risks remain consistently larger under RFF, while linear regression maintains more balanced drawdowns (typically between 5–10 s.d.).

At the longest horizon ($T = 120$), the advantage of parsimony is decisive. Linear regression strategies produce their strongest results across all predictor sets here: EUR ($SR = 0.648$, $t = 2.571$), GBP ($SR = 0.464$, $t = 2.924$), and AUD ($SR = 0.314$, $t = 1.980$) all record statistically significant Sharpe ratios, while NOK and SEK also deliver positive and marginally significant performance. By contrast, Ridge–RFF remains persistently weak, with negative or near-zero Sharpe ratios for most currencies and large drawdowns (e.g., JPY at 9.171 s.d., CHE at 8.364 s.d.). Even where RFF produces positive values (such as GBP or JPY), these gains are small and insignificant, and never competitive with linear regression.

Overall, the Monetary+ results reinforce the fragility of complexity in FX timing. Ridge–RFF occasionally produces marginally positive outcomes in small samples (e.g., EUR at $T = 12$), but these gains vanish quickly as training samples grow. Linear regression, by contrast, benefits systematically from the richer information set, delivering strong and statis-

Table 6: Market timing performance metrics for single currency portfolios under Monetary and nonmonetary (monetary+) fundamentals

		SR	t	IR v. Mkt	t	IR v. RW	t	IR v. LIN	t	Skew	Max Loss
Panel A: Training window size $T = 12$											
LIN	AUD	0.019	0.132	0.032	0.225	0.007	0.046			−1.372	8.976
	CAD	−0.201	−1.402	−0.240	−1.673*	−0.167	−1.159			−5.587	14.173
	CHE	−0.056	−0.368	−0.073	−0.479	−0.091	−0.601			−3.950	10.843
	EUR	0.092	0.455	0.098	0.486	0.077	0.383			2.043	7.212
	GBP	−0.187	−1.305	−0.197	−1.371	−0.259	−1.802*			−2.496	10.083
	JPY	−0.093	−0.649	−0.120	−0.833	−0.144	−1.002			−0.721	7.280
	NOK	−0.187	−1.309	−0.184	−1.283	−0.188	−1.311			−21.551	23.198
	SEK	−0.184	−1.286	−0.172	−1.199	−0.177	−1.234			−3.750	12.836
RFF	AUD	−0.139	−0.974	−0.145	−1.012	−0.141	−0.981	−0.138	−0.965	1.114	7.906
	CAD	0.033	0.232	0.030	0.210	0.044	0.306	0.053	0.365	1.390	7.645
	CHE	−0.026	−0.173	−0.038	−0.250	−0.038	−0.248	−0.028	−0.187	0.171	9.832
	EUR	0.216	1.075	0.219	1.087	0.209	1.036	0.212	1.049	1.400	5.051
	GBP	−0.020	−0.140	−0.018	−0.124	−0.064	−0.443	−0.001	−0.005	0.257	6.006
	JPY	−0.051	−0.352	−0.079	−0.546	−0.172	−1.196	−0.043	−0.298	−0.746	11.293
	NOK	−0.009	−0.063	−0.010	−0.067	−0.004	−0.026	−0.006	−0.043	−0.428	7.206
	SEK	0.004	0.028	0.014	0.099	−0.014	−0.101	0.011	0.074	0.458	6.978
Panel B: Training window size $T = 60$											
LIN	AUD	0.139	0.928	0.137	0.913	0.079	0.529			0.627	8.209
	CAD	−0.080	−0.532	−0.071	−0.472	−0.037	−0.249			−2.409	10.416
	CHE	−0.157	−0.989	−0.207	−1.300	−0.189	−1.191			−8.163	15.487
	EUR	0.051	0.233	0.044	0.199	0.094	0.427			−2.693	7.807
	GBP	0.404	2.703***	0.384	2.565**	0.360	2.399**			2.955	5.463
	JPY	0.083	0.555	0.072	0.482	0.066	0.437			−0.827	8.077
	NOK	0.192	1.285	0.178	1.189	0.193	1.288			−2.371	9.154
	SEK	0.209	1.400	0.223	1.485	0.200	1.335			−3.436	10.064
RFF	AUD	−0.247	−1.652*	−0.251	−1.672*	−0.257	−1.713*	−0.237	−1.581	−1.496	7.203
	CAD	−0.042	−0.278	−0.042	−0.280	−0.030	−0.199	−0.044	−0.291	0.727	5.670
	CHE	−0.184	−1.158	−0.227	−1.426	−0.189	−1.191	−0.166	−1.043	−2.768	10.402
	EUR	−0.284	−1.294	−0.295	−1.337	−0.278	−1.262	−0.280	−1.271	−4.224	10.132
	GBP	0.098	0.658	0.098	0.652	0.069	0.462	0.050	0.331	1.916	4.719
	JPY	−0.032	−0.211	−0.035	−0.230	−0.109	−0.724	−0.042	−0.280	−1.126	9.685
	NOK	−0.150	−1.001	−0.149	−0.992	−0.152	−1.013	−0.163	−1.085	−1.018	7.296
	SEK	−0.145	−0.967	−0.132	−0.877	−0.166	−1.110	−0.158	−1.050	−0.901	6.841
Panel C: Training window size $T = 120$											
LIN	AUD	0.314	1.980**	0.311	1.958*	0.298	1.877*			2.470	4.331
	CAD	0.106	0.668	0.112	0.706	0.162	1.020			1.069	9.341
	CHE	0.027	0.160	−0.010	−0.061	0.011	0.062			−5.780	14.288
	EUR	0.648	2.571**	0.667	2.632***	0.658	2.595***			1.126	4.795
	GBP	0.464	2.924***	0.501	3.153***	0.435	2.731***			2.944	4.977
	JPY	0.079	0.499	0.060	0.375	0.045	0.283			0.975	5.039
	NOK	0.308	1.941*	0.306	1.927*	0.327	2.058**			−2.442	11.616
	SEK	0.302	1.904*	0.305	1.920*	0.305	1.915*			0.078	6.926
RFF	AUD	−0.193	−1.218	−0.193	−1.215	−0.202	−1.271	−0.164	−1.028	−1.121	7.100
	CAD	−0.074	−0.469	−0.073	−0.461	−0.072	−0.456	−0.086	−0.541	−1.236	5.849
	CHE	−0.065	−0.384	−0.084	−0.496	−0.070	−0.413	−0.064	−0.375	−1.202	8.364
	EUR	−0.263	−1.045	−0.286	−1.127	−0.239	−0.942	−0.182	−0.705	−2.314	7.278
	GBP	0.143	0.903	0.144	0.903	0.118	0.741	0.103	0.645	1.976	3.658
	JPY	0.035	0.222	0.032	0.200	−0.061	−0.381	0.027	0.171	−1.284	9.171
	NOK	0.035	0.220	0.037	0.231	0.040	0.254	0.009	0.057	−0.563	7.446
	SEK	−0.062	−0.394	−0.059	−0.374	−0.085	−0.534	−0.081	−0.510	−0.986	5.501

