Exchange rate predictability and dynamic Bayesian learning

Joscha Beckmann*
University of Greifswald
Kiel Institute for the World Economy

Gary Koop†
University of Strathclyde

Dimitris Korobilis‡
University of Essex

Rainer Schüssler§
University of Rostock

October 4, 2019

Abstract

This paper considers how an investor in the foreign exchange market can exploit predictive information by means of flexible Bayesian inference. Using a benchmark vector autoregressive model, the investor is able to revise each period past predictive mistakes and learn about important data features such as parameter instability and model switching. The proposed methodology is specified in order to reflect a wide array of established empirical and theoretical patterns of exchange rates. In a thorough investigation of monthly exchange rate predictability for ten countries, we find that an investor using the proposed flexible methodology for dynamic asset allocation achieves significant economic gains relative to benchmark strategies. In particular, we find strong evidence for sparsity, fast model switching and exploiting the exchange rate cross-section.

Keywords: Exchange rates; Bayesian vector autoregression; forecasting; dynamic portfolio allocation; economic fundamentals

JEL Classification: C11; D83; F31; G12; G15; G17

*Department of Economics, University of Greifswald, Friedrich-Loeffler-Straße 70, 17489 Greifswald. E: joscha.beckmann@uni-greifswald.de. T: +49 (0)3834 420-2482.

†Corresponding author.
Department of Economics, University of Strathclyde, 199 Cathedral Street, Glasgow.
E: gary.koop@strath.ac.uk. T: +44 (0)141 548 3840.

‡Essex Business School, Wivenhoe Park, Colchester, CO4 3SQ, UK.
E: d.korobilis@essex.ac.uk. T: +44 (0) 1206 873 072.

§Department of Economics and Social Sciences, University of Rostock, Ulmenstrasse 69, 18057 Rostock.
E: rainer.schuessler@uni-rostock.de. T: +49 (0)381 498 4340.

1 Introduction

Understanding and predicting the evolution of global exchange rates has long been a key component of the research agenda in international economics and finance. Yet, the early finding by Meese and Rogoff (1983) that structural models cannot offer predictability superior to that of a random walk has not been convincingly overturned. The voluminous existing literature on exchange rate forecasting adopts many different econometric methods. Broadly speaking, these differences fall in the following categories. First, they differ in whether they are multivariate (e.g. building a Vector Autoregressive, VAR, model involving a cross-section of exchange rates for many countries) or univariate. Second, they differ in which predictors they use. Third, they differ in how they treat the fact that there may be many potential predictors, most of which are unimportant. Fourth, they differ in whether they allow for dynamic model change (i.e. whether the best forecasting model can involve different predictors at different points in time) or not. Fifth, they differ in whether they allow for parameter change (both in VAR or regression coefficients and in volatilities) or not.

In the present paper, we adopt an econometric approach that allows for a general treatment of each of these five categories. That is, its most flexible specification is a high-dimensional multivariate time series model involving the full cross-section of exchange rates, a large number of exogenous predictors and time-variation in coefficients and volatilities. But our algorithm allows for decisions relating to these categories to be made in a data-based fashion using dynamic model selection methods. That is, the estimation procedure automatically decides whether to set a coefficient on a predictor or a VAR lag to be zero (or not). Most importantly, it does so in a dynamic manner, allowing for different forecasting models to be used at different points of time. Thus, decisions about specification choices (i.e. different predictors, different VARs, different degrees of model switching) are all made automatically in a time-varying fashion.

The econometric methods are based on Koop and Korobilis (2013) but, as described in Section 4, we improve on and extend them in important directions of relevance for our empirical application. In an exchange rate forecasting exercise involving 10 countries and a large set of predictors, we demonstrate the ability of our modelling approach to forecast exchange rates using both statistical criteria (i.e. log predictive likelihoods) and economic criteria (i.e. the economic gains that an investor would make from adopting our approach). We find that an investor would make substantial economic gains from adopting our dynamic model learning strategy relative to a multivariate random walk benchmark.

The remainder of the paper is organized as follows. Section 2 provides a literature review and uses it to motivate various aspects of our modelling strategy. Section 3 discusses the data while Section 4 lays out our econometric methods. Section 5 describes the dynamic asset allocation strategy we use for the economic evaluation of our forecasts. Section 6 presents and discusses our empirical results. Section 7 provides some robustness checks and Section 8 concludes. We present technical details of our econometric methods along with additional empirical results and further details regarding the underlying data in an online appendix.

2 Literature review

In this section, we briefly review the relevant literature, with a focus on the five stylazie described in the Introduction. The literature on exchange rate forecasting is voluminous and, hence, for the sake of brevity, we refer the reader to Rossi (2013) which is an excellent survey of the literature up to 2013. Rossi (2013) concludes that exchange rate predictability largely depends on choice of sample, currency, and modelling strategy.[1] However, it is often hard to beat a simple random walk specification.

A large part of the existing literature attempts to use macroeconomic fundamentals (e.g. Purchasing Power Parity, PPP, or Uncovered Interest Rate Parity, UIP) to forecast exchange rates with little success. A possible explanation for this poor performance is developed in Bacchetta and van Wincoop (2004). These authors conjecture that market participants

[1] See also Sarno (2005) for an early comprehensive survey on the major challenges in exchange rate modelling.

attach excessive weight to observable fundamentals that deviate from their long-run trend. As a result of this, agents quickly switch between models over time (Bacchetta and van Wincoop, 2006; Markiewicz, 2012). Thus, different fundamentals may be relevant only for short periods at different points in time depending on deviations from long-run trends. This motivates a dynamic learning framework such as ours which allows for dynamic model change and time-varying parameters. Fast model switching has also been found to be of crucial importance in the empirical exchange rate literature. Sarno and Valente (2009) discuss how the fact that there is evidence of a weak link between in-sample fit and out-of-sample predictability complicates the choice of selecting an appropriate model even if fundamentals contain valuable information about the path of the exchange rate.[2] In this regard, we want to have a method which allows for the quick and transparent disentanglement of the informational content of various fundamentals in a dynamic fashion. Our modelling approach allows for this.

The survey paper of Rossi (2013) also discusses the empirical evidence relating to parameter change and other nonlinearities. The evidence is not strong, although some papers (Byrne, Korobilis, and Ribeiro, 2016; Canova, 1993; Rossi, 2006) find some benefits from allowing for time variation in parameters in some cases. Thus, in this paper, we adopt an approach which can decide, in a data-based fashion, whether parameters should be time-varying or not.

The question of whether there are benefits of working with a multivariate time series model such as a VAR involving a cross-section of exchange rates is also debated. Such an approach has the advantage that it exploits information in the co-movements and common dynamics in exchange rates. There is some evidence that doing so can improve exchange rate forecasts. Carriero, Kapetanios, and Marcellino (2009) work with a large Bayesian VAR involving a cross-section of exchange rates and finds forecast improvements over univariate

[2]Decomposing out-of-sample forecasting performance of exchange rate returns into the three components i) time variation, ii) (in-sample) predictive content and iii) overfitting, Rossi and Sekhposyan (2011) find the lack of predictive content to be the major explanation for the lack of short-term forecasting ability, while they consider instabilities as a main reason for the lack of medium-term (one-year ahead) forecasting ability.

methods. Abbate and Marcellino (2018) extend Carriero, Kapetanios, and Marcellino (2009) by allowing for time-varying coefficients and volatilities and find the latter to be particularly useful in improving forecast performance. These considerations suggest that working with VARs with time-varying volatilities is potentially important and our modelling approach does so. However, unlike our approach, this area of the literature uses single models and does not allow for dynamic model switching.

