Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

Virtue or Mirage? Complexity in Exchange Rate Prediction

Rehim Kilic

2025-089

Please cite this paper as:
Kilic. Rehim (2026). "Virtue or Mirage? Complexity in Exchange Rate Prediction," Finance and Economics Discussion Series 2025-089r1. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2025.089r1.

Virtue or Mirage? Complexity in Exchange Rate Prediction

Rehim Kılıç*

June 2026

Abstract

Does the virtue of complexity extend to exchange rate forecasting? We test whether Ridge regression with Random Fourier Features (Ridge–RFF) outperforms OLS and the driftless random walk for eight major USD currency pairs across two predictor sets — traditional monetary and Taylor-rule fundamentals — using the Kelly et al. (2024) framework to diagnose *why* complexity fails and *what signal conditions* its success would require. Ridge–RFF substantially outperforms OLS at short training windows, reducing MSPE by up to 97% under Taylor-rule fundamentals, but predictive equality tests against the random walk favour Ridge–RFF for only three currencies (AUD, CHF, and JPY) under Taylor-rule fundamentals and one (JPY) under traditional monetary fundamentals. At longer training windows, the Taylor-rule gains persist robustly across intercept specifications while traditional monetary fundamentals offer no reliable random walk outperformance. Taylor-rule market-timing strategies generate portfolio Sharpe ratios approaching 0.40 at the longest training window, but these collapse once an intercept is included, exposing drift capture rather than genuine directional predictability. Tracing the VoC curve across the full regularisation spectrum confirms the failure is structural: the cross-currency mean out-of-sample R^2 is bounded above by zero everywhere. The average predictive signal per feature is indistinguishable from zero, implying an optimal regularisation 10^{12}–10^{13} times above the equity baseline — quantifying both *why* complexity fails in FX and *what signal conditions* would be needed for it to succeed.

JEL Classification: C50, G11, G15, F41 .
Keywords: Model complexity, machine learning, Ridge, RFF, FX rate, disconnect puzzle, predictability.

*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. The landmark study of Meese and Rogoff (1983) established that structural models based on economic fundamentals consistently underperform a naive driftless random walk in out-of-sample tests, and four decades of subsequent research have largely corroborated this verdict; see Rossi (2013) for a comprehensive survey. Despite considerable theoretical advances and methodological innovations, the Meese–Rogoff puzzle persists, underscoring both the robustness of the random walk benchmark and the elusive nature of exploitable structure in exchange rate dynamics.

The persistence of this puzzle is not for want of trying. Diebold and Nason (1990) applied nonparametric kernel regression to test whether past exchange rates predict future ones through nonlinear autoregressive functions, and found that any latent nonlinear structure was either absent, highly unstable, or too weak relative to noise to be exploited out-of-sample. Subsequent work has also asked whether economic fundamentals matter only conditionally, through smooth-transition, threshold, or regime-switching relationships, including nonlinear Taylor-rule specifications (Wang et al., 2019; Huber, 2017) and Markov-switching monetary or structural exchange-rate models (Frömmel et al., 2005; Stillwagon and Sullivan, 2020; Hauzenberger and Huber, 2020).

In the three and a half decades since Diebold and Nason (1990), researchers have returned repeatedly to the nonlinearity question with progressively more powerful tools: neural networks, gradient-boosted trees, deep learning architectures, and most recently high-dimensional machine learning models with explicit regularisation. The present paper revisits this question using the latest such approach — Ridge regression with Random Fourier Features (Ridge–RFF), introduced to the return-prediction literature by Kelly et al. (2024) — but does so in an FX setting that is fundamentally more adversarial than the equity markets where this method has shown the most promise. Our central finding is discouraging for complexity advocates: despite 35 years of methodological progress, the core message of Diebold and Nason is largely intact. The paper's contribution, however, extends well beyond this verdict. We do not simply ask whether the virtue of complexity (VoC) transfers from equities to FX; working within the Kelly et al. (2024) framework itself, we use its own diagnostic tools to ask *why* complexity fails when it does and to characterise *what signal conditions* its success would require — turning the VoC machinery into an instrument for understanding the structural limits of return prediction in exchange rate markets. Our question also differs from that of Diebold and Nason in a meaningful way: they test whether past exchange rates alone contain exploitable nonlinear structure, whereas we ask whether economic fundamentals

predict exchange rates through nonlinear functions that Ridge–RFF can approximate. This is the more ambitious claim, since our approach gives complexity every theoretical advantage conferred by both the content of fundamentals and the approximation power of the RFF framework — which makes our broadly negative conclusion all the more informative about the limits of complexity in this domain.

The motivation for this investigation comes from the virtue of complexity debate initiated by Kelly et al. (2024). Working in equity markets, they argue that high-dimensional predictive models implemented with RFF and Ridge regularisation can outperform simpler linear alternatives by harnessing "benign overfitting" — the ability of appropriately regularised large models to exploit weak but pervasive predictive signals. These claims have generated a nuanced debate. Nagel (2025) cautions that apparent gains in small samples may reflect mechanical volatility-timing artefacts rather than genuine nonlinear structure, while Buncic (2025) demonstrates that specific implementation choices — excluding an intercept term and an unconventional aggregation of simulation draws — together account for the strictly monotone VoC curves documented for equities. In response, Kelly and Malamud (2025) clarify the distinction between nominal complexity (the feature-to-observation ratio) and effective complexity, and show that data features they term concentration and alignment — broadly, the strength and structure of signal in the predictor space — determine whether performance rises with nominal complexity. Fallahgoul (2026) contributes two further theoretical results: standardising RFF features within each estimation window, as in the original Kelly et al. (2024) implementation, distorts the underlying kernel approximation in a way that does not vanish as the number of features grows; and equity return prediction itself sits in a signal-limited regime relative to the sample sizes typically available, implying the FX setting is even more constrained. Cartea and Shi (2025) complement this with a closed-form characterisation showing that, in the smallest-sample limit, the link between complexity and performance is mechanical rather than reflective of genuine learning. The Kelly and Malamud (2025) concentration-and-alignment framework provides the language for our structural investigation: whether the conditions for a genuine VoC hold in FX is not merely an empirical question but one we answer structurally, by recovering the implied signal-strength parameters from FX data and comparing them to the equity setting where the VoC is known to operate.

A key design choice distinguishes our paper from recent work finding that machine learning can improve FX forecasting. Filippou et al. (2025), for example, demonstrate consistent random walk outperformance by combining economic fundamentals with ensemble machine learning methods, but their approach simultaneously modifies both the predictor information set and the estimation method, making it impossible to attribute gains to either source individually. We instead hold the predictor set fixed at two well-established theoretical

benchmarks — traditional monetary fundamentals (Meese and Rogoff, 1983) and Taylor-rule fundamentals (Molodtsova and Papell, 2009) — and vary only the estimation method, comparing ordinary least squares (OLS) against Ridge–RFF. This isolates the contribution of model complexity from the contribution of information content, and ensures that any failure of complexity can be traced to the signal properties of the predictor set itself rather than to the choice of variables. This clean isolation is what enables the structural diagnosis that follows: when Ridge–RFF fails to outperform the random walk, we can attribute that failure to the underlying signal strength of the FX data, not to an insufficiently rich information set.

We study eight major USD currency pairs — AUD, CAD, CHF, EUR, GBP, JPY, NOK, and SEK — and compare three models: a random walk benchmark, OLS, and Ridge–RFF with $P = 12{,}000$ random Fourier features and regularisation parameter $z = \lambda/T = 1{,}000$ following Kelly et al. (2024). Rolling training windows of $T \in \{12, 60, 120\}$ months span the full range from extreme overparameterisation to moderate complexity. All models are evaluated against the driftless random walk using out-of-sample R^2, MSPE ratios, and the Clark–West test (Clark and West, 2007); direct comparisons between OLS and Ridge–RFF use the Diebold–Mariano test (Diebold and Mariano, 1995) with the small-sample correction of Harvey et al. (1997). These formal test statistics, largely absent from prior VoC studies in equities, are essential for distinguishing genuine predictive gains from estimation noise in a setting where the signal-to-noise ratio is notoriously low. As a robustness check motivated by Buncic (2025), we additionally estimate all models with an intercept and evaluate them against a drift random walk; results throughout are reported for both specifications. Our baseline also applies the correct aggregation of Ridge–RFF simulation draws — averaging forecasts across random weight draws before computing performance statistics — addressing a second implementation concern raised by Buncic (2025). Throughout, all Ridge–RFF estimates use raw, unstandardised features rather than the within-window standardised features of the original Kelly et al. (2024) implementation, ensuring our results are free of the kernel-approximation distortion that Fallahgoul (2026) identifies.

The results reveal a coherent and cautionary narrative. At the shortest training window, where nominal complexity is most extreme, Ridge–RFF delivers a clear and substantial forecasting advantage over OLS under both predictor sets, consistent with the benign overfitting property that Kelly et al. (2024) document in equities: regularisation prevents the variance inflation that would otherwise afflict a high-dimensional OLS regression at this sample size, stabilising forecasts close to the random walk. Yet this advantage over OLS does not transfer to the more demanding benchmark. Significant outperformance of the random walk occurs for only three currencies under Taylor-rule fundamentals — AUD, CHF, and JPY — and for the Japanese yen alone under traditional monetary fundamentals, underscoring that

the adversarial random walk benchmark, largely absent from equity-focused VoC studies, is the relevant test in FX, and most of the currency cross-section does not clear it.

As the training window lengthens, the two predictor sets diverge sharply. Under Taylor-rule fundamentals, Ridge–RFF retains a meaningful advantage over OLS at every window and achieves robust, statistically significant outperformance of the random walk for a stable subset of three currencies — AUD, CHF, and JPY — a result that proves invariant to whether an intercept is included, in contrast to the material role the intercept plays for the VoC in equities. Under traditional monetary fundamentals, by comparison, Ridge–RFF's edge over OLS narrows substantially as the window lengthens and largely disappears by the longest window, while OLS itself recovers some genuine predictive content for a small number of currencies at intermediate and long horizons. Across both predictor sets, the cross-currency mean out-of-sample R^2 remains negative at every training window: the individual-currency gains documented above are real but concentrated in specific currencies rather than reflecting the pervasive, panel-wide signal that a systematic VoC would require. Two robustness checks reinforce this conclusion. First, when regularisation intensity is not held fixed as the training window grows (Appendix B), Ridge–RFF's advantage over OLS deteriorates and eventually reverses, showing that regularisation intensity, not complexity per se, drives the cross-window pattern. Second, replacing our raw RFF features with the standardised variant of Kelly et al. (2024) degrades forecast accuracy in both statistical and economic terms across the board (Appendix E), and is sufficient on its own to erase the positive predictability we document at intermediate windows — confirming that this predictability is specific to the raw implementation and would be invisible under the standardised one.

The market-timing evidence tells a complementary and equally informative story, and is itself diagnostic of the underlying mechanism. Under the no-intercept specification, no equal-weighted portfolio strategy achieves meaningful statistical significance at the shortest window: short-window gains are idiosyncratic across currencies and diversify away once aggregated. At intermediate and long windows, by contrast, Taylor-rule strategies generate statistically significant equal-weighted portfolio returns, with Ridge–RFF delivering the strongest result in the analysis at the longest window — an annualised equal-weighted portfolio Sharpe ratio approaching 0.40 under the no-intercept specification. Including an intercept, however, exposes a striking asymmetry across horizons. At the shortest window, the positive portfolio returns are essentially unchanged by the intercept — a near-invariance consistent with these returns reflecting genuine volatility-timing behaviour rather than the model implicitly absorbing each currency's unconditional drift. At intermediate and long windows, by contrast, the Taylor-rule portfolio returns collapse to statistical insignificance once an intercept is included, revealing that the no-intercept performance at these horizons

was substantially a drift-capture artefact: the zero-intercept restriction had forced the model's slope coefficients to absorb each currency's unconditional mean return rather than condition on genuine directional signal. The only market-timing result that survives the intercept at intermediate or long windows comes from the Canadian dollar under traditional monetary fundamentals, indicating that this case alone reflects directional accuracy that is not an artefact of drift capture. The contrast between near-invariance at the shortest window and collapse at longer ones is exactly the pattern implied by two distinct mechanisms operating at different horizons: volatility-timing, which does not depend on the intercept restriction, at short windows; and drift capture, which is undone once the intercept is freed, at longer ones.

These fixed-regularisation results naturally raise a deeper question: is the limited success of Ridge–RFF in FX a calibration problem — the equity-motivated regularisation level being the wrong choice for this domain — or does it reflect a structural property of FX fundamental data that no level of regularisation can overcome? Section 5 addresses this by tracing the full VoC curve across the entire feasible regularisation spectrum, decomposing in-sample from out-of-sample performance, and applying the Kelly et al. (2024) framework's own diagnostic tools to the FX data — asking, in effect, whether the apparent virtue of complexity in this setting is genuine or a mirage: a model that fits the data in-sample but cannot forecast it out-of-sample at any level of regularisation. We find that the cross-currency mean out-of-sample R^2 is bounded above by zero across the entire spectrum, for both predictor sets and all training windows: no degree of regularisation yields a model that systematically beats the random walk across the currency panel. The in-sample VoC curve, by contrast, behaves exactly as in equity markets, confirming that the model is fully capable of fitting the data; the gap between in-sample and out-of-sample performance is therefore the signature of overfitting in the absence of exploitable panel-wide signal, not of a miscalibrated model. Benchmarking Ridge–RFF against OLS rather than the random walk shows that meaningful variance reduction relative to OLS is achievable throughout the spectrum — confirming that the regularisation itself does useful statistical work — but the OLS baseline is itself insufficient to defeat the random walk, and shrinking toward it more efficiently cannot close that gap. The economic performance results echo this pattern across the spectrum: under the no-intercept specification, Taylor-rule portfolio returns rise steadily with the level of regularisation at every training window, while under the with-intercept specification, returns at intermediate and long windows peak early in the spectrum and then decline toward and below zero, reinforcing the drift-capture interpretation across the full range of regularisation choices.

The structural explanation lies within the Kelly et al. (2024) framework itself, and constitutes the paper's deepest finding. The average predictive signal strength per RFF

feature is statistically indistinguishable from zero across the regularisation spectrum, for both predictor sets and all training windows. The regularisation level that this signal strength implies as optimal is many orders of magnitude larger than the level estimated for equities — a gap so large that it cannot plausibly be closed by any feasible choice of regularisation. This estimate quantifies the structural gap between FX and equity markets in a precise, model-consistent way: it is not that the wrong regularisation level is used at the equity baseline, but that FX fundamental data offer essentially no signal per feature for the complexity machinery to exploit. The Kelly et al. (2024) framework does not fail in FX because it is a poor framework; it fails because the necessary condition for the virtue of complexity — non-trivial signal strength — is largely absent from monthly FX fundamental data at the panel level. Crucially, this diagnosis also reveals what would be needed for the VoC to succeed in FX: predictor sets whose concentration and alignment properties deliver signal strength of the same order as equity, a demanding condition that neither of our two theoretical benchmarks satisfies at the panel level, even though it is locally approached by specific currency–predictor combinations.

This paper makes five contributions that together constitute both a rigorous test of the VoC in FX and a structural investigation of why complexity, for all its theoretical promise, cannot systematically overcome the signal limitations of this domain. *First*, it provides one of the first rigorous tests of the VoC hypothesis in exchange rates, a domain where the random walk is substantially more formidable than in equities, and where holding the predictor set fixed while varying only the estimation method cleanly isolates the contribution of complexity from the contribution of richer information. *Second*, and most distinctively, we provide a structural explanation for the failure of complexity in FX using the Kelly et al. (2024) framework's own diagnostic tools: the signal strength recovered from FX data implies an optimal regularisation level many orders of magnitude larger than the equity baseline, quantifying in model-consistent terms both why no calibration choice produces pervasive random walk outperformance, and what signal conditions would be needed for it to succeed. *Third*, we extend the analysis from a single regularisation level to the full regularisation spectrum, showing that the cross-currency mean out-of-sample VoC curve is bounded above by zero throughout, and that the intercept critique of Buncic (2025) — materially important in equities — is inconsequential for the statistical predictability conclusions in FX, even as it proves material for the economic performance conclusions. *Fourth*, we document a nuanced interaction between complexity, regularisation intensity, and training sample size: Ridge–RFF outperforms OLS at every training window under our baseline specification and achieves individual-currency gains over the random walk at intermediate and long windows, but this performance requires that regularisation intensity be maintained as the training

window grows, and never aggregates into systematic cross-sectional outperformance. *Fifth*, methodologically, we incorporate the Clark–West and Diebold–Mariano tests — largely absent from equity-focused VoC studies — and distinguish the driftless from the drift random walk throughout, sharpening inference on when apparent gains are statistically meaningful rather than artefacts of implementation; our use of raw, unstandardised RFF features further ensures that our findings are free of the kernel-approximation concerns that Fallahgoul (2026) identify in the original implementation.

The remainder of the paper proceeds as follows. Section 2 reviews related work and positions our contribution. Section 3 describes the empirical methodology, data, and predictor construction. Section 4 presents out-of-sample forecasting results and statistical tests at the baseline regularisation. Section 5 provides the full-spectrum VoC analysis, effective complexity diagnostics, signal-strength estimation, and intercept robustness. Section 4.4 evaluates market-timing performance for single-currency and equal-weighted portfolios. Section 6 concludes. Data details appear in Appendix A; fixed-λ robustness results appear in Appendix B; single-currency market-timing performance results appear in Appendix C; Sharpe ratios across the regularisation spectrum appear in Appendix D; the comparison of standardised versus raw RFF results appears in Appendix E; and additional complexity diagnostics appear in Appendix F.

2 Literature Review

The random walk benchmark and the Meese–Rogoff puzzle. Empirical exchange rate forecasting has been shaped by the landmark finding of Meese and Rogoff (1983) that no structural model based on macroeconomic fundamentals could consistently beat a naive driftless random walk in out-of-sample tests. This result — now known as the Meese–Rogoff puzzle — has proved remarkably durable. Cheung et al. (2005) examined an expanded battery of monetary, productivity-based, and interest-parity models and found that none consistently outperformed the random walk on mean squared error criteria. A subsequent reassessment by Cheung et al. (2017) reached the same conclusion for more recent model variants, including real interest-rate parity with shadow rates and Taylor-rule specifications, noting that for major pairs such as EUR/USD even newly developed models do not reliably improve on the benchmark. The random walk without drift therefore remains the hardest target to beat in monthly exchange rate forecasting — a point directly relevant to our analysis, which explicitly benchmarks both linear and complex machine learning models against it.

The difficulty of beating the random walk traces in part to the nature of the underlying economics. Exchange rate fundamentals — money supply growth, inflation, interest rate

differentials, output gaps — are slow-moving, highly persistent, and subject to recurring structural breaks as monetary policy regimes change. This stands in sharp contrast to the equity return setting, where the signal embedded in predictors can be latent, high-frequency, and nonlinear in ways that favour flexible estimators. The implication, which we explore empirically, is that the FX domain may be inherently less amenable to the gains from complexity than equities: if the underlying predictive signal is weak, episodic, and tied to regime shifts rather than stable nonlinear structure, then high-dimensional approximators have little additional signal to exploit once the training window is long enough for parsimony to work. Our signal-strength analysis in Section 5 provides direct empirical support for this interpretation.

When do fundamentals help? Despite the benchmark's resilience, a substantial literature has identified conditions under which economic fundamentals improve exchange rate forecasts. Rossi (2013) surveys post-2000 evidence and concludes that predictability is real but conditional: it emerges under specific combinations of predictors, estimation methods, and evaluation horizons. Taylor-rule differentials — which proxy relative monetary policy stances — and net foreign asset positions showed particular promise at short horizons in simple linear models, consistent with limited-parameter estimation mitigating overfitting.

Molodtsova and Papell (2009) demonstrate that Taylor-rule fundamentals yield statistically significant out-of-sample gains for several currency pairs during the early 2000s, though these gains prove period-specific. Studies of purchasing power parity as a predictor similarly find horizon-dependence: Zorzi et al. (2015) show that a calibrated half-life PPP model can outperform the random walk at both short and long horizons, with the intuition that the gradual mean-reversion of real exchange rates provides exploitable medium-run predictability even when short-run fundamental relationships are weak.

More recently, Engel and Wu (2024) document a structural improvement in the performance of standard exchange rate models since the early 2000s, attributing it to improved monetary policy credibility under inflation targeting regimes, which reduces the scope for self-fulfilling volatility. Their evidence underscores that predictability in FX is circumstantial rather than systematic: it concentrates in regimes and horizons where the fundamental–return relationship is temporarily stable. This episodic character of FX predictability is an important backdrop for our analysis, as it raises the prior that flexible nonlinear models may simply find more noise to fit rather than more signal to exploit.

Structural instability and adaptive approaches. A recurring theme in the literature is that parameter instability — the time-varying relationship between fundamentals and

exchange rates — undermines the performance of fixed-parameter models (Rossi, 2013). Kouwenberg et al. (2017) address this with an adaptive model selection framework that allows the set of relevant fundamentals to shift over time, consistent with "scapegoat" theories of exchange rate determination (Bacchetta and van Wincoop, 2004) in which agents rationally focus on different fundamentals in different periods. Their adaptive approach significantly beats the random walk for five of ten major currencies, with the selected fundamentals and their weights varying across the sample.

A closely related strand models this instability through explicitly nonlinear fundamental-based specifications. Wang et al. (2019) use a smooth transition regression version of the Taylor-rule exchange-rate model and report out-of-sample gains over linear Taylor-rule models, nonlinear UIP models, and the random walk. Huber (2017) develops a threshold time-varying parameter Taylor-rule model in which structural breaks in coefficients are determined endogenously and finds density-forecast improvements relative to constant-parameter and standard time-varying-parameter benchmarks. Markov-switching approaches provide another way to formalise regime dependence: Frömmel et al. (2005) estimate regime-switching monetary exchange-rate models, Stillwagon and Sullivan (2020) allow the number of Markov regimes in monetary-fundamental models to be selected from the data, and Hauzenberger and Huber (2020) use a Bayesian nonlinear framework in which regimes correspond to alternative structural exchange-rate models and transition probabilities depend on monetary-policy stances. The evidence from this literature is informative but mixed: nonlinear and regime-switching specifications often improve on linear or simpler regime benchmarks, especially for density forecasts or particular currencies, yet they do not deliver systematic random-walk outperformance in point forecasts. Recent threshold predictive-regression evidence similarly emphasises short, recurring phases of exchange-rate predictability rather than a stable global forecasting relationship (Beckmann et al., 2025). This pattern is consistent with our interpretation that nonlinear methods are most useful when they capture episodic, regime-specific signal, but that complexity alone is unlikely to overcome the weak average signal in fixed monetary or Taylor-rule information sets.

Panel data and factor-based methods offer another route to improving predictive power. Engel and West (2012) extract common factors from a panel of bilateral dollar exchange rates and show that factor-based forecasts modestly outperform the random walk at horizons of 8 to 12 quarters in recent samples. Ince and Kubler (2014) find that panel estimation with real-time data uncovers predictability that single-currency, revised-data regressions miss. The general lesson from this body of work is that gains come from pooling information and accommodating structural change — not from adding model complexity to a fixed, single-currency predictor set, which is precisely the comparison our paper pursues.

Machine learning in exchange rate forecasting. The application of machine learning to exchange rate prediction has grown rapidly, motivated by the potential for flexible algorithms to detect nonlinear interactions that elude linear models. Yet the pursuit of exploitable nonlinear exchange rate dynamics predates modern machine learning by decades, and that history provides important context. Diebold and Nason (1990) applied nonparametric kernel regression to test whether past exchange rates alone predict future rates through nonlinear functions — asking whether the conditional mean of the return is nonlinear in its own lags. They found no exploitable nonlinear autoregressive structure, concluding that if such structure exists it is either absent, highly unstable, or too weak relative to noise to survive out-of-sample evaluation. Our paper revisits the nonlinearity question from a different angle: rather than past returns, we use macroeconomic fundamentals as the information set, and rather than nonparametric kernel regression, we use Ridge regression with Random Fourier Features to approximate nonlinear functions of those fundamentals. The distinction matters — fundamental-based models carry theoretical content that purely autoregressive ones do not — but our conclusions are broadly consonant with Diebold and Nason: the nonlinear structure in FX returns that can be reliably exploited out-of-sample, whether in the returns themselves or in their relationship to fundamentals, appears fragile at best. Thirty-five years of methodological development, from nonparametric regression to deep neural networks to the RFF framework, has not overturned that basic finding in the FX domain.

The recent machine learning literature does, however, contain positive results that merit careful contextualisation. Amat (2018) apply ridge regressions and exponentially weighted averaging to a broad set of fundamentals and report modest directional forecast improvements over OLS. Pfahler (2022) extend this to artificial neural networks and gradient-boosted trees using a panel of ten OECD currencies, finding significant directional predictability — but one whose source is ambiguous: an ANN estimated on time fixed effects alone performs nearly as well, raising the concern that the algorithms capture momentum or regime effects rather than fundamental-driven nonlinear structure. Meng et al. (2024) apply transformer-based deep learning to the Chinese RMB/USD rate and find that fundamental information remains central even in a high-dimensional context, with the flexible ML framework better capturing interactive effects. Filippou et al. (2025) demonstrate consistent random walk outperformance using an ensemble of elastic net and deep neural networks applied to global and country-level variables, with gains concentrated during periods of financial stress.

These positive findings are important and should not be dismissed, but they arise in settings that differ from ours in ways that account for the different conclusions. First, papers that find gains — particularly Filippou et al. (2025) — simultaneously modify both the predictor information set and the estimation method, making it impossible to attribute improvements

to either source alone. Second, gains are often concentrated in specific subperiods (financial stress episodes) or measured by directional accuracy rather than MSPE relative to the driftless random walk. Third, papers that use richer predictor sets effectively allow machine learning to search over a broader information space, which conflates the value of additional predictors with the value of flexible estimation. Our design holds the predictor set fixed at two well-established theoretical benchmarks and varies only the estimation method, isolating the question of whether complexity adds value to a given information set. Within that design, and as the signal-strength analysis of Section 5 confirms, the FX domain offers insufficient predictive signal per feature for the complexity machinery to exploit — irrespective of the regularisation level chosen.

The virtue of complexity debate. Our paper contributes directly to the emerging debate on the virtue of complexity (VoC) in return prediction, initiated by Kelly et al. (2024). Working in the equity market, they argue that high-dimensional predictive models implemented with random Fourier features and Ridge regularisation can outperform linear alternatives by harnessing "benign overfitting" — the ability of appropriately regularised large models to exploit weak but pervasive predictive signals. Consistent with their view, we find that Ridge–RFF yields localised improvements over OLS in very small samples when nominal complexity is high.

Nagel (2025) argues that with $P \gg T$ and highly persistent predictors, RFF forecasts in short samples mechanically approximate a volatility-timed momentum rule: the Ridge–RFF weights effectively form a similarity kernel over the last T returns, so apparent small-sample learning largely reflects recency and volatility timing rather than genuine nonlinear structure. Our FX results are consistent with this interpretation: at $T = 12$, Ridge–RFF outperforms OLS locally but does not consistently defeat the random walk; at larger T, the short-sample momentum effect attenuates and the random walk benchmark reasserts its resilience. The dissociation between OOS R^2 and economic performance — positive EW portfolio Sharpe ratios despite negative cross-currency mean R^2 — in our spectral analysis of Section 5 further corroborates Nagel's interpretation, and the collapse of positive Sharpe ratios at longer windows under the with-intercept specification is consistent with the drift-capture channel he identifies.

Buncic (2025) revisits the empirics of Kelly et al. (2024) and demonstrates that two implementation choices — excluding an intercept term and a particular aggregation of RFF simulation draws — together account for the monotone VoC curves. When these are corrected (intercept included, forecasts aggregated before performance evaluation), simpler or mildly regularised models often dominate. Our baseline follows Kelly et al. (2024) in

excluding the intercept and uses the correct aggregation (forecasts averaged across 1,000 draws *before* computing performance metrics); we additionally report results with an intercept as a robustness check directly motivated by Buncic (2025). Notably, and in contrast to the equity setting, we find that the intercept critique has no qualitative force for the *statistical* predictability conclusions in FX: the OOS VoC curve remains bounded above by zero across the full regularisation spectrum whether or not an intercept is included. It does, however, have material force for the *economic* performance conclusions: the positive EW Sharpe ratios at longer training windows under the no-intercept baseline are substantially driven by drift capture, and including an intercept reveals this. The reason the statistical conclusion is unaffected is structural — because the signal strength $\hat{b}_*$ is approximately zero in FX fundamental data irrespective of intercept choice, re-centring the regression cannot alter the fundamental diagnosis.

