Out-of-Sample Exchange Rate Prediction: A Machine Learning Perspective

Ilias Filippou
Washington University in St. Louis
iliasfilippou@wustl.edu

David E. Rapach*
Saint Louis University
dave.rapach@gmail.com

Mark P. Taylor
Washington University in St. Louis
mark.p.taylor@wustl.edu

Guofu Zhou
Washington University in St. Louis
zhou@wustl.edu

April 5, 2022

*Corresponding author. Send correspondence to David Rapach, Chaifetz School of Business, Saint Louis University, 3674 Lindell Boulevard, St. Louis, MO 63108-3397; e-mail: dave.rapach@gmail.com.

We thank conference and seminar 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, 2021 Vienna Symposium on Foreign Exchange Markets, 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, as well as 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.

Out-of-Sample Exchange Rate Prediction: A Machine Learning Perspective

Abstract

We establish the out-of-sample predictability of monthly exchange rates via machine learning techniques based on 70 predictors capturing country characteristics, global variables, and their interactions. To better guard against overfitting in our high-dimensional and noisy data environment, we adjust "off-the-shelf" implementations of machine learning techniques to induce adequate shrinkage. The resulting forecasts consistently outperform the no-change benchmark, which has proven difficult to beat. Variable importance analysis indicates that country characteristics are important for forecasting, once they interact with global variables. Machine learning forecasts also markedly improve the performance of a carry trade portfolio, especially since the Global Financial Crisis.

JEL classifications: C45, F31, F37, G11, G12, G15

Keywords: Short-horizon exchange rate predictability, Panel predictive regression, Elastic net, Deep neural network, Variable importance, Partial dependence plot, Individual conditional expectation curves, Carry trade

1. Introduction

The specter of Meese and Rogoff (1983) continues to haunt international finance: despite an array of theoretical models linking fundamentals to exchange rates, it is difficult to outperform the no-change benchmark forecast of the exchange rate on an out-of-sample basis at short horizons, such as the monthly horizon. A possible explanation for this difficulty is that existing studies typically rely on only a few predictors, while a relatively large number of predictors is potentially relevant for anticipating future exchange rates. Furthermore, monthly exchange rate fluctuations likely contain only a small predictable component, similarly to equities or any asset traded in a reasonably competitive market. However, with respect to equity returns, the apparent consensus is that monthly out-of-sample stock return predictability exists to a statistically and economically significant degree (e.g., Rapach and Zhou 2013); in contrast, whether monthly out-of-sample exchange rate predictability exists remains an open question (e.g., Kilian and Taylor 2003; Rossi 2013).[1]

In this paper, we find that, similarly to equity returns, monthly exchange rates are predictable on an out-of-sample basis, once we employ a large set of predictor variables and appropriate methods. Specifically, we use a rich information set, panel framework, and machine learning techniques to construct monthly out-of-sample forecasts of US dollar exchange rates for a group of developed countries.[2] The information set is comprised of ten country characteristics and six global variables. The country characteristics include various macroeconomic and financial variables, such as inflation, interest rate, unemployment rate gap, and valuation ratio differentials, which can be motivated by relative purchasing power parity (PPP), uncovered interest parity (UIP), the Taylor (1993) rule, and uncovered equity parity (UEP), among other theories. The global variables include economic and monetary policy

[1] There is stronger evidence of out-of-sample exchange rate predictability at longer horizons, such as a year or more (e.g., Mark 1995; Eichenbaum, Johannsen, and Rebelo 2021).

[2] Machine learning is becoming popular in finance for predicting equity returns with large information sets (e.g., Rapach, Strauss, and Zhou 2013; Chinco, Clark-Joseph, and Ye 2019; Rapach et al. 2019; Freyberger, Neuhierl, and Weber 2020; Gu, Kelly, and Xiu 2020; Kozak, Nagel, and Santosh 2020; Bryzgalova, Pelger, and Zhu 2021; Chen, Pelger, and Zhu 2021; Cong et al. 2021; Han et al. 2021; Liu, Zhou, and Zhu 2021; Dong et al. 2022; Avramov, Cheng, and Metzker forthcoming).

uncertainty indices (Baker, Bloom, and Davis 2016), a geopolitical risk index (Caldara and Iacoviello forthcoming), as well as measures of global foreign exchange (FX) volatility, illiquidity, and correlation (Menkhoff et al. 2012a; Mueller, Stathopoulos, and Vedolin 2017). To allow for the predictive relationships between the country characteristics and exchange rate changes to vary with global conditions, we interact the country characteristics with the global variables, producing a set of 70 predictors for our panel predictive regression framework.[3] Previous studies typically focus on a limited number of country characteristics and do not consider interactions with global variables, which we find to be important. Considering an extensive set of predictors and including all of them in the model recognizes that we cannot know a priori the most relevant exchange rate predictors.

We begin with a linear panel specification for generating out-of-sample forecasts. Given our high-dimensional setting with 70 predictors and the intrinsically large unpredictable component in monthly exchange rate changes, it is vital to guard against overfitting when estimating the panel predictive regressions. By construction, conventional ordinary least squares (OLS) estimation maximizes the fit of the model over the estimation (or training) sample, which can lead to overfitting the model to the training sample and thus poor out-of-sample performance. To mitigate overfitting, we estimate the panel predictive regressions via the elastic net (ENet; Zou and Hastie 2005), a refinement of the popular least absolute shrinkage and selection operator (LASSO; Tibshirani 1996) from machine learning. The ENet is a penalized regression technique that shrinks the estimated coefficients towards zero, thereby alleviating overfitting. The penalty term in the ENet includes both an ℓ_1 component—as in the LASSO—and an ℓ_2 component—as in ridge regression (Hoerl and Kennard 1970); the former permits shrinkage to zero, so that the ENet also performs variable selection. Instead of conventional M-fold cross validation, we use the extended regularization information criterion (ERIC; Hui, Warton, and Foster 2015), which is a modification of the Bayesian information criterion (BIC; Schwarz 1978), to select (or tune) the ENet's shrinkage

[3]In the context of predicting individual stock returns, Gu, Kelly, and Xiu (2020) similarly consider a large number of firm characteristics, which they interact with a set of macroeconomic variables.

hyperparameter (λ). The ERIC is a more stringent method for tuning λ, in the sense that it tends to select a larger value for λ and thus induces more shrinkage, thereby better guarding against overfitting in our high-dimensional and noisy data environment.

We further guard against overfitting by imposing additional restrictions, which can be viewed as economic constraints. After standardizing the predictors, we set the intercept terms in the panel predictive regressions to zero, thereby imposing the restrictions that the average exchange rate changes are zero. These restrictions are consistent with the data and help to guard against overfitting by reducing the parameter space.[4] We also pool the data in our panel framework, which imposes slope homogeneity restrictions across countries to further reduce the parameter space. Reducing the number of parameters that we need to estimate helps to improve out-of-sample performance in light of the bias-variance trade-off.

To permit nonlinearities when predicting exchange rates, we also generate forecasts based on a deep neural network (DNN). DNNs are popular machine learning models that allow for complex nonlinear predictive relationships via a network architecture with multiple hidden layers containing neurons activated by predictive signals. To better guard against overfitting, we move beyond an "off-the-shelf" implementation of deep learning by setting the intercept terms for the weights to zero.[5] In addition, we include ℓ_1 and ℓ_2 penalty components in the objective function and employ dropout (Hinton et al. 2012; Srivastava et al. 2014) when training the DNN. We also consider an ensemble forecast that takes the average of the linear ENet and DNN forecasts. The ensemble forecast recognizes that it is difficult to know a priori the best individual forecast, and it allows us to take advantage of potential complementarities between the linear ENet and DNN forecasts. To the best of our knowledge, we are the first to simultaneously employ the ENet, deep learning, and an ensemble of linear and nonlinear forecasts for out-of-sample exchange rate prediction.

[4] Nagel (2021) emphasizes the importance of using economic insights when applying machine learning tools for empirical asset pricing.

[5] The parameters of a DNN are typically referred to as weights.

Mimicking the situation of a forecaster in real time, we generate monthly out-of-sample forecasts based on the 70 predictors by recursively estimating the linear panel predictive regression and DNN models each month. Based on data availability (and after allowing for a ten-year initial training sample), the out-of-sample period spans 1995:01 to 2020:09. We compute forecasts for the entire out-of-sample period for the United Kingdom, Switzerland, Japan, Canada, Australia, New Zealand, Sweden, Norway, and Denmark; for the Euro area, the out-of-sample period begins in 2000:02. We refer to the exchange rates for these ten countries as the G10.[6] For Germany, Italy, France, and the Netherlands, the out-of-sample period ends in 1998:12, corresponding to their adoption of the Euro.

We find that the information in the rich set of predictors is indeed useful for outperforming the stringent no-change benchmark forecast over the 1995:01 to 2020:09 out-of-sample period, provided that we adequately guard against overfitting. As expected, a forecast based on conventional OLS estimation of the linear panel predictive regression substantially underperforms the no-change benchmark in terms of mean squared prediction error (MSPE) for all countries, a clear manifestation of overfitting. In contrast, the linear ENet, DNN, and ensemble forecasts outperform the no-change benchmark for twelve, 13, and all 14 countries, respectively. Based on the Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistic, which measures the proportional decrease in MSPE for a competing forecast vis-à-vis a benchmark, the improvements in forecast accuracy are quantitatively large in the context of the extensive literature surveyed by Rossi (2013). The improvements in MSPE vis-à-vis the no-change benchmark are also statistically significant in many cases according to the Clark and West (2007) test. The ensemble forecast delivers the best overall performance: the R^2_{OS} statistics are positive for all of the countries, are above 1% (2%) for ten (six) countries, and reach as high as 3.12% (for the United Kingdom); the R^2_{OS} statistic is also a sizable 1.88% (significant at the 1% level) for the entire group of countries taken together.

[6] The G11 currencies are the US dollar, Euro, British pound, Swiss franc, Japanese yen, Canadian dollar, Australian dollar, New Zealand dollar, Swedish krona, Norwegian krone, and Danish krone. With the US dollar serving as the base currency, we label our set of ten exchange rates the G10.

Based on the graphical device of Goyal and Welch (2003, 2008), the machine learning forecasts outperform the no-change benchmark on a consistent basis over time. The outperformance is particularly strong during the worst phase of the Global Financial Crisis in late 2008. The finding that out-of-sample exchange rate predictability is stronger during periods of crisis is reminiscent of the literature on aggregate stock market return predictability, which finds that out-of-sample predictability is stronger during deep recessions (e.g., Rapach, Strauss, and Zhou 2010; Henkel, Martin, and Nardari 2011; Dangl and Halling 2012).

Overall, by utilizing a rich information set—while adequately guarding against overfitting—our machine learning approach makes considerable progress in solving the short-horizon no-predictability puzzle, providing among the best monthly exchange rate forecasts available to date.

We further extend our analysis along two dimensions. First, we use variable importance analysis to glean insight into the economic forces driving the ensemble forecast. Specifically, we use the recently developed variable importance metric of Greenwell, Boehmke, and McCarthy (2018), which is based on partial dependence plots (PDPs; Friedman 2001). We find that popular predictors from the literature, such as inflation and government bill yield differentials, are important when they interact with global FX volatility. We also find that the importance of a number of predictors increases substantially over time, often near the advent of the Global Financial Crisis, indicating that the underlying data-generating process evolves over time. This is perhaps not surprising, as central banks around the world engaged in historically unprecedented actions in response to the Global Financial Crisis that had significant international effects (e.g., Neely 2015). In addition, we compute individual conditional expectation (ICE) curves (Goldstein et al. 2015) and corresponding PDPs. Among other things, the ICE cuves show that the relationships predicted by relative PPP and UIP hold to a greater degree as global FX volatility increases.[7]

[7] Our finding that UIP holds to a greater degree as FX volatility increases is reminiscent of the finding of Clarida, Davis, and Pedersen (2009) that the slope coefficients in UIP regressions change across low and high option-implied volatility regimes.

Second, we explore the economic implications of out-of-sample exchange rate predictability for carry trade investment strategies, which receive considerable attention from both academic researchers and practitioners. The popular carry trade entails going long (short) currencies with relatively high (low) interest rates. Beginning with Hansen and Hodrick (1980), Bilson (1981), and Fama (1984), a voluminous literature finds that UIP does not hold, creating scope for a profitable carry trade.[8] Compared to a variety of investment strategies, conventional carry trade portfolios deliver impressive Sharpe ratios prior to the Global Financial Crisis (e.g., Burnside, Eichenbaum, and Rebelo 2011; Lustig, Roussanov, and Verdelhan 2011).[9] However, they suffered large losses in late 2008, and their performance has since deteriorated in the wake of the Global Financial Crisis (e.g., Melvin and Taylor 2009; Jordà and Taylor 2012; Daniel, Hodrick, and Lu 2017; Melvin and Shand 2017).[10]

The usual carry trade simply uses the bill yield differential to forecast the currency excess return, so that—in the spirit of Meese and Rogoff (1983)—it sets the forecast of the exchange rate change to zero (Jordà and Taylor 2012). In this vein, we construct an optimal portfolio for a mean-variance investor who allocates across available foreign currencies by relying on bill yield differentials to forecast foreign currency excess returns. We label this an unconditional optimal (U-OPT) portfolio, as the investor ignores exchange rate predictability when forecasting excess returns and sets the expected exchange rate change to zero. The U-OPT portfolio delivers impressive performance before the Global Financial Crisis. However, it suffers large losses in late 2008, and its cumulative return is essentially flat thereafter.[11]

We also construct a conditional optimal (C-OPT) portfolio, in which the mean-variance investor augments the bill yield differential with the ensemble forecast of the exchange rate

[8] See Froot and Thaler (1990), Taylor (1995), and Burnside (2018) for surveys of UIP.

[9] Studies that explore risk-based explanations for carry trade returns include Burnside et al. (2011), Lustig, Roussanov, and Verdelhan (2011), Menkhoff et al. (2012a), Dobrynskaya (2014), Jurek (2014), Lettau, Maggiori, and Weber (2014), Dahlquist and Hasseltoft (2020), and Ready, Roussanov, and Ward (forthcoming).

[10] Brunnermeier, Nagel, and Pedersen (2009) provide an explanation for carry trade crashes based on funding-constrained speculators.

[11] Daniel, Hodrick, and Lu (2017) also construct an optimal carry trade portfolio for a mean-variance investor who uses interest rate differentials to forecast currency excess returns. Our U-OPT portfolio performs similarly to theirs. A conventional carry trade portfolio that sorts currencies based on bill yield differentials and goes long (short) the fifth (first) quintile performs even worse than the U-OPT portfolio.

change to forecast the currency excess return, thereby attempting to exploit exchange rate predictability by using the conditioning information in the predictors when allocating across currencies.[12] The C-OPT portfolio provides substantial economic value to the investor vis-à-vis the U-OPT portfolio, generating an annualized increase in certainty equivalent return of 340 basis points. The superior performance of the C-OPT portfolio is evident both before and after the Global Financial Crisis, although it is especially apparent starting in late 2008. The C-OPT portfolio experiences a smaller loss than the U-OPT portfolio in September of 2008, generates much larger gains in the last three months of 2008, and performs well subsequently. Consistent with a safe-haven role for the US dollar, the ensemble forecast predicts a substantial depreciation for many foreign currencies in late 2008, which leads to markedly different allocations for the C-OPT vis-à-vis the U-OPT portfolio. The C-OPT portfolio also generates substantial alpha both before and after the crisis in the context of the Lustig, Roussanov, and Verdelhan (2011) currency factor model.

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 discusses the specification and estimation of the models used to generate the out-of-sample exchange rate forecasts. Section 4 presents our main results on the accuracy of the out-of-sample exchange rate forecasts. Section 5 provides results for variable importance analysis and ICE curves. Section 6 discusses the construction of the U-OPT and C-OPT carry trade portfolios and analyzes their performance. Section 7 concludes.

2. Data

This section describes the data used in our analysis. Section A1 of the Internet Appendix provides further details on the data sources and construction of the variables.

[12] The results are similar when the investor uses the linear ENet or DNN forecast (as reported in Section 6). Della Corte, Sarno, and Tsiakas (2009) construct mean-variance optimal portfolios for a US investor who allocates across US, British, German, and Japanese short-term bonds using a handful of fundamentals to forecast exchange rates. Jordà and Taylor (2012) use a small number of fundamentals to improve carry trade strategies (but not in a mean-variance optimal framework). Della Corte, Jeanneret, and Patelli (2020) consider a mean-variance investor who allocates between US short-term bonds and a Euro-denominated cash account using a credit-implied risk premium to forecast the exchange rate.

2.1. 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 et al. 2011; Lustig, Roussanov, and Verdelhan 2011). Our sample consists of the following 14 countries: the United Kingdom, Switzerland, Japan, Canada, Australia, New Zealand, Sweden, Norway, Denmark, the Euro area, Germany, Italy, France, and the Netherlands.[13] Germany, Italy, France, and the Netherlands are replaced by the Euro area after the Euro's introduction in 1999. We refer to the group of ten countries excluding Germany, Italy, France, and the Netherlands as the G10.

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, Roussanov, and Verdelhan 2011; Menkhoff et al. 2012a,b). 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 $s_{i,t} = \log(S_{i,t})$. Table 1 reports summary statistics for the 14 log exchange rate changes. The second column reports the sample period for each country. With the exceptions of 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 of the countries, with the exception of the Euro area, where the sample begins in 1999:02, corresponding to the introduction of the Euro in January of 1999.

The annualized means in the third column of Table 1 are generally small in magnitude. In fact, none of the means is significant at conventional levels. The annualized volatilities in the fourth column are typically sizable; apart from Canada (7.34%), they range from 9.55% (Euro area) to 12.04% (New Zealand). With the exceptions of Switzerland and Japan, the exchange rate changes are positively skewed (fifth column), while all of the exchange rate

[13] Our universe of exchange rates is similar to the sample of developed countries employed in other studies (e.g., Lustig, Roussanov, and Verdelhan 2011; Menkhoff et al. 2012a), with the exception of Belgium, which we exclude due to a lack of data availability for the country characteristics in Section 2.2.

(1) Country	(2) Sample period	(3) Ann. mean	(4) Ann. vol.	(5) Skewness	(6) Excess kurtosis	(7) 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 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 $\left(\sqrt{12}\right)$.

changes exhibit excess kurtosis (sixth column). The autocorrelations in the last column are all relatively small in magnitude. Overall, the summary statistics in Table 1 reflect well-known empirical features of exchange rates.