Another issue which arises when we have many potential predictors is the need for some method for ensuring parsimony so as to avoid overfitting and poor out-of-sample results. Indeed, even in univariate models, papers such as Ackermann, Pohl, and Schmedders (2016) find parameter estimation error to be substantial and, hence, they use no predictors when building a diversified FX portfolio. Instead they focus solely on exploiting volatility timing. However, several recent papers have successfully used data reduction methods, priors or model averaging methods to minimize overfitting concerns. Abbate and Marcellino (2018) include a large number of exogenous predictors and uses principal component methods to reduce their dimensionality. Other techniques have been successfully used including elastic nets (Li, Tsiakas, and Wang, 2015), gradient boosting (Berge, 2014) and model averaging/selection (Della Corte, Sarno, and Tsiakas, 2008; Della Corte and Tsiakas, 2012; Kouwenberg, Markiewicz, Verhoeks, and Zwinkels, 2017).

Our approach reflects these concerns, ensuring parsimony through the use of dynamic model averaging and selection methods. We stress that, relative to the univariate literature cited, our methods are multivariate and dynamic, allowing for the selection of different parsimonious models at different points in time. The dynamic feature we share with Byrne, Korobilis, and Ribeiro (2016). However, the latter paper is univariate and, thus, does not exploit the cross-country information as does our multivariate approach.

The recent literature has also explored the implications of exchange rate predictability (or a lack thereof) for the investor wishing to build an investment portfolio involving various exchange rates. See, for instance, Abhyankar, Sarno, and Valente (2005) Bacchetta and van Wincoop (2006), Bacchetta and van Wincoop (2013), Della Corte, Sarno, and Tsiakas (2008) and Markiewicz (2012). This is another theme that we investigate in our work when specifying our models and evaluating their performance.

Motivated by these considerations, our econometric approach takes the perspective of an investor who learns from past mistakes. We formalize this setting econometrically using the notion of dynamic Bayesian learning. In it, the investor can adapt to a new forecasting environment each time period by switching to a new model. The decision to switch is based on past forecast errors. The result is an extremely flexible framework that learns quickly from recent forecast performance. Our empirical framework has several desirable features. First, our set of models are VARs with exogenous predictors which allow for time-varying volatilities and, in some cases, time-varying VAR coefficients.[3] We work with a large number of the VARs and allow for switching between them. Thus, coefficients, volatilities and fundamentals relevant for forecasting change adaptively each time period depending on recent forecast performance. Hence, our approach can adapt to abrupt structural changes or sudden shifts in the investor's information set. Our estimation methods are Bayesian so that the investor's decisions account for parameter uncertainty. At the same time Bayesian methods offer a natural setting for imposing statistical shrinkage which, as discussed above, has been shown to be important for exchange rate predictability when working with large numbers of predictors and a large cross-section of exchange rates. We develop a learning methodology that allows also for optimal statistical shrinkage each period, such that exchange rate predictions are sufficiently regularized. For example, in periods where VAR lags and macroeconomic fundamentals have low predictive ability, our shrinkage methods can remove them from the forecasting model resulting in a multivariate random walk model with time-varying volatility. Such a specification is broadly similar to the unobserved components forecasting model of Stock and Watson (2007). In other words, our dynamic learning approach has a choice between a wide range of popular specifications used previously in

[3] In the discussion below we will simply refer to our models as VARs, although we stress that they do have exogenous variables and time-varying parameters.

the literature and automatically chooses the best one in a dynamic fashion.

It is also worth mentioning that, unlike some other approaches, We allow for model incompleteness (see, e.g., Billio, Casarin, Ravazzolo, and Van Dijk (2013)). That is, we do not assume that one of our entertained VARs reflects the correct data generating process.

To preview our empirical results, we do find that model switching has a big role to play. At most points in time only one or a few predictors are relevant for forecasting with UIP typically being the most useful one. But there are also several periods where a simple multivariate random walk with stochastic volatility is the best forecasting model. From an investor's perspective, this leads to quickly changing portfolio weights. We find such an investor would experience substantial economic gains relative to the random walk model with time-varying volatility. A risk-averse mean-variance investor is willing to pay an annualized fee of several hundred basis points (after transaction costs) for switching from the dynamic portfolio strategy implied by the random walk with constant volatility model to the dynamic asset allocation implied by our VAR-based approach. Similarly, we find that the annualized Sharpe ratio after transaction costs increases substantially from adopting our approach.

3 Data

All of our individual model configurations are VARs (or extensions thereof) which involve a cross-section of exchange rates as dependent variables. Some models also include additional exogenous predictors. For reasons discussed below, we present results for both a long sample and short sample of data.

We use the G10 set of the world's most heavily traded currencies: the Australian dollar (AUD), the Canadian dollar (CAD), the Euro (EUR)[4], the Japanese yen (JPY), the New Zealand dollar (NZD), the Norwegian krone (NOK), the Swedish krona (SWK), the Swiss franc (SWF), the Great Britain pound sterling (GBP) and the US dollar. All currencies are expressed in terms of the US dollar and are end-of-month exchange rates which enter the

[4]We use German instead of Euro data prior to 1999.

model as discrete returns. Thus, we have nine exchange rates, each relative to the US dollar, entering our VAR.

Our long data set contains only monthly exchange rate data and runs from 1973:01 until 2016:12. We also work with a shorter sample which runs from 1986:01 until 2016:12. With this data set we include additional predictors. These include the Uncovered Interest Parity (UIP), the percentage change in stock prices over the past 12 months (STOCK GROWTH), the difference between long and short term interest rates (INT DIFF) and the percentage change in the nominal oil price (OIL). UIP, STOCK GROWTH and INT DIFF have been widely used in studies such as Wright (2008) and previous research shows that US dollar exchange rates are affected by the price of oil (Lizardo and Mollick, 2010). With regards to the interest rates, we use one-month LIBOR and Eurodeposit interest rates as well as 10 Year government bonds.

Why work with two different sample spans? The longer sample starting in 1973 spans the entire time period after the breakdown of the Bretton Woods agreement. For this longer period, we only use exchange rate data in our models and do not consider additional predictors. We do this for two reasons. First, this enables us to evaluate whether exploiting the cross-section dimension of exchange rates results in economic gains over the longest available sample. We expect this to happen since it is well-established that exchange rates co-move due to cross-country arbitrage. Secondly, several fundamentals which could potentially be used as exchange rate predictors are either not available or heavily revised. The 1970s included several periods of economic turbulence, such as the oil price shock and changes in exchange rate arrangements for some currencies such as Sweden or Norway. We note that these periods are not part of our out-of-sample evaluation period. The shorter sample, which begins in 1986, covers a period where all exchange rates are largely freely floating and where real-time data availability of the additional predictors is less of a concern. In particular, we are able to obtain data on several variables which reflect fundamentals for a wide range of countries. Furthermore, this shorter sample excludes large exchange rate adjustments by central banks as observed for Sweden in the early 1980s. Analyzing both a short and a long sample provides an implicit robustness check when it comes to disentangling the relevant features of our empirical framework.

