Economic Fundamentals and Short-Run Exchange Rate Prediction: A Machine-Learning Perspective

Ilias Filippou, David E. Rapach, Mark P. Taylor, and Guofu Zhou*

August 19, 2026

Abstract

The exchange rate predictability puzzle—that fundamentals fail to forecast short-horizon exchange rate movements out of sample—has long resisted resolution. We show that fundamentals do matter, but their effects vary with global financial conditions. Using a rich set of country characteristics interacted with global state variables and machine-learning techniques to handle the high-dimensional data, we document that interactions between country characteristics and foreign exchange market volatility are particularly important, consistent with a "flight to fundamentals" during periods of financial stress. Our forecasts deliver substantial economic value and offer new insight into the carry trade.

*Ilias Filippou, ifilippou@fsu.edu, Florida State University Herbert Wertheim College of Business. David E. Rapach, david.rapach@atl.frb.org, Federal Reserve Bank of Atlanta Research Division. Mark P. Taylor, mark.p.taylor@wustl.edu, Washington University in St. Louis Olin Business School and Centre for Economic Policy Research (CEPR). Guofu Zhou, zhou@wustl.edu, Washington University in St. Louis Olin Business School. We thank conference participants at the 2019 City University of Hong Kong Workshop in Econometrics and Statistics, 2019 International Symposium on Forecasting, NIESR/CFM/OMFIF 2019 Workshop on Modeling the Macroeconomy in Risky Times, 2020 Wolfe Virtual Global Quantitative and Macro Investment Conference, 2021 International Workshop on Financial Markets and Nonlinear Dynamics, 2021 Financial Economics Meeting: Crisis Challenges, and 2021 Vienna Symposium on Foreign Exchange Markets; and seminar participants at Aarhus University, Bayes Business School, BI Norwegian Business School, Chinese University of Hong Kong, Chinese University of Hong Kong-Shenzhen, Dongbei University of Finance and Economics, Federal Reserve Bank of Atlanta, Federal Reserve Board, HEC Montreal, IE University, Jinan University, Laval University, London Business School, Shanghai University of Finance and Economics, Syracuse University, University of Liverpool, University of North Carolina at Charlotte, and Washington University in St. Louis. We also thank Svetlana Bryzgalova, Victor DeMiguel, Ana-Maria Fuertes, Federico Gavazzoni, Paolo Giordani, Yufeng Han, Ai He, Dashan Huang, Jiahan Li, Sicong (Allen) Li, Malvina Marchese, Kate Phylaktis, Dagfinn Rime, and Michael Weber (Vienna Symposium discussant) for valuable comments; and Bertrand Gruss for kindly providing data on net commodity returns. We are especially grateful to the editor (Stephan Siegel) and an anonymous referee for detailed and positive comments that significantly improved the paper. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Economic Fundamentals and Short-Run Exchange Rate Prediction: A Machine-Learning Perspective

I. Introduction

Exchange rates affect the cost of trade, the returns to international investors, and the transmission of monetary policy across borders, yet their short-horizon movements have resisted consistent prediction. The seminal study of Meese and Rogoff (1983) showed that forecasts based on macroeconomic fundamentals fail to outperform a simple no-change (i.e., random walk without drift) benchmark, and the four decades of research that followed have produced extensive evidence of predictability only at longer horizons (e.g., Mark (1995), Clarida and Taylor (1997), Clarida, Sarno, Taylor, and Valente (2003), Kilian and Taylor (2003), Eichenbaum, Johannsen, and Rebelo (2021)). Evidence of out-of-sample predictability at short horizons exists but is more scattered, typically focused on a single fundamental, a single currency, or a limited time span, and robust evidence that fundamental-based forecasts can outperform the no-change benchmark at the monthly frequency over a reasonably long period and for a sizable group of exchange rates remains elusive. This persistent difficulty at short horizons—the finding that fundamentals do not seem to matter for short-horizon exchange rate movements and that the simple no-change forecast has proven so difficult to outperform—is the *exchange rate predictability puzzle* surveyed by Rossi (2013).

This paper shows that fundamentals do matter for short-horizon exchange rate movements—but their effects vary systematically with global financial conditions. For a panel of ten developed economies for 1995 to 2020, we find that monthly US dollar exchange rates are consistently and significantly predictable out of sample once the predictive role of fundamentals is allowed to depend on the global state. Earlier work has shown that allowing for nonlinearity and time variation in predictive relations is important for exchange rate prediction (Kilian and Taylor (2003), Clarida et al. (2003)), but the existing literature has typically been confined to a small number of fundamentals and limited forms of nonlinearity. Our framework combines ten country-level fundamentals (i.e., characteristics), six global state variables, and their interactions in a high-dimensional panel predictive regression, which we estimate using machine-learning techniques to guard against overfitting in our noisy data environment.

Our predictor set captures a wide range of channels through which fundamentals can move exchange rates. The ten country characteristics, measured relative to the United States, include inflation rates, unemployment gaps, interest rates at three maturities, equity market valuation ratios, and currency-specific volatility and skewness; each is grounded in a well-known theory of exchange rate determination and is observable to foreign exchange (FX) market participants in real time. The six global state variables comprise indices of economic policy uncertainty, monetary policy uncertainty, and geopolitical risk, together with measures of global FX market volatility, illiquidity, and correlation. We consider numerous country characteristics and global state variables because we cannot know a priori which ones matter most for exchange rate prediction. We then interact every country characteristic with every global state variable, yielding 70 predictors. This lets the predictive effect of each fundamental vary smoothly with the state of the global financial system—something prior work, which typically analyzes nonlinearity within a single fundamental at a time, has not been able to do.

After reserving the first ten years of data for an initial training sample, we generate monthly out-of-sample forecasts of exchange rate changes by recursively re-estimating the prediction models each month. The out-of-sample period runs from 1995:01 to 2020:09 (2000:02 to 2020:09) for the United Kingdom, Switzerland, Japan, Canada, Australia, New Zealand, Sweden, Norway, and Denmark (Euro area). Due to their adoption of the euro, data for Germany, Italy, France, and the Netherlands end in 1998:12 and are excluded from the out-of-sample analysis (although we use them for training the prediction models).

We estimate the baseline panel predictive regression—which includes the ten country characteristics and their 60 interactions with the six global state variables—via the elastic net (ENet, Zou and Hastie (2005)). The ENet guards against overfitting by shrinking coefficients toward zero and performs variable selection by shrinking some coefficients to exactly zero. The interaction terms generate *continuous regime switching*: the predictive effect of each country characteristic varies linearly with each global state variable. To allow for richer nonlinearity, we also fit a deep neural network (DNN) on the same 70 predictors and report an ensemble forecast that averages the ENet and DNN forecasts to harness complementarities between the approaches.[1]

The machine-learning forecasts consistently and significantly outperform the no-change benchmark. Out-of-sample gains are particularly evident when financial distress and market turbulence increase sharply, including during the Global Financial Crisis. Based on the out-of-sample R^2 statistic (Fama and French (1989), Campbell and Thompson (2008)), which measures the proportional decrease in mean squared prediction error (MSPE) for a competing forecast vis-à-vis a benchmark, the gains are large in the context of the existing literature. The improvements are statistically significant in nearly all cases according to the Clark and West (2007) test. For the ensemble forecast, out-of-sample R^2 statistics are above 1% (2%) for eight (four) countries, reach 3.41% for the United Kingdom, and produce a composite out-of-sample R^2 of 1.86% across the ten countries (significant at the 1% level). Forecast accuracy also rises significantly with global FX volatility, consistent with a tightening of the link between fundamentals and exchange rates when FX markets become more turbulent.

Beyond predictive performance itself, we investigate the economic drivers underlying the improvements in forecast accuracy. Using the recursive ENet coefficient estimates and the Greenwell, Boehmke, and McCarthy (2018) variable-importance measure based on partial dependence plots (Friedman (2001)), we identify a core set of predictors that consistently stand out in the fitted prediction models. Two interactions involving global FX volatility are particularly important. The first is the bill yield differential interacted with global FX volatility: the relation

[1] While machine learning is increasingly common in finance (e.g., Rapach, Strauss, and Zhou (2013), Gu, Kelly, and Xiu (2020), Avramov, Cheng, and Metzker (2023), Chen, Pelger, and Zhu (2024)), to our knowledge, our paper is the first to combine the ENet, deep learning, and ensemble forecasts for out-of-sample exchange rate prediction.

between the bill yield differential and expected US dollar appreciation posited by uncovered interest parity (UIP) strengthens as global FX volatility rises. The second is the inflation differential interacted with global FX volatility: the relation posited by relative purchasing power parity (PPP) likewise strengthens with global FX volatility. Both interactions point to a “flight to fundamentals” in periods of increasing FX market stress, consistent with Clarida, Davis, and Pedersen (2009) and Hsu, Taylor, Wang, and Li (2024), (2025), who document that carry-trade profitability varies substantially over time and declines markedly when global FX volatility rises. Other predictors that contribute meaningfully include the dividend yield differential and the interaction between the stock market time-series momentum differential and geopolitical risk. Taken together, these findings indicate that the link between fundamentals and exchange rates is conditional, tightening as global financial conditions deteriorate.

Next, we analyze the economic value of the machine-learning forecasts from an investment perspective. We consider a mean-variance US investor with a coefficient of relative risk aversion of five who allocates across the US dollar and all available foreign currencies, computing each foreign currency excess return forecast as the bill yield differential minus the expected US dollar appreciation. We label the resulting portfolio the ML-Opt (“machine-learning optimized”) portfolio when the investor uses a machine-learning forecast of US dollar appreciation, and the Bench-Opt (“benchmark-optimized”) portfolio when the investor uses the no-change benchmark forecast. In the latter case, because the mean-variance investor uses the bill yield differential alone to forecast the currency excess return, the Bench-Opt portfolio can be viewed as an optimal carry-trade portfolio.[2]

[2] A conventional carry-trade portfolio underperforms the Bench-Opt portfolio, while the optimal carry-trade portfolio of Daniel, Hodrick, and Lu (2017) delivers similar performance to our Bench-Opt portfolio.

Across the different machine-learning forecasts, the ML-Opt portfolios deliver increases in annualized average utility gains of 198 (DNN) to 450 (baseline ENet) basis points over the Bench-Opt portfolio. The gains are present throughout the sample but become especially large starting in late 2008, when the Bench-Opt portfolio's performance sharply deteriorates—a pattern that tracks the well-documented decline of conventional carry-trade strategies in the wake of the crisis (e.g., Melvin and Taylor (2009), Burnside, Eichenbaum, and Rebelo (2011b), Lustig, Roussanov, and Verdelhan (2011), Jordà and Taylor (2012), Daniel et al. (2017), Melvin and Shand (2017)). Importantly, the ML-Opt portfolios continue to generate sizable, statistically significant alphas in the currency factor model of Lustig et al. (2011) both before and after the crisis, even as their carry-factor exposures collapse post-2008—direct evidence that the prediction models pick up time-varying risk exposures.[3]

Our work on resolving the exchange rate predictability puzzle contributes to a time-series forecasting literature that runs parallel to a more recent cross-sectional literature studying exchange rates and fundamentals (e.g., Lustig et al. (2011), Colacito, Riddiough, and Sarno (2020), Dahlquist and Hasseltoft (2020), Dahlquist and Pénasse (2022), Filippou and Taylor (2023)). These cross-sectional studies construct long-short currency portfolios by sorting on lagged fundamentals—such as interest rate differentials, output gaps, and Taylor (1993) rule variables—to assess their pricing relevance. In contrast to cross-sectional studies, we generate monthly out-of-sample point forecasts of exchange rate changes, directly analyzing their forecast accuracy as well as their economic value via our ML-Opt and Bench-Opt portfolios. In sum, we

[3]Other portfolio strategies for investing in foreign currencies, some of which are based on mean-variance optimization, include Della Corte, Sarno, and Tsiakas (2009), Jordà and Taylor (2012), Sager and Taylor (2014), Chernov, Dahlquist, and Lochstoer (2023), and Della Corte, Jeanneret, and Patelli (2023).

go beyond the ordinal information in currency sorts by using point forecasts of exchange rate changes, which incorporate information about both the sign and magnitude of expected movements in a time-series forecasting context to address the exchange rate predictability puzzle. The relatively small number of available currencies also limits the number of fundamentals that can be analyzed in a sort-based framework. Our predictive regression approach instead accommodates a large number of characteristics and their interactions with global state variables.

The remainder of the paper is organized as follows. Section II describes the data. Section III discusses model specification and construction of the machine-learning forecasts. Section IV presents results on out-of-sample forecast accuracy. Section V analyzes the economic drivers of the forecasts. Section VI constructs the mean-variance optimal portfolios and measures the economic value of the machine-learning forecasts. Section VII concludes.

II. Data

This section describes the data used in our analysis.

Exchange Rates. We begin with daily exchange rate data from Barclays and Reuters via Datastream. We convert daily spot exchange rates to a monthly frequency using end-of-month values (e.g., Burnside, Eichenbaum, Kleshchelski, and Rebelo (2011a), Lustig et al. (2011)). Our sample consists of currencies for the Euro area and 13 countries: the United Kingdom, Switzerland, Japan, Canada, Australia, New Zealand, Sweden, Norway, Denmark, Germany, Italy, France, and the Netherlands. Our universe of exchange rates is similar to the sample of developed countries employed in other studies (e.g., Lustig et al. (2011), Menkhoff, Sarno, Schmeling, and

Schrimpf (2012a)). The Euro area replaces Germany, Italy, France, and the Netherlands after the Euro's introduction in 1999. We use $S_{i,t}$ to denote the month-t spot exchange rate for country i, expressed as the number of country-i currency units per US dollar (e.g., Lustig et al. (2011), Menkhoff et al. (2012a), Menkhoff, Sarno, Schmeling, and Schrimpf (2012b)). An increase in $S_{i,t}$ thus represents an appreciation of the US dollar. The country-i log exchange rate change is denoted by $\Delta s_{i,t} = s_{i,t} - s_{i,t-1}$, where $s_{i,t} = \ln(S_{i,t})$. Except for four countries, the sample ends in 2020:09; for France, Germany, Italy, and the Netherlands, the sample ends in 1998:12, the last month for which these countries had their own currencies. Based on data availability for the predictors, the sample begins in 1985:01 for all countries, except for the Euro area, where the sample begins in 1999:02, corresponding to the introduction of the Euro in January of 1999.[4]

Country Characteristics. We consider a large set of country characteristics based on a variety of macroeconomic and financial variables, all of which are based on data available to FX market participants. Section A.I of the Appendix provides more detailed descriptions of the predictors and the motivation for their inclusion. Specifically, we consider ten monthly country characteristics: (1) *inflation differential* ($\text{INF}_{i,t}$); (2) *unemployment rate gap differential* ($\text{UN}_{i,t}$); (3) *bill yield differential* ($\text{BILL}_{i,t}$); (4) *note yield differential* ($\text{NOTE}_{i,t}$); (5) *bond yield differential* ($\text{BOND}_{i,t}$); (6) *dividend yield differential* ($\text{DP}_{i,t}$); (7) *price-earnings ratio differential* ($\text{PE}_{i,t}$); (8) *stock market time-series momentum differential* ($\text{SRET12}_{i,t}$); (9) *idiosyncratic volatility* ($\text{IV}_{i,t}$); (10) *idiosyncratic skewness* ($\text{IS}_{i,t}$). All the differentials are measured as the country-i value minus the US value. IV and IS are computed using fitted residuals for the Lustig et al. (2011) two-factor

[4]Table IA.1 in the Internet Appendix reports summary statistics for the log exchange rate changes, which show well-known empirical features such as near-zero means, sizable volatilities, and low autocorrelations.

model estimated using daily data for month t for country-i log currency excess returns. The characteristics are based on popular predictors from the literature that relate to well-known theories, such as relative PPP, Taylor-rule fundamentals, UIP, and uncovered equity parity (UEP).

Global State Variables. Given that we are generating short-horizon forecasts, allowing for state variables that may influence short-term market sentiment is important. We term these global variables, although the first two relate mainly to the United States, because they all likely affect US dollar exchange rates. We consider six monthly global variables: (1) *economic policy uncertainty* (EPU_t, Baker, Bloom, and Davis (2016)); (2) *monetary policy uncertainty* (MPU_t, Baker et al. (2016)); (3) *geopolitical risk* (GR_t, Caldara and Iacoviello (2022)); (4) *global FX volatility* (GVOL_t, Menkhoff et al. (2012a)); (5) *global FX illiquidity* (GILL_t, Menkhoff et al. (2012a)); (6) *global FX correlation* (GCOR_t, Mueller, Stathopoulos, and Vedolin (2017)).[5] Each captures general conditions that can affect the predictive ability of the country characteristics.

III. Panel Predictive Regressions

In this section, we specify the baseline panel predictive regression and DNN models used to construct the out-of-sample forecasts and the ensemble forecast.

[5]We follow Menkhoff et al. (2012a) and Mueller et al. (2017) by measuring GVOL, GILL, and GCOR as the residuals from fitted first-order autoregressive processes. To avoid look-ahead bias, we only use data available at the time of forecast formation when fitting the autoregressive processes and computing the residuals.

A. Baseline Specification

We denote the collection of characteristics in Section II for country i and month t by $\{z_{j,i,t}\}_{j=1}^{J}$, where $J = 10$, for $i = 1, \ldots, N$, and $t = 1, \ldots, T$.[6] The collection of global state variables in Section II is denoted by $\{g_{k,t}\}_{k=1}^{K}$, where $K = 6$. We interact each country characteristic with the set of global variables, yielding the set $\{\{z_{j,i,t}\, g_{k,t}\}_{k=1}^{K}\}_{j=1}^{J}$. After standardizing all the predictors, our baseline panel predictive regression is given by

$$(1) \qquad \Delta s_{i,t+1} = \sum_{j=1}^{J} \beta_j^z \underbrace{\left[\frac{1}{\sigma_{j,i}^z}\left(z_{j,i,t} - \mu_{j,i}^z\right)\right]}_{\tilde{z}_{j,i,t}} + \sum_{j=1}^{J}\sum_{k=1}^{K} \beta_{j,k}^{z,g} \underbrace{\left[\frac{1}{\sigma_{j,i,k}^{z,g}}\left(z_{j,i,t}\, g_{k,t} - \mu_{j,i,k}^{z,g}\right)\right]}_{\widetilde{z_{j,i,t}\, g_{k,t}}} + \varepsilon_{i,t+1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T-1$, where $\tilde{z}_{j,i,t}$ $(\widetilde{z_{j,i,t}\, g_{k,t}})$ is the standardized country characteristic (interaction term), $\mu_{j,i}^z$ $(\mu_{j,i,k}^{z,g})$ is the mean of $z_{j,i,t}$ $(z_{j,i,t}\, g_{k,t})$, $\sigma_{j,i}^z$ $(\sigma_{j,i,k}^{z,g})$ is the standard deviation of $z_{j,i,t}$ $(z_{j,i,t}\, g_{k,t})$, and $\varepsilon_{i,t+1}$ is a zero-mean disturbance term.

Equation (1) is high dimensional, as it includes $J(K+1) = 10 \times 7 = 70$ predictors. We include many more predictors than existing studies of exchange rate predictability, which typically consider only a limited number of country characteristics. The inclusion of a broad set of characteristics is germane, since we do not know a priori which are relevant for forecasting exchange rates. It also helps to guard against data mining or snooping, as we include an extensive set of potentially relevant fundamentals in the model from the start. In addition to numerous country characteristics, we include their interactions with the set of global variables, which allows for the predictive relations between exchange rate changes and fundamentals to vary with global

[6] To simplify the notation, we assume that the panel of data is balanced. For the application in Section IV, the panel is unbalanced. It is straightforward to modify the notation for an unbalanced panel.

economic conditions (as detailed below). To our knowledge, such interactions have not previously been considered in the out-of-sample exchange rate predictability literature.

Equation (1) also pools the data, thereby restricting the response of $\Delta s_{i,t+1}$ to $\tilde{z}_{j,i,t}$ and $\widetilde{z_{j,i,t}\, g_{k,t}}$ to be the same across i. These slope homogeneity restrictions reduce the parameter space to help guard against overfitting. While the slope homogeneity restrictions are unlikely to hold exactly—potentially inducing bias in the parameter estimates—they also decrease the variance of the parameter estimates, which can improve out-of-sample performance in light of the bias-variance trade-off. Whether pooling is beneficial is ultimately an empirical issue determined by the bias-variance trade-off; for our application, we find that pooling leads to better time-series forecasts of individual exchange rates.[7]

Furthermore, observe that equation (1) does not include an intercept term. Because the predictors are in deviation form, we effectively impose the restrictions that the country-specific means for $\Delta s_{i,t}$ for $i = 1, \ldots, N$ are zero. In other words, equation (1) is tantamount to a fixed-effects specification in which the country-specific means for the log exchange rate changes are all zero. These zero restrictions are consistent with economic intuition and the data, and they further reduce the number of parameters that we need to estimate, again helping to guard against overfitting in light of the bias-variance trade-off.

Regime Switching. The inclusion of the interaction terms in equation (1) effectively permits a form of regime switching in the predictive relations between the country characteristics and exchange rate changes. To see this, note that the response of $\Delta s_{i,t+1}$ to a change in $z_{j,i,t}$ in

[7]Viewed another way, pooling substantially increases the estimation sample size from $T-1$ to $N(T-1)$, thereby reducing the variance of the parameter estimates.

equation (1) is given by

$$(2) \qquad \frac{\partial \Delta s_{i,t+1}}{\partial z_{j,i,t}} = \underbrace{\beta_j^z \, \frac{1}{\sigma_{j,i}^z} + \sum_{k=1}^{K} \beta_{j,k}^{z,g} \, \frac{1}{\sigma_{j,i,k}^{z,g}} \, g_{k,t}}_{\text{slope coefficient for } z_{j,i,t}}$$

for $j = 1, \ldots, J$ and $i = 1, \ldots, N$. According to equation (2), the slope coefficient for $z_{j,i,t}$ varies linearly with $g_{k,t}$ for $k = 1, \ldots, K$. For example, if $\beta_{j,k}^{z,g} > 0$, then the response of $\Delta s_{i,t+1}$ to a change in $z_{j,i,t}$ increases linearly with $g_{k,t}$. We refer to equation (1) and the concomitant response in equation (2) as encapsulating *continuous regime switching*. Setting $\beta_{j,k}^{z,g} = 0$ for $j = 1, \ldots, J$ and $k = 1, \ldots, K$ in equation (1) (i.e., excluding the interaction terms) produces a panel predictive regression based on the country characteristics alone:

$$(3) \qquad \Delta s_{i,t+1} = \sum_{j=1}^{J} \beta_j^z \, \tilde{z}_{j,i,t} + \varepsilon_{i,t+1},$$

where the number of predictors is reduced to $J = 10$. For equation (3), the derivative in equation (2) simplifies to the first term on the right-hand-side ($\beta_j^z \, (1/\sigma_{j,i}^z)$), so regime switching is absent.

Elastic Net. An out-of-sample forecast of $\Delta s_{i,T+1}$ based on the baseline panel predictive regression in equation (1) and data available through T is given by

$$(4) \qquad \widehat{\Delta s}_{i,T+1|T} = \sum_{j=1}^{J} \hat{\beta}_j^z \, \tilde{z}_{j,i,T} + \sum_{j=1}^{J} \sum_{k=1}^{K} \hat{\beta}_{j,k}^{z,g} \, \widetilde{z_{j,i,T} \, g_{k,T}},$$

where $\hat{\beta}_j^z$ and $\hat{\beta}_{j,k}^{z,g}$ are estimates of β_j^z and $\beta_{j,k}^{z,g}$, respectively, in equation (1) for $j = 1, \ldots, J$ and $k = 1, \ldots, K$ based on data available through T. The plugged-in predictors on the right-hand-side of equation (4), $\tilde{z}_{j,i,T}$ and $\widetilde{z_{j,i,T}\, g_{k,T}}$, are standardized using data through T. By estimating the slope coefficients and standardizing the predictors using data available through T (i.e., the time of forecast formation), we avoid "look-ahead" bias in the forecast.

Even after imposing the slope homogeneity and zero-mean restrictions in equation (1), because the panel predictive regression is high dimensional and the unpredictable component in monthly exchange rate changes is inherently large, conventional ordinary least squares (OLS) estimation of the parameters β_j^z and $\beta_{j,k}^{z,g}$ for $j = 1, \ldots, J$ and $k = 1, \ldots, K$ is highly susceptible to overfitting. Thus, we consider penalized (or regularized) regression, which shrinks the parameter estimates toward zero to reduce their variance and improve out-of-sample performance according to the bias-variance trade-off.

We employ penalized regression via the ENet (Zou and Hastie (2005)), a popular and powerful machine-learning technique that mitigates overfitting by including a penalty term in the objective function for estimating equation (1). Collecting the parameters in the vector $\underset{J(K+1)}{\boldsymbol{\beta}} := \left[\begin{array}{cccccccccc} \beta_1^z & \cdots & \beta_J^z & \beta_{1,1}^{z,g} & \cdots & \beta_{1,K}^{z,g} & \cdots & \beta_{J,1}^{z,g} & \cdots & \beta_{J,K}^{z,g} \end{array}\right]'$, the ENet objective function can be expressed as

$$\begin{aligned} \underset{\boldsymbol{\beta} \in \mathbb{R}^{J(K+1)}}{\arg\min}\ & \frac{1}{2N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \left[\Delta s_{i,t+1} - \left(\sum_{j=1}^{J} \beta_j^z\, \tilde{z}_{j,i,t} + \sum_{j=1}^{J} \sum_{k=1}^{K} \beta_{j,k}^{z,g}\, \widetilde{z_{j,i,t}\, g_{k,t}} \right) \right]^2 + \\ & \underbrace{\lambda \left[0.5(1-\delta) \|\boldsymbol{\beta}\|_2^2 + \delta \|\boldsymbol{\beta}\|_1 \right]}_{\text{penalty term}}, \end{aligned} \tag{5}$$

where $\lambda \geq 0$ is a regularization hyperparameter that governs the degree of shrinkage; $\|\cdot\|_1$ and $\|\cdot\|_2$ are the ℓ_1 and ℓ_2 norms, respectively; and δ is a blending hyperparameter for the ℓ_1 and ℓ_2

components in the penalty term.[8] When $\lambda = 0$, there is no shrinkage, so the ENet and OLS objective functions coincide; when $\lambda > 0$, the penalty term shrinks the parameter estimates toward zero to mitigate overfitting. The ENet penalty term in equation (5) includes both ℓ_1 (Least Absolute Shrinkage and Selection Operator or LASSO, Tibshirani (1996)) and ℓ_2 (ridge, Hoerl and Kennard (1970)) components. The former permits shrinkage to zero (for sufficiently large λ), so the ENet performs variable selection (i.e., can produce a sparse model).[9] We tune the hyperparameters λ and δ via the Extended Regularized Information Criterion (ERIC, Hui, Warton, and Foster (2015)), which is a relatively stringent information criterion, because it tends to induce more shrinkage than cross-validation and other information criteria.[10] Considering grids of values for λ and δ, we select the values that minimize the ERIC. Given our high-dimensional and noisy data environment, the ERIC appears effective for tuning the hyperparameters.

B. Deep Neural Network

The baseline specification in equation (1) permits nonlinearities in the predictors through the interaction terms. To allow for additional nonlinearities, we consider the following general

[8]Again, we only use data at the time of forecast formation when standardizing the predictors in equation (5), so there is no look-ahead bias.

[9]Ridge regression sets $\delta = 0$ in equation (5), so it does not shrink coefficient estimates to precisely zero, which results in a large number of predictors having relatively small effects, while the ENet corresponds to a sparse model. The ENet provides better overall out-of-sample performance for our application than ridge regression.

[10]The ERIC is given by $N(T-1)\log\big(\text{SSR}_{\lambda,\delta}/[N(T-1)]\big)+\text{df}_{\lambda,\delta}\log\big([N(T-1)\hat{\sigma}^2_{\lambda,\delta}]/\lambda\big)$, where $\text{SSR}_{\lambda,\delta}$ ($\text{df}_{\lambda,\delta}$) is the sum of squared residuals (effective degrees of freedom) for the ENet-fitted model based on λ and δ, and $\hat{\sigma}^2_{\lambda,\delta} = \text{SSR}_{\lambda,\delta}/[N(T-1)]$.

panel predictive regression:

$$\Delta s_{i,t+1} = f(\tilde{\boldsymbol{x}}_{i,t}) + \varepsilon_{i,t+1} \tag{6}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T-1$, where $\underset{J(K+1)}{\tilde{\boldsymbol{x}}_{i,t}} := [\; \tilde{z}_{1,i,t} \;\; \cdots \;\; \tilde{z}_{J,i,t} \;\; \widetilde{z_{1,i,t}\, g_{1,t}} \;\; \cdots \;\; \widetilde{z_{1,i,t}\, g_{K,t}} \;\; \cdots \;\; \widetilde{z_{J,i,t}\, g_{1,t}} \;\; \cdots \;\; \widetilde{z_{J,i,t}\, g_{K,t}} \;]'$ and $f(\tilde{\boldsymbol{x}}_{i,t})$ is a generic function constituting the conditional expectation function. We model $f(\tilde{\boldsymbol{x}}_{i,t})$ in equation (6) via a DNN, a popular machine-learning device for flexibly modeling complex relations between predictors and a target variable. A feedforward neural network is composed of multiple layers. The first is the input layer, which is simply the set of predictors. One or more hidden layers are next, where each hidden layer contains multiple neurons with a nonlinear activation function. The input layer is densely connected to the first hidden layer, and each subsequent hidden layer is densely connected to its preceding hidden layer. The neurons produce signals that are passed forward through the hidden layers. Finally, an output layer translates the signals from the last hidden layer into a prediction.

It is well known that a neural network with a single hidden layer and a sufficiently large number of neurons can, theoretically, approximate any smooth function (e.g., Hornik, Stinchcombe, and White (1989)). However, neural networks with multiple hidden layers (i.e., DNNs) and a more limited number of neurons in each layer often perform better in practice than neural networks with one hidden layer and a large number of neurons (e.g., Goodfellow, Bengio, and Courville (2016), Gu et al. (2020)). Our DNN has four hidden layers of 16, 8, 4, and 2 neurons, respectively, with dense connections between adjacent layers and a single output node producing the exchange rate change forecast. The specification of 16 neurons in the first hidden layer is a compromise between two popular rules of thumb: half or the square root of the number of predictors. Section IA.2 of the Internet Appendix provides further details on the architecture, activation function, and estimation strategy, including the steps that we take to guard against overfitting in our noisy data environment. As in the baseline panel predictive regression in equation (1), the DNN predicts a value of zero for the exchange rate change when all of the (standardized) predictors are at their mean value of zero.

C. Ensemble

We also compute an ensemble forecast by taking the average of the ENet and DNN forecasts, which provides a convenient strategy for blending the different forecast strategies. An average of multiple individual forecasts is known as a "voting ensemble" in the machine-learning literature. An ensemble forecast recognizes that we do not know the best individual model a priori and allows us to take advantage of potential complementarities between models.

IV. Out-of-Sample Performance

In this section, we analyze the accuracy of the out-of-sample machine-learning forecasts and their ability to resolve the short-horizon exchange rate predictability puzzle.

A. Forecast Construction

Mimicking the situation of a forecaster in real time, we generate exchange rate forecasts for the 1995:01 to 2020:09 out-of-sample period as follows. Reserving the first ten years of data (1985:01 to 1994:12) for the initial training sample, we estimate the baseline panel predictive regression in equation (1) and the DNN in Section III.B using data available through 1994:12. We then use the fitted prediction models and 1994:12 predictor values for each country to compute forecasts for 1995:01. Next, we re-estimate the models using data available through 1995:01 and use the fitted models and 1995:01 predictor values for each country to generate forecasts for 1995:02. We continue in this fashion through the end of the out-of-sample period, providing us with a set of exchange rate forecasts for the available countries for each of the 309 months comprising the out-of-sample period. By retraining the prediction models each month as new data become available, we update the fitted models in a timely manner; similarly, we also regularly retune the hyperparameters.[11] Note that there is no look-ahead bias in the forecasts, as we only use data available at the time of forecast formation when training the models (and standardizing the predictors).

For Germany, Italy, France, and the Netherlands, there are only 48 monthly forecasts (1995:01 to 1998:12) available for evaluation, due to these countries joining the Euro area in 1999:01. Thus, we do not include these four countries when evaluating forecast performance. However, we use data for these countries when training the prediction models, as this increases the number of training-sample observations, thereby improving estimation efficiency. After imposing a minimum requirement of twelve monthly observations before a currency is included in the panel predictive regression, there are 248 forecasts (2000:02 to 2020:09) available for the Euro area. For the remaining nine countries, forecasts are available for the entire 1995:01 to 2020:09 out-of-sample period (309 observations). In addition to reporting results for the individual countries, we report results for the entire collection of countries taken together (excluding Germany, Italy, France, and the Netherlands, resulting in $248 + 9 \times 309 = 3{,}029$ observations).

[11] See Sections IA.2 and IA.4.A of the Internet Appendix for details.

B. Test Statistics

We compare competing forecasts that incorporate information from economic fundamentals to the no-change benchmark, the most stringent benchmark for out-of-sample exchange rate prediction (Rossi (2013)). For country i, we can conveniently compare the relative accuracy of a competing forecast vis-à-vis the no-change benchmark in terms of MSPE in the time-series dimension using the out-of-sample R^2 statistic (Fama and French (1989), Campbell and Thompson (2008)):

$$R^2_{i,\text{OS}} = 1 - \frac{\sum_{t=T_1+1}^{T_2} \left(\hat{e}^{\text{Compete}}_{i,t|t-1}\right)^2}{\sum_{t=T_1+1}^{T_2} \left(\hat{e}^{\text{Bench}}_{i,t|t-1}\right)^2}, \tag{7}$$

where $\hat{e}^{j}_{i,t|t-1} = \Delta s_{i,t} - \widehat{\Delta s}^{j}_{i,t|t-1}$ for $j = \text{Bench}, \text{Compete}$; $\widehat{\Delta s}^{\text{Bench}}_{i,t|t-1} = 0$ is the no-change benchmark forecast; $\widehat{\Delta s}^{\text{Compete}}_{i,t|t-1}$ is a competing forecast; T_1 is the last observation for the initial in-sample period; and T_2 is the last available observation for the out-of-sample period. The $R^2_{i,\text{OS}}$ statistic measures the proportional reduction in MSPE for a competing forecast vis-à-vis the benchmark. Because the predictable component in monthly exchange rate changes is intrinsically small, the $R^2_{i,\text{OS}}$ statistic will necessarily be small. Nevertheless, even a seemingly small degree of monthly exchange rate predictability can be economically meaningful, as we show in Section VI and as is the case for stock return predictability (Campbell and Thompson (2008)).

To test whether the competing forecast provides a statistically significant improvement in MSPE relative to the benchmark when comparing nested models, we compute the Clark and West (2007) adjusted version of the popular Diebold and Mariano (1995) and West (1996) statistic. Specifically, we use the Clark and West (2007) statistic to test (in population) H_0:

$\text{MSPE}^{\text{Bench}} \leq \text{MSPE}^{\text{Compete}}$ ($R^2_{i,\text{OS}} \leq 0$) against H_A: $\text{MSPE}^{\text{Bench}} > \text{MSPE}^{\text{Compete}}$ ($R^2_{i,\text{OS}} > 0$). For country i, the Clark and West (2007) statistic can be conveniently computed using the t-statistic corresponding to the intercept term ($a_{0,i}$) in the following regression:

$$\text{(8)} \qquad \underbrace{\left(\hat{e}^{\text{Bench}}_{i,t|t-1}\right)^2 - \left(\hat{e}^{\text{Compete}}_{i,t|t-1}\right)^2 + \left(\widehat{\Delta s}^{\text{Bench}}_{i,t|t-1} - \widehat{\Delta s}^{\text{Compete}}_{i,t|t-1}\right)^2}_{l_{i,t}} = a_{0,i} + \varepsilon_{i,t}$$

for $t = T_1 + 1, \ldots, T_2$, where $l_{i,t}$ is the adjusted loss differential. Clark and West (2007) show that the t-statistic corresponding to $a_{0,i}$ in equation (8) is well approximated by the standard normal distribution under the null hypothesis.

We also compute a composite out-of-sample R^2 statistic based on the entire collection of countries taken together:

$$\text{(9)} \qquad R^2_{\text{All,OS}} = 1 - \frac{\sum_{i=1}^{N} \sum_{t=T_1+1}^{T_2} \left(\hat{e}^{\text{Compete}}_{i,t|t-1}\right)^2}{\sum_{i=1}^{N} \sum_{t=T_1+1}^{T_2} \left(\hat{e}^{\text{Bench}}_{i,t|t-1}\right)^2}.$$

For testing H_0: $R^2_{\text{All,OS}} \leq 0$ against H_A: $R^2_{\text{All,OS}} > 0$, we compute a pooled version of the Clark and West (2007) statistic by calculating the t-statistic corresponding to the intercept term (a_0) in the following regression:

$$\text{(10)} \qquad l_{i,t} = a_0 + \varepsilon_{i,t}$$

for $i = 1, \ldots, N$ and $t = T_1 + 1, \ldots, T_2$, where we cluster the standard errors by country and time. A caveat to the $R^2_{\text{All,OS}}$ statistic is that it can reflect cross-sectional as well as time-series predictability, while, in line with the exchange rate predictability puzzle, we are interested in the latter. With this caveat in mind, we report the $R^2_{\text{All,OS}}$ statistic in equation (9) to provide an overall measure of the accuracy of the exchange rate forecasts for the entire group of countries taken together.