2.2. Country Characteristics

We consider ten monthly country characteristics computed using macroeconomic and financial data from Global Financial Data and the Organization for Economic Cooperation and Development:

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

Unemployment rate gap differential ($\mathrm{UN}_{i,t}$). 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.[14]

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

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

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

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

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

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

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

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

[14] 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.

The country characteristics include a variety of macroeconomic and financial variables, all of which are based on data readily available to FX market participants.[15] The inflation differential relates to relative PPP, while the inflation and unemployment rate gap differentials constitute Taylor (1993) rule fundamentals (e.g., Engel and West 2005; Molodtsova and Papell 2009), which appear to perform better for exchange rate prediction than fundamentals based on the traditional monetary model (Frenkel 1976; Mussa 1976). The bill yield differential relates to the voluminous UIP literature, while longer-term yield differentials are considered by Ang and Chen (2011) and Chen and Tsang (2013) in the context of yield curves. Hau and Rey (2006) and Cenedese et al. (2016), among others, employ valuation ratio differentials to analyze UEP. The other country characteristics represent additional financial variables that are potentially relevant to market participants.

2.3. Global Variables

We consider six global variables:

Economic policy uncertainty (EPU_t). Baker, Bloom, and Davis (2016) economic policy uncertainty index based on coverage frequencies in ten major US newspapers.

Monetary policy uncertainty (MPU_t). Baker, Bloom, and Davis (2016) monetary policy uncertainty index based on coverage frequencies in ten major US newspapers.

Geopolitical risk (GR_t). Caldara and Iacoviello (forthcoming) geopolitical risk index based on newspaper coverage.

Global FX volatility (GVOL_t). Following Menkhoff et al. (2012a), global FX volatility is the average for the month of the daily cross-sectional averages of the absolute values of log exchange rate changes.

[15] We account for the publication lag in the consumer price index and unemployment rate, so that $\mathrm{INF}_{i,t}$ and $\mathrm{UN}_{i,t}$ correspond to data for month $t-1$ that are reported in month t.

Global FX illiquidity (GILL_t). Following Menkhoff et al. (2012a), global FX illiquidity is the average for the month of the daily cross-sectional averages of the bid-ask spreads for the currencies.

Global FX correlation (GCOR_t). Similarly to Mueller, Stathopoulos, and Vedolin (2017), we measure global FX correlation as the average of the realized correlations for all currency pairs computed using daily log currency excess returns for the month.

The global variables capture general economic conditions that potentially affect the predictive ability of the country characteristics.[16]

3. Panel Predictive Regressions

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

3.1. Linear Specification

We collect the month-t characteristics for country i and month-t global variables in the following vectors:

$$\underset{(Z\times 1)}{\boldsymbol{z}_{i,t}} = [\ \text{INF}_{i,t} \quad \text{UN}_{i,t} \quad \text{BILL}_{i,t} \quad \text{NOTE}_{i,t} \quad \text{BOND}_{i,t} \quad \text{DP}_{i,t} \quad \text{PE}_{i,t} \quad \text{SRET12}_{i,t} \quad \text{IV}_{i,t} \quad \text{IS}_{i,t}\ ]', \tag{1}$$

$$\underset{(G\times 1)}{\boldsymbol{g}_{t}} = [\ \text{EPU}_t \quad \text{MPU}_t \quad \text{GR}_t \quad \text{GVOL}_t \quad \text{GILL}_t \quad \text{GCOR}_t\ ]', \tag{2}$$

respectively, for $i = 1, \dots, N$ and $t = 1, \dots, T$, where N (T) is the number of countries (time-series observations). The vector of predictors for country i is comprised of the country

[16] We follow Menkhoff et al. (2012a) and Mueller, Stathopoulos, and Vedolin (2017) by measuring GVOL_t, GILL_t, and GCOR_t as the residuals from fitted first-order autoregressive processes. We only use data available at the time of forecast formation when fitting the autoregressive processes and computing the residuals. Bakshi and Panayotov (2013) and Filippou and Taylor (2017) use aggregated country characteristics and global variables to predict conventional carry trade portfolio returns, while we forecast individual exchange rate changes.

characteristics and the characteristics interacted with each global variable:

$$\underset{(K\times 1)}{\boldsymbol{x}_{i,t}} = [\ \boldsymbol{z}_{i,t}' \quad \boldsymbol{h}_{i,t}'\ ]', \tag{3}$$

where

$$\underset{(ZG\times 1)}{\boldsymbol{h}_{i,t}} = \boldsymbol{z}_{i,t} \otimes \boldsymbol{g}_t, \tag{4}$$

$\otimes$ is the Kronecker product, and $K = Z(G+1)$. Since $Z = 10$ and $G = 6$ in Equations (1) and (2), respectively, we have $K = 70$ predictors for each country. We standardize each of the country-i predictors using its county-specific mean and variance:

$$\tilde{\boldsymbol{x}}_{i,t} = (\boldsymbol{x}_{i,t} - \bar{\boldsymbol{x}}_{i\cdot}) \oslash \hat{\boldsymbol{\sigma}}_{i\cdot}, \tag{5}$$

where

$$\bar{\boldsymbol{x}}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{i,t}, \tag{6}$$

$$\hat{\boldsymbol{\sigma}}_{i\cdot} = \left[\frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{x}_{i,t} - \bar{\boldsymbol{x}}_{i\cdot})^2\right]^{0.5}, \tag{7}$$

$\oslash$ indicates element-wise division, and the power operations in Equation (7) are element-wise. The panel predictive regression is given by

$$\Delta s_{i,t+1} = \tilde{\boldsymbol{x}}_{i,t}'\boldsymbol{b} + \varepsilon_{i,t+1} \quad \text{for } i = 1, \ldots, N;\ t = 1, \ldots, T, \tag{8}$$

where $\boldsymbol{b} = [\ b_1 \quad \cdots \quad b_K\ ]'$ is the K-vector of slope coefficients, and $\varepsilon_{i,t+1}$ is a zero-mean disturbance term. Equation (8) pools the data, thereby imposing slope homogeneity restrictions. Such restrictions reduce the parameter space to help guard against overfitting. Furthermore, we do not include an intercept term in Equation (8). Because the predictors are in deviation form, this means that we effectively impose the restrictions that the country-specific means for $\Delta s_{i,t}$ for $i = 1, \ldots, N$ are zero.[17] These restrictions are consistent with the data and further reduce the number of parameters that we need to estimate, again helping to guard against overfitting.

It is convenient to express the panel predictive regression in matrix notation as

$$\Delta \boldsymbol{s} = \tilde{\boldsymbol{X}} \boldsymbol{b} + \boldsymbol{\varepsilon}, \tag{9}$$

where

$$\underset{(NT\times 1)}{\Delta \boldsymbol{s}} = [\ \Delta \boldsymbol{s}_{1\cdot}' \quad \cdots \quad \Delta \boldsymbol{s}_{N\cdot}'\ ]', \tag{10}$$

$$\underset{(T\times 1)}{\Delta \boldsymbol{s}_{i\cdot}} = [\ \Delta s_{i,2} \quad \cdots \quad \Delta s_{i,T+1}\ ]', \tag{11}$$

$$\underset{(NT\times K)}{\tilde{\boldsymbol{X}}} = [\ \tilde{\boldsymbol{X}}_{1\cdot}' \quad \cdots \quad \tilde{\boldsymbol{X}}_{N\cdot}'\ ]', \tag{12}$$

$$\underset{(T\times K)}{\tilde{\boldsymbol{X}}_{i\cdot}} = [\ \tilde{\boldsymbol{x}}_{i,1} \quad \cdots \quad \tilde{\boldsymbol{x}}_{i,T}\ ]', \tag{13}$$

$$\underset{(NT\times 1)}{\boldsymbol{\varepsilon}} = [\ \boldsymbol{\varepsilon}_{1\cdot}' \quad \cdots \quad \boldsymbol{\varepsilon}_{N\cdot}'\ ]', \tag{14}$$

$$\underset{(T\times 1)}{\boldsymbol{\varepsilon}_{i\cdot}} = [\ \varepsilon_{i,2} \quad \cdots \quad \varepsilon_{i,T+1}\ ]'. \tag{15}$$

For simplicity, we assume a balanced panel in the notation. When we estimate $\boldsymbol{b}$ in Equation (9) for our application, we have an unbalanced panel at some points; it is straightforward to adjust the notation accordingly.

Note that we include more predictors than existing studies of exchange rate predictability, which typically consider only a limited number of country characteristics. This is important, as we cannot know a priori which characteristics are the most relevant for forecasting exchange rates. It also helps to guard against data mining or snooping, as we include an

[17] In other words, Equation (8) is tantamount to a fixed-effects specification in which the country-specific means for the log exchange rate changes are all zero.

extensive set of potentially relevant characteristics in the model from the start. In addition to numerous country characteristics, we include interactions of the country characteristics with a set of global variables. To the best of our knowledge, such interactions have not been considered in the literature on out-of-sample exchange rate predictability.

An out-of-sample forecast of $\Delta s_{i,T+1}$ based on the panel predictive regression in Equation (8) and data available through T is given by

$$\widehat{\Delta s}_{i,T+1|T} = \tilde{\boldsymbol{x}}_{i,T}' \hat{\boldsymbol{b}}_{1:T}, \tag{16}$$

where $\hat{\boldsymbol{b}}_{1:T}$ is an estimate of $\boldsymbol{b}$ based on data available through T. Because the panel predictive regression is high dimensional and the unpredictable component in monthly exchange rate changes is inherently large, OLS estimation of $\boldsymbol{b}$ is highly susceptible to overfitting.

The ENet (Zou and Hastie 2005), a refinement of the popular LASSO (Tibshirani 1996), is a machine learning technique based on penalized (or regularized) regression. It mitigates overfitting by including a penalty term in the objective function for estimating $\boldsymbol{b}$ in Equation (9):

$$\underset{\boldsymbol{b} \in \mathbb{R}^K}{\arg\min} \ \frac{1}{2NT} \|\Delta \boldsymbol{s} - \tilde{\boldsymbol{X}} \boldsymbol{b}\|_2^2 + \lambda P_\delta(\boldsymbol{b}), \tag{17}$$

where $\lambda \geq 0$ is a regularization hyperparameter governing the degree of shrinkage,

$$P_\delta(\boldsymbol{b}) = 0.5(1 - \delta)\|\boldsymbol{b}\|_2^2 + \delta \|\boldsymbol{b}\|_1, \tag{18}$$

δ is a blending hyperparameter for the ℓ_1 and ℓ_2 components of the penalty term, and

$$\|\boldsymbol{v}\|_1 = \sum_{j=1}^{J} |v_j|, \tag{19}$$

$$\|\boldsymbol{v}\|_2 = \left(\sum_{j=1}^{J} v_j^2 \right)^{0.5} \tag{20}$$

are the ℓ_1 and ℓ_2 norms, respectively, for a generic J-dimensional vector $\boldsymbol{v} = [\begin{array}{ccc} v_1 & \cdots & v_J \end{array}]'$. When $\lambda = 0$, there is no shrinkage, so that the LASSO and OLS objective functions coincide. The ENet penalty term in Equation (17) includes both ℓ_1 (LASSO) and ℓ_2 (ridge) components. The ℓ_1 component permits shrinkage to zero (for sufficiently large λ), so that the LASSO performs variable selection.[18] Following the recommendation of Hastie and Qian (2016), we set $\delta = 0.5$.

An important step in implementing the ENet (or LASSO) is tuning the hyperparameter λ. The most popular procedure is M-fold cross validation. However, the selection of the number of folds is somewhat arbitrary, and conventional M-fold cross validation does not necessarily respect the time-series dimension of the data. Alternatively, information criteria can provide an effective strategy for tuning λ (e.g., Zou, Hastie, and Tibshirani 2007; Wang, Li, and Leng 2009; Fan and Tang 2013; Flynn, Hurvich, and Simonoff 2013; Hui, Warton, and Foster 2015). The challenge is to find the proper balance between responding to relevant information in the predictors while inducing sufficient shrinkage to adequately guard against overfitting. We tune λ via the ERIC (Hui, Warton, and Foster 2015):

$$\text{ERIC} = NT \log\left(\frac{\text{SSR}_{\lambda,\delta}}{NT}\right) + \text{df}_{\lambda,\delta} \log\left(\frac{NT\hat{\sigma}^2_{\lambda,\delta}}{\lambda}\right), \tag{21}$$

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 $\hat{\sigma}^2_{\lambda,\delta} = \text{SSR}_{\lambda,\delta}/(NT)$. Equation (21) modifies the penalty term in the BIC to include λ. Considering a grid of values for λ, we select the value that minimizes Equation (21). The ERIC is a relatively stringent information criterion in the sense that it tends to induce more shrinkage than cross validation and other information

[18] We implement elastic net estimation of the linear model in Equation (8) via the `glmnet` package in `R`. The LASSO is adept at selecting relevant predictors in some settings (e.g., Zhang and Huang 2008; Bickel, Ritov, and Tsybakov 2009; Meinshausen and Yu 2009). However, it tends to arbitrarily select one predictor from a group of highly correlated predictors. The ENet alleviates this tendency by including both ℓ_1 and ℓ_2 components in the penalty term for the objective function.

criteria, such as the BIC. Given our high-dimensional and noisy data environment, the ERIC appears effective for tuning λ in light of the bias-variance trade-off.[19]

3.2. Deep Neural Network

To this point, we permit a degree of nonlinearity in the predictive regression via the interaction terms involving the country characteristics multiplied by the global variables; however, the specification in Equation (8) remains linear in the parameters. In this section, we allow for more complex predictive relationships by generalizing the linear specification in Equation (8):

$$\Delta s_{i,t+1} = f(\tilde{\boldsymbol{x}}_{i,t}) + \varepsilon_{i,t+1} \quad \text{for } i = 1, \ldots, N;\ t = 1, \ldots, T. \tag{22}$$

We then model $f(\tilde{\boldsymbol{x}}_{i,t})$ via a DNN, a popular machine learning device.

A feedforward neural network architecture is comprised of multiple layers. The first is the input layer, which is simply the set of predictors, and which we denote by $x_1, \ldots, x_{K_0}$. One or more hidden layers are next. Each hidden layer l contains K_l neurons, each of which takes predictive signals from the neurons in the previous hidden layer to produce another signal:

$$h_m^{(l)} = g\left(\omega_{m,0}^{(l)} + \sum_{j=1}^{K_{l-1}} \omega_{m,j}^{(l)} h_j^{(l-1)}\right) \quad \text{for } m = 1, \ldots, K_l;\ l = 1, \ldots, L, \tag{23}$$

where $h_m^{(l)}$ is the mth neuron in the lth hidden layer;[20] $\omega_{m,0}^{(l)}, \ldots, \omega_{m,K_{l-1}}^{(l)}$ are weights; and $g(\cdot)$ is an activation function. The output layer is a linear function that translates the signals

[19]Section A2 of the Internet Appendix discusses different validation methods for tuning λ in Equation (17), including conventional M-fold cross validation. As shown in Table A1 of the Internet Appendix, forecasts based on the ERIC generally perform better than those based on the different validation methods, as well as the BIC. A cross-validation method that respects the time-series dimension of the data and the BIC also perform well overall, while conventional five-fold cross validation performs poorly.

[20]For the first hidden layer, $h_j^{(0)} = x_j$ for $j = 1, \ldots, K_0$.

from the last hidden layer into a prediction:

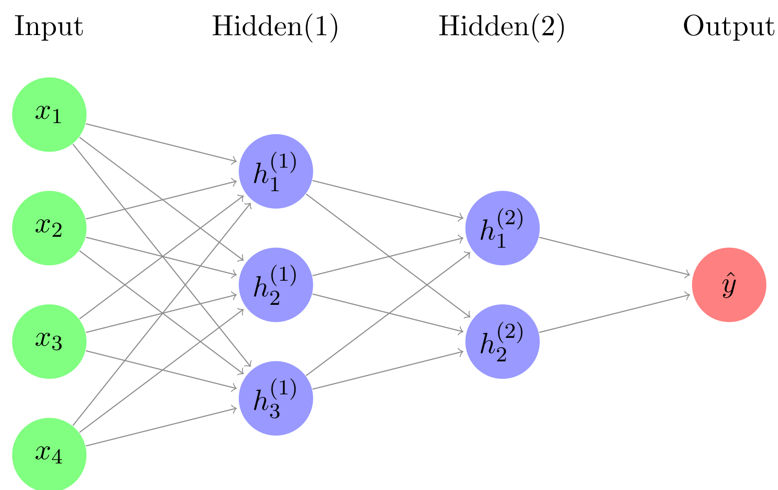
$$\widehat{y} = \omega_0^{(L+1)} + \sum_{j=1}^{K_L} \omega_j^{(L+1)} h_j^{(L)}, \tag{24}$$

where $\widehat{y}$ 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):[21]

$$g(x) = \begin{cases} x & \text{if } x > 0, \\ 0.01x & \text{if } x \leq 0. \end{cases} \tag{25}$$

Intuitively, Equation (25) activates a neuronal connection in response to the strength of the signal, thereby relaying the signal forward through the network.

The following diagram provides a schematic for a relatively simple feedforward neural network architecture with four predictors and two hidden layers, where the first and second hidden layers contain three and two neurons, respectively. The diagram shows that the four predictors in the input layer feed through to provide signals to each of the three neurons in the first hidden layer; the neurons in the first hidden layer subsequently feed through to provide signals to each of the two neurons in the second hidden layer. The neurons in the second hidden layer provide a final set of signals for the output layer.



[21] 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.