The forecast evaluation period when using the long sample spans the time from 1990:01 to 2016:12 for a total of 324 observations. When using the short sample, the evaluation period runs from 1996:01 to 2016:12 for a total of 252 observations.

The online data appendix provides sources and other details about the data.

4 Econometric methods

Our dynamic Bayesian learning methodology involves working with many VARs which differ with respect to included VAR lags, exogenous predictors, shrinkage intensity and the amount of time variation in parameters. In this section, we first describe our econometric methods for working with a single VAR, before expanding our methodology to the case of many models. We build on the approach of Koop and Korobilis (2013) for estimating large Bayesian VARs with time-varying parameters. We extend their methods in several directions so as to obtain a methodology appropriate for exchange rate forecasting in light of the empirical issues discussion in Section 2. First, we include exogenous predictors and fundamentals into the VAR and specify how they enter and leave the model by means of a flexible shrinkage prior. Second, we use a recursively updated Wishart posterior estimator of the covariance matrix that generalizes to the time-varying case standard Bayesian results for constant covariance matrix estimators. The Wishart Matrix Discounting (WMD) scheme we adopt builds on West and Harrison (1997) and Triantafyllopoulos (2011). In contrast to the point heuristic estimator used in Koop and Korobilis (2013), the WMD estimator allows to characterize the full posterior distribution of the VAR covariance matrix. As a consequence, predictions and investor decisions fully incorporate parameter uncertainty as in Barberis (2000). Third, we employ a real-time data-adaptive procedure for estimating the degree of time-variation in our dynamic model learning strategy following Beckmann and Schüssler (2016).

4.1 The VAR

Our starting point is a time-varying parameter VAR with exogenous variables that can be written as a general regression model of the form

$$y_t = x_t\beta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_t) \tag{1}$$

$$\beta_{t+1} = \beta_t + u_t, \quad u_t \sim N(0, \Omega_t), \tag{2}$$

where y_t is an $M \times 1$ vector containing observations on M time series variables (in our case, discrete exchange-rate returns for nine countries). x_t is a matrix where each row contains predetermined variables in each VAR equation, namely an intercept, (lagged) exogenous variables, and p lags of each of the M variables. We divide the set of exogenous variables into two groups: N_x denotes the number of variables which are asset specific and considered as relevant only for a specific exchange rate. For instance, in the equation for the UK currency the UIP for the UK belongs in this class. N_{xx} denotes the number of non asset-specific variables which are supposed to be potentially relevant for all currencies in the setting (e.g. oil price changes). Thus, we have, $k = M(1 + p \cdot M + N_x + N_{xx})$ elements in β_t.[5] Following a large literature in economics and finance[6] we assume that β_t evolves as a multivariate random walk without drift, with covariance matrix Ω_t of dimension $k \times k$.

Complete details of the statistical methods we use for estimating a single VAR are given in the Technical Appendix. Here we briefly outline the main ideas. In a changing environment, the investor needs to learn about changes in intercepts (average conditional returns), VAR lags, regression coefficients (effects of fundamentals and other exogenous predictors) and time-varying volatilities and correlations between assets (risk). For the Bayesian investor,

[5] For our main results we set the lag length $p = 6$.

[6] See Byrne, Korobilis, and Ribeiro (2016) or Dangl and Halling (2012) and references therein.

the quantity of interest is next period's multivariate predictive return distribution:

$$p\left(y_t|y^{t-1}\right) = \iint p\left(y_t|y^{t-1}, \beta_t, \Sigma_t\right) p\left(\beta_t|y^{t-1}, \Sigma_t\right) p\left(\Sigma_t|y^{t-1}\right) d\beta_t d\Sigma_t, \tag{3}$$

where $y^{t-s} = (y_1, ..., y_{t-s})'$ denotes the observations through time $t-s$. We obtain the marginal predictive return distribution (3) by integrating out the uncertainty about coefficients (β_t) and the observational covariance matrix (Σ_t). For details we refer the reader to the Technical Appendix.

Assuming Σ_t and Ω_t are known, standard Bayesian methods for state space models involving the Kalman filter can be used to estimate β_t and obtain the predictive distribution of the returns.

In practice, the econometrician/investor does not observe Σ_t and Ω_t. In small models, these parameters can be estimated with Markov Chain Monte Carlo (MCMC) methods using approaches such as Chib, Nardari, and Shephard (2006). However, when working with larger models MCMC methods become too computationally demanding. Accordingly, we rely on exponential discounting methods. These are filtering methods in which Σ_t and Ω_t are updated by looking at recent data and discounting more distant observations at a higher rate. Thus, if an abrupt change occurs, parameter estimates can adapt at a faster rate compared to an investor who tracks parameters based on the whole, equally weighted, sample of data. Exponential discounting methods are well established in the state space literature (see West and Harrison (1997)).

The mechanics behind the discounting approach is described in the Technical Appendix.[7] The key point to note here is that they involve the use of discount factors δ and λ to control the dynamics of Σ_t and Ω_t, respectively. These two discount factors control how quickly/slowly investors learn from past forecasting performance. When $\delta = 1$ (similarly for λ), then the investor uses all available historical observations, equally weighted, to update

[7] Discount factors are well established; see the J.P. Morgan/Reuters (1996) `Riskmetrics` model, and Dangl and Halling (2012) for an application in stock return predictability. For a general treatment, see West and Harrison (1997).

volatilities and parameters. For values less than one, older observations are exponentially penalized, giving more weight to recent observations. As we work with monthly data, we set $\delta = 0.97$, following J.P. Morgan/Reuters (1996) and, hence, allowing for time-varying observational volatilities and covariances. For our main results, we select constant slope coefficients, setting $\lambda = 1$ and investigate the effects of time-varying slope parameters ($\lambda = 0.99$) as a robustness check. The choice of constant slope parameters for our main results is motivated by the empirical finding that in medium-sized VARs such as ours, it is common to find strong evidence for time-varying error variances, but little evidence in favor of time-varying VAR coefficients (Koop and Korobilis, 2013). Time-varying VAR coefficients may be even detrimental for portfolio performance in the case of FX portfolios (Abbate and Marcellino, 2018).[8]

The investor/econometrician needs to specify prior beliefs about the initial state β_0. The prior we use is of the form $\beta_0 \sim N(0, \Omega_0)$. The amount of prior information that one imposes on the initial state can markedly affect our ability to track parameters successfully and produce accurate predictions. Here we follow a large literature in economics and finance that specifies Ω_0 using Minnesota prior shrinkage. The Minnesota prior is the most popular prior for Bayesian VARs with Banbura, Giannone and Reichlin (2010) being an early example of its use with a large Bayesian VARs and Koop and Korobilis (2013) using it with large TVP-VARs.