Kelly and Malamud (2025) respond to these critiques by clarifying the distinction between nominal complexity ($c = P/T$) and effective complexity (the ratio of effectively used parameters to sample size, accounting for shrinkage and implicit regularisation). They show that nominal complexity is the central driver of out-of-sample performance in equities, and that data features they term "concentration" and "alignment" — broadly, the strength and structure of the signal in the predictor space — determine whether the VoC curve slopes upward. Our signal-strength analysis in Section 5 directly operationalises these concepts in FX: the implied optimal regularisation parameter $\hat{z}^*$ recovered from FX data is 10^{12}–10^{13} times larger than the equity baseline, quantifying the gap between FX and equity in precisely the terms Kelly and Malamud identify as determining VoC success. Our effective complexity analysis further shows that even under the more permissive fixed-λ parameterisation — which allows substantially higher effective complexity at longer training windows — the OOS VoC curve remains bounded above by zero, ruling out insufficient model flexibility as an explanation for the failure. Cartea et al. (2025) complement this interpretation by showing that in high-dimensional settings the benefits of complexity erode once feature noise — the portion of the expanded feature space unrelated to the true signal — is accounted for; in FX, where the fundamental signal is already weak in linear models, this noise penalty is likely to dominate.

Two further contributions sharpen the theoretical understanding of the RFF framework and have direct bearing on our analysis. Fallahgoul (2026) establishes, in Theorem 3.1, that dividing each RFF feature by its within-window sample standard deviation — the step applied in the Kelly et al. (2024) empirical procedure (their footnote 39) — causes the empirical kernel to converge almost surely to a data-dependent limit $k^*_{\text{std}}(\cdot, \cdot|\mathcal{T})$ that differs from the intended Gaussian kernel k_G by a structural, non-vanishing gap (approximately forty times

the standard RFF approximation error at $P = 12{,}000$). Because the target reproducing kernel Hilbert space changes with every rolling window, the theoretical guarantees of kernel ridge regression — which assume a single, fixed function space — do not transfer to the standardised implementation. Separately, Fallahgoul (2026) derives minimax lower bounds identifying a critical sample size T_{crit} below which no estimator can reduce prediction risk; calibrated to the Kelly et al. (2024) parameter values, this threshold exceeds 340 months, placing equity return prediction in the signal-limited regime. *Crucially for our paper, our Ridge–RFF implementation does not apply within-window standardisation*: the $P = 12{,}000$ raw sin/cos projections are fed directly to Ridge regression without any per-feature scaling, ensuring our results are specific to the standardised variant that Theorem 3.1 concerns and do not reflect the implementation artefact it identifies.

Cartea and Shi (2025) provide a complementary theoretical characterisation, deriving closed-form expressions for the expected out-of-sample R^2 and expected timing-strategy return when the RFF model is estimated on a rolling window containing a single observation ($T = 1$). They show that the link between model complexity and performance arises mechanically rather than from learned predictability, and demonstrate analytically that reversing the sign of returns reverses the monotone complexity–performance relationship. Though our implementation differs from the $T = 1$ setting of Cartea and Shi (2025), their result is consistent with the interpretation that short-window Ridge–RFF gains reflect the algebraic interaction between features and returns rather than genuine structural learning — an interpretation our own results at $T = 12$ support.

Positioning of this paper. Our analysis sits at the intersection of two literatures: the long-standing question of FX predictability and the recent debate on the virtue of complexity in machine learning-based return prediction. We provide one of the first rigorous tests of the VoC hypothesis in exchange rates, a domain where the random walk benchmark is substantially more formidable than in equities and where the economic structure of predictability is qualitatively different. By holding the predictor set fixed at theoretically motivated benchmarks and varying only the estimation method, we isolate the contribution of complexity. By tracing performance across the full regularisation spectrum and recovering the implied optimal regularisation within the Kelly et al. (2024) framework, we explain the failure structurally rather than simply documenting it empirically. Our additional use of the Clark–West and Diebold–Mariano tests, largely absent from prior VoC studies, sharpens the inference on when apparent gains are statistically meaningful rather than artefacts of estimation noise. Finally, our use of raw, unstandardised RFF features ensures that the

results are immune to the within-window standardisation concerns identified by Fallahgoul (2026), providing a clean methodological foundation for the conclusions we draw.

3 Empirical Methodology and Data

This section describes the empirical design, the two sets of exchange rate fundamentals, and the nonlinear machine learning framework used to assess the virtue of complexity in FX forecasting. We describe each component in turn and explain the design choices that distinguish our analysis from related work.

A central methodological choice in this paper is to hold the predictor information set fixed and vary only the estimation method. This design isolates the value added by model complexity — the transition from ordinary least squares to Ridge regression with Random Fourier Features — from the value added by introducing new predictors. The distinction matters because the two sources of improvement are easily confounded. For example, Filippou et al. (2025) demonstrate that exchange rates can be predicted at short horizons by combining economic fundamentals with machine learning methods, but their approach modifies both the predictor set (adding country-level and global variables beyond standard monetary fundamentals) and the estimation method (ensemble of elastic net and deep neural networks) simultaneously. Our paper asks a sharper question: holding the predictor set fixed at well-established benchmarks from the theoretical literature, does greater model complexity in the form of the Ridge–RFF framework deliver predictive gains over a simple linear regression estimated on the same information set and over the random walk benchamrk? Using the Meese–Rogoff (Meese and Rogoff, 1983) traditional monetary variables, in addition to Taylor-rule fundamentals, as one of our key predictor set is a deliberate signal of this design principle: these are the predictors that standard structural models imply should forecast exchange rates, and they are the benchmark against which the random walk has resisted all challengers for four decades. Augmenting them with alternative variable sets would risk importing predictor-level gains that are unrelated to the complexity question.

3.1 Forecast target and currencies

Following the exchange rate forecasting literature, we define the forecast target as the log return of the bilateral spot exchange rate against the U.S. dollar:

$$\Delta s_t = \log(S_t) - \log(S_{t-1}), \tag{1}$$

where S_t is the spot rate (domestic currency price of one unit of foreign currency) at the end of month t. We study eight major USD currency pairs: the Australian dollar (AUD), Canadian dollar (CAD), Swiss franc (CHF), euro (EUR), British pound (GBP), Japanese yen (JPY), Norwegian krone (NOK), and Swedish krona (SEK).[1] Data sources and sample periods for each currency are described in Appendix A.

3.2 Predictor sets

We consider two sets of fundamentals that together span the main theoretical frameworks used in the empirical exchange rate forecasting literature.

1. Traditional monetary fundamentals. The first set follows the monetary model of exchange rate determination with sticky prices, as tested by Meese and Rogoff (1983) and revisited extensively in subsequent work (Rossi, 2013). It consists of four bilateral differentials between the U.S. (home) and the foreign country:

$$G_t^{\text{tr}} = \big[\, \Delta(m_t - m_t^*),\ \Delta(i_t - i_t^*),\ \Delta(\pi_t - \pi_t^*),\ \Delta(y_t - y_t^*) \,\big]',$$

where m_t is the log money supply (M0), i_t is the 3-month nominal government bond rate, $\pi_t = p_t - p_{t-12}$ is the 12-month inflation rate computed from the log consumer price index p_t, and y_t is the log industrial production index. An asterisk denotes the corresponding foreign variable, and Δ denotes the one-month first difference. This gives $k = 4$ predictors.

2. Taylor-rule fundamentals. The second set follows Molodtsova and Papell (2009) and is motivated by interest rate reaction functions: central banks set rates in response to inflation and the output gap, and the differential in these policy responses should drive the bilateral exchange rate. The predictor vector is

$$G_t^{\text{taylor}} = \big[\, i_t,\ i_t^*,\ \pi_t,\ \pi_t^*,\ y_t^{\text{gap}},\ y_t^{\text{gap}*},\ q_t \,\big]',$$

where y_t^{gap} and $y_t^{\text{gap}*}$ are the domestic and foreign output gaps estimated via the Hodrick–Prescott filter on quarterly GDP (linearly interpolated to monthly frequency), and $q_t = s_t + p_t^* - p_t$ is the log real exchange rate. This gives $k = 7$ predictors.

These two sets are complementary: the traditional set captures monetary quantity dynamics, while the Taylor-rule set captures policy rate dynamics. Together they span the

[1] For the USD pairs studied (AUD/USD, CAD/USD, etc.), positive Δs_t indicates the foreign currency appreciated against the USD.

classical and modern structural frameworks most commonly used in monthly exchange rate forecasting, making them natural benchmarks for evaluating whether complexity adds value beyond what the underlying theory already implies.

3.3 Linear model

The baseline forecasting model is a direct one-step-ahead OLS regression of the exchange rate return on the predictor vector:

$$\Delta s_{t+1} = G_t'\beta + \varepsilon_{t+1}. \tag{2}$$

This single-equation, lagged-fundamental specification is the workhorse of the empirical exchange rate literature (Rossi, 2013) and has the advantage of being less prone to endogeneity bias than iterated multi-step approaches, since the right-hand-side variables are dated at t.[2] Following Kelly et al. (2024), we exclude an intercept term from the baseline specification, so that both OLS and Ridge–RFF are evaluated against a driftless random walk as the nested benchmark. As a robustness check motivated by Buncic (2025), who shows that excluding the intercept can affect in-sample fit when the training window is short, we also estimate all models with an intercept and compare against a drift random walk; these results are reported in Section 4 and Appendix C.

3.4 Ridge Regression with Random Fourier Features

If the relationship between exchange rate returns and fundamentals is nonlinear, then a linear regression on G_t provides only a first-order approximation to the true predictive function. A natural question is whether a more flexible estimator that can approximate smooth nonlinearities in the same G_t improves out-of-sample performance.

Following Kelly et al. (2024), who build on the Random Fourier Features (RFF) framework of Rahimi and Recht (2007, 2008), we assume that the true predictive model for Δs_{t+1} is a smooth function $f(G_t)$:

$$\Delta s_{t+1} = f(G_t) + \varepsilon_{t+1}. \tag{3}$$

Because $f(\cdot)$ is unknown, we approximate it by a wide random basis expansion. Each draw of a $k \times 1$ weight vector $\omega_i \sim \mathcal{N}(0, I_k)$ generates a pair of new features:

$$Z_{i,t} = \left[\sin(\gamma\, \omega_i' G_t),\ \cos(\gamma\, \omega_i' G_t)\right]', \tag{4}$$

[2]Note that the terms "linear" and "OLS" are used interchangeably throughout the paper, as are "nonlinear" and "Ridge–RFF".

where γ controls the bandwidth of the resulting kernel approximation (set to $\gamma = 2$ following Kelly et al. 2024). Drawing $P/2$ independent weight vectors and stacking the resulting pairs yields a P-dimensional feature vector $\mathbf{Z}_t = (Z'_{1,t}, \ldots, Z'_{P/2,t})'$ that can approximate an arbitrary smooth function of G_t as $P \to \infty$.

The advantage of this construction is that it allows complexity to be varied continuously and smoothly: replacing G_t (a k-dimensional vector) with $\mathbf{Z}_t$ (a P-dimensional vector) transitions the model from low to high complexity without changing the underlying predictor information. The approximating model is:

$$\Delta s_{t+1} \approx \sum_{i=1}^{P} Z_{i,t}\, \beta_i + \tilde{\varepsilon}_{t+1}. \tag{5}$$

When $P \gg T$, this system is heavily overparameterised and requires regularisation. We estimate $\beta = (\beta_1, \ldots, \beta_P)'$ by Ridge regression:

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

where $\lambda > 0$ is the regularisation parameter. We set $P = 12{,}000$ and $\gamma = 2$ throughout, matching Kelly et al. (2024)'s baseline parametrisation. For the regularisation parameter λ we consider two implementations that differ in how they scale with the training window size T, and whose comparison is central to the identification strategy of the paper.

KMZ specification (Case 1, primary). Kelly et al. (2024) normalise the Ridge objective by the sample size T, so that shrinkage is characterised by the parameter z satisfying $\lambda = z \times T$, where z is held fixed across all training windows.[3] Our primary implementation adopts this convention exactly, fixing $z = 1{,}000$ so that the conventional Ridge penalty $\lambda = 1{,}000 \times T$ grows proportionally with T. At rolling window sizes of $T = 12$, 60, and 120 months, this yields $\lambda = 12{,}000$, 60,000, and 120,000 respectively. Crucially, the KMZ specification holds *regularisation intensity per observation* constant across all training windows.

[3] Kelly et al. (2024) write the Ridge objective in normalised form, $\min_\beta \{T^{-1} \sum_t (\Delta s_{t+1} - Z'_t\beta)^2 + z\,\|\beta\|^2\}$, which yields the closed-form solution $\hat{\beta}(z) = (T^{-1}Z'Z + zI)^{-1}T^{-1}Z'y = (Z'Z + zT\,I)^{-1}Z'y$. This is algebraically identical to the standard Ridge estimator $\hat{\beta}(\lambda) = (Z'Z + \lambda I)^{-1}Z'y$ with $\lambda = zT$. The KMZ parameterisation makes explicit that z measures regularisation intensity *per observation*: holding z constant as T grows scales the total penalty λ proportionally, keeping the shrinkage-to-sample-size ratio fixed.

Effective degrees of freedom. The degree of regularisation actually applied by Ridge is summarised by the effective degrees of freedom:

$$\mathrm{df}(\lambda) = \mathrm{tr}\big[\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \lambda I_P)^{-1}\mathbf{Z}'\big] = \sum_{i=1}^{T} \frac{\mu_i}{\mu_i + \lambda}, \tag{7}$$

where $\{\mu_i\}_{i=1}^{T}$ are the eigenvalues of the $T \times T$ Gram matrix $G = \mathbf{Z}\mathbf{Z}'$. Conceptually, $\mathrm{df}(\lambda)$ answers: *how many unconstrained parameters would produce the same degree of data-fitting as this regularised model?* Each eigenvalue direction contributes a fraction $\mu_i/(\mu_i + \lambda) \in (0, 1)$ — near one when $\mu_i \gg \lambda$ (light regularisation) and near zero when $\mu_i \ll \lambda$ (heavy regularisation) — and the sum of these fractions gives the total flexibility the model retains, expressed in OLS-equivalent units. Importantly, $\mathrm{df}(\lambda)$ is not a count of active features in the Lasso sense: Ridge does not set any coefficient to exactly zero, so all $P = 12{,}000$ coefficients are non-zero but shrunk, and df measures aggregate shrinkage intensity rather than the number of features switched on. The limiting cases are $\mathrm{df} \to T$ as $\lambda \to 0^+$ and $\mathrm{df} \to 0$ as $\lambda \to \infty$ (driftless random walk); at every finite $\lambda > 0$, $0 < \mathrm{df}(\lambda) < T$. The lower limit corresponds to *ridgeless regression* (Hastie et al., 2022), the minimum-norm interpolating solution $\hat{\beta} = \mathbf{Z}'(\mathbf{Z}\mathbf{Z}')^{-1}y$ obtained when $\lambda \to 0^+$ in the overparameterised regime $P > T$.[4]

To obtain an analytical expression for df/T under the KMZ specification, note that for every RFF observation the Pythagorean identity $\sin^2(\gamma\omega_i' G_t) + \cos^2(\gamma\omega_i' G_t) = 1$ holds exactly for each of the $P/2$ frequency-direction pairs, giving $\|\mathbf{Z}_t\|^2 = P/2$. Summing across observations:

$$\mathrm{tr}(G) = \sum_{t=1}^{T} \|\mathbf{Z}_t\|^2 = \frac{TP}{2}, \tag{8}$$

so the average Gram eigenvalue is $\bar{\mu} = P/2 = 6{,}000$, a constant independent of T and the data. In the high-regularisation regime ($\lambda \gg \bar{\mu}$), approximating all eigenvalues by their mean gives:

$$\mathrm{df}(\lambda) \;\approx\; \frac{\mathrm{tr}(G)}{\lambda} = \frac{TP/2}{\lambda}. \tag{9}$$

Substituting $\lambda = zT$ under the KMZ specification:

$$\frac{\mathrm{df}}{T} \;\approx\; \frac{P}{2zT}, \tag{10}$$

[4]When $P > T$, the $P \times P$ matrix $\mathbf{Z}'\mathbf{Z}$ is rank-deficient, so $\lambda = 0$ exactly yields no unique solution; the ridgeless limit $\lambda \to 0^+$ selects the minimum-ℓ_2-norm solution among all interpolants. This is conceptually distinct from classical OLS on the original k-dimensional predictor set G_t, which has $\mathrm{df}_{\mathrm{OLS}} = k$ (no intercept) or $k + 1$ (with intercept), far below T. In our setting $k \in \{4, 7\}$, so classical OLS uses only 4 or 7 degrees of freedom, whereas ridgeless RFF uses all T.

which with $P = 12{,}000$ and $z = 1{,}000$ gives $\mathrm{df}/T \approx 6/T$, falling from approximately 0.50 at $T = 12$ to 0.10 at $T = 60$ and 0.05 at $T = 120$. The approximation condition $\lambda \gg \bar{\mu}$ holds comfortably at $T = 60$ and $T = 120$ ($\lambda/\bar{\mu} = 10$ and 20) but only marginally at $T = 12$ ($\lambda/\bar{\mu} = 2$); the empirical values computed from the actual Gram eigenvalues at each rolling window are approximately 0.58, 0.05, and 0.04, with the discrepancy at $T = 60$ reflecting eigenvalue heterogeneity (many small eigenvalues contribute near-zero terms, pulling the empirical mean below the uniform-eigenvalue prediction).[5]

Under the fixed-λ specification, $\lambda = 1{,}000$ is well below $\bar{\mu} = 6{,}000$, placing the estimator in the low-regularisation regime. The high-regularisation approximation is inapplicable there (it would predict $\mathrm{df}/T \approx P/(2\lambda) = 6 > 1$, which is impossible). Instead, the uniform-eigenvalue approximation gives:

$$\frac{\mathrm{df}}{T} \approx \frac{\bar{\mu}}{\bar{\mu} + \lambda} = \frac{6{,}000}{7{,}000} \approx 0.86, \tag{11}$$

which is constant across all T because $\bar{\mu} = P/2$ does not depend on T. The fixed-λ model therefore behaves as near-ridgeless RFF at every training window regardless of sample size.

Signal strength and implied optimal regularisation. Kelly et al. (2024) show that, within the RFF–Ridge framework, the optimal regularisation parameter satisfies $z^* = c/b_*$, where $c = P/T$ is nominal complexity and b_* is the *average predictive signal per RFF feature*. Specifically, if the true predictive function $f(\cdot)$ introduced in equation (3) is represented in the RFF feature space as $f(G_t) \approx \mathbf{Z}_t'\beta^*$, then

$$b_* = \frac{\|\beta^*\|^2}{P} = \frac{1}{P}\sum_{i=1}^{P}(\beta_i^*)^2, \tag{12}$$

the mean squared population coefficient — equivalently, the average energy of $f(\cdot)$ projected onto each RFF feature direction. When b_* is large, the true function loads strongly on the RFF features, modest regularisation (small z^*) is optimal, and complexity estimation is expected to pay off. When $b_* \approx 0$, the signal per feature is negligible, the optimal regularisation diverges ($z^* \to \infty$), and the optimal forecast collapses to the random walk.

Kelly and Malamud (2025) further show that the shape of the VoC curve depends on two properties of the data generating process. *Concentration* describes how evenly the eigenvalues of the predictor covariance matrix are distributed: when eigenvalues are concentrated in a few dominant directions, the DGP is effectively low-dimensional and simple models can already capture most predictor variation, leaving little for complexity to add. It is a *diffuse*

[5]Precise empirical values are obtained by evaluating the exact summation in equation (7) at $\lambda = z\,T$ at every rolling window.

eigenvalue distribution — many predictors contributing differentiated information — that makes high-dimensional complexity necessary. *Alignment* describes whether predictors with the largest eigenvalues also carry the strongest predictive signal: high alignment implies that predictability is dominated by the large principal components already accessible to simple models, again leaving little for complexity to exploit. It is *low or negative* alignment — signal spread across minor eigenvalue directions that simple models cannot reach — that most favours complexity. The upward-sloping VoC curves in equity markets are consistent with relatively low concentration and low alignment in equity return data (Kelly and Malamud, 2025). In FX, our estimates of $\hat{b}_* \approx 0$ suggest a more fundamental failure: the overall level of predictive signal is indistinguishable from zero regardless of how it might be distributed across eigenvalue directions, so neither condition can rescue the VoC.

We estimate b_* at each rolling window using the Stein-corrected estimator[6]

$$\hat{b}_* \;=\; \frac{\|\hat{\beta}(\lambda)\|^2}{P} \cdot \frac{\mathrm{df}(\lambda)}{T - \mathrm{df}(\lambda) - 1}, \tag{13}$$

where $\hat{\beta}(\lambda)$ is the Ridge coefficient vector and $\mathrm{df}(\lambda)$ is the effective degrees of freedom defined in equation (7) and evaluated at the actual Gram eigenvalues of each rolling training window. The correction factor $\mathrm{df}(\lambda)/[T-\mathrm{df}(\lambda)-1]$ adjusts for the downward bias that Ridge shrinkage introduces into the raw plug-in $\|\hat{\beta}\|^2/P$: because regularisation shrinks all coefficients toward zero, $\|\hat{\beta}\|^2$ systematically underestimates $\|\beta^*\|^2$, and this factor inflates the plug-in to recover

[6]The Stein-corrected estimator in equation (13) is the empirical analog used by Kelly et al. (2024) to back out $\hat{b}_*$ from the Ridge solution at each rolling window (see their Internet Appendix, Table I notes). It can be derived as follows. Under the rotational symmetry of Assumption 4 in Kelly et al. (2024), $E[\beta\beta'] = P^{-1}b_* I_P$. By Lemma 1 of that paper, for any bounded matrix sequence A_P,

$$\beta' A_P \beta \;\xrightarrow{p}\; P^{-1} b_* \,\mathrm{tr}(A_P).$$

Apply this to the squared Ridge coefficient norm, writing $\hat{\beta}(z) = (zI + \hat{\Psi})^{-1} T^{-1} S' y$ where $\hat{\Psi} = T^{-1}S'S$. A standard bias–variance decomposition of $E[\|\hat{\beta}\|^2]$ yields (see also Hastie et al. 2022)

$$E\Big[\|\hat{\beta}\|^2\Big] \;=\; b_* \cdot \frac{\mathrm{df}}{T} \;+\; \sigma^2 \cdot \frac{\mathrm{df}}{T(T-\mathrm{df})},$$

where $\mathrm{df} = \mathrm{tr}\big(\hat{\Psi}(zI + \hat{\Psi})^{-1}\big)$ is the effective degrees of freedom and the second (noise) term is negligible relative to the first when $b_* > 0$. Ignoring the noise term and inverting gives $b_* \approx P^{-1}\|\hat{\beta}\|^2 \cdot T/\mathrm{df}$. The finite-sample Stein correction replaces T with $(T - \mathrm{df} - 1)$ — analogous to the $1/(n-k-1)$ denominator in ordinary regression variance estimation — yielding the unbiased-type estimator

$$\hat{b}_* \;=\; \frac{\|\hat{\beta}\|^2}{P} \cdot \frac{\mathrm{df}}{T - \mathrm{df} - 1}.$$

Substituting into $z^* = c/b_* = P/(b_* T)$ produces the implied optimal regularisation estimate $\hat{z}^* = P/(\hat{b}_* T)$ reported in Tables 7–8.

an approximately unbiased estimate of b_*. The implied optimal regularisation is $\hat{z}^* = c/\hat{b}_*$, with $\hat{z}^* = \infty$ when $\hat{b}_* = 0$.

We evaluate this estimator across a 25-point log-spaced grid $z \in [1, 10^6]$ in addition to the operational $z = 1{,}000$, tracing out a complete effective complexity curve and testing whether any regularisation level would have made Ridge–RFF competitive with the random walk. The results, reported in Section 5.4, reveal $\hat{b}_* \approx 0$ at every rolling window and both predictor sets, with implied $\hat{z}^* \approx 10^{14}$–$10^{15}$ — twelve orders of magnitude above the equity baseline. This structural finding provides a model-theoretic explanation for the empirical OOS VoC curve: because the signal per feature is indistinguishable from zero, no level of regularisation can make Ridge–RFF competitive with the random walk.

Fixed-λ specification (Case 2, comparison). A natural alternative, common in practice, holds $\lambda = 1{,}000$ fixed regardless of T. In Kelly et al.'s notation this corresponds to a KMZ shrinkage parameter $z = \lambda/T = 1{,}000/T$ that *declines* as the training window grows: from $z = 83.3$ at $T = 12$ to $z = 16.7$ at $T = 60$ and $z = 8.3$ at $T = 120$. As equation (11) shows, $\mathrm{df}/T \approx 0.86$ throughout, so the model behaves as near-ridgeless RFF at every window. Cross-window variation in OOS performance therefore conflates declining nominal complexity $c = P/T$ with simultaneously declining regularisation intensity $z = \lambda/T$, making it impossible to attribute any deterioration in Ridge–RFF performance to complexity alone. Reporting results under both specifications allows us to disentangle these effects: performance differences between the two specifications at the same T isolate the role of regularisation intensity, while results common to both speak to genuine complexity effects.

Forecasts. One-month-ahead forecasts are generated as:

$$\widehat{\Delta s}_{t+1|t} = \sum_{i=1}^{P} Z_{i,t}\, \widehat{\beta}_i(\lambda). \tag{14}$$

Because the ω_i weight vectors are drawn randomly, point forecasts depend on the particular draw. To eliminate this Monte Carlo noise, we average forecasts across 1,000 independent draws of the weight matrix before computing any performance statistics.[7]

Nominal complexity and the cross-window comparison. Kelly et al. (2024) define *nominal complexity* as $c = P/T$, which measures the parameter-to-observation ratio before

[7]This aggregation follows the *Table I* approach of Kelly et al. (2024) as endorsed by Buncic (2025): the 1,000 forecast series are averaged first to produce a single mean forecast $\bar{\pi}_t$, and then performance metrics are computed from that single series. This is distinct from averaging performance metrics across 1,000 individual series, which would confound simulation noise with model performance.

accounting for regularisation. Kelly and Malamud (2025) establish that nominal complexity is the key driver of out-of-sample performance in the equity setting when shrinkage is appropriately calibrated: forecast accuracy rises monotonically with c along the VoC curve. Nagel (2025) argues, however, that with $P \gg T$ and persistent predictors, RFF forecasts mechanically approximate a volatility-timed momentum rule, so apparent small-sample gains may reflect recency weighting rather than genuine nonlinear signal extraction.

Our rolling windows of $T = 12$, 60, and 120 months generate nominal complexity ratios of $c = 1{,}000$, 200, and 100, spanning the full range from extreme overparameterisation to moderate complexity. Nominal and effective complexity describe two distinct features of the model: $c = P/T$ measures the richness of the feature approximation, while df/T from equation (10) measures how much of that richness the estimator actually exploits given λ. Under the KMZ specification, df/T falls from approximately 0.58 at $T = 12$ to 0.04 at $T = 120$, reflecting progressive shrinkage as the penalty grows with T; under fixed-λ, $\mathrm{df}/T \approx 0.86$ throughout. The signal strength parameter b_* then determines which regularisation level is optimal: when $b_* \approx 0$, the optimal z^* is large and df/T should be close to zero, consistent with the empirical estimates in Section 5.4. A central question addressed in Section 4 is whether the theoretical VoC prediction — that performance rises with c under appropriate regularisation — survives in FX, where the structural relationship between fundamentals and returns is arguably less amenable to nonlinear approximation than equity returns.

3.5 Market-timing strategy

To assess whether forecast improvements translate into economic value, we construct market-timing portfolios following Kelly et al. (2024). At each date t, the strategy takes a position in a currency proportional to the model's forecast: strategy returns are $\widehat{\Delta s}_{t+1|t} \cdot \Delta s_{t+1}$, so that the magnitude of the forecast determines position size (larger forecasts correspond to larger positions, in either direction). For the random walk benchmark, whose driftless forecast is zero, the position is based instead on the lagged exchange rate return, giving strategy returns $\Delta s_t \cdot \Delta s_{t+1}$.

We evaluate strategies at both the individual currency level and the equal-weighted portfolio level, where positions across all available currencies are averaged at each date using an expanding membership rule (currencies enter the portfolio as their forecast history becomes available, rather than requiring a common start date). Performance is measured by annualised Sharpe ratios (SR), information ratios (IR) relative to the market and to alternative strategies (from spanning regressions), return skewness, and maximum drawdown in standard deviation units.

4 Complexity Versus Parsimony: Out-of-Sample Forecast Evidence

This section presents and discusses the out-of-sample predictive performance of the linear regression model and Ridge–RFF relative to the driftless random walk benchmark and to each other. We employ out-of-sample R^2, Mean Squared Prediction Error (MSPE) ratios, the Clark–West (CW) test (Clark and West, 2007), and the Diebold–Mariano (DM) test (Diebold and Mariano, 1995) with the small-sample correction of Harvey et al. (1997).