Theoretically, a neural network with a single hidden layer and sufficiently large number of neurons can approximate any smooth function under a reasonable set of assumptions (e.g., Cybenko 1989; Funahashi 1989; Hornik, Stinchcombe, and White 1989; Hornik 1991; Barron 1994). However, neural networks with three or more hidden layers (i.e., DNNs) and a more limited number of neurons in each layer often perform better than neural networks with one or two hidden layers (i.e., shallow neural networks) and a larger number of neurons in each layer (e.g., Goodfellow, Bengio, and Courville 2016; Rolnick and Tegmark 2018). We specify a DNN containing four hidden layers with 16, eight, four, and two neurons, respectively. Our specification of 16 neurons in the first hidden layer is a compromise between two popular rules of thumb, namely, half or the square root of the number of predictors.[22]

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 (17). Computationally efficient algorithms based on stochastic gradient descent (SGD) are available for estimating the DNN weights. We use the recently developed 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 different times with different seeds and take an average of the forecasts across the three fitted DNNs.[23]

Given our noisy data environment, we take additional steps to further guard against overfitting. We set the intercept terms for the weights ($\omega_{m,0}^{(l)}$ for $m = 1, \ldots, K_l$; $l = 1, \ldots, L$ and $\omega_0^{(L+1)}$) to zero.[24] These restrictions are analogous to the absence of an intercept term

[22]We also considered DNNs with three hidden layers and different numbers of nodes in the hidden layers and obtained similar results.

[23]We estimate the DNN using the `keras` package in `R` and set the batch size and number of epochs to 32 and 500, respectively. To tune the ℓ_1 and ℓ_2 regularization hyperparameters for the DNN, we consider a grid of values for each and use observations for the last 30% of months for the available data at the time of forecast formation as a validation sample. We select the vector of hyperparameter values that minimizes the objective function for the validation sample.

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

in the linear panel predictive regression in Equation (8). By setting the intercept terms for the weights to zero, as in Equation (8), we ensure 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 Equation (22), so that we impose the homogeneity restrictions that the weights are the same across countries. Imposing the slope homogeneity and zero intercept restrictions substantially reduces the number of weights that we need to estimate, thereby helping to guard against overfitting. In addition, we employ dropout (Hinton et al. 2012; Srivastava et al. 2014), which randomly drops a portion of the neurons in a hidden layer when training the model via the SGD algorithm and has been found to be useful for mitigating overfitting. We use a dropout rate of 0.5 for each of the first three hidden layers.

3.3. Ensemble

We also compute an ensemble forecast by taking the average of the linear ENet and DNN forecasts. Such a forecast is known as a voting ensemble in the machine learning literature. An ensemble forecast recognizes that we cannot know a priori the best individual model, and it allows us to take advantage of any complementarities between the models.

4. Out-of-Sample Performance

In this section, we analyze the accuracy of the out-of-sample exchange rate forecasts.

4.1. Beating the No-Change Benchmark

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 for the initial training sample, we estimate the linear panel predictive regression in Equation (8) and DNN using avaliable data from the beginning of the sample through

1994:12. We then use the fitted models and 1994:12 predictor values for each country to compute forecasts of exchange rate changes for each available country for 1995:01. Next, we re-estimate the models using available data through 1995:01; we then use the fitted models and 1995:01 predictor values for each country to generate forecasts for each available country 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. Each month we compute linear OLS and linear ENet forecasts based on OLS and ENet estimation, respectively, of the linear specification in Equation (8), as well as the DNN and ensemble forecasts. By retraining the forecasting models each month as new data become available, we update the fitted models in a timely manner. 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.[25]

For Germany, Italy, France, and the Netherlands, there are only 48 monthly forecasts (1995:01 to 1998:12) available for evaluation, due to the countries joining the Euro area in 1999:01. 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 each individual country, we report results for the entire collection of forecasts taken together ($4 \times 48 + 248 + 9 \times 309 = 3{,}221$ observations).

Figure 1 depicts the linear OLS forecasts, while Figures 2 and 3 show the linear ENet and DNN forecasts, respectively. The linear OLS forecasts are quite volatile, hinting at substantive overfitting for conventional OLS estimation in our high-dimensional and noisy data environment. The linear ENet forecasts are considerably less volatile, reflecting the strong shrinkage property of the ENet with ERIC hyperparameter tuning. The DNN forecasts are

[25] This includes when we standardize the predictors in Equation (5).

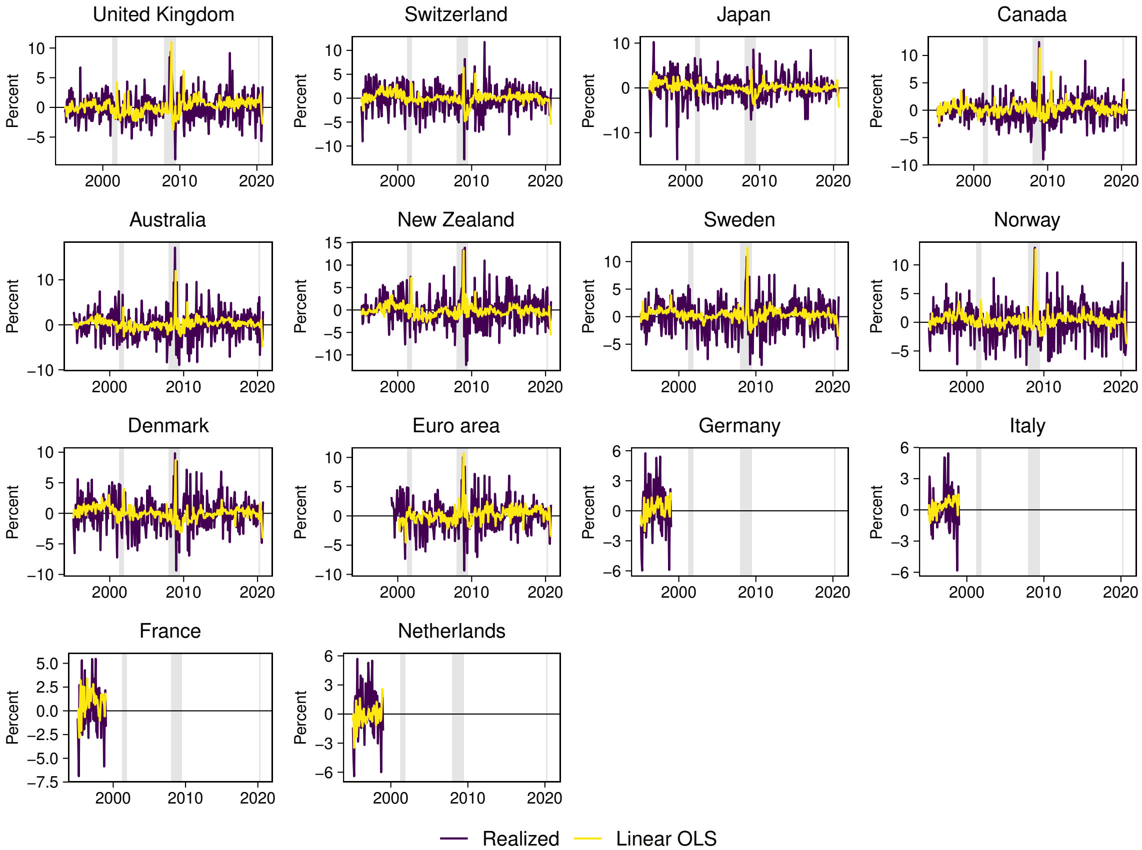


Figure 1. Linear OLS forecasts of log exchange rate changes. Each panel depicts the realized value and linear OLS forecast for the monthly log exchange rate change for the country in the panel heading. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

also much less volatile than the linear OLS forecasts, in line with the steps we take to guard against overfitting, as described in Section 3.2.

Next, we compare the linear OLS, linear ENet, DNN, and ensemble forecasts to the no-change benchmark, which is the most stringent benchmark for exchange rate prediction (Rossi 2013). We assess the accuracy of the exchange rate forecasts in terms of MSPE. We can conveniently compare the relative accuracy of a competing forecast to the no-change benchmark using the Campbell and Thompson (2008) R^2_{OS} statistic:

$$R^2_{\mathrm{OS}} = 1 - \frac{\sum_{t=T_1+1}^{T_2} \left(\hat{e}^{\mathrm{Compete}}_{i,t|t-1} \right)^2}{\sum_{t=T_1+1}^{T_2} \left(\hat{e}^{\mathrm{Bench}}_{i,t|t-1} \right)^2}, \tag{26}$$

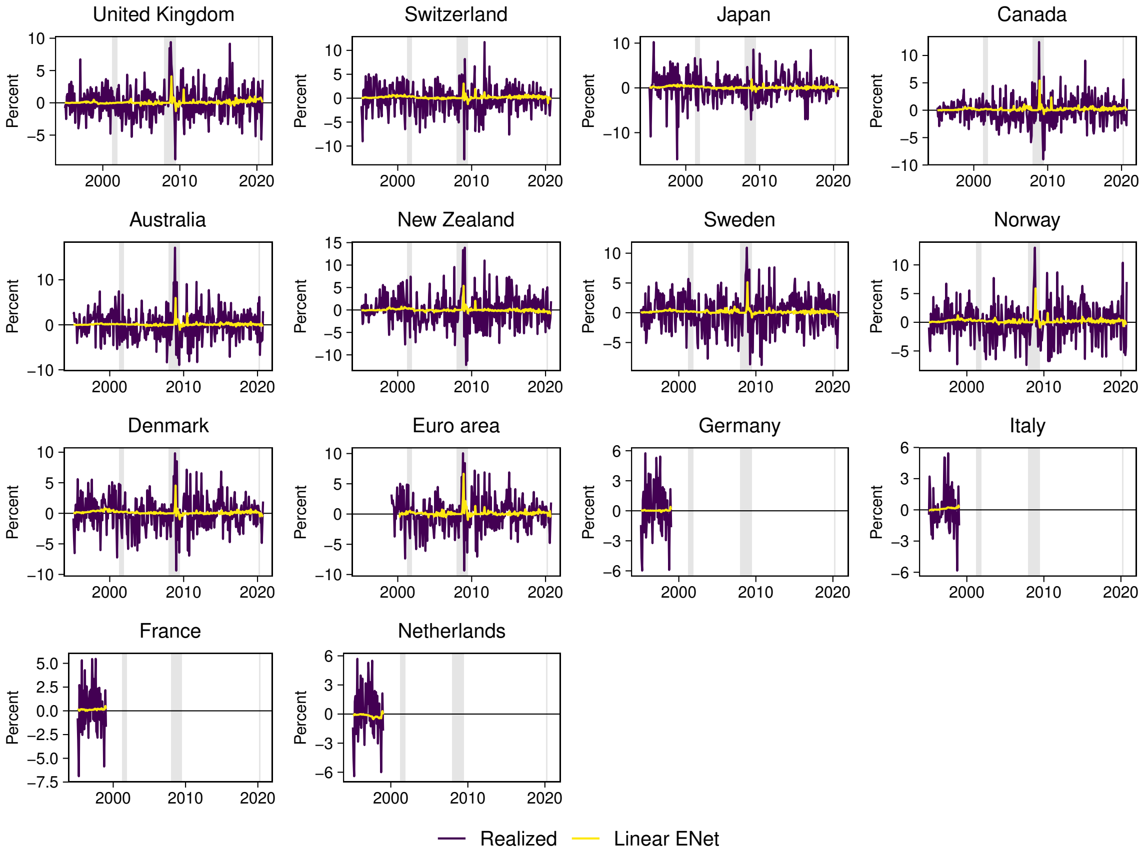


Figure 2. Linear ENet forecasts of log exchange rate changes. Each panel depicts the realized value and linear ENet forecast for the monthly log exchange rate change for the country in the panel heading. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

where

$$\hat{e}^{\text{Compete}}_{i,t|t-1} = \Delta s_{i,t} - \widehat{\Delta s}^{\text{Compete}}_{i,t|t-1}, \tag{27}$$

$$\hat{e}^{\text{Bench}}_{i,t|t-1} = \Delta s_{i,t} - \underbrace{\widehat{\Delta s}^{\text{Bench}}_{i,t|t-1}}_{=0}, \tag{28}$$

$\widehat{\Delta s}^{\text{Bench}}_{i,t|t-1} = 0$ is the no-change benchmark forecast, $\widehat{\Delta s}^{\text{Compete}}_{i,t|t-1}$ is a competing forecast (linear OLS, linear ENet, DNN, or ensemble), 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.[26] The R^2_{OS} statistic measures the proportional reduction in MSPE for a competing forecast vis-à-vis

[26]For our application, T_1 and T_2 correspond to 1994:12 and 2020:09, respectively.

the benchmark. Because the predictable component in monthly exchange rate changes is intrinsically small, the R^2_{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 6 below and as is the case for stock return predictability (Campbell and Thompson 2008). To get a sense of whether the competing forecast provides a statistically significant improvement in MSPE relative to the benchmark, we compute the Clark and West (2007) adjusted version of the Diebold and Mariano (1995) and West (1996) statistic.[27]

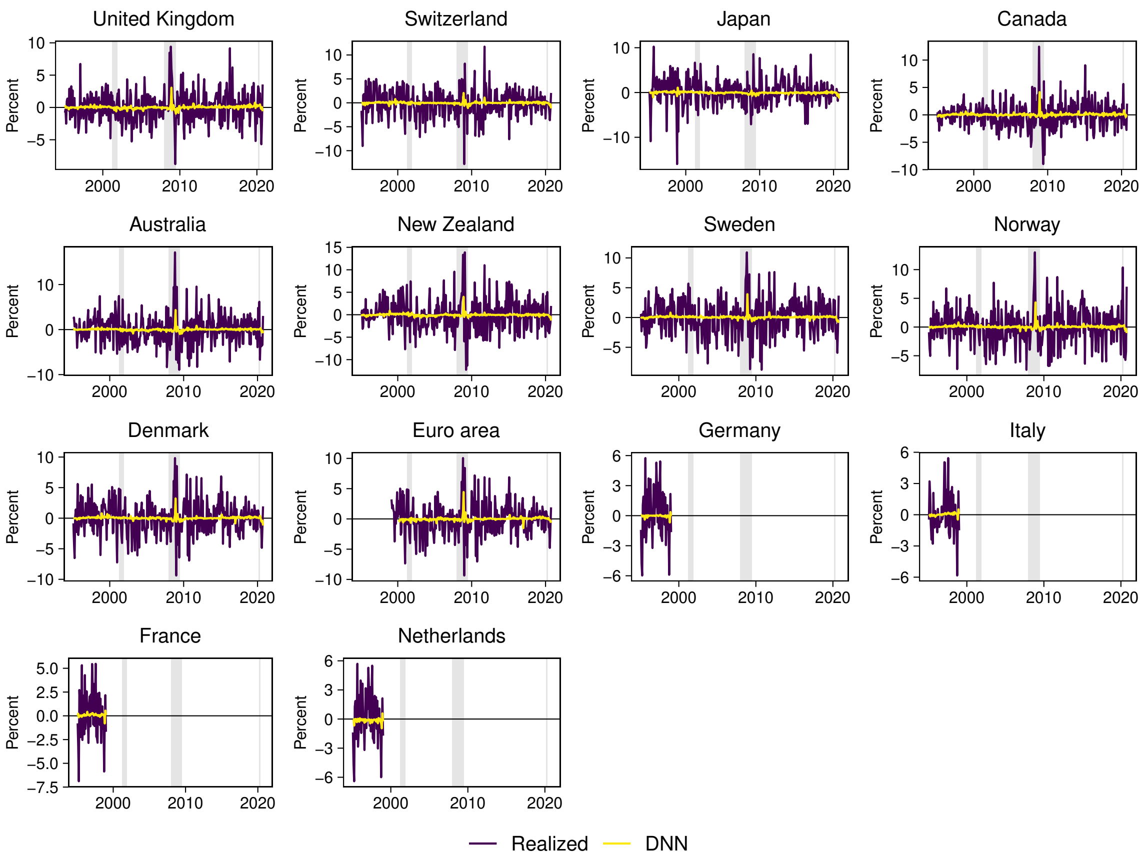


Figure 3. DNN forecasts of log exchange rate changes. Each panel depicts the realized value and DNN forecast for the monthly log exchange rate change for the country in the panel heading. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

[27]As shown by Clark and McCracken (2001) and McCracken (2007), the Diebold-Mariano-West statistic has a non-standard asymptotic distribution when comparing forecasts from nested models (as is the case for our application). In particular, the Diebold-Mariano-West statistic can be severely undersized when comparing nested forecasts, meaning that it can have little power to detect a significant improvement in forecast accuracy.

Table 2 reports R^2_{OS} statistics for the linear OLS, linear ENet, DNN, and ensemble forecasts. The R^2_{OS} statistics are all negative in the fourth column of Table 2, so that the linear OLS forecast always fails to outperform the no-change benchmark in terms of MSPE. The negative R^2_{OS} statistics are often sizable in magnitude for the individual countries, and the statistic is -9.21% for all of the countries taken together in the last row. The overfitting in the linear OLS forecast suggested by Figure 1 thus translates into poor out-of-sample performance in terms of forecast accuracy in Table 2.

(1) Country	(2) Out-of-sample period	(3) Obs.	(4) Linear OLS	(5) Linear ENet	(6) DNN	(7) Ensemble
United Kingdom	1995:01–2020:09	309	-12.08	2.77*	2.95**	3.12**
Switzerland	1995:01–2020:09	309	-13.40	1.03	2.65**	2.11**
Japan	1995:01–2020:09	309	-11.31	-0.25	1.30*	0.75*
Canada	1995:01–2020:09	309	-19.15	-0.40	0.60	0.42
Australia	1995:01–2020:09	309	-4.30	1.04*	1.60***	1.48**
New Zealand	1995:01–2020:09	309	-7.77	1.90*	2.16**	2.16**
Sweden	1995:01–2020:09	309	-3.87	2.88**	2.57**	2.86**
Norway	1995:01–2020:09	309	-3.26	2.45**	1.69**	2.27**
Denmark	1995:01–2020:09	309	-6.57	2.23***	1.76***	2.18***
Euro area	2000:02–2020:09	248	-13.24	1.12*	1.48**	1.75**
Germany	1995:01–1998:12	48	-11.68	0.45	-0.16	0.23
Italy	1995:01–1998:12	48	-13.37	0.24	2.36*	1.40
France	1995:01–1998:12	48	-37.73	0.62	1.47	1.18
Netherlands	1995:01–1998:12	48	-20.73	0.88	0.32	0.82
All	1995:01–2020:09	3,221	-9.21	1.48***	1.85***	1.88***

Table 2. R^2_{OS} statistics (%). The table reports Campbell and Thompson (2008) R^2_{OS} statistics in percent for forecasts of monthly log exchange rate changes. 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 R^2_{OS} 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 R^2_{OS} statistics, *, **, and *** indicate that the reduction in MSPE is significant at the 10%, 5%, and 1% level, respectively, according to the Clark and West (2007) test. The linear OLS, linear ENet, DNN, and ensemble forecasts incorporate the information in 70 predictors.