The Minnesota prior is typically controlled by a single shrinkage parameter, see Bańbura, Giannone, and Reichlin (2010) and citations therein. In order to deal with prior sensitivity associated selecting this shrinkage parameter, Giannone, Lenza, and Primiceri (2015) and Koop and Korobilis (2013) use information in the data to learn about its value.[9] We adopt a similar approach and allow the degree of shrinkage in the Minnesota prior to adaptively

[8] In a broader context, Chan and Eisenstat (2018) find that gains in time-varying parameter VARs compared to constant VARs stem from time variation in the error covariances rather than from time variation in VAR coefficients.

[9] **?** is a recent study pointing to the importance of prior sensitivity in Bayesian VARs and the need to use sensible strategies to minimize arbitrariness.

change over time. Furthermore, we extend this prior to allow for richer shrinkage patterns. Instead of having one shrinkage parameter for all VAR coefficients, we allow for multiple shrinkage parameters. When working with the short sample, there are seven of these, $\gamma_1, ..., \gamma_7$, that control the shrinkage of different sets of coefficients in the VAR. In particular, we have a shrinkage parameter for intercepts, own lags, cross lags and each individual asset-specific and non asset-specific variable. When working with the long sample, which does not include the exogenous predictors, we only have the first three of these. We choose a value for each of them from a grid of values which includes zero. Note that setting $\gamma_i = 0$ implies that the i^{th} variable (or block of variables) is excluded from the model. The variable (or block of variables) may refer to exogenous variables, VAR lags or the intercepts. Hence, our method allows for the exclusion of model elements such as VAR lags or predictors if this is empirically warranted.

Full details of our statistical methods are given in the Technical Appendix.

4.2 Dynamic model learning

Estimation of a particular VAR involves setting each of $\delta, \lambda, \gamma_1, ..., \gamma_7$ to a particular value. In practice, we fix δ and λ and consider a grid of values for each of $\gamma_1, ..., \gamma_7$ to allow for variable exclusion and different degrees of shrinkage intensity (see the Technical Appendix). If we consider every possible combination of values taken from all of these grids we have 32 choices of shrinkage parameters when using the long sample and 512 choices for the short sample. We interpret a choice as defining a model that the investor has at their disposal at each point of time upon which they could base their portfolio allocation. In order to allow for the investor to make an optimal choice each period t, we use the notion of dynamic model learning (DML). Dynamic model learning involves selecting, at each point in time, the model specification with the highest discounted joint log predictive likelihood at that time. The predictive likelihood is a measure of out-of-sample forecasting ability that takes into account the entire predictive distribution; see Geweke and Amisano (2012). The individual model configuration with the highest discounted joint log predictive likelihood is used in order to obtain the predictive mean and covariance matrix. These are a crucial input in portfolio optimization. Our motivation for using learning based on past forecast performance is that it potentially allows for a different model at each point in time. Such a feature is likely particularly useful in times of abrupt change. If we were to use a single VAR, gradual parameter changes are accommodated if the discount factors δ and λ are below one. But this is not the same as switching between entirely different models as dynamic model learning allows for.

In this dynamic model learning setting, the discounted joint predictive likelihood (DPL) can be calculated as

$$DPL_{t|t-1,j} = \prod_{i=1}^{t-1} \left[p_j \left(y_{t-i} | y^{t-i-1} \right) \right]^{\alpha^i},$$

where $p_j \left(y_{t-i} | y^{t-i-1} \right)$ denotes the predictive likelihood of model j in period i and $t|t-1$ subscripts refer to estimates made of time-t quantities given information available at time $t-1$. Hence, model j will receive a higher value at a given point in time if it has forecast well in the recent past, using the predictive likelihood (i.e., the predictive density evaluated at the actual outcome) as the evaluation criterion. The interpretation of "recent past" is controlled by the the discount factor α, reflecting exponential decay. For example, if $\alpha = 0.95$, forecast performance three years ago receives approximately 15% as much weight as the forecast performance last period. If $\alpha = 0.90$, then forecast performance three years ago receives only about 2% as much weight. The case $\alpha = 1$ implies no discounting and the discounted predictive likelihood is then proportional to the marginal likelihood. Lower values of α are associated with more rapid switching between models. We consider a range of values for α and, at each point in time, choose the best value for it. In this way, we can allow for times of fast model switching and times of slow model switching.

At time τ, we choose the best value for α as the one which has produced the model with the highest product of predictive likelihoods[10] in the past from $t = 1, ..., \tau$. We consider the

[10] We stress that we are not using the DPL when choosing between different values for α. The DPL is

following grid of values: $\alpha \in \{0.20; 0.40; 0.50; 0.60; 0.70; 0.80; 0.90; 0.95; 0.99; 1\}$.

5 Dynamic asset allocation

5.1 Portfolio allocation

We design an international asset allocation strategy that involves trading the US dollar and nine other currencies. Consider a US investor who builds a portfolio by allocating their wealth between ten bonds: one domestic (US), and the nine foreign bonds. The US bond return is r_f. Define $y_t = (y_{1,t}, ..., y_{9,t})'$. At each period, the foreign bonds yield a riskless return in the local currency but a risky return due to currency fluctuations in US dollars. The expectation of the risky return from the investment in country $i's$ bonds, $r_{i,t}$, at time $t-1$ is equal to $E_{t-1}(r_{i,t}) = int_{i,t-1} + y_{i,t}$.[11] The only risk the US investor is exposed to is foreign exchange (FX) risk. Every period the investor takes two steps. First, they use the currently selected model (i.e., the model with the highest discounted sum of predictive likelihoods) to forecast the one-period ahead exchange rate returns and the predictive covariance matrix. Second, using these predictions, they dynamically rebalance their portfolio by calculating the new optimal weights. This setup is designed to assess the economic value of exchange rate predictability and to dissect which sources of information are valuable for asset allocation.

We evaluate our models within a dynamic mean-variance framework, implementing a maximum expected return strategy. That is, we consider an investor who tries to find the point on the efficient frontier with the highest possible (ex-ante) return, subject to achieving a target conditional volatility and a given horizon of the investor (one-month ahead for our main results). Define $r_t = (r_{1,t}, ..., r_{9,t})'$, $\mu_{t|t-1} = E_{t-1}(r_t)$ as its expectation. The portfolio allocation problem involves choosing weights, $w_t = (w_{1,t}, ..., w_{9,t})'$ attached to each of the 9

only used to select the best model for a given value of α.