This table reports one-month-ahead out-of-sample market-timing strategy performance metrics for single-currency portfolios under monetary+ fundamentals. See, notes for Table 5.

tically significant Sharpe ratios for multiple currencies at $T = 120$. Importantly, drawdown patterns show that RFF strategies entail higher downside risk across all horizons, further weakening the economic case for complexity. In this predictor environment, parsimony is not only more reliable but also economically safer, underscoring the limited virtue of complexity once richer fundamentals and longer histories are available.

Table 7 presents timing strategy performance under Taylor rule fundamentals. Compared to traditional and Monetary+ predictors, the Taylor rule set provides the most consistent evidence of economic value for linear regression, particularly as the sample size grows, while Ridge–RFF offers selective short-sample advantages that erode quickly.

In the short-window case ($T = 12$), linear regression produces a mixed profile. While CAD and CHE perform poorly (negative Sharpe ratios around –0.15 and –0.04), several currencies show encouraging results. Notably, JPY achieves a Sharpe ratio of 0.249 ($t = 1.740$), significant at the 10% level, while GBP, AUD, NOK, and SEK all record small but positive Sharpe ratios. Ridge–RFF delivers competitive, and in some cases stronger, results: JPY (SR=0.257, $t = 1.796$) and AUD (SR=0.159, $t = 1.114$) show robust short-horizon performance, and CHE and EUR also post positive ratios. Nevertheless, these gains come at the cost of greater instability in other currencies: for example, CAD underperforms markedly, with negative Sharpe ratios across specifications and a large maximum loss exceeding 12 s.d.

At the medium horizon ($T = 60$), linear regression consolidates its advantage. Significant gains emerge for GBP (SR=0.293, $t = 1.961$), NOK (SR=0.261, $t = 1.746$), and EUR (SR=0.237, $t = 1.081$), while AUD and JPY also show positive, though less significant, results. Ridge–RFF produces respectable outcomes for JPY (SR=0.274, $t = 1.835$) and AUD (SR=0.175, $t = 1.174$), but overall performance weakens elsewhere: GBP, NOK, and SEK all fall into negative territory, with drawdowns consistently higher than those observed under linear regression. For example, CAD records a maximum loss of 12.4 s.d. under RFF versus a much smaller 10.9 s.d. under LIN. These patterns highlight the fragility of complexity in medium-sized samples.

At the longest horizon ($T = 120$), the contrast is most pronounced. Linear regression delivers its strongest and most statistically reliable results across all fundamentals. EUR achieves a Sharpe ratio of 0.528 ($t = 2.094$), while AUD also performs strongly (SR=0.354, $t = 2.235$). Positive and significant gains extend to CAD and NOK, while most other currencies remain at least weakly positive. Ridge–RFF, by contrast, deteriorates markedly. Apart from JPY (SR=0.284, $t = 1.787$) and CHE (SR=0.238, $t = 1.406$), most currencies record negative Sharpe ratios, with especially poor results for EUR (SR=–0.234, $t = -0.927$). Moreover, maximum losses for RFF remain elevated, often exceeding those of linear regression, underscoring the higher tail risk associated with complexity.