As shown in the fifth column of Table 2, the linear ENet forecast evinces markedly better out-of-sample performance. The R^2_{OS} statistics are positive for twelve of the 14 countries, and the reductions in MSPE vis-à-vis the no-change benchmark are significant at conventional levels for seven of the counties. Eight of the R^2_{OS} statistics for the individual countries are above 1% and reach as high as 2.88% (Sweden), so that they are sizable in the context of the literature surveyed by Rossi (2013). For the complete set of countries taken together in the last row, the R^2_{OS} statistic is 1.48% (significant at the 1% level). According to the sixth column, the DNN forecast also performs well. The R^2_{OS} statistics are positive for 13 of the 14 countries, including all ten of the G10 countries (for which at least 248 out-of-sample observations are available); the improvements in MSPE are significant for nine of the G10 countries (as well as Italy). The R^2_{OS} statistics for the individual countries are sizable, as eleven (five) are above 1% (2%). Taking all of the countries together, the R^2_{OS} statistic is 1.85% (significant at the 1% level) in the last row for the DNN forecast.

By combining the linear ENet and DNN forecasts, the ensemble approach in the last column of Table 2 provides the best overall performance. The R^2_{OS} statistics are positive for all of the countries, and the improvements in MSPE vis-à-vis the no-change benchmark are significant at conventional levels for nine of the G10 countries. The R^2_{OS} statistics are also typically sizable, with ten (six) above 1% (2%) and as large as 3.12% (United Kingdom). For the entire set of countries taken together, the R^2_{OS} statistic is 1.88% (significant at the 1% level), which is the highest value in the last row of Table 2. Note that the R^2_{OS} statistics for the ensemble forecast are often larger or nearly as large as the highest corresponding statistics in the fifth and sixth columns, pointing to meaningful complementarities in the linear and nonlinear forecasts.[28]

To examine the performance of the linear ENet, DNN, and ensemble forecasts over time, Figure 4 employs the graphical device of Goyal and Welch (2003, 2008). The figure portrays cumulative differences in squared prediction errors for the no-change benchmark vis-à-vis

[28]Forecasts generated by estimating models using country-specific data are less accurate than forecasts based on pooling, so that the parameter homogeneity restrictions improve out-of-sample performance.

the linear 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: we compare the height of the curve at the beginning and end of the segment corresponding to the subsample; if the curve is higher (lower) at the end of the segment, then the competing (benchmark) forecast has a lower MSPE for the subsample. A forecast that always outperforms the benchmark will thus have a curve with a uniformly positive slope. Of course, given that exchange rate changes have a large unpredictable component, this ideal is unattainable in practice. Realistically, we seek a forecast with a curve that is predominantly positively sloped and does not have extended segments with negative slopes.

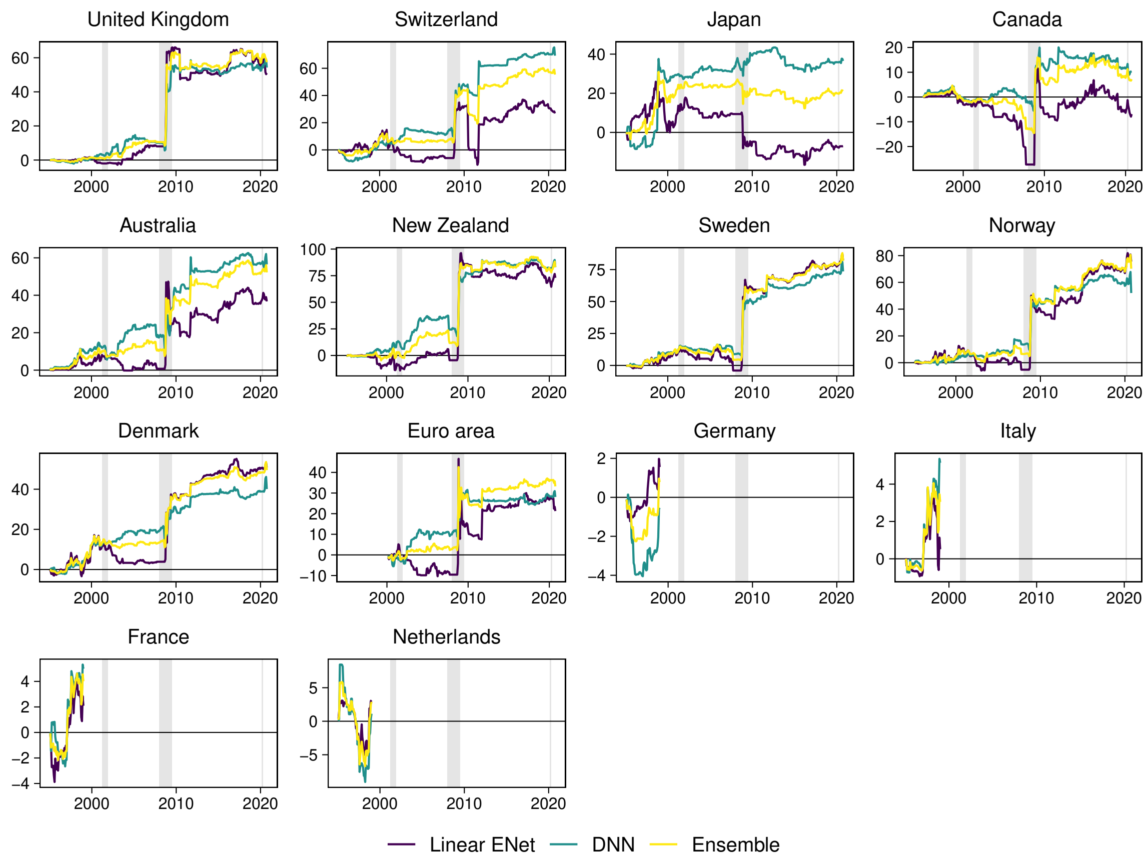


Figure 4. Cumulative differences in squared prediction errors. Each panel depicts the cumulative difference in squared prediction errors for the log exchange rate change for the country in the panel heading. The difference is computed for the no-change benchmark relative to the linear ENet, DNN, and ensemble forecasts (considered in turn). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

According to Figure 4, the curves are positively sloped for much of the time for the different countries—especially the G10 countries—so that the linear ENet, DNN, and ensemble forecasts outperform the no-change benchmark on a consistent basis over time. By blending the linear ENet and DNN forecasts, the ensemble approach generally provides the most consistent out-of-sample gains. The improvements in accuracy provided by the linear 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 that the information in the predictors becomes especially important during the crisis. Sizable gains are evident before the crisis in a number of cases, and the gains are quite consistent for many countries after 2008 through the end of the out-of-sample period. Compared to the linear ENet forecast, the DNN forecast substantially improves performance around the crisis for Switzerland, Japan, and Canada. Overall, Table 2 and Figure 4 indicate that the information in a rich set of predictors can be used to generate sizable and consistent improvements in out-of-sample forecasting accuracy, provided that we adequately guard against overfitting in our high-dimensional and noisy data setting.[29]

The no-change forecast is the standard benchmark in the out-of-sample exchange rate predictability literature, as it has proven difficult to beat consistently at short horizons. We check the robustness of our results to 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 INF and UN as predictors and is fitted via OLS. The second is a forecast computed from a linear panel predictive regression that uses INF and BILL as predictors. The latter benchmark is motivated by recent studies that include similar variables in predictive regressions when analyzing cur-

[29] An emerging literature (e.g., Engel and Wu 2019; Kremens and Martin 2019; Adrian and Xie 2020; Jiang, Krishnamurthy, and Lustig 2021; Lilley et al. forthcoming) finds evidence of exchange rate predictability around the Global Financial Crisis using variables related to a safe-haven role for the US dollar. Based on data availability, these studies use relatively short samples and/or analyze predictability at horizons of one quarter to multiple years. In contrast, we consider a lengthy out-of-sample period (beginning well before the Global Financial Crisis and extending to 2020:09) and focus on monthly predictability. As discussed in Section A3 of the Internet Appendix, our results are consistent with exchange rate predictability during the crisis corresponding to a safe-haven role for the US dollar.

rency risk premia (e.g., Chernov and Creal 2021; Chernov, Dahlquist, and Lochstoer 2021; Eichenbaum, Johannsen, and Rebelo 2021; Dahlquist and Pénasse forthcoming).[30] As shown in Tables A2 and A3 of the Internet Appendix, the results are very similar to those in Table 2, as our linear ENet, DNN, and ensemble forecasts outperform the alternative benchmarks for the vast majority of countries, often by a substantial margin.

5. Variable Importance

It is of economic interest to know the importance of the individual predictors in the models underlying the forecasts. To this end, we compute variable importance measures for the fitted models via the approach of Greenwell, Boehmke, and McCarthy (2018), which is based on PDPs (Friedman 2001).

A PDP measures the relationship between the expected value of the target 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

$$\begin{aligned} f_q(x_q) &= E_{\boldsymbol{x}_{C(q)}}\left[\hat{f}\big(x_q, \boldsymbol{x}_{C(q)}\big)\right] \\ &= \int_{\boldsymbol{x}_{C(q)}} \hat{f}\big(x_q, \boldsymbol{x}_{C(q)}\big) p_{C(q)}\big(\boldsymbol{x}_{C(q)}\big)\, d\boldsymbol{x}_{C(q)}, \end{aligned} \tag{29}$$

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

$$p_{C(q)}\big(\boldsymbol{x}_{C(q)}\big) = \int_{x_q} p(x_q, \boldsymbol{x}_{C(q)})\, dx_q \tag{30}$$

[30] We obtain similar results when we use the real exchange rate in place of INF.

is the joint marginal probability density for $\boldsymbol{x}_{C(q)}$. Equation (29) is typically estimated via Monte Carlo integration using the training data, which in our panel data setting is given by

$$\bar{f}_q(x_q) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{f}\big(x_q, \boldsymbol{x}_{i,t,C(q)}\big). \tag{31}$$

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 $\{x_{q,j}\}_{j=1}^{J}$. By construction, the PDP is a straight line for a linear model. For the ensemble of models, the PDP is the average of the measures in Equation (31) across the fitted models comprising the ensemble.

Greenwell, Boehmke, and McCarthy (2018) propose a PDP-based measure of predictor importance:

$$\mathcal{I}(x_q) = \left\{ \left(\frac{1}{J-1} \right) \sum_{j=1}^{J} \left[\bar{f}_q(x_{q,j}) - \frac{1}{J} \sum_{j=1}^{J} \bar{f}_q(x_{q,j}) \right]^2 \right\}^{0.5}. \tag{32}$$

Intuitively, Equation (32) 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 that the variable importance measure is zero. As the variability of the PDP increases, the importance measure in Equation (32) likewise increases.

Figure 5 depicts sequences of variable importance measures in Equation (32) for the 70 predictors for the ensemble forecast. The sequence corresponds to the variable importance measure for the fitted models used to generate the ensemble forecast for the date on the horizontal axis. We denote predictors that are the interaction between two variables with a period between the two variables (e.g., the interaction between INF and EPU is INF.EPU). Figure 5 allows us to see which of the 70 predictors are generally important in the fitted models underlying the ensemble forecast and how the importance of the predictors evolves over time.[31]

[31] The results are similar for variable importance meaures based on SHAP values (Lundberg and Lee 2017).

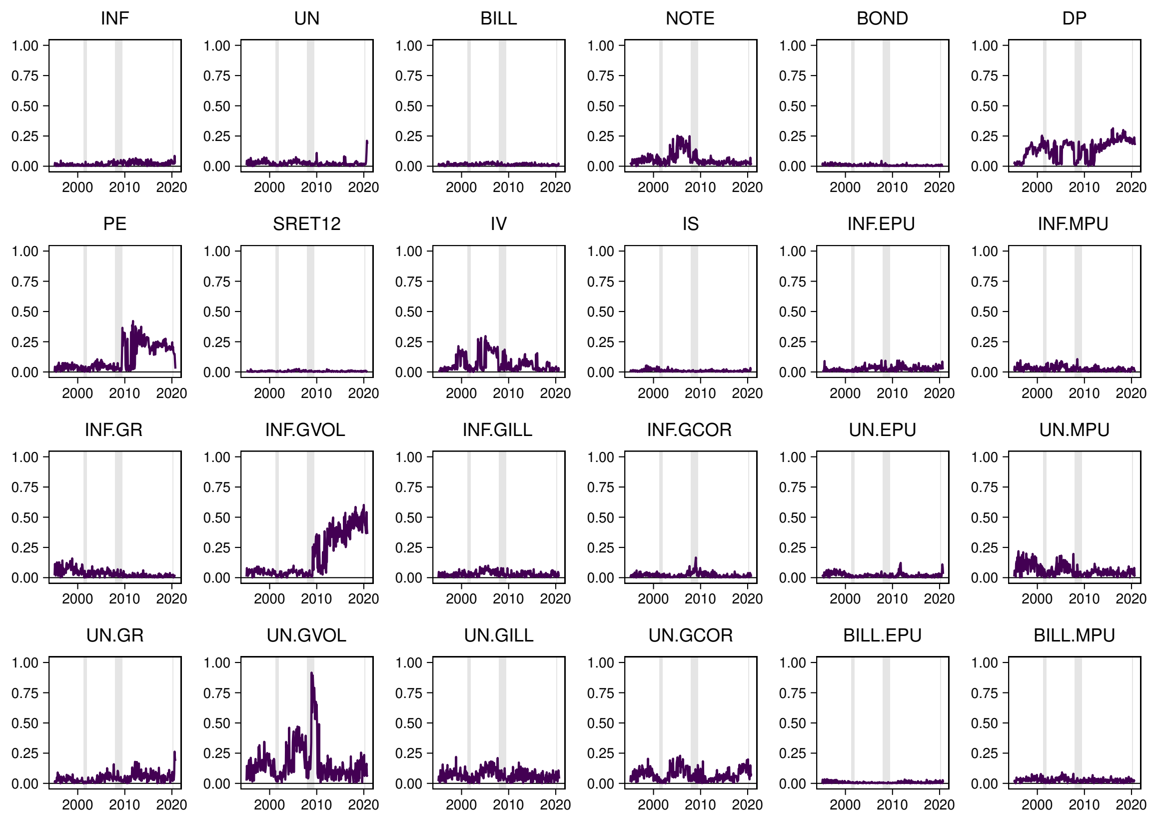


Figure 5. Variable importance. Each panel depicts the sequence of variable importance measures in Equation (32) for the predictor in the panel heading and ensemble forecast. The sequence corresponds to fitted models used to generate the ensemble forecast for the date on the horizontal axis.

With respect to the individual country characteristics, NOTE, DP, PE, and IV appear to be the most important, although their importance varies over time. NOTE and IV are relatively important from the mid to late 2000s, while PE becomes notably more important starting in the later stages of the Great Recession. DP appears more consistent in importance over time.

Turning to the interacted predictors, those involving GVOL are important in a number of cases. For example, INF.GVOL is limited in importance before the Global Financial Crisis, but it increases sharply in importance during the Great Recession and generally continues to increase in importance thereafter. UN.GVOL is important on a more consistent basis, but also exhibits a sharp spike during the Great Recession. BILL.GVOL displays a nearly

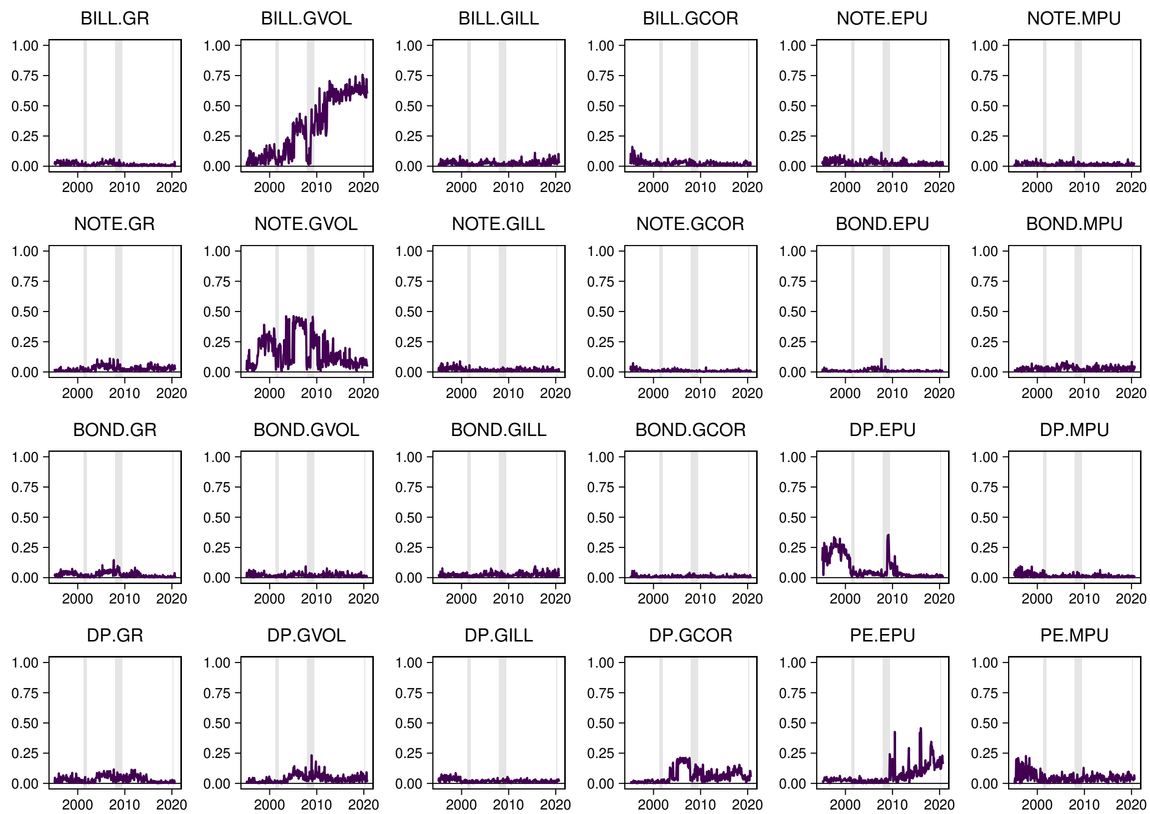


Figure 5 (continued).

monotonic increase in importance over the out-of-sample period, while NOTE.GVOL appears quite important from the late 1990s to around 2015, after which it declines in importance. IV.GVOL is also moderately important throughout much of out-of-sample period, including the run-up to the Great Recession.