[11] We use $y_{i,t}$, the discrete exchange rate returns, rather than log returns Δs_t, as, in the context of portfolio optimization, it is important to distinguish discrete and log returns.

foreign bonds (with $1-\sum_{i=1}^{9} w_{i,t}$ being the weight attached to the domestic bond):

$$\begin{aligned} & \max_{w_t} \left\{ \mu_{p,t|t-1} = w_t^{'} \mu_{t|t-1} + (1-w_t^{\prime}\iota) r_f - \tau \left(\iota^{\prime} \left| w_t - w_{t-1} \circ \frac{1+r_t}{1+r_{p,t}} \right| \right) \right\} \\ & \text{subject to} \\ \left(\sigma_p^*\right)^2 \quad = \quad & \underbrace{w_t^{\prime} \frac{\delta n_{t-1}}{\delta n_{t-1}-2} \left(x_{t-1} \Omega_{t|t-1} x_{t-1}^{'} + Q_{t|t-1} \right) w_t}_{\text{estimate of the predictive covariance matrix}}, \end{aligned}$$

where $\mu_{p,t|t-1}$ is the conditional expected portfolio return and $\left(\sigma_p^*\right)^2$ the target portfolio variance. ι is a vector of ones and the arguments of the predictive covariance matrix are all produced by our estimation algorithm (see the Technical Appendix for definitions). We also here and below use notation where the portfolio return before transaction costs is

$$R_{p,t} = 1 + r_{p,t-1} = 1 + \left(1 - w_{t-1}^{'} \iota\right) r_f + w_{t-1}^{'} r_t.$$

In addition, we let $R_{p,t}^{TC}$ denote period-t gross return after transaction costs, τ. Our specification of the portfolio allocation problem takes into account proportional transaction costs, τ, ex ante (i.e., at the time of the portfolio construction).[12] Following Della Corte and Tsiakas (2012), we set $\tau = 0.0008$. For our main results, we choose $\sigma_p^* = 10\%$ as target portfolio volatility of the conditional portfolio returns.

5.2 Evaluation of economic utility

5.2.1 Quadratic utility

Our econometric model provides forecasts of the mean vector of returns and the covariance matrix. To assess the economic utility of the forecasts, we employ the method proposed by West, Edison, and Cho (1993). In a mean-variance framework with quadratic utility, we can express the investor's realized utility in period t as

[12] Maurer and Pezzo (2018) show the importance of treating transaction costs in FX portfolios ex ante rather than ex post. Doing so avoids unnecessary trading and reduces transaction costs.

$$U(W_t) = W_t - \frac{\rho}{2} W_t^2 = W_{t-1} R_{p,t} - \frac{\rho W_{t-1}^2}{2} (R_{p,t})^2 ,$$

where W_t is the investor's wealth in t, ρ determines their risk preferences.

The investor's degree of relative risk aversion $\theta_t = \frac{\rho W_t}{1-\rho W_t}$ is set to a constant value θ. We choose $\theta = 2$ for our main results (and $\theta = 6$ for robustness checks). Then, the average realized utility, $U(\cdot)$, can be employed to consistently estimate the expected utility achieved by a given level of initial wealth (West, Edison, and Cho, 1993). With initial wealth W_0, the average utility for an investor can be expressed as

$$\overline{U}(\cdot) = W_0 \left\{ \sum_{t=0}^{T-1} R_{p,t+1}^{TC} - \frac{\theta}{2(1+\theta)} \left(R_{p,t+1}^{TC} \right)^2 \right\} . \tag{4}$$

The advantage of the representation above is that, for a fixed value of θ, the relative risk aversion is constant and utility is linearly homogenous in wealth. In contrast, for standard quadratic utility without restrictions on θ, relative risk aversion would be increasing in wealth, which is not likely to represent a typical investor's preferences. Here, having constant relative risk aversion, we can set $W_0 = \$1$.

5.2.2 Performance measures

Our main evaluation criterion is based on the dynamic mean-variance framework and quadratic utility. Comparing two competing forecasting models involves comparing the average utilities generated by the respective forecasting models. We assess the economic value of different forecasting approaches by equating the average utility generated by a portfolio strategy which is based on (a particular version of) the VAR approach and the average utility achieved by a portfolio strategy relying on a simple random walk. Φ is the the maximum (monthly) performance fee an investor is willing to pay to switch from the random walk to the specific VAR configuration. The estimated value of Φ ensures that the following equation holds:

$$\sum_{t=0}^{T-1}\left\{\left(R_{p,t+1}^{TC,*}-\Phi^{TC}\right)-\frac{\theta}{2\left(1+\theta\right)}\left(R_{p,t+1}^{TC,*}-\Phi^{TC}\right)^{2}\right\}=\sum_{t=0}^{T-1}\left\{R_{p,t+1}^{TC}-\frac{\theta}{2\left(1+\theta\right)}\left(R_{p,t+1}^{TC}\right)^{2}\right\}, \tag{5}$$

where $R_{p,t+1}^{TC,*}$ is the gross portfolio return constructed using the expected return and covariance forecasts from the dynamically selected best model configuration and $R_{p,t+1}^{TC}$ is implied by the benchmark random walk (without drift) model. The superscript TC indicates that all quantities are computed after adjusting for transaction costs.

As a second measure of economic utility, we report the Sharpe ratio. Despite its popularity as a risk measure, it is well known that the Sharpe ratio comes with a few drawbacks in the context of evaluating dynamic portfolio strategies (see, for example, Marquering and Verbeek (2004) or Han (2006)). This is why we primarily rely on performance fees as an evaluation criterion, while Sharpe ratios are reported as a complementary measure.

6 Empirical results

6.1 Evidence on model switching and sparsity

When using the long sample, our most flexible approach allows for dynamic model learning over a set of 32 different VAR models and ten different values of α using the methods described in Section 4. Using the short sample, dynamic model learning involves 512 different VAR models and ten different values of α. When all the VARs involve only exchange rates, we use the label DML in the tables and discussion below. We also consider many special cases of VAR as noted in the table. For instance, DML ($\alpha = 0.99$) means that α is fixed to 0.99 rather than being selected from a grid of values. "DML without cross lags" means that the coefficients on the cross lags are set to zero in all VARs. We implement such restrictions through restricting our vector of shrinkage parameters, $\gamma_1, ..., \gamma_7$. For instance, to delete the cross lags we set γ_3, the shrinkage parameter on the cross lags, to be zero.

For the short sample, which includes many potential predictors, we use the term "DML with ALL REGRESSORS" to denote the case where DML is being done over all specification choices including all of the exogenous predictors. We also consider several restricted versions of DML which involves dynamic model learning over only some of the predictors. DML with OIL, for example, means that OIL is the only possible exogenous variable variable which could be chosen.

The main focus of this paper is on how well these specifications perform in terms of our dynamic asset allocation problem. However, before doing this, we present a few results illustrating how the dynamic model learning strategy is working using the most flexible specification.

Dynamic model learning is to be preferred over static Bayesian model learning only if the optimal forecasting model is changing over time. Figure 1 shows that it does so in our application when using the long sample. This result is similar when using the short sample, but the figure for this case is left out for the sake of brevity. TFrequency of model change for DML-UIP. The vertical axis represents the model configurations $1, ..., 512$. The red line depicts the evolution of the selected model configuration for $\alpha = 1$. The grey line shows the evolution of the selected model configuration when is dynamically chosen from the grid of values $\alpha \in \{0.2; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 0.95; 1$

Figures 2 and 3 show which blocks of variables are included at each point in time for the long sample and short sample, respectively. A key point to note is that sparsity is empirically warranted. Using the long sample, 54.94% of the time, neither own lags nor cross lags are chosen and thus the selected model collapses temporarily to a multivariate random walk 54.94% of the time. For the shorter sample, the random walk is selected in 30.16% of the time. It is apparent from both figures that in most cases, if a variable of block of variables is selected, it tends to remain selected for a few time periods.

frequency_model_change-eps-converted-to.pdf

Figure 1: Frequency of model change.