C. Forecast Accuracy: Outperforming the Random Walk

Table 1 reports the $R^2_{i,\text{OS}}$ statistic in equation (7) for different forecasts and the ten countries for which we have at least 248 monthly observations for the forecast evaluation period. It also reports the $R^2_{\text{All,OS}}$ statistic in equation (9) for the ten countries taken together. The fourth and fifth columns of the table report results for forecasts based on the baseline panel predictive regression in equation (1) estimated via OLS ("baseline OLS") and the ENet ("baseline ENet"), respectively, while the sixth column provides results for the panel predictive regression without interaction terms in equation (3) estimated via the ENet ("no-interaction ENet"). Finally, the seventh and eighth columns report results for the DNN in Section III.B and the ensemble forecast in Section III.C (an average of the baseline ENet and DNN forecasts), respectively.[12]

[Insert Table 1 approximately here]

The $R^2_{i,\text{OS}}$ and $R^2_{\text{All,OS}}$ statistics are all negative in the fourth column of Table 1, so the baseline OLS forecast always fails to outperform the no-change benchmark. The negative $R^2_{i,\text{OS}}$ statistics are often quite sizable in magnitude for the individual countries, and the $R^2_{\text{All,OS}}$ statistic is -8.74% for the ten countries taken together in the last row. Its poor out-of-sample performance indicates that the baseline OLS forecast substantially overfits the data, which is not surprising in our high-dimensional and noisy data environment.

[12]Figures IA.1 to IA.4 in the Internet Appendix depict the baseline OLS, baseline ENet, DNN, and ensemble forecasts, respectively.

As shown in the fifth column of Table 1, employing shrinkage to mitigate overfitting proves highly beneficial, as the baseline ENet forecast evinces markedly better out-of-sample performance than the baseline OLS forecast. Specifically, the $R^2_{i,\text{OS}}$ statistics are positive for eight of the ten countries for the baseline ENet forecast, and the reduction in MSPE relative to the no-change benchmark is significant at conventional levels for seven of the eight countries. Seven (four) of the $R^2_{i,\text{OS}}$ statistics are above 1% (2%) and reach as high as 3.24% (United Kingdom), so they are sizable in the context of the literature on out-of-sample exchange rate predictability. For the complete set of countries taken together, the $R^2_{\text{All,OS}}$ statistic is 1.51% (significant at the 5% level).

The sixth column of Table 1 reports results for the no-interaction ENet forecast. This forecast is based on the panel predictive regression without interaction terms in equation (3), which, as discussed in Section III.A, is akin to restricting the slope coefficients for the country characteristics to be constant (i.e., no regime switching). The $R^2_{i,\text{OS}}$ statistics are positive for five of the ten countries, and all five are significant at conventional levels. Taking the ten countries together, the $R^2_{\text{All,OS}}$ is negative (-0.29%). Overall, there is some evidence that using country characteristics without regime switching outperforms the no-change benchmark. However, comparing the results in the fifth and sixth columns of Table 1, the $R^2_{i,\text{OS}}$ and $R^2_{\text{All,OS}}$ statistics are all larger for the baseline ENet forecast vis-à-vis the no-interactions ENet forecast, often by a substantial margin. Thus, we have strong evidence that allowing for regime switching in the responses of exchange rate changes to characteristics improves out-of-sample predictability.

According to the seventh column, the DNN forecast performs well. The $R^2_{i,\text{OS}}$ statistics are positive for all countries, with significant improvements in MSPE relative to the no-change benchmark for eight of them. Like those for the baseline ENet forecast, the $R^2_{i,\text{OS}}$ statistics for the individual countries are sizable, and eight (four) are above 1% (2%), with a maximum of 3.03% (United Kingdom). For the ten countries considered jointly, the $R^2_{\text{All,OS}}$ statistic is 1.74% for the DNN forecast (significant at the 1% level).

By combining the baseline ENet and DNN forecasts, the ensemble approach in the last column of Table 1 provides the best overall performance. The $R^2_{i,\text{OS}}$ statistics are positive for all ten countries, and the improvements in MSPE vis-à-vis the no-change benchmark are significant for eight of them. The $R^2_{i,\text{OS}}$ statistics are also often quite sizable, with eight (four) above 1% (2%), and the maximum is 3.41% (United Kingdom). For the ten countries taken together, the $R^2_{\text{All,OS}}$ statistic is 1.86% (significant at the 1% level). Note that for a number of countries, the $R^2_{i,\text{OS}}$ statistic for the ensemble forecast is larger than the highest corresponding statistic in the fifth and seventh columns, pointing to meaningful complementarities in the baseline ENet and DNN forecasts.[13]

[Insert Figure 1 approximately here]

To examine the performance of the baseline ENet, DNN, and ensemble forecasts over time, Figure 1 shows rolling 60-month $R^2_{i,\text{OS}}$ and $R^2_{\text{All,OS}}$ statistics for the baseline ENet, DNN, and ensemble forecasts (considered in turn). The horizontal axes in Figure 1 correspond to the end of the 60-month rolling window. Figure 1 demonstrates that our machine-learning approach consistently generates out-of-sample gains over time. Except for Japan, the clear majority of the rolling 60-month $R^2_{i,\text{OS}}$ statistics are positive for the individual countries, and the $R^2_{\text{All,OS}}$ statistics for the ensemble forecast are positive for nearly every 60-month window. The rolling $R^2_{i,\text{OS}}$ and $R^2_{\text{All,OS}}$ statistics are quite large for 60-month windows that include the nadir of the Global

[13]Table IA.4 in the Internet Appendix reports $R^2_{i,\text{OS}}$ statistics for the four countries that joined the Euro (Germany, Italy, France, and the Netherlands), for which only 48 monthly observations are available for forecast evaluation.

Financial Crisis in late 2008, with values above 6% in many cases. Substantive out-of-sample gains are also evident outside of the crisis. For example, for numerous countries, including Switzerland, Australia, Sweden, Norway, and Denmark, the $R^2_{i,\mathrm{OS}}$ statistics are relatively large for windows ending in the early 2000s. In addition, $R^2_{i,\mathrm{OS}}$ statistics are sizable for windows ending in the later part of the out-of-sample period for Switzerland, Japan, Sweden, Norway, and Denmark. Along this line, the $R^2_{i,\mathrm{OS}}$ statistics for 60-month windows ending around 2004 to 2006 are substantive for countries such as the United Kingdom, Canada, Australia, and New Zealand.[14] Overall, Table 1 and Figure 1 indicate that the information in our large set of predictors and machine-learning methods can be used to generate sizable and consistent improvements in out-of-sample forecast accuracy.

While our approach performs well when the US dollar is used as the base currency, which is consistent with its dominant role as the primary settlement currency in international trade and global financial markets, we find less predictive power when other currencies serve as the base (such as the Japanese yen and Swiss franc, as examined in Section IA.3 of the Internet Appendix). Understanding the economic forces underlying this asymmetry is an interesting avenue for future research.

Section A.II of the Appendix lists a variety of robustness checks regarding forecast accuracy (with detailed results reported in the Internet Appendix). Overall, alternative approaches do not perform as well as the baseline ENet, DNN, and ensemble forecasts in terms of forecast accuracy as reported in Table 1.

[14]Figure IA.5 in the Internet Appendix depicts rolling 60-month Clark and West (2007) statistics for the rolling $R^2_{i,\mathrm{OS}}$ and $R^2_{\mathrm{All,OS}}$ statistics in Figure 1. Many of the rolling Clark and West (2007) statistics are significant, including outside of the crisis, despite the substantial reduction in the number of observations available for testing.

D. Forecast Accuracy and Global FX Volatility

Figure 1 indicates that the machine-learning forecasts perform particularly well relative to the no-change benchmark during periods of increases in FX market turbulence, such as during the worst phase of the Global Financial Crisis in late 2008. To investigate the relation between out-of-sample exchange rate predictability and FX market volatility more formally, we estimate an augmented version of equation (8):

$$l_{i,t} = a_{i,0} + a_{i,1}\text{GVOL}_t + \varepsilon_{i,t} \tag{11}$$

for $i = 1, \ldots, N$ and $t = T_1 + 1, \ldots, T_2$, where $l_{i,t}$ is the Clark and West (2007) adjusted loss differential for country i for the no-change benchmark forecast vis-à-vis the competing forecast (baseline ENet, DNN or ensemble), and GVOL_t is our measure of global FX volatility. To test whether the relative accuracy of the competing forecast increases as global FX volatility rises, we test H_0: $a_{i,1} \leq 0$ against H_A: $a_{i,1} > 0$. We also consider a composite version of equation (11) for all the countries taken together, which is an augmented version of equation (10):

$$l_{i,t} = a_0 + a_1\text{GVOL}_t + \varepsilon_{i,t} \tag{12}$$

for $i = 1, \ldots, N$ and $t = T_1 + 1, \ldots, T_2$. To again test whether the relative accuracy of the competing forecast increases with global FX volatility, we test H_0: $a_1 \leq 0$ against H_A: $a_1 > 0$. We cluster the standard errors by country and time for equation (12).

[Insert Table 2 approximately here]

Table 2 reports estimates of $a_{i,1}$ and a_1 in equations (11) and (12), respectively. For the baseline ENet forecast in the fourth column, the coefficient estimates are positive for all countries except Japan; apart from Japan, nine of the estimates are significant, eight of them at the 1% level. The coefficient estimates are positive for all countries for both the DNN and ensemble forecasts in the fifth and sixth columns, respectively; six (nine) of the estimates are significant for the DNN (ensemble) forecast, with nine of the estimates significant at the 1% level for the ensemble forecast. Taking the ten countries together, the pooled coefficient estimates are positive and significant (at the 1% level) for the baseline ENet, DNN, and ensemble forecasts in the last row of Table 2. In sum, we find extensive evidence that the relative accuracy of the baseline ENet, DNN, and ensemble forecasts significantly increases as global FX volatility increases.

V. Model Interpretation

The previous section provides evidence that our baseline panel predictive regression approach based on a large set of economic fundamentals and machine learning significantly outperforms the random walk without drift or no-change exchange rate forecast at the monthly horizon. This section examines the relative importance of the underlying fundamentals in the fitted prediction models driving this result. We focus on ENet estimation of the baseline panel predictive regression, as the ENet enhances model interpretation by performing variable selection.

Figure 2 shows the full-sample (1985:01 to 2020:09) ENet estimates of the slope coefficients for the baseline panel predictive regression. Although our out-of-sample forecasts are based only on data available at the time of forecast formation, it is informative to examine the full-sample coefficient estimates to get a sense of the leading drivers of short-run exchange rate movements over the full sample.

[Insert Figure 2 approximately here]

We refer to the country characteristics and global variables by their abbreviations in Section II. We denote predictors that are the interaction between two variables with a period between them (e.g., the interaction between BILL and GVOL is BILL.GVOL). The predictors are standardized so that the magnitudes of the estimated slope coefficients are comparable across predictors. ENet estimation produces a sparse model in Figure 2: ten of the 70 predictors (14%) are selected for the full-sample estimation.[15] The following predictors are selected by the ENet: UN, DP, PE, INF.GVOL, UN.GVOL, BILL.GVOL, PE.EPU, PE.GILL, SRET12.GR, and IV.GCOR.

In terms of magnitude, the most important predictor in Figure 2 is BILL.GVOL, with a coefficient estimate of 0.17. The positive coefficient estimate implies that an increase in GVOL increases the expected appreciation of the US dollar for a given increase in BILL; see equation (2). Since an increase in BILL corresponds to a relative decrease in the US bill yield, the relation between expected US dollar appreciation and BILL posited by UIP becomes stronger as global FX volatility increases. The importance of BILL.GVOL in the fitted predictive regression helps to explain the performance of the carry trade over time. The carry trade effectively relies on UIP not holding; specifically, a positive BILL is not fully offset (or worse) by a subsequent US dollar appreciation. The positive and sizable coefficient estimate indicates that, in more quiescent times (i.e., when GVOL is relatively low), investors can profit from the carry trade; however, when GVOL spikes, investors are exposed to large losses. This aligns with Clarida et al. (2009), who provide evidence that carry-trade returns decline sharply when FX volatility is in its highest

[15] Figure IA.8 in the Internet Appendix depicts the full-sample OLS estimates of the slope coefficients, and a comparison of the OLS and ENet estimates shows that the latter induces substantial shrinkage.

quartile of observed values.[16] In sum, ENet estimation of the baseline panel predictive regression indicates that, with respect to BILL and UIP, economic fundamentals become more relevant for exchange rate prediction as GVOL increases, which can be interpreted as a “flight to fundamentals” during periods of rising FX market stress.

INF.GVOL is also selected by the ENet, and the coefficient estimate is fairly sizable (0.09). The positive coefficient estimate indicates that, as GVOL increases, the expected appreciation of the US dollar increases for a given increase in INF. As an increase in INF constitutes a relative decline in the US inflation rate, the relation between expected US dollar appreciation and INF predicted by relative PPP strengthens as GVOL increases. Similarly to BILL.GVOL, we find a flight to fundamentals as the FX market becomes more turbulent.

With respect to the coefficient estimates for the country characteristics by themselves (i.e., not interacted with global variables), the coefficient estimate is sizable for DP (0.14). The positive estimate is consistent with the direction posited by UEP: an increase in DP represents a relative decrease in the expected return on US stocks; thus, an increase in the expected US dollar appreciation is needed to bring the expected US stock return in line with the expected foreign return (when measured in the same currency). The positive coefficient estimate for DP is also consistent with the theoretical model of Hau and Rey (2006), which is based on portfolio rebalancing as investors limit their FX exposure. Although another equity market variable, PE, is selected by the ENet, its coefficient estimate is much smaller in magnitude (0.02) than that for

[16]It is also consistent with Hsu et al. (2024), (2025), who show that the use and profitability of carry-trade strategies varies markedly over our sample period.

DP.[17] UN is also selected by the ENet. Its coefficient estimate is positive, although it is limited in magnitude (0.06). According to the positive estimate, an increase in UN leads to an increase in the expected appreciation of the US dollar, which is in line with the prediction of Taylor-rule fundamentals linking monetary policy to exchange rate fluctuations and the empirical evidence in Molodtsova and Papell (2009) and Filippou and Taylor (2023).[18]

Finally, the ENet selects two additional predictors involving interactions, SRET12.GR and IV.GCOR, in Figure 2. The coefficient estimate for SRET12.GR is -0.13. SRET12 measures relative time-series momentum (Moskowitz, Ooi, and Pedersen (2012)) in equity markets, where an increase in SRET12 signals a relative decrease in the expected return on US stocks. Under UEP and/or the Hau and Rey (2006) model, this leads to an increase in the expected US dollar appreciation. The negative coefficient estimate for SRET12.GR indicates that UEP and/or the FX exposure effect become weaker as geopolitical risk increases, which can be interpreted as international equity markets becoming more segmented as GR increases. The coefficient estimate for IV.GCOR is also -0.13. An increase in IV is typically associated with increasing limits to arbitrage (e.g., Menkhoff et al. (2012b)), thereby making holding foreign currency more risky and leading to an expected US dollar depreciation in order to increase the compensation for bearing FX risk. The negative coefficient estimate for IV.GCOR implies that the expected US dollar depreciation—and thus the increase in compensation for bearing FX risk—for a given increase in IV increases in magnitude as GCOR increases, which is consistent with an increase in global FX correlation exacerbating the exposure to FX risk (e.g., Mueller et al. (2017)).

[17]Two predictors involving the interactions of PE with global variables are selected, but their coefficient estimates are limited in magnitude (0.06 and -0.06 for PE.EPU and PE.GILL, respectively).

[18]The ENet selects UN.GVOL, but the coefficient is quite small in magnitude (-0.03).

Overall, the full-sample ENet coefficient estimates in Figure 2 point to a variety of economic fundamentals, including their interactions with global variables, as relevant predictors of exchange rate changes. We interpret the coefficient estimates for the predictors selected by the ENet in light of economic theory. As discussed in Section III.A, the predictors involving interactions allow for continuous regime switching in terms of the effects of fundamentals on expected exchange rate changes. With regard to quantitative importance, the interaction between BILL and GVOL stands out, highlighting a flight to fundamentals as FX market turbulence rises.

Section IA.5.A of the Internet Appendix analyzes recursive ENet estimates of the slope coefficients for the fitted baseline panel predictive regressions used to generate the out-of-sample baseline ENet forecasts for the 1995:01 to 2020:09 out-of-sample period. For the fitted DNNs used to generate the out-of-sample DNN and ensemble forecasts, we measure variable importance using the approach of Greenwell et al. (2018), which is based on partial dependence plots (PDPs, Friedman (2001)), as shown in Section IA.5.B of the Internet Appendix.[19] In terms of variable importance, the results paint a broadly similar picture to that for the full-sample baseline ENet estimates in Figure 2, including the crucial role of a number of interaction terms. The consistency in the variable-importance results across the baseline panel predictive regression estimated using the ENet and the general panel predictive regression modeled via the DNN stands out, demonstrating the importance and robustness of nonlinear interactions, including those involving global FX volatility. We examine regime switching governed by GVOL in more detail in Section IA.6 of the Internet Appendix.

[19]The recursive ENet estimates are shown in Figure IA.9, and recursive variable-importance measures for the fitted models underlying the DNN and ensemble forecasts are depicted in Figures IA.10 and IA.11, respectively, in the Internet Appendix.

VI. Economic Value

In this section, we evaluate the economic value of the baseline ENet, DNN, and ensemble forecasts from an investment perspective. Generating superior portfolio returns relative to a no-change benchmark does not necessarily equate with outperforming the benchmark in terms of MSPE (e.g., Melvin, Prins, and Shand (2013)). Having established that our machine-learning forecasts outperform the random walk without drift in terms of statistical accuracy, we examine their performance as inputs for portfolio construction. This dual assessment provides both deeper insight into the quality of our machine-learning forecasts and comprehensive measures of their economic value to investors.

A. Portfolio Construction

Consider a US investor who goes long (short) the currency for country i (US dollar). We can approximate the log excess return for the investment as $rx_{i,t+1} \approx (r_{i,t} - r_{\text{US},t}) - \Delta s_{i,t+1}$, where $rx_{i,t}$ is the month-t log currency excess return for country i, and $r_{i,t}$ ($r_{\text{US},t}$) is the government bill yield for country i (the United States). Assume that the investor has mean-variance preferences. The investor allocates monthly across the US dollar and all available foreign currencies. At the end of month T, the investor's objective function is

$$\max_{\boldsymbol{w}_{T+1|T}} \boldsymbol{w}'_{T+1|T}\, \hat{\boldsymbol{\mu}}_{T+1|T} - 0.5\gamma\, \boldsymbol{w}'_{T+1|T}\, \hat{\boldsymbol{\Sigma}}_{T+1|T}\, \boldsymbol{w}_{T+1|T}, \tag{13}$$

where

$$\widehat{rx}_{i,T+1|T} = (r_{i,T} - r_{\text{US},T}) - \widehat{\Delta s}_{i,T+1|T}, \tag{14}$$

$\widehat{\Delta s}_{i,T+1|T}$ ($\widehat{rx}_{i,T+1|T}$) is the investor's exchange rate change (currency excess return) forecast for country i, $\hat{\boldsymbol{\mu}}_{T+1|T} = \left[\begin{array}{ccc} \widehat{rx}_{1,T+1|T} & \cdots & \widehat{rx}_{N,T+1|T} \end{array}\right]'$, $\hat{\boldsymbol{\Sigma}}_{T+1|T}$ is the investor's estimate of the covariance matrix for the currency excess returns, $\boldsymbol{w}_{T+1|T} = \left[\begin{array}{ccc} w_{1,T+1|T} & \cdots & w_{N,T+1|T} \end{array}\right]'$ is the N-vector of portfolio weights that will be in effect for month $T+1$, $w_{i,T+1|T}$ is the allocation to currency i (the US dollar allocation is $1 - \boldsymbol{w}'_{T+1|T}\boldsymbol{\iota}_N$, where $\boldsymbol{\iota}_N$ is an N-vector of ones), and γ is the coefficient of relative risk aversion. As is common among practitioners, we assume that the investor uses an exponentially weighted moving average (EWMA) estimator for $\hat{\boldsymbol{\Sigma}}_{T+1|T}$ (with a decay factor of 0.92). To keep the portfolio weights in plausible ranges when investing in multiple assets, we impose the restrictions that $-0.5 \leq w_{i,T+1|T} \leq 0.5$ for $i = 1, \ldots, N$. We assume that $\gamma = 5$; the results are similar for reasonable alternative γ values.

We measure the economic value of a machine-learning exchange rate forecast vis-à-vis the no-change benchmark by considering two cases. In the first case, the Bench-Opt (benchmark optimized) portfolio, the investor uses the no-change benchmark forecast ($\widehat{\Delta s}_{i,T+1|T} = 0$) in equation (14), so they ignore exchange rate predictability and set the expected exchange rate change to zero (i.e., they use the stringent no-change benchmark forecast from the literature). In the second case, the ML-Opt (machine-learning optimized) portfolio, the investor uses the baseline ENet, DNN, or ensemble forecast for $\widehat{\Delta s}_{i,T+1|T}$ in equation (14), so they exploit exchange rate predictability by using the information in the predictors when allocating to foreign currencies. The Bench-Opt portfolio can be viewed as an optimal carry-trade portfolio, in the sense that it uses the bill yield differential to forecast the currency excess return ($\widehat{\Delta s}_{i,T+1|T} = 0$ in equation (14)) in a mean-variance optimization framework (e.g., Daniel et al. (2017)). The ML-Opt portfolio refines the Bench-Opt portfolio by augmenting the bill yield differential with a machine-learning exchange rate forecast to predict the currency excess return.

We construct out-of-sample portfolio weights as follows. We first use data through 1994:12 to compute the EWMA covariance matrix estimate and baseline ENet, DNN, or ensemble forecasts for 1995:01. We then solve equation (13) using the EWMA covariance estimate and baseline ENet, DNN, or ensemble (no-change) forecasts to compute the ML-Opt (Bench-Opt) portfolio weights for 1995:01. Next, we use data through 1995:01 to generate the EWMA covariance estimate and baseline ENet, DNN, or ensemble forecasts for 1995:02, and we compute the ML-Opt and Bench-Opt portfolio weights for 1995:02. We proceed in this manner through the end of the out-of-sample period, thereby mimicking the situation of an investor in real time. By comparing the performance of the ML-Opt portfolio to that of the Bench-Opt portfolio, we can gauge the economic value of the machine-learning forecasts to an investor.[20]

We compute the annualized average utility gain for the investor when they use the ML-Opt in place of the Bench-Opt portfolio:

$$(15) \qquad \text{Gain} = 12\big[\big(\overline{rx}_{\text{ML-Opt}} - 0.5\gamma\sigma^2_{\text{ML-Opt}}\big) - \big(\overline{rx}_{\text{Bench-Opt}} - 0.5\gamma\sigma^2_{\text{Bench-Opt}}\big)\big],$$

where $\overline{rx}_{\text{ML-Opt}}$ ($\overline{rx}_{\text{Bench-Opt}}$) and $\sigma_{\text{ML-Opt}}$ ($\sigma_{\text{Bench-Opt}}$) are the mean and volatility, respectively, for the monthly ML-Opt (Bench-Opt) portfolio excess return over the out-of-sample period. Equation

[20]We obtain similar results when we compute the portfolio excess return using simple (instead of log) currency excess returns based on the relevant forward and spot rates.

(15) is the annualized gain in certainty equivalent return, which can be interpreted as the maximum annualized portfolio management fee that the investor would be willing to pay to have access to the ML-Opt vis-à-vis the Bench-Opt portfolio. It measures the economic value of a machine-learning forecast to the investor.

B. Portfolio Performance

Table 3 reports portfolio performance metrics for the ML-Opt and Bench-Opt portfolios. In addition to the annualized average utility gain for the ML-Opt portfolio relative to Bench-Opt portfolio, Table 3 reports annualized means, volatilities, and Sharpe ratios for the ML-Opt and Bench-Opt portfolio excess returns, as these are often of interest for analyzing foreign-currency (including carry-trade) portfolios.

[Insert Table 3 approximately here]

Panel A of Table 3 reports results for the full 1995:01 to 2020:09 out-of-sample period. The ML-Opt portfolios based on the baseline ENet, DNN, and ensemble forecasts all perform very well, delivering annualized Sharpe ratios between 0.80 (DNN) and 0.97 (baseline ENet), all of which are significant at the 1% level. The Bench-Opt portfolio, which ignores exchange rate predictability, performs worse, with a Sharpe ratio of 0.64 (significant at the 1% level). The ML-Opt portfolios also provide substantive economic value to the investor vis-à-vis the Bench-Opt portfolio, with sizable average utility gains, ranging from 198 (DNN) to 450 (baseline ENet) basis points.

As discussed earlier, while carry-trade portfolios perform well prior to the Global Financial Crisis, they experience large losses in late 2008, and their performance generally deteriorates following the crisis. With this in mind, Panels B and C of Table 3 report results for the 1995:01 to 2008:08 and 2008:09 to 2020:09 subsamples, respectively. The start of the second subsample coincides with the bankruptcy of Lehman Brothers on September 15, 2008 during the worst phase of the Global Financial Crisis.

As discussed in Section VI.A, the Bench-Opt portfolio can be viewed as an optimal carry-trade portfolio. For the first subsample in Panel B, the average return and Sharpe ratio for the Bench-Opt portfolio are considerably higher than their values for the full sample, with an annualized average return of 10.93% and a Sharpe ratio of 1.06 (significant at the 1% level). As shown in Panel C, beginning in September of 2008, the average return and Sharpe ratio decline dramatically, with values of 1.37% and 0.15, respectively (the latter is insignificant at conventional levels). For comparison, we construct a standard (i.e., non-optimized) long-short carry-trade portfolio that sorts currencies based on bill yield differentials into quintiles and goes long (short) the highest (lowest) interest-rate currencies each month, following the construction of the carry factor in Lustig et al. (2011). The conventional carry-trade portfolio generates annualized Sharpe ratios of 0.33, 0.77, and -0.08 for the 1995:01 to 2020:09, 1995:01 to 2008:08, and 2008:09 to 2020:09 out-of-sample periods, respectively, so it suffers a sharp deterioration in performance starting in September of 2008.[21] Although the Bench-Opt portfolio performs better than the conventional carry-trade portfolio, like the latter, it still experiences a substantive decline in performance starting in late 2008.

Turning to the ML-Opt portfolios, in terms of the Sharpe ratio, the ML-Opt portfolios perform similarly to the Bench-Opt portfolio for the first subsample in Panel B of Table 3. Nevertheless, the ML-Opt portfolios still provide significant economic value to the investor

[21] The Sharpe ratio is significant at the 10% (1%) level for the full out-of-sample period (first subsample) and insignificant at conventional levels for the second subsample.

relative to the Bench-Opt portfolio, with annualized average utility gains of 82 (ensemble) to 95 (DNN) basis points. Differences in performance between the Bench-Opt and ML-Opt portfolios become much more marked for the second subsample in Panel C. The Sharpe ratios for the ML-Opt portfolios remain relatively high in the second subsample. Furthermore, the annualized average utility gains accruing to the ML-Opt portfolios are especially sizable for the second subsample, ranging from 308 (DNN) to 846 (baseline ENet) basis points.

In sum, Table 3 shows that the machine-learning exchange rate forecasts deliver substantive economic value from an investment perspective. The economic value of the forecasts is especially evident beginning in late 2008 during the worst phase of the Global Financial Crisis. While the performance of optimal and conventional carry-trade portfolios declines precipitously starting in late 2008, by harnessing the information in economic fundamentals and their interactions with global variables (especially global FX volatility), the ML-Opt portfolios continue to generate substantial investment gains through the end of the out-of-sample period.

[Insert Figure 3 approximately here]

To provide additional perspective on the economic value of the machine-learning exchange rate forecasts, Figure 3 depicts log cumulative excess returns for the ML-Opt portfolios as well as the Bench-Opt and conventional carry-trade portfolios. The three ML-Opt portfolios and Bench-Opt portfolio perform somewhat similarly through the summer of 2008, although the ML-Opt portfolios fare substantially better in the late 1990s and early 2000s in the wake of the Asian financial and Long-Term Capital Management crises. While the ML-Opt and Bench-Opt portfolios experience losses in September of 2008 during the Lehman bankruptcy, their subsequent performances differ markedly, in line with Panel C of Table 3. The Bench-Opt portfolio suffers additional sizable losses later in 2008 in Figure 3, and its cumulative return essentially “flatlines” thereafter. The conventional carry-trade portfolio (which is not based on an optimization framework) suffers an even larger drop in late 2008 compared to the Bench-Opt portfolio and also flatlines subsequently. In sharp contrast, the ML-Opt portfolios make a strong recovery in late 2008 and continue to produce gains consistently thereafter, especially the ML-Opt portfolio based on the baseline ENet forecast. As indicated along the right-hand-side of Figure 3, the log cumulative return profiles translate into sizable differences in terminal wealth. For an investor starting with $1 at the end of 1994:12, they would only have $4.61 ($2.02) at the end of 2020:09 based on the Bench-Opt (conventional carry-trade) portfolio; if they instead relied on an ML-Opt portfolio, they would have $21.76, $9.11, or $13.28 using the baseline ENet, DNN, or ensemble forecast, respectively.

Section A.III of the Appendix lists a variety of robustness checks regarding portfolio performance (with detailed results reported in the Internet Appendix). Overall, the results show that the strong performance of the ML-Opt portfolios is robust along numerous dimensions.

VII. Conclusion

This paper makes substantial progress in resolving the exchange rate predictability puzzle. Monthly US dollar exchange rates for a panel of ten developed economies over the period 1995 to 2020 are consistently and significantly predictable out of sample once the predictive role of fundamentals is allowed to depend on global financial conditions. We achieve this by combining a large information set—ten country characteristics, six global state variables, and their 60 interactions—with machine-learning techniques (the elastic net and a deep neural network) that handle the high-dimensional, noisy data. The improvements in forecast accuracy are particularly large during the Global Financial Crisis.

Model-interpretation tools identify the channels through which fundamentals do their work. We show that as global FX volatility rises, fundamentals—in particular, the predictive relation posited by UIP between the bill yield differential and expected US dollar appreciation—become increasingly important, consistent with a “flight to fundamentals” as financial stress spikes.

The economic value of the machine-learning forecasts is substantive. A mean-variance US investor who uses our forecasts (the ML-Opt portfolios) realizes annualized utility gains of 198 to 450 basis points over a benchmark investor who relies on the bill yield differential alone (the Bench-Opt portfolio, which can be viewed as an optimal carry-trade portfolio). The gains are present throughout the sample but become especially large starting in late 2008, when the Bench-Opt portfolio’s performance sharply deteriorates—a pattern that mirrors the well-documented decline of conventional carry-trade strategies after the Global Financial Crisis. Importantly, the ML-Opt portfolios continue to earn sizable and statistically significant alphas in the currency factor model of Lustig et al. (2011) both before and after the crisis, even as their carry-factor exposures collapse post-2008—direct evidence that the prediction models pick up time-varying risk exposures.

The link between fundamentals and exchange rates is conditional, and it strengthens as global financial conditions deteriorate. Indeed, allowing the predictive role of fundamentals to depend on the global state is essential for resolving the exchange rate predictability puzzle.

Appendix

A.I. Predictor Data

This section describes the predictor data used in our analysis. Section IA.1 of the Internet Appendix provides further details on the data sources and construction of the variables.

A. Country Characteristics

We consider the following ten monthly country characteristics:

Inflation differential ($\text{INF}_{i,t}$). Difference in consumer price index (CPI) inflation rates for country i and the United States.

Unemployment rate gap differential ($\text{UN}_{i,t}$). Difference in unemployment rate gaps for country i and the United States. The unemployment rate gap is the cyclical component of the unemployment rate computed using the Christiano and Fitzgerald (2003) band-pass filter for periodicities between six and 96 months.[22]

Bill yield differential ($\text{BILL}_{i,t}$). Difference in three-month government bill yields for country i and the United States.

Note yield differential ($\text{NOTE}_{i,t}$). Difference in five-year government note yields for country i and the United States.

[22]To avoid look-ahead bias, we compute the unemployment rate gap only using data available at the time of forecast formation. The Hodrick and Prescott (1997) filter is often used to compute output and unemployment rate gaps. Because we are interested in out-of-sample forecasting, we use the Christiano and Fitzgerald (2003) band-pass filter, as it performs better at the right-hand endpoint.

Bond yield differential ($\text{BOND}_{i,t}$). Difference in ten-year government bond yields for country i and the United States.

Dividend yield differential ($\text{DP}_{i,t}$). Difference in dividend yields for country i and the United States.

Price-earnings ratio differential ($\text{PE}_{i,t}$). Difference in price-earnings ratios for country i and the United States.

Stock market time-series momentum differential ($\text{SRET12}_{i,t}$). Difference in cumulative twelve-month stock market returns for country i and the United States.

Idiosyncratic volatility ($\text{IV}_{i,t}$). Integrated volatility computed using the fitted residuals for the Lustig et al. (2011) two-factor model estimated using daily data for month t for country-i log currency excess returns.

Idiosyncratic skewness ($\text{IS}_{i,t}$). Integrated skewness computed using the fitted residuals for the Lustig et al. (2011) two-factor model estimated using daily data for month t for country-i log currency excess returns.

The ten country characteristics span a range of macroeconomic and financial fundamentals. The inflation differential (INF) relates to relative PPP, while the relative inflation and unemployment rate gap differential (UN) together can be linked to Taylor-rule fundamentals (Taylor (1993)) and the potential impact of monetary policy on exchange rates (e.g., Engel and West (2005), Molodtsova and Papell (2009), Filippou and Taylor (2023)). The bill, note, and bond yield differentials (BILL, NOTE, and BOND, respectively) relate to the voluminous UIP literature (e.g., Froot and Thaler (1990), Taylor (1995), Burnside (2019)) as well as the related literature on carry trades (e.g., Burnside et al. (2011b), Burnside (2019)), while longer-term yield differentials are considered by Ang and Chen (2010) and Chen and Tsang (2013) in the context of yield curves and the exchange rate. The inclusion of the equity market variables (DP, PE, and SRET12) is motivated by Hau and Rey (2006) and Cenedese, Payne, Sarno, and Valente (2016), among others, who employ equity market variables, such as valuation ratio differentials, to analyze uncovered equity parity (UEP) and the effect of portfolio rebalancing on exchange rate movements.[23] Idiosyncratic volatility (IV) refers to the component of a currency's return variability that is unique to that currency and not driven by global or systematic factors. Similarly, idiosyncratic skewness (IS) captures the asymmetry in a currency's return distribution that cannot be explained by common market influences, reflecting the likelihood of sudden, large, currency-specific depreciations or appreciations. Both IV and IS potentially represent currency market risks for which investors demand compensation. High idiosyncratic volatility and negative skewness can deter arbitrage and contribute to excess returns. These features make them relevant indicators of currency-specific risk premia and potential limits to arbitrage (e.g., Menkhoff et al. (2012b), Filippou, Gozluklu, and Taylor (2018)).

The CPIs and unemployment rates underlying the inflation differentials and unemployment rate gap differentials, respectively, are published with a lag, which can differ across countries. The US CPI and unemployment rate are both subject to a one-month publication lag. Table IA.2 in the Internet Appendix reports publication lags for the CPIs and unemployment rates for the 14 non-US countries and how $\text{INF}_{i,t}$ and $\text{UN}_{i,t}$ are adjusted to account for the

[23] Cenedese et al. (2016) use the dividend yield, price-earnings ratio, and cumulative twelve-month stock return differentials as fundamentals for analyzing UEP, as these are leading stock return predictors.

publication lags to avoid look-ahead bias. No publication lags exist for the variables underlying the other eight country characteristics.[24]

B. Global State Variables

We consider the following six monthly global state variables:

Economic policy uncertainty (EPU_t). We use the Legacy Three-Component US Economic Policy Uncertainty (EPU) index from Baker et al. (2016). We focus on the US EPU index because the monthly Global Economic Policy Uncertainty (GEPU) index starts from January 1997. This composite index combines three distinct sources of policy-related uncertainty: (1) news-based component, which tracks monthly article counts across ten leading US newspapers; (2) tax policy component; (3) forecaster disagreement component.

Monetary policy uncertainty (MPU_t). The US Monetary Policy Uncertainty (MPU) index, also developed by Baker et al. (2016), which uses newspaper coverage frequency data from the same ten major US newspapers as the EPU, quantifies uncertainty specifically related to monetary policy.

Geopolitical risk (GR_t). The Geopolitical Risk (GR) index constructed by Caldara and Iacoviello (2022) quantifies global geopolitical uncertainty using automated text analysis of ten major international newspapers[25]

[24] We use final revised CPI and unemployment rate data, as vintage data are only available starting in 1999 from the OECD. Our results are likely to be little affected by this, as CPI and unemployment rate data are not substantially revised (and financial data are not revised).

[25] The EPU, MPU, and GR indices have been constructed and used in academic papers published since 2016.