In addition to GVOL, interactions of UN with some of the other global variables are relatively important, including UN.MPU, UN.GR, UN.GILL, and UN.GCOR. Other interacted predictors worth noting in terms of importance are DP.EPU, EP.GCOR, PE.EPU, PE.GR, PE.GILL, SRET12.MPU, and IV.GCOR. Among these, PE.GILL, SRET12.MPU, and IV.GCOR stand out by exhibiting marked increases in importance beginning in the mid to late 2000s. This is especially evident for PE.GILL, which displays a sharp increase near the end of the Great Recession and remains elevated in importance thereafter.

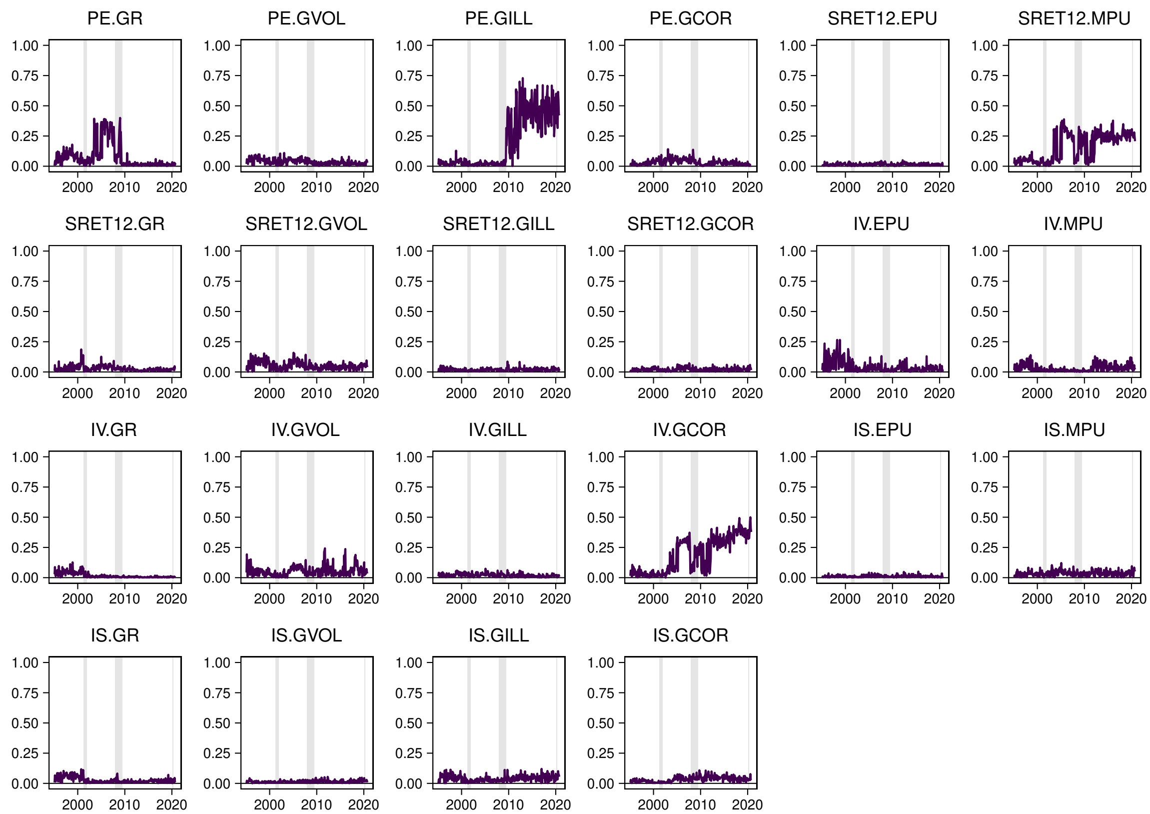


Figure 5 (continued).

To explore the nature of the relationships between the predictors and the exchange rate change forecasts, Figure 6 shows centered ICE (c-ICE) curves (Goldstein et al. 2015) and PDPs for the fitted models underlying the ensemble forecast for 2020:09 (which is the forecast based on the longest span of data). PDPs are useful for characterizing the average relationship between a predictor and the forecast for the target variable, which can help to identify nonlinearities in predictive relationships (on average). A drawback to PDPs is that the process of averaging in Equation (31) can mask nonlinearities involving interactions between predictors in the predictive relationships.

A given ICE curve provides additional information regarding potential nonlinearities by depicting the predictive relationship between the target variable and a grid of values for a particular predictor and a given vector of values for all of the other predictors. For the ensemble forecast, we take the average of the predicted values for the fitted linear ENet and

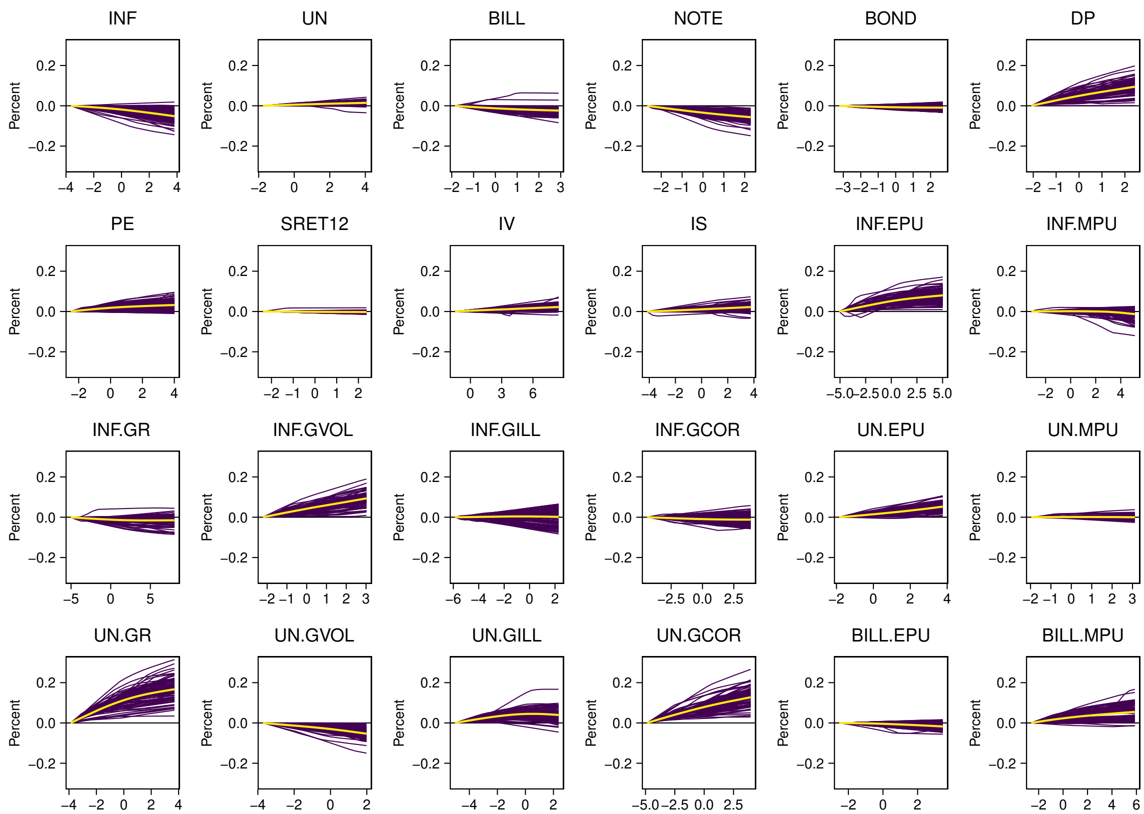


Figure 6. Centered ICE curves. Each panel depicts c-ICE curves for the predictor in the panel heading and ensemble forecast. The c-ICE curves are for fitted models used to generate the ensemble forecast for September 2020. The yellow line is the average of the curves and corresponds to the PDP.

DNN models. A collection of ICE curves for a particular predictor depicts multiple curves corresponding to different vectors of values for all of the other predictors. In this way, ICE curves provide additional information on potential nonlinearities that PDPs can obfuscate. To facilitate interpretation, we follow convention and center the ICE curves at the minimum value of the grid of predictor values.

For a given predictor, if we include c-ICE curves for all vectors of observations in the sample used to fit the linear ENet and DNN models underlying the ensemble forecast for 2020:09 (4,733 observations), then the sheer number of c-ICE curves in a given plot would essentially create a mass that is difficult to interpret. Instead, again following convention, we

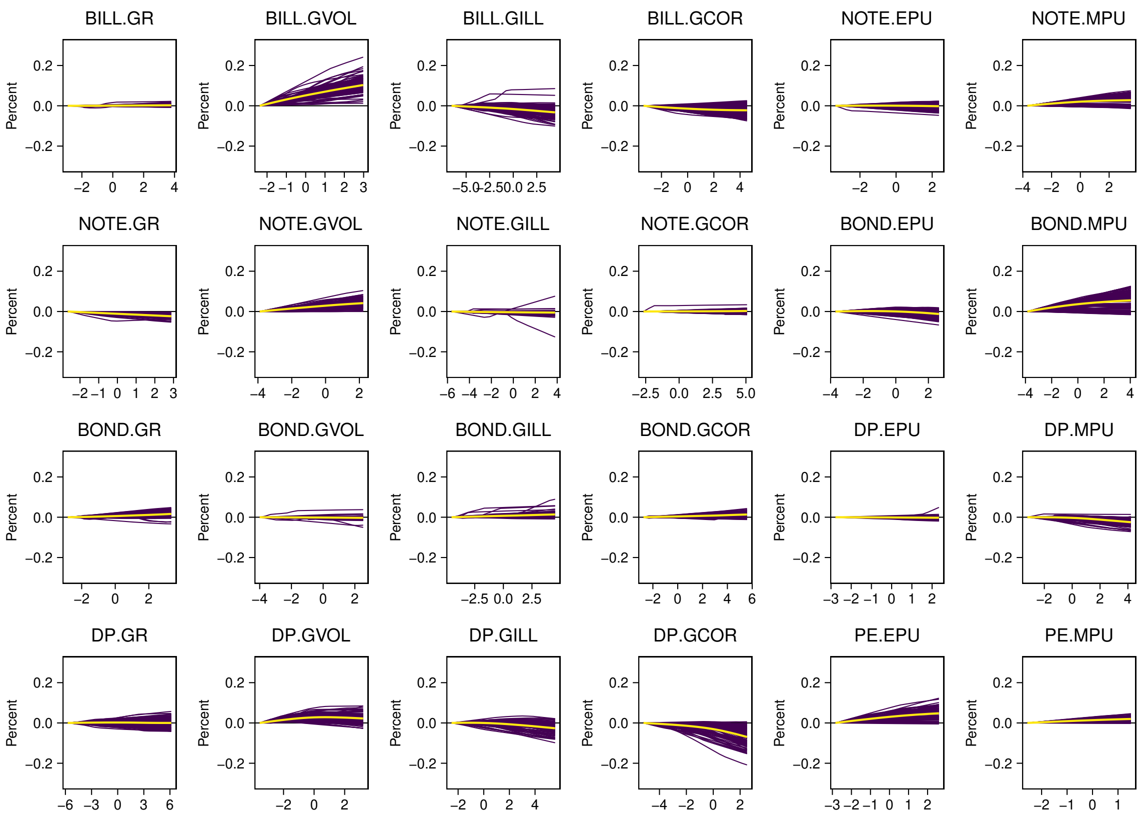


Figure 6 (continued).

randomly select a subset of observations (100) for computing c-ICE curves and corresponding PDPs in Figure 6.[32]

If there are no interaction effects in the marginal relationship between a given predictor and the target, then the c-ICE curves (as well as the PDP curve) will all coincide, and the marginal predictive relationship is additive with respect to the other predictors.[33] By construction, this is the case for the linear ENet, with the additional implication that each c-ICE curve will also be linear, so that any nonlinearities in Figure 6 are due to the influence of the fitted DNN on the ensemble forecast.

[32]The fitted linear ENet and DNN models underlying the c-ICE curves in Figure 6 are based on all of the available observations.

[33]The interactions we refer to here are those beyond the interactions inherent in the 60 interacted predictors among the 70 total predictors.

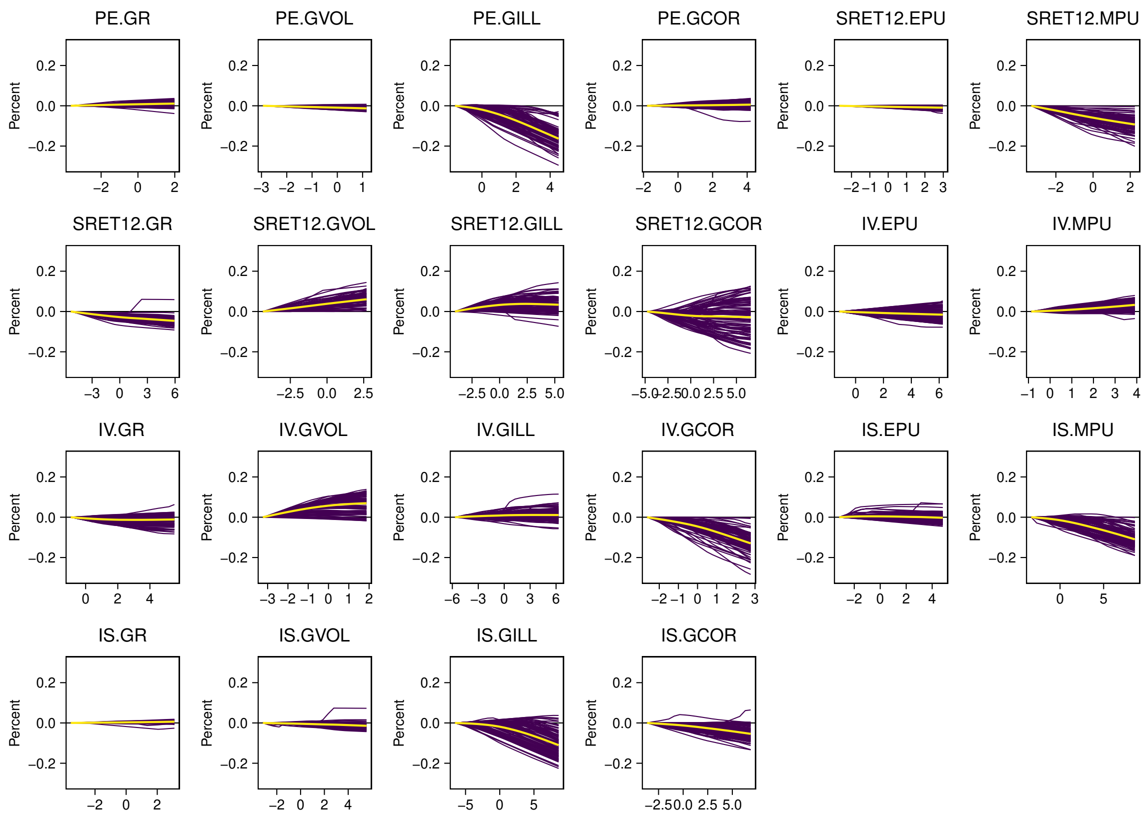


Figure 6 (continued).

According to Figure 6, the marginal relationships on average as depicted by the PDPs are often fairly close to linear for many predictors, although there are a number of cases where the slope of the PDP changes markedly (e.g., IS.GILL). As anticipated, the c-ICE curves display more extensive evidence of nonlinearities. They do so in two ways. First, many of the individual c-ICE curves show substantial changes in the slopes of the predictive relationships, so that the marginal relationship can be significantly nonlinear for certain vector of values for the other predictors, even when the average marginal relationship is close to linear. Second, numerous c-ICE curve plots spread out significantly as the value of the predictor changes, so that the marginal relationship changes substantially depending on the vector of values for the other predictors, indicating important interaction effects.

INF and BILL are perhaps the most popular exchange rate predictors in the literature. Neither variable on its own appears important over time in Figure 5. However, INF and BILL are often quite important when they interact with GVOL, as INF.GVOL and BILL.GVOL are frequently important predictors in Figure 5. Furthermore, INF.GVOL and BILL.GVOL generally become more important over time. The c-ICE curves and PDPs for INF.GVOL and BILL.GVOL in Figure 6 are positively sloped, so that the expected US dollar appreciation in response to an increase in INF or BILL becomes larger as GVOL increases. According to relative PPP, an increase in INF leads to an increase in expected US dollar appreciation, so that the prediction of PPP becomes more relevant as GVOL increases. Similarly, UIP predicts that an increase in BILL corresponds to an increase in expected US dollar appreciation, meaning that the relationship predicted by UIP also holds to a greater degree as GVOL increases (as it does, e.g., during the Global Financial Crisis). Overall, we find that popular fundamentals matter when they interact with global variables, especially global volatility.

6. Carry Trade Portfolios

In this section, we analyze the economic value of the linear ENet, DNN, and ensemble forecasts in the context of carry trade portfolios.