The figure displays the frequency of model change over time using the long sample. The vertical axis represents the model configurations $1, \ldots, 32$. The red line depicts the evolution of the selected model configuration for $\alpha = 1$. The grey line shows the evolution of the selected model configuration when is dynamically chosen from the grid of values $\alpha \in \{0.2, 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 0.95; 0.99; 1\}$.

Figure 2: Inclusion of (blocks of) variables in the long sample.

The figure displays which blocks of variables are included at each point in time using the long sample. "Included" means the respective γ_i is not 0.

Figure 3: Inclusion of (blocks of) variables in the short sample.

The figure displays which blocks of variables are included at each point in time using the short sample. "Included" means the respective γ_i is not 0.

6.2 Out-of-sample forecast evaluation

The previous sub-section established that our DML approach was picking up model change. But the key issue is whether this is important for dynamic portfolio selection and forecast performance. In this sub-section we report various performance measures using three economic criteria and one statistical criterion: the performance fee after transaction costs (Φ^{TC}), [13] the Sharpe ratio before and after transactions costs (SR and SR^{TC}) as well as the average joint predictive log likelihood (PLL). The latter measures the accuracy of the density forecasts. The economic criteria are benchmarked relative to a random walk without drift and constant volatility. Hence, the performance fee Φ^{TC} is the annualized fee a risk-averse investor (with risk aversion $\theta = 2$) is willing to pay for switching from a dynamic portfolio strategy based on the multivariate random walk model with constant error covariance matrix

[13] As an alternative performance measure we also investigated the manipulation-proof performance measure proposed by Goetzmann, Ingersoll, Spiegel, and Welch (2007). The advantage of this criterion is that we neither have to assume a particular return distribution nor a certain utility function. The results compared to the reported quadratic utility case are very similar and available upon request.

compared to one that uses a more flexible forecasting strategy.

In addition to DML, we consider a range of restricted versions thereof so as to investigate which aspects of our approach are most important. That is, we can disentangle whether exclusion of certain sets of variables or individual variables or the dynamics of the model selection procedure are the most important features. Table 1 and Table 2 contain our main results.

Table 1: Evaluation of forecasting results (long sample).

The table summarizes the economic and statistical evaluation of our forecasts from the DML and restricted versions thereof for the period from 1990 : 01 to 2016 : 12. We measure statistical significance for differences in performance fees and log scores using the (one-sided) Diebold and Mariano (1995) t-test using heteroskedasticity and autocorrelation robust (HAC) standard errors. We evaluate whether the Sharpe ratio of a model is different from that of the random walk (with constant volatility) benchmark using the (one-sided version of the) Ledoit and Wolf (2008) bootstrap test. We compute the Ledoit and Wolf (2008) test statistics with a serial correlation-robust variance, using the pre-whitened quadratic spectral estimator of Andrews and Monahan (1992). One star indicates significance at 10% level; two stars significance at 5% level; and three stars significance at 1% level.

	Φ^{TC}	SR	SR^{TC}	PLL
DML	485**	1.08**	0.92**	22.05
Type of restrictions: VAR lags				
DML without own lags	365*	0.82*	0.72*	21.86
DML without cross lags	278	0.80	0.66	21.78
Type of restrictions: Random walk				
Random walk (without drift) with time-varying (co-)variance	17	0.47	0.46	21.65
Type of restrictions: Model selection dynamics				
DML ($\alpha = 1$)	−255	0.35	0.19	21.65
DML ($\alpha = 0.99$)	−194	0.40	0.24	21.65
DML ($\alpha = 0.95$)	60	0.59	0.45	21.68
DML ($\alpha = 0.90$)	238	0.78	0.65	21.88
DML ($\alpha = 0.80$)	485**	1.08**	0.92**	22.05
DML ($\alpha = 0.70$)	478*	1.07**	0.90*	22.06
DML ($\alpha = 0.60$)	486*	1.12**	0.93**	22.06
DML ($\alpha = 0.50$)	409*	1.04**	0.84**	22.05
DML ($\alpha = 0.40$)	276	0.93*	0.70	22.03
DML ($\alpha = 0.20$)	181	0.85	0.60	21.98

The DML approach can be seen to perform very well in terms of our economic performance indicators when using the long sample. The annualized performance fee after transaction costs is 485 basis points (bps) and the annualized Sharpe ratio is 1.08 before transaction costs and 0.92 after transaction costs. These figures are the highest or nearly highest of any in Table 1 and are substantially better than most alternatives. With the

Table 2: Evaluation of forecasting results (short sample).

See notes to Table 1.

	Φ^{TC}	SR	SR^{TC}	PLL
DML	327	1.01*	0.82*	22.02*
With regressors				
DML with OIL	199	0.89	0.70	22.03*
DML with UIP	464*	1.12**	0.93**	22.01*
DML with INT_DIFF	388*	1.06*	0.88*	22.02*
DML with STOCK_GROWTH	368*	1.06*	0.88*	22.06*
DML with ALL REGRESSORS	397*	1.02*	0.87*	22.04*
Type of restrictions: VAR lags				
DML without own lags	98	0.72	0.60	21.97
DML without cross lags	200	0.86	0.79	21.78*
Type of restrictions: Random walk				
RW without drift and time-varying (co-)variance	5	0.54	0.53	21.72*
Type of restrictions: Model selection dynamics				
DML ($\alpha = 1$)	−427	0.34	0.11	21.69
DML ($\alpha = 0.99$)	−464	0.28	0.08	21.66
DML ($\alpha = 0.95$)	−167	0.51	0.34	21.79*
DML ($\alpha = 0.90$)	98	0.77	0.60	21.96*
DML ($\alpha = 0.80$)	266	0.94	0.76	22.02*
DML ($\alpha = 0.70$)	327	1.01*	0.82*	22.02*
DML ($\alpha = 0.60$)	251	0.97	0.75	22.02*
DML ($\alpha = 0.50$)	84	0.82	0.60	21.98
DML ($\alpha = 0.40$)	−31	0.71	0.48	21.96
DML ($\alpha = 0.20$)	11	0.75	0.52	21.94

long sample, the economic performance improvements are statistically significant relative to the random walk with constant volatility benchmark. With the short sample, for the Sharpe ratios, we are also finding statistically significant improvements. When looking at the statistical performance, the joint predictive log likelihoods also shows that DML is among the best.

In the introduction to this paper, we emphasized that a large body of research on exchange rate forecasting found that nothing beats a random walk. In terms of pure forecast performance we achieve some gains in terms of density forecasting accuracy as indicated by the PLLs, however our results do not not greatly undermine this story in that we find the improvements in forecast performance relative to the random walk is not statistically significant. However, in terms of portfolio management we are finding large and statistically significant improvements relative to the random walk.