Table 7: Market timing performance metrics for single currency portfolios under Taylor rule fundamentals

		SR	t	IR v. Mkt	t	IR v. RW	t	IR v. LIN	t	Skew	Max Loss
Panel A: Training window size $T = 12$											
LIN	AUD	0.085	0.594	0.088	0.615	0.067	0.469			3.437	6.282
	CAD	−0.147	−1.027	−0.154	−1.076	−0.103	−0.719			−1.190	6.819
	CHE	−0.038	−0.251	−0.045	−0.295	−0.106	−0.702			0.794	6.075
	EUR	−0.018	−0.090	−0.020	−0.101	−0.056	−0.277			0.005	7.402
	GBP	0.136	0.952	0.137	0.955	0.061	0.425			1.662	4.356
	JPY	0.249	1.740*	0.227	1.577	0.150	1.041			1.472	4.484
	NOK	0.059	0.409	0.044	0.309	0.076	0.532			4.001	5.761
	SEK	0.121	0.846	0.086	0.601	0.083	0.580			1.239	6.421
RFF	AUD	0.159	1.114	0.152	1.059	0.163	1.136	0.140	0.977	0.732	7.139
	CAD	−0.094	−0.656	−0.089	−0.620	0.130	0.904	−0.067	−0.464	−4.513	12.845
	CHE	0.190	1.255	0.182	1.202	0.151	0.995	0.222	1.466	0.421	5.209
	EUR	0.134	0.665	0.131	0.652	0.151	0.748	0.152	0.752	−0.539	4.787
	GBP	0.018	0.124	0.025	0.174	−0.084	−0.582	−0.044	−0.304	−0.796	6.812
	JPY	0.257	1.796*	0.244	1.699*	0.074	0.515	0.167	1.158	−0.885	10.405
	NOK	−0.152	−1.059	−0.147	−1.022	−0.177	−1.236	−0.175	−1.218	−1.440	6.636
	SEK	−0.006	−0.041	−0.006	−0.040	−0.077	−0.534	−0.043	−0.302	−0.457	5.613
Panel B: Training window size $T = 60$											
LIN	AUD	0.120	0.801	0.110	0.738	0.042	0.280			2.612	6.044
	CAD	0.053	0.355	0.070	0.469	0.142	0.951			0.645	10.855
	CHE	0.047	0.294	−0.047	−0.293	0.025	0.160			−4.328	13.051
	EUR	0.237	1.081	0.263	1.195	0.327	1.485			−3.892	10.332
	GBP	0.293	1.961**	0.280	1.873*	0.230	1.535			1.625	5.190
	JPY	0.208	1.391	0.195	1.301	0.104	0.691			0.762	6.647
	NOK	0.261	1.746*	0.248	1.653*	0.292	1.949*			0.724	4.689
	SEK	0.100	0.669	0.079	0.530	0.032	0.213			−0.979	11.530
RFF	AUD	0.175	1.174	0.171	1.144	0.102	0.682	0.135	0.903	0.070	8.271
	CAD	−0.114	−0.765	−0.110	−0.734	0.074	0.496	−0.141	−0.941	−4.616	12.426
	CHE	0.193	1.220	0.169	1.061	0.203	1.277	0.188	1.182	0.144	5.382
	EUR	−0.027	−0.123	−0.024	−0.110	0.033	0.150	−0.150	−0.679	−1.045	5.318
	GBP	−0.065	−0.432	−0.048	−0.323	−0.167	−1.112	−0.157	−1.048	−0.796	6.287
	JPY	0.274	1.835*	0.265	1.769*	0.089	0.595	0.212	1.411	−0.798	10.498
	NOK	−0.083	−0.557	−0.083	−0.554	−0.117	−0.779	−0.190	−1.268	−1.243	6.667
	SEK	0.011	0.071	0.012	0.077	−0.084	−0.562	−0.025	−0.170	−0.310	5.489
Panel C: Training window size $T = 120$											
LIN	AUD	0.354	2.235**	0.351	2.212**	0.349	2.198**			2.173	6.429
	CAD	0.147	0.930	0.159	0.999	0.228	1.433			2.181	10.080
	CHE	0.080	0.470	0.052	0.307	0.067	0.396			−0.329	4.905
	EUR	0.528	2.094**	0.572	2.259**	0.563	2.219**			0.926	4.678
	GBP	0.119	0.753	0.126	0.796	0.043	0.270			0.005	7.143
	JPY	0.040	0.255	0.036	0.224	−0.055	−0.344			−3.554	11.821
	NOK	0.156	0.983	0.156	0.984	0.186	1.171			1.316	4.698
	SEK	−0.007	−0.042	0.001	0.008	0.000	−0.002			−1.364	7.251
RFF	AUD	0.179	1.132	0.177	1.115	0.135	0.853	0.027	0.172	0.175	7.937
	CAD	−0.125	−0.792	−0.125	−0.785	0.042	0.262	−0.174	−1.095	−4.496	11.805
	CHE	0.238	1.406	0.227	1.333	0.253	1.488	0.225	1.324	0.539	5.624
	EUR	−0.234	−0.927	−0.233	−0.920	−0.156	−0.614	−0.372	−1.453	−1.180	4.823
	GBP	−0.086	−0.542	−0.095	−0.600	−0.179	−1.128	−0.122	−0.767	−0.738	6.241
	JPY	0.284	1.787*	0.267	1.676*	0.071	0.445	0.283	1.781*	−1.241	10.610
	NOK	−0.089	−0.563	−0.089	−0.562	−0.079	−0.499	−0.135	−0.849	−1.222	6.306
	SEK	0.011	0.071	0.013	0.080	−0.071	−0.445	0.013	0.081	−0.314	5.205

This table reports one-month-ahead out-of-sample market-timing strategy performance metrics for single-currency portfolios under Taylor rule fundamentals. See, notes for Table 5.