6.1. Portfolio Construction

Consider a US investor who goes long (short) the currency for country i (US dollar). Based on covered interest parity, we can approximate the log excess return for the investment as

$$rx_{i,t+1} \approx (r_{i,t} - r_{\mathrm{US},t}) - \Delta s_{i,t+1}, \tag{33}$$

where $rx_{i,t}$ is the month-t log currency excess return for country i, and $r_{i,t}$ ($r_{\mathrm{US},t}$) is the government bill yield for country i (the United States). Assume that the investor has mean-variance preferences and allocates monthly across available foreign currencies. At the end of month T, the investor's objective function is given by

$$\max_{\boldsymbol{w}_{T+1|T}} \boldsymbol{w}'_{T+1|T}\hat{\boldsymbol{\mu}}_{T+1|T} - 0.5\gamma\boldsymbol{w}'_{T+1|T}\hat{\boldsymbol{\Sigma}}_{T+1|T}\boldsymbol{w}_{T+1|T}, \tag{34}$$

where

$$\hat{\boldsymbol{\mu}}_{T+1|T} = [\ \widehat{rx}_{1,T+1|T} \quad \cdots \quad \widehat{rx}_{N,T+1|T}\ ]', \tag{35}$$

$$\widehat{rx}_{i,T+1|T} = (r_{i,T} - r_{\text{US},T}) - \widehat{\Delta s}_{i,T+1|T}, \tag{36}$$

$\widehat{\Delta s}_{i,T+1|T}$ ($\widehat{rx}_{i,T+1|T}$) is the investor's exchange rate change (currency excess return) forecast for country i, $\hat{\boldsymbol{\Sigma}}_{T+1|T}$ is the investor's estimate of the covariance matrix for the currency excess returns, $\boldsymbol{w}_{T+1|T} = [\ w_{1,T+1|T} \quad \cdots \quad w_{N,T+1|T}\ ]'$ is the N-vector of portfolio weights that will be in effect for month $T+1$, 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 $\hat{\boldsymbol{\Sigma}}_{T+1|T}$. We assume that $\gamma = 5$; the results are qualitatively similar for reasonable alternative γ values. To keep the portfolio weights in a plausible range, we also impose the restrictions that $-0.5 \leq w_{i,T+1|T} \leq 0.5$ for $i = 1, \ldots, N$.

We consider two cases, which differ with respect to the exchange rate forecast used to compute $\widehat{rx}_{i,T+1|T}$ in Equation (36). In the C-OPT portfolio case, the investor uses the linear ENet, DNN, or ensemble forecast for $\widehat{\Delta s}_{i,T+1|T}$ in Equation (36), so that they attempt to exploit exchange rate predictability by using the conditioning information in the predictors when allocating across currencies. In the U-OPT portfolio case, the investor uses the no-change benchmark forecast ($\widehat{\Delta s}_{i,T+1|T} = 0$) in Equation (36), so that they ignore exchange rate predictability and set the expected exchange rate change to zero; in this case, the investor simply uses the the bill yield differential (which is known at T) to predict the excess return.

We construct out-of-sample portfolio weights as follows. We first use data through 1994:12 to compute the EWMA estimate of the covariance matrix and linear ENet, DNN, or ensemble exchange rate forecast for 1995:01. We then solve Equation (34) using the EWMA covariance matrix estimate and linear ENet, DNN, or ensemble (no-change) exchange rate forecast to compute the C-OPT (U-OPT) portfolio weights for 1995:01. Next, we use data through 1995:01 to generate the EWMA covariance matrix estimate and linear ENet, DNN, or ensemble forecast for 1995:02; we then compute the C-OPT and U-OPT portfolio weights for 1995:02. We proceed in this fashion through the end of the out-of-sample period, so that we mimic the situation of an investor in real time. By comparing the performance of the C-OPT portfolio—which uses the information in the 70 predictors to forecast exchange rate changes—to that of the U-OPT portfolio—which assumes that exchange rate changes are not predictable—we can gauge the economic value of exchange rate predictability to an investor.[34]

Specifically, we compute the annualized average utility gain for the investor when they use the C-OPT in lieu of the U-OPT portfolio:

$$\text{Gain} = 12\big[\overline{rx}_{\text{C-OPT}} - 0.5\gamma\sigma^2_{\text{C-OPT}} - \big(\overline{rx}_{\text{U-OPT}} - 0.5\gamma\sigma^2_{\text{U-OPT}}\big)\big], \tag{37}$$

where $\overline{rx}_{\text{C-OPT}}$ ($\overline{rx}_{\text{U-OPT}}$) and $\sigma_{\text{C-OPT}}$ ($\sigma_{\text{U-OPT}}$) are the mean and volatility, respectively, for the monthly C-OPT (U-OPT) portfolio excess return over the out-of-sample period. Equation (37) is the annualized increase in certainly equivalent return, which can be interpreted as the annualized portfolio management fee that the investor would be willing to pay to have access to the C-OPT vis-à-vis the U-OPT portfolio.

6.2. Portfolio Performance

Table 3 reports the annualized average utility gain in Equation (37) when the investor uses the C-OPT instead of the U-OPT portfolio, as well as annualized means, volatilities, and Sharpe ratios for the C-OPT and U-OPT portfolio excess returns. In addition to the full

[34]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.

1995:01 to 2020:09 out-of-sample period, the table reports results for the 1995:01 to 2008:08 and 2008:09 to 2020:09 subsamples. The start of the second subsample coincides with the bankruptcy of Lehman Brothers on September 15, 2008 at the height of the Global Financial Crisis.

(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				
C-OPT, linear ENet	3.30%	11.52%	12.74%	0.90***
C-OPT, DNN	3.34%	10.75%	11.39%	0.94***
C-OPT, ensemble	3.40%	11.02%	11.76%	0.94***
U-OPT	–	6.73%	10.14%	0.66***
Panel B: 1995:01 to 2008:08 out-of-sample period				
C-OPT, linear ENet	0.58%	13.01%	12.28%	1.06***
C-OPT, DNN	2.10%	13.82%	11.06%	1.25***
C-OPT, ensemble	0.93%	12.63%	11.02%	1.15***
U-OPT	–	11.47%	10.61%	1.08***
Panel C: 2008:09 to 2020:09 out-of-sample period				
C-OPT, linear ENet	6.27%	9.83%	13.26%	0.74**
C-OPT, DNN	4.68%	7.27%	11.71%	0.62**
C-OPT, ensemble	6.09%	9.20%	12.55%	0.73**
U-OPT	–	1.37%	9.39%	0.15

Table 3. Portfolio performance. The table reports portfolio performance metrics for a mean-variance US investor with a relative risk aversion coefficient of five who allocates monthly across available foreign currencies. For the C-OPT (U-OPT) portfolio, the investor uses the linear ENet, DNN, or ensemble (no-change) exchange rate forecast when predicting the currency excess return. The linear ENet, DNN and ensemble forecasts incorporate the information in 70 predictors. The second column reports the annualized increase in certainty equivalent return when the investor uses the C-OPT instead of the U-OPT portfolio. Statistical significance for the Sharpe ratio is based on the Bao (2009) procedure; *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Panel A of Table 3 reports performance measures for the full out-of-sample period. The C-OPT portfolios based on the linear ENet, DNN, and ensemble forecasts all perform quite well, delivering annualized Sharpe ratios between 0.90 (DNN) and 0.94 (DNN and ensemble), all of which are significant at the 1% level. The U-OPT portfolio, which ignores exchange rate predictability, also performs well, with an annualized Sharpe ratio of 0.66 (significant at the 1% level). Nevertheless, as evinced by the average utility gains in the second column, the C-OPT portfolios provide substantive economic value to the investor vis-à-vis the U-OPT portfolio: the investor realizes sizable annualized increases in certainty equivalent return, ranging from 330 (linear ENet) to 340 (ensemble) basis points.[35]

The results for the U-OPT portfolio in Panel A of Table 3 mask stark differences in the portfolio's performance over time. For the first subsample in Panel B, the average return and Sharpe ratio are considerably higher than their values for the full sample, with an annualized average return of 11.47% and Sharpe ratio of 1.08 (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).

In terms of the Sharpe ratio, the C-OPT portfolios perform fairly similarly to the U-OPT portfolio for the first subsample in Panel B of Table 3. The annualized Sharpe ratios for the C-OPT portfolios range from 1.06 (linear ENet) to 1.25 (DNN), all of which are significant at the 1% level. The C-OPT portfolios still provide additional economic value to the investor, with annualized average utility gains of 58 (linear ENet) to 210 (DNN) basis points. Differences in performance between the U-OPT and C-OPT portfolios become much more marked for the second subsample in Panel C. The Sharpe ratios in the last column are approximately four to five times higher for the C-OPT portfolios compared to that for the U-OPT portfolio, with values ranging from 0.62 (DNN) to 0.74 (linear ENet), all of which are significant at the 5% level. The annualized average utility gains accruing to the C-OPT

[35]For the nine countries for which we have forecasts for the full 1995:01 to 2020:09 out-of-sample period, we also compute performance measures for U-OPT portfolios and C-OPT portfolios based on the ensemble forecast for non-US domestic investors. As shown in Table A4 of the Internet Appendix, with the exception of Switzerland, the C-OPT portfolios deliver substantial economic value to non-US investors, with annualized average utility gains typically well above 200 basis points.

portfolios are especially sizable for the second subsample, ranging from 468 (DNN) to 627 (linear ENet) basis points.[36]

We also construct a conventional carry trade portfolio that sorts currencies into quintiles according to interest rate differentials and then takes equally weighted long (short) positions in the currencies in the fifth (first) quintile. This is tantamount to the carry trade risk factor in Lustig, Roussanov, and Verdelhan (2011). The conventional carry portfolio generally does not perform as well as the U-OPT portfolio, with 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.[37]

Next, we check the extent to which transaction costs affect the results. We compute portfolio excess returns using bid and ask quotes from Datastream for the relevant forward and spot rates (e.g., Lustig, Roussanov, and Verdelhan 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 et al. 2012b; Neely and Weller 2013). Specifically, investors trade 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.[38] 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 4 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 monontonically as we move from Panel A (25% of the bid-ask spread) to Panel D (full bid-ask

[36]Providing further evidence of overfitting, when the investor uses the linear OLS exchange rate forecast in Equation (36), the annualized average utility gains are −1.87%, −2.48%, and −1.29% for the 1995:01 to 2020:09, 1995:01 to 2008:08, and 2008:09 to 2020:09 out-of-sample periods, respectively.

[37]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.

[38]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.

(1) Portfolio	(2) Annualized average utility gain	(3) Annualized mean	(4) Annualized volatility	(5) Annualized Sharpe ratio
Panel A: 25% of bid-ask spread				
C-OPT, linear ENet	3.19%	10.40%	12.98%	0.80***
C-OPT, DNN	3.17%	9.51%	11.56%	0.82***
C-OPT, ensemble	3.22%	9.80%	11.98%	0.82***
U-OPT	–	5.59%	10.18%	0.55***
Panel B: 50% of bid-ask spread				
C-OPT, linear ENet	3.11%	9.34%	12.98%	0.72***
C-OPT, DNN	3.11%	8.47%	11.55%	0.73***
C-OPT, ensemble	3.14%	8.74%	11.97%	0.73***
U-OPT	–	4.61%	10.17%	0.45**
Panel C: 75% of bid-ask spread				
C-OPT, linear ENet	3.03%	8.28%	12.97%	0.64***
C-OPT, DNN	3.06%	7.43%	11.53%	0.64***
C-OPT, ensemble	3.06%	7.68%	11.96%	0.64***
U-OPT	–	3.63%	10.16%	0.36*
Panel D: Full bid-ask spread				
C-OPT, linear ENet	2.95%	7.21%	12.96%	0.56***
C-OPT, DNN	3.00%	6.39%	11.52%	0.55**
C-OPT, ensemble	2.98%	6.61%	11.95%	0.55***
U-OPT	–	2.64%	10.15%	0.26

Table 4. 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 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 C-OPT (U-OPT) portfolio, the investor uses the linear ENet, DNN, or ensemble (no-change) exchange rate forecast when predicting the currency excess return. The linear ENet, DNN, and ensemble exchange rate forecasts incorporate the information in 70 predictors. The second column reports the annualized increase in certainty equivalent return when the investor uses the C-OPT instead of the U-OPT portfolio. Statistical significance for the Sharpe ratio is based on the Bao (2009) procedure; *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

spread). Nevertheless, they remain quite sizable for the C-OPT portfolios: they are greater than or equal to 0.80 (0.72) for 25% (50%) of the bid-ask spread in Panel A (B); for 75% of the bid-ask spread in Panel C, they still are all 0.64.[39] The C-OPT portfolios also continue to provide substantive economic value to the investor vis-à-vis the U-OPT portfolio in terms of the annualized average utility gains in the second column. Even for 75% of the bid-ask spread, the gains are all above 300 basis points. In sum, the performance of the C-OPT portfolios—and thus the economic value of out-of-sample exchange rate predictability—is robust to reasonable assumptions regarding transaction costs.

To provide additional perspective on relative performance, Figure 7 depicts log cumulative excess returns for the C-OPT portfolio based on the ensemble forecast, as well as the U-OPT and conventional carry trade portfolios.[40] The C-OPT and U-OPT portfolios perform somewhat similarly through the summer of 2008, although the C-OPT portfolio fares substantially better in the late 1990s and early 2000s in the wake of the Asian financial and Long-Term Capital Management crises. While both 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 U-OPT portfolio suffers more sizable losses later in 2008 in Figure 7, and its cumulative return essentially "flatlines" thereafter. The conventional carry portfolio, which is not based on an optimization framework, suffers an even larger drop in late 2008 compared to the U-OPT portfolio and also flatlines subsequently. In sharp contrast, the C-OPT portfolio makes a strong recovery in late 2008 and continues to produce gains on a consistent basis thereafter.[41]

The relatively strong performance of the C-OPT portfolio in late 2008 in Figure 7 aligns with the large out-of-sample gains accruing to the ensemble forecasts vis-à-vis the no-change benchmark during that time in Figure 4. The U-OPT portfolio simply uses the bill yield

[39] All of the Sharpe ratios for the C-OPT portfolios are significant at the 1% level in Panels A through C of Table 4.

[40] The results are qualitatively similar for the C-OPT portfolios based on the linear ENet and DNN forecasts.

[41] It is interesting to note that the C-OPT portfolio also performs considerably better than the U-OPT and conventional carry trade portfolios in Figure 7 near the advent of the COVID-19 crisis in early 2020.

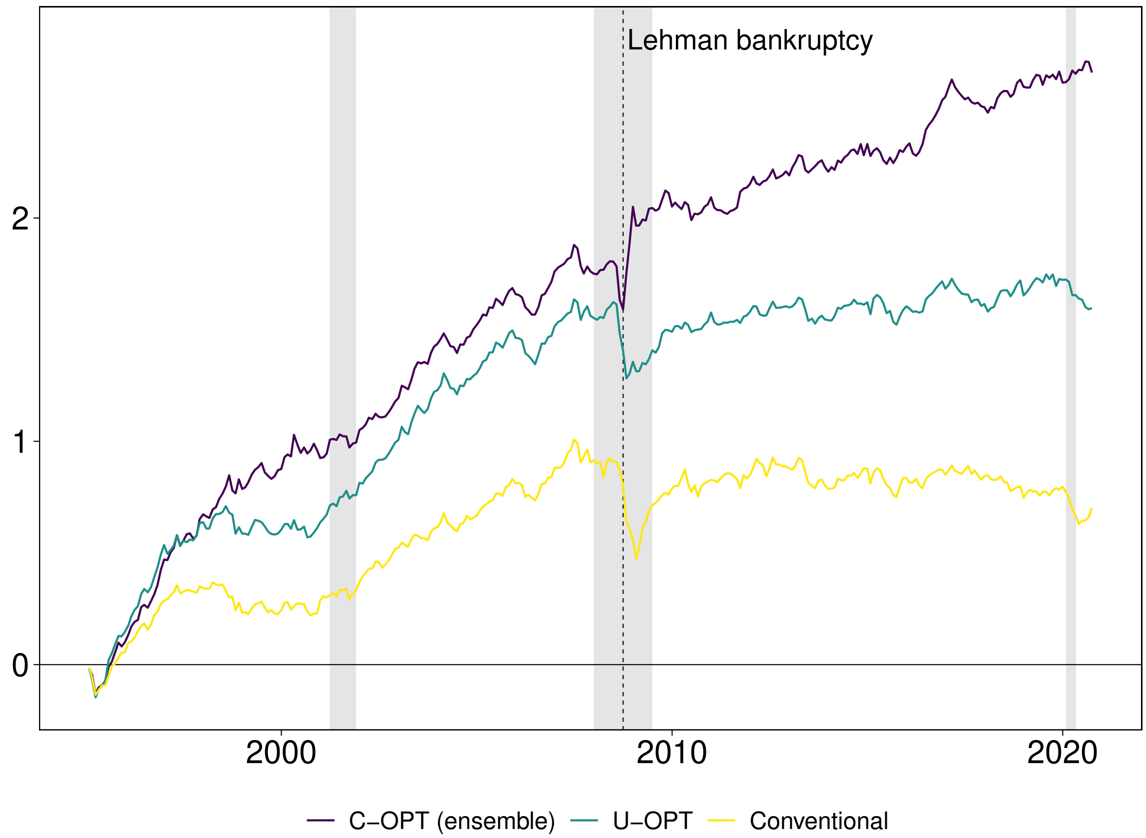


Figure 7. Log cumulative portfolio excess returns. The figure depicts log cumulative excess returns for the C-OPT portfolio based on the ensemble forecast, U-OPT portfolio, and a conventional carry trade portfolio. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

differential to forecast the currency excess return in Equation (36)—in line with the no-change benchmark forecast—while the C-OPT portfolio incorporates the information in the predictors via the ensemble forecast. The accuracy gains generated by the ensemble forecasts relative to the no-change benchmark in late 2008 and beyond in Figure 4 translate into economic gains in the form of improved portfolio performance in Figure 7.

Further evidence on the links between exchange rate predictability and the carry portfolios is furnished by Figure 8, which portrays the currency weights for the C-OPT portfolio based on the ensemble forecast and the U-OPT portfolio. As the figure illustrates, by incorporating information from the predictors, the ensemble forecast often leads to substantially different allocations. The differences in allocations produced by the ensemble forecast vis-à-vis the no-change benchmark in Equation (36) deliver improved carry trade portfolio performance in Table 3 and Figure 7.