From a statistical perspective, it is worth noting that PLLs indicate that, if one were to select one single best model for the entire evaluation period, it might be the random walk with time-varying error covariance for the long sample. This has an value for the PLL of 21.72 which is nearly as high as the DML methods. But as we have seen, DML, which allows for the selected model to change over time, would choose this model roughly half the time. DML also leads to a much better economic performance. This demonstrates that our very flexible learning mechanism is able to efficiently switch between different model configurations in real time.

It is also worth noting that the joint predictive likelihood and the two economic criteria broadly agree with respect to the ranking of the approaches. [14]

Results using the short sample are broadly similar to those for the long sample. For the DML we find the annualized performance fee after transaction costs is 327 bps and the annualized Sharpe ratio is 1.01 before transaction costs and 0.82 after transaction costs. Including exogenous regressors can lead to very slight improvements in terms of the PLL, but much stronger improvements when using the economic evaluation criteria. For instance, the annualized performance fee after transactions costs increases to 397 when all the regressors are considered. Among the exogenous regressors, including UIP leads to the largest improvements in the economic performance measures. But, with the exception of OIL, the other regressors also lead to improvements. These patterns are in line with those in Figure 3. It is also worth noting that we are now finding DML to offer statistically significant improvements in all criteria (PLL as well as economic criteria) relative to a multivariate random walk.

For the sake of brevity, we do not report results for individual models (other than the multivariate random walk). But it is worth noting that, for the short sample, the best single model (out of the 512 considered) is the individual VAR with no intercept ($\gamma_1 = 0$), including

[14] This finding aligns with the results documented by Cenesizoglu and Timmermann (2012). They report broad agreement between density forecast measures and economic performance measures based on the entire predictive density. In contrast, they note that there is typically a weak link between point forecast evaluation criteria and economic evaluation criteria.

own lags ($\gamma_2 = 0.1$) but no cross lags ($\gamma_3 = 0$) and no additional regressors. This specification has an average PLL of 21.73. But the DML has a higher average PLL of 22.02. And the economic performance of the DML is substantially stronger than the random walk model, which again indicates the benefits of allowing for the forecasting model to change over time. Overall, the patterns for using the short and long samples are similar to one another.

We now turn to an investigation of which aspects are most important in leading to DML's good economic performance. An important finding is that the random walk without drift and a time-varying covariance provides poor forecasts in both VAR settings with the performance fee essentially dropping to zero. Importance of VAR lags is identified in both samples. Neglecting own lags or cross lags is detrimental for portfolio performance. Cross lags are relatively more important in the long than the short sample.[15].

We next delineate the effect of restrictions on the tuning parameter α. Previously, in Figure 1, we showed that the optimal model can rapidly change over time. The results in Tables 1 and 2 relating to α show the benefits of this for forecasting. Fixing $\alpha = 1$ rather than choosing the value of α in real time leads to very poor forecasting results in both the long and short samples. Allowing for lower values of α and, thus, more model switching leads to higher values of the log scores, and in particular, to higher performance fees and Sharpe ratios. In fact, the highest performance fee and Sharpe ratio is obtained when $\alpha = 0.80$ in the long sample and $\alpha = 0.70$ in the short sample. Thus, large economic and statistical losses occur if the investor does not emphasize the most recent forecast performance when selecting the forecasting model on which to base their asset allocation decision. Altogether, we are finding the choice of the discount factor α to be a very important one.

6.3 When does DML perform best?

Given the evidence that global factors such as volatility affect the returns of momentum and carry trade strategies (Menkhoff, Sarno, Schmeling, and Schrimpf, 2012; Bakshi and

[15]This finding is interesting in light of recent of evidence in Berg and Mark (2015) that third-country fundamentals improve the explanatory power of conventional fundamental exchange rate models.

Panayotov, 2013), it is interesting to see if these are also associated with DML's good performance. We use the short sample and the DML with ALL REGRESSORS specification. We analyze whether its performance (relative to the random walk) can be explained by shocks in FX volatility for the G10 countries measured by the corresponding J.P. Morgan index ($\Delta FXVOL$), common disagreement in FX markets for the G10 currencies ($FXDIS$) or shocks in volatility in equity markets (ΔVIX).[16] To this end, we regress differences in utility ($\Delta Utility$) on an intercept and these three additional regressors. Utility is defined in Section 5.2.1. We obtain the following fitted regression line (t-statistics are in parentheses):

$$\Delta\widehat{Utility} = \underset{(1.51)}{0.0109} + \underset{(3.26)}{0.0031}\Delta FXVOL \underset{(-0.20)}{-0.00005}\,\Delta VIX + \underset{(1.43)}{0.0037}FXDIS. \tag{6}$$

The results indicate that the outperformance of the flexible model is higher in times of positive FX volatility shocks. This finding is in contrast to pure carry trade strategies which typically perform poorly in times of high FX volatility (see, e.g., Menkhoff, Sarno, Schmeling, and Schrimpf (2012) or Burnside, Eichenbaum, Kleshchelski, and Rebelo (2010)). Against this background, our proposed strategy seems to provide high diversification benefits to FX style strategies such as pure carry. The outperformance of the flexible DML strategy also seems to be higher in times of increased disagreement among professional forecasters although this coefficient is insignificant at standard levels of significance. However, volatility shocks as measured by the VIX have no association with the change in utility.

6.4 Analysis of portfolio returns

In this subsection, we present some additional evidence relating to the portfolio constructed using DML.

Figure 4 compares the evolution of wealth for an investor who begins with one dollar and

[16] Disagreement is approximated by the recursively drawn first principal component of the one month-ahead exchange rate disagreement proxied by the absolute difference of the strongest and weakest forecast among professional survey participants.

relies on DML to the wealth of an investor who uses a multivariate random walk with constant covariance to construct their porfolio. As is evident from the figure, the outperformance of DML is large, with the most striking gains around the time of the subprime crisis.

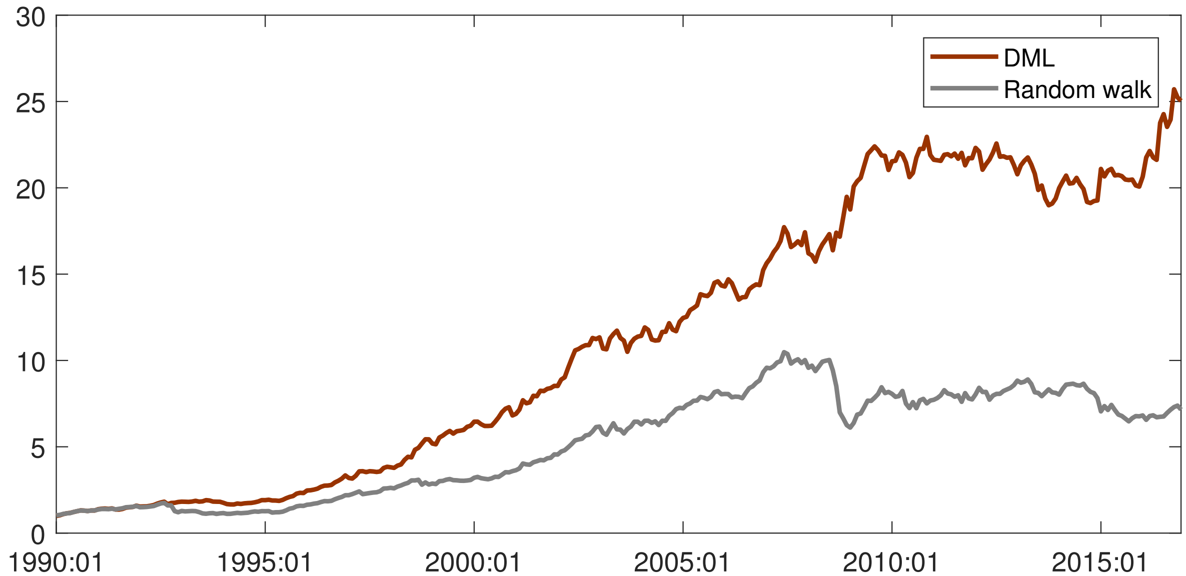

Figure 4: Evolution of wealth in the long sample.