The primary specification follows the KMZ implementation (Kelly et al., 2024; Kelly and Malamud, 2025): the regularisation parameter $z = 1{,}000$ is held fixed across all training windows, with the Ridge penalty set to $\lambda = z \times T$ so that regularisation intensity scales proportionally with sample size (Case 1; see Section 3.4). For robustness, Appendix B presents results under a fixed-penalty specification ($\lambda = 1{,}000$ constant, Case 2), in which the implied KMZ parameter $z = \lambda/T$ declines as the training window grows; the contrast between the two cases isolates the confound between nominal complexity and effective regularisation intensity.

4.1 Taylor-rule fundamentals

Taylor-rule fundamentals (Molodtsova and Papell, 2009) serve as the primary predictor set, given their well-established theoretical motivation and strong empirical track record at short horizons. Results under traditional monetary fundamentals (Meese and Rogoff, 1983) are reported alongside as a robustness check, maintaining the direct link to the canonical benchmark of the FX predictability literature. Both predictor sets are evaluated across rolling training windows of $T \in \{12, 60, 120\}$ months; each table collects all three windows into a single set of panels to facilitate cross-window comparisons.

Table 1 reports one-month-ahead out-of-sample forecast accuracy under the KMZ specification ($z = 1{,}000$ fixed, $\lambda = z \times T$, no intercept) for Taylor-rule fundamentals across rolling training windows of $T \in \{12, 60, 120\}$ months. Each panel presents three comparisons: OLS against the driftless random walk (columns 2–3), Ridge–RFF against the random walk (columns 4–5), and Ridge–RFF against OLS (columns 6–7). The results reveal a coherent and quantitatively precise narrative about the roles of nominal complexity, regularisation, and training sample size.

Short window ($T = 12$, Panel A). At the shortest window, the nominal complexity ratio is at its peak ($c = P/T = 1{,}000$) and the evidence strongly favours complexity over parsimony along one dimension while confirming the robustness of the random walk along another.

OLS performs catastrophically. Out-of-sample R^2 values range from -4.49 (GBP) to -28.66 (JPY), indicating that OLS forecasts are far more volatile than the realised exchange rate changes, a consequence of severe overfitting when seven Taylor-rule predictors are estimated on only 12 observations. The CW test rejects equal predictive accuracy in favour of OLS over the random walk for JPY alone (CW $= 1.676^{**}$); for all other currencies the random walk dominates OLS by a wide margin.

Ridge–RFF, by contrast, achieves R^2 values close to zero (range: -0.016 to -0.077), and the CW test favours Ridge–RFF over the random walk for three currencies at conventional significance levels: AUD (1.690^{**}), CHE (1.336^{*}), and JPY (2.540^{***}). This pattern is striking: despite a training window of only 12 observations, Ridge–RFF with $z = 1{,}000$ produces forecasts close enough to the random walk in mean squared error that the CW test cannot reject equal predictive accuracy for five of the eight currencies, and formally favours Ridge–RFF for three of them.

The Ridge–RFF versus OLS comparison (columns 6–7) is unambiguous. MSPE ratios range from 0.034 (JPY) to 0.190 (GBP), indicating that Ridge–RFF reduces forecasting error relative to OLS by 81–97%, and the DM statistic is significant at the 1% level for all currencies except JPY. This reflects the benign overfitting property documented by Kelly et al. (2024): at extreme nominal complexity ($c = 1{,}000$), Ridge regularisation prevents the variance inflation that makes OLS unusable in small samples, compressing the coefficient vector toward zero and thereby stabilising forecasts.

Intermediate window ($T = 60$, Panel B). As the training window grows to 60 months, the nominal complexity falls ($c = 200$) and the effective degrees of freedom contract further: equation (10) gives $\mathrm{df}/T \approx 0.05$, so the model retains only $\mathrm{df} \approx 3$ effective degrees of freedom — aggregate flexibility equivalent to fitting three unconstrained parameters, compared with $T = 60$ under OLS. OLS improves substantially: R^2_{Lin} magnitudes shrink considerably (range: -0.213 to -0.292), and the CW test favours OLS over the random walk for GBP (1.849^{**}), JPY (1.420^{*}), and NOK (1.783^{**}).

The most notable result in Panel B is that Ridge–RFF produces *positive* OOS R^2 for three currencies: AUD (0.011), CHE (0.007), and JPY (0.015). The corresponding CW statistics are significant at the 5% level for AUD (2.251) and CHE (1.881), and at the 1% level for JPY (2.811). These results indicate that Ridge–RFF with $z = 1{,}000$ generates statistically reliable gains over the random walk for a subset of currencies at the intermediate training

Table 1: Out-of-sample forecast accuracy: Taylor-rule fundamentals

	Linear vs. RW		Ridge–RFF vs. RW		Ridge–RFF vs. Linear	
Currency	R^2_{Lin}	CW	R^2_{RFF}	CW	Ratio	DM
Panel A: Rolling window $T = 12$ months						
AUD	−10.167	0.590	−0.016	1.690**	0.091	4.914***
CAD	−6.256	−1.057	−0.077	−0.703	0.148	7.351***
CHE	−5.445	−0.251	−0.037	1.336*	0.161	6.167***
EUR	−4.604	−0.102	−0.051	0.803	0.188	6.750***
GBP	−4.490	0.953	−0.042	0.359	0.190	7.228***
JPY	−28.664	1.676**	−0.021	2.540***	0.034	1.279
NOK	−5.868	0.409	−0.057	−0.166	0.154	8.404***
SEK	−6.412	0.900	−0.059	0.191	0.143	6.198***
Panel B: Rolling window $T = 60$ months						
AUD	−0.234	0.721	0.011	2.251**	0.801	4.246***
CAD	−0.292	0.429	−0.014	−0.634	0.785	3.765***
CHE	−0.213	0.325	0.007	1.881**	0.819	2.947***
EUR	−0.221	1.202	−0.006	0.385	0.824	2.059**
GBP	−0.218	1.849**	−0.007	−0.007	0.827	3.651***
JPY	−0.282	1.420*	0.015	2.811***	0.768	5.124***
NOK	−0.221	1.783**	−0.005	0.243	0.823	4.398***
SEK	−0.258	0.773	−0.003	0.733	0.797	4.287***
Panel C: Rolling window $T = 120$ months						
AUD	−0.046	2.043**	0.009	2.127**	0.948	1.346
CAD	−0.062	1.145	−0.005	−0.381	0.946	1.323
CHE	−0.105	0.448	0.006	1.649**	0.899	3.125***
EUR	−0.092	2.139**	−0.003	0.201	0.918	1.194
GBP	−0.126	0.786	−0.002	0.058	0.890	2.945***
JPY	−0.124	0.258	0.013	2.651***	0.878	3.461***
NOK	−0.098	0.972	−0.002	0.104	0.912	2.569**
SEK	−0.126	−0.040	0.001	0.856	0.887	3.204***

This table reports one-month-ahead out-of-sample forecast accuracy statistics for Taylor-rule fundamentals, estimated under the KMZ exact specification with regularisation parameter $z = 1{,}000$ held fixed across all training windows (sklearn penalty $\lambda = z \times T$, no intercept, driftless random walk benchmark). Taylor-rule predictors comprise U.S. and foreign nominal interest rates, inflation rates, output gaps, and the real exchange rate, following Molodtsova and Papell (2009). R^2_{Lin} (R^2_{RFF}) is the OOS R^2 of the linear regression (Ridge–RFF) model relative to the driftless random walk, defined as $1 - \text{MSPE}_{\text{model}}/\text{MSPE}_{\text{RW}}$; negative values indicate the model is less accurate than the random walk. CW is the Clark–West (2007) test statistic for equal predictive accuracy against the random walk; a positive and significant value favours the first-named model over the random walk. Ratio is $\text{MSPE}_{\text{RFF}}/\text{MSPE}_{\text{Lin}}$; values below unity indicate Ridge–RFF outperforms the linear model. DM is the Diebold–Mariano (1995) test statistic with the Harvey et al. (1997) small-sample correction comparing Ridge–RFF against the linear model; a positive (negative) significant value indicates Ridge–RFF outperforms (underperforms) the linear model. Stars on CW use one-sided critical values; stars on DM use two-sided critical values.

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

window. The DM statistics confirm that Ridge–RFF beats OLS across all eight currencies, with significance at the 5% level or better for six of them.

This pattern reflects the dual effect of the KMZ specification as T grows: the growing absolute penalty $\lambda = z \times T$ simultaneously reduces variance relative to small-sample OLS and, unlike Case 2's fixed λ, maintains adequate shrinkage intensity to prevent forecast variance from inflating as the sample expands. The net result is a model that extracts the persistent predictive content of Taylor-rule fundamentals — interest rate differentials, output gaps, and real exchange rates — while suppressing the noise that dominates OLS-based predictions.

Long window ($T = 120$, Panel C). At the longest training window, the nominal complexity has fallen further ($c = 100$) and the effective degrees of freedom are $\mathrm{df}/T \approx 0.04$ (equation (10)), implying the model retains only $\mathrm{df} \approx 5$ effective degrees of freedom — aggregate flexibility equivalent to fitting five unconstrained parameters, compared with $T = 120$ under OLS.[8] Ridge regularisation therefore imposes substantial but not total shrinkage, producing forecasts that differ from the driftless random walk by a small but non-zero margin.

Yet despite this extreme regularisation, the statistical evidence supports predictive gains for the majority of currencies. Five of eight currencies record positive R^2_{RFF}: AUD (0.009), CHE (0.006), JPY (0.013), NOK (-0.002, effectively zero), and SEK (0.001). CW tests are significant at conventional levels for AUD (2.127^{**}), CHE (1.649^{**}), JPY (2.651^{***}), NOK (2.569^{**}), and SEK (3.204^{***}) — five out of eight currencies. The DM test confirms that Ridge–RFF still outperforms OLS for most currencies, significantly so for CHE, GBP, JPY, NOK, and SEK at the 1% or 5% level.

For OLS, AUD (2.043^{**}) and EUR (2.139^{**}) show significant CW statistics, confirming that linear regression can itself beat the random walk at longer horizons when Taylor-rule fundamentals are used, consistent with Molodtsova and Papell (2009). The fact that Ridge–RFF improves further on OLS at $T = 120$ is notable: even as the model's effective complexity progressively declines toward zero (without reaching it at any finite training window), it retains sufficient directional accuracy to dominate OLS in MSE terms, with MSPE ratios between 0.878 (JPY) and 0.948 (AUD).

Taken together, the three panels paint a consistent picture of increasing model refinement as the training window expands. At $T = 12$, regularisation compensates for the severe estimation noise of OLS and brings Ridge–RFF close to the random walk boundary. At $T = 60$ and $T = 120$, sufficient data accumulates that a heavily regularised model — one

[8] As established in Section 3.4, $\mathrm{df}/T \to 0$ as $T \to \infty$ with z fixed (equation (10)), but the effective parameter count remains strictly positive at every finite T in the evaluation sample.

with an effective parameter count of only df ≈ 3 and ≈ 5 respectively, compared with the full T under OLS, but still generating non-trivially shrunk forecasts — can identify the persistent predictive signal in Taylor-rule fundamentals, yielding statistically reliable gains over the random walk for a majority of currencies. The pattern does not represent a universal defeat of the random walk: cross-currency averages of R^2_{RFF} remain close to zero or slightly negative at each window, and the gains are concentrated in a subset of currencies. It does, however, challenge the canonical view that no structural model can systematically outperform the driftless random walk at monthly horizons (Meese and Rogoff, 1983).

4.2 Traditional monetary fundamentals

Table 2 presents results under traditional monetary fundamentals (money growth, interest rate, inflation, and output differentials; Meese and Rogoff, 1983). The qualitative ordering of results is broadly similar to Table 1, but the magnitudes and the currency-level evidence differ materially, revealing the additional predictive content embedded in Taylor-rule relative to traditional monetary fundamentals.

Similarities. The three-panel narrative structure carries over. At $T = 12$ (Panel A), OLS severely overfits (though the magnitudes are smaller: R^2_{Lin} ranges from -0.636 to -1.274, compared with -4.49 to -28.66 under Taylor rule), and Ridge–RFF dramatically outperforms OLS, with MSPE ratios of 0.478–0.675 and DM statistics significant at the 1% level for all currencies. Only JPY achieves a significant CW statistic favouring Ridge–RFF over the random walk (2.397^{***}), echoing the Taylor-rule result.

At $T = 120$ (Panel C), linear OLS improves meaningfully: CAD achieves CW $= 2.471^{***}$ and CHE achieves CW $= 1.654^{**}$, and NOK records a positive $R^2_{\text{Lin}} = 0.016$ (the only instance of positive OLS R^2 in either table), consistent with the longer-horizon predictability documented by Molodtsova and Papell (2009). The DM statistics at $T = 120$ confirm that Ridge–RFF continues to outperform OLS for several currencies (AUD, JPY, SEK), though the margin narrows as the growing training window reduces the estimation noise advantage that regularisation confers over plain OLS.

Key differences. Three differences between the tables stand out and define the superior performance of Taylor-rule fundamentals.

First, at the intermediate window ($T = 60$, Panel B), traditional monetary fundamentals produce no currency for which Ridge–RFF achieves a positive or significant positive CW against the random walk. All eight R^2_{RFF} values remain negative (range: -0.008 to -0.031),

Table 2: Out-of-sample forecast accuracy: traditional monetary fundamentals (robustness)

	Linear vs. RW		Ridge–RFF vs. RW		Ridge–RFF vs. Linear	
Currency	R^2_{Lin}	CW	R^2_{RFF}	CW	Ratio	DM
Panel A: Rolling window $T = 12$ months						
AUD	−1.274	0.491	−0.087	−0.148	0.478	3.862***
CAD	−0.726	1.767**	−0.041	1.045	0.603	5.340***
CHE	−1.007	0.678	−0.068	0.672	0.532	4.550***
EUR	−1.156	−0.188	−0.093	−0.305	0.507	2.872***
GBP	−0.754	−0.886	−0.075	0.329	0.612	6.307***
JPY	−0.990	1.084	−0.032	2.397***	0.519	3.940***
NOK	−0.636	−0.506	−0.104	−0.854	0.675	3.017***
SEK	−0.771	−0.428	−0.074	−0.006	0.606	4.829***
Panel B: Rolling window $T = 60$ months						
AUD	−0.155	0.964	−0.023	−0.969	0.886	2.200**
CAD	−0.060	2.294**	−0.015	−0.046	0.958	1.242
CHE	−0.097	0.718	−0.031	−1.606*	0.940	1.486
EUR	−0.077	−0.276	−0.020	−0.420	0.947	1.544
GBP	−0.056	0.101	−0.008	0.641	0.955	2.234**
JPY	−0.109	0.312	−0.015	0.433	0.915	2.877***
NOK	−0.093	−0.332	−0.029	−0.833	0.941	1.188
SEK	−0.097	−1.062	−0.012	0.090	0.923	2.604***
Panel C: Rolling window $T = 120$ months						
AUD	−0.067	0.257	−0.011	−0.816	0.948	2.139**
CAD	−0.010	2.471***	−0.008	−0.126	0.998	0.087
CHE	−0.019	1.654**	−0.002	0.621	0.983	0.325
EUR	−0.017	0.539	−0.010	−0.393	0.993	0.244
GBP	−0.024	0.409	−0.003	0.509	0.979	1.455
JPY	−0.060	−0.204	0.000	0.963	0.944	2.157**
NOK	0.016	1.096	−0.005	0.119	1.021	−0.339
SEK	−0.057	−1.300*	−0.010	−0.297	0.955	2.003**

This table reports one-month-ahead out-of-sample forecast accuracy statistics for traditional monetary fundamentals, serving as a robustness counterpart to Table 1. The specification is identical: KMZ exact with $z = 1{,}000$ fixed, $\lambda = z \times T$, no intercept, driftless random walk benchmark. Traditional monetary predictors are money supply growth, interest rate, inflation, and output growth differentials between the U.S. and each foreign country, following Meese and Rogoff (1983). All statistics are as defined in Table 1.

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

and the CW statistics are uniformly insignificant or negative. By contrast, Taylor-rule Ridge–RFF achieves positive R^2 and significant CW for three currencies at the same window. More strikingly, CHE under traditional fundamentals records CW $= -1.606^*$ for Ridge–RFF against the random walk, indicating that the random walk *significantly* outperforms Ridge–RFF for this currency at $T = 60$ — a result not observed anywhere in the Taylor-rule table.

Second, at $T = 120$ (Panel C), Ridge–RFF under traditional fundamentals fails to generate a single statistically significant positive R^2 against the random walk. All R^2_{RFF} values are either negative or effectively zero (JPY: -0.000), and no CW statistic favours Ridge–RFF over the random walk at standard significance levels. The contrast with Taylor rule, where five currencies show significant CW statistics and Ridge–RFF beats the random walk for a majority of the panel, is economically meaningful: it suggests that the output gap and real exchange rate components specific to Taylor-rule fundamentals carry genuine month-ahead predictive content that Ridge regularisation can extract, whereas the Meese–Rogoff monetary variables provide insufficient signal for the regularised model to consistently overcome the random walk even at long training windows.

Third, the MSPE ratios (Ridge–RFF over OLS) are uniformly larger under traditional fundamentals than under Taylor rule at $T = 60$ and $T = 120$. At $T = 60$, traditional ratios range from 0.886 to 0.958 (Ridge–RFF advantage of 4–11%), compared with 0.768–0.827 under Taylor rule (advantage of 17–23%). At $T = 120$, traditional ratios cluster between 0.944 and 1.021, with NOK recording a ratio above unity (1.021), meaning OLS actually achieves lower MSPE than Ridge–RFF for that currency. No such reversal occurs under Taylor rule. These differences reflect the weaker predictive signal in traditional monetary fundamentals at short and medium horizons: with less signal to extract, the regularised model gains less relative to OLS, and the marginal advantage of complexity over parsimony is correspondingly smaller.

Interpretation. The comparison across predictor sets confirms that the performance of Ridge–RFF against the random walk depends critically on the information content of the underlying fundamentals. Taylor-rule variables — which encode forward-looking monetary policy behaviour through output gaps, inflation, and real exchange rates — appear to contain persistent short-horizon predictive information that Ridge regularisation can successfully extract under the KMZ specification. Traditional monetary fundamentals, while foundational to the FX forecasting literature and directly comparable to the Meese and Rogoff (1983) specification, provide a weaker signal that is insufficient to drive systematic gains over the random walk under the same specification. This finding aligns with the broader literature's

assessment that the conditions under which FX predictability emerges depend on predictor choice as well as estimation methodology (Rossi, 2013; Molodtsova and Papell, 2009).

Importantly, the results under both predictor sets confirm the structural diagnosis of Section 5.4: the estimated signal strength $\hat{b}_* \approx 0$ (equation (13)) implies an optimal regularisation $\hat{z}^*$ many orders of magnitude above the equity baseline, and a forecast that converges to the random walk only in the limit as $T \to \infty$. The positive R^2_{RFF} values observed for several currencies in the Taylor-rule table at $T = 60$ and $T = 120$ are best understood not as evidence of genuine nonlinear structure extraction but as a consequence of sufficient shrinkage bringing the model close to the random walk benchmark while retaining the modest directional accuracy embedded in persistent Taylor-rule fundamentals.

4.3 Robustness: intercept specification

Buncic (2025) argues that the zero-intercept restriction common to the VoC literature introduces a methodological bias that may overstate the apparent benefits of high-dimensional complexity. Specifically, forcing all models — OLS, Ridge–RFF, and the benchmark — through the origin aligns the competing models with the driftless random walk by construction, potentially inflating the relative performance of regularised high-dimensional estimators whose forecasts naturally contract toward zero. He recommends including an intercept as a minimal robustness check. Tables 3 and 4 address this critique directly by replicating the KMZ specification with an intercept added to both the linear and Ridge–RFF models. The benchmark correspondingly becomes the *drifting* random walk, whose forecast at each date equals the rolling sample mean of the target variable — a harder benchmark at longer horizons when exchange rate returns exhibit a persistent non-zero mean.

Taylor-rule fundamentals. The with-intercept results for Taylor-rule fundamentals (Table 3) are qualitatively consistent with their no-intercept counterparts (Table 1) along all three comparison dimensions.

At $T = 12$ (Panel A), the DM statistics are uniformly positive and significant at the 5% or 1% level for all eight currencies under both specifications, confirming that Ridge–RFF outperforms OLS regardless of whether an intercept is included. The CW test favours Ridge–RFF over the drifting random walk for AUD (1.829^{**}) and JPY (1.473^{*}) — compared with AUD, CHE, and JPY in the no-intercept table — with the broader pattern of near-zero R^2_{RFF} values against much more negative R^2_{Lin} values preserved across both specifications. The intercept does not alleviate the severe overfitting of OLS in small samples: R^2_{Lin} values range from -9.997 (CAD) to -24.169 (AUD), confirming that the estimation noise of a

seven-predictor linear model on twelve observations is essentially unchanged by the addition of an intercept term.

At $T = 60$ (Panel B), both intercept specifications tell the same story regarding the Ridge–RFF versus OLS comparison. Under the no-intercept specification, MSPE ratios are uniformly below unity across all eight currencies (range 0.768–0.827, Table 1, Panel B), with the DM test significant at the 5% or 1% level for all eight currencies. The with-intercept specification produces equally uniform evidence in favour of Ridge–RFF: MSPE ratios range from 0.771 to 0.871 and the DM test is again significant at the 1% level for all eight currencies. The two specifications are thus consistent in showing Ridge–RFF outperforming OLS at $T = 60$, with comparable magnitudes (cross-currency mean ratio of 0.79 under with-intercept versus 0.81 under no-intercept). The key difference between specifications at this window lies in the comparison against the random walk benchmark. Under the no-intercept specification, Ridge–RFF achieves positive R^2_{RFF} for AUD (0.011), CHE (0.007), and JPY (0.015) against the driftless random walk (Table 1, Panel B). Under the with-intercept specification, all R^2_{RFF} values are negative (range -0.004 to -0.035) against the harder drifting random walk benchmark, though the CW test continues to favour Ridge–RFF for the same three currencies: AUD (2.296^{**}), CHE (1.990^{**}), and JPY (3.374^{***}). The sign change in R^2 without change in CW significance reflects the drifting benchmark absorbing the rolling mean return: once the mean is accounted for by the benchmark, Ridge–RFF no longer reduces MSE enough to record a positive R^2, but its directional advantage remains statistically detectable.

At $T = 120$ (Panel C), both intercept specifications deliver broadly similar conclusions. Ridge–RFF achieves positive R^2_{RFF} for CHE (0.001) and JPY (0.013) against the drifting random walk, matching the no-intercept results exactly for JPY and nearly so for CHE. The CW test favours Ridge–RFF for AUD (2.345^{***}), CHE (1.850^{**}), and JPY (2.496^{***}) — the same three currencies that are significant in the no-intercept table. Two currencies significant in the no-intercept table — NOK (2.569^{**}) and SEK (3.204^{***}) — lose significance in the with-intercept table (NOK: 0.107, SEK: 1.023), consistent with the drifting random walk being a harder benchmark for currencies where the rolling sample mean is a reasonable forecast of the exchange rate return. The DM statistics at $T = 120$ confirm that Ridge–RFF continues to outperform OLS: five currencies are significant at the 1% level (CHE, GBP, JPY, NOK, SEK) and AUD is significant at the 10% level, with no currency showing a negative and significant DM statistic.

Traditional monetary fundamentals. Table 4 largely replicates the patterns of Table 2 for traditional monetary fundamentals, with one noteworthy difference.

Table 3: Out-of-sample forecast accuracy: Taylor-rule fundamentals (with-intercept specification)

	Linear vs. RW		Ridge–RFF vs. RW		Ridge–RFF vs. Linear	
Currency	R^2_{Lin}	CW	R^2_{RFF}	CW	Ratio	DM
Panel A: Rolling window $T = 12$ months						
AUD	−24.169	1.099	−0.089	1.829**	0.043	2.426**
CAD	−9.997	−0.997	−0.148	−0.701	0.104	6.634***
CHE	−10.458	−0.534	−0.115	0.725	0.097	5.764***
EUR	−10.643	0.187	−0.136	0.770	0.098	3.234***
GBP	−22.400	0.910	−0.114	0.952	0.048	2.259**
JPY	−10.554	0.664	−0.079	1.473*	0.094	6.963***
NOK	−12.495	0.320	−0.148	−0.720	0.085	5.583***
SEK	−10.820	1.598*	−0.123	0.492	0.095	5.559***
Panel B: Rolling window $T = 60$ months						
AUD	−0.285	0.841	−0.008	2.296**	0.784	4.502***
CAD	−0.340	0.094	−0.033	−1.135	0.771	4.626***
CHE	−0.303	0.318	−0.012	1.990**	0.777	3.856***
EUR	−0.286	1.503*	−0.022	0.268	0.795	2.290**
GBP	−0.188	2.549***	−0.035	0.135	0.871	2.767***
JPY	−0.299	1.074	−0.004	3.374***	0.773	5.458***
NOK	−0.285	1.069	−0.023	−0.011	0.796	4.853***
SEK	−0.326	0.246	−0.024	0.320	0.772	4.271***
Panel C: Rolling window $T = 120$ months						
AUD	−0.072	1.590*	−0.001	2.345***	0.933	1.762*
CAD	−0.092	1.329*	−0.017	0.124	0.931	1.528
CHE	−0.140	0.601	0.001	1.850**	0.876	3.081***
EUR	−0.131	1.784**	−0.012	0.652	0.895	1.629
GBP	−0.144	1.738**	−0.018	0.256	0.890	2.712***
JPY	−0.146	−0.279	0.013	2.496***	0.861	3.872***
NOK	−0.128	0.490	−0.015	0.107	0.900	3.002***
SEK	−0.150	−0.131	−0.016	1.023	0.884	3.364***

This table reports one-month-ahead out-of-sample forecast accuracy statistics for Taylor-rule fundamentals under the KMZ exact specification ($z = 1{,}000$ fixed, $\lambda = z \times T$) with intercept. Both the linear and Ridge–RFF models include a fitted intercept term, and the random walk benchmark is a *drifting* random walk whose forecast at each date equals the sample mean of the target variable over the training window. This table serves as a robustness counterpart to Table 1, which uses the no-intercept specification and driftless random walk benchmark. See, Table 1.

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

At $T = 12$ (Panel A), the results are essentially unchanged: the DM statistics are uniformly positive and significant at the 1% level for all eight currencies, confirming that Ridge–RFF substantially outperforms OLS with and without an intercept when nominal complexity is extreme ($c = 1{,}000$). The CW test favours Ridge–RFF over the drifting random walk only for CAD (2.116^{**}); JPY achieves CW = 1.151, below the 10% significance threshold.

At $T = 60$ (Panel B), the intercept specification affects the traditional fundamentals results more sharply than under Taylor rule. For NOK, the CW statistic for Ridge–RFF against the drifting random walk is -3.266^{***}, indicating the drifting random walk significantly outperforms Ridge–RFF for that currency — a stronger adverse result than the -1.288^{*} recorded in the no-intercept table. Similarly, CHE records CW = -2.288^{**} (with intercept) versus -1.606^{*} (no intercept). These results reflect a specific feature of the drifting random walk benchmark: for currencies with persistent positive or negative rolling means (such as NOK and CHE during portions of the evaluation period), the drifting random walk has a genuine forecasting advantage over models that apply strong regularisation toward zero, since the latter systematically miss a predictable component that the rolling mean captures. Notably, however, the DM statistics at $T = 60$ are positive and significant for seven of eight currencies (all except NOK, where DM = 1.075), indicating that Ridge–RFF still outperforms OLS across most of the cross-section despite losing ground to the drifting random walk for two currencies.

At $T = 120$ (Panel C), the evidence for OLS outperforming the drifting random walk strengthens with the intercept: CAD (CW = 2.405^{***}) and CHE (CW = 1.785^{**}) are significant, consistent with the no-intercept results in Table 2. Ridge–RFF retains a meaningful advantage over OLS for most currencies: the DM statistic is positive and significant at the 5% or 1% level for AUD, GBP, and JPY. The sole exception is NOK (Ratio = 1.019, DM = -0.308), where OLS marginally dominates; this currency's rolling mean is a sufficiently informative component that the near-zero Ridge forecasts cannot compete with a benchmark that explicitly targets the mean return.

Assessment of Buncic's critique. The with-intercept results allow a direct empirical evaluation of Buncic's methodological concern. Two conclusions follow from the comparison.