Overall, the Taylor rule results are consistent with a broader narrative: Ridge–RFF can provide isolated improvements in very small samples, particularly for JPY and AUD, but these advantages weaken and often reverse as the sample size expands. By contrast, linear regression steadily strengthens with longer histories, producing statistically significant gains for multiple major currencies at $T = 120$. Importantly, RFF strategies entail higher and more volatile drawdowns across all horizons, further reducing their appeal in practical portfolio settings. The evidence suggests that under Taylor rule fundamentals, parsimony dominates complexity: linear regression extracts the predictive content of these fundamentals in a stable and economically meaningful manner, while Ridge–RFF adds little and frequently introduces greater risk.

Taken together, the single-currency results paint a consistent picture of the limited and fragile nature of complexity in exchange rate forecasting. Ridge–RFF offers selective benefits in very small samples with richer fundamentals, echoing the notion of "benign overfitting" whereby regularization allows complex models to exploit weak signals. Yet these advantages are narrow in scope: as sample sizes grow, Ridge–RFF not only fails to add value but often underperforms both linear regression and the random walk, with larger volatility and deeper drawdowns. By contrast, linear regression steadily improves with richer fundamentals and longer samples, culminating in robust, statistically significant gains under Taylor-rule fundamentals at $T = 120$. This progression underscores the broader lesson that parsimony and data availability ultimately dominate complexity in the context of exchange rate predictability. The "virtue of complexity," while occasionally visible, does not translate into systematic or durable economic value; rather, the enduring strength lies in well-specified linear models and, in many contexts, the random walk benchmark.

5.2 Equal-Weighted Currency Portfolio

Table 8 reports performance metrics for the equal-weighted market-timing strategy under RW, Linear regression, and Ridge-RFF.[5] With the shortest training window of $T = 12$, Ridge-RFF demonstrates some promise in generating economically meaningful portfolio returns, while linear regression performs poorly relative to both RFF and the random walk. For example, under Taylor rule fundamentals, Ridge-RFF achieves the highest Sharpe ratio (0.503) with a statistically significant t-statistic (2.501), clearly outperforming both the linear model and RW (SR = 0.232, $t = 1.181$). Similarly, under traditional fundamentals, RFF delivers a

[5]Because the equal-weighted portfolio requires all eight currencies, the evaluation period begins only after the euro (EUR) enters the sample in 1999. Results using an alternative specification—where the portfolio includes seven currencies before 2000 and all eight thereafter—are very similar and therefore omitted for brevity, but available upon request.

Table 8: Equal Weighted Portfolio Market timing trading strategy performance

		SR	t	IR v. Mkt	t	IR v. RW	t	IR v. LIN	t	Skew	Max Loss
Panel A: Training window size $T = 12$											
	RW	0.232	1.181	0.232	1.149					−0.469	7.413
Trad	LIN	−0.499	−1.953*	−0.500	−1.946*	−0.516	−2.008**			−1.823	5.909
	RFF	0.446	1.747*	0.449	1.747*	0.351	1.349	0.475	1.851*	−0.026	3.488
Mon+	LIN	−0.256	−1.276	−0.251	−1.242	−0.266	−1.318			−13.900	15.973
	RFF	−0.020	−0.099	−0.006	−0.032	−0.240	−1.173	−0.003	−0.016	0.179	4.545
Taylor	LIN	0.200	0.997	0.213	1.056	0.092	0.456			0.476	5.495
	RFF	0.503	2.501**	0.516	2.557**	0.164	0.802	0.450	2.226**	0.039	3.891
Panel B: Training window size $T = 60$											
	RW	0.345	1.701*	0.345	1.566					1.430	3.095
Trad	LIN	−0.367	−1.236	−0.363	−1.213	−0.363	−1.213			0.272	3.894
	RFF	0.095	0.319	0.108	0.360	0.045	0.150	0.135	0.450	0.036	2.434
Mon+	LIN	0.321	1.464	0.313	1.414	0.303	1.366			−0.233	3.665
	RFF	−0.286	−1.303	−0.250	−1.133	−0.372	−1.667*	−0.382	−1.715*	0.340	2.531
Taylor	LIN	0.467	2.129**	0.505	2.287**	0.336	1.516			−0.113	3.551
	RFF	0.387	1.761*	0.402	1.818*	0.084	0.377	0.217	0.974	0.275	3.321
Panel C: Training window size $T = 120$											
	RW	0.361	1.413	0.361	1.427					1.671	2.858
Trad	LIN	−0.032	−0.081	−0.030	−0.074	−0.049	−0.123			−0.535	3.802
	RFF	−0.663	−1.669*	−0.671	−1.658*	−0.678	−1.682*	−0.668	−1.659*	0.013	2.803
Mon+	LIN	0.298	1.181	0.303	1.196	0.317	1.245			−0.048	5.237
	RFF	−0.114	−0.452	−0.114	−0.452	−0.258	−1.008	−0.250	−0.972	0.449	2.359
Taylor	LIN	0.220	0.875	0.217	0.856	0.164	0.646			−0.478	4.723
	RFF	0.496	1.970**	0.506	1.997**	0.283	1.110	0.465	1.831*	0.544	2.906