Global FX volatility (GVOL_t). Following Menkhoff et al. (2012a), global FX volatility (GVOL) is computed as the monthly average of daily cross-sectional averages of the absolute values of log exchange rate changes, across all currencies.

Global FX illiquidity (GILL_t). Following Menkhoff et al. (2012a), global FX illiquidity (GILL) is measured as the monthly average of daily cross-sectional averages of the bid-ask spreads, across all currencies.

Global FX correlation (GCOR_t). Similarly to Mueller et al. (2017), we measure global FX correlation (GCOR) as the monthly average of the realized correlations for all currency pairs computed using daily log currency excess returns.

Each of these global variables captures general economic conditions that potentially affect the predictive ability of the country characteristics.

The US EPU and MPU indices provide quantitative measures of uncertainty surrounding US macroeconomic and monetary policy. The GR index captures both actual geopolitical events and rising concerns about future instability, making it a forward-looking gauge of global risk sentiment. In the context of FX markets, these indices serve as forward-looking indicators of global risk sentiment and policy credibility. Elevated economic or monetary policy uncertainty and geopolitical risk can trigger significant capital flows as investors seek safe-haven assets.

GVOL captures the overall intensity of exchange rate fluctuations in the global market and serves as a proxy for changing uncertainty and risk conditions. Increases in GVOL often coincide with macroeconomic shocks or financial instability and can induce significant changes in investor

Nevertheless, the news articles underlying the indices have been publicly available since the start of our sample in 1985:01, so it is reasonable to include them in the information set available to market participants.

behavior, such as greater risk aversion or retreat from speculative strategies, thereby influencing currency movements. As such, GVOL is a key state variable for modeling the conditional dynamics of exchange rate predictability. Indeed, Clarida et al. (2009) find that carry-trade returns, which are dominated by exchange rate fluctuations, are strongly related to FX market volatility.

GILL captures the overall trading cost and ease of transacting in the FX market, serving as a proxy for aggregate market liquidity conditions. Spikes in GILL indicate deteriorating liquidity (i.e., wider bid-ask spreads and more costly or difficult trading), which often coincides with periods of heightened uncertainty and market stress. Increases in GCOR typically reflect increases in systemic risk or market-wide sentiment shifts, where currencies respond collectively to global shocks. In contrast, declines in correlation suggest a market environment driven more by idiosyncratic, country-specific shocks. Along with GVOL, GILL and GCOR help to capture time-varying FX market conditions.

The collection of global state variables likely contains overlapping information, making them redundant to a certain degree. We initially include all of them, as it is difficult to know a priori which global state variables are the most relevant for out-of-sample exchange rate prediction.[26]

A.II. Robustness Checks: Forecast Accuracy

The Internet Appendix reports detailed results for the following robustness checks regarding forecast accuracy (with the relevant sections and tables and/or figures in the Internet

[26]The correlation matrix for the global state variables is provided in Table IA.3 in the Internet Appendix. The correlations range in magnitude from 0.001 (for EPU and GILL) to 0.50 (EPU and MPU).

Appendix provided in parentheses): alternative hyperparameter tuning methods for ENet estimation of the baseline panel predictive regression (Section IA.4.A and Table IA.7); panel version of the random features model in Kelly, Malamud, and Zhou (2024) (Section IA.4.B and Table IA.8); ridge estimation of the baseline panel predictive regression (Section IA.4.C and Table IA.9); baseline ENet, DNN, and ensemble forecasts generated by estimating models using country-specific data (i.e., not pooling; Section IA.4.D and Table IA.10); a well-known machine-learning model based on decision trees, random forests (Breiman (2001)) (Section IA.4.E and Table IA.11); additional country characteristics from the literature (Section IA.4.F and Tables IA.12 and IA.13). The results for the baseline ENet, DNN, and ensemble forecasts in Table 1 are also robust to alternative benchmark forecasts based on popular predictors from the literature (Section IA.4.G and Tables IA.14 and IA.15). To provide additional perspective on the consistency of the out-of-sample gains, we employ the graphical device of Goyal and Welch (2008) based on cumulative differences in squared prediction errors; the results support those in Figure 1 (Section IA.4.H and Figure IA.6). Furthermore, we show that our results are consistent with a recent literature that finds exchange rate predictability related to a safe-haven role for the US dollar (Section IA.4.I, Figure IA.7, and Table IA.5).

A.III. Robustness Checks: Portfolio Performance

The Internet Appendix reports additional results regarding portfolio performance and the portfolio weights, including for the following: average annualized utility gains for the mean-variance investor when they use versions of the Bench-Opt and ML-Opt portfolios to allocate between the US dollar and an individual foreign currency (Section IA.7.A and Table

IA.17); performance measures for the Bench-Opt and ML-Opt portfolios for non-US domestic investors (Table IA.18); transaction costs (Section IA.7.B and Table IA.19); time-series plots of the ML-Opt and Bench-Opt portfolio weights (Figures IA.12 to IA.14); links between the exchange rate forecasts and portfolio weights in late 2008 (Section IA.7.C and Figures IA.15 to IA.20); alphas in the context of the Lustig et al. (2011) currency factor model that includes dollar and carry-trade risk factors (Section IA.7.D and Table IA.20). Along the lines of Chernov et al. (2023), we also compute conditional maximum Sharpe ratios, which relate to the conditional price of currency risk (Section IA.7.E and Figure IA.21). Finally, we find that the performance of the ML-Opt portfolios significantly improves relative to the Bench-Opt portfolios as GVOL increases, analogously to the extensive evidence in Table 2 (Section IA.7.F and Table IA.21).

References

Ang, A., and J. Chen. "Yield Curve Predictors of Foreign Exchange Returns." SSRN Working Paper No. 1542342 (2010).

Avramov, D.; S. Cheng; and L. Metzker. "Machine Learning Versus Economic Restrictions: Evidence from Stock Return Predictability." *Management Science*, 69 (2023), 2587–2619.

Baker, S. R.; N. Bloom; and S. J. Davis. "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics*, 131 (2016), 1593–1636.

Bao, Y. "Estimation Risk-Adjusted Sharpe Ratio and Fund Performance Ranking under a General Return Distribution." *Journal of Financial Econometrics*, 7 (2009), 152–173.

Breiman, L. "Random Forests." *Machine Learning*, 45 (2001), 5–32.

Burnside, A. C. "Exchange Rates, Interest Parity, and the Carry Trade." *Oxford Research Encyclopedia of Economics and Finance*. August 28, 2019.

Burnside, A. C.; M. S. Eichenbaum; I. Kleshchelski; and S. T. Rebelo. “Do Peso Problems Explain the Returns to the Carry Trade?” *Review of Financial Studies*, 24 (2011a), 853–891.

Burnside, A. C.; M. S. Eichenbaum; and S. T. Rebelo. “Carry Trade and Momentum in Currency Markets.” *Annual Review of Finance and Economics*, 3 (2011b), 511–535.

Caldara, D., and M. Iacoviello. “Measuring Geopolitical Risk.” *American Economic Review*, 112 (2022), 1194–1225.

Campbell, J. Y., and S. B. Thompson. “Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?” *Review of Financial Studies*, 21 (2008), 1509–1531.

Cenedese, G.; R. Payne; L. Sarno; and G. Valente. “What Do Stock Markets Tell Us About Exchange Rates?” *Review of Finance*, 20 (2016), 1045–1080.

Chen, L.; M. Pelger; and J. Zhu. “Deep Learning in Asset Pricing.” *Management Science*, 70 (2024), 714–750.

Chen, Y.-C., and K. P. Tsang. “What Does the Yield Curve Tell Us About Exchange Rate Predictability?” *Review of Economics and Statistics*, 95 (2013), 185–205.

Chernov, M.; M. Dahlquist; and L. A. Lochstoer. “Pricing Currency Risks.” *Journal of Finance*, 78 (2023), 693–730.

Christiano, L. J., and T. J. Fitzgerald. “The Band Pass Filter.” *International Economic Review*, 44 (2003), 435–465.

Clarida, R. H.; J. Davis; and N. Pedersen. “Currency Carry Trade Regimes: Beyond the Fama Regression.” *Journal of International Money and Finance*, 28 (2009), 1375–1389.

Clarida, R. H.; L. Sarno; M. P. Taylor; and G. Valente. “The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond.” *Journal of International Economics*, 60 (2003), 61–83.

Clarida, R. H., and M. P. Taylor. “The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors.” *Review of Economics and Statistics*, 79 (1997), 353–361.

Clark, T. E., and K. D. West. "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." *Journal of Econometrics*, 138 (2007), 291–311.

Colacito, R.; S. J. Riddiough; and L. Sarno. "Business Cycles and Currency Returns." *Journal of Financial Economics*, 137 (2020), 659–678.

Dahlquist, M., and H. Hasseltoft. "Economic Momentum and Currency Returns." *Journal of Financial Economics*, 136 (2020), 152–167.

Dahlquist, M., and J. Pénasse. "The Missing Risk Premium in Exchange Rates." *Journal of Financial Economics*, 143 (2022), 697–715.

Daniel, K.; R. J. Hodrick; and Z. Lu. "The Carry Trade: Risks and Drawdowns." *Critical Finance Review*, 6 (2017), 211–262.

Della Corte, P.; A. Jeanneret; and E. Patelli. "A Credit-Based Theory of the Currency Risk Premium." *Journal of Financial Economics*, 149 (2023), 473–496.

Della Corte, P.; L. Sarno; and I. Tsiakas. "An Economic Evaluation of Empirical Exchange Rate Models." *Review of Financial Studies*, 22 (2009), 3491–3530.

Diebold, F. X., and R. S. Mariano. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13 (1995), 253–263.

Eichenbaum, M. S.; B. K. Johannsen; and S. T. Rebelo. "Monetary Policy and the Predictability of Nominal Exchange Rates." *Review of Economic Studies*, 88 (2021), 192–228.

Engel, C., and K. D. West. "Exchange Rates and Fundamentals." *Journal of Political Economy*, 113 (2005), 485–517.

Fama, E. F., and K. R. French. "Business Conditions and Expected Returns on Stocks and Bonds." *Journal of Financial Economics*, 25 (1989), 23–49.

Filippou, I.; A. E. Gozluklu; and M. P. Taylor. "Global Political Risk and Currency Momentum." *Journal of Financial and Quantitative Analysis*, 53 (2018), 2227–2259.

Filippou, I., and M. P. Taylor. “Forward-Looking Policy Rules and Currency Premia.” *Journal of Financial and Quantitative Analysis*, 58 (2023), 449–483.

Friedman, J. H. “Greedy Function Approximation: A Gradient Boosting Machine.” *Annals of Statistics*, 29 (2001), 1189–1232.

Froot, K. A., and R. H. Thaler. “Anomalies: Foreign Exchange.” *Journal of Economic Perspectives*, 4 (1990), 179–192.

Goodfellow, I.; Y. Bengio; and A. Courville. *Deep Learning*. MIT Press, Cambridge, MA (2016).

Goyal, A., and I. Welch. “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction.” *Review of Financial Studies*, 21 (2008), 1455–1508.

Greenwell, B. M.; B. C. Boehmke; and A. J. McCarthy. “A Simple and Effective Model-Based Variable Importance Measure.” Working Paper arXiv:1805.04755v1 (2018).

Gu, S.; B. Kelly; and D. Xiu. “Empirical Asset Pricing via Machine Learning.” *Review of Financial Studies*, 33 (2020), 2223–2273.

Hau, H., and H. Rey. “Exchange Rates, Equity Prices, and Capital Flows.” *Review of Financial Studies*, 19 (2006), 273–317.

Hodrick, R. J., and E. C. Prescott. “Postwar U.S. Business Cycles: An Empirical Investigation.” *Journal of Money, Credit, and Banking*, 19 (1997), 1–16.

Hoerl, A. E., and R. W. Kennard. “Ridge Regression: Applications to Nonorthogonal Problems.” *Technometrics*, 12 (1970), 69–82.

Hornik, K.; M. Stinchcombe; and H. White. “Multilayer Feedforward Networks Are Universal Approximators.” *Neural Networks*, 2 (1989), 359–366.

Hsu, P.-H.; M. P. Taylor; Z. Wang; and Y. Li. “The Out-of-Sample Performance of Carry Trades.” *Journal of International Money and Finance*, 143 (2024), 103042.

Hsu, P.-H.; M. P. Taylor; Z. Wang; and Y. Li. "On the Profitability of Influential Carry-Trade Strategies: Data-Snooping Bias and Post-Publication Performance." *Journal of Empirical Finance*, 83 (2025), 101640.

Hui, F. K. C.; D. I. Warton; and S. D. Foster. "Tuning Parameter Selection for the Adaptive Lasso Using ERIC." *Journal of the American Statistical Society*, 110 (2015), 262–269.

Jordà, O., and A. M. Taylor. "The Carry Trade and Fundamentals: Nothing to Fear but FEER Itself." *Journal of International Economics*, 88 (2012), 74–90.

Kelly, B.; S. Malamud; and K. Zhou. "The Virtue of Complexity in Return Prediction." *Journal of Finance*, 79 (2024), 459–503.

Kilian, L., and M. P. Taylor. "Why Is It So Difficult to Beat the Random Walk Forecast of Exchange Rates?" *Journal of International Economics*, 60 (2003), 85–107.

Lustig, H.; N. L. Roussanov; and A. Verdelhan. "Common Risk Factors in Currency Markets." *Review of Financial Studies*, 24 (2011), 3731–3777.

Mark, N. C. "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability." *American Economic Review*, 85 (1995), 201–218.

Meese, R. A., and K. Rogoff. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics*, 14 (1983), 3–24.

Melvin, M.; J. Prins; and D. Shand. "Forecasting Exchange Rates: An Investor Perspective." In *Handbook of Economic Forecasting*, Vol. 2A, G. Elliott; and A. Timmermann, eds. Elsevier, Amsterdam (2013), 721–750.

Melvin, M., and D. Shand. "When Carry Goes Bad: The Magnitude, Causes, and Duration of Currency Carry Unwinds." *Financial Analysts Journal*, 73 (2017), 121–144.

Melvin, M., and M. P. Taylor. "The Crisis in the Foreign Exchange Market." *Journal of International Money and Finance*, 28 (2009), 1317–1330.

Menkhoff, L.; L. Sarno; M. Schmeling; and A. Schrimpf. "Carry Trades and Global Foreign Exchange Volatility." *Journal of Finance*, 67 (2012a), 681–718.

Menkhoff, L.; L. Sarno; M. Schmeling; and A. Schrimpf. "Currency Momentum Strategies." *Journal of Financial Economics*, 106 (2012b), 660–684.

Molodtsova, T., and D. H. Papell. "Out-of-Sample Exchange Rate Predictability with Taylor Rule Fundamentals." *Journal of International Economics*, 77 (2009), 167–180.

Moskowitz, T. J.; Y. H. Ooi; and L. H. Pedersen. "Time Series Momentum." *Journal of Financial Economics*, 104 (2012), 228–250.

Mueller, P.; A. Stathopoulos; and A. Vedolin. "International Correlation Risk." *Journal of Financial Economics*, 126 (2017), 270–299.

Rapach, D. E.; J. K. Strauss; and G. Zhou. "International Stock Return Predictability: What Is the Role of the United States?" *Journal of Finance*, 68 (2013), 1633–1662.

Rossi, B. "Exchange Rate Predictability." *Journal of Economic Literature*, 51 (2013), 1063–1119.

Sager, M., and M. P. Taylor. "Generating Currency Trading Rules from the Term Structure of Forward Foreign Exchange Premia." *Journal of International Money and Finance*, 44 (2014), 230–250.

Taylor, J. B. "Discretion Versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*, 39 (1993), 195–214.

Taylor, M. P. "The Economics of Exchange Rates." *Journal of Economic Literature*, 33 (1995), 13–47.

Tibshirani, R. "Regression Shrinkage and Selection via the LASSO." *Journal of the Royal Statistical Society. Series B (Methodological)*, 58 (1996), 267–288.

West, K. D. "Asymptotic Inference about Predictive Ability." *Econometrica*, 64 (1996), 1067–1084.

Zou, H., and T. Hastie. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67 (2005), 301–320.

TABLE 1

Out-of-Sample R^2 Statistics

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; 0.00 indicates less than 0.005 in absolute value; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. "Baseline OLS" ("Baseline ENet") is the baseline panel predictive regression in equation (1) estimated via OLS (the ENet); "No-interaction ENet" is the panel predictive regression without interaction terms in equation (3) estimated via the ENet; "DNN" is the deep neural network in Section III.B; "Ensemble" is an average of the baseline ENet and DNN forecasts.

1	2	3	4	5	6	7	8
Country	Out-of-sample period	Obs.	Baseline OLS	Baseline ENet	No-interaction ENet	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	−14.13	3.24**	0.28**	3.03**	3.41**
Switzerland	1995:01–2020:09	309	−14.20	0.37	−0.81	2.15**	1.60**
Japan	1995:01–2020:09	309	−12.36	−0.38	−2.01	0.28	0.19
Canada	1995:01–2020:09	309	−19.75	0.00	−3.10	0.91	0.87
Australia	1995:01–2020:09	309	−3.97	1.47**	0.77*	1.89***	1.83**
New Zealand	1995:01–2020:09	309	−8.22	2.15**	0.21*	1.96**	2.19**
Sweden	1995:01–2020:09	309	−3.35	1.86**	−0.20	1.69**	1.95**
Norway	1995:01–2020:09	309	−3.21	1.91**	0.20*	1.55**	1.93**
Denmark	1995:01–2020:09	309	−5.78	2.23***	1.02**	2.04***	2.34***
Euro area	2000:02–2020:09	248	−11.21	2.38**	−0.50	2.11**	2.65**
All	1995:01–2020:09	3,029	−8.74	1.51**	−0.29	1.74***	1.86***

TABLE 2

Global FX Volatility and Relative Forecast Accuracy

The table reports OLS estimates of $a_{i,1}$ for the regression, $l_{i,t} = a_{i,0} + a_{i,1}\text{GVOL}_t + \varepsilon_{i,t}$, where $l_{i,t}$ is the Clark and West (2007) adjusted loss differential for country i for the no-change benchmark forecast vis-à-vis the competing forecast in the column heading, and GVOL_t is global FX volatility. We test H_0: $a_{i,1} \leq 0$ against H_A: $a_{i,1} > 0$. "All" corresponds to the pooled regression, $l_{i,t} = a_0 + a_1\text{GVOL}_t + \varepsilon_{i,t}$. We test H_0: $a_1 \leq 0$ against H_A: $a_1 > 0$; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

1	2	3	4	5	6
Country	Out-of-sample period	Obs.	Baseline ENet	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	7.22^{***}	0.71	3.96^{***}
Switzerland	1995:01–2020:09	309	3.76^{**}	2.90^{**}	3.33^{***}
Japan	1995:01–2020:09	309	-0.88	2.76^{***}	0.94
Canada	1995:01–2020:09	309	8.15^{***}	1.37	4.76^{***}
Australia	1995:01–2020:09	309	11.20^{***}	3.72^{***}	7.46^{***}
New Zealand	1995:01–2020:09	309	8.91^{***}	1.16	5.04^{***}
Sweden	1995:01–2020:09	309	7.57^{***}	0.95	4.26^{***}
Norway	1995:01–2020:09	309	9.89^{***}	2.41^{*}	6.15^{***}
Denmark	1995:01–2020:09	309	6.50^{***}	3.37^{***}	4.93^{***}
Euro area	2000:02–2020:09	248	20.87^{***}	8.22^{***}	14.55^{***}
All	1995:01–2020:09	3,029	8.05^{***}	2.64^{***}	5.35^{***}

TABLE 3

Portfolio Performance: Allocating Across All Available Foreign Currencies

The table reports portfolio performance metrics for a mean-variance US investor with a relative risk aversion coefficient of five who allocates monthly across the US dollar and all available foreign currencies. For the ML-Opt (Bench-Opt) portfolio, the investor uses the baseline ENet, DNN, or ensemble (no-change) exchange rate forecast when predicting the currency excess return. The second column reports the annualized increase in certainty equivalent return when the investor uses the ML-Opt instead of the Bench-Opt portfolio. Statistical significance for the Sharpe ratio is based on the Bao (2009) procedure; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

1 Portfolio	2 Annualized average utility gain	3 Annualized mean	4 Annualized volatility	5 Annualized Sharpe ratio
Panel A. 1995:01 to 2020:09 out-of-sample period				
ML-Opt, baseline ENet	4.50%	12.89%	13.30%	0.97***
ML-Opt, DNN	1.98%	9.27%	11.54%	0.80***
ML-Opt, ensemble	3.04%	10.84%	12.39%	0.87***
Bench-Opt	–	6.45%	12.39%	0.64***
Panel B. 1995:01 to 2008:08 out-of-sample period				
ML-Opt, baseline ENet	0.89%	13.36%	12.93%	1.03***
ML-Opt, DNN	0.95%	12.36%	11.19%	1.11***
ML-Opt, ensemble	0.82%	12.60%	11.82%	1.07***
Bench-Opt	–	10.93%	10.61%	1.06***
Panel C. 2008:09 to 2020:09 out-of-sample period				
ML-Opt, baseline ENet	8.46%	12.35%	13.75%	0.90***
ML-Opt, DNN	3.08%	5.77%	11.88%	0.49*
ML-Opt, ensemble	5.44%	8.85%	13.03%	0.68***
Bench-Opt	–	1.37%	9.39%	0.15

FIGURE 1

Rolling Out-of-Sample R^2 Statistics

The figure depicts rolling 60-month $R^2_{i,\mathrm{OS}}$ and $R^2_{\mathrm{All,OS}}$ statistics in percent. The horizontal axis gives the end of the 60-month period. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

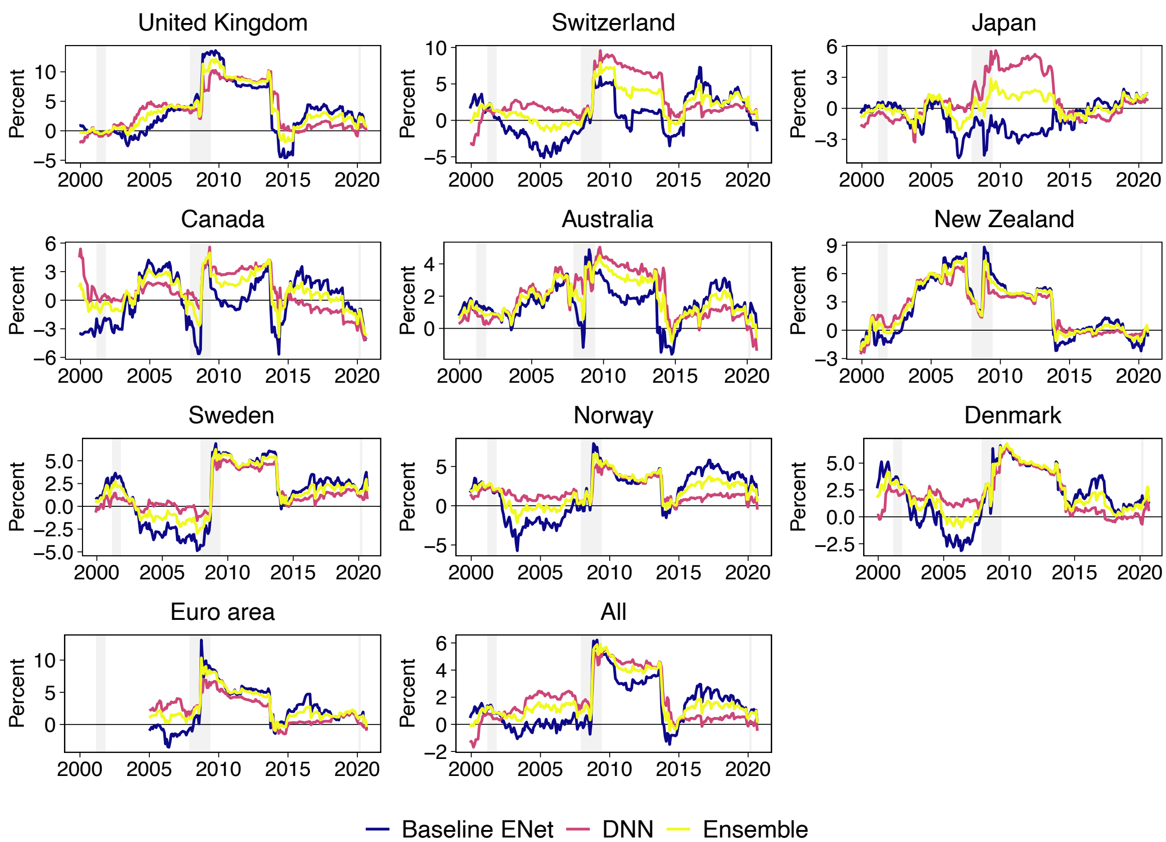


FIGURE 2

Full-Sample Baseline ENet Coefficient Estimates

The figure depicts ENet estimates of the slope coefficients for the baseline panel predictive regression and the 1985:01 to 2020:09 sample period. The predictors are standardized before estimation.

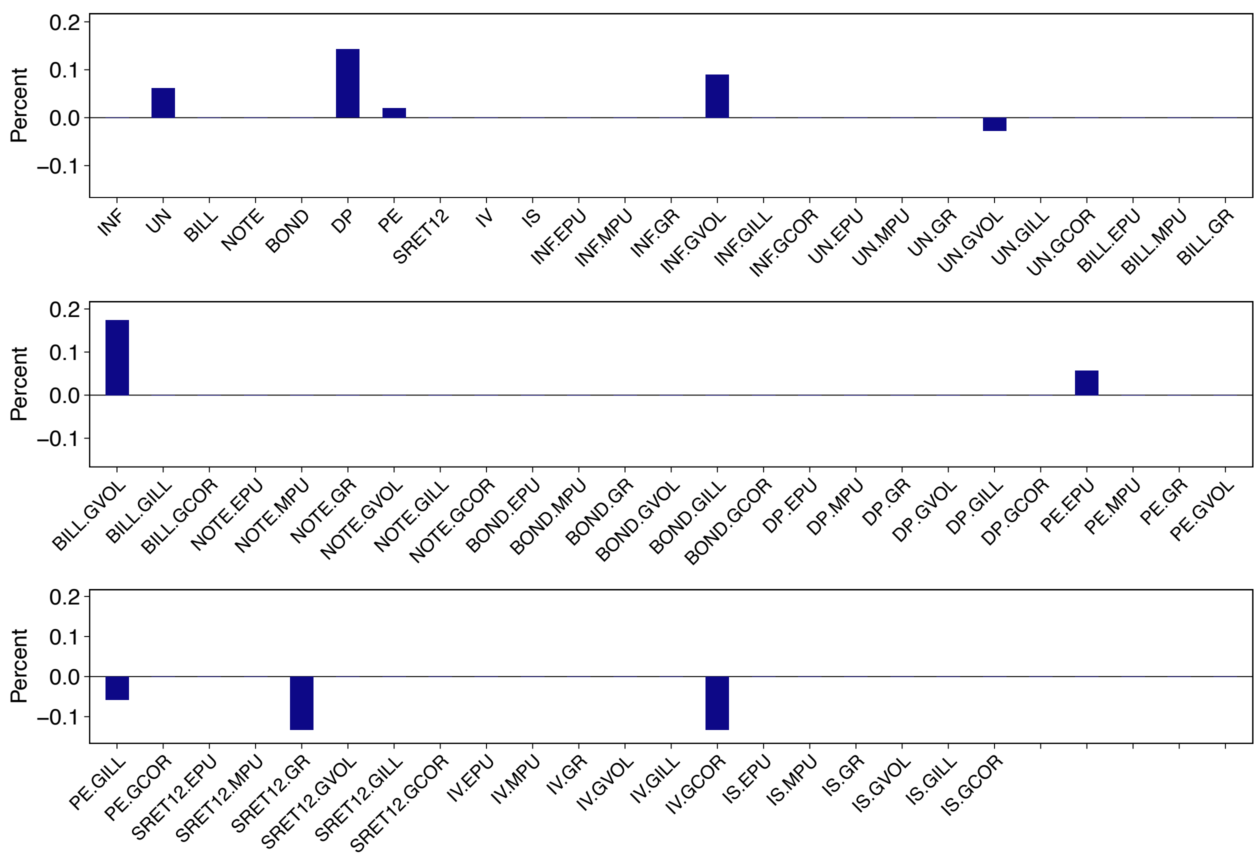
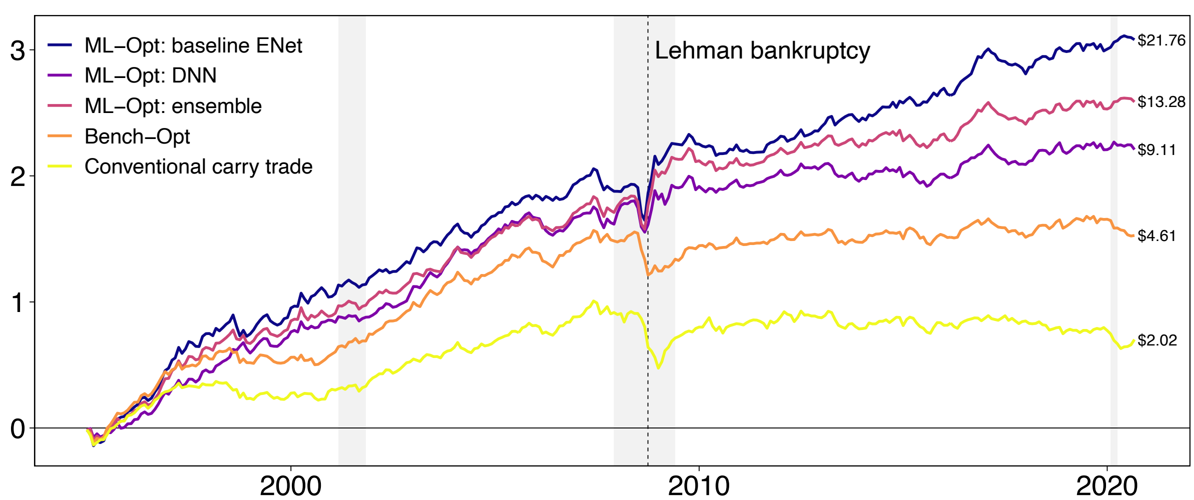


FIGURE 3

Log Cumulative Portfolio Excess Returns

The figure depicts log cumulative excess returns for the ML-Opt, Bench-Opt, and conventional carry-trade portfolios. The evaluation period is 1995:01 to 2020:09. The numbers on the right are terminal wealth values, assuming that an investor begins with $1. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.



Internet Appendix for

"Economic Fundamentals and Short-Run Exchange Rate Prediction: A Machine-Learning Perspective"

August 19, 2026

IA.1. Data

This section provides details on the data sources and construction of the country characteristics and global variables.

A. Exchange Rates

Daily bid and ask spot and forward exchange rates for 14 countries are from Barclays and Reuters via Datastream. Datastream country mnemonics for the currencies are as follows:

Country	Mnemonic	Country	Mnemonic
United Kingdom	GBP	Norway	NOK
Switzerland	CHF	Denmark	DKK
Japan	JPY	Euro area	EUR
Canada	CAD	Germany	DEM
Australia	AUD	Italy	ITL
New Zealand	NZD	France	FRF
Sweden	SEK	Netherlands	NLG

Bid spot price Ticker BB***SP(EB), where "***" indicates the country mnemonic.

Ask spot price Ticker BB***SP(EO).

Bid forward price Ticker BB***1F(EB).

Ask forward price Ticker BB***1F(EO).

We use the midpoint of end-of-month bid and ask spot prices to compute the month-t spot exchange rate for country i, which we denote by $S_{i,t}$ and is expressed as the number of country-i currency units per US dollar. We then calculate the country-i log exchange rate change: $\Delta s_{i,t} = s_{i,t} - s_{i,t-1}$, where $s_{i,t} = \log(S_{i,t})$. The set of exchange rates is similar to the sample of developed countries employed in other studies (e.g., Lustig, Roussanov, and Verdelhan (2011), Menkhoff, Sarno, Schmeling, and Schrimpf (2012a)), except Belgium, which we exclude due to a lack of data availability for the country characteristics. Table IA.1 reports summary statistics for the log exchange rate changes for the 14 countries.

B. Country Characteristics

Country characteristics are computed using data from Finaeon and the Organization for Economic Cooperation and Development (OECD). Finaeon and OECD country mnemonics are as follows:

Country	Mnemonic	Country	Mnemonic
United Kingdom	GBR	Denmark	DNK
Switzerland	CHE	Euro area	EUR
Japan	JPN	Germany	DEU
Canada	CAN	Italy	ITA
Australia	AUS	France	FRA
New Zealand	NZL	Netherlands	NLD
Sweden	SWE	United States	USA
Norway	NOR		

Inflation differential (INF) Difference in inflation rates for a given country and the United States. Inflation rates are computed from consumer price index (CPI) data from Finaeon (ticker CP***M); CPI data for the Euro area (EA19) are from the OECD (available at https://data.oecd.org/price/inflation-cpi.htm).

Unemployment rate gap differential (UN) Difference in unemployment rate gaps for a given country and the United States. The unemployment rate gap is the cyclical component of the unemployment rate computed using the Christiano and Fitzgerald (2003) band-pass filter for periodicities between six and 96 months. Unemployment rates are from Finaeon (ticker UN***M); the unemployment rate for the Euro area (EA19) is from the OECD (available at https://data.oecd.org/unemp/unemployment-rate.htm).

Bill yield differential (BILL) Difference in government bill yields for a given country and the United States. Government bill yields are three-month Treasury bill yields from Finaeon (ticker IT***3D; for the Euro area, IBEUR3D).

Note yield differential (NOTE) Difference in government note yields for a given country and the United States. Government note yields are five-year government bond yields from Finaeon (ticker IG***5D).

Bond yield differential (BOND) Difference in government bond yields for a given country and the United States. Government bond yields are ten-year government bond yields from Finaeon (ticker IG***10D).

Dividend yield differential (DP) Difference in dividend yields for a given country and the United States. Dividend yields are from Finaeon (ticker SY***YM; for the United Kingdom, _DFTASD; for Canada, SYCANYTM; for the Netherlands, SYNLDYAM).

Price-earnings differential (PE) Difference in price-earnings ratios for a given country and the

United States. Price-earnings ratios are from Finaeon (ticker SY***PM; for the United Kingdom, _PFTASD; for Japan, SYJPNPTM; for Canada, SYCANPTM).

Stock market time-series momentum differential (SRET12) Difference in cumulative twelve-month stock market returns for a given country and the United States. Twelve-month cumulative returns are computed from total return indices from Finaeon. Finaeon tickers for total return indices are as follows:

Country	Ticker	Country	Ticker
United Kingdom	_TFTASD	Denmark	_OMXCGID
Switzerland	_SSHID	Euro area	_DMIEU0D
Japan	_TOPXDVD	Germany	_CDAXD
Canada	_TRGSPTSE	Italy	_BCIPRD
Australia	_AORDAD	France	_TRSBF250D
New Zealand	_NZGID	Netherlands	_AAXGRD
Sweden	_OMXSBGI	United States	_SPXTRD
Norway	_OSEAXD		

Idiosyncratic volatility (IV) Similarly to Filippou, Gozluklu, and Taylor (2018), we construct month-t IV as follows. We first compute daily log currency excess returns using daily spot and forward rates from Datastream for the 14 countries that we analyze. We then construct daily dollar ($\mathrm{MKT}_{\mathrm{FX}}$) and carry trade ($\mathrm{HML}_{\mathrm{FX}}$) risk factors for the Lustig et al. (2011) two-factor model. The daily dollar risk factor is the cross-sectional average of the daily log currency excess returns. To construct the daily carry trade risk factor, we first sort the currencies into six portfolios based on the previous day's forward discount and take equally weighted long (short) positions in the currencies in the last (first) portfolio; the carry trade factor is the log return for the long-short portfolio. Each month, we regress daily log currency excess returns for country i on a constant and the MKT_{FX} and HML_{FX} factors:

$$rx^i_{t,d} = \alpha^i + \beta^i_{\text{MKT}_{\text{FX}},t}\text{MKT}_{\text{FX},t,d} + \beta^i_{\text{HML}_{\text{FX}},t}\text{HML}_{\text{FX},t,d} + \varepsilon^i_{t,d}, \tag{1}$$

where $rx^i_{t,d}$ is the day-d log currency excess return for country i for month t, and $\text{MKT}_{\text{FX},t,d}$ ($\text{HML}_{\text{FX},t,d}$) is the day-$d$ return for month t for the dollar (carry trade) factor. Idiosyncratic volatility is defined as

$$\text{IV}_{i,t} = \left[\frac{1}{T_{i,t}}\sum_{d=1}^{T_{i,t}}\left(\hat{\varepsilon}^i_{t,d}\right)^2\right]^{0.5}, \tag{2}$$

where $\hat{\varepsilon}^i_{t,d}$ is the fitted ordinary least squares residual for equation (1), and $T_{i,t}$ is the number of daily log currency excess return observations available for country i for month t.