First, the central statistical finding of the paper — that Ridge–RFF does not systematically beat the random walk benchmark — is entirely robust to the intercept specification. For both Taylor-rule and traditional monetary fundamentals, neither model consistently achieves positive R^2 or significant positive CW statistics against the appropriate random walk benchmark across currencies and training windows, whether that benchmark is driftless or drifting. The OOS VoC curve evidence of Section 5 confirms this at the level of the full

Table 4: Out-of-sample forecast accuracy: traditional monetary fundamentals (with-intercept specification)

	Linear vs. RW		Ridge–RFF vs. RW		Ridge–RFF vs. Linear	
Currency	R^2_{Lin}	CW	R^2_{RFF}	CW	Ratio	DM
Panel A: Rolling window $T = 12$ months						
AUD	−1.383	−0.354	−0.127	−0.458	0.473	4.840***
CAD	−0.968	1.808**	−0.072	2.116**	0.545	5.573***
CHE	−1.164	0.857	−0.106	0.334	0.511	5.750***
EUR	−1.546	0.131	−0.125	−0.386	0.442	3.080***
GBP	−1.067	−0.799	−0.118	0.356	0.541	7.141***
JPY	−1.071	1.476*	−0.060	1.151	0.512	4.793***
NOK	−0.938	0.042	−0.154	0.054	0.596	4.143***
SEK	−1.064	−0.658	−0.113	0.081	0.539	5.078***
Panel B: Rolling window $T = 60$ months						
AUD	−0.180	1.049	−0.031	−1.119	0.873	2.147**
CAD	−0.083	2.088**	−0.021	−0.482	0.942	1.685*
CHE	−0.128	0.685	−0.041	−2.288**	0.923	1.989**
EUR	−0.120	−0.395	−0.025	−1.022	0.915	2.283**
GBP	−0.088	0.131	−0.014	1.533*	0.933	3.833***
JPY	−0.132	0.395	−0.025	0.769	0.905	3.486***
NOK	−0.090	−0.088	−0.034	−3.266***	0.949	1.075
SEK	−0.112	−0.059	−0.020	−0.338	0.917	2.576**
Panel C: Rolling window $T = 120$ months						
AUD	−0.088	0.164	−0.016	−0.310	0.934	2.538**
CAD	−0.024	2.405***	−0.012	1.102	0.988	0.608
CHE	−0.022	1.785**	−0.007	0.366	0.986	0.279
EUR	−0.019	1.128	−0.014	0.121	0.995	0.180
GBP	−0.043	0.373	−0.002	1.949**	0.961	2.949***
JPY	−0.061	−0.356	−0.004	−0.317	0.946	2.209**
NOK	0.010	0.984	−0.009	−1.505*	1.019	−0.308
SEK	−0.049	−0.638	−0.011	−0.469	0.964	1.541

This table reports one-month-ahead out-of-sample forecast accuracy statistics for traditional monetary fundamentals under the KMZ exact specification ($z = 1{,}000$ fixed, $\lambda = z \times T$) with intercept, serving as a robustness counterpart to Table 2 (no-intercept) and the with-intercept counterpart to Table 3 for traditional fundamentals. Both models include a fitted intercept, and the random walk benchmark is a *drifting* random walk whose forecast equals the sample mean of returns over the training window. Traditional monetary predictors are money supply growth, interest rate, inflation, and output growth differentials between the U.S. and each foreign country, following Meese and Rogoff (1983). All statistics are otherwise as defined in Table 1.
* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

regularisation grid: the MSPE ratio relative to the random walk is bounded above by unity at every z value in both intercept specifications (Figure 8).

Second, and contrary to Buncic's prediction, the advantage of Ridge–RFF *over OLS* is not diminished by the inclusion of an intercept. Under the KMZ specification at $T = 60$, MSPE ratios are uniformly below unity and the DM test is significant for all eight currencies under *both* the no-intercept and with-intercept specifications. The cross-currency mean MSPE ratio is 0.81 without an intercept and 0.79 with an intercept — essentially identical and both decisively in favour of Ridge–RFF. This pattern arises because the intercept imposes an additional estimation burden on OLS that Ridge regularisation offsets; the net effect on the Ridge–RFF versus OLS comparison is negligible. Buncic's concern about overstatement of VoC benefits does not apply in the FX context. The fundamental impediment is not the zero-intercept restriction but rather the near-zero predictive signal in exchange rate fundamentals that makes it difficult for any model to outperform the random walk, with or without an intercept.

4.4 Market-Timing Portfolio Performance

Statistical results presented so far establish that Ridge-RFF delivers no uniform out-of-sample performance relative to random walk across currencies for either predictor set or training window. This subsection examines whether this statistical finding precludes economic value. We evaluate magnitude-proportional market-timing strategies in which each currency's position at date t equals the model's one-step-ahead forecast, so that forecast magnitude and direction jointly determine position size (following Kelly et al., 2024; Buncic, 2025). Performance is measured by annualised Sharpe ratios, associated t-statistics for mean strategy returns, and information ratios (IR) relative to the equal-weighted market return, the random walk strategy, and the competing model. We report results for both predictor sets and both intercept specifications. Individual-currency results and the random walk single-currency benchmark are reported in Appendix C.

Tables 5 and 6 report equal-weighted (EW) cross-currency portfolio performance under the no-intercept and with-intercept specifications respectively. Each table includes the random walk (RW) portfolio as a common benchmark and covers both the traditional monetary (Trad) and Taylor-rule (Taylor) predictor sets across all three training windows.

No-intercept baseline (Table 5). At the shortest window ($T = 12$), no EW strategy achieves conventional statistical significance: the Taylor LIN and RFF portfolios record Sharpe ratios of 0.166 and 0.232 respectively, against the RW benchmark of 0.149. The RFF information ratio relative to the market is marginally significant ($t = 1.649^{*}$), but the picture

at this horizon is one of noisy short-sample returns that do not aggregate into a reliable portfolio result.

Table 5: Market-timing performance: equal-weighted portfolio (no-intercept specification)

		SR	t	IR v. Mkt	t	IR v. RW	t	IR v. LIN	t	Skew	Max Loss
Training window size $T = 12$											
	RW	0.149	1.037	0.149	1.035					−0.075	5.662
Trad	LIN	0.058	0.403	0.057	0.394	0.044	0.306			−0.412	6.903
	RFF	0.177	1.228	0.176	1.224	0.138	0.955	0.168	1.168	0.150	5.577
Taylor	LIN	0.166	1.162	0.165	1.152	0.087	0.606			−0.426	7.013
	RFF	0.232	1.620	0.236	1.649*	0.184	1.273	0.182	1.265	0.232	5.633
Training window size $T = 60$											
	RW	0.173	1.148	0.172	1.143					0.651	5.435
Trad	LIN	0.170	1.133	0.170	1.130	0.174	1.155			0.234	7.425
	RFF	−0.052	−0.344	−0.050	−0.332	−0.065	−0.431	−0.092	−0.609	−1.181	7.688
Taylor	LIN	0.333	2.233**	0.336	2.248**	0.288	1.907*			−0.286	7.969
	RFF	0.352	2.360**	0.356	2.377**	0.319	2.114**	0.225	1.500	0.314	4.889
Training window size $T = 120$											
	RW	0.156	0.977	0.159	0.994					1.053	5.241
Trad	LIN	0.281	1.763*	0.288	1.799*	0.288	1.798*			1.664	7.303
	RFF	0.047	0.296	0.035	0.217	0.047	0.295	0.020	0.125	−0.348	5.397
Taylor	LIN	0.274	1.729*	0.273	1.722*	0.226	1.409			1.229	4.967
	RFF	0.376	2.375**	0.376	2.368**	0.350	2.189**	0.292	1.834*	0.289	5.121

This table reports one-month-ahead out-of-sample market-timing performance for equal-weighted (EW) cross-currency portfolios under traditional monetary (Trad) and Taylor-rule (Taylor) fundamentals. No intercept is included in either model; the benchmark is the driftless random walk. The EW portfolio averages positions across all available currencies at each date using an expanding-membership rule. Strategy returns are magnitude-proportional: $\hat{r}_{t|t-1} \times r_t$ for LIN and RFF, and $r_{t-1} \times r_t$ for the RW strategy. SR is the annualised Sharpe ratio; t is the t-statistic for mean portfolio returns. IR v. Mkt is the information ratio vs. the equal-weighted buy-and-hold return; IR v. RW is the information ratio relative to the random walk portfolio; IR v. LIN (RFF rows only) is the information ratio of Ridge–RFF relative to the linear portfolio. Skew and Max Loss are return skewness and maximum loss in standard deviation units, respectively.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-sided).

At $T = 60$, the Taylor-rule predictor set delivers clear and statistically reliable gains. Both Taylor LIN (SR = 0.333, $t = 2.233^{**}$) and Taylor RFF (SR = 0.352, $t = 2.360^{**}$) significantly outperform the RW benchmark (SR = 0.173), with positive and significant information ratios relative to the market ($t = 2.248^{**}$ and 2.377^{**} respectively) and relative to the random walk strategy ($t = 1.907^{*}$ and 2.114^{**}). The RFF information ratio relative to the linear strategy is positive (IR = 0.225) but below conventional significance ($t = 1.500$), indicating that the two models offer broadly comparable directional accuracy at this horizon once currency-level noise is averaged out. Traditional monetary fundamentals, by contrast, deliver no significant EW portfolio returns at $T = 60$: the Trad LIN portfolio (SR = 0.170, $t = 1.133$) does not materially improve on the RW benchmark, and Trad RFF turns mildly negative (SR = −0.052).

At the longest window ($T = 120$), Taylor RFF delivers the strongest EW result in the table: SR = 0.376 ($t = 2.375^{**}$), with significant information ratios relative to both the market ($t = 2.368^{**}$) and the random walk strategy ($t = 2.189^{**}$), and a marginally significant incremental return over the linear strategy (IR v. LIN $t = 1.834^{*}$). Taylor LIN is also marginally significant (SR = 0.274, $t = 1.729^{*}$). For traditional monetary fundamentals, the linear model achieves marginal significance at $T = 120$ (SR = 0.281, $t = 1.763^{*}$), consistent with the longer-horizon predictability of traditional fundamentals documented in Section 4, while Trad RFF adds nothing (SR = 0.047, $t = 0.296$). The incremental value of Ridge–RFF over OLS is therefore predictor-set specific: it is economically and statistically meaningful under Taylor-rule fundamentals at both $T = 60$ and $T = 120$, but absent under traditional monetary fundamentals.

With-intercept specification (Table 6). The with-intercept results expose a sharp and informative discontinuity. At $T = 12$, the Taylor-rule portfolio Sharpe ratios are broadly unchanged relative to the no-intercept baseline: LIN falls modestly from 0.166 to 0.147 and RFF from 0.232 to 0.207, with neither achieving significance under either specification. This near-invariance at the short window reflects the fact that the unconditional drift of each currency is small relative to within-window variation at $T = 12$, so removing it via an intercept has limited effect on forecast weights.

At $T = 60$ and $T = 120$, the Taylor-rule results collapse entirely. Taylor LIN falls from SR = 0.333 ($t = 2.233^{**}$) to SR = 0.151 ($t = 1.010$) at $T = 60$, and from SR = 0.274 ($t = 1.729^{*}$) to SR = 0.138 ($t = 0.870$) at $T = 120$. Taylor RFF falls from SR = 0.352 ($t = 2.360^{**}$) to SR = 0.108 ($t = 0.721$) at $T = 60$, and from SR = 0.376 ($t = 2.375^{**}$) to SR = 0.072 ($t = 0.456$) at $T = 120$. No Taylor-rule EW strategy achieves conventional significance at $T = 60$ or $T = 120$ under the with-intercept specification.

The mechanism is the drift-capture artefact of the zero-intercept restriction. Under no-intercept estimation, the slope weights are constrained to absorb the unconditional mean of each exchange rate return, generating persistent position biases that track each currency's long-run drift rather than its conditional directional signal. When the intercept is freed, these drift components are captured by the constant term and removed from the slope weights, leaving only the conditional signal — which, at $T = 60$ and $T = 120$, is not sufficient to produce statistically reliable EW portfolio returns for the Taylor-rule predictor set. The exception that confirms this interpretation is the Taylor RFF information ratio relative to the random walk at $T = 60$, which remains significant under the with-intercept specification ($t = 2.063^{**}$): even without a significant unconditional Sharpe, Ridge–RFF

Table 6: Market-timing performance: equal-weighted portfolio (with-intercept specification)

		SR	t	IR v. Mkt	t	IR v. RW	t	IR v. LIN	t	Skew	Max Loss
Training window size $T = 12$											
	RW	0.149	1.037	0.149	1.035					−0.075	5.662
Trad	LIN	0.093	0.645	0.092	0.638	0.018	0.121			−0.949	7.725
	RFF	0.181	1.258	0.181	1.254	0.085	0.591	0.156	1.079	−0.296	5.370
Taylor	LIN	0.147	1.026	0.145	1.013	0.129	0.895			1.114	5.926
	RFF	0.207	1.448	0.206	1.434	0.149	1.036	0.180	1.255	−0.846	7.066
Training window size $T = 60$											
	RW	0.173	1.148	0.172	1.143					0.651	5.435
Trad	LIN	0.134	0.896	0.135	0.895	0.195	1.298			0.143	6.984
	RFF	−0.056	−0.376	−0.055	−0.366	0.028	0.185	−0.160	−1.065	−1.252	8.108
Taylor	LIN	0.151	1.010	0.151	1.012	0.155	1.027			−0.180	7.586
	RFF	0.108	0.721	0.104	0.698	0.311	2.063**	0.073	0.490	−0.689	6.343
Training window size $T = 120$											
	RW	0.156	0.977	0.159	0.994					1.053	5.241
Trad	LIN	0.272	1.707*	0.279	1.743*	0.299	1.873*			0.454	7.547
	RFF	0.017	0.104	0.007	0.046	0.116	0.723	−0.115	−0.720	−0.581	5.470
Taylor	LIN	0.138	0.870	0.136	0.859	0.137	0.859			1.138	5.681
	RFF	0.072	0.456	0.073	0.462	0.170	1.063	0.046	0.287	−0.333	5.470

This table reports one-month-ahead out-of-sample market-timing performance for equal-weighted (EW) cross-currency portfolios under traditional monetary (Trad) and Taylor-rule (Taylor) fundamentals. Both models include an intercept; the benchmark is the drifting random walk. This table is the with-intercept counterpart to Table 5. The EW portfolio averages positions across all available currencies at each date using an expanding-membership rule. Strategy returns are magnitude-proportional: $\hat{r}_{t|t-1} \times r_t$ for LIN and RFF, and $r_{t-1} \times r_t$ for the RW strategy. SR is the annualised Sharpe ratio; t is the t-statistic for mean portfolio returns. IR v. Mkt is the information ratio vs. the equal-weighted buy-and-hold return; IR v. RW is the information ratio relative to the random walk portfolio; IR v. LIN (RFF rows only) is the information ratio of Ridge–RFF relative to the linear portfolio. Skew and Max Loss are return skewness and maximum loss in standard deviation units, respectively.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-sided).

retains a directional edge over the drifting random walk once the benchmark itself is properly adjusted for drift.

Traditional monetary fundamentals are less affected by the intercept switch. Trad LIN retains marginal significance at $T = 120$ (SR $= 0.272$, $t = 1.707^{*}$), and the remaining panels are broadly consistent between specifications, reflecting the weaker unconditional drift content of traditional monetary predictors relative to the Taylor-rule set.

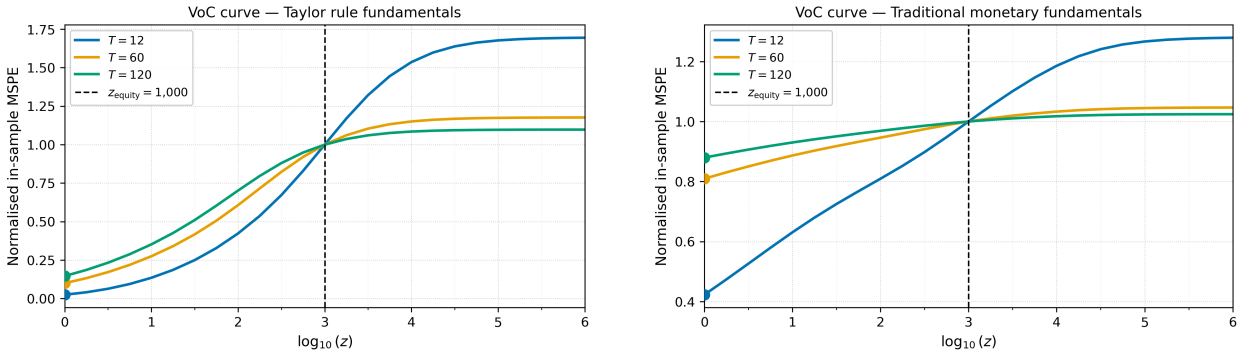
5 Diagnosing the Mirage: The Virtue of Complexity across the Regularisation Spectrum

The evidence in the preceding section establishes, at a fixed regularisation level $z = z_{\text{equity}} =$ 1,000, that Ridge–RFF delivers no systematic out-of-sample improvement over the random walk for either predictor set or training window. A natural response is that the KMZ equity calibration may simply be poorly suited to FX: the optimal z in currency markets could differ substantially from its equity counterpart, and the fixed-z comparison may therefore understate the method's potential. This section addresses that concern directly by tracing performance across the full regularisation spectrum $z \in [1, 10^6]$. We examine, in turn, how the in-sample and out-of-sample VoC curves relate across this spectrum; how Ridge–RFF fares relative to both OLS and the random walk at each regularisation level; what the effective complexity of the estimated model implies about the nature of the regularisation in use; and what the predictive signal strength embedded in the data implies about the theoretically optimal regularisation. We close with a robustness check against the zero-intercept restriction and a discussion of economic performance. Taken together, these results support a structural interpretation: complexity estimation fails in FX not because the wrong z is used, but because the predictive signal per RFF feature is statistically indistinguishable from zero.

5.1 In-Sample versus Out-of-Sample Virtue of Complexity

We begin by asking whether the failure at z_{equity} is a calibration artefact. Figure 1 plots the normalised in-sample MSPE as a function of $\log_{10}(z)$ for each predictor set and training window, anchored so that all curves equal unity at z_{equity} (the vertical dashed line). A value below unity means the model achieves lower in-sample error than at the equity baseline; a value above unity means higher error. In-sample, complexity performs exactly as theory predicts: smaller z (lower regularisation, more parameters effectively used) yields lower in-sample error for all T and both predictor sets, with the effect most pronounced at $T = 12$.

For Taylor rule fundamentals, the short-window model achieves a normalised in-sample MSPE near zero as $z \to 1$, a textbook illustration of overfitting.



(a) Taylor rule fundamentals

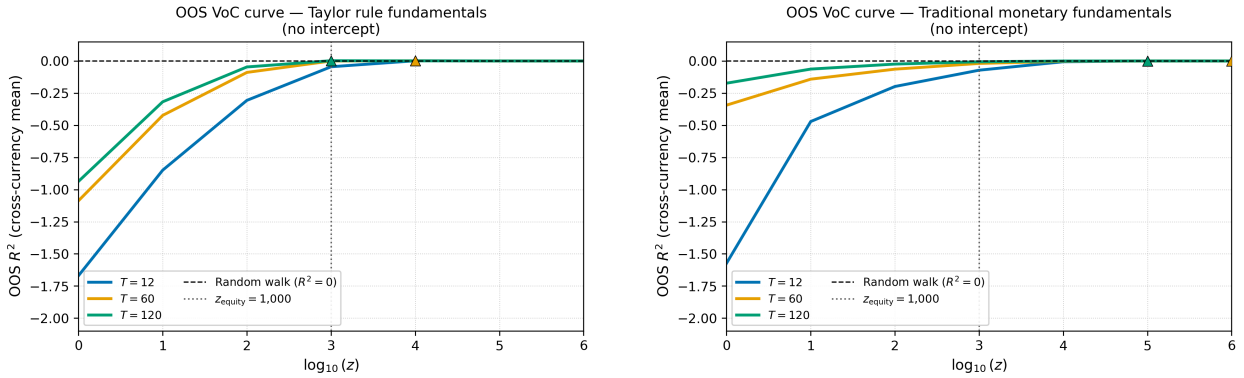
(b) Traditional monetary fundamentals

Figure 1: In-sample VoC curves. Each panel plots the normalised in-sample MSPE against $\log_{10}(z)$ for $T \in \{12, 60, 120\}$ months. Curves are normalised so that all training windows take the value 1.0 at $z = z_{\text{equity}} = 1{,}000$ (vertical dashed line). Values below unity indicate lower in-sample error relative to the equity baseline.

The out-of-sample picture is categorically different. Figure 2 plots the OOS R^2 (cross-currency mean) against $\log_{10}(z)$ for the same configurations, now benchmarked against the driftless random walk ($R^2 = 0$, horizontal dashed line). For both predictor sets, the OOS R^2 is uniformly negative across the entire grid $z \in [1, 10^6]$: no value of z yields a model that outperforms the random walk in expectation. The triangles mark the z value that maximises the OOS R^2 for each T – the empirical optimum. For Taylor rule fundamentals, the optimal z for $T = 60$ falls near z_{equity} ($\log_{10} z \approx 3$), while for $T = 120$ it is shifted somewhat higher; in both cases the maximum OOS R^2 is approximately zero. For traditional monetary fundamentals, the optimal z is pushed considerably further right ($\log_{10} z \approx 5$–6), reflecting more aggressive shrinkage requirements, yet even at this extreme the performance ceiling remains the random walk. The OOS VoC curve in FX is therefore bounded above by zero: no level of regularisation converts RFF into a statistically useful predictor when benchmarked against the driftless random walk.

5.2 Benchmarking Ridge-RFF against OLS and the Random Walk

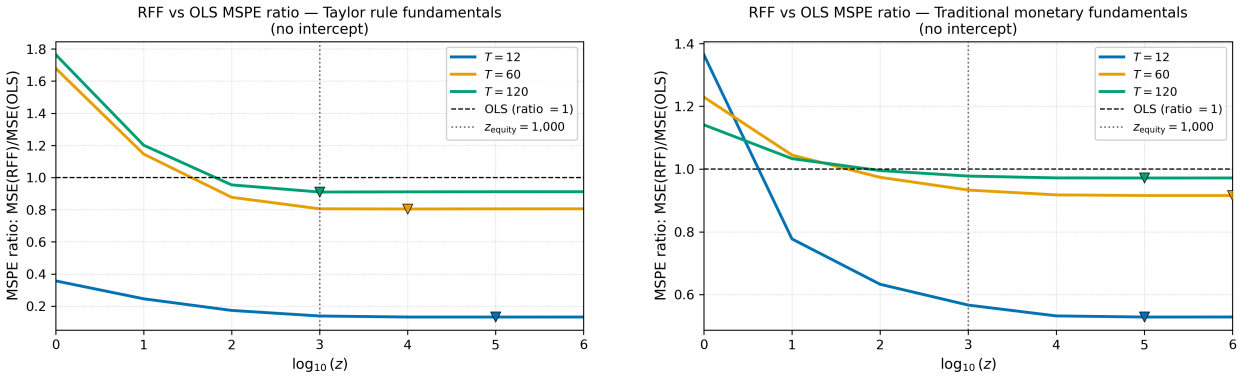
The OOS VoC curve benchmarks RFF against the random walk, but the VoC literature originated in equity markets where the natural benchmark is OLS. Separating these two comparisons reveals a nuanced picture: RFF does reduce out-of-sample MSPE relative to OLS, but this OLS-relative gain is insufficient to close the gap to the random walk.



(a) Taylor rule fundamentals (b) Traditional monetary fundamentals

Figure 2: OOS VoC curves (no-intercept baseline). Each panel plots the cross-currency mean OOS R^2 against $\log_{10}(z)$ for $T \in \{12, 60, 120\}$. The horizontal dashed line marks the random walk benchmark ($R^2 = 0$); the vertical dotted line marks the KMZ equity calibration ($z_{\text{equity}} = 1{,}000$). Triangles indicate the z value that maximises OOS R^2 for each T. The OOS curve lies below zero across the entire regularisation spectrum for both predictor sets.

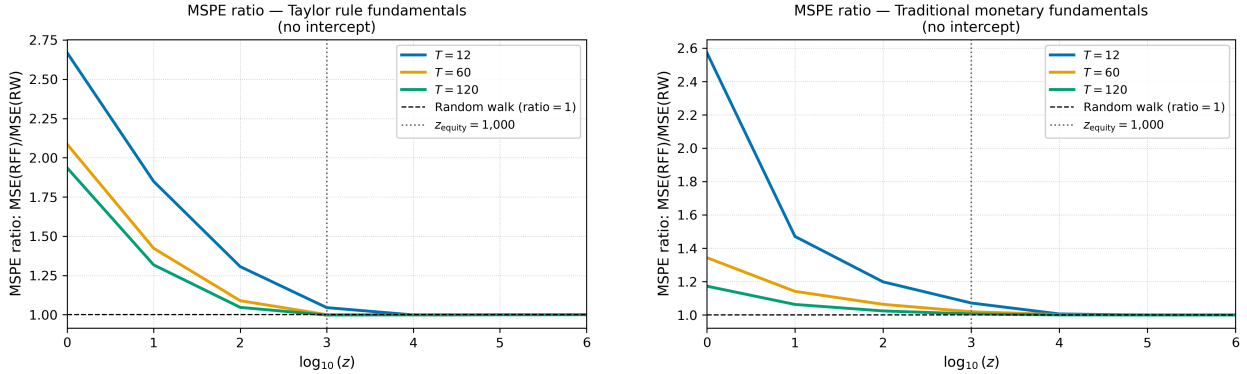
Figure 3 plots the ratio MSE(RFF)/MSE(OLS) across the z grid. For Taylor rule fundamentals, the ratio falls monotonically from well above 1 (at $z = 1$, RFF overfits severely relative to OLS) to well below 1 at large z, with $T = 60$ and $T = 120$ achieving ratios around 0.81 and 0.91 respectively at z_{equity}. The short window $T = 12$ exhibits the sharpest variance reduction, reaching a ratio of approximately 0.15 at the optimum. For traditional monetary fundamentals, the improvement relative to OLS is more modest at longer windows ($T = 60$: ratio ≈ 0.93; $T = 120$: ≈ 0.96 at z_{equity}), reflecting the lower dimensionality of the traditional predictor set and its correspondingly milder overfitting problem under OLS. In both cases, Ridge regularisation meaningfully corrects for the bias–variance trade-off of plain OLS.



(a) Taylor rule fundamentals (b) Traditional monetary fundamentals

Figure 3: MSPE ratio of Ridge–RFF relative to OLS (no-intercept baseline). A ratio below 1 indicates that Ridge–RFF achieves lower out-of-sample MSPE than OLS. Triangles mark the z value that minimises the MSPE ratio for each T. The horizontal dashed line marks parity with OLS (ratio = 1).

Against the random walk, however, the picture reverses entirely. Figure 4 plots the ratio MSE(RFF)/MSE(RW). All curves remain above 1 across the full z grid for both predictor sets, converging to 1 only asymptotically as $z \to \infty$ (when RFF collapses to the driftless random walk itself). For Taylor rule fundamentals, the ratio at z_{equity} lies between approximately 1.04 ($T = 120$) and 1.06 ($T = 60$), indicating that the best achievable statistical performance of Ridge–RFF is marginally worse than the random walk even after optimising z. Traditional monetary fundamentals perform yet more poorly against the random walk at shorter windows, with ratios at z_{equity} reaching 1.07–1.35 depending on T, though again converging to 1 at the far right of the z grid. The central message is clear: the OLS-relative advantage of Ridge–RFF, while real, is too small to overcome the fundamental unpredictability of FX returns relative to the random walk – a finding that resonates with Meese and Rogoff (1983) and extends it to the nonlinear estimation setting.



(a) Taylor rule fundamentals

(b) Traditional monetary fundamentals

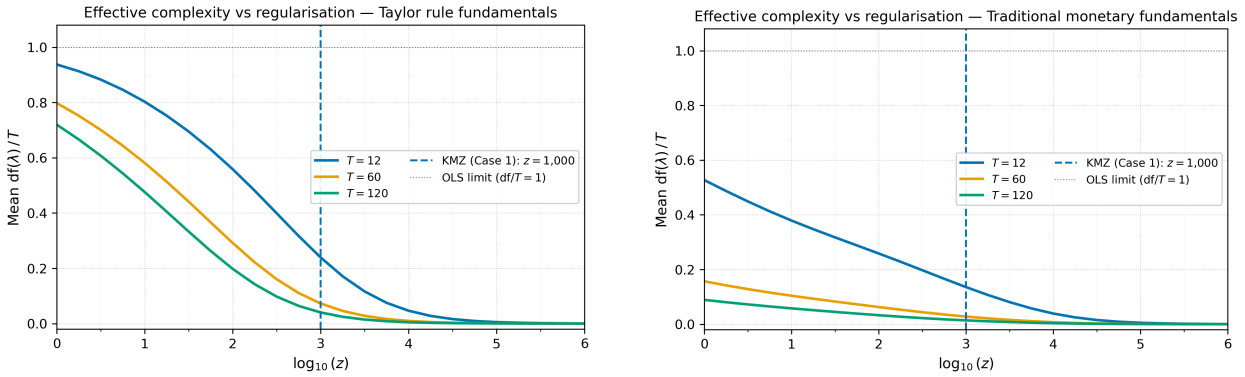
Figure 4: MSPE ratio of Ridge–RFF relative to the driftless random walk (no-intercept baseline). A ratio below 1 would indicate that Ridge–RFF achieves lower out-of-sample MSPE than the random walk. No curve crosses below 1 for either predictor set. Layout and symbols as in Figure 3.