This table reports performance metrics for equal-weighted currency market-timing portfolios constructed from one-month-ahead forecasts under the random walk (RW), linear (LIN), and Ridge–RFF (RFF) models across three sets of fundamentals, traditional monetary (Trad), monetary+ (Mon+), and Taylor rule (Taylor). Reported statistics include the Sharpe ratio (SR) with t-statistics for mean returns, the Information Ratio relative to the market (IR v. Mkt), the Information Ratio of LIN relative to RW, and the Information Ratio of RFF relative to both RW and LIN, each with corresponding t-statistics. We also report skewness of returns (Skew) and the maximum portfolio loss (Max Loss) in standard deviation units. Statistical significance of t-statistics is denoted by *, **, and *** at the 10%, 5%, and 1% levels, respectively.

positive Sharpe ratio (0.446, $t = 1.747$) and significant information ratios relative to OLS, while the linear model produces negative Sharpe ratios (–0.499, $t = -1.953$), substantially underperforming RW. By contrast, under monetary+ fundamentals, both models fail to improve upon RW: Ridge-RFF generates a near-zero Sharpe ratio (–0.020, insignificant), and OLS produces strongly negative performance (–0.256, $t = -1.276$), while RW remains positive (SR = 0.232).

Importantly, the drawdown evidence reinforces the fragility of these results. Ridge-RFF under Taylor fundamentals exhibits relatively modest losses (Max Loss ≈ 3.89), while OLS portfolios under monetary+ fundamentals suffer extremely large drawdowns (Max Loss ≈ 15.97) with very large negative skew. Even under traditional fundamentals, OLS drawdowns are deeper than those of RFF or RW. Thus, while Ridge-RFF occasionally improves on OLS and even outperforms RW in settings such as Taylor fundamentals, these gains are neither broad-based nor robust across predictor sets, and the strategy remains vulnerable to sharp losses depending on model choice and information set.

With larger samples, performance dynamics shift. In the evaluation period corresponding to $T = 60$ in Panel B of Table 8, the random walk benchmark sets a higher bar (SR = 0.345, $t = 1.70$). Linear regression begins to show meaningful economic value under Taylor rule fundamentals (SR = 0.467, $t = 2.13$), significantly outperforming both the random walk and RFF. Ridge-RFF continues to deliver moderate gains under Taylor rule fundamentals (SR = 0.387, $t = 1.76$), but its edge over linear regression largely disappears, with information ratios against linear model weak or insignificant. Under traditional and monetary+ fundamentals, RFF underperforms (Sharpe ratios near zero or negative), while linear regression strategies provide at best modest gains. Drawdowns provide further evidence of fragility: although linear portfolios under Taylor fundamentals and RW achieve positive Sharpe ratios in this training window, they also suffer moderate losses (≈ 3.55 to 3.10), while RFF portfolios show similar downside risk (≈ 3.32). In contrast, RFF under traditional or monetary+ fundamentals yields both weak returns and low information resilience, with even negative and statistically significant information ratios relative to both random walk and linear strategies.

At the longest training horizon of $T = 120$ reported in Panel C of the Table, the reversal is stark. The random walk continues to yield stable if modest returns (SR = 0.361), while Ridge-RFF generally deteriorates, producing negative or insignificant Sharpe ratios under traditional and monetary+ fundamentals. For example, RFF under traditional fundamentals generates a Sharpe ratio of –0.66 ($t = -1.67$), significantly underperforming both the random walk and linear regression, with drawdowns (≈ 2.8) comparable to linear but without compensating returns. By contrast, linear regression produces positive Sharpe ratios under monetary+ (0.30) and Taylor rule fundamentals (0.20). Importantly, Ridge-RFF under Taylor

rule fundamentals still produces a positive and statistically significant Sharpe ratio (0.496, $t = 1.97$) with better information ratio relative to linear portfolio and relatively less elevated drawdown risk, suggesting some residual benefit of complexity in this setting.

Overall, the portfolio results align closely with the out-of-sample forecast evidence. At short training windows ($T = 12$), Ridge-RFF can exploit extreme nominal complexity to deliver economically meaningful trading gains, particularly under Taylor rule fundamentals, with comparatively controlled drawdowns. At intermediate horizons ($T = 60$), the advantage shifts toward linear regression, which produces higher Sharpe ratios and stronger statistical significance, especially under Taylor rule predictors, though both linear and RFF portfolios still exhibit moderate downside risks. By $T = 120$, Ridge-RFF strategies generally underperform both linear regression and the random walk, with the exception of modest gains under Taylor fundamentals, and all models continue to be constrained by drawdowns that limit their practical applicability.

Taken together, these results reinforce the broader conclusion that the "virtue of complexity" in FX prediction is fragile and sample-dependent. Complexity delivers relative improvements in very small samples, but these do not persist as sample size grows. Moreover, the drawdown evidence highlights that even when complexity improves Sharpe ratios, the resulting strategies remain vulnerable to sizeable losses, which undermines their appeal in practical portfolio settings. In terms of economic value, parsimony—captured by simple linear regression or even the random walk—remains more robust and reliable for constructing profitable and risk-managed market-timing portfolios.