Idiosyncratic skewness (IS) Defined as

$$\text{IS}_{i,t} = \left(\frac{1}{T_{i,t}-2}\right)\frac{\sum_{d=1}^{T_{i,t}}\left(\hat{\varepsilon}^i_{t,d}\right)^3}{(\text{IV}_{i,t})^3}, \tag{3}$$

where $\text{IV}_{i,t}$ is given by equation (2).[1]

CPI and unemployment rate data underlying INF and UN are published with a lag. For the United States, both series are published with a one-month lag. Table IA.2 reports publication lags for CPI and unemployment rate data, respectively, for the 14 non-US countries. The last column

[1]The construction of idiosyncratic volatility and skewness follows Goyal and Santa-Clara (2003), Fu (2009), Boyer, Mitton, and Vorkink (2010), and Chen and Petkova (2012).

of the table indicates the adjustment made to the timing of the country characteristic to account for the publication lag.

CPI data for Australia and New Zealand are reported at the quarterly frequency and published in the first month of the subsequent quarter. We proceed as follows to compute monthly inflation rates for these two countries that account for the publication lag. Consider, for example, forming a forecast of the exchange rate change for January based on data available through December. Because the CPI is not available for the fourth quarter, we use the CPI inflation for the third quarter divided by three (to convert it to a monthly frequency). When forming the next forecast for February based on data available through January, we use the CPI inflation rate for the fourth quarter divided by three. We use the same inflation rate when forming the forecasts for March and April based on data available through February and March, respectively. Unemployment rate data for New Zealand are also reported at the quarterly frequency and published in the first month of the subsequent quarter. We proceed in the same manner to compute a monthly unemployment rate for New Zealand that accounts for the publication lag, except that we do not need to divide the quarterly unemployment rate by three.

C. Global Variables

Economic policy uncertainty (EPU) Baker, Bloom, and Davis (2016) Legacy Three-Component US Economic Policy Uncertainty index based on coverage frequencies in ten major US newspapers (available at https://www.policyuncertainty.com/us_monthly.html).

Monetary policy uncertainty (MPU) Baker et al. (2016) US Monetary Policy Uncertainty index based on coverage frequencies in ten major US newspapers (available at https://www.policyuncertainty.com/monetary.html).

Geopolitical risk (GR) Caldara and Iacoviello (2022) geopolitical risk index based on newspaper coverage (available at https://www.policyuncertainty.com/gpr.html).

Global foreign exchange (FX) volatility (GVOL) As in Menkhoff et al. (2012a), month-t global FX volatility is defined as

$$\text{GVOL}_t = \frac{1}{T_t} \sum_{d \in T_t} \left[\sum_{k \in K_{t,d}} \left(\frac{|\Delta s_{t,d}^k|}{K_{t,d}} \right) \right], \tag{4}$$

where $\Delta s_{t,d}^k$ is the day-d log change in the exchange rate for country k and month t, $K_{t,d}$ is the number of currencies available for day d in month t, and T_t is the number of days in month t. We use the 48 countries from Menkhoff et al. (2012a) to compute GVOL.

Global FX illiquidity (GILL) As in Menkhoff et al. (2012a), month-t global FX illiquidity is defined as

$$\text{GILL}_t = \frac{1}{T_t} \sum_{d \in T_t} \left[\sum_{k \in K_{t,d}} \left(\frac{|\text{BAS}_{t,d}^k|}{K_{t,d}} \right) \right], \tag{5}$$

where $\text{BAS}_{t,d}^k$ is the day-d bid-ask exchange rate spread (in percent) for country k and month t. We use the 48 countries from Menkhoff et al. (2012a) to compute GILL.

Global FX correlation (GCOR) Similarly to Mueller, Stathopoulos, and Vedolin (2017), month-t global FX currency correlation is defined as

$$\text{GCOR}_t = \frac{1}{N_t^{\text{comb}}} \sum_{i=1}^{N_t} \Big(\sum_{j>i} \text{RC}_t^{i,j} \Big), \tag{6}$$

where $\text{RC}_t^{i,j}$ is the realized correlation between log currency excess returns for countries i and j based on daily data for month t, N_t^{comb} is the number of combinations of currencies (i, j) for month t, and N_t is the number of available currencies for month t. We use the 14 countries in Section IA.1.A to compute GCOR.

IA.2. Deep Neural Network

This section provides details on the deep neural network (DNN) outlined in Section III.B of the paper. A feedforward neural network architecture is comprised of multiple layers. The first is the input layer, which is simply the set of predictors, denoted by $\tilde{x}_1, \ldots, \tilde{x}_{P_0}$. One or more hidden layers are next. Each hidden layer l contains P_l neurons, each of which takes predictive signals from the neurons in the previous hidden layer to produce a subsequent signal:

$$h_m^{(l)} = g\big(\omega_{m,0}^{(l)} + \sum_{j=1}^{P_{l-1}} \omega_{m,j}^{(l)} h_j^{(l-1)}\big) \quad \text{for } m = 1, \ldots, P_l;\ l = 1, \ldots, L, \tag{7}$$

where $h_m^{(l)}$ is the mth neuron in the lth hidden layer;[2] $\omega_{m,0}^{(l)}, \ldots, \omega_{m,P_{l-1}}^{(l)}$ are weights; and $g(\cdot)$ is a nonlinear activation function. The output layer is a linear function that translates the signals from

[2]For the first hidden layer, $h_j^{(0)} = \tilde{x}_j$ for $j = 1, \ldots, P_0$.

the last hidden layer into a prediction:

$$\hat{y} = \omega_0^{(L+1)} + \sum_{j=1}^{P_L} \omega_j^{(L+1)} h_j^{(L)}, \tag{8}$$

where $\hat{y}$ denotes the forecast of the target variable. For the activation function, we use the leaky rectified linear unit (LReLU) function (Maas, Hannun, and Ng (2013)):[3]

$$g(x) = \begin{cases} x & \text{if } x > 0, \\ 0.01x & \text{if } x \leq 0. \end{cases} \tag{9}$$

Intuitively, equation (9) activates a neuronal connection in response to the strength of the signal, thereby relaying the signal forward through the network.

Training a DNN entails estimating the weights. We estimate the weights by minimizing an objective function based on the training sample MSPE. To better guard against overfitting, we augment the objective function with ℓ_1 and ℓ_2 penalty terms, similarly to equation (5) from the paper. Computationally efficient algorithms based on stochastic gradient descent (SGD) are available to estimate the DNN weights. We use the powerful Adam algorithm (Kingma and Ba (2015)). Due to the stochastic nature of the SGD algorithm, the estimated weights depend on the seed for the random number generator. To reduce the dependency of the DNN forecast on the seed, we train the model three times with different seeds and take an average of the forecasts across the three fitted DNNs.

Given our noisy data environment, we take additional steps to further guard against

[3]The LReLU refines the conventional ReLU (which sets $g(x) = 0$ if $x \leq 0$) to help prevent the DNN from "dying" during training, meaning that the neurons in the DNN never activate.

overfitting the DNN. We set the intercept terms for the weights ($\omega_{m,0}^{(l)}$ for $m = 1, \ldots, P_l$; $l = 1, \ldots, L$ and $\omega_0^{(L+1)}$) to zero.[4] These restrictions are analogous to the absence of an intercept term in the baseline panel predictive regression in equation (1) from the paper, ensuring that the DNN predicts a value of zero for the exchange rate change when all of the (standardized) predictors are at their mean value of zero. We continue to pool the data in the DNN to impose the homogeneity restrictions that the weights are the same across countries. Imposing both the zero intercept and slope homogeneity restrictions substantially reduces the number of weights that we need to estimate, thereby helping to guard against overfitting. In addition, we employ dropout (Srivastava, Hinton, Krizhevsky, Sutskever, and Salakhutdinov (2014)), which randomly drops a portion of the neurons in a hidden layer when training the model via the SGD algorithm and is helpful in mitigating overfitting.

We tune several hyperparameters for the DNN—including the ℓ_1 and ℓ_2 regularization hyperparameters, dropout rate, and batch size—by considering a grid of values for each and using observations for the last 30% of months for the available data at the time of forecast formation as a validation sample, thereby respecting the time-series dimension of the panel data. We select the vector of hyperparameter values that minimizes the objective function for the validation sample. To help keep the hyperparameters relevant, we tune them every 36 months.

[4]The intercept terms for the weights are typically called "bias" terms in the machine-learning literature, although it does not have its traditional econometric meaning in this context.

IA.3. Non-US Dollar Base Currencies

As shown in Section IV.C of the paper, our approach performs well when the US dollar is used as the base currency, which is consistent with its dominant role as the primary settlement currency in international trade and global financial markets. In this section, we consider alternative base currencies, namely, the Japanese yen and Swiss franc. For the former (latter), the exchange rate is the number of country-i currency units per Japanese yen (Swiss franc). We measure the country characteristics as the country-i value minus the Japanese or Swiss value. For each of the alternative base currencies, we generate out-of-sample forecasts for 1995:01 to 2020:09 as described in the paper.

Tables IA.5 and IA.6 report results when the Japanese yen and Swiss franc, respectively, serve as the base currency. Our machine-learning approach also works reasonably well for forecasting exchange rate changes in Table IA.5 when the Japanese yen serves as the base currency. However, compared to the results reported in Table 1 from the paper, the degree of exchange rate predictability is not as strong as when the US dollar is the base currency. When the Swiss franc serves as the base currency, the exchange rate forecasts are substantially less accurate in Table IA.6 compared to Table 1 from the paper. Overall, we find less predictive power when other currencies, such as the Japanese yen and Swiss franc, serve as the base currency. As mentioned in Section IV.C of the paper, understanding the economic reasons for such differences is an interesting avenue for future research.

IA.4. Robustness Checks: Forecast Accuracy

A. Elastic Net Hyperparameter Tuning

Conventional M-fold cross validation (e.g., as implemented in the `glmnet` package in `R`) randomly divides the observations for the training sample into M non-overlapping and generally non-contiguous blocks or folds of approximately equal size. For each combination of δ and λ in a two-way grid, we first fit the baseline predictive regression in equation (1) from the paper via the elastic net (ENet, Zou and Hastie (2005)) using the available observations after excluding those in the first fold; for each combination of δ and λ, we use the fitted model to compute predictions for the observations in the first fold and calculate the mean squared prediction error (MSPE). Next, for each combination of δ and λ, we fit the model via the ENet using available observations after excluding those in the second fold; for each combination of δ and λ, we use the fitted model to compute predictions for the observations in the second fold and calculate the MSPE. We continue in this fashion through the remaining folds. Finally, for each combination of δ and λ, we compute the average of the MSPEs across the M folds and select the combination of δ and λ that minimizes the average MSPE.

A potential drawback to conventional M-fold cross validation in our context is that it does not take into account the time-series nature of panel data. Specifically, models can be fitted using observations that temporally follow those in a validation fold. To address this issue, we consider additional tuning methods based on validation in a panel data context. We refer to the first as time-series validation. Suppose that we want to generate forecasts for month $t + 1$ using data available through month t. We divide the panel data into observations corresponding to the first 70% and last 30% of months from the start of the sample through month t, where observations from the last 30% of months comprise the validation period. The split preserves the temporal ordering of the data. For each combination of δ and λ, we fit the baseline predictive regression in equation (1) from the paper via the ENet using data for the first 70% of months, use the fitted model to compute predictions for the validation period, and calculate the MSPE. We select the combination of δ and λ that minimizes the MSPE.

To implement cross validation (i.e., multiple validation periods) in a manner that recognizes the time-series nature of panel data, we implement what we call time-series cross validation in a panel data context. Again suppose that we want to generate forecasts for month $t+1$ using data available through month t. We begin by using panel data observations for the first 25% of months from the start of the sample through month t to fit the baseline predictive regression in equation (1) from the paper via the ENet for each combination of δ and λ. For each combination of δ and λ, we then use the fitted model to compute predictions for panel data observations for the next 25% of months (first fold) and calculate the MSPE. Next, we use panel data for the first 50% of months from the start of the sample through month t to fit the model via the ENet for each combination of δ and λ; for each combination of δ and λ, we use the fitted model to compute predictions for panel data observations for the next 25% of months (second fold) and calculate the MSPE. In the last iteration, we use panel data for the first 75% of months from the start of the sample through month t to fit the model via the ENet for each combination of δ and λ; for each combination of δ and λ, we use the fitted model to compute predictions for panel data observations for the last 25% of months (third fold) and calculate the MSPE. Finally, for each combination of δ and λ, we compute the average of the MSPEs across the three folds and select the combination of δ and λ that minimizes the average MSPE.

Table IA.7 reports out-of-sample R^2 statistics for baseline ENet forecasts that use different methods for tuning δ and λ. For all methods, we retune the hyperparameters every month. The values in the fourth through seventh columns are typically less than the corresponding values in the fifth column of Table 1 from the paper.

B. Random Features Forecast

We also consider forecasts based on the random features model in Kelly, Malamud, and Zhou (2024), which is essentially a neural network with one hidden layer (with a very large number of nodes) and random weights for the hidden layer. As shown in Table IA.8, the random features forecasts do not perform as well as the DNN forecast in the seventh column of Table 1 from the paper.

C. Ridge Forecast

Table IA.9 reports out-of-sample R^2 statistics for a forecast based on ridge estimation of the baseline predictive regression in equation (1) from the paper as well as an ensemble forecast comprised of the baseline ridge and DNN forecasts. Overall, the ridge-based forecasts are less accurate than the corresponding baseline ENet and ensemble forecasts reported in Table 1 from the paper, so allowing for sparsity provides better out-of-sample performance for our application.

D. Country-Specific Data

We also generate baseline ENet, DNN, and ensemble forecasts by estimating models using country-specific data. As shown in Table IA.10, forecasts generated using country-specific data are nearly always substantially less accurate than the forecasts based on pooling in Table 1 from the paper. Thus, for out-of-sample time-series exchange rate prediction, information appears better processed via homogeneity of the model parameters across countries in light of the bias-variance trade-off.

E. Random Forests

In addition to the baseline ENet and DNN forecasts, we consider a popular machine-learning method based on decision trees, random forests (Breiman (2001)). Like DNNs, random forests are powerful machine-learning techniques that perform well in a variety of domains. However, although they perform reasonably well, random forests are not as effective for forecasting monthly exchange rates as ENet estimation of the baseline predictive regression and the DNN. We explain why this is the case.

A random forest is built up from regression trees, which are machine-learning devices that incorporate nonlinearities via multi-way interactions and higher-order effects of the predictors. To construct a regression tree, the predictor space is sequentially split into regions, typically using the classification and regression tree (CART) algorithm (Breiman, Friedman, Stone, and Olshen (1984)). The final set of regions is referred to as "terminal nodes" or "leaves," and the prediction is the average value of the target in a given leaf. The forecast corresponding to a regression tree with U leaves can be expressed as

$$\widehat{\Delta s}^{\text{RT}}_{i,T+1|T} = \sum_{u=1}^{U} \overline{\Delta s}_u \mathbf{1}_u(\tilde{\boldsymbol{x}}_{i,t}\,;\boldsymbol{\eta}_u), \quad (10)$$

where the indicator function $\mathbf{1}_u(\tilde{\boldsymbol{x}}_{i,t}\,;\boldsymbol{\eta}_u) = 1$ if $\tilde{\boldsymbol{x}}_{i,t} \in R_u(\boldsymbol{\eta}_u)$ for the uth region denoted by R_u

(which is determined by the parameter vector $\boldsymbol{\eta}_u$) and zero otherwise, and $\overline{\Delta s}_u$ is the average value of the target observations in R_u for the training sample based on data through T.

A large (i.e., "deep") regression tree can often capture complex nonlinear relations in the data, but it is susceptible to overfitting due to the high variance of the tree in light of the bias-variance trade-off. To reduce the variance, a random forest averages forecasts across many deep regression trees, where each tree is constructed based on a bootstrap sample of the original data using a randomly selected subset of the predictors for each split. Using a randomly selected subset of the predictors "decorrelates" the trees to further reduce the variance. Indexing the bootstrap samples by b, the random forest forecast can be expressed as

$$\widehat{\Delta s}^{\text{RF}}_{i,T+1:T} = \frac{1}{B}\sum_{b=1}^{B}\left[\sum_{u=1}^{U}\overline{\Delta s}_u^{(b)}\mathbf{1}_u^{(b)}(\tilde{\boldsymbol{x}}_{i,t}\,;\boldsymbol{\eta}_u)\right], \tag{11}$$

where B is the number of bootstrap samples, and $\overline{\Delta s}_u^{(b)}$ and $\mathbf{1}_u^{(b)}(\tilde{\boldsymbol{x}}_{i,t}\,;\boldsymbol{\eta}_u)$ are the analogs to $\overline{\Delta s}_u$ and $\mathbf{1}_u(\tilde{\boldsymbol{x}}_{i,t}\,;\boldsymbol{\eta}_u)$, respectively, in equation (10) for the bth bootstrap sample. As with the DNN, we tune various hyperparameters for the random forest by considering a grid of values for each and using observations for the last 30% of months for the available data at the time of forecast formation as a validation sample and select the vector of hyperparameter values that minimizes the objective function for the validation sample.[5]

Table IA.11 reports out-of-sample R^2 statistics for the random forest forecast as well as an ensemble-all forecast that is an average of the baseline ENet, DNN, and random forest forecasts.

[5] We fit the random forests using the ranger package in R. We tune the following set of hyperparameters for the random forest as defined in the ranger package documentation: "num.trees," "mtry," "min.node.size," and "sample.fraction."

Overall, the random forest forecast performs reasonably well. The $R^2_{i,\text{OS}}$ statistics are positive for eight of the ten individual countries, and all positive $R^2_{i,\text{OS}}$ statistics are significant at conventional levels. Taking all of the countries together, the $R^2_{\text{All,OS}}$ statistic is 1.10% for the random forest, which is significant at the 1% level.

Compared to the baseline ENet and DNN forecasts in Table 1 from the paper, the random forest forecast in Table IA.11 generally does not perform as well, as the $R^2_{\text{All,OS}}$ statistics for the baseline ENet and DNN forecasts are larger vis-à-vis that for the random forest. A likely explanation for the better overall relative performance of the baseline ENet and DNN forecasts is due to our imposition of the zero restrictions for the country mean exchange rate changes for these forecasts, as described in Section III of the paper. The restrictions are in line with the data and motivated by the stringency of the no-change benchmark forecast in the literature. Given the noise in monthly exchange rate changes, restricting the country means to zero is highly beneficial in helping to prevent overfitting in light of the bias-variance trade-off. However, because of the nature of decision trees, restricting the country mean exchange rates to zero is not possible, so we are unable to refine the random forest forecast in this manner, which generally adversely affects out-of-sample performance compared to the baseline ENet and DNN forecasts.[6]

[6]It is interesting to note that the ensemble-all forecast in the last column of Table IA.11, which is a combination of the baseline ENet, DNN, and random forest forecasts, produces an $R^2_{\text{All,OS}}$ statistic of 2.13% (significant at the 1% level) in the last row of Table IA.11, which is very close to the corresponding $R^2_{\text{All,OS}}$ statistic (1.86%) for the ensemble forecast based on the baseline ENet and DNN forecasts in Table 1 from the paper.

F. Additional Predictors

There are additional predictors of exchange rate changes proposed in the literature that we do not include in our baseline analysis. For example, Della Corte, Jeanneret, and Patelli (2023) consider the credit-implied risk premium for 16 Euro-area countries, showing that it predicts US dollar-euro exchange rate changes on an in-sample basis (but they do not consider out-of-sample forecast accuracy). However, this measure is confined to the Euro area, so it does not provide sufficient coverage for our sample.

Various studies document a link between commodity returns and exchange rates. A number of these studies find that exchange rate changes predict commodity returns (e.g., Chen, Rogoff, and Rossi (2010)), while fewer find that commodity returns contain information about future exchange rate changes (e.g., Kohlscheen, Avalos, and Schrimpf (2017)). To account for the latter, we augment our baseline specification with a measure of commodity returns and assess its incremental predictive power.

To this end, we begin by constructing a country-level commodity price index following the methodology of Gruss and Kebhaj (2019). For each country and month, we compute a weighted average of log commodity price changes using each commodity's net export share as the weight, with the weights lagged two years to account for publication delays. This yields a terms-of-trade index that we label CXP, which can be viewed as a measure of net commodity returns. For the Euro area, we aggregate the indices of Germany, France, Italy, and the Netherlands using time-varying GDP weights, treating the bloc as a single pseudo-country from January 1999 onward. We compute the difference between the CXP for country i and the United

States. CXP enters our panel of data as an additional country characteristic for forecasting the log exchange rate change.

As a preliminary analysis, we examine (in-sample) dynamic links between exchange rate changes and CXP using Granger causality tests. Panel A of Table IA.12 reports results using country-level data with the lag orders selected via the AIC (considering a maximum of twelve), while Panel B provides results for various assumed lag orders; finally, Panel C reports results for the Dumitrescu and Hurlin (2012) panel test of Granger causality. Overall, the evidence in Table IA.12 strongly favors a unidirectional relation from exchange rate changes to net commodity returns, consistent with the prevailing view in the literature for developed countries. In Panel A, the null hypotheses that the exchange rate change does not Granger cause CXP is rejected for nine of 14 countries at conventional significance levels; the reverse null, that CXP does not Granger cause the exchange rate change, is only rejected for two countries (Canada and France at the 5% and 10% levels, respectively). Panel B confirms the asymmetry across lag orders of one, three, six, and twelve months: between 43% and 57% of countries exhibit significant Granger causality from the exchange rate change to CXP, compared to at most 14% in the reverse direction. Panel C further reinforces these findings at the panel level, with the Dumitrescu and Hurlin (2012) $\bar{\bar{Z}}$ statistics rejecting the null that the exchange rate change does not Granger cause CXP at the 1% level for all reported lag orders, while none of the statistics are significant for the reverse null.

Given that the results in Table IA.12 point to a dominant flow of information from exchange rate changes to net commodity returns (rather than the reverse), the incremental forecasting ability of CXP for exchange rate changes may be limited. To investigate this issue, Table IA.13 reports out-of-sample R^2 statistics when CXP is included among the set of country characteristics. Consistent with the Granger causality evidence, the out-of-sample gains relative to the results in Table 1 from the paper are modest and uneven across countries. The baseline ENet, DNN, and ensemble forecasts provide positive and often statistically significant improvements in forecast accuracy for commodity-exporting economies such as Australia, New Zealand, Sweden, and Norway. For a non-commodity-exporting economy such as Japan, the out-of-sample R^2 statistics are all negative in the last three columns of Table IA.13. The composite out-of-sample R^2 statistics in the last row produce statistically significant improvements in forecast accuracy of 1.07%, 0.92%, and 1.27% for the baseline ENet, DNN, and ensemble forecasts, respectively. However, these statistics do not exceed those in Table 1 from the paper, in line with the literature indicating that exchange rate changes are more important for predicting net commodity returns. Thus, we omit CXP from the main analysis in the paper.[7]

G. Alternative Benchmarks

We check the robustness of the results in Table 1 from the paper against two other benchmarks based on popular predictors from the literature. The first benchmark uses Taylor (1993) rule fundamentals. Specifically, we compute a forecast based on a linear panel predictive regression that uses the inflation differential and unemployment rate gap differential as predictors and is fitted via OLS. The second is a forecast computed from a linear panel predictive regression that uses the inflation differential and bill yield differential as predictors and again is fitted via OLS. The latter is motivated by recent studies that include similar variables in predictive regressions when analyzing currency markets (e.g., Chernov and Creal (2021), Eichenbaum, Johannsen, and Rebelo (2021), Dahlquist and Pénasse (2022), Chernov, Dahlquist, and Lochstoer

[7]Our results are in line with Jeanneret and Sokolovski (2026), who find that exchange rate predictability based on changes in country-specific commodity export or import prices is limited to illiquid emerging currencies.

(2023)).[8] As shown in Tables IA.14 and IA.15, the results are very similar to those in Table 1 from the paper, as the baseline ENet, DNN, and ensemble forecasts significantly outperform the alternative benchmarks for nearly all individual countries, often by a substantial margin, as well as when the ten countries are considered jointly.

H. Cumulative Differences in Squared Prediction Errors

To provide additional perspective on the consistency of the out-of-sample gains in Figure 1 from the paper, Figure IA.6 employs the graphical device of Goyal and Welch (2008). The figure portrays cumulative differences in squared prediction errors for the no-change benchmark forecast vis-à-vis the baseline ENet, DNN, and ensemble forecasts (considered in turn). Each curve conveniently allows for a comparison of forecast accuracy in terms of MSPE for any subsample: if the curve is higher (lower) at the end of the segment corresponding to the subsample, then the competing (benchmark) forecast has a lower MSPE for the subsample. Thus, a forecast that always outperforms the benchmark will have a curve with a uniformly positive slope. Of course, given that exchange rate changes have a large unpredictable component, this ideal is practically unattainable. Realistically, we seek a forecast with a curve that is predominantly positively sloped and does not have extended segments with negative slopes.

According to Figure IA.6, the curves are positively sloped for much of the time for the different countries, so the baseline ENet, DNN, and ensemble forecasts outperform the no-change benchmark on a consistent basis over time. By blending the baseline ENet and DNN forecasts, the ensemble approach generally provides the most consistent out-of-sample gains. The

[8]We obtain similar results when we use the real exchange rate in place of the inflation differential.

improvements in accuracy provided by the baseline ENet, DNN, and ensemble forecasts vis-à-vis the no-change benchmark are often particularly strong during the worst phase of the Global Financial Crisis in late 2008, so the information in the predictors becomes especially important during the crisis. Sizable gains are also evident before the crisis in a number of cases, and the gains are quite consistent and substantive for many countries after 2008 through the end of the out-of-sample period. Compared to the baseline ENet forecast, the DNN forecast substantially improves performance around the crisis for Switzerland, Japan, and Canada.

I. Safe-Haven Currency

A spate of recent papers (e.g., Kremens and Martin (2019), Adrian and Xie (2020), Jiang, Krishnamurthy, and Lustig (2021), Lilley, Maggiori, Neiman, and Schreger (2022)) finds evidence of exchange rate predictability around the Global Financial Crisis based on the US dollar's perception as a safe-haven currency. Due to data availability, these studies analyze predictability at a quarterly horizon or longer and/or employ relatively short samples. Although we focus on a monthly horizon and consider a longer out-of-sample period, our machine-learning forecasts appear to capture exchange rate predictability around the crisis relating to a safe-haven role for the US dollar. Specifically, the baseline ENet, DNN, and ensemble forecasts in Figures IA.2, IA.3, and IA.4, respectively, portend strong depreciations for many countries' currencies in late 2008 during the worst phase of the crisis; as shown in Figure 1 from the paper and Figure IA.6, the machine-learning forecasts substantially outperform the no-change benchmark during that time.

We also examine links between the set of predictors selected by the baseline ENet and the Lilley et al. (2022) capital flow measure, which is available at the quarterly frequency and is based on the change in US holdings of foreign bonds. Lilley et al. (2022) find that their measure is significantly related to various proxies for global risk appetite, as well as the change in a broad US dollar index from 2007 to 2017, which they interpret as evidence of the US dollar's role as a safe-haven currency during the crisis. We investigate links between the ten predictors selected by the ENet in the fitted baseline panel predictive regression based on data through 2020:08 (corresponding to the forecast for the final month, 2020:09) and the change in US foreign bond holdings.[9] We average the ten predictors over the three months comprising a quarter, as well as across countries. We then regress the change in US foreign bond holdings on the set of ten time- and country-aggregated predictors for 2007:1 to 2019:2.[10]

Figure IA.7 shows the (standardized) change in US foreign bond holdings, together with the fitted values for the regression. As a group, the ten predictors selected by the ENet in the fitted baseline panel predictive regression are significantly related to the change in US foreign bond holdings at the 5% level, and the predictors collectively explain 38% of the variation in the capital flow measure. The fitted values track the actual values quite closely during the Global Financial Crisis, providing further evidence that relevant predictors selected by the ENet contain information pertaining to a safe-haven role for the US dollar.

As discussed in Section IA.3, we also computed exchange rate forecasts with the Japanese yen replacing the US dollar as the base currency. While our machine-learning approach also works reasonably well for forecasting exchange rate changes with the Japanese yen as the base

[9]The ten predictors selected by the ENet are UN, DP, PE, INF.GVOL, UN.GVOL, BILL.GVOL, PE.EPU, PE.GILL, SRET12.GR, and IV.GCOR.

[10]Data for the change in US foreign bond holdings from Lilley et al. (2022) are available from the Global Capital Allocation Project website.

currency (see Table IA.5), the degree of exchange rate predictability is not as strong as when the US dollar is the base currency in Table 1 from the paper. This provides further evidence of a safe-haven (and/or dominant currency) role for the US dollar.

IA.5. Model Interpretation

A. Recursive Baseline ENet Estimates

As a complement to Figure 2 from the paper, Figure IA.9 depicts recursive ENet estimates of the slope coefficients for the fitted baseline panel predictive regressions used to generate the out-of-sample baseline ENet forecasts for the 1995:01 to 2020:09 out-of-sample period. For the full sample, recall that the ENet selects UN, DP, PE, INF.GVOL, UN.GVOL, BILL.GVOL, PE.EPU, PE.GILL, SRET12.GR, and IV.GCOR. Among these predictors, Figure IA.9 shows that DP, BILL.GVOL, and SRET12.GR are selected the vast majority of the time over the out-of-sample period, and the signs of their coefficient estimates always match those for the full-sample estimates. The magnitude of coefficient estimate for DP is fairly consistent over time, while that for BILL.GVOL displays a general upward trend. The coefficient for SRET12.GR increases in magnitude in the early 2000s and remains reasonably sizable through the rest of the out-of-sample period.

Turning to the other predictors selected by the ENet for the full sample, UN is consistently selected from around the mid 2000s through 2010. Thereafter, it is selected less often until toward the end of the out-of-sample period. The magnitude of the coefficient estimate for UN is largest near the Great Recession as well as the recession in early 2020 corresponding to the COVID-19 crisis. PE is selected on a regular basis starting in the early 2000s, and its coefficient estimate remains limited in magnitude over time. INF.GVOL is selected starting near the end of Great Recession and continues to be selected thereafter, with the coefficient estimate evincing a slight upward trend. While UN.GVOL is always selected from the early 2000s through the early 2010s, and its coefficient estimate is typically sizable in magnitude (especially in the wake of the Great Recession), it is only infrequently selected outside of this period, and the magnitude of the estimate is considerably smaller. PE.EPU and PE.GILL are selected starting in 2009, with the former (latter) selected sporadically (regularly) through the end of the out-of-sample period, and the magnitudes of their coefficient estimates are typically limited. Beginning in the early 2000s, IV.GCOR is almost always selected, and its coefficient estimate is quite sizable from the early 2000s through 2010 and again near the end of the out-of-sample period.

There are some predictors that are not selected by the ENet for the full sample in Figure 2 from the paper although they are earlier in the out-of-sample period in Figure IA.9. These include IV, NOTE.GVOL, and DP.GCOR, which are selected fairly frequently, as well as INF.GCOR, UN.EPU, BILL.GCOR, DP.GVOL, SRET12.MPU, IV.GR, and IV.GVOL, which are selected very rarely (and their coefficient estimates are very small in magnitude). The remaining 50 predictors are never selected by the ENet over the out-of-sample period. Figure IA.9 helps to explain the benefits of the ENet for improving out-of-sample exchange rate prediction. We cannot know a priori which country characteristics (or fundamentals) and their interactions with global variables are the most relevant for predicting exchange rate changes. Thus, we consider a large set of country characteristics and interactions, resulting in a large-dimensional predictive regression. However, estimating a large dimensional model, in conjunction with an inherently small signal-to-noise ratio, renders conventional OLS estimation highly susceptible to in-sample overfitting and thus poor out-of-sample performance. To mitigate overfitting, we employ regularized regression via the ENet, which shrinks the coefficient estimates, many of them to exactly zero, to produce a sparse model that focuses on the most relevant predictors. As shown in Section IV.C of the paper, the strategy of considering a large number of country characteristics in conjunction with machine-learning methods produces significant improvements in forecast accuracy when it comes to out-of-sample prediction of short-run exchange rate changes. According to Figure IA.9, together, UN, DP, PE, BILL.GVOL, SRET12.GR, and IV.GCOR are all consistently selected by the ENet since the early 2000s; among these predictors, DP, BILL.GVOL, SRET12.GR, and IV.GCOR also exhibit relatively large coefficient estimates on a regular basis, identifying them as the most relevant predictors of short-run exchange rate changes over the out-of-sample period.

B. DNN and Ensemble Forecasts

The ENet is a machine-learning device that facilitates model interpretation by performing model selection, and the ENet coefficient estimates for the baseline panel predictive regression in Figure 2 from the paper and Figure IA.9 are tantamount to variable-importance measures. However, when we model the conditional expectation function in equation (6) from the paper using the DNN in Section III.B of the paper, the fitted model becomes a "black box." Thus, we need to employ model-interpretation tools developed in the machine-learning literature. For the fitted DNNs used to generate the out-of-sample DNN and ensemble forecasts, we measure variable importance using the approach of Greenwell, Boehmke, and McCarthy (2018), which is based on partial dependence plots (PDPs, Friedman (2001)).

A PDP measures the relation between the expected value of the dependent variable and a given predictor for any fitted model, including "black-box" models like DNNs. Let $\hat{f}(\boldsymbol{x})$ denote the prediction function for a generic fitted model, where $\boldsymbol{x}$ denotes the vector of predictors. The PDP for predictor x_q is defined as

$$
\begin{aligned}
f_q(x_q) &= E_{\boldsymbol{x}_{C(q)}}\left[\hat{f}(x_q, \boldsymbol{x}_{C(q)})\right] \\
&= \int_{\boldsymbol{x}_{C(q)}} \hat{f}(x_q, \boldsymbol{x}_{C(q)})\, p_{C(q)}(\boldsymbol{x}_{C(q)})\, d\boldsymbol{x}_{C(q)},
\end{aligned} \tag{12}
$$

where $\boldsymbol{x}_{C(q)} = \boldsymbol{x} \setminus x_q$, and

$$
p_{C(q)}(\boldsymbol{x}_{C(q)}) = \int_{x_q} p(x_q, \boldsymbol{x}_{C(q)})\, dx_q \tag{13}
$$

is the joint marginal probability density for $\boldsymbol{x}_{C(q)}$. Equation (12) is typically estimated via Monte Carlo integration using the training data, which in our panel data setting is given by

$$
\bar{f}_q(x_q) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{f}(x_q, \boldsymbol{x}_{i,t,C(q)}). \tag{14}
$$

It is common to analyze the PDP for a grid of x_q values in the range of training-sample observations, which we denote by $\{x_{q,j}\}_{j=1}^{J}$. For the ensemble of models, the PDP is the average of the measures in equation (14) across the fitted models comprising the ensemble.

Greenwell et al. (2018) propose a PDP-based measure of variable (or predictor) importance:

$$\mathcal{I}(x_q) = \left\{[1/(J-1)]\sum_{j=1}^{J}\left[\bar{f}_q(x_{q,j}) - (1/J)\sum_{j=1}^{J}\bar{f}_q(x_{q,j})\right]^2\right\}^{0.5}. \tag{15}$$

Intuitively, equation (15) uses the sample standard deviation of the PDP to measure its "flatness." If the expected value of the target does not change as the predictor changes, then the PDP is flat, so the variable importance measure is zero. As the variability of the PDP increases, the importance measure in equation (15) likewise increases.

Figures IA.10 and IA.11 depict PDP-based variable-importance measures for the sequence of fitted prediction models underlying the DNN and ensemble forecasts, respectively. In terms of variable importance, the figures paint a broadly similar picture to that for the recursive baseline ENet estimates in Figure IA.9, including the crucial role of a number of interaction terms.[11]

IA.6. A Closer Look at Regime Switching

Section III.A of the paper explains how the interaction terms in the baseline panel predictive regression in equation (1) from the paper allow for continuous regime switching in the response of the exchange rate change to a change in a country characteristic; see equation (2) from the paper. Section V of the paper shows that the interaction terms, especially BILL.GVOL, play important roles in the fitted prediction models. Motivated by our findings, as well as evidence in Clarida, Davis, and Pedersen (2009), we explore regime switching in more detail in

[11]The variable-importance measures in Figures IA.10 and IA.11 are always positive by construction. They can be compared to the absolute values of the recursive baseline ENet estimates in Figure IA.9.

this section, focusing on GVOL as the global variable governing regime switching. We augment the interaction terms with threshold or logistic smooth transition mechanisms, thereby considering richer forms of regime switching. For consistency, we continue to specify the panel predictive regressions in terms of standardized predictors.