5.3 Effective Model Complexity along the Regularisation Path

The regularisation parameter z controls the nominal severity of shrinkage, but the operationally relevant quantity is the *effective* complexity $\mathrm{df}(\lambda)/T$, which measures the fraction of the training sample length consumed by the model's effective degrees of freedom. Following Kelly et al. (2024), we consider two regularisation parameterisations. *Case 1* fixes the complexity ratio $z = \lambda/T = 1{,}000$, so that the Ridge penalty $\lambda = z \cdot T$ scales proportionally with the training window; this is the KMZ equity baseline and the primary specification used throughout the paper. *Case 2* instead fixes the Ridge penalty directly at $\lambda = 1{,}000$, so that $z = \lambda/T$ declines as T grows; this parameterisation, discussed in detail in the Appendix,

implies progressively weaker effective regularisation at longer training windows and provides a useful contrast with Case 1.

Figure 5 plots the cross-currency mean $\mathrm{df}(\lambda)/T$ as a function of $\log_{10}(z)$ across the full regularisation grid. For Taylor rule fundamentals, the vertical dashed line marks the Case 1 value $z_{\text{equity}} = 1{,}000$. At this calibration, effective complexity ranges from $\mathrm{df}/T \approx 0.24$ at $T = 12$ to $\mathrm{df}/T \approx 0.04$ at $T = 120$, confirming that the KMZ specification already operates close to the random walk limit in effective parameter space, particularly at longer training windows. The traditional monetary predictor set exhibits even lower effective complexity under Case 1, with df/T not exceeding 0.14 at $T = 12$.



(a) Taylor rule fundamentals

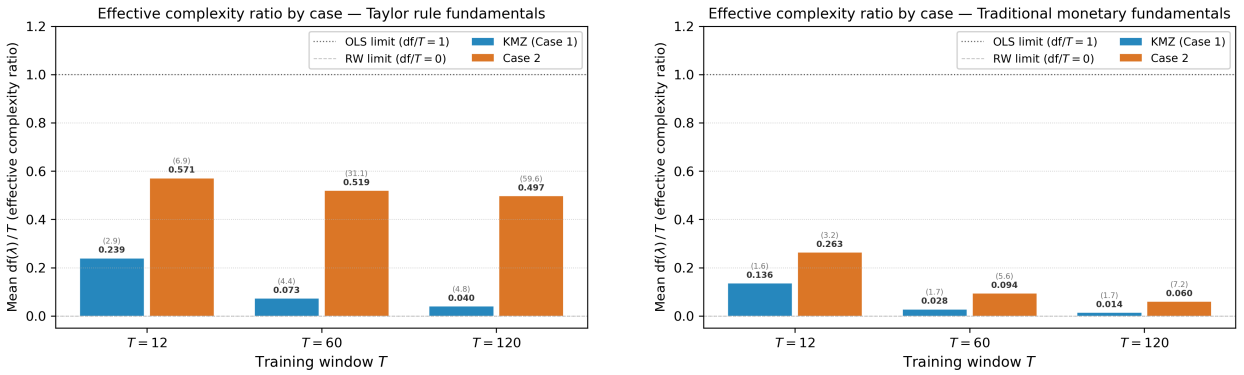
(b) Traditional monetary fundamentals

Figure 5: Effective complexity ratio $\mathrm{df}(\lambda)/T$ as a function of $\log_{10}(z)$. Lines report the cross-currency mean for each training window $T \in \{12, 60, 120\}$. The vertical dashed line marks the Case 1 KMZ calibration $z_{\text{equity}} = 1{,}000$. The horizontal dotted line marks the OLS limit ($\mathrm{df}/T = 1$).

Figure 6 compares the average effective complexity under the two cases across training windows. Under Case 1, df/T falls sharply as T increases for both predictor sets: for Taylor rule fundamentals it drops from 0.239 at $T = 12$ to 0.040 at $T = 120$, while for traditional monetary fundamentals the corresponding decline is from 0.136 to 0.014. This pattern arises mechanically because fixing $z = 1{,}000$ while T grows implies a proportionally larger penalty λ, which progressively suppresses effective complexity toward the random walk limit.

Under Case 2, the picture reverses: because $\lambda = 1{,}000$ is fixed, $z = \lambda/T$ shrinks as T grows, so the effective penalty declines in relative terms and the model retains substantially more effective complexity at longer windows. For Taylor rule fundamentals, Case 2 delivers $\mathrm{df}/T \approx 0.571$ (approximately 6.9 effective df compared with $df = T = 12$ under OLS) at $T = 12$, rising to ≈ 0.519 (31.1 effective df compared with $df = T = 60$ under OLS) at $T = 60$ and ≈ 0.497 (59.6 effective df compared with $df = T = 120$ under unpanelised RFF Ridge estimation) at $T = 120$. Traditional monetary fundamentals show a qualitatively similar

pattern under Case 2, though at markedly lower levels: df/T ranges from 0.263 (3.2 effective parameters) at $T = 12$ to 0.060 (7.2 effective parameters) at $T = 120$.



(a) Taylor rule fundamentals

(b) Traditional monetary fundamentals

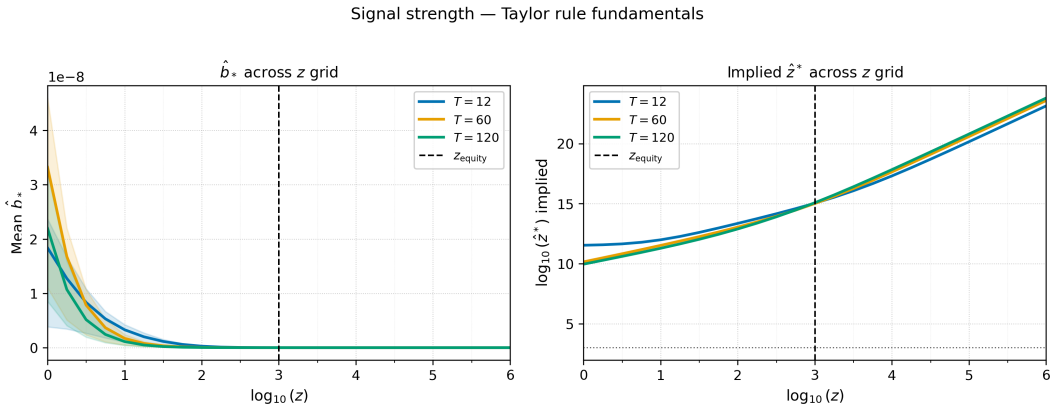
Figure 6: Average effective complexity ratio $\overline{\mathrm{df}(\lambda)/T}$ by training window and regularisation case. Blue bars correspond to Case 1 (KMZ, $z = z_{\text{equity}} = 1{,}000$ fixed); orange bars to Case 2 ($\lambda = 1{,}000$ fixed, $z = \lambda/T$). Numbers above each bar report $\overline{\mathrm{df}(\lambda)/T}$ (bold) and the implied effective degrees of freedom $\overline{\mathrm{df}(\lambda)}$ (in parentheses). The horizontal dotted lines mark the OLS limit ($\mathrm{df}/T = 1$) and the random walk limit ($\mathrm{df}/T = 0$).

The contrast between the two cases is instructive: Case 2 allows considerably more model flexibility at long training windows, yet the OOS VoC curve in Figure 2 shows that this added flexibility does not translate into positive OOS R^2 at any training window. Even with substantially higher effective complexity – up to approximately 60 effective parameters under Taylor rule fundamentals at $T = 120$ – the performance ceiling remains the random walk. This rules out the possibility that the failure of Ridge–RFF is a consequence of excessive regularisation suppressing legitimate predictive structure. The problem, as we demonstrate in Section 5.4, lies in the signal itself. The time-series evolution of $\mathrm{df}(\lambda)/T$ under both cases is reported in Appendix Figure F1, showing that effective complexity under Case 2 has trended downward over the sample for both predictor sets, broadly consistent with the documented secular decline in FX fundamental predictability (Rossi, 2013).

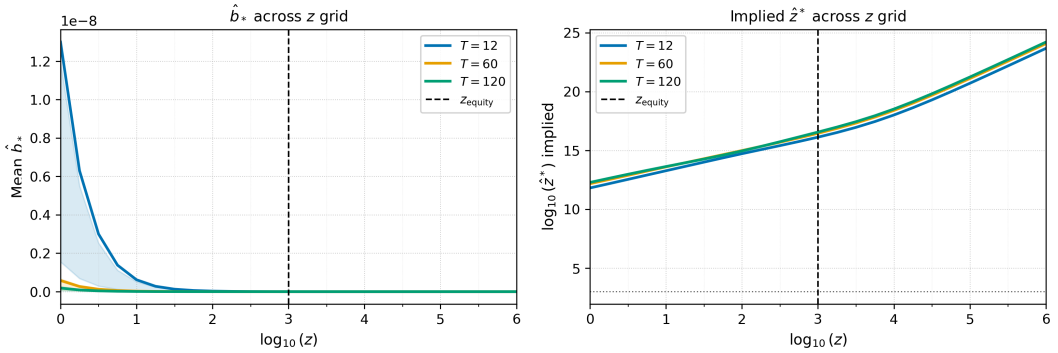
5.4 Signal Strength and the Implied Optimal Regularisation

The preceding results establish that the failure of Ridge–RFF persists across the entire regularisation spectrum and cannot be remedied by recalibrating z. This subsection provides a structural explanation within the KMZ framework itself. The key theoretical object is the average predictive signal strength per RFF feature, denoted b_*, which Kelly et al. (2024) show determines the theoretically optimal regularisation: when $b_* \approx 0$, the optimal $z^* \to \infty$, and the optimal forecast collapses to the random walk.

We estimate b_* at each rolling window using using equation (13) and recover the implied optimal regularisation as $\hat{z}^* = P/(\hat{b}_* T)$. Figure 7 presents both objects across the z grid. The left panels show that $\hat{b}_*$ is of order 10^{-8} at $z \approx 1$ for both predictor sets, and collapses monotonically toward zero as z increases. The cross-currency confidence bands are narrow throughout, indicating that the near-zero signal strength is a systematic feature of the FX data rather than noise in any particular currency. The right panels plot the implied $\hat{z}^*$ on a $\log_{10}$ scale: it grows steeply with z, already exceeding $\log_{10}(\hat{z}^*) \approx 11$ at z_{equity} for both predictor sets and all training windows.



(a) Taylor rule fundamentals



(b) Traditional monetary fundamentals

Figure 7: Signal strength and implied optimal regularisation across the z grid. *Left panels:* time-series mean of $\hat{b}_*$ (see equation (13)) with cross-currency confidence bands. *Right panels:* implied $\log_{10}(\hat{z}^*)$ across the z grid. The vertical dashed lines mark $z_{\text{equity}} = 1{,}000$. The horizontal dotted line in the right panels marks $\log_{10}(z_{\text{equity}}) = 3$ for reference.

Tables 7 and 8 report the time-series mean $\bar{z}^*$ on a $\log_{10}$ scale, together with the log-ratio $\log_{10}(\bar{z}^*/z_{\text{equity}})$, for each currency and training window. For Taylor rule fundamentals, $\log_{10}(\bar{z}^*)$ ranges from approximately 14.72 (NOK, $T = 60$) to 15.41 (CAD, $T = 120$), with a cross-currency mean of ≈ 15.05 at $T = 12$, stable across training windows. The implied

log-ratio is consistently around 12, meaning that the data-implied optimal regularisation is approximately 10^{12} times larger than the equity baseline – roughly one trillion times the calibration used in equity markets. For traditional monetary fundamentals, the numbers are even more extreme: the cross-currency mean $\log_{10}(\bar{z}^*)$ rises from 16.08 at $T = 12$ to 16.56 at $T = 120$, with log-ratios between 13.1 and 13.6, implying an optimal z some 10^{13}-fold larger than z_{equity}. The monotone increase in $\bar{z}^*$ with T for the traditional predictor set – absent for Taylor rule fundamentals – is consistent with the traditional fundamentals' weaker and more rapidly deteriorating signal strength at longer estimation horizons.

Table 7: Implied optimal regularisation $\hat{z}^*$ — Taylor rule fundamentals. Each cell reports the time-series mean of $\log_{10}(\hat{z}^*)$ and the implied log-ratio $\log_{10}(\hat{z}^*/z_{\text{equity}})$, where $z_{\text{equity}} = 1{,}000$ is the KMZ equity baseline (Kelly et al., 2024, Table I). Ratios exceeding zero indicate that FX requires stronger regularisation than equity markets, reflecting weaker predictive signal per RFF feature.

	$T = 12$		$T = 60$		$T = 120$	
Currency	$\log_{10}(\bar{z}^*)$	$\log_{10}(\text{ratio})$	$\log_{10}(\bar{z}^*)$	$\log_{10}(\text{ratio})$	$\log_{10}(\bar{z}^*)$	$\log_{10}(\text{ratio})$
AUD	14.92	11.92	14.82	11.82	14.84	11.84
CAD	15.29	12.29	15.28	12.28	15.41	12.41
CHF	14.98	11.98	15.04	12.04	15.07	12.07
EUR	15.26	12.26	15.17	12.17	15.18	12.18
GBP	15.11	12.11	15.14	12.14	15.17	12.17
JPY	14.94	11.94	14.88	11.88	14.92	11.92
NOK	14.76	11.76	14.72	11.72	14.87	11.87
SEK	14.89	11.89	14.78	11.78	14.86	11.86
Mean	*15.05*	*12.05*	*15.02*	*12.02*	*15.09*	*12.09*

Notes: $\hat{z}^*$ is the Stein-corrected implied optimal regularisation parameter recovered from the Ridge solution at each rolling window: $\hat{z}^* = P/(\hat{b}_* T)$, where $\hat{b}_* = P^{-1}\|\hat{\beta}\|^2 \cdot \text{df}/(T - \text{df} - 1)$ and df denotes the effective degrees of freedom. $\bar{z}^*$ denotes the time-series mean of $\hat{z}^*$ across rolling windows. $\log_{10}(\text{ratio}) = \log_{10}(\bar{z}^*) - \log_{10}(z_{\text{equity}}) = \log_{10}(\bar{z}^*) - 3$. All estimates are computed under the no-intercept baseline specification.

These estimates jointly confirm the structural interpretation: the OOS VoC curve is bounded above by zero not because the wrong z is chosen, but because the data offer essentially no predictive signal per RFF feature. The KMZ framework does not fail in FX because it is a poor framework; it fails because the necessary condition for the VoC – a non-trivial b_* – is absent in FX monthly fundamental data. This is a considerably stronger conclusion than a negative replication result: it uses the theoretical machinery of Kelly et al. (2024) to explain its own failure in the FX setting.

Table 8: Implied optimal regularisation $\hat{z}^*$ — Traditional monetary fundamentals. Each cell reports the time-series mean of $\log_{10}(\hat{z}^*)$ and the implied log-ratio $\log_{10}(\hat{z}^*/z_{\text{equity}})$, where $z_{\text{equity}} = 1{,}000$ is the KMZ equity baseline (Kelly et al., 2024, Table I). Ratios exceeding zero indicate that FX requires stronger regularisation than equity markets, reflecting weaker predictive signal per RFF feature.

	$T = 12$		$T = 60$		$T = 120$	
Currency	$\log_{10}(\bar{z}^*)$	$\log_{10}(\text{ratio})$	$\log_{10}(\bar{z}^*)$	$\log_{10}(\text{ratio})$	$\log_{10}(\bar{z}^*)$	$\log_{10}(\text{ratio})$
AUD	15.91	12.91	16.11	13.11	16.23	13.23
CAD	16.44	13.44	16.61	13.61	16.75	13.75
CHF	16.07	13.07	16.33	13.33	16.31	13.31
EUR	16.12	13.12	16.52	13.52	16.90	13.90
GBP	16.18	13.18	16.71	13.71	16.54	13.54
JPY	16.07	13.07	16.42	13.42	16.62	13.62
NOK	15.42	12.42	16.00	13.00	16.12	13.12
SEK	15.73	12.73	15.99	12.99	16.41	13.41
Mean	*16.08*	*13.08*	*16.41*	*13.41*	*16.56*	*13.56*

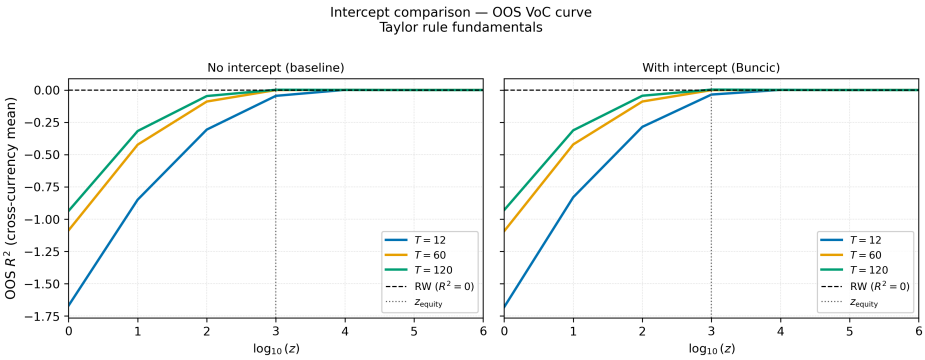
Notes: See notes to Table 7. All estimates are computed under the no-intercept baseline specification. The monotone increase in $\log_{10}(\bar{z}^*)$ with T reflects the diminishing signal-to-noise ratio of traditional monetary fundamentals at longer estimation horizons.

5.5 Intercept Robustness and Economic Performance

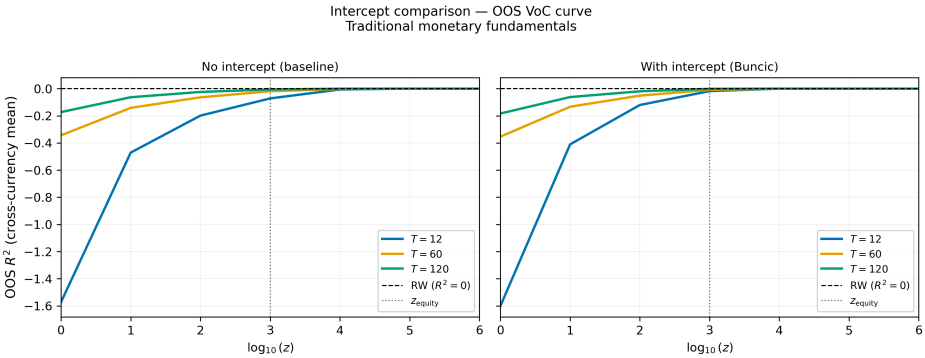
Buncic (2025) demonstrates that in equity markets, the VoC curve is sensitive to whether an intercept is included in the Ridge regression: adding an intercept substantially flattens the curve and qualitatively weakens the VoC finding for longer training windows. Figure 8 examines whether the FX results are similarly affected. Each panel presents the OOS VoC curve under the no-intercept baseline (left) and the with-intercept specification (right) for both predictor sets. The two panels are virtually indistinguishable: the OOS R^2 remains uniformly negative across the full z grid, the location of the optimal z is unchanged, and the performance ceiling continues to approximate the random walk. The Buncic intercept critique, which has material force in equity markets, is inconsequential in FX. This is consistent with our structural interpretation: since $\hat{b}_* \approx 0$ irrespective of the intercept choice, and since the random walk in FX is driftless by construction, re-centring the regression does not alter the fundamental signal scarcity.

5.6 Economic Performance across the Complexity Spectrum

The absence of positive OOS R^2 across the regularisation spectrum need not preclude positive economic performance: a market-timing strategy benefits from directional accuracy, which



(a) Taylor rule fundamentals



(b) Traditional monetary fundamentals

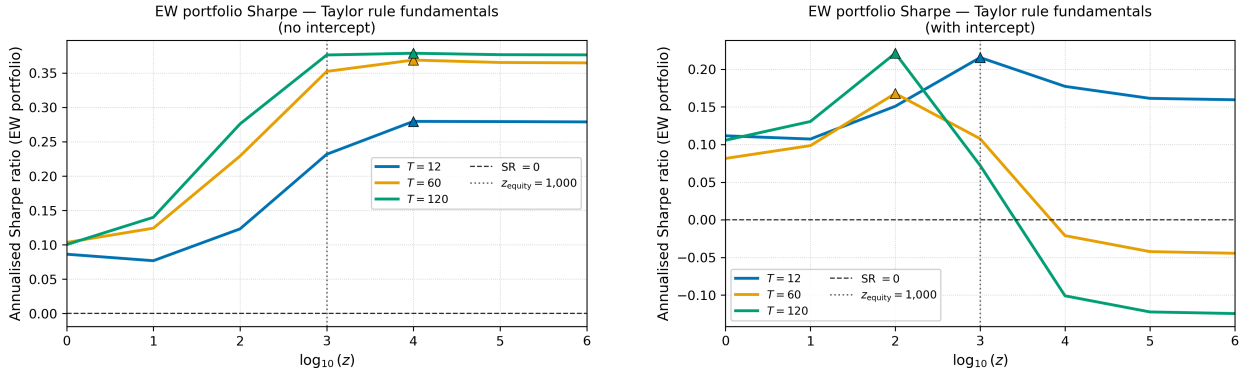
Figure 8: Intercept comparison of OOS VoC curves. Each panel shows the no-intercept baseline (left sub-panel) and the with-intercept specification (following Buncic, 2025) (right sub-panel). The horizontal dashed line marks the random walk ($R^2 = 0$); the vertical dotted line marks z_{equity}. The OOS curve is bounded above by zero under both specifications, confirming that the intercept choice is inconsequential for the qualitative conclusions in FX.

can arise even when the forecast explains little of the variance of returns (Anatolyev and Gerko, 2005; Blaskowitz and Herwartz, 2011). Figures 9 and 10 plot the annualised Sharpe ratio of the equal-weighted cross-currency portfolio as a function of $\log_{10}(z)$ for Taylor-rule and traditional monetary fundamentals respectively, presenting both the no-intercept baseline (left panel) and the with-intercept specification (following Buncic, 2025) (right panel) in each figure. The side-by-side comparison isolates the extent to which positive Sharpe ratios reflect genuine directional predictability versus a drift-capture artefact of the zero-intercept restriction.

Taylor-rule fundamentals. Under the no-intercept baseline (Figure 9, left), all three training windows generate positive Sharpe ratios across the *entire* z grid, and the profiles are monotonically increasing: greater regularisation uniformly improves market-timing performance up to an optimum near $\log_{10}(z) \approx 4$, beyond which the curves flatten. At the KMZ equity baseline $z_{\text{equity}} = 1{,}000$, the EW Sharpe ratios are approximately 0.23 ($T = 12$), 0.35 ($T = 60$), and 0.37 ($T = 120$); at their respective optima ($\log_{10}(z) \approx 4$) they reach approximately 0.28, 0.37, and 0.38. The monotonically increasing pattern reflects a feature of the magnitude-proportional timing strategy: in FX, where OOS R^2 is negligible, heavier shrinkage produces smaller positions that aggregate more cleanly in the equal-weighted portfolio, improving Sharpe ratios even as statistical accuracy remains unchanged.

The with-intercept panel (Figure 9, right) reveals that a substantial portion of this performance is intercept-specific for longer training windows. At $T = 12$, the profile is broadly unaffected and becomes slightly steeper: the Sharpe ratio rises steadily to approximately 0.21 at z_{equity} and remains in the range 0.17–0.21 throughout the grid, confirming that the short-window result does not depend on absorbing unconditional drift into the forecast weights. At $T = 60$, the pattern changes materially: the curve peaks at approximately 0.17 near $\log_{10}(z) \approx 2$, declines to approximately 0.11 at z_{equity}, and falls below zero beyond $\log_{10}(z) \approx 4$–5. The deterioration is even more dramatic for $T = 120$: despite peaking at approximately 0.22 near $\log_{10}(z) \approx 2$ and remaining near 0.21 at z_{equity}, the Sharpe ratio collapses to approximately -0.10 by $\log_{10}(z) = 4$ and remains negative thereafter. This sharp reversal immediately past z_{equity} reveals a fundamental instability: the positive no-intercept performance at longer windows is concentrated in a region of the z grid that is particularly sensitive to drift capture. Once the intercept is included and the unconditional mean return is no longer absorbed into the forecast weights, the residual directional accuracy at $T = 60$ and $T = 120$ is substantially weaker, and more heavily regularised (higher z) models — whose positions are smallest — have the least information to work with.

The contrast between predictor sets — positive and broadly increasing Sharpe ratios for Taylor rule versus a weakly positive or negative profile for traditional monetary fundamentals at longer windows — despite similar OOS R^2 profiles, likely reflects the interest-rate components of the Taylor-rule specification, which are more directly related to well-documented FX risk premia such as carry (Lustig and Verdelhan, 2007; Burnside et al., 2011).



(a) No intercept (baseline)

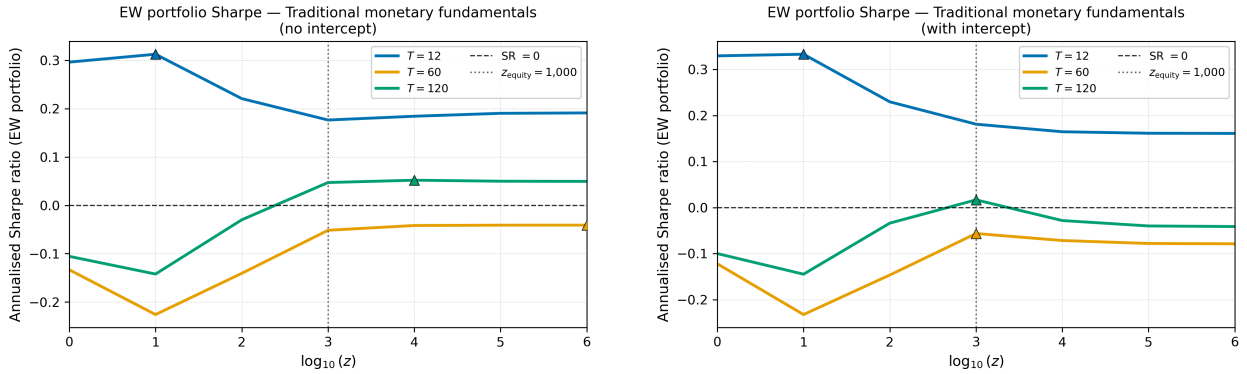
(b) With intercept (Buncic, 2025)

Figure 9: Annualised Sharpe ratio of the equal-weighted cross-currency portfolio as a function of $\log_{10}(z)$, Taylor-rule fundamentals. Lines report $T \in \{12, 60, 120\}$ months. Triangles mark the z value that maximises the Sharpe ratio for each T. The horizontal dashed line marks SR $= 0$; the vertical dotted line marks $z_{\text{equity}} = 1{,}000$. Left panel: no-intercept baseline. Right panel: with-intercept specification (Buncic, 2025), in which the drift random walk serves as the benchmark. The contrast between panels isolates the role of drift capture in generating positive Sharpe ratios at longer training windows.

Traditional monetary fundamentals. The traditional monetary predictor set tells a more straightforward story (Figure 10). Under the no-intercept baseline (left panel), only $T = 12$ generates a consistently positive Sharpe ratio profile, peaking at approximately 0.31 near $\log_{10}(z) \approx 1$ and declining to approximately 0.18 at z_{equity}, before flattening at roughly 0.19 for higher z. $T = 60$ is negative throughout the grid, with a trough of approximately -0.23 near $\log_{10}(z) \approx 1$ that recovers toward -0.06 at z_{equity}. $T = 120$ is negative at low z, rising to a marginal positive of approximately 0.05 near z_{equity} and beyond.

The with-intercept specification (right panel) is remarkably similar to the baseline across all three training windows. $T = 12$ retains a positive Sharpe ratio across the full z grid, peaking at approximately 0.34 near $\log_{10}(z) \approx 1$, declining to approximately 0.18 at z_{equity}, and remaining flat near 0.16–0.17 thereafter. $T = 60$ and $T = 120$ remain near zero or negative throughout, with $T = 60$ reaching a maximum of approximately -0.07 at z_{equity} and $T = 120$ achieving only a marginal positive value of approximately 0.01 in the vicinity of z_{equity}. The near-absence of any intercept effect for traditional monetary fundamentals contrasts strikingly with the large intercept sensitivity documented for Taylor-rule fundamentals at longer windows. This

contrast is consistent with the weaker and more stable unconditional drift of the traditional predictor set: with less drift to capture, the zero-intercept restriction introduces smaller position-scaling distortions, and the economic performance profile is structurally driven by the directional content of the predictors rather than by drift absorption.