In comparing the equal-weighted portfolio performance with the single-currency portfolio results reported in Appendix 5.1, a consistent picture emerges. The equal-weighted portfolio tends to smooth out idiosyncratic gains and losses that appear in single-currency strategies, particularly for Ridge-RFF at short horizons. For example, while single-currency portfolios occasionally deliver strong Sharpe ratios under Ridge-RFF in small samples (e.g., JPY or CAD at $T = 12$), the equal-weighted portfolio translates these into only modest gains, with performance highly dependent on predictor choice. Likewise, linear regression strategies that perform reasonably well for select currencies under Taylor rule fundamentals in larger samples ($T = 60$ or $T = 120$) produce more muted gains once aggregated into the equal-weighted portfolio. Importantly, the drawdown evidence shows that portfolio aggregation reduces extreme tail risks present in single-currency strategies (e.g., very large losses for NOK or SEK), but at the same time it dilutes pockets of predictability that individual currencies occasionally display. Taken together, the results indicate that while individual currencies may at times exhibit local predictability under complex models, these signals are too weak or

inconsistent to scale into robust portfolio-level profits, leaving the random walk and simple linear benchmarks difficult to beat in practice.

Overall, market timing portfolio results align closely with the out-of-sample forecast evidence. At short training windows ($T = 12$), Ridge-RFF can exploit extreme nominal complexity to deliver economically meaningful trading gains, particularly under Taylor rule fundamentals. At intermediate horizons ($T = 60$), the advantage shifts toward linear regression, which produces higher Sharpe ratios and stronger statistical significance, especially under Taylor rule predictors. By $T = 120$, Ridge-RFF strategies generally underperform both linear regression and the random walk, with the exception of modest gains under Taylor fundamentals.

Overall, these results reinforce the broader conclusion that the "virtue of complexity" in FX prediction is fragile and sample-dependent. Complexity delivers relative improvements in very small samples, but these do not persist as sample size grows. In terms of economic value, parsimony—captured by simple linear regression or even the random walk—remains more robust and reliable for constructing profitable market-timing portfolios.

The portfolio evidence closely mirrors the statistical forecast results. At short training windows, Ridge-RFF capitalizes on high nominal complexity to outperform linear regression in terms of both forecast accuracy and portfolio Sharpe ratios, particularly under Taylor rule fundamentals. Yet, these improvements remain fragile and never extend to consistent outperformance of the random walk. At intermediate horizons, linear regression gains ground statistically—reflected in smaller negative or occasionally positive OOS R^2 values and significant CW test rejections—and this translates directly into superior portfolio performance, with higher Sharpe ratios and stronger information ratios relative to both Ridge-RFF and the random walk. At long horizons, the pattern is decisive: Ridge-RFF underperforms on both statistical and economic fronts, while linear regression delivers modest but more reliable gains, at times even surpassing the random walk benchmark.

Taken together, these findings underscore a critical point: statistical evidence of predictability translates into economic value only when it is robust across sample sizes and predictor sets. Complexity through Ridge-RFF provides localized benefits in very small samples, but fails to deliver durable portfolio gains. By contrast, parsimony—embodied by linear regression and the random walk—offers greater consistency, reinforcing the conclusion that the "virtue of complexity" is limited in FX markets, both statistically and economically.

Across predictor sets and horizons, the evidence in FX is clear. At the most complex setting ($T = 12$), Ridge–RFF occasionally improves upon OLS, but neither model delivers systematic gains over the random walk (RW) benchmark. With $T = 60$, OLS tightens markedly—yielding smaller (sometimes positive) OOS R^2 and sporadic CW rejections in

its favor—while Ridge–RFF's advantages dissipate. At $T = 120$, the pattern is decisive: Ridge–RFF underperforms OLS and fails to beat RW; OLS is modestly better but still rarely dominant. Portfolio-level results echo the forecast statistics: when complexity is maximized and samples are shortest, Ridge–RFF can raise Sharpe ratios relative to OLS under specific fundamentals, yet these local gains do not scale into robust outperformance of RW; as T rises, OLS becomes the economically stronger (and more stable) alternative, and Ridge–RFF's edge vanishes. Taken together, RW remains hard to beat in FX at the monthly horizon, and the "virtue" of high nominal complexity is, at best, fragile and sample-dependent.

These findings connect tightly to the evolving debate on complexity in equity return prediction. First, Kelly and Malamud (2025) respond to recent critiques by clarifying that the central object is the complexity ratio $c = P/T$, distinguishing nominal from effective complexity and emphasizing how implicit regularization can allow large models to learn even in small samples; they further stress that VoC is a theory of out-of-sample performance under misspecification and is not about recovering a "true" model. They also highlight data features—"concentration" and "alignment"—that govern whether performance should rise with complexity. In our FX setting, the absence of systematic gains relative to RW as T grows suggests that either concentration, alignment, or both, differ materially from the equity contexts where VoC curves slope upward, consistent with their framework.

Second, Nagel (2025) argues that with $P >> T$ and highly persistent predictors, ridgeless RFF forecasts mechanically approximate a volatility-timed momentum rule—weights become a similarity kernel over the last T returns—so apparent small-sample "learning" largely reflects recency and volatility timing rather than extraction of genuine signals. Our FX results are potentially consistent with this interpretation: at $T = 12$, Ridge–RFF can outperform OLS (a local advantage) but does not consistently defeat RW; at larger T, the momentum-like edge recedes and parsimony wins, aligning with the view that small-T overparameterization explores only a thin subspace and cannot reliably uncover stable predictability.