A. Interaction

Consider the following panel predictive regression with an interaction term:

$$\Delta s_{i,t+1} = \beta^z \underbrace{\frac{1}{\sigma_i^z}\left(z_{i,t} - \mu_i^z\right)}_{\tilde{z}_{i,t}} + \beta^{z,g} \underbrace{\frac{1}{\sigma_i^{z,g}}\left(z_{i,t}\, g_t - \mu_i^{z,g}\right)}_{\widetilde{z_{i,t}\, g_t}} + \varepsilon_{i,t+1}, \tag{16}$$

where μ_i^z ($\mu_i^{z,g}$) is the mean of $z_{i,t}$ ($z_{i,t}\, g_t$), and σ_i^z ($\sigma_i^{z,g}$) is the standard deviation of $z_{i,t}$ ($z_{i,t}\, g_t$). For notational simplicity, we consider only a single country characteristic and a single global variable; it is straightforward to extend equation (16) to include additional country characteristics and global variables, as in equation (1) from the paper. We can re-write equation (16) as

$$\Delta s_{i,t+1} = \underbrace{\left(\beta^z \frac{1}{\sigma_i^z} + \beta^{z,g} \frac{1}{\sigma_i^{z,g}}\, g_t\right)}_{\text{slope coefficient for } z_{i,t}} z_{i,t} - \left(\beta^z \frac{1}{\sigma_i^z}\, \mu_i^z + \beta^{z,g} \frac{1}{\sigma_i^{z,g}}\, \mu_i^{z,g}\right) + \varepsilon_{i,t+1}. \tag{17}$$

Equation (17) shows that the inclusion of the interaction term in equation (16) is equivalent to allowing for the slope coefficient for $z_{i,t}$ to vary continuously with g_t in a linear manner, which is what we term continuous regime switching in Section III.A of the paper. To compare different regime-switching specifications, as in equation (2) from the paper, it is useful to consider $\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}}$, which, for equation (17), is given by

$$\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}} = \beta^{z}\,\frac{1}{\sigma_i^{z}} + \beta^{z,g}\,\frac{1}{\sigma_i^{z,g}}\,g_t. \tag{18}$$

If $\beta^{z,g} > 0$, then $\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}}$ increases linearly with g_t. For example, if $z_{i,t}$ is $\text{BILL}_{i,t}$, g_t is GVOL_t, and $\beta^{z,g} > 0$, then the slope coefficient for $\text{BILL}_{i,t}$ increases continuously with GVOL_t, as discussed in Section V of the paper.

B. Interaction with Threshold

Next, we consider an interaction specification with a threshold mechanism:

$$\begin{aligned}\Delta s_{i,t+1} &= \beta^{z}\,\underbrace{\frac{1}{\sigma_i^{z}}\big(z_{i,t} - \mu_i^{z}\big)}_{\tilde{z}_{i,t}} + \beta^{z,g}\,\underbrace{\frac{1}{\sigma_i^{z,g}}\big[z_{i,t}\,g_t\,\mathbb{I}(g_t\,;g^{\star}) - \mu_i^{z,g}\big]}_{\widetilde{z_{i,t}\,g_t\,\mathbb{I}(g_t;\,g^{\star})}} + \varepsilon_{i,t+1}\\ &= \underbrace{\left[\beta^{z}\,\frac{1}{\sigma_i^{z}} + \beta^{z,g}\,\frac{1}{\sigma_i^{z,g}}\,g_t\,\mathbb{I}(g_t\,;g^{\star})\right]}_{\text{slope coefficient for } z_{i,t}} z_{i,t} - \left(\beta^{z}\,\frac{1}{\sigma_i^{z}}\,\mu_i^{z} + \beta^{z,g}\,\frac{1}{\sigma_i^{z,g}}\,\mu_i^{z,g}\right) + \varepsilon_{i,t+1},\end{aligned} \tag{19}$$

where $\mathbb{I}(g_t\,;g^{\star}) = 1$ if $g_t \geq g^{\star}$ and zero otherwise, $g^{\star}$ is the threshold parameter for the global variable, and $\mu_i^{z,g}$ ($\sigma_i^{z,g}$) is the mean (standard deviation) of $z_{i,t}\,g_t\,\mathbb{I}(g_t\,;g^{\star})$. Based on equation (19),

$$\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}} = \beta^{z}\,\frac{1}{\sigma_i^{z}} + \beta^{z,g}\,\frac{1}{\sigma_i^{z,g}}\,g_t\,\mathbb{I}(g_t\,;g^{\star}). \tag{20}$$

According to equation (20) and assuming that $\beta^{z,g} > 0$, $\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}}$ is constant at $\beta^{z}\frac{1}{\sigma_i^{z}}$ for $g_t < g^{\star}$, jumps to $\beta^{z}\frac{1}{\sigma_i^{z}} + \beta^{z,g}\frac{1}{\sigma_i^{z,g}}g^{\star}$ at $g_t = g^{\star}$, and increases linearly with g_t (with a slope of $\beta^{z,g}\frac{1}{\sigma_i^{z,g}}$) for

$g_t > g^\star$. If we again assume that $z_{i,t}$ is $\text{BILL}_{i,t}$, g_t is GVOL_t, and $\beta^{z,g} > 0$, then the slope coefficient for $\text{BILL}_{i,t}$ is constant for $\text{GVOL}_t < \text{GVOL}^\star$, jumps to a higher value when $\text{GVOL}_t = \text{GVOL}^\star$, and increases continuously with GVOL_t for $\text{GVOL}_t > \text{GVOL}^\star$.[12]

C. Interaction with Logistic Smooth Transition

Analogously to combining the interaction specification with a threshold mechanism in equation (19), we also consider combining the interaction specification with a logistic smooth transition (LST) mechanism:

$$
\begin{aligned}
\Delta s_{i,t+1} &= \beta^z \underbrace{\frac{1}{\sigma_i^z}\big(z_{i,t} - \mu_i^z\big)}_{\tilde{z}_{i,t}} + \beta^{z,g} \underbrace{\frac{1}{\sigma_i^{z,g}}\big[z_{i,t}\, g_t\, \mathbb{L}(g_t\,;\kappa) - \mu_i^{z,g}\big]}_{\widetilde{z_{i,t}\, g_t\, \mathbb{L}(g_t;\kappa)}} + \varepsilon_{i,t+1} \\
&= \underbrace{\left[\beta^z \frac{1}{\sigma_i^z} + \beta^{z,g} \frac{1}{\sigma_i^{z,g}}\, g_t\, \mathbb{L}(g_t\,;\kappa)\right]}_{\text{slope coefficient for } z_{i,t}} z_{i,t} - \left(\beta^z \frac{1}{\sigma_i^z}\, \mu_i^z + \beta^{z,g} \frac{1}{\sigma_i^{z,g}}\, \mu_i^{z,g}\right) + \varepsilon_{i,t+1},
\end{aligned}
\tag{21}
$$

where $\mathbb{L}(g_t\,;\kappa) = \big[1 + \exp\big(-\kappa \frac{1}{\sigma_g}(g_t - \mu_g)\big)\big]^{-1}$ for $\kappa > 0$ is the logistic function, μ_g (σ_g) is the mean (standard deviation) of g_t, and $\mu_i^{z,g}$ ($\sigma_i^{z,g}$) is the mean (standard deviation) of $z_{i,t}\, g_t\, \mathbb{L}(g_t\,;\kappa)$. The curvature parameter κ controls how sharply the transition occurs around μ_g: larger values of κ produce a sharper transition, while smaller values produce a more gradual one. Based on

[12]When $z_{i,t}\, g_t\, \mathbb{I}(g_t;\, g^\star)$ is simplified to $z_{i,t}\, \mathbb{I}(g_t;\, g^\star)$ in equation (19), the panel predictive regression becomes a conventional threshold specification. In this case, $\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}}$ is a step function: the slope coefficient equals $\beta^z \frac{1}{\sigma_i^z}$ ($\beta^z \frac{1}{\sigma_i^z} + \beta^{z,g} \frac{1}{\sigma_i^{z,g}}$) for $g_t < g^\star$ ($g_t \geq g^\star$).

equation (21),

$$\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}} = \beta^{z}\,\frac{1}{\sigma_i^{z}} + \beta^{z,g}\,\frac{1}{\sigma_i^{z,g}}\,g_t\,\mathbb{L}(g_t\,;\kappa). \tag{22}$$

As $g_t \to -\infty$, $\mathbb{L}(g_t\,;\kappa) \to 0$, so $\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}}$ approaches $\beta^{z}\frac{1}{\sigma_i^{z}}$. For $\beta^{z,g} > 0$, as g_t increases, $\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}}$ in equation (22) smoothly increases according to $g_t\,\mathbb{L}(g_t\,;\kappa)$. As $g_t \to \infty$, $\mathbb{L}(g_t\,;\kappa) \to 1$; thus, in the limit, $\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}}$ increases linearly as g_t increases (with slope $\beta^{z,g}\frac{1}{\sigma_i^{z,g}}$). In addition, as $\kappa \to \infty$, $\frac{\partial \Delta s_{i,t+1}}{\partial z_{i,t}}$ in equation (22) approaches the expression in equation (20) (for $g^{\star} = \mu_g$). Continuing to assume that $z_{i,t}$ is $\text{BILL}_{i,t}$, g_t is GVOL_t, and $\beta^{z,g} > 0$, the slope coefficient for $\text{BILL}_{i,t}$ approaches a constant value as $\text{GVOL}_t \to -\infty$; as GVOL_t increases, the slope coefficient continuously increases according to $\text{GVOL}_t\,\mathbb{L}(\text{GVOL}_t\,;\kappa)$; as $\text{GVOL}_t \to \infty$, the slope increases linearly.[13]

D. Out-of-Sample Results

To compare the different regime-switching specifications in Sections IA.6.A, IA.6.B, and IA.6.C, we generate out-of-sample exchange rate change forecasts based on the different specifications, where the panel predictive regressions include all ten country characteristics as predictors as well as each characteristic interacted with GVOL. Because each panel predictive regression has a relatively large number of predictors (20) and given the inherently large unpredictable component in exchange rate changes, we continue to generate forecasts by estimating the panel predictive regressions via the ENet.

[13]When $z_{i,t}\,g_t\,\mathbb{L}(g_t;\,\kappa)$ is simplified to $z_{i,t}\,\mathbb{L}(g_t;\,\kappa)$ in equation (21), the panel predictive regression becomes a conventional logistic smooth transition specification, with the slope coefficient taking the familiar S-shape with a minimum (maximum) value of $\beta^{z}\frac{1}{\sigma_i^{z}}$ ($\beta^{z}\frac{1}{\sigma_i^{z}} + \beta^{z,g}\frac{1}{\sigma_i^{z,g}}$).

When estimating the interaction with threshold specification in Section IA.6.B, we need to determine a value for the threshold $g^\star$. Recall from Section III.A of the paper that ENet estimation requires tuning the hyperparameters λ and δ, for which we use the ERIC. Our strategy is to treat $g^\star$ as an additional hyperparameter to tune; that is, we use the ERIC to tune λ, δ, and $g^\star$ jointly for ENet estimation. This has the advantage of setting $g^\star$ to improve out-of-sample performance in light of the bias-variance trade-off. We follow a similar strategy for estimating the interaction with LST specification in Section IA.6.C, tuning λ, δ, and the LST curvature parameter κ jointly via the ERIC. To our knowledge, our paper is the first to propose setting the threshold or curvature parameter ($g^\star$ or κ, respectively) in the transition function for regime-switching models by treating them as additional hyperparameters to be tuned. Thus, we make a methodological contribution to out-of-sample forecasting with regime-switching specifications for models that are trained using machine-learning methods such as the ENet.

Table IA.16 reports the out-of-sample results. For reference, the fourth column reproduces the results from the sixth column of Table 1 from the paper, which correspond to the panel prediction regression with the ten country characteristics only (i.e., no interaction terms). According to the results in the fifth column of Table IA.16, including interaction terms with GVOL in the panel predictive regression substantially improves out-of-sample performance: the $R^2_{\text{All,OS}}$ statistic goes from negative (-0.29%) to 0.61% (significant at the 5% level), and nearly all of the country-level $R^2_{i,\text{OS}}$ statistics increase (some markedly, e.g., the United Kingdom and New Zealand). The last two columns of Table IA.16 show that augmenting the interaction terms with threshold or LST mechanisms generally produces additional out-of-sample gains. The $R^2_{\text{All,OS}}$ statistics are 1.02% and 0.99% for the interaction with threshold and interaction with LST specifications, respectively (both significant at the 5% level). For countries such as Switzerland, Australia, New Zealand, Sweden, Norway, Denmark, and the Euro area, the $R^2_{i,\text{OS}}$ statistics evince considerable increases in the last two columns.[14] Overall, the results in Table IA.16 highlight the relevance of GVOL for capturing FX market conditions that affect the responses of exchange rate changes to economic fundamentals, especially BILL, reiterating the findings in Sections IV.D and V of the paper.

IA.7. Robustness Checks: Portfolio Performance

A. Allocating Between the US Dollar and an Individual Foreign Currency

We consider a scenario where a mean-variance investor allocates monthly between the US dollar and an individual foreign currency i. At the end of month T, the investor's objective function is given by

$$\max_{w_{i,T+1|T}} w_{i,T+1|T}\, \widehat{rx}_{i,T+1|T} - 0.5\gamma\, w^2_{i,T+1|T}\, \hat{\sigma}^2_{i,T+1|T}, \tag{23}$$

where $\widehat{rx}_{i,T+1|T}$ is given by equation (14) from the paper, $w_{i,T+1|T}$ is the allocation to currency i that will be in effect for month $T+1$ (the US dollar allocation is $1 - w_{i,T+1|T}$), and $\hat{\sigma}^2_{i,T+1|T}$ is the investor's estimate of the variance for the currency excess return. We assume that the investor uses an exponentially weighted moving average (EWMA) estimator for $\hat{\sigma}^2_{i,T+1|T}$ (with a decay factor of 0.92). As in the paper, we assume that $\gamma = 5$. To keep the allocation in a range that appears

[14]Conventional threshold and LST specifications (footnotes 12 and 13, respectively) generally do not outperform the no-change benchmark forecast. Thus, allowing for interactions appears important for short-run out-of-sample exchange rate prediction.

plausible for an investor allocating to a single asset, we impose the restriction that $-1 \leq w_{i,T+1|T} \leq 1.5$, similarly to related studies on stock return predictability.

We consider the two cases from Section VI.A in the paper. In the first case, the Bench-Opt (benchmark optimized) portfolio, the investor sets the expected exchange rate change to the no-change benchmark forecast ($\widehat{\Delta} s_{i,T+1|T} = 0$) in equation (14) from the paper. In the second case, the ML-Opt (machine-learning optimized) portfolio, the investor uses the baseline ENet, DNN, or ensemble forecast for the exchange rate change in equation (14) from the paper. To measure the economic value of a machine-learning forecast to the investor, we compute the annualized average utility gain in equation (15) from the paper.

Table IA.17 reports the annualized average utility gain when the investor allocates between the US dollar and an individual foreign currency. The gains accruing to the investor when they use the ML-Opt instead of the Bench-Opt portfolio are all positive and typically quite sizable. The clear majority are above 200 basis points and a number exceed 300 basis points. Table IA.17 demonstrates that the improvements in the statistical accuracy provided by the machine-learning forecasts, as reflected by the $R^2_{i,\mathrm{OS}}$ statistics in Table 1 from the paper, also produce substantial economic value from an investment perspective. The results demonstrate the value of the forecasts at the individual currency level in terms of direction (i.e., when to buy and sell a currency) as well as magnitude.

B. Transaction Costs

We examine the effects of transaction costs by computing portfolio excess returns using bid and ask quotes from Datastream for the relevant forward and spot rates (e.g., Lustig et al.

(2011)). In this case, the investor is assumed to pay the bid and ask prices reported in Datastream. A number of studies document that the bid-ask spreads offered by Datastream are unrealistically high (e.g., Lyons (2001), Neely, Weller, and Ulrich (2009), Menkhoff, Sarno, Schmeling, and Schrimpf (2012b), Neely and Weller (2013)). Specifically, investors trade at the best quoted price at each point in time, making the full spread in Datastream considerably higher than the effective spread for FX market participants.[15] To more accurately reflect the relevant transaction costs faced by traders, we follow Goyal and Saretto (2009) and Menkhoff et al. (2012b) and also report results for currency excess returns computed using 25%, 50%, and 75% of the quoted bid-ask spreads from Datastream.

Table IA.19 reports portfolio performance measures adjusted for transaction costs for the 1995:01 to 2020:09 out-of-sample period. As expected, the annualized Sharpe ratios decrease monotonically as we move from Panel A (25% of the bid-ask spread) to Panel D (full bid-ask spread). Nevertheless, they remain quite sizable for the ML-Opt portfolios: they are greater than or equal to 0.69 (0.61) for 25% (50%) of the bid-ask spread in Panel A (B); for 75% of the bid-ask spread in Panel C, they still range from 0.53 (DNN) to 0.73 (baseline ENet).[16] The ML-Opt portfolios also continue to provide substantive economic value to the investor vis-à-vis the Bench-Opt portfolio in terms of the annualized average utility gains in the second column. Even for 75% of the bid-ask spread, the gains range from 170 (DNN) to 424 (baseline ENet) basis

[15] The FX market is one of the most liquid markets, with low transaction costs and no natural short-selling constraints. According to the 2016 Bank for International Settlements Triennial Survey, average daily turnover in the FX market is five trillion US dollars.

[16] With the exception of the DNN for the full bid-ask spread in Panel D, all of the Sharpe ratios for the ML-Opt portfolios are significant at the 1% level in Panels A through D of Table IA.19.

points. In sum, the performance of the ML-Opt portfolios—and thus the economic value of out-of-sample exchange rate predictability—is robust to reasonable assumptions regarding transaction costs.

C. A Closer Look at Late 2008

Figure 3 from the paper indicates that late 2008 is a perilous time for the Bench-Opt portfolio (as well as the conventional carry portfolio). Figures IA.15 to IA.20 provide additional insight into the sources of the poor performance of the Bench-Opt portfolio in late 2008 as well as how the ML-Opt portfolios improve performance. Figure IA.15 to IA.17 show the currency excess return forecasts for September through December of 2008 for the baseline ENet, DNN, and ensemble log exchange rate change forecasts, respectively, that serve as inputs in the portfolio optimization problem in equation (13) from the paper. The figures also show the currency excess return forecasts for the Bench-Opt portfolio along with the realized excess returns. Because the Bench-Opt portfolio uses the no-change benchmark exchange rate forecast, the benchmark currency excess return forecast in equation (14) from the paper is simply the bill yield differential; the ML-Opt portfolio augments the bill yield differential with the baseline ENet, DNN, or ensemble exchange rate change forecast. Figures IA.18 to IA.20 display the currency weights for the portfolios.

For September of 2008, the machine-learning and benchmark currency excess return forecasts in Figures IA.15 to IA.17 lead to allocations of similar sizes in Figures IA.18 to IA.20 for most countries. The machine-learning currency excess return forecasts generate a notable difference in allocation for the Euro area, where the ML-Opt portfolios (Bench-Opt portfolio) take short (long) positions; the ML-Opt portfolios based on the DNN and ensemble forecasts take long positions for Japan, while the Bench-Opt portfolio takes a short position. As shown in Figures IA.15 to IA.17, with the exceptions of Japan and Canada, all of the realized currency excess returns are negative for September of 2008; the negative returns are large in magnitude for Australia, New Zealand, Sweden, Norway, Denmark, and the Euro area. On the basis of the allocations and realized currency excess returns, the excess return for the Bench-Opt portfolio is -7.57% in September of 2008, while the losses are smaller for the ML-Opt portfolio, ranging from -4.02% (ensemble) to -5.25% (baseline ENet). The differences in allocations signaled by the machine-learning forecasts—especially for the Euro area—help to limit portfolio losses in the month of the Lehman bankruptcy.

The currency excess return forecasts diverge more sharply during October of 2008 in Figures IA.15 to IA.17: with the exception of Japan, the benchmark currency excess return forecasts are all positive, while the machine-learning forecasts are negative for all ten countries for the baseline ENet forecast and nearly all negative for the DNN and ensemble forecasts. The differences in currency excess return forecasts give rise to a number of markedly different allocations during October of 2008 in Figures IA.18 to IA.20; for example, the ML-Opt (Bench-Opt) portfolios based on the baseline ENet and ensemble forecasts exhibit sizable short (long) positions for the United Kingdom, New Zealand, Norway, Denmark, and the Euro area. With the exception of Japan, all of the realized currency excess returns are negative for October of 2008 in Figures IA.15 to IA.17, and the negative returns are larger in magnitude than those for September of 2008.[17] The different allocations prompted by the machine-learning forecasts

[17]Note the difference in scales for the vertical axes across the top two panels of Figures IA.15 to IA.17.

vis-à-vis the no-change benchmark enable the ML-Opt portfolios to substantively outperform the Bench-Opt portfolio during October of 2008: the latter suffers a large loss of -11.60%, while the ML-Opt portfolios realize substantial gains of 6.29% (DNN) to 24.26% (baseline ENet). The situation for November of 2008 is fairly similar to that for October in terms of the currency excess return forecasts and allocations in Figures IA.15 to IA.20. Because the negative realized currency excess returns are typically smaller in magnitude, the Bench-Opt realizes an excess return of 1.91% during the month; however, the information in the machine-learning forecasts leads to much larger excess returns for the ML-Opt portfolios, ranging from 12.36% (ensemble) to 12.90% (DNN).

The environment appears to normalize to an extent in December of 2008, in the sense that the discrepancies between the machine-learning and benchmark forecasts in Figures IA.15 to IA.17, as well as those between the ML-Opt and Bench-Opt portfolio weights in Figures IA.18 to IA.20, are more muted. Nevertheless, important differences remain (e.g., the portfolio weights for Switzerland and Norway). In addition, with the exception of the United Kingdom, all the realized currency excess returns are positive in December of 2008. The Bench-Opt portfolio experiences an excess return of 5.64% for the month, while the excess returns are around three times as large for the ML-Opt portfolios (19.07%, 15.93%, and 17.58% for the baseline ENet, DNN, and ensemble, respectively).[18]

Figures IA.15 to IA.20 help to explain how exchange rate predictability—as captured by our machine-learning forecasts—is valuable to an investor during the worst part of the Global

[18]For September through December of 2008, the excess returns for the conventional carry trade portfolio are -6.33%, -15.00%, -3.06%, and -4.77%, respectively, in line with the sharp drop in the portfolio's cumulative excess return in Figure 3 from the paper.

Financial Crisis in late 2008. By anticipating depreciations in many foreign currencies, the machine-learning forecasts lead to sizable negative positions in many currencies, enabling the ML-Opt portfolios to avoid the large losses suffered by a Bench-Opt portfolio and even realize substantive gains.

D. Alphas

We examine whether the ML-Opt and Bench-Opt portfolios generate alpha in the context of the Lustig et al. (2011) currency factor model. The model includes dollar and carry trade risk factors, denoted by MKT_{FX} and HML_{FX}, respectively. The dollar factor is an equally weighted average of the available currency excess returns for the month, while the carry trade risk factor is the conventional carry portfolio defined in the paper. Table IA.20 reports factor model estimation results for the ML-Opt and Bench-Opt portfolios. We again report results for the full out-of-sample period as well as the 1995:01 to 2008:08 and 2008:09 to 2020:09 subsamples.

For the full 1995:01 to 2020:09 out-of-sample period in Panel A of Table IA.20, the ML-Opt portfolios generate large annualized alphas of 11.53%, 7.98%, and 9.51% for the baseline ENet, DNN, and ensemble forecasts, respectively, all of which are significant at the 1% level and substantially larger than that for the Bench-Opt portfolio (4.27%). The Bench-Opt portfolio evinces considerable exposure to the carry factor (0.75, significant at the 1% level) for the full out-of-sample period. The ML-Opt portfolios also exhibit sizable exposures to the carry factor (0.40 for all three forecasts, which are significant at the 5% or 1% level), but they are about half as large as that for the Bench-Opt portfolio.

Panels B and C of Table IA.20 reveal important differences in performance for the

Bench-Opt portfolio across the two subsamples. For the 1995:01 to 2008:08 subsample in Panel B, the Bench-Opt portfolio exhibits substantial exposure to the carry trade factor (0.92, significant at the 1% level), and it generates a sizable annualized alpha of 4.51% (significant at the 1% level). Reminiscent of Table 3 and Figure 3 from the paper, the Bench-Opt portfolio's performance deteriorates sharply for the 2008:09 to 2020:09 subsample in Panel C. It continues to display substantial exposure to the carry factor (0.64, significant at the 1% level), while its annualized alpha declines to only 1.71% (insignificant at conventional levels). In contrast, the ML-Opt portfolios deliver impressive annualized alphas for both subsamples in Panels B and C of Table IA.20, all of which are above 585 basis points and significant. An interesting pattern emerges with respect to the exposures of the ML-Opt portfolios to the carry factor. The exposures are quite sizable for the first subsample, ranging from 0.64 (DNN) to 0.76 (baseline ENet), and all are significant at the 1% level. For the second subsample, the exposures become much smaller, ranging from 0.02 (baseline ENet) to 0.19 (DNN), and none is significant. Thus, the information contained in the 70 predictors leads the investor to effectively disconnect fully from a conventional carry strategy for the second subsample. Taken together, the pre- and post-2008 exposures provide direct evidence that the model picks up time-varying risk exposures.

E. Conditional Maximum Sharpe Ratios

Finally, we examine the economic implications of our machine-learning forecasts for the conditional maximum Sharpe ratio (CMSR), which relates to the conditional price of currency risk. Chernov et al. (2023) recently emphasize that it is feasible to construct the ex ante mean-variance efficient portfolio for individual currencies, as the cross section of currencies is relatively small compared to that for individual equities. The sheer size of the cross section for equities makes estimating the covariance matrix with reasonable precision problematic, so it is practically infeasible to compute a reliable mean-variance efficient portfolio for asset pricing based on individual stocks. In contrast, constructing a reliable covariance matrix estimate is eminently workable for the limited cross section of individual currencies. Combined with a vector of currency excess return forecasts, we can construct an ex ante conditional mean-variance efficient (CMVE) portfolio based on information through month t.

Along the lines of studies such as Ackermann, Pohl, and Schmedders (2017), Daniel, Hodrick, and Lu (2017), Chernov et al. (2023), and Maurer, To, and Tran (2023), we construct a CMVE portfolio by computing the portfolio that is fully invested in the individual foreign currencies and maximizes the expected Sharpe ratio (i.e., the tangency portfolio). Specifically, to estimate the CMSR for month $T+1$ based on information available through month T, we compute portfolio weights based on the following objective function:

$$\max_{\boldsymbol{w}_{T+1|T}} \left(\boldsymbol{w}'_{T+1|T}\hat{\boldsymbol{\mu}}_{T+1|T}\right)\left(\boldsymbol{w}'_{T+1|T}\hat{\boldsymbol{\Sigma}}\boldsymbol{w}_{T+1|T}\right)^{-0.5} \quad \text{subject to } \boldsymbol{\iota}'_N\boldsymbol{w}_{T+1|T} = 1, \tag{24}$$

where $\hat{\boldsymbol{\mu}}_{T+1|T}$ is the vector of currency excess return forecasts, $\hat{\boldsymbol{\Sigma}}$ is the estimate of the covariance matrix for the excess returns, and $\boldsymbol{\iota}_N$ is an N-dimensional vector of ones. To ensure that the portfolio weights lie in reasonable ranges, we also impose the restrictions that $-1 \leq w_{i,T+1|T} \leq 1$ for $i = 1, \ldots, N$. Denoting the solution by $\boldsymbol{w}^*_{T+1|T}$, the CMSR is given by

$$\text{CMSR}_{T+1|T} = \left(\boldsymbol{w}^{*\prime}_{T+1|T}\,\hat{\boldsymbol{\mu}}_{T+1|T}\right)\left(\boldsymbol{w}^{*\prime}_{T+1|T}\,\hat{\boldsymbol{\Sigma}}_{T+1|T}\,\boldsymbol{w}^*_{T+1|T}\right)^{-0.5}, \tag{25}$$

which can be viewed as a measure of the conditional price of currency risk. We are interested in how the information in our broad array of predictors—as embodied in the machine-learning exchange rate forecasts—affects the CMSR and price of currency risk. Because our estimates of the CMSR are out-of-sample, they are more relevant for investors in real time.

Figure IA.21 plots the annualized CMSR when $\hat{\boldsymbol{\mu}}_{T+1|T}$ uses the baseline ENet, DNN, ensemble and no-change benchmark forecasts of the exchange rate change (considered in turn) when forming the currency excess return forecasts. As in Section VI of the paper, $\hat{\boldsymbol{\Sigma}}$ is based on an EWMA estimator. The CMSRs based on the machine-learning forecasts move together fairly closely over time, ranging from approximately 0.3 to over four. They indicate an economically sizable CMSR as well as economically significant fluctuations in the conditional price of currency risk. Periods with an elevated CMSR based on the machine-learning forecasts are often associated with times of stress in the global economy. For example, the CMSRs evince marked increases in the mid-to-late 2000s during the Global Financial Crisis and concomitant Great Recession. Sharp increases are also evident in early 2020 corresponding to the advent of the COVID-19 pandemic. Compared to those based on the machine-learning forecasts, the CMSR based on the no-change benchmark forecast tends to be lower and less volatile. Overall, our approach for forecasting exchange rates, which employs a large information set and machine learning, yields a sensibly behaved conditional price of currency risk, with notable increases during periods of global economic stress.

F. Portfolio Performance and Global FX Volatility

In Section IV.D of the paper, we examine the relation between global FX volatility and the statistical accuracy of the machine-learning forecasts relative to the no-change benchmark. The results in Table 2 from the paper provide extensive evidence that the relative accuracy of the machine-learning forecasts significantly improves as global FX volatility increases. Analogously, we investigate the relation between global FX volatility and the economic value of the machine-learning forecasts vis-à-vis the no-change benchmark. To do so, equation (11) from the paper becomes

$$rx_{i,t}^{\text{ML-Opt}} - rx_{i,t}^{\text{Bench-Opt}} = a_{i,0} + a_{i,1}\text{GVOL}_t + \varepsilon_{i,t}, \tag{26}$$

where $rx_{i,t}^{\text{ML-Opt}}$ ($rx_{i,t}^{\text{Bench-Opt}}$) is the month-$t$ excess return for the mean-variance US investor who allocates monthly between the US dollar and the foreign currency for country i using a machine-learning exchange rate change (no-change benchmark) forecast, corresponding to the objective function in equation (23). We also estimate a version of equation (26) for the scenario where the US investor allocates monthly across the US dollar and all available foreign currencies, corresponding to the objective function in equation (13) from the paper:

$$rx_t^{\text{ML-Opt}} - rx_t^{\text{Bench-Opt}} = a_0 + a_1\text{GVOL}_t + \varepsilon_t, \tag{27}$$

where $rx_t^{\text{ML-Opt}}$ ($rx_t^{\text{Bench-Opt}}$) is the month-t excess return for the mean-variance portfolio based on a machine-learning exchange rate change (no-change benchmark) forecast. To facilitate the interpretation of the results, we standardize GVOL_t in equations (26) and (27).

Table IA.21 reports the results. With the exception of Japan for the baseline ENet forecast, all of the slope coefficient estimates are positive and significant (nearly all at the 1% level). Thus, we have strong evidence that the performance of the ML-Opt portfolios significantly improves relative to the Bench-Opt portfolios as GVOL increases, analogously to the extensive evidence in Table 2 from the paper that the statistical accuracy of the machine-learning forecasts significantly improves relative to the no-change benchmark as GVOL increases. The magnitudes of the slope coefficient estimates are also economically sizable. For example, for the portfolios that invest in all available foreign currencies and the baseline ENet forecast, a one-standard-deviation increase in GVOL is associated on average with a 131 basis point increase in the monthly excess return difference between the ML-Opt and Bench-Opt portfolios. Many of the other slope coefficients are around 50 basis points or larger, again signaling quite large increases in excess return differences on average.

References

Ackermann, F.; W. Pohl; and K. Schmedders. "Optimal and Naive Diversification in Currency Markets." *Management Science*, 63 (2017), 3147–3529.

Adrian, T., and P. Xie. "The Non-U.S. Bank Demand for U.S. Dollar Assets." IMF Working Paper No. 20/101 (2020).

Baker, S. R.; N. Bloom; and S. J. Davis. "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics*, 131 (2016), 1593–1636.

Bao, Y. "Estimation Risk-Adjusted Sharpe Ratio and Fund Performance Ranking under a General Return Distribution." *Journal of Financial Econometrics*, 7 (2009), 152–173.

Boyer, B.; T. Mitton; and K. Vorkink. "Expected Idiosyncratic Skewness." *Review of Financial Studies*, 23 (2010), 169–202.

Breiman, L. “Random Forests.” *Machine Learning*, 45 (2001), 5–32.

Breiman, L.; J. Friedman; C. J. Stone; and R. A. Olshen. *Classification and Regression Trees*. CRC Press, New York (1984).

Caldara, D., and M. Iacoviello. “Measuring Geopolitical Risk.” *American Economic Review*, 112 (2022), 1194–1225.

Chen, Y.-C.; K. S. Rogoff; and B. Rossi. “Can Exchange Rates Forecast Commodity Prices?” *Quarterly Journal of Economics*, 125 (2010), 1145–1194.

Chen, Z., and R. Petkova. “Does Idiosyncratic Volatility Proxy for Risk Exposure?” *Review of Financial Studies*, 25 (2012), 2745–2787.

Chernov, M., and D. Creal. “The PPP View of Multihorizon Currency Risk Premiums.” *Review of Financial Studies*, 34 (2021), 2728–2772.

Chernov, M.; M. Dahlquist; and L. A. Lochstoer. “Pricing Currency Risks.” *Journal of Finance*, 78 (2023), 693–730.

Christiano, L. J., and T. J. Fitzgerald. “The Band Pass Filter.” *International Economic Review*, 44 (2003), 435–465.

Clarida, R. H.; J. Davis; and N. Pedersen. “Currency Carry Trade Regimes: Beyond the Fama Regression.” *Journal of International Money and Finance*, 28 (2009), 1375–1389.

Clark, T. E., and K. D. West. “Approximately Normal Tests for Equal Predictive Accuracy in Nested Models.” *Journal of Econometrics*, 138 (2007), 291–311.

Dahlquist, M., and J. Pénasse. “The Missing Risk Premium in Exchange Rates.” *Journal of Financial Economics*, 143 (2022), 697–715.

Daniel, K.; R. J. Hodrick; and Z. Lu. “The Carry Trade: Risks and Drawdowns.” *Critical Finance Review*, 6 (2017), 211–262.

Della Corte, P.; A. Jeanneret; and E. Patelli. “A Credit-Based Theory of the Currency Risk Premium.” *Journal of Financial Economics*, 149 (2023), 473–496.

Dumitrescu, E.-I., and C. Hurlin. "Testing for Granger Non-Causality in Heterogeneous Panels." *Economic Modelling*, 29 (2012), 1450–1460.

Eichenbaum, M. S.; B. K. Johannsen; and S. T. Rebelo. "Monetary Policy and the Predictability of Nominal Exchange Rates." *Review of Economic Studies*, 88 (2021), 192–228.

Filippou, I.; A. E. Gozluklu; and M. P. Taylor. "Global Political Risk and Currency Momentum." *Journal of Financial and Quantitative Analysis*, 53 (2018), 2227–2259.

Friedman, J. H. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics*, 29 (2001), 1189–1232.

Fu, F. "Idiosyncratic Risk and the Cross-Section of Expected Stock Returns." *Journal of Financial Economics*, 91 (2009), 24–27.

Goyal, A., and P. Santa-Clara. "Idiosyncratic Risk Matters!" *Journal of Finance*, 58 (2003), 975–1007.

Goyal, A., and A. Saretto. "Cross-Section of Option Returns and Volatility." *Journal of Financial Economics*, 94 (2009), 310–326.

Goyal, A., and I. Welch. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies*, 21 (2008), 1455–1508.

Greenwell, B. M.; B. C. Boehmke; and A. J. McCarthy. "A Simple and Effective Model-Based Variable Importance Measure." Working Paper arXiv:1805.04755v1 (2018).

Gruss, B., and S. Kebhaj. "Commodity Terms of Trade: A New Database." IMF Working Paper No. 19/21 (2019).

Jeanneret, A., and V. Sokolovski. "Looking Under the Hood of Commodity Currency Predictability." *Management Science*, forthcoming.

Jiang, Z.; A. Krishnamurthy; and H. Lustig. "Foreign Safe Asset Demand and the Dollar Exchange Rate." *Journal of Finance*, 76 (2021), 1049–1089.

Kelly, B.; S. Malamud; and K. Zhou. “The Virtue of Complexity in Return Prediction.” *Journal of Finance*, 79 (2024), 459–503.

Kingma, D. P., and J. Ba. In *Third Annual International Conference on Learning Representations*. “Adam: A Method for Stochastic Optimization.” (2015).

Kohlscheen, E.; F. Avalos; and A. Schrimpf. “When the Walk Is Not Random: Commodity Prices and Exchange Rates.” *International Journal of Central Banking*, 13 (2017), 121–158.

Kremens, L., and I. Martin. “The Quanto Theory of Exchange Rates.” *American Economic Review*, 109 (2019), 810–843.

Lilley, A.; M. Maggiori; B. Neiman; and J. Schreger. “Exchange Rate Reconnect.” *Review of Economics and Statistics*, 104 (2022), 845–855.

Lustig, H.; N. L. Roussanov; and A. Verdelhan. “Common Risk Factors in Currency Markets.” *Review of Financial Studies*, 24 (2011), 3731–3777.

Lyons, R. K. *The Microstructure Approach to Exchange Rates*. MIT Press, Cambridge, MA (2001).

Maas, A. L.; A. Y. Hannun; and A. Y. Ng. In *Proceedings of the 30th International Conference on Machine Learning*. “Rectifier Nonlinearities Improve Neural Network Acoustic Models.”, 3–8 (2013).

Maurer, T.; T.-D. To; and N.-K. Tran. “Market Timing and Predictability in FX Markets.” *Review of Finance*, 27 (2023), 223–246.

Menkhoff, L.; L. Sarno; M. Schmeling; and A. Schrimpf. “Carry Trades and Global Foreign Exchange Volatility.” *Journal of Finance*, 67 (2012a), 681–718.