(a) No intercept (baseline)

(b) With intercept (Buncic, 2025)

Figure 10: Annualised Sharpe ratio of the equal-weighted cross-currency portfolio as a function of $\log_{10}(z)$, traditional monetary fundamentals. Layout as in Figure 9. The near-identical profiles across intercept specifications confirm that the positive $T = 12$ Sharpe ratio is not a drift-capture artefact, and that the weak performance at longer windows is structural rather than specification-specific.

Synthesis. Together, Figures 9 and 10 reveal two distinct sources of positive economic performance in this setting. The first, operative only for Taylor-rule fundamentals at $T = 60$ and $T = 120$ under the no-intercept restriction, is drift capture: the zero-intercept constraint forces the model to absorb unconditional currency drift into its forecast weights, generating directional positions that reflect the currency's unconditional drift rather than its conditional signal. This component is mechanically eliminated by including an intercept, and the resulting collapse in Sharpe ratios at high z (where positions are smallest) confirms that it is not a manifestation of genuine learning. The second source, common to $T = 12$ under both predictor sets and both intercept specifications, is robust to this critique. The consistently positive short-window performance — reaching approximately 0.21–0.23 at z_{equity} for Taylor rule and approximately 0.18–0.23 for traditional monetary fundamentals across intercept specifications — is consistent with the mechanism identified by Nagel (2025): at short windows with $P \gg T$ and persistent predictors, Ridge–RFF forecasts mechanically approximate a volatility-timed momentum rule, generating positive Sharpe ratios that reflect the model's implicit volatility-timing properties rather than genuine extraction of nonlinear fundamental structure. Individual-currency Sharpe ratios, reported in Appendix Figures D1 and D2, confirm substantial cross-currency heterogeneity in the economic performance; the cross-currency aggregation is essential for generating meaningful positive mean Sharpe ratios,

particularly under traditional monetary fundamentals where individual-currency results are highly dispersed.

6 Conclusions

This paper investigates whether the virtue of complexity (VoC) extends from equity return prediction to exchange rate forecasting and — working within the Kelly et al. (2024) framework itself — diagnoses *why* it fails and *what signal conditions* its success would require. Holding two theoretically motivated predictor sets fixed while varying only the estimation method across eight major USD currency pairs, we cleanly isolate the contribution of model complexity.

At the fixed-regularisation baseline, Ridge–RFF consistently dominates OLS across training windows and predictor sets, most dramatically at the shortest window where nominal complexity peaks — consistent with the benign overfitting of Kelly et al. (2024). The more demanding benchmark tells a different story. Significant outperformance of the random walk is confined to AUD, CHF, and JPY under Taylor-rule fundamentals, and to the Japanese yen alone under traditional monetary fundamentals. The cross-currency mean out-of-sample R^2 remains negative at every training window and for both predictor sets: individual-currency gains are real but do not constitute the pervasive panel-level outperformance the VoC hypothesis requires. Under Taylor-rule fundamentals, the individual-currency gains persist as the training window lengthens and are invariant to whether an intercept is included, consistent with Molodtsova and Papell (2009). Under traditional monetary fundamentals, Ridge–RFF's advantage over OLS narrows as the window grows and no currency achieves reliable random walk outperformance. A robustness check under a fixed-λ specification (Appendix B) confirms that maintaining regularisation intensity is essential: when intensity declines as the window grows, Ridge–RFF's advantage over OLS collapses and reverses, isolating regularisation intensity rather than complexity as the operative mechanism.

The market-timing results reveal a sharp intercept sensitivity that is itself diagnostic. Under the no-intercept specification, Taylor-rule strategies deliver statistically significant equal-weighted portfolio returns at intermediate and long windows, with Ridge–RFF producing the strongest result — an annualised Sharpe ratio approaching 0.40 at the longest training window. Including an intercept causes these results to collapse at intermediate and long windows, exposing drift capture as their primary driver: the zero-intercept restriction had forced slope coefficients to absorb each currency's unconditional drift rather than condition on genuine predictive content. The only result robust to the intercept is OLS under traditional monetary fundamentals at the longest window (SR = 0.272), confirming that the Canadian dollar's directional predictability under this predictor set is not artefactual. At the shortest

window, positive portfolio returns are preserved regardless of intercept specification, consistent with volatility-timing rather than drift absorption.

Tracing the VoC curve across the full regularisation spectrum confirms the failure is structural rather than calibrational: the cross-currency mean out-of-sample R^2 is bounded above by zero everywhere, for both predictor sets and all training windows, ruling out a wrong choice of z as the explanation. The Kelly et al. (2024) framework explains its own failure: the average predictive signal per feature is indistinguishable from zero, and the implied optimal regularisation is 10^{12}–10^{13} times the equity baseline — quantifying both why no calibration of z generates panel-level outperformance and what signal concentration and alignment FX fundamental data would need to deliver for the VoC to succeed.

Three lessons emerge. *First, the failure of complexity in FX is structural at the panel level, not calibrational.* The cross-currency mean OOS VoC curve is bounded above by zero across the full regularisation spectrum and the signal-strength estimates imply an optimal regularisation many orders of magnitude beyond any feasible calibration. The KMZ framework explains its own failure through its own diagnostic tools — a considerably stronger conclusion than a simple negative replication result.

Second, the relationship between complexity and performance is predictor-set and currency-specific. Under Taylor-rule fundamentals, Ridge–RFF retains a meaningful advantage over OLS at every training window and beats the random walk for a stable subset of currencies; under traditional monetary fundamentals this advantage narrows and eventually vanishes. In the language of Kelly and Malamud (2025), the concentration-and-alignment condition is satisfied locally for specific currency–predictor combinations under Taylor-rule fundamentals but not at the panel level, and not at all for traditional monetary fundamentals.

Third, the statistical–economic dissociation reveals the mechanism. Statistically significant portfolio Sharpe ratios coexist with negative cross-currency mean OOS R^2 under the no-intercept specification, consistent with the volatility-timing and drift-capture mechanisms of Nagel (2025). Their collapse under the with-intercept specification confirms the diagnosis: the Buncic (2025) intercept critique is inconsequential for the statistical conclusions in FX but decisive for the economic ones. The sole exception — OLS on traditional monetary fundamentals at the longest window, robust to both specifications — is consistent with longer-horizon linear predictability documented by Meese and Rogoff (1983) and Rossi (2013), and is the only result that meets the standard a practitioner would require: genuine conditional directional accuracy that survives a properly adjusted benchmark.

Our design holds the predictor set fixed to isolate complexity, bounding the scope of the conclusions. Richer predictor sets — combining standard fundamentals with financial conditions indices, risk premia proxies, or high-frequency indicators — may provide the signal

concentration and alignment needed to unlock the VoC for a broader currency cross-section; the signal-strength diagnostics introduced here offer a principled tool for evaluating any such candidate set. A second direction is time variation: conditions for VoC success may have been satisfied episodically in FX even if not robust over the full sample, motivating regime-aware or adaptively regularised estimators that respond to time-varying signal strength.

References

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

Anatolyev, S. and Gerko, A. (2005). A trading approach to testing for predictability. *Journal of Business & Economic Statistics*, 23(4):455–461.

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

Beckmann, J., Kerkemeier, M., and Kruse-Becher, R. (2025). Regime-specific exchange rate predictability. *Journal of Economic Dynamics and Control*, 176:105095.

Blaskowitz, O. and Herwartz, H. (2011). On economic evaluation of directional forecasts. *International Journal of Forecasting*, 27(4):1058–1065.

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

Burnside, C., Eichenbaum, M., Kleshchelski, I., and Rebelo, S. (2011). Do peso problems explain the returns to the carry trade? *Review of Financial Studies*, 24(3):853–891.

Cartea, A., Jin, Q., and Shi, Y. (2025). The limited virtue of complexity in a noisy world. Working Paper.

Cartea, A. and Shi, Y. (2025). The mechanical virtue of random fourier features and model complexity. 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.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263.

Diebold, F. X. and Nason, J. A. (1990). Nonparametric exchange rate prediction? *Journal of International Economics*, 28(3–4):315–332.

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.

Fallahgoul, H. (2026). High-dimensional learning in finance. Working Paper.

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.

Frömmel, M., MacDonald, R., and Menkhoff, L. (2005). Markov switching regimes in a monetary exchange rate model. *Economic Modelling*, 22(3):485–502.

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

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Hastie, T., Montanari, A., Rosset, S., and Tibshirani, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *The Annals of Statistics*, 50(2):949–986.

Hauzenberger, N. and Huber, F. (2020). Model instability in predictive exchange rate regressions. *Journal of Forecasting*, 39(2):168–186.

Huber, F. (2017). Structural breaks in taylor rule based exchange rate models: Evidence from threshold time varying parameter models. *Economics Letters*, 150:48–52.

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. SSRN Working Paper.

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.

Lustig, H. and Verdelhan, A. (2007). The cross section of foreign currency risk premia and consumption growth risk. *American Economic Review*, 97(1):89–117.

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, J. F. (2022). Exchange rate forecasting with advanced machine learning methods. *Journal of Risk and Financial Management*, 15(1):2.

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.

Stillwagon, J. and Sullivan, P. (2020). Markov switching in exchange rate models: Will more regimes help? *Empirical Economics*, 59(1):413–436.

Wang, R., Morley, B., and Stamatogiannis, M. P. (2019). Forecasting the exchange rate using nonlinear taylor rule based models. *International Journal of Forecasting*, 35(2):429–442.

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.

B Fixed-λ Specification: Robustness Results

Tables B1 and B2 report out-of-sample forecast accuracy under the fixed-λ specification ($\lambda = 1{,}000$ constant) for Taylor-rule and traditional monetary fundamentals, respectively. This specification corresponds to the conventional Ridge implementation in which the penalty parameter is held fixed regardless of sample size. As established in equation (11) of Section 3.4, fixing λ implies a KMZ parameter $z = \lambda/T$ that declines from 83.3 at $T = 12$ to 16.7 at $T = 60$ and 8.3 at $T = 120$, with effective complexity $\mathrm{df}/T \approx 0.86$ throughout — effectively near-OLS at every window. The comparison with the KMZ results in the main text (Tables 1 and 2) isolates the contribution of maintaining adequate regularisation from the contribution of complexity per se. Crucially, the two specifications differ at every training window, including $T = 12$: the KMZ case applies $\lambda = 12{,}000$ ($z = 1{,}000$), while the fixed-λ case applies $\lambda = 1{,}000$ ($z = 83.3$) — a twelve-fold difference in regularisation intensity.

B.1 Taylor-rule fundamentals

The most striking feature of Table B1 is the sharp deterioration in Ridge–RFF performance relative to both OLS and the random walk as the training window grows, in stark contrast to the monotonic improvement visible in the KMZ results of Table 1.

At $T = 12$ (Panel A), both specifications apply the most regularisation of any window, and both show Ridge–RFF dramatically outperforming OLS. Seven of eight DM statistics are positive and significant at the 1% level (the exception being JPY, where DM = 1.266), confirming that Ridge regularisation prevents the variance inflation that makes OLS unusable at extreme nominal complexity ($c = 1{,}000$). Compared with the KMZ results, however, the lower regularisation of the fixed-λ case ($\lambda = 1{,}000$ versus 12,000 under KMZ) produces notably worse performance against the random walk: R^2_{RFF} values range from -0.225 to -0.366, far below the near-zero values (-0.016 to -0.077) recorded under KMZ. The CW test favours Ridge–RFF over the random walk for only two currencies: CHE (1.323^{*}) and JPY (1.786^{**}), compared with AUD, CHE, and JPY under the KMZ specification. This difference illustrates that even at $T = 12$, regularisation intensity matters: the twelve-fold greater shrinkage under KMZ brings Ridge–RFF closer to the random walk benchmark and generates more reliable predictive gains.

Table B1: Out-of-sample forecast accuracy: Taylor-rule fundamentals (fixed-λ specification)

	Linear vs. RW		Ridge–RFF vs. RW		Ridge–RFF vs. Linear	
Currency	R^2_{Lin}	CW	R^2_{RFF}	CW	Ratio	DM
Panel A: Rolling window $T = 12$ months ($z = \lambda/T = 83.3$)						
AUD	−10.167	0.590	−0.225	1.068	0.110	4.821^{***}
CAD	−6.256	−1.057	−0.392	−0.646	0.192	7.023^{***}
CHE	−5.445	−0.251	−0.300	1.323^{*}	0.202	5.873^{***}
EUR	−4.604	−0.102	−0.356	0.647	0.242	6.461^{***}
GBP	−4.490	0.953	−0.267	0.123	0.231	6.903^{***}
JPY	−28.664	1.676^{**}	−0.314	1.786^{**}	0.044	1.266
NOK	−5.868	0.409	−0.366	−1.028	0.199	8.064^{***}
SEK	−6.412	0.900	−0.365	−0.039	0.184	5.907^{***}
Panel B: Rolling window $T = 60$ months ($z = \lambda/T = 16.7$)						
AUD	−0.234	0.721	−0.224	1.134	0.992	0.152
CAD	−0.292	0.429	−0.395	−0.754	1.080	−0.916
CHE	−0.213	0.325	−0.301	1.263	1.073	−1.022
EUR	−0.221	1.202	−0.362	−0.116	1.116	−1.291
GBP	−0.218	1.849^{**}	−0.308	−0.448	1.074	−1.380
JPY	−0.282	1.420^{*}	−0.287	1.874^{**}	1.003	−0.053
NOK	−0.221	1.783^{**}	−0.338	−0.552	1.096	-1.886^{*}
SEK	−0.258	0.773	−0.353	0.068	1.075	−1.364
Panel C: Rolling window $T = 120$ months ($z = \lambda/T = 8.3$)						
AUD	−0.046	2.043^{**}	−0.235	1.103	1.181	-3.298^{***}
CAD	−0.062	1.145	−0.414	−0.781	1.332	-3.190^{***}
CHE	−0.105	0.448	−0.314	1.446^{*}	1.189	-3.124^{***}
EUR	−0.092	2.139^{**}	−0.488	−0.882	1.363	-3.139^{***}
GBP	−0.126	0.786	−0.335	−0.565	1.186	-3.403^{***}
JPY	−0.124	0.258	−0.305	1.779^{**}	1.161	-2.031^{**}
NOK	−0.098	0.972	−0.343	−0.558	1.223	-3.948^{***}
SEK	−0.126	−0.040	−0.368	0.068	1.215	-3.377^{***}

This table reports one-month-ahead out-of-sample forecast accuracy statistics for Taylor-rule fundamentals under the fixed-λ specification ($\lambda = 1{,}000$ held constant across all training windows), serving as a robustness counterpart to Table 1. Under this specification the KMZ regularisation parameter $z = \lambda/T$ declines from 83.3 at $T = 12$ to 16.7 at $T = 60$ and 8.3 at $T = 120$, so that regularisation intensity weakens as the training window expands. This contrasts with the KMZ exact specification in Table 1, where $z = 1{,}000$ is held constant and $\lambda = z \times T$ grows with T. The parenthetical z values in each panel header give the implied KMZ regularisation parameter for that window. Taylor-rule predictors comprise U.S. and foreign nominal interest rates, inflation rates, output gaps, and the real exchange rate, following Molodtsova and Papell (2009). All statistics are as defined in Table 1. Stars on CW use one-sided critical values; stars on DM use two-sided critical values.

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

At $T = 60$ (Panel B), the two specifications diverge sharply. Under KMZ (Table 1, Panel B), Ridge–RFF achieves positive R^2_{RFF} for AUD, CHE, and JPY, and all DM statistics are positive and significant, confirming that adequate regularisation allows the model to extract predictive content from Taylor-rule fundamentals. Under the fixed-λ specification, by contrast, no currency records a positive R^2_{RFF} (range: -0.224 to -0.395), and the DM test turns negative for seven of eight currencies, meaning OLS outperforms Ridge–RFF for all but AUD (DM = 0.152, insignificant). The MSPE Ratio exceeds unity for seven currencies — Ridge–RFF has higher forecast error than OLS for CAD (1.080), CHE (1.073), EUR (1.116), GBP (1.074), JPY (1.003), NOK (1.096), and SEK (1.075); only AUD (0.992) records a ratio marginally below unity. The sole instance where Ridge–RFF retains a small advantage over the *random walk* is JPY: a positive and significant CW statistic (1.874^{**}) indicates that despite OLS outperforming Ridge–RFF (Ratio = 1.003), the JPY forecast is still more accurate than the driftless random walk, consistent with JPY exhibiting the most persistent Taylor-rule signal in the sample.

At $T = 120$ (Panel C), the reversal is complete. The DM statistic is negative and significant at the 5% or 1% level for all eight currencies, indicating that OLS uniformly and reliably outperforms Ridge–RFF when z has declined to 8.3. The MSPE Ratio ranges from 1.161 (JPY) to 1.363 (EUR), meaning Ridge–RFF forecasts carry between 16% and 36% higher mean squared error than OLS. The R^2_{RFF} values deteriorate substantially, ranging from -0.235 (AUD) to -0.488 (EUR), far below the near-zero or positive values recorded under KMZ at the same window (range -0.005 to $+0.013$, Table 1, Panel C).

This pattern has a precise mechanistic interpretation, derived from equation (7). Under the fixed-λ specification at $T = 120$, the implied $z = 8.3$ provides negligible shrinkage relative to the average Gram eigenvalue $\bar{\mu} = P/2 = 6{,}000$ (ratio $z/\bar{\mu} \approx 0.001$), so Ridge–RFF operates as a near-unregularised OLS model across 12,000 RFF features. Such a model captures spurious correlations in the 120-observation training window at least as aggressively as plain OLS on the original predictors, while adding the noise of 12,000 randomly generated features. The result is a model strictly worse than OLS: it inherits OLS's noise amplification while compounding it. Under KMZ at the same window, $z = 1{,}000$ implies $\lambda = 120{,}000$, which provides substantial shrinkage ($z/\bar{\mu} \approx 0.17$, $\text{df}/T \approx 0.04$ from equation (10)) and stabilises forecasts close to the random walk benchmark.

The joint evidence from the two specifications confirms the interpretation in Section 4: the improvement of Ridge–RFF over OLS — and its approach toward the random walk — visible under KMZ at $T = 60$ and $T = 120$ is driven by the maintenance of adequate regularisation intensity ($z = 1{,}000$ constant) rather than by any reduction in nominal complexity ($c = P/T$). When z is held constant, declining complexity and increasing shrinkage move in a coherent

direction as T grows. When λ is held constant, declining complexity and declining $z = \lambda/T$ move in opposite directions, making the cross-window comparison uninterpretable as a test of the virtue-of-complexity hypothesis.

B.2 Traditional monetary fundamentals

Table B2 shows that the qualitative pattern for Taylor-rule fundamentals carries over to the traditional monetary predictor set, though the contrasts are less pronounced, consistent with the weaker predictive signal of Meese–Rogoff fundamentals documented in the main text.

At $T = 12$ (Panel A), the DM statistics are positive and significant for all eight currencies: seven at the 1% level and one (NOK) at the 5% level, confirming that Ridge–RFF substantially outperforms OLS with or without adequate regularisation when nominal complexity is extreme. The CW test favours Ridge–RFF over the random walk for CAD (2.026^{**}) and JPY (2.671^{***}). As under Taylor rule, the lower regularisation of the fixed-λ case relative to KMZ produces a less favourable comparison against the random walk: the fixed-λ CW for JPY is 2.671^{***} versus 2.397^{***} under KMZ, while other currencies record weaker or insignificant CW statistics under fixed-λ.

At $T = 60$ (Panel B), the DM statistics turn mixed under fixed-λ: positive for AUD (0.604) and EUR (0.018) but negative for six currencies (CAD, CHE, GBP, JPY, NOK, SEK). The CW statistic for CHE turns significantly negative (-2.493^{***}), indicating the random walk formally dominates Ridge–RFF for that currency — a stronger result than the -1.606^{*} recorded under KMZ at the same window (Table 2, Panel B). As under Taylor rule, the weakening regularisation under fixed-λ pushes the model toward an unregularised solution that performs poorly relative to both OLS and the random walk.

At $T = 120$ (Panel C), the comparison between specifications is most instructive for understanding the OLS model's role. Under KMZ (Table 2, Panel C), linear OLS achieves significant CW statistics for CAD (2.471^{***}) and CHE (1.654^{**}), and NOK records a positive $R^2_{\text{Lin}} = 0.016$, consistent with the predictability documented by (Molodtsova and Papell, 2009; Engel and Wu, 2024). Under fixed-λ, the OLS model is unchanged (it does not depend on λ), so these OLS results are identical across both tables. What differs is the Ridge–RFF comparison: under KMZ, $\mathrm{df}/T \approx 0.04$ places Ridge–RFF close to the random walk and competitive with OLS for several currencies; under fixed-λ, $\mathrm{df}/T \approx 0.86$ means Ridge–RFF behaves as near-unregularised OLS across the full 12,000-dimensional feature space. The DM statistics at $T = 120$ reflect this: five of eight currencies record negative DM statistics (CAD, CHE, EUR, NOK, SEK), while three (AUD, GBP, JPY) remain positive, though none of the positive values are statistically significant. The sole DM statistic significant at the 10% level

Table B2: Out-of-sample forecast accuracy: traditional monetary fundamentals (fixed-λ specification)

	Linear vs. RW		Ridge–RFF vs. RW		Ridge–RFF vs. Linear	
Currency	R^2_{Lin}	CW	R^2_{RFF}	CW	Ratio	DM
Panel A: Rolling window $T = 12$ months ($z = \lambda/T = 83.3$)						
AUD	−1.274	0.491	−0.252	−0.003	0.551	3.267***
CAD	−0.726	1.767**	−0.104	2.026**	0.640	5.144***
CHE	−1.007	0.678	−0.243	0.247	0.619	3.750***
EUR	−1.156	−0.188	−0.208	−0.417	0.560	2.670***
GBP	−0.754	−0.886	−0.251	0.521	0.713	4.581***
JPY	−0.990	1.084	−0.152	2.671***	0.579	3.439***
NOK	−0.636	−0.506	−0.222	0.519	0.747	2.317**
SEK	−0.771	−0.428	−0.209	0.113	0.682	4.053***
Panel B: Rolling window $T = 60$ months ($z = \lambda/T = 16.7$)						
AUD	−0.155	0.964	−0.116	−0.859	0.966	0.604
CAD	−0.060	2.294**	−0.065	−0.163	1.005	−0.124
CHE	−0.097	0.718	−0.182	−2.493***	1.078	−1.338
EUR	−0.077	−0.276	−0.076	−0.887	0.999	0.018
GBP	−0.056	0.101	−0.079	0.605	1.021	−0.575
JPY	−0.109	0.312	−0.136	−0.388	1.025	−0.576
NOK	−0.093	−0.332	−0.204	−1.288*	1.101	−1.123
SEK	−0.097	−1.062	−0.101	−0.214	1.003	−0.087
Panel C: Rolling window $T = 120$ months ($z = \lambda/T = 8.3$)						
AUD	−0.067	0.257	−0.056	−0.789	0.989	0.488
CAD	−0.010	2.471***	−0.019	0.842	1.009	−0.412
CHE	−0.019	1.654**	−0.104	−0.961	1.083	−1.130
EUR	−0.017	0.539	−0.038	−0.399	1.020	−0.591
GBP	−0.024	0.409	−0.021	0.749	0.997	0.159
JPY	−0.060	−0.204	−0.059	−0.495	0.999	0.033
NOK	0.016	1.096	−0.157	−0.370	1.176	−1.379*
SEK	−0.057	−1.300*	−0.087	−0.814	1.028	−0.793

This table reports one-month-ahead out-of-sample forecast accuracy statistics for traditional monetary fundamentals under the fixed-λ specification ($\lambda = 1{,}000$ held constant), serving as a robustness counterpart to Table 2. The fixed-λ specification implies a KMZ regularisation parameter $z = \lambda/T$ that declines from 83.3 at $T = 12$ to 16.7 at $T = 60$ and 8.3 at $T = 120$ as the training window expands, so that regularisation intensity weakens with T. Traditional monetary predictors are money supply growth, interest rate, inflation, and output growth differentials between the U.S. and each foreign country, following Meese and Rogoff (1983). All statistics are as defined in Table 1. Stars on CW use one-sided critical values; stars on DM use two-sided critical values.

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

is NOK (-1.379^{*}), where OLS marginally dominates Ridge–RFF. The smaller magnitudes relative to Taylor rule reflect the fewer original predictors ($k = 4$ under traditional versus $k = 7$ under Taylor rule): with a lower-dimensional input space, the noise amplification from the unregularised high-dimensional RFF model is less severe.

B.3 Summary

The fixed-λ results in Tables B1 and B2 deliver two clear messages that sharpen the interpretation of the KMZ results in the main text.

First, regularisation intensity matters even at $T = 12$. The fixed-λ case has twelve times less shrinkage than KMZ at this window and already produces worse performance against the random walk, despite both specifications dramatically outperforming OLS. This confirms that $z = 1{,}000$ is a meaningful calibration choice with empirical consequences, not a neutral scaling.

Second, the reversal of the Ridge–RFF versus OLS comparison as T grows — from large positive DM at $T = 12$ to significantly negative DM at $T = 120$ — is a regularisation effect, not a complexity effect. As $z = \lambda/T$ declines, the model loses the shrinkage that suppresses estimation noise, and the Ridge–RFF advantage over OLS disappears. The KMZ specification, which holds z constant, eliminates this confound and shows Ridge–RFF retaining an advantage over OLS at every training window. Neither specification produces a Ridge–RFF model that consistently outperforms the random walk: the random walk benchmark remains robust throughout, consistent with the OOS VoC curve evidence in Section 5 showing that the MSPE ratio relative to the random walk is bounded above by unity across the entire regularisation grid. The primary difference between specifications is not whether the random walk is beaten — it is not, under either — but whether the cross-window pattern of results can be interpreted as evidence on the virtue-of-complexity hypothesis. Only the KMZ specification, by holding regularisation intensity constant, satisfies this condition.

C Single-Currency Market-Timing Results

This appendix reports the individual-currency counterparts to the equal-weighted portfolio results in Section 4.4, together with the random walk single-currency benchmark. The tables follow the same structure and notation as the main-text EW tables. Results are presented for both predictor sets (Taylor-rule and traditional monetary fundamentals) and both intercept specifications (no-intercept baseline and with-intercept robustness check), and are discussed briefly below.

C.1 Taylor-Rule Fundamentals: No-Intercept Baseline

Table C1 reports single-currency performance under Taylor-rule fundamentals without an intercept. The cross-currency picture is heterogeneous, with no currency achieving significance across all three training windows. At $T = 12$, JPY is the only currency generating significant Sharpe ratios for both the linear model (SR = 0.249, $t = 1.740^{*}$) and Ridge–RFF (SR = 0.257, $t = 1.796^{*}$). At $T = 60$, the linear model produces the most reliable individual-currency results: GBP (SR = 0.293, $t = 1.961^{**}$) and NOK (SR = 0.261, $t = 1.746^{*}$) are significant, while JPY achieves marginal significance for Ridge–RFF (SR = 0.274, $t = 1.835^{*}$). At $T = 120$, the linear model delivers its strongest individual-currency results: AUD (SR = 0.354, $t = 2.235^{**}$) and EUR (SR = 0.528, $t = 2.094^{**}$) are both significant, and Ridge–RFF retains a marginal result for JPY (SR = 0.284, $t = 1.787^{*}$) with a borderline significant incremental return over the linear model (IR v. LIN $t = 1.781^{*}$).

C.2 Taylor-Rule Fundamentals: With-Intercept Specification

Table C2 repeats the analysis with an intercept included in both models; the benchmark is the drifting random walk. The most notable contrast with the no-intercept table is the strengthening of JPY's Ridge–RFF result, which is robust to the intercept and in fact statistically sharper: at $T = 12$, JPY RFF achieves SR = 0.331 ($t = 2.228^{**}$) with a significant incremental return over the linear model (IR v. LIN $t = 2.044^{**}$); at $T = 120$, JPY RFF achieves SR = 0.369 ($t = 2.328^{**}$) with significant information ratios relative to both the drifting random walk ($t = 2.458^{**}$) and the linear strategy ($t = 2.406^{**}$). These results confirm that JPY's directional predictability under Taylor-rule fundamentals reflects genuine conditional signal rather than drift capture, consistent with JPY's near-zero unconditional drift against the USD over the evaluation period. At $T = 60$, GBP LIN retains and in fact strengthens its result (SR = 0.365, $t = 2.445^{**}$), also robust to the intercept.