Third, Buncic (2025) revisits Kelly et al. (2024)'s empirics and shows that two implementation choices—zero-intercept and aggregation of RFF draws—shape the monotone VoC patterns; when an intercept is included and aggregation is handled coherently, simpler (or mildly regularized) models often dominate. Our FX evidence parallels that spirit without including an intercept: as T expands, OLS (parsimonious, with effectively lower c) overtakes RFF, and portfolio performance improves without escalating nominal complexity. Notably, all three equity-based papers debate buy-and-hold versus timing and do not employ the RW forecasting benchmark we adopt in FX; thus, our findings contribute new evidence under a benchmark that is notoriously difficult to surpass in currencies.

In FX, complexity's benefits are not universal: when $c = P/T$ is extreme and T is short, Ridge–RFF can look relatively good versus OLS but not versus RW; as effective sample size grows, parsimony becomes economically preferable. This pattern is consistent with theories that allow small-sample learning under high complexity but predict data-dependent VoC curves, and with critiques highlighting the mechanical momentum content of overparameterized, short-window RFF. The upshot is that domain (FX vs. equities), sample length, and the benchmark all matter. Our results therefore refine the "virtue of complexity" claim by showing its limits in FX: when judged against a strong RW comparator, complexity delivers—at most—localized advantages that do not generalize into durable economic value.

6 Conclusions

This paper investigates whether the recently proposed "virtue of complexity"—originally documented in equity return prediction—extends to foreign exchange rate forecasting. Using nonlinear Ridge regressions augmented with Random Fourier Features (RFF), we conduct a comprehensive 1-month ahead out-of-sample forecasting exercise under alternative sets of economic fundamentals, across rolling training windows of 12, 60, and 120 months. Benchmarks include both traditional linear regression models and the random walk, with additional evaluation through market-timing portfolio strategies.

The results provide a measured perspective on the role of complexity in FX forecasting. With short training windows, where nominal complexity ($T = c/T$) is highest, Ridge–RFF can deliver relative gains over linear regression, including lower MSPE ratios and improved Diebold–Mariano statistics in select currencies. Yet, these improvements are fragile and never translate into robust gains against the random walk benchmark. As the training window expands, Ridge–RFF performance deteriorates—first losing parity with linear regression at $T = 60$ and then decisively underperforming at $T = 120$. In contrast, linear models gradually improve with larger samples, occasionally outperforming the random walk under richer predictor sets (especially under Monetary+ fundamentals, supporting recent results in Engel and Wu, 2024), though the benchmark remains difficult to beat consistently. Market-timing portfolio results reinforce this message: RFF-based strategies show local gains in very small samples, especially under Taylor rule fundamentals, but linear strategies provide more stable and economically meaningful returns as sample sizes grow.

Relative to the recent literature, our findings add several insights. Consistent with Kelly et al. (2024), we observe that complexity can yield localized benefits in small samples with richer fundamentals, reflecting the idea of "benign overfitting." However, our results show that these benefits are limited in scope and highly sensitive to training sample size, predictor choice,

and benchmark. Importantly, the random walk—absent from equity-based studies—remains the dominant comparator in FX, and neither Ridge–RFF nor linear regression consistently surpass it. This highlights the domain-specific limits of complexity's appeal.

Our evidence also connects to the evolving debate. Kelly and Malamud (2025) clarify that the key distinction is between nominal complexity (p/T) and effective complexity (parameters effectively used after shrinkage). In their framework, nominal complexity remains central for understanding out-of-sample performance. Our results suggest that, in FX, high nominal complexity does not translate into effective predictive gains against the random walk, particularly as sample size grows. Instead, parsimony appears to dominate at longer training windows. Meanwhile, the critiques of Nagel (2025) and Buncic (2025) resonate strongly with our findings. Nagel argues that RFF-based forecasts often mimic volatility-timed momentum rules in short samples, which aligns with our evidence that Ridge–RFF advantages fade quickly as T expands. Buncic shows that methodological restrictions can exaggerate complexity's benefits, and when relaxed, simpler models often prevail—mirroring our result that OLS reasserts itself in larger samples.

Taken together, our analysis suggests three key lessons. First, complexity in FX forecasting is conditional and fragile, offering advantages only in small samples with specific predictor sets, and never consistently over the random walk. Second, linear regression retains relevance: with richer fundamentals and larger samples, it often provides more robust statistical and economic performance. Third, the persistence of the random walk benchmark underscores the challenges of FX predictability and the risks of overstating the benefits of nonlinear complexity.

For researchers, these findings highlight the importance of benchmarking complex models against both linear regressions and the random walk, and of linking statistical performance to economic value. For practitioners and policymakers, they underscore the enduring appeal of transparency, parsimony, and robustness in FX forecasting—where signals are weak, benchmarks are strong, and the supposed virtue of complexity proves sharply limited.

Future research should build on these insights by probing further into the specific mechanisms behind complexity gains, investigating alternative nonlinear structures beyond RFF, and more carefully accounting for methodological artifacts that may confound empirical results. Moreover, adaptive or hybrid models that intelligently blend simplicity and complexity in response to data conditions may hold promise for improving forecast accuracy in a robust and interpretable manner.

References

Amat, O. (2018). Forecasting exchange rates using fundamentals and machine learning. *International Journal of Forecasting*, 34(1):123–142.

Bacchetta, M. and van Wincoop, E. (2004). Exchange rate misalignment in emerging market economies. *Journal of International Economics*, 62(1):67–95.