Menkhoff, L.; L. Sarno; M. Schmeling; and A. Schrimpf. “Currency Momentum Strategies.” *Journal of Financial Economics*, 106 (2012b), 660–684.

Mueller, P.; A. Stathopoulos; and A. Vedolin. “International Correlation Risk.” *Journal of Financial Economics*, 126 (2017), 270–299.

Neely, C. J., and P. A. Weller. "Lessons from the Evolution of Foreign Exchange Trading Strategies." *Journal of Banking and Finance*, 37 (2013), 3783–3798.

Neely, C. J.; P. A. Weller; and J. M. Ulrich. "The Adaptive Markets Hypothesis: Evidence from the Foreign Exchange Market." *Journal of Financial and Quantitative Analysis*, 44 (2009), 467–488.

Srivastava, N.; G. Hinton; A. Krizhevsky; I. Sutskever; and R. Salakhutdinov. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *Journal of Machine Learning Research*, 15 (2014), 1929–1958.

Taylor, J. B. "Discretion Versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*, 39 (1993), 195–214.

Zou, H., and T. Hastie. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67 (2005), 301–320.

TABLE IA.1

Summary Statistics

The table reports summary statistics for monthly log exchange rate changes measured against the US dollar. The country-i log exchange rate change is $\Delta s_{i,t}$, where $s_{i,t} = \log(S_{i,t})$, and $S_{i,t}$ is the month-t spot exchange rate for country i (number of country-i currency units per US dollar). The annualized mean (volatility) in the third (fourth) column is the monthly mean (standard deviation) multiplied by 12 ($\sqrt{12}$).

1	2	3	4	5	6	7
Country	Sample period	Ann. mean	Ann. vol.	Skewness	Excess kurtosis	Autocorr.
United Kingdom	1985:01–2020:09	-0.30%	9.98%	0.30	2.52	0.06
Switzerland	1985:01–2020:09	-2.90%	11.14%	-0.04	0.96	-0.01
Japan	1985:01–2020:09	-2.43%	10.88%	-0.35	1.82	0.04
Canada	1985:01–2020:09	0.02%	7.34%	0.48	3.99	-0.04
Australia	1985:01–2020:09	0.39%	11.64%	0.67	2.36	0.04
New Zealand	1985:01–2020:09	-0.92%	12.04%	0.39	1.83	-0.03
Sweden	1985:01–2020:09	-0.01%	10.98%	0.44	1.50	0.10
Norway	1985:01–2020:09	0.08%	11.03%	0.43	1.11	0.02
Denmark	1985:01–2020:09	-1.61%	10.27%	0.19	0.79	0.03
Euro area	1999:02–2020:09	-0.14%	9.55%	0.16	1.13	0.03
Germany	1985:01–1998:12	-4.55%	11.61%	0.32	0.32	0.03
Italy	1985:01–1998:12	-1.13%	11.39%	0.79	2.03	0.09
France	1985:01–1998:12	-3.89%	11.13%	0.42	0.54	0.01
Netherlands	1985:01–1998:12	-4.56%	11.57%	0.28	0.34	0.03

TABLE IA.2

Publication Lags: CPIs and Unemployment Rates

The table reports publication frequencies and lags for CPIs and unemployment rates for non-US countries. The last column shows the adjustment made to the timing of the country characteristic to account for the publication lag. Adjustments for the variables reported at a quarterly frequency are described in Section IA.1.B.

1 Country	2 Publication frequency	3 Publication lag	4 Adjustment
Panel A. CPI			
United Kingdom	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
Switzerland	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
Japan	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
Canada	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
Australia	Quarterly	—	—
New Zealand	Quarterly	—	—
Sweden	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
Norway	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
Denmark	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
Euro area	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
Germany	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
Italy	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
France	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$
Netherlands	Monthly	1 month	$\text{INF}_{i,t} \Rightarrow \text{INF}_{i,t-1}$

TABLE IA.2 (continued)

1 Country	2 Publication frequency	3 Publication lag	4 Adjustment
Panel B. Unemployment rate			
United Kingdom	Monthly	2 months	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-2}$
Switzerland	Monthly	1 month	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-1}$
Japan	Monthly	1 month	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-1}$
Canada	Monthly	1 month	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-1}$
Australia	Monthly	1 month	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-1}$
New Zealand	Quarterly	—	—
Sweden	Monthly	1 month	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-1}$
Norway	Monthly	1 month	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-1}$
Denmark	Monthly	1 month	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-1}$
Euro area	Monthly	2 months	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-2}$
Germany	Monthly	1 month	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-1}$
Italy	Monthly	2 months	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-2}$
France	Monthly	2 months	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-2}$
Netherlands	Monthly	1 month	$\text{UN}_{i,t} \Rightarrow \text{UN}_{i,t-1}$

TABLE IA.3

Correlation Matrix: Global State Variables

The table reports the correlation matrix for the global state variables defined in Section II of the paper; 0.00 indicates less than 0.005 in absolute value. The sample period is 1985:01 to 2020:09.

1	2	3	4	5	6	7
Global variable	EPU	MPU	GR	GVOL	GILL	GCOR
EPU	1	0.50	0.13	0.19	0.00	0.04
MPU	0.50	1	0.40	0.26	0.12	0.04
GR	0.13	0.40	1	−0.02	−0.03	0.06
GVOL	0.19	0.26	−0.02	1	0.33	0.27
GILL	0.00	0.12	−0.03	0.33	1	−0.03
GCOR	0.04	0.04	0.06	0.27	−0.03	1

TABLE IA.4

Out-of-Sample R^2 Statistics: Countries that Joined the Euro

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. "Baseline OLS" ("Baseline ENet") is the baseline panel predictive regression in equation (1) from the paper estimated via OLS (the ENet); "No-interaction ENet" is the panel predictive regression without interaction terms in equation (3) from the paper estimated via the ENet; "DNN" is the deep neural network in Section III.B of the paper; "Ensemble" is an average of the baseline ENet and DNN forecasts.

1	2	3	4	5	6	7	8
Country	Out-of-sample period	Obs.	Baseline OLS	Baseline ENet	No-interaction ENet	DNN	Ensemble
Germany	1995:01–1998:12	48	−9.77	0.38	1.16	−3.86	−1.65
Italy	1995:01–1998:12	48	−9.90	0.18	1.05	0.07	0.26
France	1995:01–1998:12	48	−32.97	0.68	0.74	−1.39	−0.22
Netherlands	1995:01–1998:12	48	−19.34	0.52	2.30	−3.50	−1.24

TABLE IA.5

Out-of-Sample R^2 Statistics: Japanese Yen as Base Currency

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes with the Japanese yen serving as the base currency. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. "Baseline OLS" ("Baseline ENet") is the baseline panel predictive regression in equation (1) from the paper estimated via OLS (the ENet); "DNN" is the deep neural network in Section III.B of the paper; "Ensemble" is an average of the baseline ENet and DNN forecasts.

1	2	3	4	5	6	7
Country	Out-of-sample period	Observations	Baseline OLS	Baseline ENet	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	-4.91	1.94^{**}	0.85	1.67^{**}
Switzerland	1995:01–2020:09	309	-10.77	0.51	2.02	1.65
United States	1995:01–2020:09	309	-13.35	-1.71	-1.28	-1.28
Canada	1995:01–2020:09	309	-6.03	1.15^{*}	0.54	1.09^{*}
Australia	1995:01–2020:09	309	-1.41	1.31^{*}	-0.33	0.69
New Zealand	1995:01–2020:09	309	-5.10	1.68^{*}	0.89	1.54^{*}
Sweden	1995:01–2020:09	309	1.82^{**}	1.66^{*}	1.38^{*}	1.73^{*}
Norway	1995:01–2020:09	309	-3.67	1.97^{*}	0.81	1.65^{*}
Denmark	1995:01–2020:09	309	-3.74	1.57^{*}	0.62	1.37^{*}
Euro area	2000:02–2020:09	248	-7.61	1.82	-3.01	-0.06
All	1995:01–2020:09	3,029	-4.84	1.31^{***}	0.39^{***}	1.12^{***}

TABLE IA.6

Out-of-Sample R^2 Statistics: Swiss Franc as Base Currency

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes with the Swiss franc serving as the base currency. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. "Baseline OLS" ("Baseline ENet") is the baseline panel predictive regression in equation (1) from the paper estimated via OLS (the ENet); "DNN" is the deep neural network in Section III.B of the paper; "Ensemble" is an average of the baseline ENet and DNN forecasts.

1	2	3	4	5	6	7
Country	Out-of-sample period	Observations	Baseline OLS	Baseline ENet	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	-21.57	0.67^*	0.31	0.93^*
United States	1995:01–2020:09	309	-21.70	-3.69	-1.50	-1.50
Japan	1995:01–2020:09	309	-13.17	-1.48	-1.40	-1.17
Canada	1995:01–2020:09	309	-14.29	-0.63	-0.93	-0.39
Australia	1995:01–2020:09	309	-9.74	0.11	-1.41	-0.34
New Zealand	1995:01–2020:09	309	-21.72	0.12	-1.33	-0.17
Sweden	1995:01–2020:09	309	-15.97	-0.23	-0.60	0.01
Norway	1995:01–2020:09	309	-17.27	0.91^*	-0.05	0.98^*
Denmark	1995:01–2020:09	309	-48.26	-5.89	-2.32	-3.08
Euro area	2000:02–2020:09	248	-51.52	-6.97	-3.49	-3.67
All	1995:01–2020:09	3,029	-18.98	-1.01	-1.10	-0.52^{***}

TABLE IA.7

Out-of-Sample R^2 Statistics: Different Hyperparameter Tuning Methods

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. The fourth through seventh columns report results for baseline ENet forecasts that use five-fold cross validation (5-fold CV), time-series validation (TSV), time-series cross validation (TSCV), and the BIC, respectively, to tune the regularization parameters (δ and λ), as described in IA.4.A.

1	2	3	4	5	6	7
Country	Out-of-sample period	Observations	5-fold CV	TSV	TSCV	BIC
United Kingdom	1995:01–2020:09	309	-3.99	1.75^{*}	2.46^{**}	3.35^{**}
Switzerland	1995:01–2020:09	309	-5.38	-1.02	0.96^{**}	0.13^{*}
Japan	1995:01–2020:09	309	-6.37	-0.98	0.50^{*}	-0.32
Canada	1995:01–2020:09	309	-10.81	-3.57	-1.26	-2.77
Australia	1995:01–2020:09	309	-0.98	0.33^{*}	-0.15	1.07^{**}
New Zealand	1995:01–2020:09	309	-3.83	-1.60	1.16	2.81^{**}
Sweden	1995:01–2020:09	309	-0.48	2.20^{**}	1.87^{**}	1.62^{**}
Norway	1995:01–2020:09	309	-1.25	1.98^{**}	1.43^{**}	1.57^{**}
Denmark	1995:01–2020:09	309	-0.85	0.86^{***}	0.85^{***}	2.23^{***}
Euro area	2000:02–2020:09	248	-5.07	-0.20	-0.88	1.23^{**}
All	1995:01–2020:09	3,029	-3.51	0.02^{***}	0.76^{***}	1.22^{***}

TABLE IA.8

Out-of-Sample R^2 Statistics: Random Features Forecast

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes based on the Kelly et al. (2024) random features model. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. P is the number of random features.

1	2	3	4	5	6
Country	Out-of-sample period	Observations	$P = 10{,}000$	$P = 20{,}000$	$P = 30{,}000$
United Kingdom	1995:01–2020:09	309	1.58^{*}	1.58^{**}	1.40^{*}
Switzerland	1995:01–2020:09	309	0.57	0.07	0.05
Japan	1995:01–2020:09	309	-1.69	-1.29	-1.55
Canada	1995:01–2020:09	309	-2.79	-2.14	-2.98
Australia	1995:01–2020:09	309	-0.24	0.48^{**}	-0.10
New Zealand	1995:01–2020:09	309	2.05^{**}	2.08^{**}	1.94^{**}
Sweden	1995:01–2020:09	309	1.23^{*}	1.59^{*}	1.50^{*}
Norway	1995:01–2020:09	309	-0.57	0.12^{*}	-0.15
Denmark	1995:01–2020:09	309	-1.05	-0.44	-0.72
Euro area	2000:02–2020:09	248	-4.68	-2.66	-4.14
All	1995:01–2020:09	3,029	-0.39	0.05^{**}	-0.30

TABLE IA.9

Out-of-Sample R^2 Statistics: Ridge-Based Forecasts

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. "Baseline ridge" is the baseline predictive regression in equation (1) from the paper estimated via ridge regression; "DNN" is the deep neural network in Section III.B of the paper; "Ensemble" is an average of the baseline ridge and DNN forecasts.

1	2	3	4	5	6
Country	Out-of-sample period	Observations	Baseline ridge	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	0.30^{**}	3.03^{**}	1.99^{**}
Switzerland	1995:01–2020:09	309	0.15^{*}	2.15^{**}	1.37^{**}
Japan	1995:01–2020:09	309	0.09^{*}	0.28	0.35
Canada	1995:01–2020:09	309	0.08	0.91	0.80
Australia	1995:01–2020:09	309	0.05	1.89^{***}	1.17^{**}
New Zealand	1995:01–2020:09	309	0.20^{**}	1.96^{**}	1.25^{**}
Sweden	1995:01–2020:09	309	0.15^{*}	1.69^{**}	1.10^{**}
Norway	1995:01–2020:09	309	0.15^{*}	1.55^{**}	1.06^{**}
Denmark	1995:01–2020:09	309	0.15^{**}	2.04^{**}	1.30^{***}
Euro area	2000:02–2020:09	248	0.29^{*}	2.11^{**}	1.48^{**}
All	1995:01–2020:09	3,029	0.15^{**}	1.74^{***}	1.16^{***}

TABLE IA.10

Out-of-Sample R^2 Statistics: Models Fitted Using Country-Specific Data

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. "Baseline OLS" ("Baseline ENet") is the baseline predictive regression in equation (1) from the paper estimated via OLS (the ENet); "DNN" is the deep neural network in Section III.B of the paper; "Ensemble" is an average of the baseline ENet and DNN forecasts. The models are estimated using country-specific data.

1	2	3	4	5	6	7
Country	Out-of-sample period	Observations	Baseline OLS	Baseline ENet	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	-161.92	0.00	0.83^{**}	0.48^{**}
Switzerland	1995:01–2020:09	309	-193.02	0.00	-0.07	0.13
Japan	1995:01–2020:09	309	-170.17	0.00	0.51	0.33
Canada	1995:01–2020:09	309	-65.94	0.42	0.34	0.47
Australia	1995:01–2020:09	309	-104.08	0.50^{**}	1.26^{**}	1.12^{***}
New Zealand	1995:01–2020:09	309	-134.92	1.35^{***}	-0.11	0.72^{**}
Sweden	1995:01–2020:09	309	-79.12	0.00	0.51	0.36
Norway	1995:01–2020:09	309	-93.55	0.00	0.74^{*}	0.44^{*}
Denmark	1995:01–2020:09	309	-120.32	0.00	0.16	0.20

TABLE IA.11

Out-of-Sample R^2 Statistics: Random Forest

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. The ensemble-all forecast is an average of the baseline ENet, DNN, and random forest forecasts.

1	2	3	4	5
Country	Out-of-sample period	Observations	Random forest	Ensemble-all
United Kingdom	1995:01–2020:09	309	2.14^{**}	3.72^{**}
Switzerland	1995:01–2020:09	309	1.23^{**}	2.03^{**}
Japan	1995:01–2020:09	309	-1.11	0.33
Canada	1995:01–2020:09	309	-1.60	0.73^{*}
Australia	1995:01–2020:09	309	0.62^{**}	1.85^{**}
New Zealand	1995:01–2020:09	309	2.14^{***}	2.54^{***}
Sweden	1995:01–2020:09	309	1.99^{**}	2.46^{**}
Norway	1995:01–2020:09	309	2.23^{***}	2.49^{***}
Denmark	1995:01–2020:09	309	1.49^{***}	2.64^{***}
Euro area	2000:02–2020:09	248	0.76^{**}	2.66^{**}
All	1995:01–2020:09	3,029	1.10^{***}	2.13^{***}

TABLE IA.12

Granger Causality Test Results

Panel A reports Granger causality test results for monthly log exchange rate changes (LIRET) and net commodity returns (CXP) at the country level, with the lag order selected via the AIC (considering a maximum lag of twelve). Panel B provides robustness checks across fixed lag orders, reporting the number and percent of countries for which the null of no Granger causality is rejected at the 10% level. Panel C reports results for the Dumitrescu and Hurlin (2012) panel Granger causality test based on the $\bar{Z}$ statistic; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively; 0.00 indicates less than 0.005.

1	2	3	4	5	6	7
			LIRET → CXP		CXP → LIRET	
Panel A. Country-level tests (AIC lag selection)						
Country	Obs.	Lag	F-stat	p-value	F-stat	p-value
United Kingdom	443	7	1.14	0.33	0.93	0.49
Switzerland	443	1	14.17	0.00***	0.96	0.33
Japan	443	1	0.01	0.91	0.09	0.77
Canada	429	1	31.73	0.00***	4.18	0.04**
Australia	429	5	9.91	0.00***	1.39	0.23
New Zealand	429	6	2.44	0.03**	0.41	0.87
Sweden	429	2	0.69	0.51	0.39	0.67
Norway	429	1	19.72	0.00***	2.58	0.11
Denmark	429	1	2.11	0.15	0.58	0.45
Euro area	260	1	9.56	0.00***	0.30	0.58
Germany	182	1	3.08	0.08*	0.78	0.38
Italy	182	1	4.26	0.04**	0.51	0.48
France	182	1	2.88	0.09*	2.89	0.09*
Netherlands	182	1	0.62	0.43	0.69	0.41

TABLE IA.12 (continued)

1	2	3	4	5	6	7
			LIRET $\rightarrow$ CXP		CXP $\rightarrow$ LIRET	
Panel B: Robustness checks (fixed lag orders)						
Lag	N		# significant	% significant	# significant	% significant
1	14		8	57	2	14
3	14		8	57	0	0
6	14		7	50	1	7
12	14		6	43	2	14
Panel C: Panel tests						
Lag	N		$\bar{Z}$	p-value	$\bar{Z}$	p-value
1	14		19.70	0.00***	0.58	0.28
3	14		12.83	0.00***	−0.51	0.70
6	14		10.42	0.00***	0.73	0.23
12	14		6.55	0.00***	−0.08	0.53

TABLE IA.13

Out-of-Sample R^2 Statistics: Including Net Commodity Returns

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes with net commodity returns as an additional characteristic. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. "Baseline OLS" ("Baseline ENet") is the baseline panel predictive regression in equation (1) from the paper estimated via OLS (the ENet); "DNN" is the deep neural network in Section III.B from the paper; "Ensemble" is an average of the baseline ENet and DNN forecasts.

1	2	3	4	5	6	7
Country	Out-of-sample period	Observations	Baseline OLS	Baseline ENet	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	−14.13	2.32**	1.77**	2.33**
Switzerland	1995:01–2020:09	309	−14.20	0.31	0.16	0.59
Japan	1995:01–2020:09	309	−12.36	−1.13	−0.42	−0.53
Canada	1995:01–2020:09	309	−19.75	−0.07	0.09	0.49
Australia	1995:01–2020:09	309	−3.97	0.44	1.39**	1.12**
New Zealand	1995:01–2020:09	309	−8.22	2.39**	1.12**	1.94***
Sweden	1995:01–2020:09	309	−3.35	1.33*	1.13**	1.40**
Norway	1995:01–2020:09	309	−3.21	1.52**	1.24**	1.63**
Denmark	1995:01–2020:09	309	−5.78	1.68***	1.37***	1.77***
Euro area	2000:02–2020:09	248	−11.21	1.82*	1.23**	2.09**
All	1995:01–2020:09	3,029	−8.74	1.07**	0.92***	1.27***

TABLE IA.14

Out-of-Sample R^2 Statistics: Benchmark based on INF and UN

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the benchmark forecast; for the positive R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. The benchmark forecast is based on a linear panel predictive regression with INF and UN as predictors estimated via OLS. "Baseline ENet" is the baseline predictive regression in equation (1) from the paper estimated via the ENet; "DNN" is the deep neural network in Section III.B from the paper; "Ensemble" is an average of the baseline ENet and DNN forecasts.

1	2	3	4	5	6
Country	Out-of-sample period	Observations	Baseline ENet	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	3.33^{**}	3.12^{**}	3.51^{**}
Switzerland	1995:01–2020:09	309	0.09	1.88^{**}	1.32^{*}
Japan	1995:01–2020:09	309	-0.84	-0.18	-0.27
Canada	1995:01–2020:09	309	-0.91	0.01	-0.03
Australia	1995:01–2020:09	309	0.70	1.12^{*}	1.05
New Zealand	1995:01–2020:09	309	1.82^{*}	1.63^{*}	1.86^{*}
Sweden	1995:01–2020:09	309	1.53^{*}	1.36^{*}	1.62^{*}
Norway	1995:01–2020:09	309	1.70^{**}	1.34^{*}	1.72^{**}
Denmark	1995:01–2020:09	309	2.36^{**}	2.18^{**}	2.48^{**}
Euro area	2000:02–2020:09	248	2.47^{*}	2.21^{**}	2.75^{**}
All	1995:01–2020:09	3,029	1.19^{***}	1.42^{***}	1.54^{***}

TABLE IA.15

Out-of-Sample R^2 Statistics: Benchmark based on INF and BILL

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the benchmark forecast; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. The benchmark forecast is based on a linear panel predictive regression with INF and BILL as predictors estimated via OLS. "Baseline ENet" is the baseline predictive regression in equation (1) from the paper estimated via the ENet; "DNN" is the deep neural network in Section III.B in the paper; "Ensemble" is an average of the baseline ENet and DNN forecasts.

1	2	3	4	5	6
Country	Out-of-sample period	Observations	Baseline ENet	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	3.88^{**}	3.67^{**}	4.05^{**}
Switzerland	1995:01–2020:09	309	0.23	2.02^{**}	1.46^{*}
Japan	1995:01–2020:09	309	-0.18	0.47	0.39
Canada	1995:01–2020:09	309	0.44	1.35^{*}	1.31^{*}
Australia	1995:01–2020:09	309	1.71^{*}	2.13^{***}	2.07^{**}
New Zealand	1995:01–2020:09	309	2.63^{**}	2.45^{**}	2.68^{**}
Sweden	1995:01–2020:09	309	2.07^{**}	1.90^{**}	2.16^{**}
Norway	1995:01–2020:09	309	2.21^{**}	1.85^{**}	2.23^{**}
Denmark	1995:01–2020:09	309	2.95^{***}	2.77^{***}	3.07^{***}
Euro area	2000:02–2020:09	248	2.81^{*}	2.55^{**}	3.09^{**}
All	1995:01–2020:09	3,029	1.85^{***}	2.07^{***}	2.20^{***}

TABLE IA.16

Out-of-Sample R^2 Statistics: Regime-Switching Specifications

The table reports out-of-sample R^2 statistics in percent for forecasts of monthly log exchange rate changes. The out-of-sample R^2 statistic measures the proportional reduction in MSPE for the competing forecast in the column heading vis-à-vis the no-change benchmark forecast; 0.00 indicates less than 0.005 in absolute value; for the positive out-of-sample R^2 statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% levels, respectively, according to the Clark and West (2007) test. "No interaction" is a panel predictive regression based on ten country characteristics; "Interaction" is a panel predictive regression based on ten country characteristics and their interactions with global FX volatility; "Interaction with threshold" ("Interaction with LST") augments the interaction terms with a threshold (logistic smooth transition, LST) mechanism. The panel predictive regressions are estimated via the ENet.

1	2	3	4	5	6	7
Country	Out-of-sample period	Obs.	No interaction	Interaction	Interaction with threshold	Interaction with LST
United Kingdom	1995:01–2020:09	309	0.28^{**}	3.04^{**}	2.08^{*}	2.77^{**}
Switzerland	1995:01–2020:09	309	-0.81	-0.20	-0.28	0.17^{*}
Japan	1995:01–2020:09	309	-2.01	-0.83	-1.88	-1.32
Canada	1995:01–2020:09	309	-3.10	-2.95	-2.77	-3.70
Australia	1995:01–2020:09	309	0.77^{*}	0.57^{*}	1.97^{**}	1.14^{**}
New Zealand	1995:01–2020:09	309	0.21^{*}	2.74^{**}	3.58^{**}	3.48^{**}
Sweden	1995:01–2020:09	309	-0.20	0.49	1.65^{**}	1.66^{**}
Norway	1995:01–2020:09	309	0.20^{*}	-0.01	0.23^{*}	0.67^{**}
Denmark	1995:01–2020:09	309	1.02^{**}	1.08^{**}	1.61^{***}	1.81^{***}
Euro area	2000:02–2020:09	248	-0.50	1.11^{*}	2.22^{**}	1.27^{**}
All	1995:01–2020:09	3,029	-0.29	0.61^{**}	1.02^{**}	0.99^{**}

TABLE IA.17

Annualized Average Utility Gains: Investing in Individual Foreign Currencies

The table reports annualized average utility gains in percent for a mean-variance US investor with a relative risk aversion coefficient of five who allocates monthly between the US dollar and the foreign currency in the first column. The average utility gain is the increase in certainty equivalent return when the investor uses the ML-Opt instead of the Bench-Opt portfolio. For the ML-Opt (Bench-Opt) portfolio, the investor uses the baseline ENet, DNN, or ensemble (no-change) exchange rate forecast when predicting the currency excess return.

1	2	3	4	5	6
Country	Out-of-sample period	Obs.	Baseline ENet	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	3.48	2.20	3.16
Switzerland	1995:01–2020:09	309	1.61	1.50	1.86
Japan	1995:01–2020:09	309	0.61	1.55	1.09
Canada	1995:01–2020:09	309	2.40	1.41	1.74
Australia	1995:01–2020:09	309	2.78	1.88	2.27
New Zealand	1995:01–2020:09	309	2.69	2.83	2.58
Sweden	1995:01–2020:09	309	3.16	1.93	2.71
Norway	1995:01–2020:09	309	2.42	2.06	2.38
Denmark	1995:01–2020:09	309	3.25	2.58	3.15
Euro area	2000:02–2020:09	248	3.84	2.82	3.64

TABLE IA.18

Portfolio Performance for Non-US Domestic Investors

The table reports portfolio performance metrics for a mean-variance non-US domestic investor (given in the panel heading) with a relative risk aversion coefficient of five who allocates monthly across the domestic currency and all available foreign currencies. For the ML-Opt (Bench-Opt) portfolio, the investor uses the baseline ENet, DNN, or ensemble (no-change) exchange rate forecast when predicting the currency excess return. The second column reports the annualized increase in certainty equivalent return when the investor uses the ML-Opt instead of the Bench-Opt portfolio. Statistical significance for the Sharpe ratio is based on the Bao (2009) procedure; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

1	2	3	4	5
Portfolio	Annualized average utility gain	Annualized mean	Annualized volatility	Annualized Sharpe ratio
Panel A: United Kingdom				
ML-Opt, baseline ENet	2.92%	10.50%	12.96%	0.81^{***}
ML-Opt, DNN	1.29%	8.40%	12.23%	0.67^{***}
ML-Opt, ensemble	2.43%	9.97%	12.42%	0.78^{***}
Bench-Opt	–	6.19%	10.61%	0.58^{***}
Panel B: Switzerland				
ML-Opt, baseline ENet	−0.49%	7.83%	14.74%	0.53^{***}
ML-Opt, DNN	−1.43%	6.05%	13.54%	0.45^{**}
ML-Opt, ensemble	−0.94%	7.13%	14.39%	0.50^{**}
Bench-Opt	–	6.61%	12.19%	0.54^{***}
Panel C: Japan				
ML-Opt, baseline ENet	0.48%	9.74%	14.41%	0.68^{***}
ML-Opt, DNN	0.82%	8.97%	12.76%	0.70^{***}
ML-Opt, ensemble	0.74%	9.37%	13.51%	0.69^{***}
Bench-Opt	–	7.61%	11.88%	0.64^{***}

TABLE IA.18 (continued)

1 Portfolio	2 Annualized average utility gain	3 Annualized mean	4 Annualized volatility	5 Annualized Sharpe ratio
Panel D: Canada				
ML-Opt, baseline ENet	2.29%	9.34%	12.57%	0.74***
ML-Opt, DNN	0.82%	7.41%	11.84%	0.63***
ML-Opt, ensemble	1.18%	8.07%	12.33%	0.65***
Bench-Opt	–	5.86%	10.54%	0.56***
Panel E: Australia				
ML-Opt, baseline ENet	1.51%	9.52%	12.44%	0.77***
ML-Opt, DNN	0.91%	8.47%	11.68%	0.72***
ML-Opt, ensemble	1.24%	9.08%	12.17%	0.75***
Bench-Opt	–	7.04%	10.77%	0.65***

TABLE IA.19

Portfolio Performance with Transaction Costs

The table reports portfolio performance metrics for a mean-variance US investor with a relative risk aversion coefficient of five who allocates monthly across the US dollar and all available foreign currencies for the 1995:01 to 2020:09 out-of-sample period. Results are reported for different assumptions regarding transaction costs for bid-ask spreads from Datastream. For the ML-Opt (Bench-Opt) portfolio, the investor uses the baseline ENet, DNN, or ensemble (no-change) exchange rate forecast when predicting the currency excess return. The second column reports the annualized increase in certainty equivalent return when the investor uses the ML-Opt instead of the Bench-Opt portfolio. Statistical significance for the Sharpe ratio is based on the Bao (2009) procedure; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

1	2	3	4	5
Portfolio	Annualized average utility gain	Annualized mean	Annualized volatility	Annualized Sharpe ratio
Panel A: 25% of bid-ask spread				
ML-Opt, baseline ENet	4.40%	11.85%	13.53%	0.88***
ML-Opt, DNN	1.82%	8.06%	11.61%	0.69***
ML-Opt, ensemble	2.86%	9.69%	12.59%	0.77***
Bench-Opt	–	5.37%	10.00%	0.54***
Panel B: 50% of bid-ask spread				
ML-Opt, baseline ENet	4.32%	10.86%	13.53%	0.80***
ML-Opt, DNN	1.76%	7.10%	11.60%	0.61***
ML-Opt, ensemble	2.79%	8.71%	12.58%	0.69***
Bench-Opt	–	4.46%	9.99%	0.45**

TABLE IA.19 (continued)

1	2	3	4	5
Portfolio	Annualized average utility gain	Annualized mean	Annualized volatility	Annualized Sharpe ratio
Panel C: 75% of bid-ask spread				
ML-Opt, baseline ENet	4.24%	9.88%	13.52%	0.73***
ML-Opt, DNN	1.70%	6.14%	11.60%	0.53***
ML-Opt, ensemble	2.72%	7.74%	12.57%	0.62***
Bench-Opt	–	3.57%	9.99%	0.36*
Panel D: Full bid-ask spread				
ML-Opt, baseline ENet	4.16%	8.90%	13.52%	0.66***
ML-Opt, DNN	1.65%	5.18%	11.59%	0.45**
ML-Opt, ensemble	2.65%	6.77%	12.57%	0.54***
Bench-Opt	–	2.66%	9.99%	0.27

TABLE IA.20

Alphas

The table reports Lustig et al. (2011) factor model estimation results for ML-Opt portfolios based on the baseline ENet, DNN, and ensemble forecasts as well as the Bench-Opt portfolio. For the ML-Opt (Bench-Opt) portfolios, a mean-variance US investor with a relative risk aversion coefficient of five allocates monthly across available foreign currencies using the baseline ENet, DNN, or ensemble (no-change) exchange rate forecasts when predicting the currency excess return. For the second through fourth columns, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

1	2	3	4	5
Portfolio	Annualized alpha	MKT_{FX}	HML_{FX}	Adjusted R^2
Panel A: 1995:01 to 2020:09 out-of-sample period				
ML-Opt, baseline ENet	$11.53\%^{***}$	-0.47^{***}	0.40^{**}	10.96%
ML-Opt, DNN	$7.98\%^{***}$	-0.06	0.40^{***}	10.06%
ML-Opt, ensemble	$9.51\%^{***}$	-0.28^{**}	0.40^{**}	9.15%
Bench-Opt	$4.04\%^{***}$	-0.03	0.75^{***}	52.81%
Panel B: 1995:01 to 2008:08 out-of-sample period				
ML-Opt, baseline ENet	$8.70\%^{***}$	-0.44	0.76^{***}	32.09%
ML-Opt, DNN	$7.94\%^{***}$	0.09	0.64^{***}	24.84%
ML-Opt, ensemble	$7.75\%^{***}$	-0.16	0.74^{***}	30.87%
Bench-Opt	$4.51\%^{***}$	0.18	0.92^{***}	63.49%
Panel C: 2008:09 to 2020:09 out-of-sample period				
ML-Opt, baseline ENet	$12.00\%^{**}$	-0.25	0.02	0.80%
ML-Opt, DNN	$5.85\%^{*}$	-0.05	0.19	1.11%
ML-Opt, ensemble	$8.67\%^{*}$	-0.16	0.08	-0.59%
Bench-Opt	1.71%	-0.12	0.64^{***}	45.26%

TABLE IA.21

Global FX Volatility and Portfolio Return Differences

The table reports OLS estimates of $a_{i,1}$ for the regression, $rx_{i,t}^{\text{ML-Opt}} - rx_{i,t}^{\text{Bench-Opt}} = a_{i,0} + a_{i,1}\text{GVOL}_t + \varepsilon_{i,t}$, where $rx_{i,t}^{\text{ML-Opt}}$ ($rx_{i,t}^{\text{Bench-Opt}}$) is the excess return for a mean-variance US investor with a relative risk aversion coefficient of five who allocates monthly between the US dollar and the foreign currency for country i in the first column using the machine-learning exchange rate change forecast in the column heading (no-change benchmark forecast), and GVOL_t is global FX volatility. "All" corresponds to the regression, $rx_{t}^{\text{ML-Opt}} - rx_{t}^{\text{Bench-Opt}} = a_0 + a_1\text{GVOL}_t + \varepsilon_t$, where the US investor allocates monthly across the US dollar and all available foreign currencies. The excess return differences are measured in percent, and GVOL_t is standardized; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

1	2	3	4	5	6
Country	Out-of-sample period	Obs.	Baseline ENet	DNN	Ensemble
United Kingdom	1995:01–2020:09	309	0.54^{***}	0.28^{***}	0.51^{***}
Switzerland	1995:01–2020:09	309	0.29^{***}	0.27^{***}	0.34^{***}
Japan	1995:01–2020:09	309	-0.07	0.32^{***}	0.12^{*}
Canada	1995:01–2020:09	309	0.64^{***}	0.39^{***}	0.58^{***}
Australia	1995:01–2020:09	309	0.54^{***}	0.22^{***}	0.38^{***}
New Zealand	1995:01–2020:09	309	0.49^{***}	0.22^{***}	0.44^{***}
Sweden	1995:01–2020:09	309	0.54^{***}	0.19^{***}	0.42^{***}
Norway	1995:01–2020:09	309	0.77^{***}	0.38^{***}	0.66^{***}
Denmark	1995:01–2020:09	309	0.51^{***}	0.34^{***}	0.54^{***}
Euro area	2000:02–2020:09	248	0.99^{***}	0.90^{***}	0.98^{***}
All	1995:01–2020:09	3,029	1.31^{***}	0.85^{***}	1.19^{***}

FIGURE IA.1

Baseline OLS Forecasts

Each panel depicts the realized value and baseline OLS forecast for the monthly log exchange rate change for the country in the panel heading. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

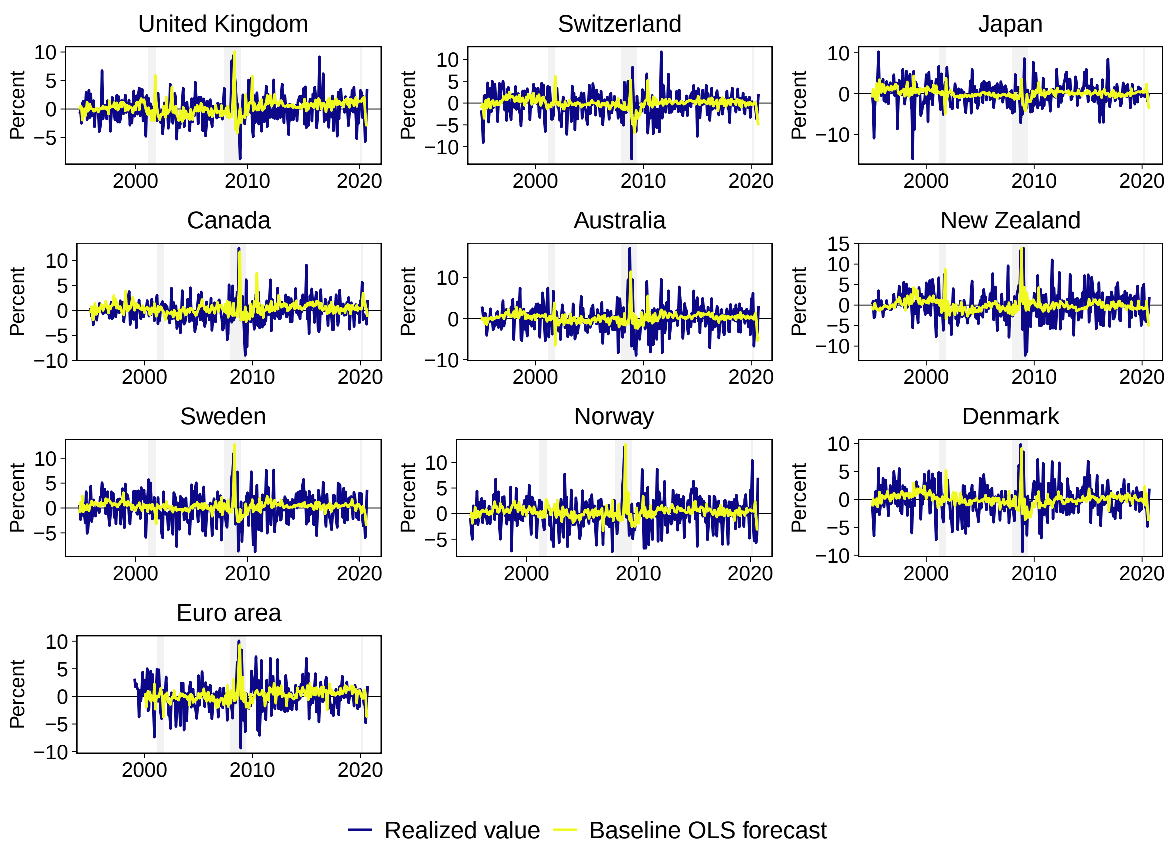


FIGURE IA.2

Baseline ENet Forecasts

Each panel depicts the realized value and baseline ENet forecast for the monthly log exchange rate change for the country in the panel heading. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