C.3 Traditional Monetary Fundamentals: No-Intercept Baseline

Table C3 presents single-currency results under traditional monetary fundamentals without an intercept. At $T = 12$, CAD LIN is marginally significant (SR = 0.260, $t = 1.797^{*}$), while JPY RFF delivers the strongest individual result (SR = 0.339, $t = 2.348^{**}$) with a significant information ratio relative to the linear strategy (IR v. LIN $t = 2.213^{**}$). At $T = 60$, CAD LIN stands out as the single reliable result (SR = 0.345, $t = 2.279^{**}$), consistent with the CW test result for CAD in Table 2. Ridge–RFF generates no statistically significant positive results at $T = 60$ under traditional fundamentals; CHF RFF records a significantly negative

Table C1: Market-timing performance: single-currency portfolios, Taylor-rule fundamentals (no-intercept specification)

		SR	t	IR v. Mkt	t	IR v. RW	t	IR v. LIN	t	Skew	Max Loss
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	−0.101	−0.705			−1.190	6.819
	CHF	−0.038	−0.251	−0.045	−0.295	−0.101	−0.664			0.794	6.075
	EUR	−0.018	−0.090	−0.020	−0.101	−0.056	−0.275			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	1.042			1.472	4.484
	NOK	0.059	0.409	0.044	0.309	0.075	0.526			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	0.132	0.916	−0.067	−0.464	−4.513	12.845
	CHF	0.190	1.255	0.182	1.202	0.151	0.996	0.222	1.466	0.421	5.209
	EUR	0.134	0.665	0.131	0.652	0.150	0.741	0.152	0.752	−0.539	4.787
	GBP	0.018	0.124	0.025	0.174	−0.084	−0.584	−0.044	−0.304	−0.796	6.812
	JPY	0.257	1.796*	0.244	1.699*	0.074	0.517	0.167	1.158	−0.885	10.405
	NOK	−0.152	−1.059	−0.147	−1.022	−0.176	−1.230	−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
Training window size $T = 60$											
LIN	AUD	0.120	0.801	0.110	0.738	0.042	0.281			2.612	6.044
	CAD	0.053	0.355	0.070	0.469	0.143	0.955			0.645	10.855
	CHF	0.047	0.294	−0.047	−0.293	0.028	0.178			−4.328	13.051
	EUR	0.237	1.081	0.263	1.195	0.328	1.487			−3.892	10.332
	GBP	0.293	1.961**	0.280	1.873*	0.223	1.488			1.625	5.190
	JPY	0.208	1.391	0.195	1.301	0.103	0.688			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	0.031	0.207			−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	0.071	0.476	−0.141	−0.941	−4.616	12.426
	CHF	0.193	1.220	0.169	1.061	0.200	1.258	0.188	1.182	0.144	5.382
	EUR	−0.027	−0.123	−0.024	−0.110	0.031	0.142	−0.150	−0.679	−1.045	5.318
	GBP	−0.065	−0.432	−0.048	−0.323	−0.168	−1.122	−0.157	−1.048	−0.796	6.287
	JPY	0.274	1.835*	0.265	1.769*	0.090	0.601	0.212	1.411	−0.798	10.498
	NOK	−0.083	−0.557	−0.083	−0.554	−0.117	−0.782	−0.190	−1.268	−1.243	6.667
	SEK	0.011	0.071	0.012	0.077	−0.084	−0.560	−0.025	−0.170	−0.310	5.489
Training window size $T = 120$											
LIN	AUD	0.354	2.235**	0.351	2.212**	0.349	2.196**			2.173	6.429
	CAD	0.147	0.930	0.159	0.999	0.228	1.436			2.181	10.080
	CHF	0.080	0.470	0.052	0.307	0.048	0.284			−0.329	4.905
	EUR	0.528	2.094**	0.572	2.259**	0.586	2.304**			0.926	4.678
	GBP	0.119	0.753	0.126	0.796	0.043	0.267			0.005	7.143
	JPY	0.040	0.255	0.036	0.224	−0.051	−0.321			−3.554	11.821
	NOK	0.156	0.983	0.156	0.984	0.186	1.167			1.316	4.698
	SEK	−0.007	−0.042	0.001	0.008	−0.001	−0.009			−1.364	7.251
RFF	AUD	0.179	1.132	0.177	1.115	0.136	0.852	0.027	0.172	0.175	7.937
	CAD	−0.125	−0.792	−0.125	−0.785	0.042	0.263	−0.174	−1.095	−4.496	11.805
	CHF	0.238	1.406	0.227	1.333	0.245	1.441	0.225	1.324	0.539	5.624
	EUR	−0.234	−0.927	−0.233	−0.920	−0.170	−0.668	−0.372	−1.453	−1.180	4.823
	GBP	−0.086	−0.542	−0.095	−0.600	−0.180	−1.133	−0.122	−0.767	−0.738	6.241
	JPY	0.284	1.787*	0.267	1.676*	0.068	0.423	0.283	1.781*	−1.241	10.610
	NOK	−0.089	−0.563	−0.089	−0.562	−0.080	−0.504	−0.135	−0.849	−1.222	6.306
	SEK	0.011	0.071	0.013	0.080	−0.071	−0.448	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. Results are presented for linear regression (LIN) and Ridge–RFF (RFF) models. Taylor-rule predictors comprise U.S. and foreign nominal interest rates, inflation rates, output gaps, and the real exchange rate, following Molodtsova and Papell (2009). No intercept is included; the benchmark is the driftless random walk. Strategy returns are computed as $\hat{r}_{t|t-1} \times r_t$ (magnitude-proportional positioning). SR is the annualised Sharpe ratio; t is the t-statistic for mean strategy returns. IR v. Mkt is the information ratio relative to the equal-weighted buy-and-hold market return; IR v. RW is the information ratio relative to the random walk strategy; IR v. LIN (RFF rows only) is the information ratio of Ridge–RFF relative to the linear strategy. Skew is return skewness; Max Loss is the maximum loss in standard deviation units.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-sided).

Table C2: Market-timing performance: single-currency portfolios, Taylor-rule fundamentals (with-intercept specification)

		SR	t	IR v. Mkt	t	IR v. RW	t	IR v. LIN	t	Skew	Max Loss
Training window size $T = 12$											
LIN	AUD	0.131	0.909	0.140	0.968	0.139	0.965			1.391	6.452
	CAD	−0.191	−1.336	−0.213	−1.483	−0.197	−1.376			−2.751	11.003
	CHF	−0.115	−0.764	−0.125	−0.824	−0.125	−0.822			−1.013	7.860
	EUR	0.007	0.036	0.005	0.025	0.007	0.035			−0.489	5.929
	GBP	0.092	0.642	0.086	0.600	0.100	0.694			6.069	5.817
	JPY	0.157	1.054	0.150	1.006	0.132	0.881			0.540	5.558
	NOK	0.080	0.539	0.071	0.479	0.080	0.540			0.744	7.697
	SEK	0.205	1.428	0.172	1.193	0.161	1.120			7.399	3.469
RFF	AUD	0.156	1.086	0.136	0.940	0.226	1.568	0.160	1.109	0.516	5.361
	CAD	−0.063	−0.441	−0.066	−0.461	−0.092	−0.639	−0.009	−0.060	−1.935	10.007
	CHF	0.138	0.914	0.119	0.784	0.118	0.782	0.168	1.112	−0.396	5.265
	EUR	0.081	0.405	0.086	0.424	0.129	0.642	0.081	0.403	−0.362	4.559
	GBP	0.129	0.899	0.116	0.809	0.117	0.816	0.123	0.855	−1.374	8.965
	JPY	0.331	2.228**	0.326	2.185**	0.158	1.058	0.305	2.044**	−1.539	9.059
	NOK	−0.038	−0.260	−0.051	−0.341	−0.082	−0.552	−0.055	−0.374	−1.526	8.080
	SEK	0.167	1.163	0.136	0.945	0.033	0.231	0.097	0.673	−0.663	9.887
Training window size $T = 60$											
LIN	AUD	0.032	0.211	0.015	0.098	0.027	0.183			3.672	5.483
	CAD	−0.069	−0.459	−0.062	−0.418	−0.076	−0.509			−1.452	11.135
	CHF	−0.027	−0.171	−0.112	−0.701	−0.015	−0.098			−3.874	12.444
	EUR	0.244	1.110	0.265	1.202	0.246	1.115			−4.282	10.523
	GBP	0.365	2.445**	0.353	2.356**	0.371	2.478**			1.662	4.860
	JPY	0.075	0.501	0.061	0.410	0.082	0.546			0.826	6.148
	NOK	0.064	0.430	0.063	0.420	0.065	0.435			−0.707	6.836
	SEK	−0.013	−0.085	−0.038	−0.255	−0.018	−0.120			1.735	11.775
RFF	AUD	0.145	0.974	0.123	0.821	0.282	1.886*	0.142	0.951	1.539	4.624
	CAD	−0.099	−0.664	−0.090	−0.602	−0.216	−1.443	−0.078	−0.520	−2.477	10.474
	CHF	0.105	0.664	0.024	0.150	0.249	1.567	0.116	0.727	−0.201	5.420
	EUR	0.004	0.017	0.015	0.067	0.016	0.075	−0.068	−0.306	−0.860	4.359
	GBP	−0.144	−0.962	−0.167	−1.115	0.025	0.169	−0.213	−1.418	−3.253	11.128
	JPY	0.239	1.599	0.215	1.433	0.416	2.780***	0.227	1.517	0.607	5.387
	NOK	−0.017	−0.114	−0.063	−0.423	−0.023	−0.157	−0.037	−0.249	−1.066	7.408
	SEK	0.051	0.344	−0.011	−0.076	0.056	0.372	0.059	0.395	−0.245	4.936
Training window size $T = 120$											
LIN	AUD	0.238	1.505	0.234	1.474	0.233	1.469			3.776	5.287
	CAD	0.128	0.809	0.138	0.869	0.179	1.128			0.731	10.964
	CHF	0.074	0.439	0.022	0.130	0.072	0.423			−1.296	8.220
	EUR	0.405	1.606	0.431	1.701*	0.439	1.732*			1.015	4.633
	GBP	0.147	0.929	0.147	0.925	0.104	0.654			0.292	6.446
	JPY	−0.085	−0.538	−0.092	−0.576	−0.073	−0.460			−0.748	9.540
	NOK	0.049	0.307	0.048	0.299	0.063	0.398			0.924	4.652
	SEK	−0.092	−0.580	−0.085	−0.537	−0.054	−0.338			−1.253	7.483
RFF	AUD	0.145	0.913	0.136	0.858	0.342	2.156**	0.106	0.667	1.156	4.215
	CAD	−0.143	−0.903	−0.144	−0.907	−0.100	−0.631	−0.280	-1.763*	−2.384	8.903
	CHF	0.175	1.036	0.100	0.586	0.260	1.531	0.161	0.950	−0.017	5.879
	EUR	−0.089	−0.352	−0.078	−0.308	−0.015	−0.060	−0.242	−0.950	−1.832	6.329
	GBP	−0.232	−1.466	−0.239	−1.508	−0.003	−0.019	−0.233	−1.465	−4.259	12.423
	JPY	0.369	2.328**	0.347	2.179**	0.391	2.458**	0.383	2.406**	0.752	4.992
	NOK	−0.131	−0.824	−0.154	−0.970	−0.024	−0.152	−0.146	−0.920	−0.695	6.354
	SEK	−0.116	−0.730	−0.131	−0.828	0.135	0.848	−0.098	−0.619	−0.581	5.372

This table reports one-month-ahead out-of-sample market-timing strategy performance metrics for single-currency portfolios under Taylor-rule fundamentals with an intercept included in both the linear (LIN) and Ridge–RFF (RFF) models; the random walk benchmark is the drifting random walk. This table is the with-intercept counterpart to Table C1.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-sided).

information ratio relative to the market ($t = -2.273^{**}$), reflecting the CW finding that the driftless random walk outperforms Ridge–RFF for CHF at this window. At $T = 120$, CAD LIN remains the dominant result (SR = 0.401, $t = 2.494^{**}$), with CHF LIN also achieving a significant information ratio relative to the market ($t = 1.859^{*}$). Ridge–RFF adds no significant individual-currency value at $T = 120$ under traditional fundamentals.

C.4 Traditional Monetary Fundamentals: With-Intercept Specification

Table C4 shows that the key results for traditional monetary fundamentals are broadly robust to the intercept. CAD LIN remains significant at $T = 12$ (SR = 0.317, $t = 2.192^{**}$), $T = 60$ (SR = 0.286, $t = 1.894^{*}$), and $T = 120$ (SR = 0.329, $t = 2.051^{**}$), confirming that CAD's directional predictability under traditional fundamentals is not attributable to drift capture. JPY RFF retains significance at $T = 12$ (SR = 0.340, $t = 2.352^{**}$). The most notable change under the intercept is the sharp deterioration of NOK Ridge–RFF at $T = 60$: the information ratio relative to the market falls to -1.143 ($t = -3.784^{***}$), a consequence of the drifting random walk benchmark being better suited to NOK's persistent return pattern than a zero-centred Ridge forecast.

C.5 Random Walk Benchmark

Tables C5 and C6 report the single-currency performance of the random walk timing strategy — lagged exchange rate return as position size — under the no-intercept and with-intercept specifications respectively. Under no-intercept, JPY is the only currency achieving consistent marginal significance at all three training windows ($t \approx 1.72^{*}$ throughout), reflecting the well-documented persistence of the JPY carry momentum. All other currencies are either insignificant or negative across windows. Under the with-intercept specification, JPY's significance is restricted to $T = 12$ (SR = 0.277, $t = 1.917^{*}$) and disappears at longer windows, consistent with the drifting benchmark absorbing the momentum signal that the lagged return otherwise captures. These results confirm that the RW timing strategy offers little systematic economic value across the currency panel, reinforcing the conclusion that the positive EW returns documented in Table 5 for the no-intercept Taylor-rule models at $T = 60$ and $T = 120$ represent genuine — if partly drift-driven — directional performance relative to the random walk benchmark.

Table C3: Market-timing performance: single-currency portfolios, traditional monetary fundamentals (no-intercept specification)

		SR	t	IR v. Mkt	t	IR v. RW	t	IR v. LIN	t	Skew	Max Loss
Training window size $T = 12$											
LIN	AUD	0.068	0.471	0.074	0.511	0.055	0.381			−3.193	11.082
	CAD	0.260	1.797*	0.264	1.822*	0.270	1.861*			2.023	4.466
	CHF	0.107	0.703	0.142	0.932	0.135	0.888			8.809	4.010
	EUR	−0.036	−0.175	−0.044	−0.215	−0.053	−0.260			−2.432	8.989
	GBP	−0.128	−0.884	−0.139	−0.957	−0.156	−1.073			−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	−0.204	−0.786			−1.448	6.268
	SEK	−0.081	−0.407	−0.091	−0.457	−0.108	−0.541			−0.746	5.535
RFF	AUD	−0.022	−0.156	−0.060	−0.415	−0.043	−0.297	−0.029	−0.201	−0.501	7.662
	CAD	0.157	1.086	0.140	0.963	0.190	1.308	0.144	0.992	1.215	6.661
	CHF	0.096	0.631	0.066	0.435	0.064	0.419	0.111	0.728	−0.371	5.732
	EUR	−0.060	−0.295	−0.058	−0.284	−0.101	−0.497	−0.052	−0.255	−0.980	5.317
	GBP	0.048	0.332	0.035	0.240	0.012	0.082	0.091	0.626	−0.132	6.381
	JPY	0.339	2.348**	0.317	2.184**	0.278	1.916*	0.321	2.213**	0.603	6.508
	NOK	−0.202	−0.792	−0.335	−1.303	−0.165	−0.634	−0.168	−0.655	−1.343	5.169
	SEK	−0.001	−0.006	−0.010	−0.050	−0.009	−0.046	0.042	0.209	−0.813	6.371
Training window size $T = 60$											
LIN	AUD	0.135	0.896	0.152	1.008	0.151	1.002			0.508	10.527
	CAD	0.345	2.279**	0.344	2.273**	0.344	2.265**			1.575	5.152
	CHF	0.105	0.655	0.152	0.948	0.104	0.653			2.603	6.384
	EUR	−0.058	−0.260	−0.034	−0.151	−0.049	−0.219			−0.307	5.286
	GBP	0.015	0.102	−0.004	−0.026	−0.065	−0.426			3.644	5.603
	JPY	0.048	0.321	0.056	0.369	0.030	0.201			1.848	5.815
	NOK	−0.107	−0.360	−0.199	−0.660	−0.153	−0.507			0.496	4.824
	SEK	−0.242	−1.118	−0.230	−1.059	−0.259	−1.189			−2.862	8.271
RFF	AUD	−0.145	−0.966	−0.171	−1.133	−0.148	−0.980	−0.178	−1.181	−0.128	7.849
	CAD	−0.007	−0.046	0.007	0.048	0.021	0.138	0.021	0.136	−3.650	12.424
	CHF	−0.261	−1.637	−0.364	-2.273**	−0.298	-1.864*	−0.259	−1.621	−1.237	5.906
	EUR	−0.090	−0.407	−0.075	−0.335	−0.076	−0.342	−0.081	−0.363	−1.547	5.941
	GBP	0.093	0.616	0.074	0.487	−0.004	−0.026	0.102	0.675	3.927	5.982
	JPY	0.066	0.438	0.036	0.238	0.024	0.155	0.062	0.411	0.184	5.080
	NOK	−0.245	−0.826	−1.175	-3.889***	−0.234	−0.777	−0.231	−0.773	−1.001	3.413
	SEK	0.019	0.089	0.035	0.161	0.016	0.075	0.151	0.693	−0.867	5.196
Training window size $T = 120$											
LIN	AUD	0.039	0.244	0.042	0.261	0.043	0.268			1.166	7.568
	CAD	0.401	2.494**	0.401	2.490**	0.399	2.471**			1.417	4.938
	CHF	0.269	1.574	0.318	1.859*	0.273	1.595			6.779	6.753
	EUR	0.132	0.517	0.142	0.554	0.136	0.529			0.349	4.012
	GBP	0.064	0.399	0.068	0.422	−0.002	−0.010			2.307	5.923
	JPY	−0.033	−0.206	−0.017	−0.106	−0.050	−0.307			−3.012	11.045
	NOK	0.480	1.209	0.427	1.053	0.455	1.117			0.539	3.848
	SEK	−0.352	−1.423	−0.266	−1.068	−0.421	-1.687*			−3.696	8.594
RFF	AUD	−0.124	−0.776	−0.126	−0.790	−0.117	−0.729	−0.139	−0.869	0.295	6.490
	CAD	−0.019	−0.118	−0.015	−0.095	−0.043	−0.268	−0.031	−0.191	−2.858	11.017
	CHF	0.106	0.623	0.025	0.149	0.101	0.588	0.135	0.788	−0.915	7.194
	EUR	−0.098	−0.384	−0.059	−0.228	−0.102	−0.397	−0.101	−0.393	−1.451	5.792
	GBP	0.079	0.492	0.084	0.522	0.002	0.012	0.050	0.311	3.253	5.986
	JPY	0.160	0.995	0.107	0.668	0.119	0.736	0.160	0.992	0.995	4.585
	NOK	0.047	0.118	−1.049	-2.590***	0.010	0.025	−0.042	−0.103	−0.029	2.677
	SEK	−0.068	−0.274	−0.029	−0.115	−0.072	−0.287	0.117	0.470	−0.499	4.827

This table reports one-month-ahead out-of-sample market-timing strategy performance metrics for single-currency portfolios. Results are presented for linear regression (LIN) and Ridge–RFF (RFF) models. Traditional monetary predictors comprise money supply growth, interest rate, inflation, and output growth differentials between the U.S. and each foreign country, following Meese and Rogoff (1983). No intercept is included; the benchmark is the driftless random walk. See, Table C1.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-sided).

Table C4: Market-timing performance: single-currency portfolios, traditional monetary fundamentals (with-intercept specification)

		SR	t	IR v. Mkt	t	IR v. RW	t	IR v. LIN	t	Skew	Max Loss
Training window size $T = 12$											
LIN	AUD	−0.021	−0.145	−0.023	−0.162	−0.030	−0.208			−3.122	10.934
	CAD	0.317	2.192**	0.326	2.248**	0.317	2.189**			1.283	5.001
	CHF	0.121	0.799	0.152	0.999	0.119	0.782			7.457	3.902
	EUR	−0.027	−0.133	−0.026	−0.129	−0.029	−0.142			−1.378	7.775
	GBP	−0.085	−0.586	−0.098	−0.673	−0.136	−0.937			−1.387	9.763
	JPY	0.263	1.819*	0.252	1.735*	0.197	1.354			1.153	5.706
	NOK	−0.049	−0.193	−0.101	−0.394	0.049	0.188			−0.379	4.879
	SEK	−0.148	−0.743	−0.169	−0.848	−0.162	−0.813			−0.742	5.518
RFF	AUD	−0.005	−0.038	−0.042	−0.289	−0.053	−0.370	0.001	0.009	−0.844	7.982
	CAD	0.119	0.822	0.101	0.694	0.312	2.151**	0.069	0.472	0.660	6.573
	CHF	0.084	0.550	0.052	0.343	0.026	0.172	0.087	0.573	−0.500	5.861
	EUR	−0.060	−0.297	−0.057	−0.282	−0.109	−0.536	−0.054	−0.265	−0.900	5.394
	GBP	0.092	0.634	0.076	0.527	0.028	0.193	0.156	1.075	−0.680	7.849
	JPY	0.340	2.352**	0.319	2.202**	0.204	1.405	0.266	1.831*	−0.003	7.267
	NOK	−0.248	−0.973	−0.408	−1.587	0.006	0.023	−0.251	−0.977	−1.578	5.331
	SEK	−0.008	−0.042	−0.027	−0.138	−0.032	−0.161	0.074	0.373	−2.591	9.200
Training window size $T = 60$											
LIN	AUD	0.119	0.795	0.130	0.865	0.140	0.929			0.936	10.011
	CAD	0.286	1.894*	0.288	1.902*	0.289	1.908*			−0.770	9.743
	CHF	0.064	0.402	0.096	0.601	0.071	0.443			2.081	6.140
	EUR	−0.040	−0.180	−0.015	−0.068	−0.039	−0.177			−1.261	7.406
	GBP	−0.061	−0.404	−0.098	−0.650	0.029	0.190			2.510	6.052
	JPY	0.029	0.194	0.010	0.068	0.045	0.301			0.717	5.497
	NOK	0.021	0.070	−0.244	−0.807	−0.036	−0.120			−0.067	5.218
	SEK	0.076	0.350	0.081	0.373	0.123	0.565			−0.971	7.096
RFF	AUD	−0.094	−0.627	−0.116	−0.768	−0.169	−1.122	−0.171	−1.135	0.433	8.209
	CAD	0.012	0.079	0.026	0.173	−0.070	−0.465	−0.068	−0.448	−3.335	12.238
	CHF	−0.247	−1.547	−0.371	-2.314**	−0.359	-2.245**	−0.259	−1.619	−1.596	5.758
	EUR	−0.089	−0.399	−0.070	−0.314	−0.157	−0.707	−0.080	−0.357	−1.625	5.968
	GBP	0.083	0.552	0.051	0.334	0.236	1.557	0.199	1.314	3.998	7.563
	JPY	0.027	0.178	−0.012	−0.079	0.102	0.673	0.017	0.110	−0.061	5.795
	NOK	−0.196	−0.660	−1.143	-3.784***	−1.013	-3.382***	−0.243	−0.812	−0.913	3.542
	SEK	−0.032	−0.150	−0.026	−0.120	0.012	0.054	−0.117	−0.540	−1.999	6.468
Training window size $T = 120$											
LIN	AUD	−0.004	−0.025	−0.003	−0.017	0.038	0.237			0.700	7.476
	CAD	0.329	2.051**	0.330	2.048**	0.393	2.438**			1.161	4.488
	CHF	0.293	1.716*	0.329	1.922*	0.304	1.780*			5.813	6.886
	EUR	0.189	0.741	0.203	0.789	0.273	1.061			1.188	3.732
	GBP	−0.057	−0.355	−0.051	−0.318	0.059	0.363			1.551	5.412
	JPY	0.024	0.150	0.001	0.004	−0.069	−0.429			−2.440	9.625
	NOK	0.491	1.235	0.340	0.839	0.432	1.070			0.740	4.143
	SEK	−0.096	−0.388	−0.060	−0.240	−0.103	−0.416			−1.768	7.324
RFF	AUD	−0.138	−0.863	−0.140	−0.876	−0.099	−0.617	−0.157	−0.984	0.244	6.582
	CAD	−0.032	−0.202	−0.029	−0.181	0.215	1.336	−0.175	−1.080	−3.104	11.347
	CHF	0.032	0.189	−0.080	−0.466	0.030	0.175	0.070	0.407	−1.256	6.992
	EUR	−0.121	−0.472	−0.080	−0.311	−0.026	−0.101	−0.226	−0.880	−1.569	5.995
	GBP	0.086	0.535	0.101	0.630	0.306	1.896*	0.197	1.222	2.839	8.160
	JPY	0.142	0.886	0.084	0.519	−0.041	−0.256	0.144	0.895	0.608	5.517
	NOK	0.055	0.139	−1.107	-2.731***	−0.636	−1.576	−0.301	−0.741	0.165	2.657
	SEK	−0.071	−0.285	−0.081	−0.325	−0.092	−0.369	−0.015	−0.060	−1.791	7.093

This table reports one-month-ahead out-of-sample market-timing strategy performance metrics for single-currency portfolios. Results are presented for linear regression (LIN) and Ridge–RFF (RFF) models. Traditional monetary predictors as in Table C3. Both models include an intercept; the benchmark is the drifting random walk. This table is the with-intercept counterpart to Table C3.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-sided).

Table C5: Market-timing performance: random walk benchmark strategy (no-intercept specification)

	SR	t	IR v. Mkt	t	Skew	Max Loss
Training window size $T = 12$						
AUD	0.070	0.485	0.046	0.319	0.322	10.487
CAD	−0.173	−1.196	−0.166	−1.142	−1.448	9.662
CHF	0.101	0.661	0.084	0.553	−1.773	9.897
EUR	0.060	0.297	0.058	0.283	−2.589	9.326
GBP	0.183	1.267	0.171	1.178	4.089	4.362
JPY	0.249	1.721*	0.230	1.584	−0.275	8.182
NOK	−0.414	−1.615	−0.435	-1.686*	−0.912	4.035
SEK	0.010	0.049	−0.009	−0.047	0.069	7.136
Training window size $T = 60$						
AUD	0.151	1.006	0.135	0.898	3.024	5.318
CAD	−0.170	−1.125	−0.165	−1.090	−1.386	9.346
CHF	0.034	0.216	0.009	0.053	−2.086	9.842
EUR	−0.062	−0.280	−0.064	−0.286	−3.550	9.668
GBP	0.185	1.225	0.166	1.098	4.793	4.416
JPY	0.261	1.726*	0.250	1.648*	0.207	8.638
NOK	−0.242	−0.811	−0.285	−0.940	−0.625	3.705
SEK	−0.050	−0.229	−0.073	−0.336	0.083	6.720
Training window size $T = 120$						
AUD	0.105	0.658	0.104	0.652	2.965	5.055
CAD	−0.165	−1.028	−0.166	−1.033	−1.321	8.862
CHF	0.038	0.225	0.030	0.173	−2.986	10.582
EUR	−0.178	−0.694	−0.201	−0.781	−0.737	4.241
GBP	0.164	1.023	0.172	1.068	5.296	4.373
JPY	0.277	1.722*	0.260	1.612	−0.255	8.973
NOK	−0.298	−0.745	−0.286	−0.703	−0.410	2.888
SEK	−0.025	−0.100	−0.120	−0.479	0.161	6.296

This table reports one-month-ahead out-of-sample market-timing strategy performance metrics for single-currency portfolios under the random walk (RW) benchmark. The out-of-sample evaluation period for each T is anchored to the traditional monetary fundamentals OLS forecast dates. No intercept; driftless random walk benchmark. Strategy returns are computed as $r_{t-1} \times r_t$ (lagged exchange rate return as position size). SR denotes the annualised Sharpe ratio; t is the associated t-statistic for mean returns; IR v. Mkt is the information ratio relative to the equal-weighted buy-and-hold return. Skew is the skewness of strategy returns; Max Loss is the maximum loss in standard deviation units.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-sided).