Buncic, D. (2025). Simplified: A closer look at the virtue of complexity in return prediction. Working Paper.

Cheung, Y. L., Chinn, M. D., and Pascual, A. (2005). Exchange rate forecasting: The role of fundamentals. *Journal of International Economics*, 65(2):123–144.

Cheung, Y. L., Chinn, M. D., and Pascual, A. (2017). Exchange rate forecasting revisited: New evidence on fundamental models. *Journal of International Money and Finance*, 72:150–170.

Engel, C. and West, K. D. (2012). Exchange rates and fundamentals: A panel data approach. *Journal of Political Economy*, 120(2):351–376.

Engel, C. and Wu, S. P. Y. (2024). Exchange rate models are better than you think, and why they didn't work in the old days. Technical Report Working Paper 32808, National Bureau of Economic Research. NBER Working Paper Series.

Filippou, I., Rapach, D. E., Taylor, M. P., and Zhou, G. (2025). Economic fundamentals and short-run exchange rate prediction: A machine-learning perspective. Working Paper.

Gilchrist, S. and Zakrajsek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720.

Ince, U. and Kubler, F. (2014). Panel data approaches in exchange rate forecasting. *Journal of International Money and Finance*, 38:125–140.

Kelly, B. and Malamud, S. (2025). Understanding the virtue of complexity.

Kelly, B., Malamud, S., and Zhou, K. (2024). The virtue of complexity in return prediction. *Journal of Finance*, 79(1):451–486.

Kouwenberg, R. et al. (2017). Time-varying predictability in exchange rates. *Journal of International Economics*, 108:509–517.

Meese, R. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14(1-2):3–24.

Meng, F. et al. (2024). Deep learning for exchange rate forecasting: A transformer approach. *Journal of Financial Data Science*, 6(1):45–67.

Molodtsova, I. and Papell, D. (2009). Exchange rate forecasting: Model evaluation and fundamentals. *Journal of International Money and Finance*, 28(3):465–487.

Nagel, S. (2025). Seemingly virtuous complexity in return prediction. Working Paper.

Pfahler, B. (2022). Machine learning models for exchange rate forecasting. Working Paper, IMF.

Rahimi, A. and Recht, B. (2007). Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20.

Rahimi, A. and Recht, B. (2008). Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning. *Advances in neural information processing systems*, 21.

Rossi, B. (2013). Exchange rate predictability: Recent evidence. *Journal of International Money and Finance*, 34(4):987–1003.

Zorzi, A. C., Muck, T., and Rubaszek, A. (2015). The role of purchasing power parity in exchange rate forecasting. *Journal of International Economics*, 96:150–167.

Appendices

A Data

As discussed in Section 3, our empirical analysis employs three sets of economic fundamentals to analyze U.S. dollar exchange rates against eight major currencies. End-of-month nominal exchange rate series are obtained from the IMF International Financial Statistics (IFS). Consumer Price Index (CPI) and industrial production index data for the United States (home country) and foreign countries are also sourced from IFS. Inflation rates are calculated as the log-difference of the CPI over the preceding 12-month period. Monthly industrial production indices serve as proxies for each country's output level.

To measure the money supply, we utilize the monetary aggregate M0, chosen for its consistent availability across the countries in our study. The end-of-month money supply data are retrieved from Haver Analytics. Nominal 3-month government bond interest rates are obtained from the Global Financial Database (GFD) and Federal Reserve Economic Data (FRED). The specific interest rate series from GFD include ITAUS3D (AUD), ITCAN3D (CAD), ITEUR3D (EUR), ITJPN3D (JPY), ITNOR3D (NOK), ITSWE3D (SEK), ITCHE3D (CHF), and ITGBR3D (GBP), while DTB3 (US) is obtained from FRED.

We construct a global risk aversion variable, denoted as Risk, by extracting the first principal component from the same five risk measures utilized by Engel and Wu (2024). These risk measures include spreads from Gilchrist and Zakrajsek (2012), Moody's Aaa and Baa corporate bond yields minus the Federal Funds rate spreads (FRED series: AAAFF and BAAFF), and Moody's Aaa and Baa corporate bond yields minus the 10-Year Treasury yield (FRED series: AAA10Y and BAA10Y). The Gilchrist and Zakrajšek (2012) spreads are available from `https://www.federalreserve.gov/econres/notes/feds-notes/updating-the-recession-risk-and-the-excess-bond-premium-20161006.html`, whereas the remaining four spreads (AAAFF, BAAFF, AAA10Y, BAA10Y) are downloaded from FRED.

Additionally, U.S. trade balance and GDP data are sourced from FRED. The trade balance data is available at monthly frequency starting from 1992 (FRED series: BOPGSTB) and quarterly prior to 1992 (BOPBGS). Quarterly trade balance and GDP series are converted into monthly frequency through linear interpolation. Finally, we compute the output gap for the United States and foreign countries using quarterly GDP data. We first apply the Hodrick-Prescott (HP) filter with a smoothing parameter of 1600, then convert the quarterly output gap series into monthly frequency through linear interpolation.

The availability of a complete dataset varies by currency, depending on data availability for the included variables across the three fundamental sets. For all currencies, the sample

period ends in October 2024. The start dates differ as follows: January 1974 for GBP and JPY, February 1975 for AUD, March 1975 for CAD, January 1980 for CHF, January 1998 for SEK, January 1999 for EUR, and January 2008 for NOK.