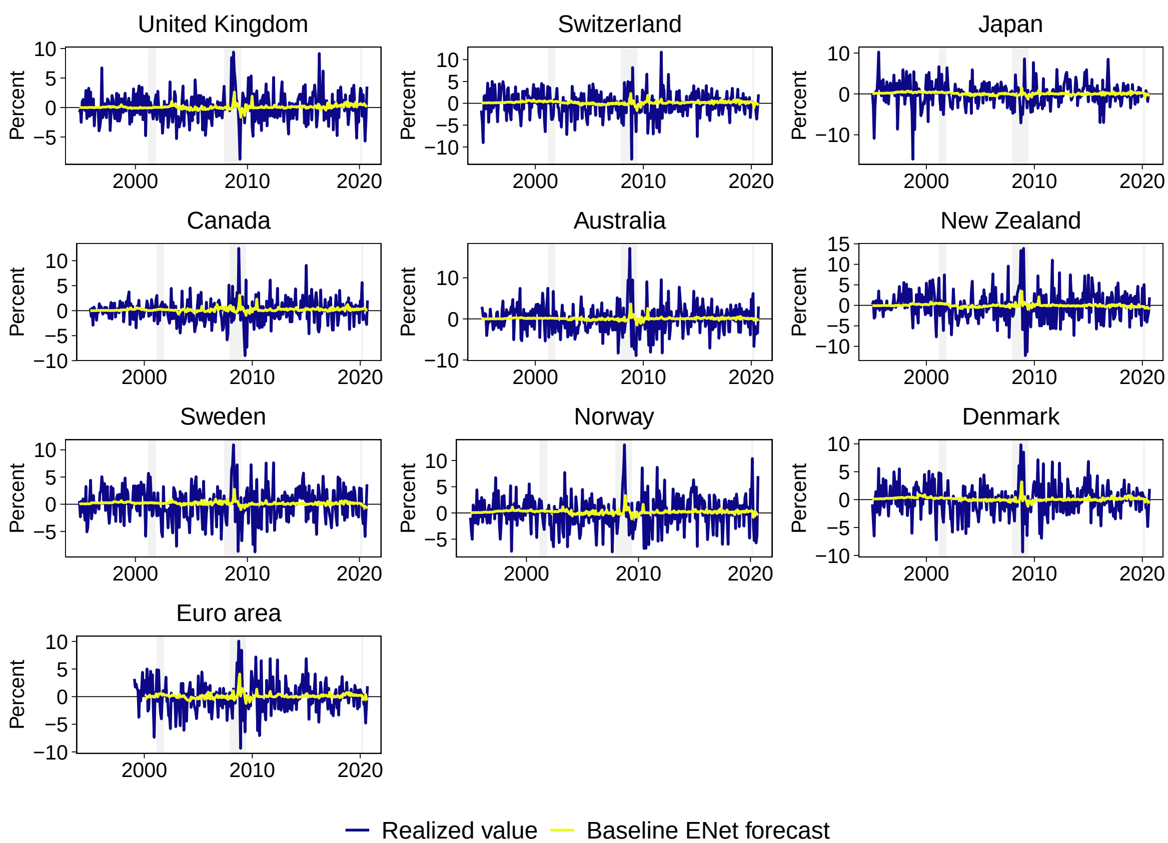


FIGURE IA.3

DNN Forecasts

Each panel depicts the realized value and DNN forecast for the monthly log exchange rate change for the country in the panel heading. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

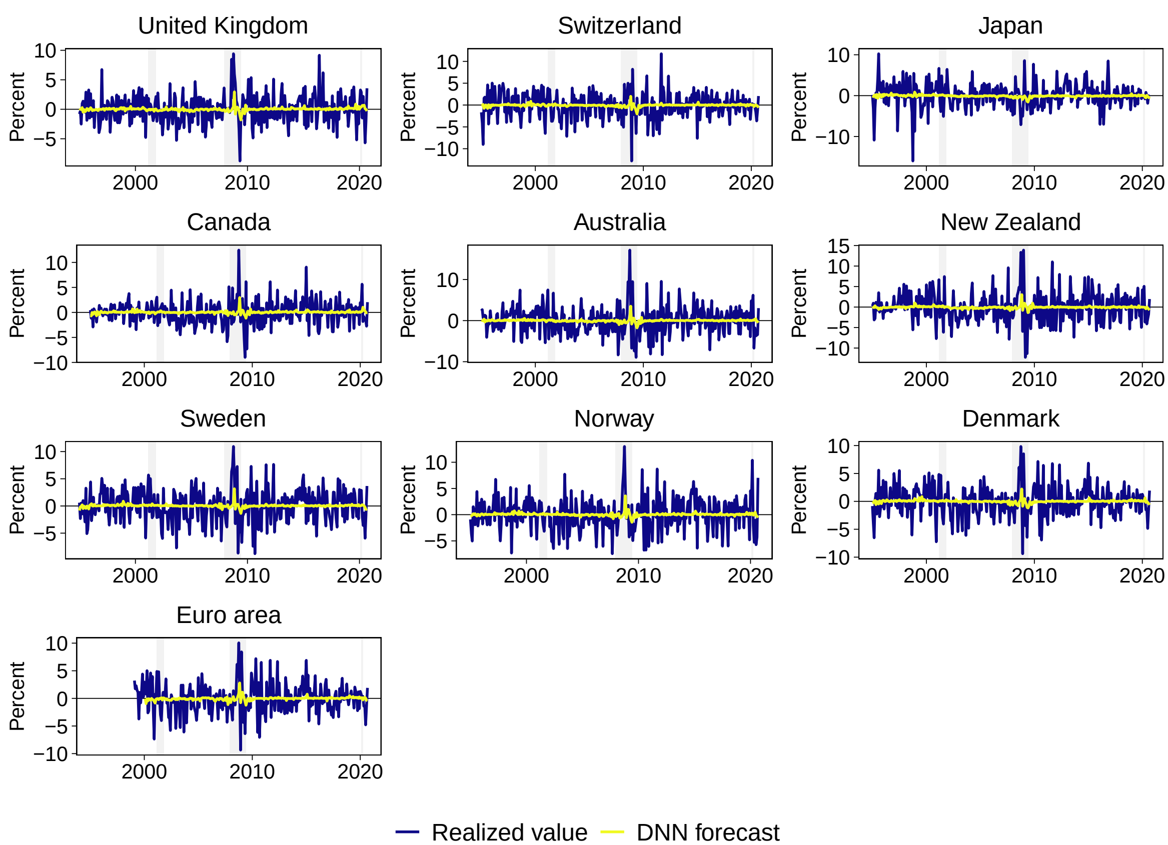


FIGURE IA.4

Ensemble Forecasts

Each panel depicts the realized value and ensemble forecast for the monthly log exchange rate change for the country in the panel heading. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

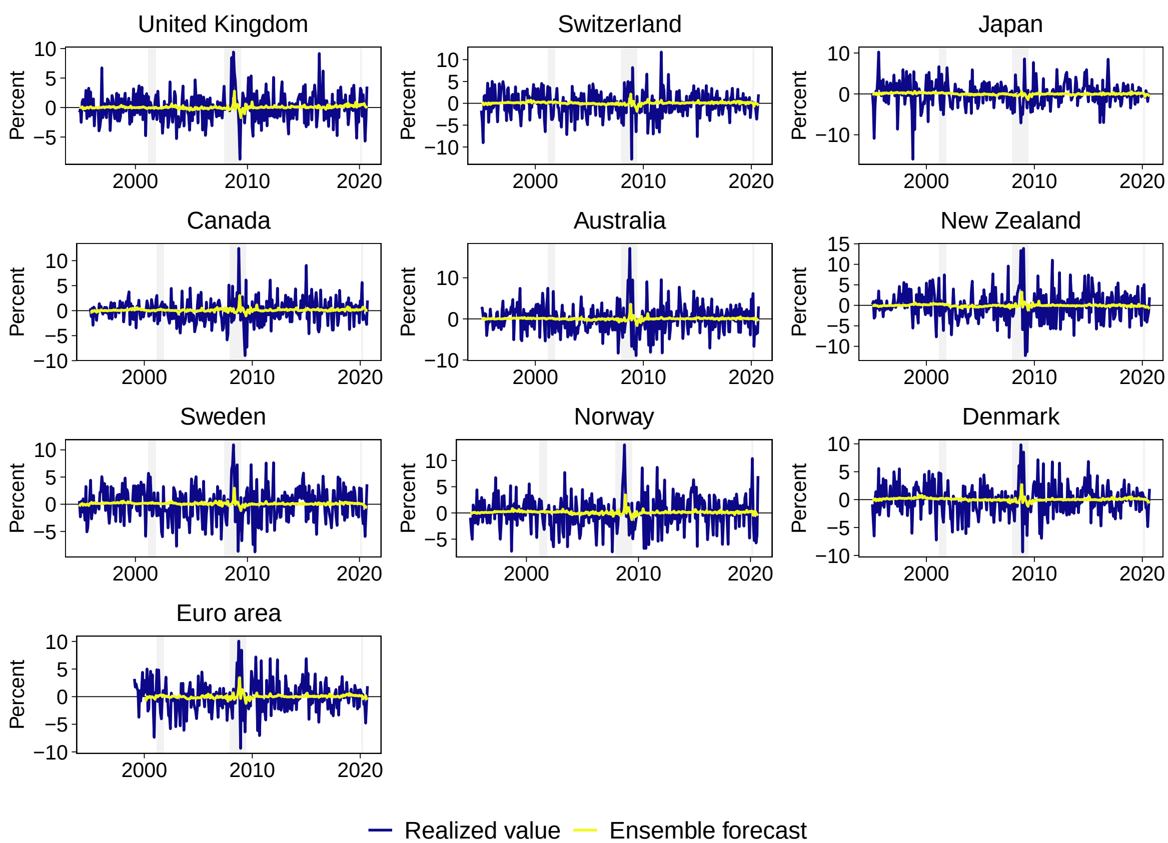


FIGURE IA.5

Rolling Out-of-Sample Clark-West Statistics

The figure depicts rolling 60-month Clark and West (2007) statistics for the baseline ENet, DNN, and ensemble forecasts. The horizontal axis gives the end of the 60-month period. The dotted horizontal line indicates the 10% critical value. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

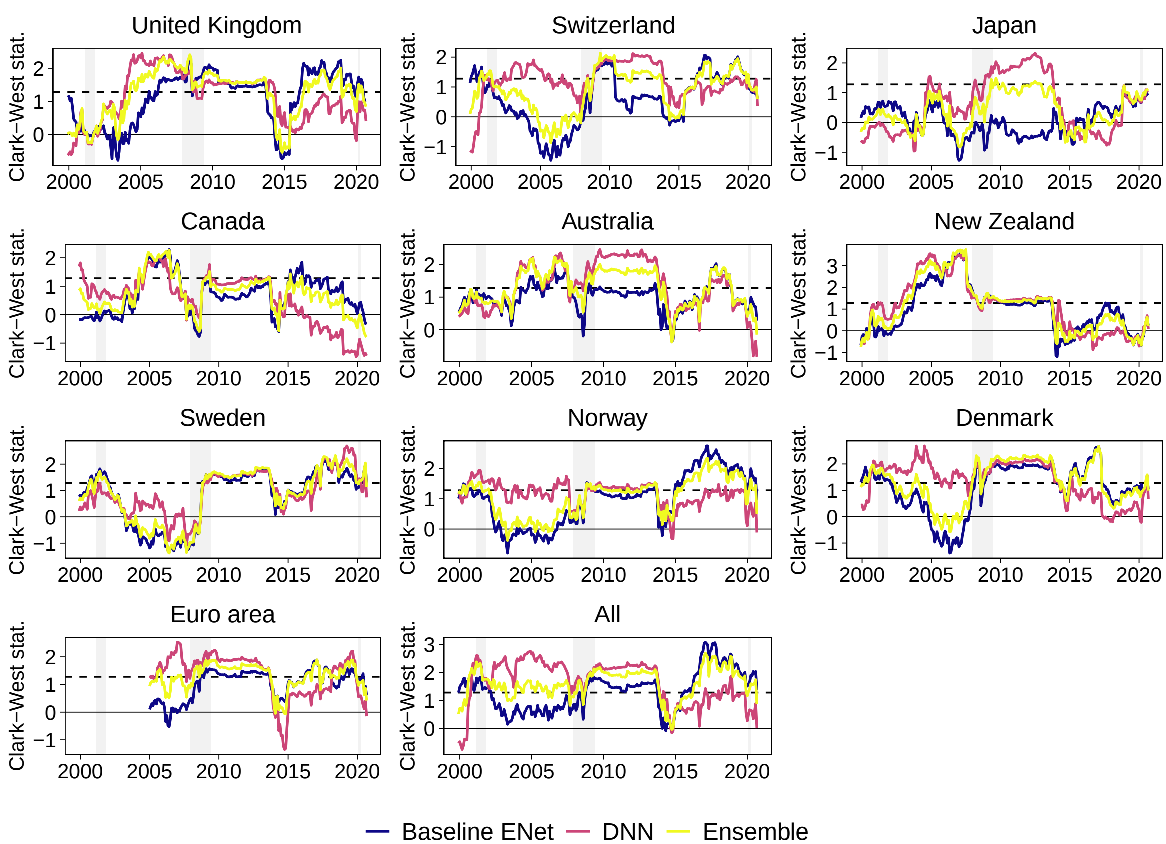


FIGURE IA.6

Cumulative Differences in Squared Prediction Errors

Each panel depicts the cumulative difference in squared prediction errors (CDSPE) for the country in the panel heading. The difference is computed for the no-change benchmark forecast relative to the baseline ENet, DNN, and ensemble forecasts (considered in turn). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

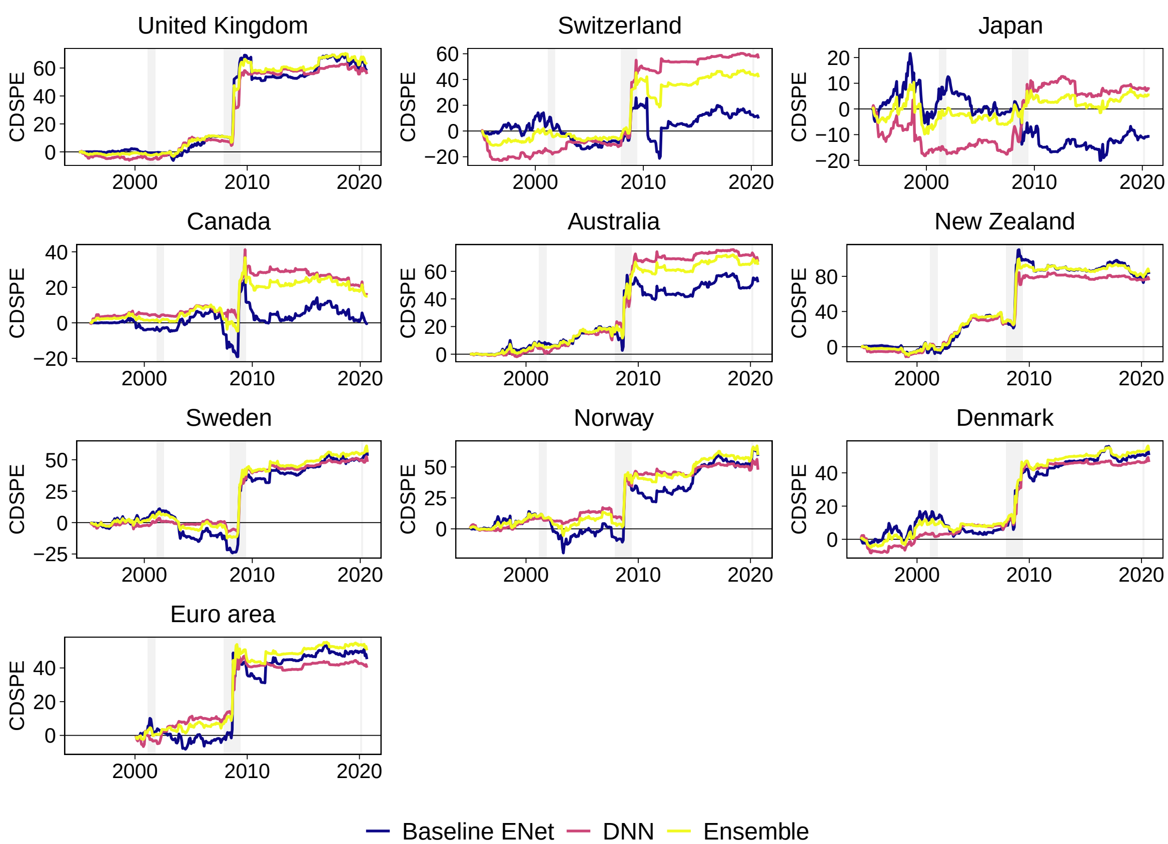


FIGURE IA.7

Change in US Foreign Bond Holdings

The figure shows actual and fitted values for a regression of the (standardized) change in US foreign bond holdings on the ten predictors selected by the ENet estimation of the baseline panel predictive regression (UN, DP, PE, INF.GVOL, UN.GVOL, BILL.GVOL, PE.EPU, PE.GILL, SRET12.GR, IV.GCOR). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

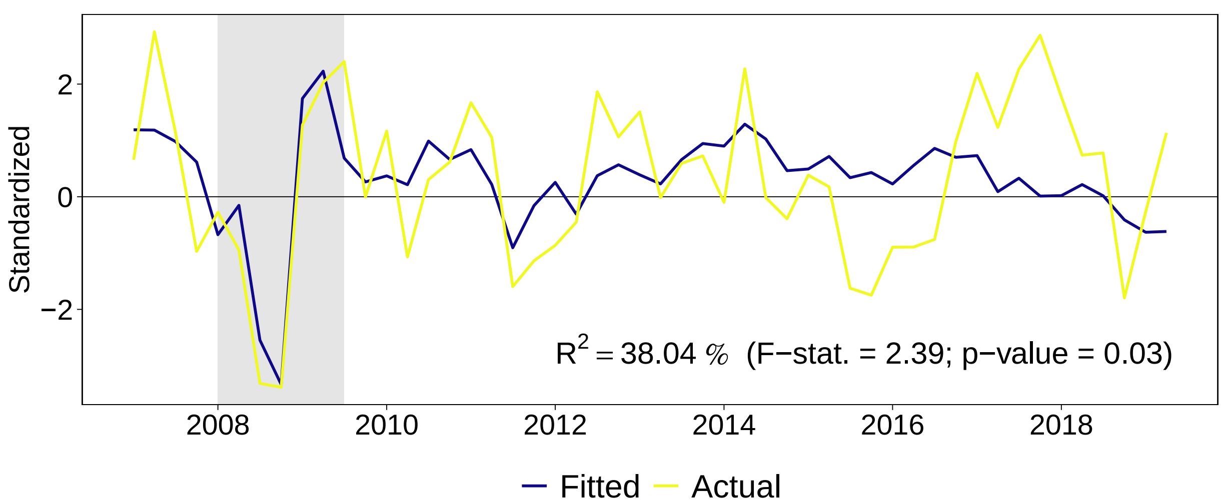


FIGURE IA.8

Full-Sample OLS Coefficient Estimates

The figure depicts OLS estimates of the slope coefficients for the baseline panel predictive regression and the 1985:01 to 2020:09 sample period. The predictors are standardized before estimation.

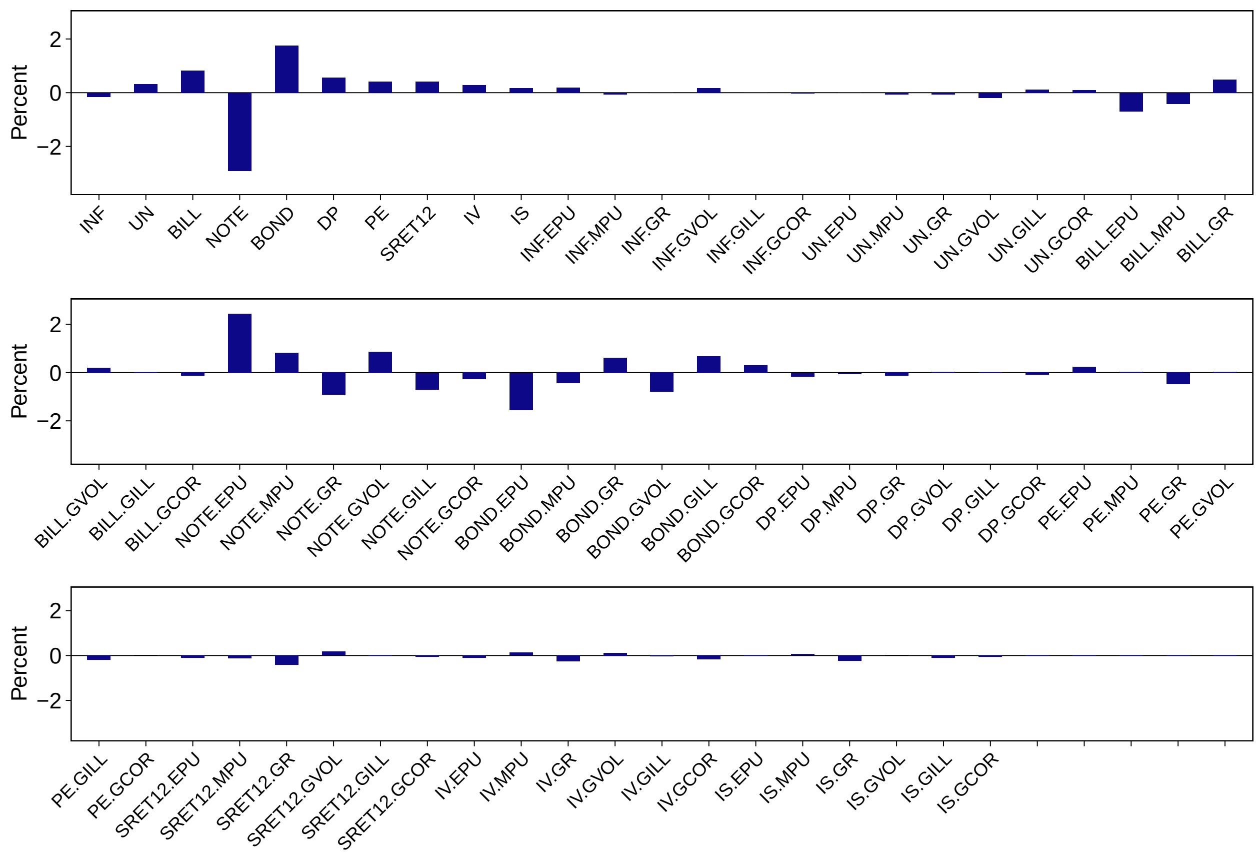


FIGURE IA.9

Recursive Baseline ENet Coefficient Estimates

Each panel depicts recursive ENet estimates of the slope coefficients for the baseline panel predictive regression for the predictor in the panel heading. Each point corresponds to the fitted model used to generate the forecast for the date on the horizontal axis. The predictors are standardized before estimation based on data available at the time of forecast formation.

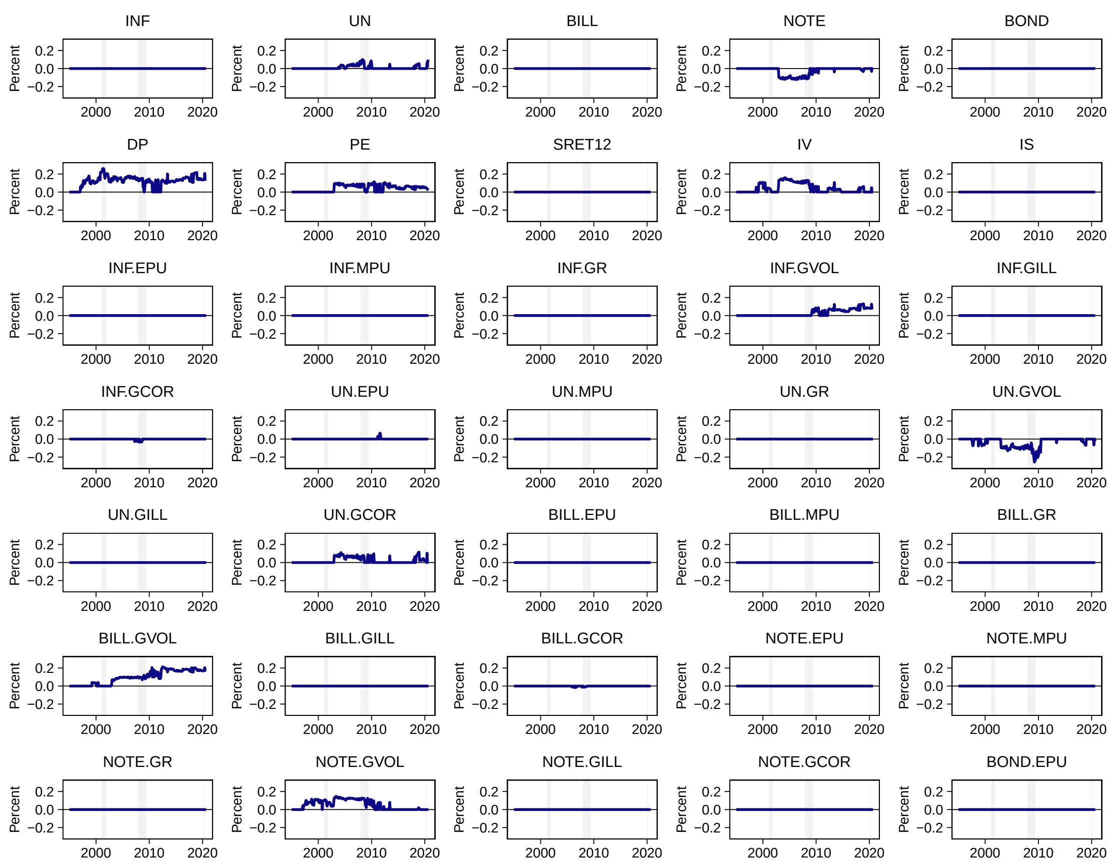


FIGURE IA.9 (continued)

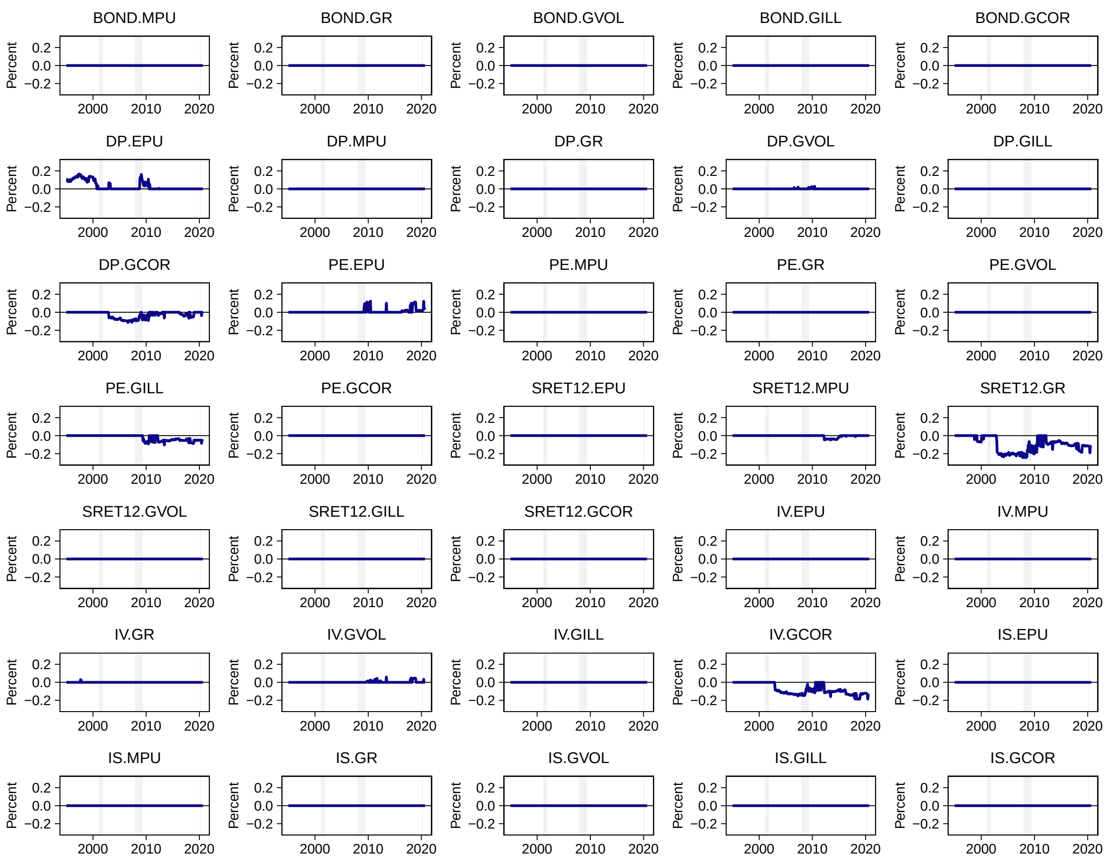


FIGURE IA.10

Recursive Variable Importance: DNN

Each panel depicts recursive PDP-based variable-importance measures for fitted DNN models for the predictor in the panel heading. Each point corresponds to the fitted model used to generate the forecast for the date on the horizontal axis.

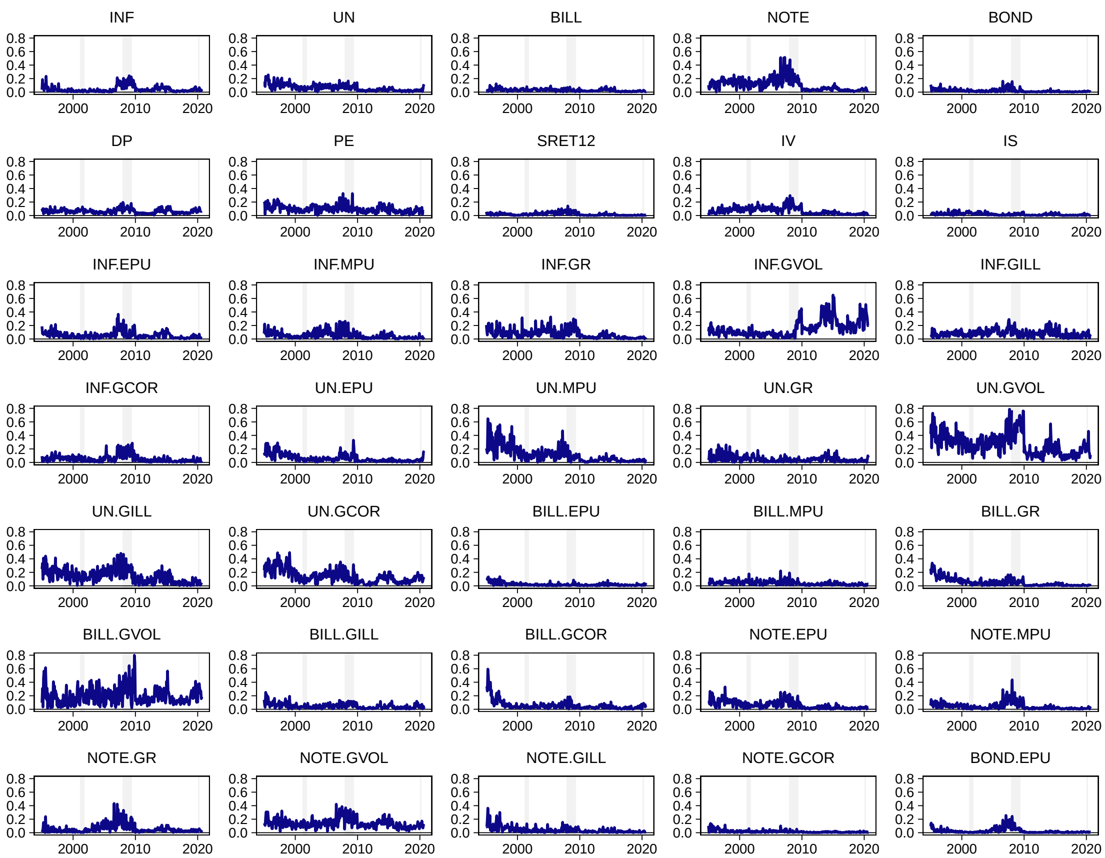


FIGURE IA.10 (continued)

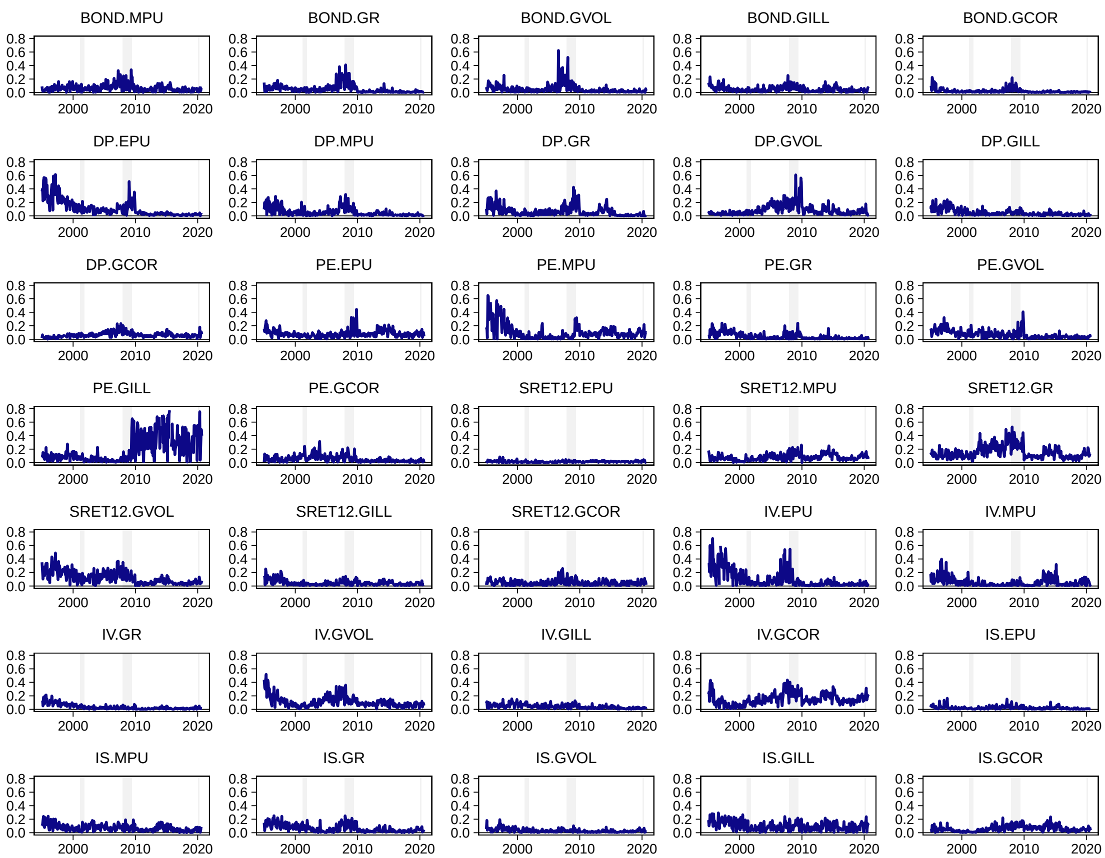


FIGURE IA.11

Recursive Variable Importance: Ensemble

Each panel depicts recursive PDP-based variable-importance measures for models underlying the ensemble forecast for the predictor in the panel heading. Each point corresponds to the fitted models used to generate the forecast for the date on the horizontal axis.