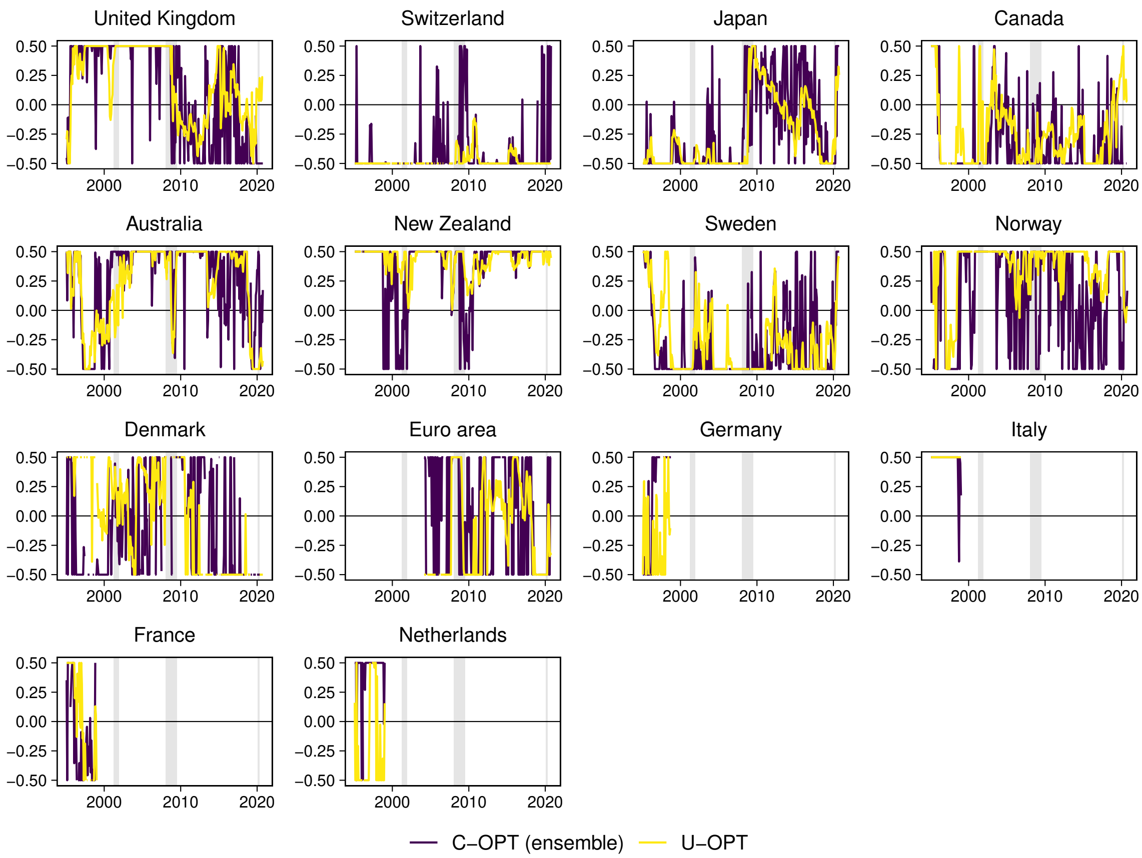


Figure 8. Portfolio weights. The figure depicts allocations to foreign currencies for the C-OPT and U-OPT portfolios. The C-OPT portfolio is based on the ensemble forecast. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

6.3. A Closer Look at Late 2008

Figure 7 indicates that late 2008 is a perilous time for the U-OPT portfolio (as well as the conventional carry portfolio). Figures 9 and 10 provide additional insight into the sources of the poor performance of the U-OPT portfolio in late 2008, as well as how the C-OPT portfolio (based on the ensemble forecast) improves performance. Figure 9 shows the currency excess return forecasts that serve as inputs in the portfolio optimization problem in Equation (34), along with the realized excess returns, for September through December of 2008. Because the U-OPT portfolio uses the no-change benchmark exchange rate forecast, the benchmark currency excess return forecast in Equation (36) is simply the bill yield differential; the C-OPT portfolio augments the bill yield differential with the ensemble forecast of the exchange rate change. Figure 10 displays the currency weights for the portfolios.

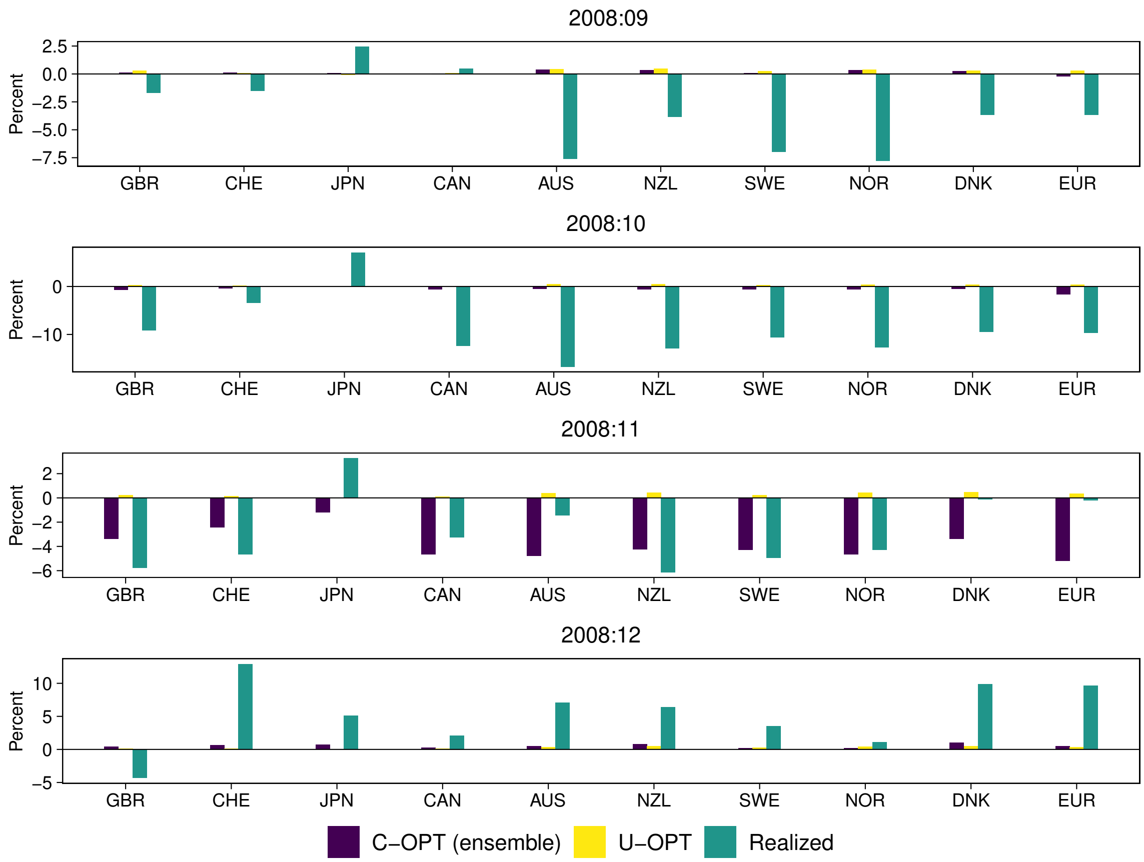


Figure 9. Currency excess returns forecasts for late 2008. The figure depicts currency excess return forecasts that serve as inputs for the C-OPT and U-OPT portfolios for the final four months of 2008. The C-OPT portfolio is based on the ensemble forecast. The figure also depicts realized values for the currency excess returns.

For September of 2008, the ensemble and benchmark currency excess return forecasts in Figure 9 lead to allocations of the same signs (and typically similar magnitudes) in Figure 10 for Switzerland, Canada, Australia, New Zealand, Sweden, Norway, and Denmark. The ensemble currency excess return forecast generates notable differences in allocations for the other countries: for the United Kingdom and Euro area, the C-OPT (U-OPT) portfolio takes short (long) positions; for Japan, the C-OPT (U-OPT) portfolio takes a long (short) position. As shown in Figure 9, 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 U-OPT portfolio is -7.57% in September of 2008, while the loss is smaller (-4.24%) for the C-OPT portfolio. The differences in allocations signaled by the ensemble forecast—especially for the United Kingdom, Japan, and Euro area—help to limit portfolio losses in the month of the Lehman bankruptcy.

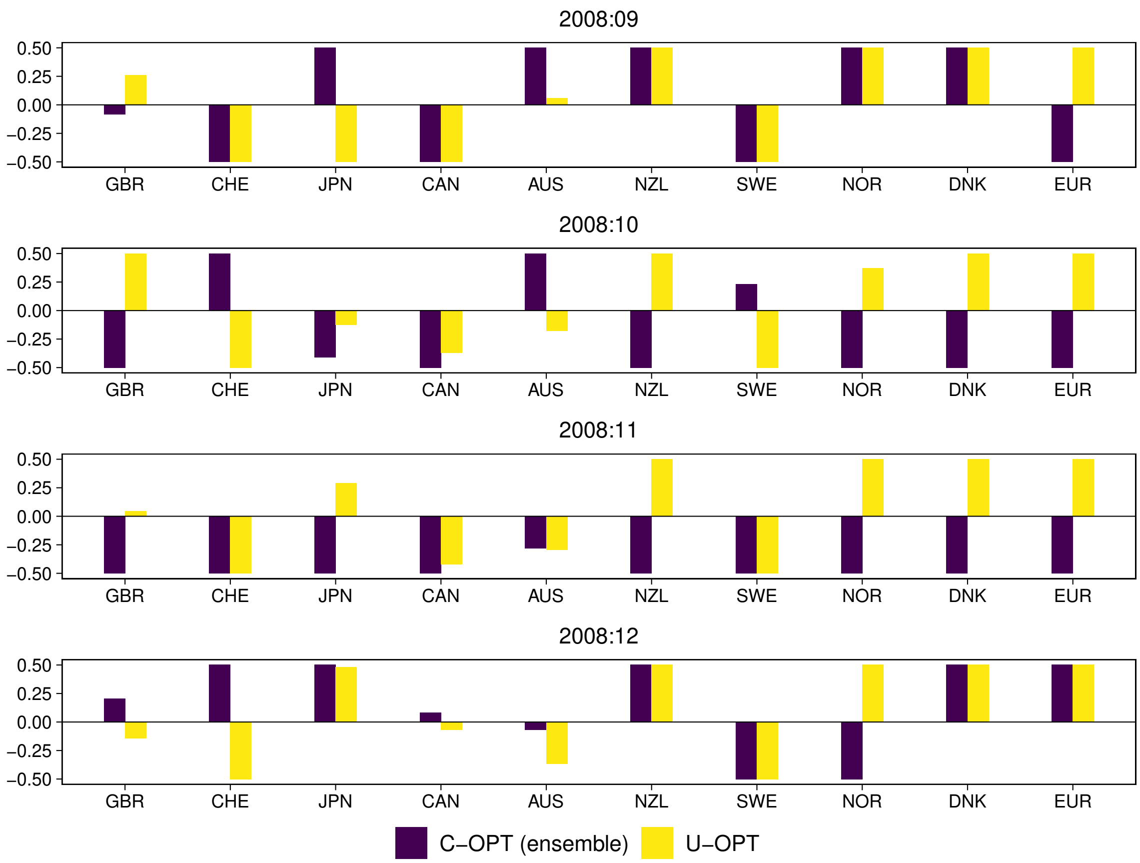


Figure 10. Portfolio weights for late 2008. The figure depicts allocations to foreign currencies for the C-OPT and U-OPT portfolios for the final four months of 2008. The C-OPT portfolio is based on the ensemble forecast.

The currency excess return forecasts diverge more sharply during October of 2008 in Figure 9: with the exception of Japan, the benchmark currency excess return forecasts are all positive, while the ensemble forecasts are negative for all ten of the countries. The differences in currency excess return forecasts give rise to a number of markedly different allocations during October of 2008 in Figure 10; most notably, the C-OPT (U-OPT) portfolio exhibits sizable short (long) positions for the United Kingdom, New Zealand, Norway, Denmark, and Euro area. With the exception of Japan, all of the realized currency excess returns are negative for October of 2008 in Figure 9, and the negative returns are larger in magnitude than those for September of 2008.[42] The different allocations prompted by the ensemble forecast vis-à-vis the no-change benchmark enable the C-OPT portfolio to substantively outperform the U-OPT portfolio during October of 2008: the latter suffers a large loss of −11.60%, while the former enjoys a substantial gain of 17.74%. 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 9 and 10. Because the negative realized currency excess returns are typically smaller in magnitude, the U-OPT realizes an excess return of 1.91% during the month; however, the information in the ensemble forecast leads to a much larger excess return of 13.46% for the C-OPT portfolio.

The environment appears to normalize to an extent in December of 2008, in the sense that the discrepancies between the ensemble and benchmark forecasts in Figure 9, as well as those between the C-OPT and U-OPT portfolio weights in Figure 10, 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 of the realized currency excess returns are positive in December of 2008. The U-OPT portfolio experiences an excess return of 5.64% for the month, while it is more than three times as large (18.48%) for the C-OPT portfolio.[43]

Figures 9 and 10 help to explain how exchange rate predictability—as captured by the ensemble forecast—is especially valuable to an investor during the worst part of the Global

[42]Note the difference in scales for the vertical axes across the top two panels of Figure 9.

[43]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 7.

Financial Crisis in late 2008. By anticipating depreciations in many foreign currencies, the ensemble forecast leads to sizable negative positions in many currencies, enabling the C-OPT portfolio to avoid the large losses suffered by a U-OPT carry strategy and even realize large gains. In essence, the C-OPT portfolio shorts the traditional carry strategy to a significant extent during the tumult of the Global Financial Crisis.

6.4. Alphas

Finally, we examine whether the C-OPT and U-OPT portfolios generate alpha in the context of the Lustig, Roussanov, and Verdelhan (2011) currency factor model. The model includes dollar and carry trade risk factors, denoted by $\mathrm{MKT_{FX}}$ and $\mathrm{HML_{FX}}$, 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 previously. Table 5 reports factor model estimation results for the C-OPT and U-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 5, the C-OPT portfolios generate large annualized alphas of 9.92%, 9.53%, and 9.61% for the linear ENet, DNN, and ensemble forecasts, respectively, all of which are significant at the 1% level and substantially larger than that for the U-OPT portfolio (4.27%). The U-OPT portfolio evinces considerable exposure to the carry factor (0.77, significant at the 1% level) for the full out-of-sample period. The C-OPT portfolios also exhibit sizable exposures to the carry factor (0.47, 0.38, and 0.42 for the linear ENet, DNN, and ensemble forecasts, respectively, all of which are significant at the 5% or 1% level), but they are about half as large as that for the U-OPT portfolio.

Panels B and C of Table 5 reveal important differences in performance for the U-OPT portfolio across the two subsamples. For the 1995:01 to 2008:08 subsample in Panel B, the U-OPT portfolio exhibits near unitary exposure to the carry trade factor (0.95, significant

(1)	(2)	(3)	(4)	(6)
Portfolio	Annualized alpha	$\mathrm{MKT}_{\mathrm{FX}}$	$\mathrm{HML}_{\mathrm{FX}}$	Adjusted R^2
Panel A: 1995:01 to 2020:09 out-of-sample period				
C-OPT, linear ENet	9.92%***	−0.56***	0.47**	16.97%
C-OPT, DNN	9.53%***	−0.05	0.38***	9.18%
C-OPT, ensemble	9.61%***	−0.29**	0.42**	11.58%
U-OPT	4.27%***	−0.03	0.77***	52.89%
Panel B: 1995:01 to 2008:08 out-of-sample period				
C-OPT, linear ENet	7.63%***	−0.47**	0.87***	46.07%
C-OPT, DNN	9.12%***	0.08	0.69***	29.12%
C-OPT, ensemble	7.48%***	−0.22	0.79***	41.81%
U-OPT	4.85%***	0.18**	0.95***	63.62%
Panel C: 2008:09 to 2020:09 out-of-sample period				
C-OPT, linear ENet	9.33%**	−0.37**	0.07	3.29%
C-OPT, DNN	7.38%**	0.02	0.09	−0.54%
C-OPT, ensemble	9.08%**	−0.11	0.06	−0.98%
U-OPT	1.71%	−0.12	0.64***	45.26%

Table 5. Alphas. The table reports Lustig, Roussanov, and Verdelhan (2011) factor model estimation results for the C-OPT portfolio based on the linear ENet, DNN, and ensemble forecasts, as well as the U-OPT portfolio. For the C-OPT (U-OPT) portfolio, a mean-variance US investor with a relative risk aversion coefficient of five allocates monthly across available foreign currencies using the linear ENet, DNN, or ensemble (no-change) exchange rate forecasts when predicting currency excess returns. The linear ENet, DNN and ensemble exchange rate forecasts incorporate the information in 70 predictors. For the second through fourth columns, *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

at the 1% level), and it generates a sizable annualized alpha of 4.85% (significant at the 1% level). Reminiscent of Table 3 and Figure 7, the U-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 C-OPT portfolios deliver impressive annualized alphas for both subsamples in Panels B and C of Table 5, all of which are well above 700 basis points and significant at the 5% or 1% level. An interesting pattern emerges with respect to the exposures of the C-OPT portfolios to the carry factor.

The exposures are quite sizable for the first subsample, ranging from 0.69 (DNN) to 0.87 (linear ENet), and all are significant at the 1% level. For the second subsample, the exposures become close to zero, and none is significant at conventional levels. The information contained in the 70 predictors thus leads the investor to effectively disconnect fully from a conventional carry strategy in the second subsample.

7. Conclusion

Short-horizon exchange rate prediction has posed an enduring challenge to researchers in international finance. In this paper, we make considerable progress in resolving the Meese and Rogoff (1983) no-predictability puzzle. Specifically, we establish predictability by showing that out-of-sample forecasts of monthly US dollar exchange rates can consistently and significantly outperform the no-change benchmark over a lengthy out-of-sample period for a group of developed countries. The outperformance is especially strong in late 2008 during the worst part of the Global Financial Crisis.

Our forecasting approach has two key elements. First, we use a panel framework together with a rich information set, which includes ten country characteristics and six global variables; after interacting the country characteristics with the global variables, we have 70 predictors for our panel predictive regressions. It is important to consider a large number of potential predictors, as we cannot know a priori which predictors are the most relevant. Furthermore, including an extensive set of predictors in the forecasting models helps to guard against data mining or snooping. Second, we employ machine learning techniques, including ENet estimation of linear models, which uses penalized regression to alleviate overfitting in our high-dimensional and noisy data setting, and a DNN, which allows for complex nonlinear predictive relationships.