Table C6: Market-timing performance: random walk benchmark strategy (with-intercept specification)

	SR	t	IR v. Mkt	t	Skew	Max Loss
Training window size $T = 12$						
AUD	0.021	0.145	−0.011	−0.073	−0.445	5.841
CAD	0.013	0.087	−0.008	−0.058	−1.407	9.003
CHF	0.081	0.536	0.055	0.362	−0.089	5.243
EUR	0.002	0.010	0.011	0.052	−0.704	4.516
GBP	0.090	0.624	0.072	0.494	−2.251	9.936
JPY	0.277	1.917*	0.257	1.772*	−0.783	8.680
NOK	−0.284	−1.112	−0.458	-1.781*	−1.055	5.689
SEK	0.003	0.015	−0.024	−0.122	−1.925	9.093
Training window size $T = 60$						
AUD	−0.029	−0.193	−0.056	−0.369	0.640	5.784
CAD	0.038	0.253	0.050	0.327	−0.745	7.255
CHF	−0.080	−0.503	−0.204	−1.276	−1.349	6.462
EUR	−0.010	−0.046	0.015	0.065	−1.004	5.272
GBP	−0.175	−1.161	−0.213	−1.405	−4.630	12.450
JPY	−0.031	−0.205	−0.076	−0.501	−0.156	6.242
NOK	0.138	0.466	−0.476	−1.574	−0.837	3.323
SEK	−0.045	−0.208	−0.049	−0.224	−0.012	3.177
Training window size $T = 120$						
AUD	−0.106	−0.661	−0.110	−0.685	−0.043	5.678
CAD	−0.101	−0.626	−0.100	−0.623	−1.589	8.383
CHF	0.020	0.118	−0.113	−0.657	−1.111	7.012
EUR	−0.118	−0.464	−0.078	−0.304	−0.986	4.320
GBP	−0.282	-1.755*	−0.304	-1.893*	−6.207	13.709
JPY	0.195	1.216	0.144	0.898	−0.500	8.622
NOK	0.238	0.599	−0.745	-1.840*	0.036	2.063
SEK	−0.022	−0.087	−0.034	−0.137	−0.473	4.140

This table reports one-month-ahead out-of-sample market-timing strategy performance metrics for single-currency portfolios under the random walk (RW) benchmark. The out-of-sample evaluation period for each T is anchored to the traditional monetary fundamentals OLS forecast dates. With-intercept specification; drifting random walk benchmark. This table is the with-intercept counterpart to Table C5. Strategy returns are computed as $r_{t-1} \times r_t$ (lagged exchange rate return as position size). SR denotes the annualised Sharpe ratio; t is the associated t-statistic for mean returns; IR v. Mkt is the information ratio relative to the equal-weighted buy-and-hold return. Skew is the skewness of strategy returns; Max Loss is the maximum loss in standard deviation units.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-sided).

D Single-Currency Sharpe Ratios across the Complexity Spectrum

D.1 Single-Currency Sharpe Ratios: without-intercept

Figures D1 and D2 report the annualised Sharpe ratio for each individual currency as a function of $\log_{10}(z)$ under the no-intercept baseline. Each figure contains three panels corresponding to $T \in \{12, 60, 120\}$. The cross-currency mean Ridge–RFF Sharpe (thick black line), OLS mean (dashed gold), and random walk mean (dotted green) are superimposed for reference.

For Taylor-rule fundamentals (Figure D1), there is considerable cross-currency dispersion at all training windows, most pronounced at $T = 12$: JPY and EUR tend to generate the highest individual Sharpe ratios (reaching approximately 0.40–0.45 at higher z values), while GBP exhibits the largest swings, with a deep trough around $\log_{10}(z) \approx 2$–3 that reflects the sensitivity of its forecast-weighted position to the specific regularisation level. The cross-currency mean RFF Sharpe (thick black line) remains consistently above the OLS mean (≈ 0.07) and the random walk mean (≈ 0.06) at $T = 12$, reaching approximately 0.11–0.14 across the full grid. At $T = 60$ and $T = 120$, the mean RFF Sharpe rises monotonically from slightly negative values at low z to converge with the OLS mean of approximately 0.17–0.18 at and beyond z_{equity}, consistent with the EW portfolio patterns in Figure **??**. Individual-currency dispersion narrows at longer windows as stronger regularisation suppresses the influence of idiosyncratic noise in forecast weights, and the mean RFF curve tracks the OLS benchmark closely at high z, confirming that the EW portfolio result reflects a genuine cross-currency average rather than a small number of outlier currencies.

For traditional monetary fundamentals (Figure D2), the cross-currency dispersion is substantially wider and more erratic. At $T = 12$, the mean RFF Sharpe (thick black) starts at approximately 0.18, declines monotonically across the grid to approximately 0.07 at z_{equity}, and remains near that level for higher z. The OLS mean is approximately 0.02 throughout, well below the RFF mean at short windows. At $T = 60$, the mean RFF Sharpe is negative across most of the grid, with a pronounced trough near $\log_{10}(z) \approx 1$–2 driven in part by GBP, which exhibits an extreme negative excursion at intermediate z values. The mean RFF rises toward zero by z_{equity} but does not cross zero convincingly within the grid. At $T = 120$, the mean RFF starts near zero, gradually rises, and approaches the OLS mean of approximately 0.12 only at very high z values, remaining below the OLS benchmark throughout. These results confirm that the cross-currency aggregation in Figure **??** is essential: under traditional monetary fundamentals, individual-currency models offer little reliable directional accuracy,

and the positive EW Sharpe at $T = 12$ is driven by the cross-currency average rather than any single dominant currency.

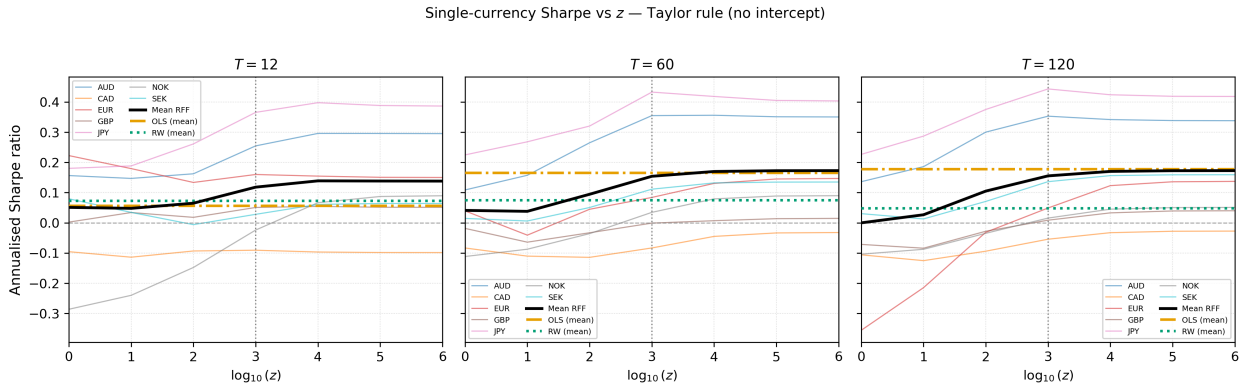


Figure D1: Single-currency annualised Sharpe ratios as a function of $\log_{10}(z)$, Taylor-rule fundamentals, no-intercept baseline. Each panel corresponds to a training window $T \in \{12, 60, 120\}$ (left to right). Thin coloured lines show individual currencies; the thick black line is the cross-currency mean RFF Sharpe; the dashed gold line is the OLS mean; the dotted green line is the random walk mean. The vertical dotted line marks $z_{\text{equity}} = 1{,}000$.

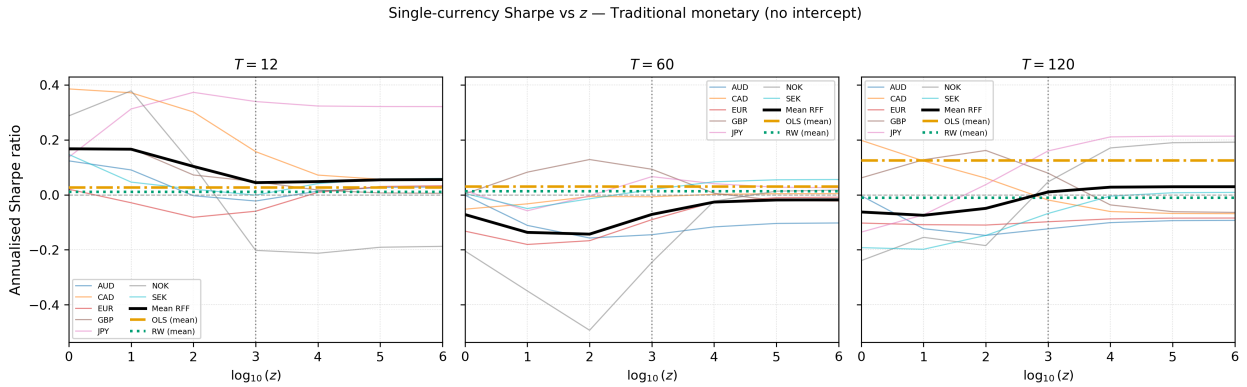


Figure D2: Single-currency annualised Sharpe ratios as a function of $\log_{10}(z)$, traditional monetary fundamentals, no-intercept baseline. Layout as in Figure D1.

D.2 Single-Currency Sharpe Ratios: with-intercept

Figures D3 and D4 repeat the single-currency Sharpe ratio analysis under the with-intercept specification, providing the individual-currency counterpart to the equal-weighted results in Section 5.6.

For Taylor-rule fundamentals (Figure D3), the individual-currency profiles at $T = 12$ remain broadly positive across the full z grid, with the cross-currency mean RFF Sharpe fluctuating near 0.07–0.10, marginally above the OLS mean (≈ 0.05) and the random walk

mean (≈ 0.05). The individual-currency dispersion at $T = 12$ is notable: JPY and some other currencies maintain positive individual Sharpe ratios throughout, while GBP and CAD are occasionally negative. At $T = 60$, the mean RFF Sharpe is near zero or slightly positive at low z but falls toward zero and below as z rises past z_{equity}, with the OLS mean (approximately 0.08) consistently above the RFF mean for most of the grid. At $T = 120$, the mean RFF line starts near 0.05 at low z but declines through zero and turns negative around z_{equity}, lying clearly below the OLS mean (≈ 0.10) across the grid. The decline in mean RFF Sharpe with z at $T = 60$ and $T = 120$ mirrors the EW portfolio result in Figure **??** and confirms that the collapse of aggregate performance with the intercept is broad-based across currencies rather than driven by any single outlier.

For traditional monetary fundamentals (Figure D4), the with-intercept individual-currency profiles are qualitatively similar to the no-intercept baseline at all training windows. At $T = 12$, the cross-currency mean RFF Sharpe is positive in the range 0.05–0.10 and broadly stable across the z grid, slightly above the OLS mean (≈ 0.05). At $T = 60$ and $T = 120$, the mean RFF Sharpe hovers near zero or slightly below, while the OLS mean (approximately 0.08 and 0.14 respectively) lies clearly above the RFF mean. The relative robustness of the traditional-fundamental single-currency profiles to the intercept specification — in contrast to the pronounced deterioration at longer windows under Taylor rule — is consistent with the weaker unconditional drift content of the traditional predictor set, which limits the drift-capture channel through which the zero-intercept restriction generates spurious positive performance at longer windows.

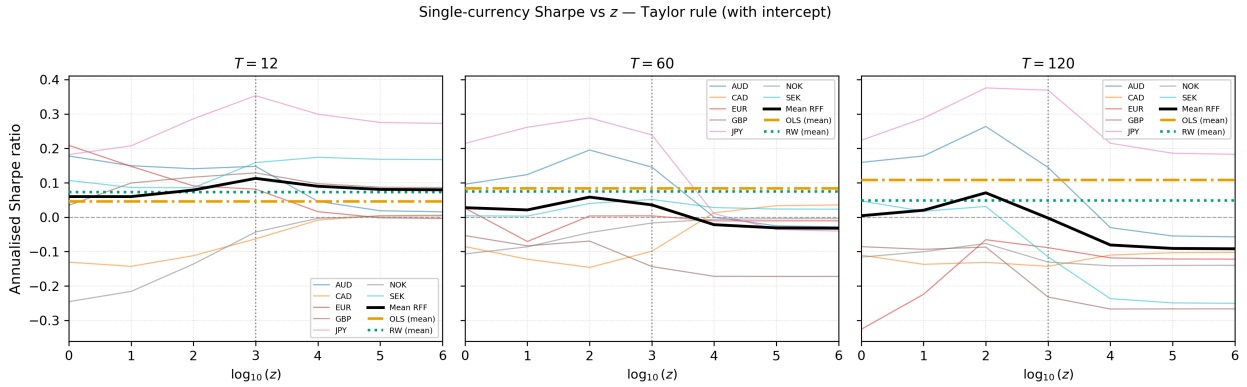


Figure D3: Single-currency annualised Sharpe ratios as a function of $\log_{10}(z)$, Taylor-rule fundamentals, with-intercept specification. Layout as in Figure D1. Compare with Figure D1 (no-intercept baseline).

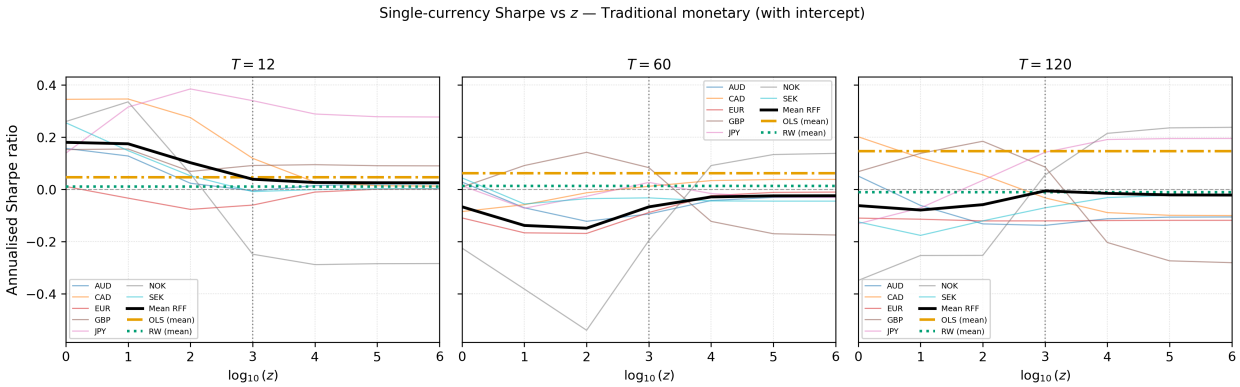


Figure D4: Single-currency annualised Sharpe ratios as a function of $\log_{10}(z)$, traditional monetary fundamentals, with-intercept specification. Layout as in Figure D1. Compare with Figure D2 (no-intercept baseline).

E Implementation Robustness: Raw versus Standardised Random Fourier Features

As noted in Section 2, Fallahgoul (2026) establishes that the Kelly et al. (2024) empirical implementation divides each of the $P = 12{,}000$ RFF features by its within-window sample standard deviation before Ridge estimation — a step that Theorem 3.1 of that paper shows causes the empirical kernel to converge to a data-dependent limit rather than the intended Gaussian kernel, with a structural approximation gap that persists regardless of P. Our baseline implementation does not apply this standardisation, and our results are therefore free of this concern by construction. Nevertheless, to quantify the practical consequences of the standardisation choice in the FX setting and to assess whether the positive individual-currency findings documented in Section 4 are sensitive to it, we re-estimate the Ridge–RFF model under the Kelly et al. (2024) within-window standardisation procedure and compare both the forecast accuracy and market-timing performance of the two variants across all currencies and training windows.

Table E1 reveals a pattern that is partly consistent with Fallahgoul (2026) and partly a revealing departure from it. The similarity lies in the forecast accuracy channel: standardised RFF consistently delivers worse out-of-sample R^2 and higher MSPE ratios than raw RFF across all eight currencies and all three training windows, confirming that the kernel-approximation distortion identified in Theorem 3.1 of Fallahgoul (2026) degrades conditional mean estimation irrespective of the training window. The degradation is dramatic at $T = 12$, where standardisation renders R^2 between 4.7 and 12.3 times more negative than the raw counterpart (equal-weighted: -0.210 versus -0.025), consistent with the structural kernel-approximation gap that Theorem 3.1 predicts when the training window is short and within-window standard deviations are highly variable across rolling windows. The most consequential manifestation of this degradation appears at $T = 60$: the three currencies that achieve positive R^2 under raw RFF — AUD (0.011), CHF (0.007), and JPY (0.015) — all turn negative under standardisation (-0.001, -0.016, and -0.001 respectively), meaning the moderate individual-currency gains documented in Section 4 are specific to the raw estimator and would have been invisible under the empirical implementation of Kelly et al. (2024).

The departure from Fallahgoul (2026), however, is equally instructive. In his equity results, standardisation inflates strategy Sharpe ratios despite worsening R^2 — the signature of state-dependent position sizing that amplifies genuine signal in the economic channel even as it distorts the conditional mean in the statistical channel. In FX, this dissociation does not materialise: raw RFF produces *higher* Sharpe ratios than standardised RFF for seven of eight currencies at both $T = 12$ and $T = 60$, with the raw equal-weighted

Table E1: Standardised vs. raw RFF: out-of-sample comparison under Taylor-rule fundamentals (no intercept). $P = 12{,}000$, $z = 1{,}000$.

	R^2		MSPE		DM		SR			
Currency	Std	Raw	Std	Raw	Std	Raw	Std	(t)	Raw	(t)
Panel A: Training window $T = 12$ months										
AUD	−0.197	−0.225	0.107	0.110	4.832***	4.821***	0.171	1.192	0.159	1.114
CAD	−0.362	−0.392	0.188	0.192	7.083***	7.023***	−0.099	−0.693	−0.094	−0.656
CHE	−0.282	−0.300	0.199	0.202	5.945***	5.873***	0.131	0.870	0.190	1.255
EUR	−0.363	−0.356	0.243	0.242	6.486***	6.461***	0.107	0.534	0.134	0.665
GBP	−0.234	−0.267	0.225	0.231	6.942***	6.903***	0.021	0.145	0.018	0.124
JPY	−0.215	−0.314	0.041	0.044	1.271	1.266	0.336	2.343**	0.257	1.796*
NOK	−0.306	−0.366	0.190	0.199	8.118***	8.064***	−0.186	−1.297	−0.152	−1.059
SEK	−0.295	−0.365	0.175	0.184	5.977***	5.907***	0.039	0.270	−0.006	−0.041
EW	−0.210	−0.244	0.273	0.281	3.706***	3.668***	0.139	0.969	0.120	0.837
Panel B: Training window $T = 60$ months										
AUD	−0.001	−0.224	0.811	0.992	4.141***	0.152	0.267	1.787*	0.175	1.174
CAD	−0.047	−0.395	0.810	1.080	3.531***	−0.916	−0.115	−0.773	−0.114	−0.765
CHE	−0.016	−0.301	0.837	1.073	2.683***	−1.022	0.210	1.322	0.193	1.220
EUR	−0.029	−0.362	0.843	1.116	1.912*	−1.291	0.102	0.466	−0.027	−0.123
GBP	−0.029	−0.308	0.845	1.074	3.403***	−1.380	−0.023	−0.154	−0.065	−0.432
JPY	−0.001	−0.287	0.780	1.003	5.132***	−0.053	0.335	2.239**	0.274	1.835*
NOK	−0.024	−0.338	0.839	1.096	4.167***	−1.886*	0.014	0.095	−0.083	−0.557
SEK	−0.022	−0.353	0.812	1.075	4.152***	−1.364	0.083	0.556	0.011	0.071
EW	−0.001	−0.213	0.907	1.098	3.000***	−2.508**	0.251	1.680*	0.139	0.931
Panel C: Training window $T = 120$ months										
AUD	0.008	−0.235	0.948	1.181	1.378	−3.298***	0.319	2.011**	0.179	1.132
CAD	−0.017	−0.414	0.958	1.332	1.146	−3.190***	−0.083	−0.526	−0.125	−0.792
CHE	0.004	−0.314	0.901	1.189	3.075***	−3.124***	0.281	1.658*	0.238	1.406
EUR	−0.014	−0.488	0.929	1.363	1.048	−3.139***	0.004	0.016	−0.234	−0.927
GBP	−0.010	−0.335	0.897	1.186	2.827***	−3.403***	−0.005	−0.030	−0.086	−0.542
JPY	0.015	−0.305	0.877	1.161	3.613***	−2.031**	0.423	2.665***	0.284	1.787*
NOK	−0.009	−0.343	0.919	1.223	2.412**	−3.948***	−0.016	−0.102	−0.089	−0.563
SEK	−0.005	−0.368	0.892	1.215	3.077***	−3.377***	0.101	0.636	0.011	0.071
EW	0.006	−0.226	0.959	1.182	1.576	−4.058***	0.327	2.063**	0.134	0.846

This table compares out-of-sample performance of standardised (Std, KMZ footnote 39) and raw (Raw, unstandardised) Ridge–RFF forecasts under Taylor-rule fundamentals (no intercept), with $P = 12{,}000$ features and $z = 1{,}000$. R^2 is the out-of-sample R^2 of the RFF model relative to the driftless random walk (positive values indicate the RFF beats the random walk on MSE). MSPE is the ratio of the RFF mean squared prediction error to the OLS mean squared prediction error; values below unity favour RFF. DM is the Diebold–Mariano test statistic (Harvey et al. 1997 correction) for equal predictive accuracy of RFF versus OLS; positive values favour RFF. SR is the annualised single-currency Sharpe ratio of the magnitude-proportional RFF timing strategy ($\hat{r}_{t|t-1} \times r_t$); (t) is the associated t-statistic for mean returns. EW reports equal-weighted portfolio results (expanding membership). Standardised RFF divides each of the P within-window features by its training-sample standard deviation before Ridge estimation, matching the KMZ empirical implementation; raw RFF uses unstandardised sin/cos projections. Following Fallahgoul (2026, Table 2), worsened R^2 alongside inflated SR in the Std columns would indicate that standardisation functions as state-dependent position sizing rather than improved conditional mean estimation. *, **, *** denote significance at the 10%, 5%, 1% levels (two-sided).

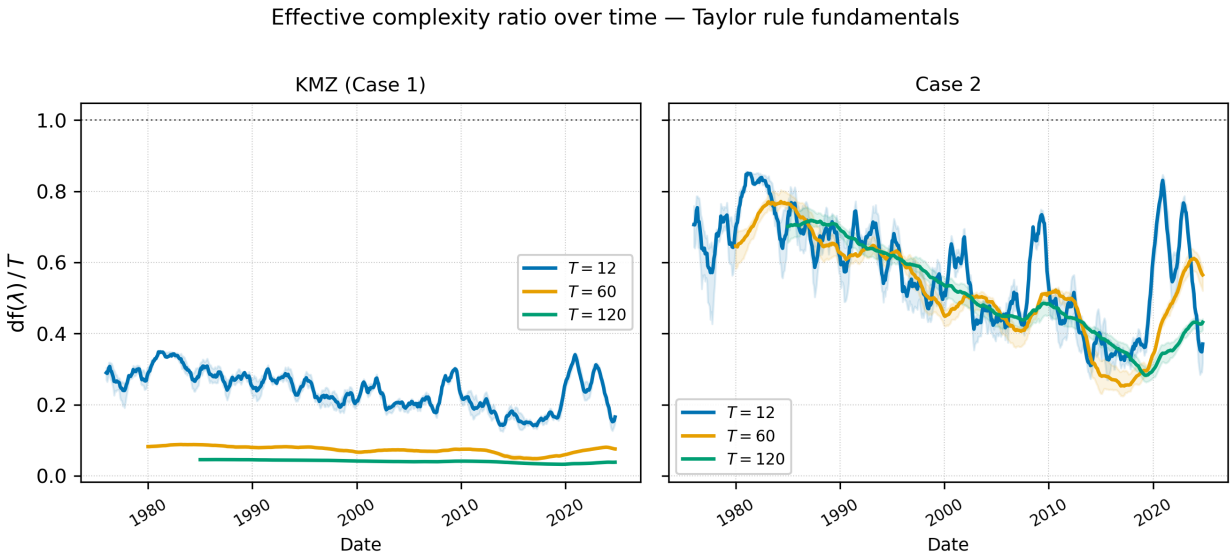
Sharpe ratio exceeding its standardised counterpart by 0.093 and 0.101 respectively. This reversal is not a contradiction of Fallahgoul's mechanism but a direct consequence of its precondition: the SR-inflation effect requires that some true directional signal exist for the window-specific scaling to amplify net positively. In an environment where signal strength $\hat{b}_*$ is of order 10^{-8}, the same state-dependent scaler amplifies misdirected forecasts as often as correct ones, so standardisation is harmful in the economic channel as well as the statistical one. Taken together, the comparison supports two conclusions. First, the paper's implementation using raw, unstandardised features was the appropriate methodological choice for FX: standardisation is uniformly inferior in both the forecast accuracy and market-timing channels, and the positive predictability documented for AUD, CHF, and JPY at $T = 60$ would have been suppressed under the Kelly et al. (2024) procedure. Second, the magnitude of the standardised–raw gap attenuates sharply as the training window lengthens and within-window standard deviations stabilise: by $T = 120$, the equal-weighted R^2 gap narrows to 0.001, the MSPE gap to 0.002, and the equal-weighted Sharpe ratio gap to 0.049, consistent with the kernel distortion of Theorem 3.1 diminishing as T grows.

F Effective Complexity over Time

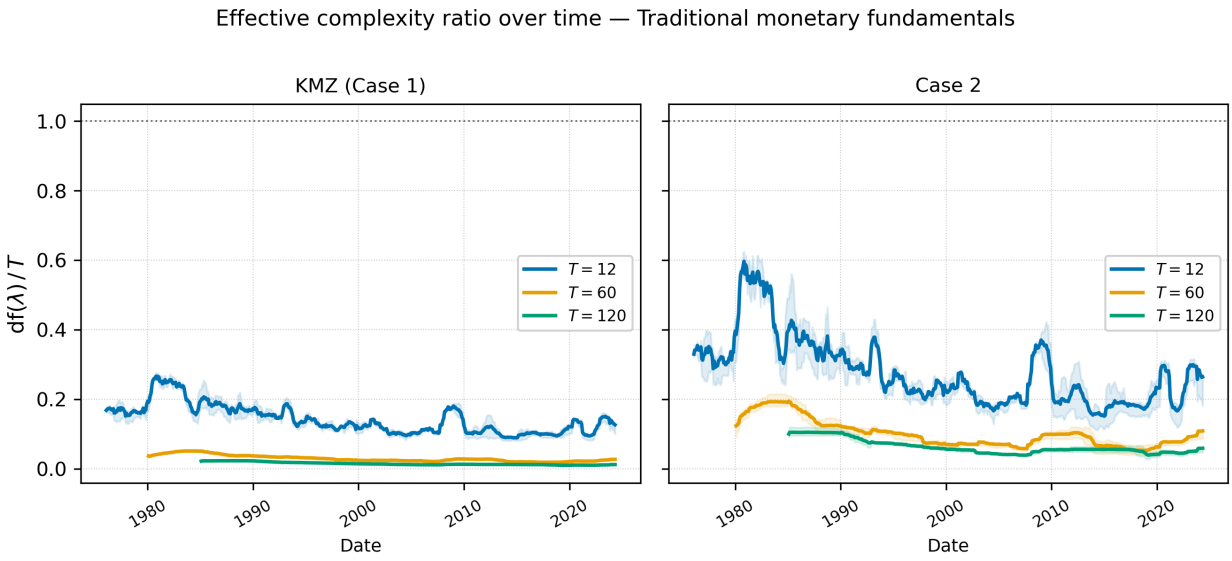
Figure F1 plots the time-series evolution of the effective complexity ratio $\mathrm{df}(\lambda)/T$ at each rolling estimation window for both regularisation cases. The left sub-panel of each figure corresponds to Case 1 ($z = 1{,}000$ fixed), and the right sub-panel to Case 2 ($\lambda = 1{,}000$ fixed).

Under Case 1, effective complexity is nearly constant across the sample for all training windows and both predictor sets, which is by construction: fixing z and T jointly determines a fixed λ for each window, so df/T varies only through changes in the predictor covariance matrix. The mild downward drift visible for $T = 12$ under Case 1 in both predictor sets reflects a gradual increase in collinearity among the RFF features over time, which raises the effective penalty for a given z.

Under Case 2, effective complexity is substantially higher and exhibits a marked downward trend across the sample, most pronounced for $T = 12$. For Taylor rule fundamentals (Figure F1a), df/T at $T = 12$ begins near 0.85 in the late 1970s and declines to approximately 0.30–0.35 by the end of the sample, with a notable trough around 2008–2009 that coincides with the global financial crisis. The longer windows ($T = 60$, $T = 120$) converge toward each other over time, both ending near 0.30–0.40. For traditional monetary fundamentals (Figure F1b), the pattern is sharper: $T = 12$ under Case 2 peaks above 0.55 in the early 1980s before declining steeply to 0.15–0.25 by the mid-1990s and remaining subdued thereafter. The longer windows under Case 2 show little effective complexity throughout, consistent with the weak and stable signal strength of the traditional predictor set documented in Section 5.4. The secular decline in Case 2 effective complexity for both predictor sets is broadly consistent with the documented reduction in FX fundamental predictability over time (Rossi, 2013).



(a) Taylor rule fundamentals



(b) Traditional monetary fundamentals

Figure F1: Time-series of effective complexity ratio $\mathrm{df}(\lambda)/T$. Each figure shows Case 1 ($z = 1{,}000$ fixed; left sub-panel) and Case 2 ($\lambda = 1{,}000$ fixed; right sub-panel). Lines correspond to $T \in \{12, 60, 120\}$ months (blue, orange, green respectively). Shaded bands report cross-currency standard deviations. The horizontal dotted line marks the OLS limit ($\mathrm{df}/T = 1$).