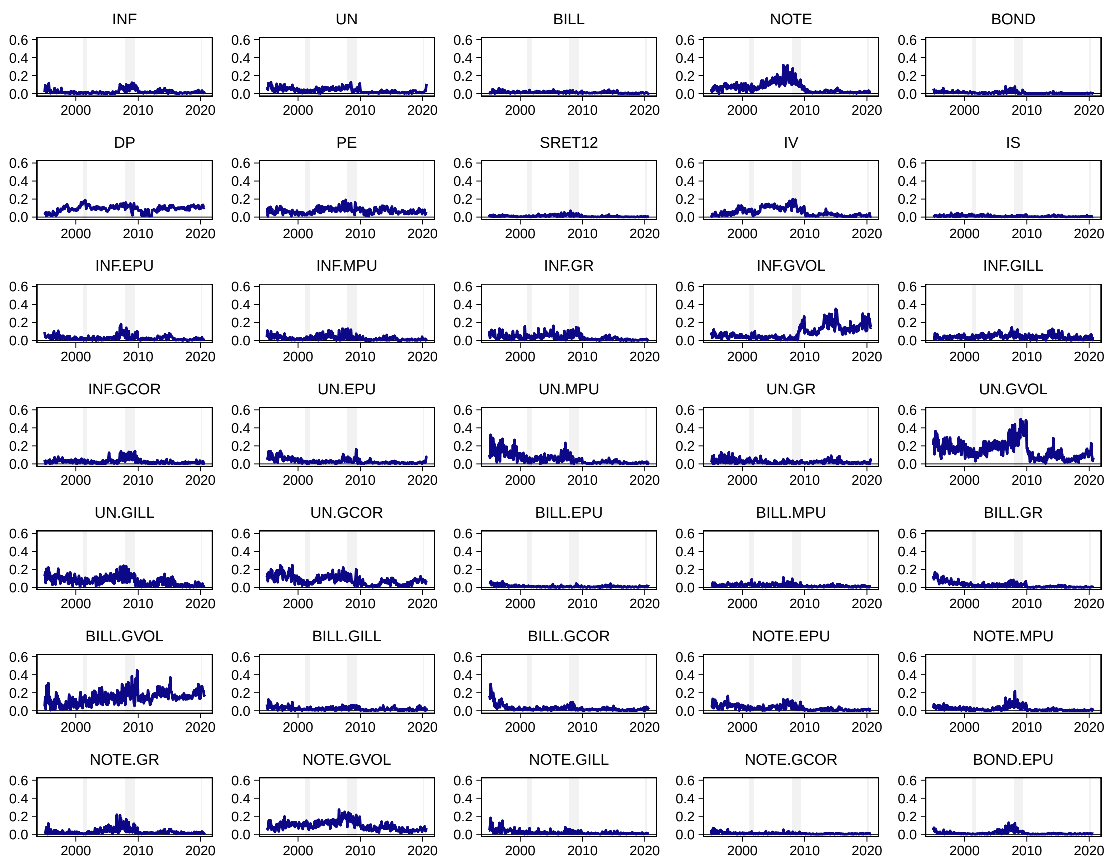


FIGURE IA.11 (continued)

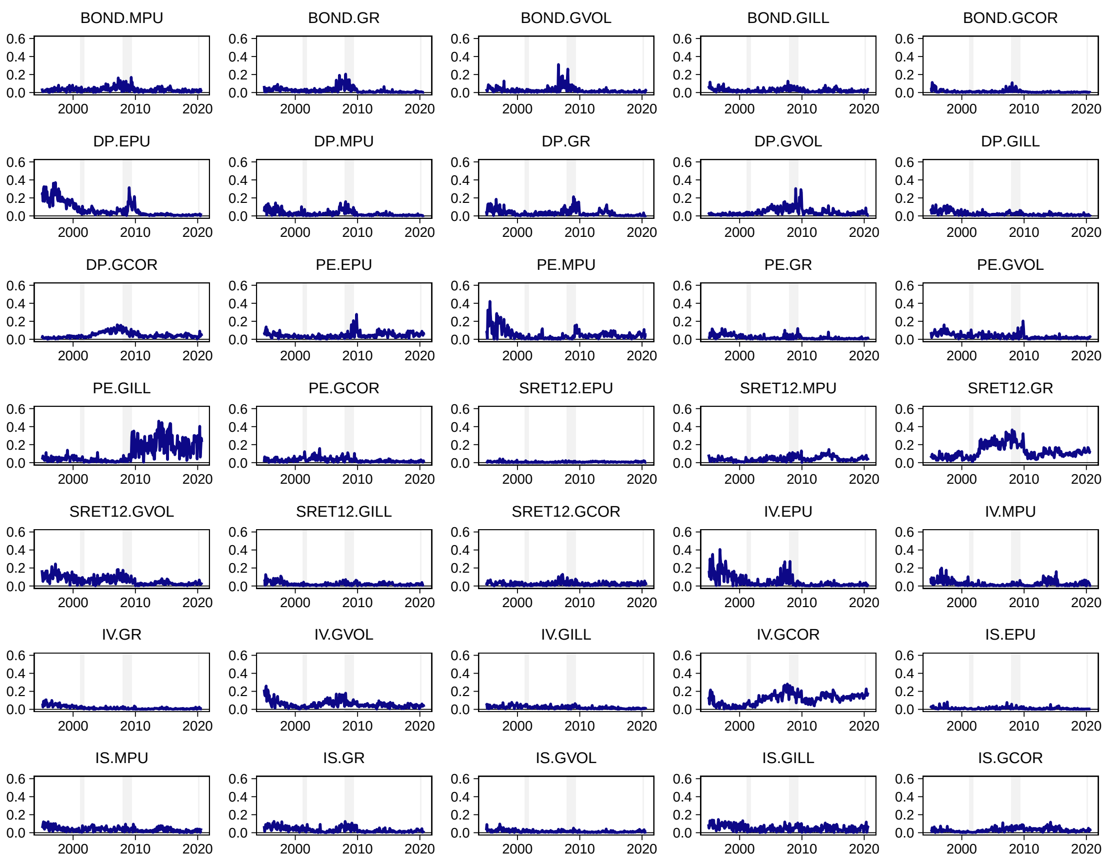


FIGURE IA.12

ML-Opt Portfolio Weights: Baseline ENet Forecasts

Each panel depicts weights for the country in the panel heading for the ML-Opt portfolio based on the baseline ENet forecast. The figure also includes weights for the Bench-Opt portfolio.

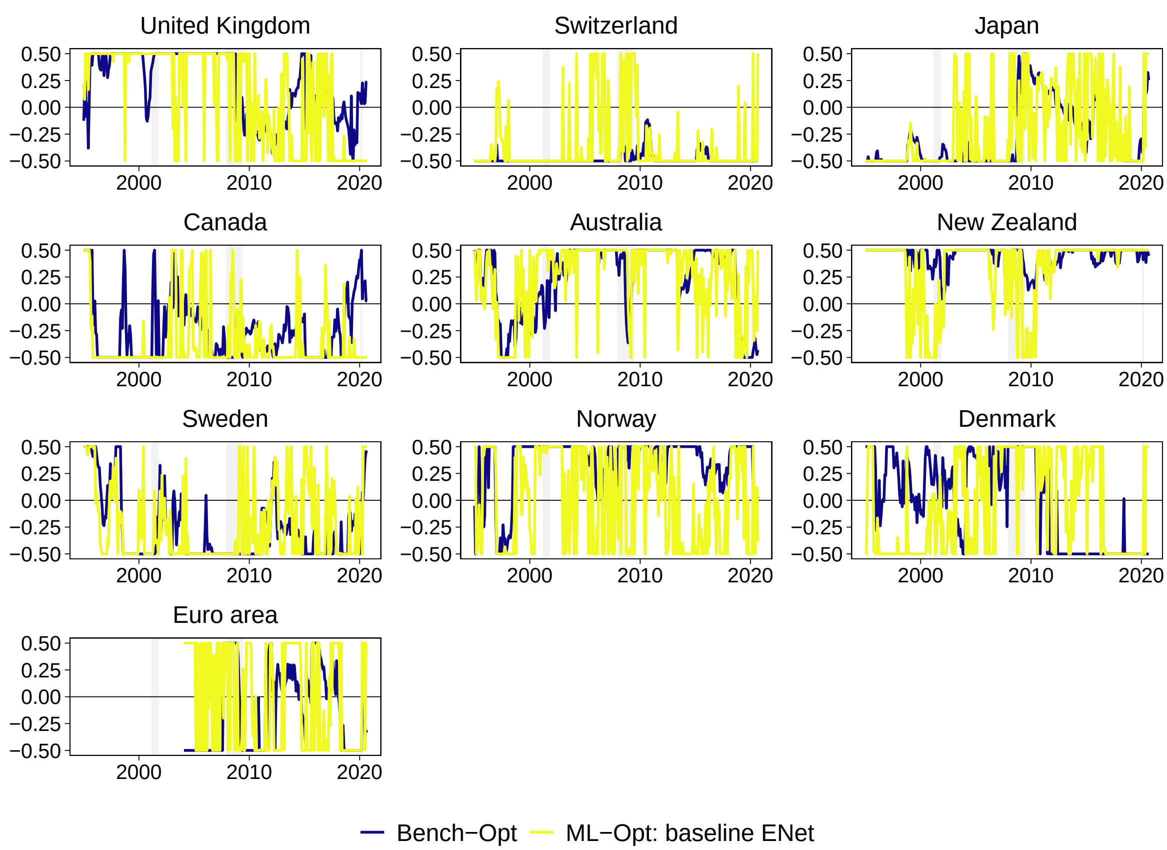


FIGURE IA.13

ML-Opt Portfolio Weights: DNN Forecasts

Each panel depicts weights for the country in the panel heading for the ML-Opt portfolio based on the DNN forecast. The figure also includes weights for the Bench-Opt portfolio.

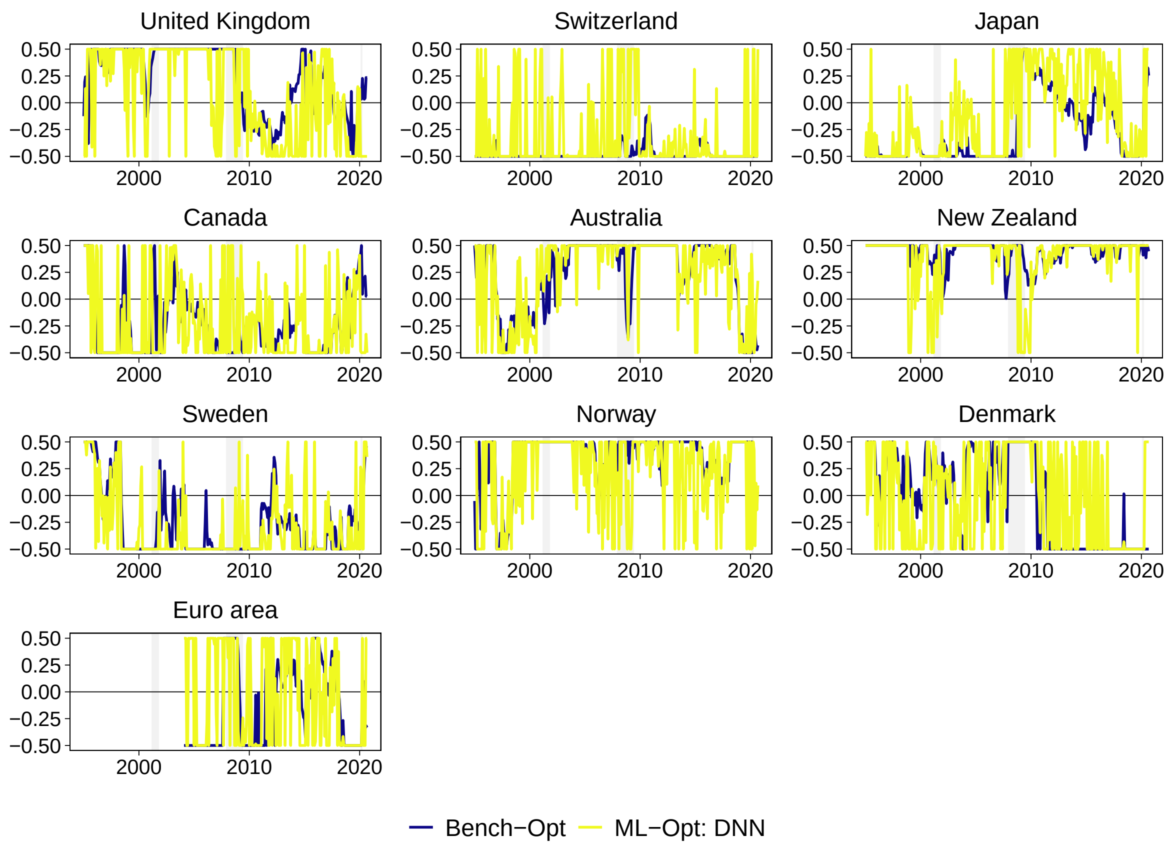


FIGURE IA.14

ML-Opt Portfolio Weights: Ensemble Forecasts

Each panel depicts weights for the country in the panel heading for the ML-Opt portfolio based on the ensemble forecast. The figure also includes weights for the Bench-Opt portfolio.

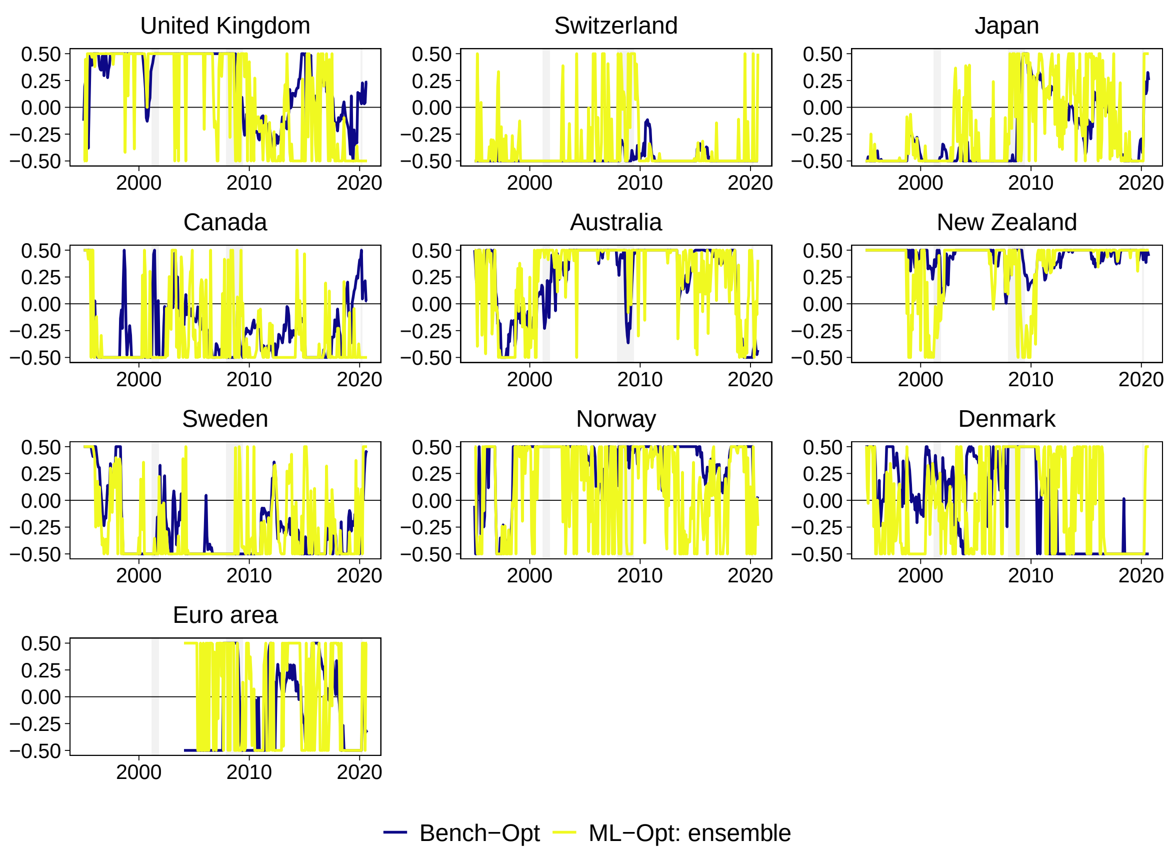


FIGURE IA.15

Currency Excess Return Forecasts in Late 2008: Baseline ENet

Each panel depicts currency excess return forecasts for the baseline ENet forecast of the exchange rate change, which serve as inputs for the ML-Opt portfolio, for the available currencies and month in the panel heading. Each panel also depicts currency excess return forecasts for the no-change benchmark forecast of the exchange rate change, which serve as inputs for the Bench-Opt portfolio, and realized currency excess returns.

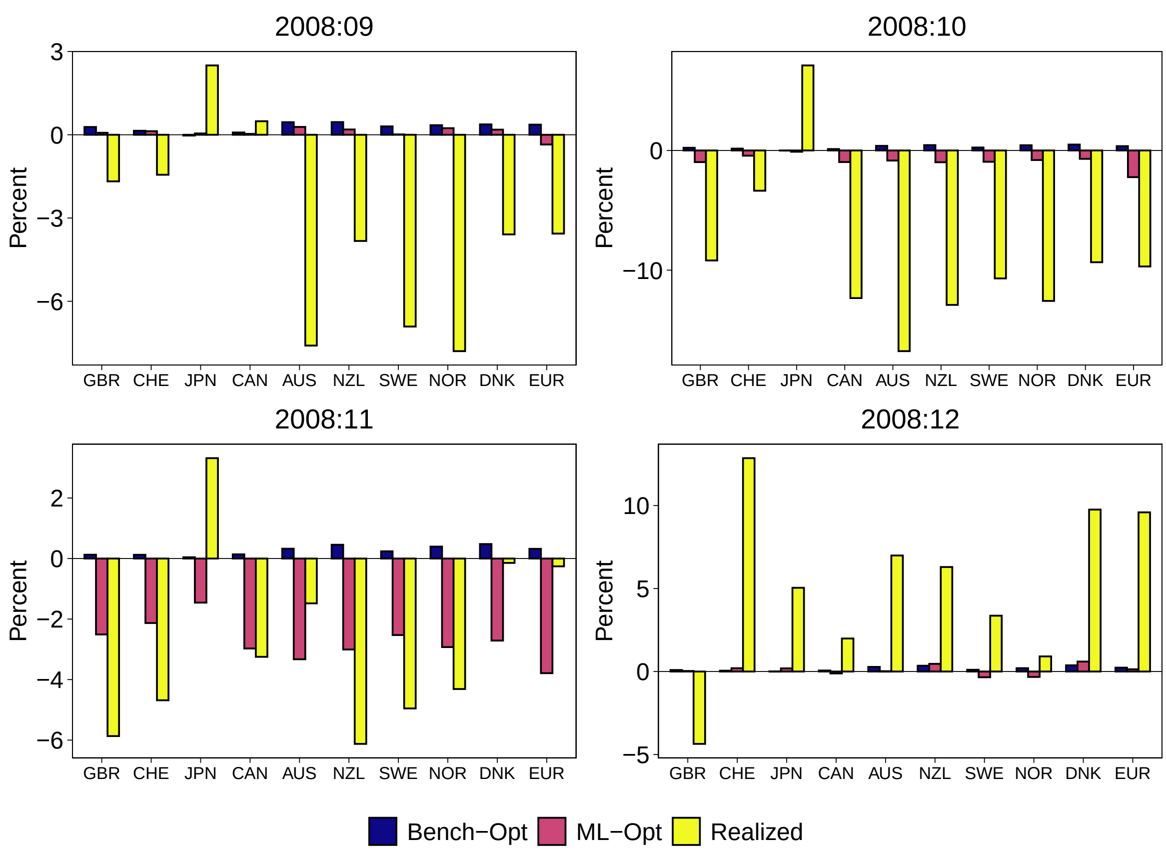


FIGURE IA.16

Currency Excess Return Forecasts in Late 2008: DNN

Each panel depicts currency excess return forecasts for the DNN forecast of the exchange rate change, which serve as inputs for the ML-Opt portfolio, for the available currencies and month in the panel heading. Each panel also depicts currency excess return forecasts for the no-change benchmark forecast of the exchange rate change, which serve as inputs for the Bench-Opt portfolio, and realized currency excess returns.

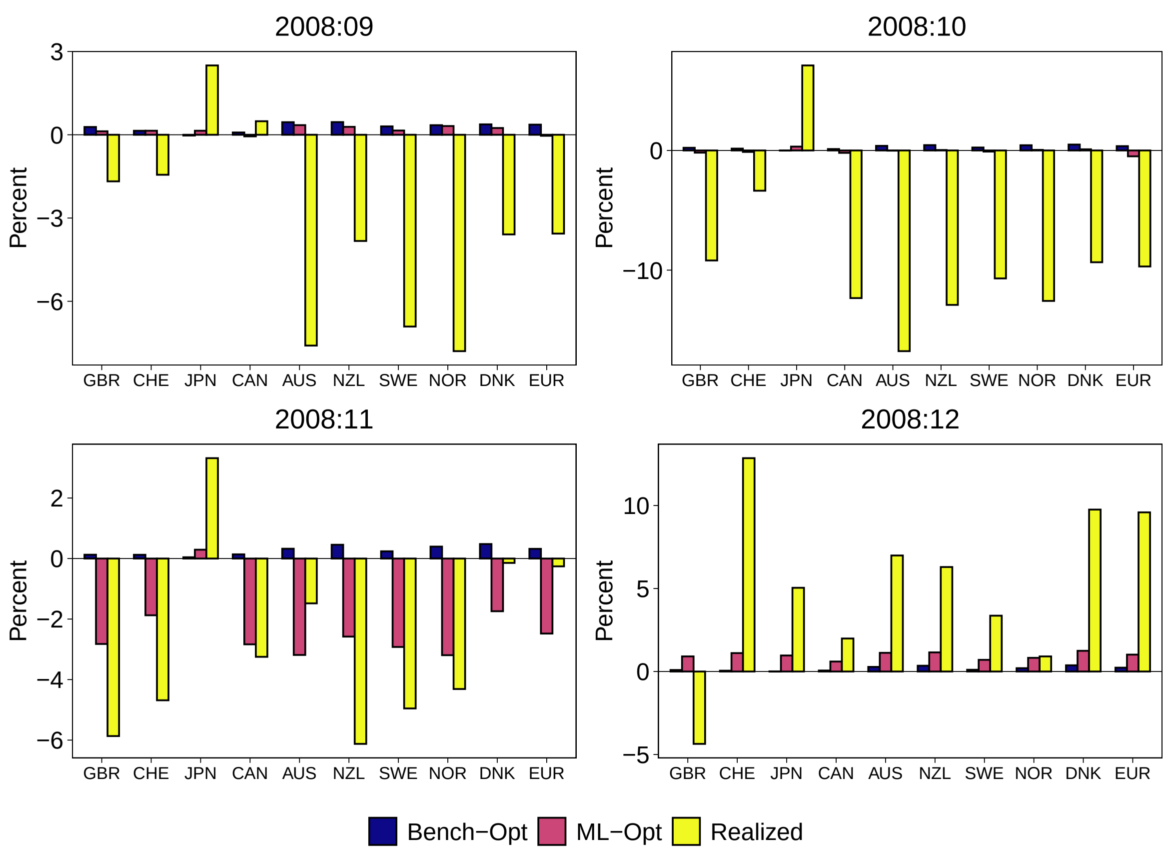


FIGURE IA.17

Currency Excess Return Forecasts in Late 2008: Ensemble

Each panel depicts currency excess return forecasts for the ensemble forecast of the exchange rate change, which serve as inputs for the ML-Opt portfolio, for the available currencies and month in the panel heading. Each panel also depicts currency excess return forecasts for the no-change benchmark forecast of the exchange rate change, which serve as inputs for the Bench-Opt portfolio, and realized currency excess returns.

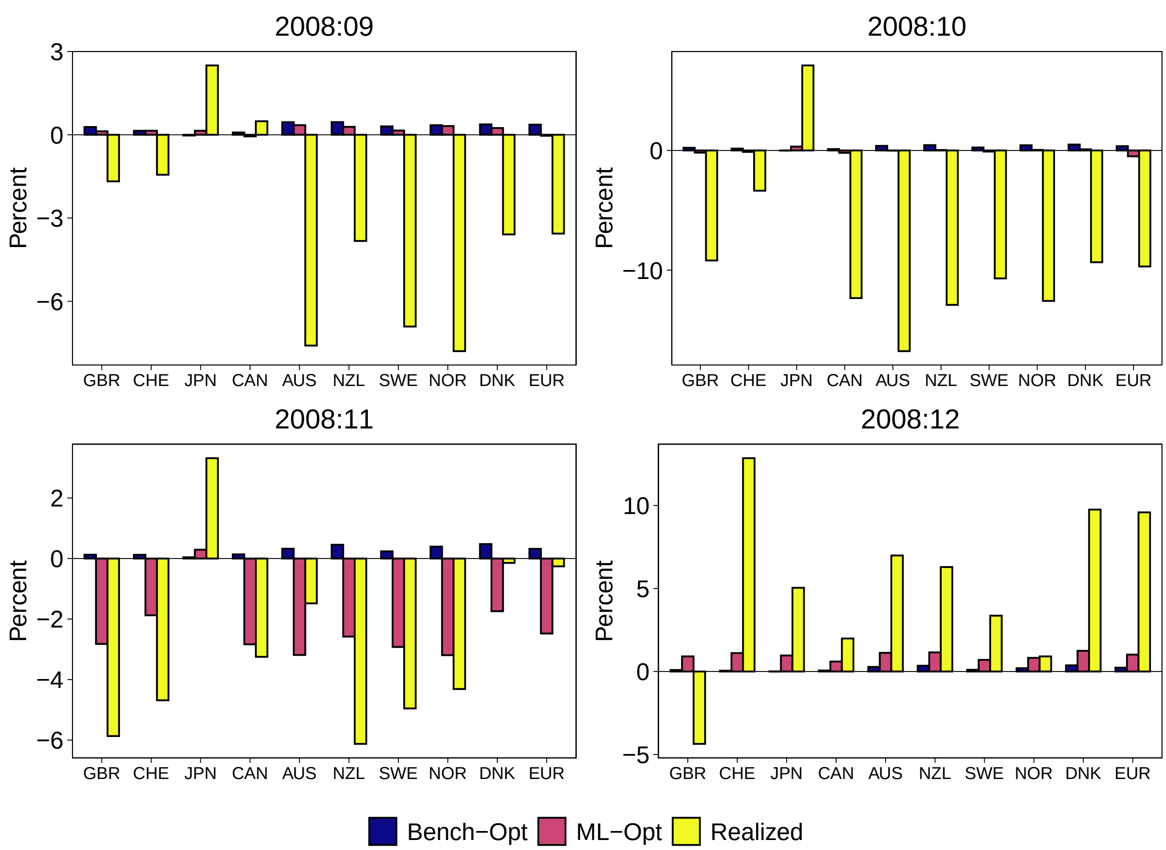


FIGURE IA.18

Portfolio Weights in Late 2008: Baseline ENet Forecast

Each panel depicts ML-Opt portfolio weights based on the baseline ENet forecast for the available currencies and month in the panel heading. Each panel also depicts Bench-Opt portfolio weights based on the no-change benchmark forecast.

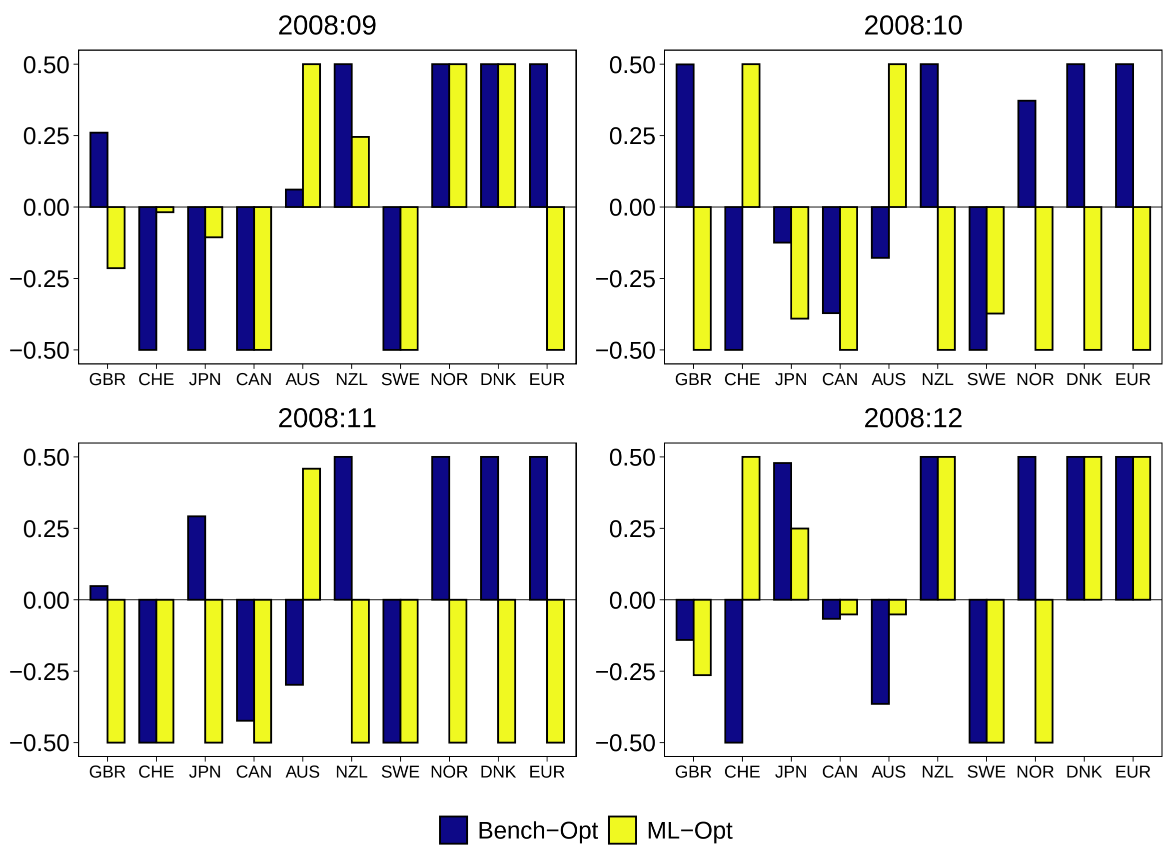


FIGURE IA.19

Portfolio Weights in Late 2008: DNN Forecast

Each panel depicts ML-Opt portfolio weights based on the DNN forecast for the available currencies and month in the panel heading. Each panel also depicts Bench-Opt portfolio weights based on the no-change benchmark forecast.

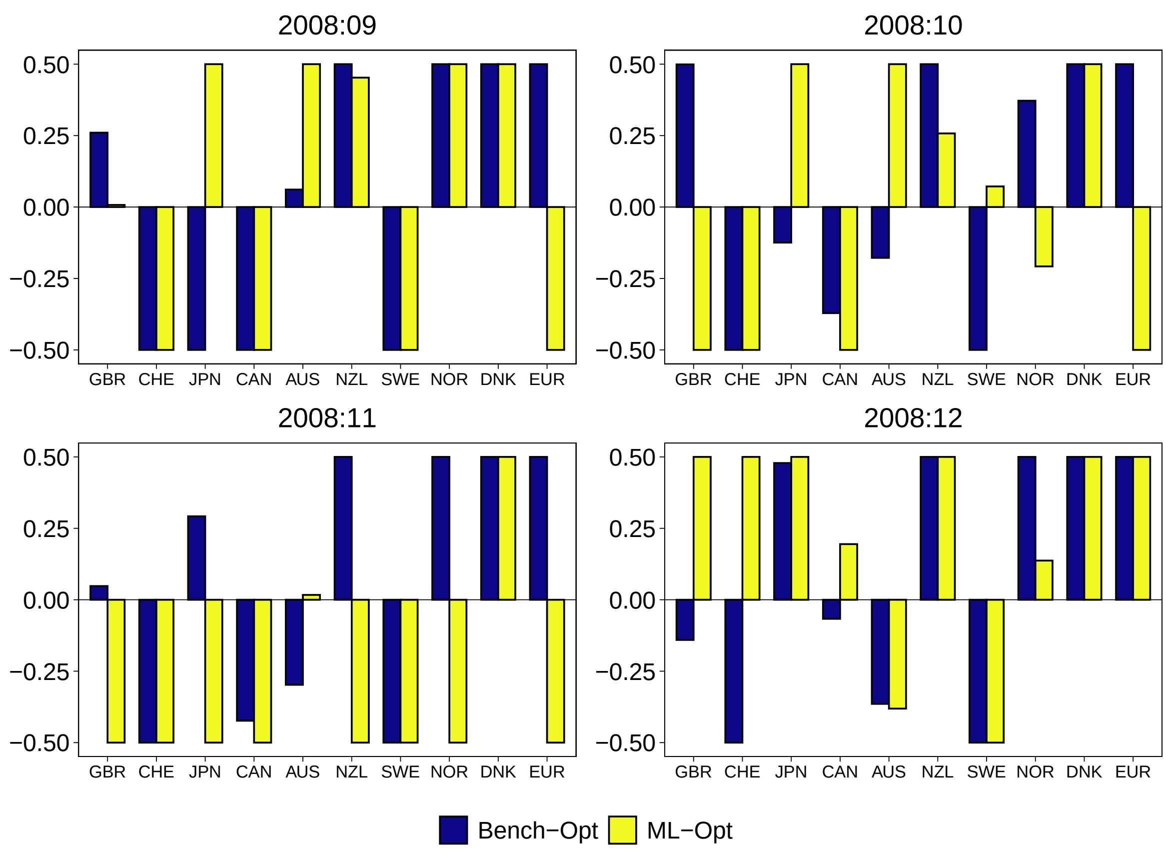


FIGURE IA.20

Portfolio Weights in Late 2008: Ensemble Forecast

Each panel depicts ML-Opt portfolio weights based on the ensemble forecast for the available currencies and month in the panel heading. Each panel also depicts Bench-Opt portfolio weights based on the no-change benchmark forecast.

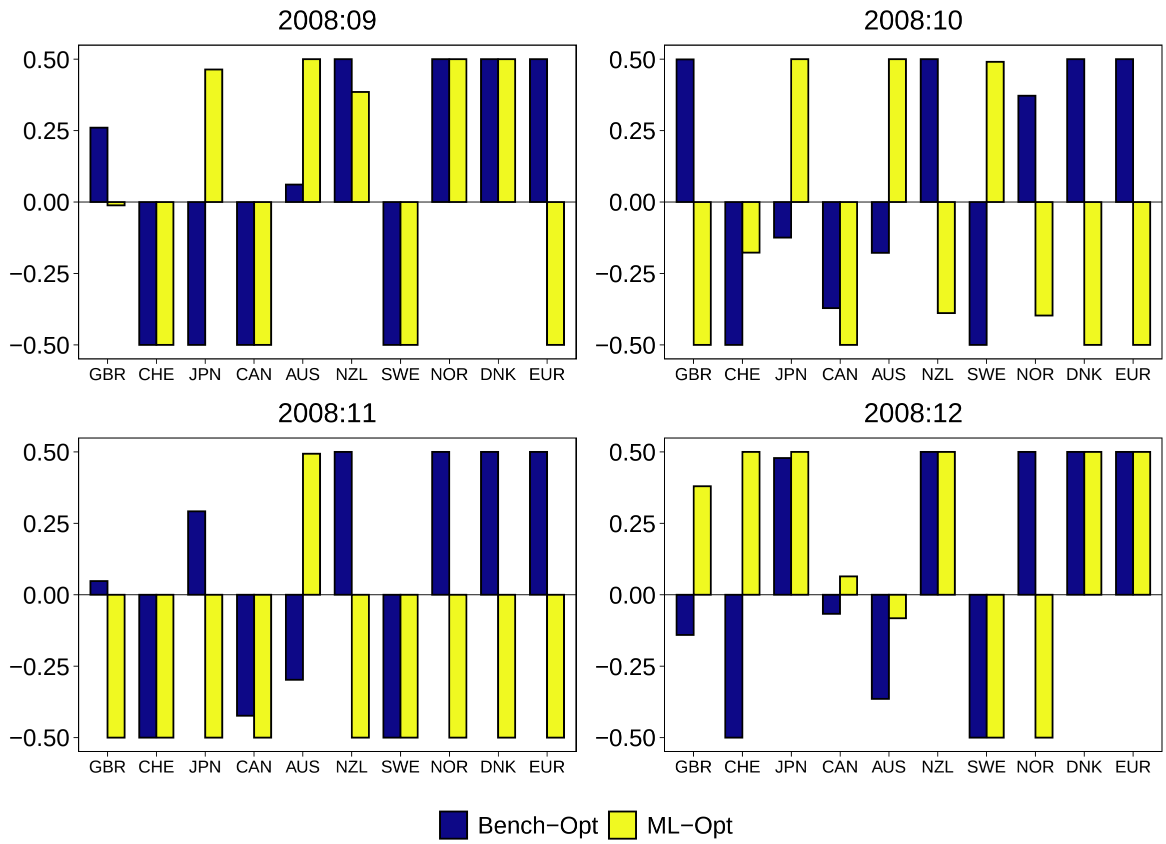


FIGURE IA.21

Conditional Maximum Sharpe Ratios

The figure depicts conditional maximum Sharpe ratios for currency excess return forecasts for the baseline ENet, DNN, ensemble, and no-change benchmark forecasts of the exchange rate change (considered in turn).