For out-of-sample exchange rate prediction, it is important to move beyond off-the-shelf implementations of machine learning techniques. Monthly exchange rate fluctuations inherently contain a large unpredictable component—which means that we are dealing with very noisy data when training models—so that we need to take additional steps to better guard against overfitting. With respect to fitting linear panel predictive regressions via the ENet, we use the ERIC to tune the shrinkage hyperparameter λ. Relative to conventional cross validation and other tuning methods, the ERIC tends to select a larger value of λ and thus induces greater shrinkage, which better guards against overfitting in our high-dimensional and noisy data environment, helping us to strike the right balance with regard of the bias-variance trade-off. We also set the intercept terms in the linear panel predictive regressions to zero, which imposes the restrictions that the mean exchange rate changes are zero. These restrictions are consistent with the data and reduce the number of parameters that we need to estimate, further helping to guard against overfitting. For the DNNs, we set the intercept terms for the weights to zero, which is analogous to setting the intercept terms in the linear panel predictive regressions to zero. This again helps to alleviate overfitting by reducing the number of parameters that we need to estimate, which is particularly useful given the large number of weights in DNNs. To further guard against overfitting when training the DNNs, we include ℓ_1 and ℓ_2 penalty terms in the objective function and employ dropout. We also consider an ensemble forecast that takes the average of the linear ENet and DNN forecasts. With respect to out-of-sample exchange rate prediction, we are the first to simultaneously use the ENet, deep learning, and an ensemble approach to blend the linear ENet and DDN forecasts. We find that the ensemble forecast exhibits the best overall performance.

The main contribution of this paper is to resolve a long-standing puzzle of the FX market by demonstrating that exchange rates are predictable on an out-of-sample basis at the monthly frequency when we use a rich information set and machine learning techniques. We also extend our analysis along two dimensions. First, to shed light on the sources of exchange rate predictability, we use variable importance analysis to investigate which predictors are the most relevant in the fitted models underlying the forecasts. We also employ ICE curves and PDPs to examine the relationships between individual predictors and expected exchange rate changes. Among other results, we find that the importance of numerous predictors changes significantly over time (often near the advent of the Global Financial Crisis), various country characteristics are important when they interact with global variables (especially global FX volatility), and the predictive relationships posited by relative PPP and UIP hold to a greater extent as global FX volatility increases. These findings can help to guide the development of asset pricing models in international finance that generate exchange rate predictability and time-varying currency risk premia (e.g., Verdelhan 2010; Bansal and Shaliastovich 2013; Farhi and Gabaix 2016). Along this line, Cujean and Hasler (2017) and Gómez-Cram (2022) explore explanations of why stock return predictability concentrates in bad times based on heightened uncertainty with heterogeneous beliefs and delayed reaction to news, respectively, during bad times. Because a number of predictors become more important near the onset of the Global Financial Crisis, our empirical results together with insights from Cujean and Hasler (2017) and Gómez-Cram (2022) appear helpful for building new models to understand exchange rate predictability.

Second, in addition to improving out-of-sample prediction in terms of MSPE, we show that our exchange rate forecasts based on a rich information set and machine learning devices provide considerable economic value to a US investor. Specifically, a mean-variance investor who allocates across foreign currencies experiences substantial increases in utility when they utilize our machine learning forecasts of exchange rate changes in lieu of the no-change benchmark forecast when predicting currency excess returns. In essence, instead of simply relying on the interest rate differential to predict the currency excess return—consistent with the no-change benchmark forecast—the investor benefits substantively by augmenting the interest rate differential with a machine learning forecast of the exchange rate change. The machine learning forecasts are especially valuable to the investor during and after the Global Financial Crisis, when the performance of conventional carry trade strategies diminishes markedly.

References

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

Ang, A. and J. Chen (2011). Yield Curve Predictors of Foreign Exchange Returns. Working Paper.

Avramov, D., S. Cheng, and L. Metzker (forthcoming). Machine Learning Versus Economic Restrictions: Evidence from Stock Return Predictability. *Management Science*.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics* 131:4, 1593–1636.

Bakshi, G. and G. Panayotov (2013). Predictability of Currency Carry Trades and Asset Pricing Implications. *Journal of Financial Economics* 110:1, 139–163.

Bansal, R. and I. Shaliastovich (2013). A Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets. *Review of Financial Studies* 26:1, 1–33.

Bao, Y. (2009). Estimation Risk-Adjusted Sharpe Ratio and Fund Performance Ranking under a General Return Distribution. *Journal of Financial Econometrics* 7:2, 152–173.

Barron, A. R. (1994). Approximation and Estimation Bounds for Artificial Neural Networks. *Machine Learning* 14:1, 115–133.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous Analysis of Lasso and Dantzig Selector. *Annals of Statistics* 37:4, 1705–1732.

Bilson, J. F. O. (1981). The "Speculative Efficiency" Hypothesis. *Journal of Business* 54:3, 435–451.

Brunnermeier, M. K., S. Nagel, and L. H. Pedersen (2009). Carry Trades and Currency Crashes. In: D. Acemoglu, K. Rogoff, and M. Woodford, eds. *NBER Macroeconomics Annual 2008*. Vol. 23. Chicago: University of Chicago Press, pp. 313–347.

Bryzgalova, S., M. Pelger, and J. Zhu (2021). Forest Through the Trees: Building Cross-Sections of Stock Returns. Working Paper.

Burnside, A. C. (2018). Exchange Rates, Interest Parity, and the Carry Trade. Working Paper.

Burnside, A. C., M. S. Eichenbaum, I. Kleshchelski, and S. T. Rebelo (2011). Do Peso Problems Explain the Returns to the Carry Trade? *Review of Financial Studies* 24:3, 853–891.

Burnside, A. C., M. S. Eichenbaum, and S. T. Rebelo (2011). Carry Trade and Momentum in Currency Markets. *Annual Review of Finance and Economics* 3:1, 511–535.

Caldara, D. and M. Iacoviello (forthcoming). Measuring Geopolitical Risk. *American Economic Review*.

Campbell, J. Y. and S. B. Thompson (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies* 21:4, 1509–1531.

Cenedese, G., R. Payne, L. Sarno, and G. Valente (2016). What Do Stock Markets Tell Us About Exchange Rates? *Review of Finance* 20:3, 1045–1080.

Chen, Y.-C. and K. P. Tsang (2013). What Does the Yield Curve Tell Us About Exchange Rate Predictability? *Review of Economics and Statistics* 95:1, 185–205.

Chen, L., M. Pelger, and J. Zhu (2021). Deep Learning in Asset Pricing. Working Paper.

Chernov, M. and D. Creal (2021). The PPP View of Multihorizon Currency Risk Premiums. *Review of Financial Studies* 34:6, 2728–2772.

Chernov, M., M. Dahlquist, and L. A. Lochstoer (2021). Pricing Currency Risks. Working Paper.

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019). Sparse Signals in the Cross-Section of Returns. *Journal of Finance* 74:1, 449–492.

Christiano, L. J. and T. J. Fitzgerald (2003). The Band Pass Filter. *International Economic Review* 44:2, 435–465.

Clarida, R., J. Davis, and N. Pedersen (2009). Currency Carry Trade Regimes: Beyond the Fama Regression. *Journal of International Money and Finance* 28:8, 1375–1389.

Clark, T. E. and M. W. McCracken (2001). Tests of Equal Forecast Accuracy and Encompassing for Nested Models. *Journal of Econometrics* 105:1, 85–110.

Clark, T. E. and K. D. West (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics* 138:1, 291–311.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2021). AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI. Working paper.

Cujean, J. and M. Hasler (2017). Why Does Return Predictability Concentrate in Bad Times? *Journal of Finance* 72:6, 2717–2757.

Cybenko, G. (1989). Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems* 2:4, 303–314.

Dahlquist, M. and H. Hasseltoft (2020). Economic Momentum and Currency Returns. *Journal of Financial Economics* 136:1, 152–167.

Dahlquist, M. and J. Pénasse (forthcoming). The Missing Risk Premium in Exchange Rates. *Journal of Financial Economics*.

Dangl, T. and M. Halling (2012). Predictive Regressions with Time-Varying Coefficients. *Journal of Financial Economics* 106:1, 157–181.

Daniel, K., R. J. Hodrick, and Z. Lu (2017). The Carry Trade: Risks and Drawdowns. *Critical Finance Review* 6:2, 211–262.

Della Corte, P., A. Jeanneret, and E. Patelli (2020). A Credit-Based Theory of the Currency Risk Premium. Working Paper.

Della Corte, P., L. Sarno, and I. Tsiakas (2009). An Economic Evaluation of Empirical Exchange Rate Models. *Review of Financial Studies* 22:9, 3491–3530.

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

Dobrynskaya, V. (2014). Downside Market Risk of Carry Trades. *Review of Finance* 18:5, 1885–1913.

Dong, X., Y. Li, D. E. Rapach, and G. Zhou (2022). Anomalies and the Expected Market Return. *Journal of Finance* 77:1, 639–681.

Eichenbaum, M. S., B. K. Johannsen, and S. T. Rebelo (2021). Monetary Policy and the Predictability of Nominal Exchange Rates. *Review of Economic Studies* 88:1, 192–228.

Engel, C. and K. D. West (2005). Exchange Rates and Fundamentals. *Journal of Political Economy* 113:3, 485–517.

Engel, C. and S. P. Y. Wu (2019). Liquidity and Exchange Rates: An Empirical Investigation. NBER Working Paper No. 25397.

Fama, E. F. (1984). Forward and Spot Exchange Rates. *Journal of Monetary Economics* 14:3, 319–338.

Fan, Y. and C. Y. Tang (2013). Tuning Parameter Selection in High Dimensional Penalized Likelihood. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 75:3, 531–552.

Farhi, E. and X. Gabaix (2016). Rare Disasters and Exchange Rates. *Quarterly Journal of Economics* 131:1, 1–52.

Filippou, I. and M. P. Taylor (2017). Common Macro Factors and Currency Premia. *Journal of Financial and Quantitative Analysis* 52:4, 1731–1763.

Flynn, C. J., C. M. Hurvich, and J. S. Simonoff (2013). Efficiency for Regularization Parameter Selection in Penalized Likelihood Estimation of Misspecified Models. *Journal of the American Statistical Association* 108:503, 1031–1043.

Frenkel, J. A. (1976). A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence. *Scandinavian Journal of Economics* 78:2, 200–224.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting Characteristics Nonparametrically. *Review of Financial Studies* 33:5, 2326–2377.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* 29:5, 1189–1232.

Froot, K. A. and R. H. Thaler (1990). Anomalies: Foreign Exchange. *Journal of Economic Perspectives* 4:3, 179–192.

Funahashi, K.-I. (1989). On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural Networks* 2:3, 183–192.

Goldstein, A., A. Kapelner, J. Bleich, and E. Pitkin (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics* 24:1, 44–65.

Gómez-Cram, R. (2022). Late to Recessions: Stocks and the Business Cycle. *Journal of Finance* 77:2, 923–966.

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

Goyal, A. and A. Saretto (2009). Cross-Section of Option Returns and Volatility. *Journal of Financial Economics* 94:2, 310–326.

Goyal, A. and I. Welch (2003). Predicting the Equity Premium with Dividend Ratios. *Management Science* 49:5, 639–654.

Goyal, A. and I. Welch (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 21:4, 1455–1508.

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

Gu, S., B. Kelly, and D. Xiu (2020). Empirical Asset Pricing via Machine Learning. *Review of Financial Studies* 33:5, 2223–2273.

Han, Y., A. He, D. E. Rapach, and G. Zhou (2021). Expected Stock Returns and Firm Characteristics: E-LASSO, Assessment, and Implications. Working Paper.

Hansen, L. P. and R. J. Hodrick (1980). Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis. *Journal of Political Economy* 88:5, 829–853.

Hastie, T. and J. Qian (2016). Glmnet Vignette. Working Paper.

Hau, H. and H. Rey (2006). Exchange Rates, Equity Prices, and Capital Flows. *Review of Financial Studies* 19:1, 273–317.

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-Varying Short-Horizon Predictability. *Journal of Financial Economics* 99:3, 560–580.

Hinton, G. E., N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov (2012). Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. Working Paper (arXiv:1207.0580).

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

Hoerl, A. E. and R. W. Kennard (1970). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* 12:1, 69–82.

Hornik, K. (1991). Approximation Capabilities of Multilayer Feedforward Networks. *Neural Networks* 4:2, 251–257.

Hornik, K., M. Stinchcombe, and H. White (1989). Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks* 2:5, 359–366.

Hui, F. K. C., D. I. Warton, and S. D. Foster (2015). Tuning Parameter Selection for the Adaptive Lasso Using ERIC. *Journal of the American Statistical Society* 110:509, 262–269.

Jiang, Z., A. Krishnamurthy, and H. Lustig (2021). Foreign Safe Asset Demand and the Dollar Exchange Rate. *Journal of Finance* 76:3, 1049–1089.

Jordà, Ò. and A. M. Taylor (2012). The Carry Trade and Fundamentals: Nothing to Fear but FEER Itself. *Journal of International Economics* 88:1, 74–90.

Jurek, J. W. (2014). Crash-Neutral Currency Carry Trades. *Journal of Financial Economics* 113:3, 325–347.

Kilian, L. and M. P. Taylor (2003). Why Is It So Difficult to Beat the Random Walk Forecast of Exchange Rates? *Journal of International Economics* 60:1, 85–107.

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

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the Cross Section. *Journal of Financial Economics* 135:2, 271–292.

Kremens, L. and I. Martin (2019). The Quanto Theory of Exchange Rates. *American Economic Review* 109:3, 810–843.

Lettau, M., M. Maggiori, and M. Weber (2014). Conditional Risk Premia in Currency Markets and Other Asset Classes. *Journal of Financial Economics* 114:2, 197–225.

Lilley, A., M. Maggiori, B. Neiman, and J. Schreger (forthcoming). Exchange Rate Reconnect. *Review of Economics and Statistics.*

Liu, Y., G. Zhou, and Y. Zhu (2021). Maximizing the Sharpe Ratio: A Genetic Programming Approach. Working paper.

Lundberg, S. M. and S.-I. Lee (2017). A Unified Approach to Interpreting Model Predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems.*

Lustig, H., N. L. Roussanov, and A. Verdelhan (2011). Common Risk Factors in Currency Markets. *Review of Financial Studies* 24:11, 3731–3777.

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

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

Mark, N. C. (1995). Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability. *American Economic Review* 85:1, 201–218.

McCracken, M. W. (2007). Asymptotics for Out of Sample Tests of Granger Causality. *Journal of Econometrics* 140:2, 719–752.

Meese, R. A. and K. Rogoff (1983). Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample? *Journal of International Economics* 14:1–2, 3–24.

Meinshausen, N. and B. Yu (2009). Lasso-Type Recovery of Sparse Representations for High-Dimensional Data. *Annals of Statistics* 37:1, 246–270.

Melvin, M. and D. Shand (2017). When Carry Goes Bad: The Magnitude, Causes, and Duration of Currency Carry Unwinds. *Financial Analysts Journal* 73:1, 121–144.

Melvin, M. and M. P. Taylor (2009). The Crisis in the Foreign Exchange Market. *Journal of International Money and Finance* 28:8, 1317–1330.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012a). Carry Trades and Global Foreign Exchange Volatility. *Journal of Finance* 67:2, 681–718.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012b). Currency Momentum Strategies. *Journal of Financial Economics* 106:3, 660–684.

Molodtsova, T. and D. H. Papell (2009). Out-of-Sample Exchange Rate Predictability with Taylor Rule Fundamentals. *Journal of International Economics* 77:2, 167–180.

Mueller, P., A. Stathopoulos, and A. Vedolin (2017). International Correlation Risk. *Journal of Financial Economics* 126:2, 270–299.

Mussa, M. (1976). The Exchange Rate, the Balance of Payments and Monetary and Fiscal Policy under a Regime of Controlled Floating. *Scandinavian Journal of Economics* 78:2, 229–248.

Nagel, S. (2021). *Machine Learning in Asset Pricing*. Princeton: Princeton University Press.

Neely, C. J. (2015). Unconventional Monetary Policy Had Large International Effects. *Journal of Banking and Finance* 52:1, 101–111.

Neely, C. J. and P. A. Weller (2013). Lessons from the Evolution of Foreign Exchange Trading Strategies. *Journal of Banking and Finance* 37:10, 3783–3798.

Neely, C. J., P. A. Weller, and J. M. Ulrich (2009). The Adaptive Markets Hypothesis: Evidence from the Foreign Exchange Market. *Journal of Financial and Quantitative Analysis* 44:2, 467–488.

Rapach, D. E., J. K. Strauss, J. Tu, and G. Zhou (2019). Industry Return Predictability: A Machine Learning Approach. *Journal of Financial Data Science* 1:3, 9–28.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies* 23:2, 821–862.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). International Stock Return Predictability: What Is the Role of the United States? *Journal of Finance* 68:4, 1633–1662.

Rapach, D. E. and G. Zhou (2013). Forecasting Stock Returns. In: G. Elliott and A. Timmermann, eds. *Handbook of Economic Forecasting*. Vol. 2A. Amsterdam: Elsevier, pp. 328–383.

Ready, R. C., N. L. Roussanov, and C. Ward (forthcoming). Commodity Trade and the Carry Trade: A Tale of Two Countries. *Journal of Finance*.

Rolnick, D. and M. Tegmark (2018). The Power of Deeper Networks for Expressing Natural Functions. In: *Sixth Annual International Conference on Learning Representations*.

Rossi, B. (2013). Exchange Rate Predictability. *Journal of Economic Literature* 51:4, 1063–1119.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics* 6:2, 461–464.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15:56, 1929–1958.

Taylor, J. B. (1993). Discretion Versus Policy Rules in Practice. *Carnegie-Rochester Conference Series on Public Policy* 39:1, 195–214.

Taylor, M. P. (1995). The Economics of Exchange Rates. *Journal of Economic Literature* 33:1, 13–47.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)* 58:1, 267–288.

Verdelhan, A. (2010). A Habit-Based Explanation of the Exchange Rate Risk Premium. *Journal of Finance* 65:1, 123–146.

Wang, H., B. Li, and C. Leng (2009). Shrinkage Tuning Parameter Selection with a Diverging Number of Parameters. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 71:3, 671–683.

West, K. D. (1996). Asymptotic Inference about Predictive Ability. *Econometrica* 64:5, 1067–1084.

Zhang, C.-H. and J. Huang (2008). The Sparsity and Bias of the Lasso Selection in High-Dimensional Linear Regression. *Annals of Statistics* 36:4, 1567–1594.

Zou, H. and T. Hastie (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67:2, 301–320.

Zou, H., T. Hastie, and R. Tibshirani (2007). On the 'Degrees of Freedom' of the Lasso. *Annals of Statistics* 35:5, 2173–2192.