DML does not restrict the weights associated with each currency when constructing the portfolio. However, some investment strategies do restrict portfolio weights. Table 3 presents the effects of some important portfolio restrictions on economic performance for the long and short samples. These are constructed using our DML methods, but impose different restrictions on the portfolio weights.

Forcing the portfolio weights for the US Dollar to zero (labelled Dollar-neutral strategy in the table) leads to a loss in economic utility and a lower Sharpe ratio. Although, in the long sample these losses are small. This shows that the success of the DML strategy is not mainly driven by movements of the US dollar. Restricting portfolio weights to a given interval might reduce portfolio turnover and lead to a decrease in transaction costs. In Table 3, we present results for various restrictions of this sort. Our results, however, do not support portfolio restrictions. While there is a moderate decrease (or even slight increase in the case of the short sample) in portfolio performance when imposing rather lenient restrictions $[-1; 1]$, imposing tighter restrictions is clearly detrimental for portfolio performance. Choosing equal weights leads to even worse forecast performance.[17]

Table 3: Restrictions on portfolio weights.
The table summarizes the effects of restrictions on portfolio weights on the economic performance.

	Long Sample			Short Sample		
	Φ^{TC}	SR	SR^{TC}	Φ^{TC}	SR	SR^{TC}
Dollar-neutral	335	0.96*	0.78*	105	0.81	0.61
Without weight restrictions	485**	1.08**	0.92**	327	1.01*	0.82*
Weight restriction: $[-1; 1]$	345*	0.96**	0.81**	393	1.05	0.86
Weight restriction:$[-0.5; 0.5]$	346*	0.96**	0.83**	150	0.84	0.68
Weight restriction: $[-0.25; 0.25]$	203	0.83**	0.74**	−27	0.65	0.54
Weight restriction:$[-0.1; 0.1]$	−93	0.67	0.60	−323	0.32	0.25
Equal weights	−383	0.01	0.00	−552	−0.11	−0.11

Table 4 presents some characteristics of the monthly portfolio returns produced using DML. Multiplying the volatility figures in the table by $\sqrt{12}$ produces annualized volatilities of 10.36% and 10.29% for the long and short samples, respectively. These are very close to the annual portfolio target volatility of 10%. For the investor, another attractive feature is that there is little skewness of the returns. Moreover, the first-order autocorrelation of returns and squared returns are near zero, indicating that there is not much information left to be exploited from the original time series. In particular, the fact that the autocorrelation of squared portfolio returns is near zero points to fact that our modelling approach to time-varying volatilities has worked well. The fact that returns are uncorrelated with stock returns (S&P500) makes our FX strategy a strong diversifier to equity portfolios. Overall, the characteristics of the portfolio returns are highly attractive for risk management purposes.

[17] The finding that equal weights are outperformed by mean-variance optimization in FX markets aligns with the results of Ackermann, Pohl, and Schmedders (2016).

Table 4: Statistics of the portfolio returns.

The table summarizes characteristics of the monthly portfolio returns for both the long sample and the short sample over their respective forecasting evaluation period.

	Long Sample	Short Sample
Mean return (in %)	1.16	1.04
Volatility (in %)	2.99	2.96
Skewness	0.07	−0.23
Kurtosis	3.37	3.16
Positive returns (> 0 in %)	64	65
First-order autocorrelation of returns	0.08	0.10
First-order autocorrelation of squared returns	−0.03	−0.04
Correlation to S&P500 returns	−0.04	−0.01

6.5 Further Empirical Work and Robustness Checks

The online Empirical Appendix contains more empirical evidence that reinforces the story told above. That is, as an econometric approach DML is performing well and, if used to construct an investment portfolio, would yield higher levels of utility than a simple benchmark.

The Empirical Appendix shows that the coverage of our predictive densities is good and carries out a variety of robustness checks. The latter include: i) a prior sensitivity analysis, ii) a use of different sets of predictors, iii) a use of different degrees of risk aversion and iv) an investigation of the effect of making our exogenous predictors into endogenous variables in the VAR. In all cases, we find the modelling approach taken in this paper to be robust and better than other plausible specification or prior choices. The Empirical Appendix also presents evidence against allowing for time-variation in the VAR coefficients which justifies our choice of $\lambda = 1$ in this paper.

7 Concluding remarks

This paper has proposed a new multivariate forecasting approach for exchange rate returns. Our dynamic Bayesian learning approach enables us to quickly detect model changes over time and achieves computational feasibility by using decay factors. A major conceptual advantage of our approach over univariate models is that we obtain the input for the inherently multivariate portfolio optimization problem in a natural manner without having to rely on additional assumptions and procedures for mapping the forecast output into portfolio weights.

We have evaluated the economic value of our exchange rate forecasts in a dynamic asset allocation framework. Relying on our forecasting method, an investor achieves sizable utility gains by capturing short-lived predictability and taking into account several dimensions of uncertainty. We find strong evidence for sparsity and fast model switching. We also find that third-currency effects matter in the sense that economics gains materialize through exploiting the cross-section of exchange rates. Sparsity and the frequent model shifts align with the implications of the theoretical and empirical exchange rate literature.

References

ABBATE, A., AND M. MARCELLINO (2018): "Point, interval and density forecasts of exchange rates with time varying parameter models," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(1), 155–179.

ABHYANKAR, A., L. SARNO, AND G. VALENTE (2005): "Exchange rates and fundamentals: evidence on the economic value of predictability," *Journal of International Economics*, 66(2), 325–348.

ACKERMANN, F., W. POHL, AND K. SCHMEDDERS (2016): "Optimal and naive diversification in currency markets," *Management Science*, 63(10), 3347–3360.

ANDREWS, D. W., AND J. C. MONAHAN (1992): "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica: Journal of the Econometric Society*, pp. 953–966.

BACCHETTA, P., AND E. VAN WINCOOP (2004): "A scapegoat model of exchange-rate fluctuations," *American Economic Review*, 94(2), 114–118.

——— (2006): "Can information heterogeneity explain the exchange rate determination puzzle?," *American Economic Review*, 96(3), 552–576.

——— (2013): "On the unstable relationship between exchange rates and macroeconomic fundamentals," *Journal of International Economics*, 91(1), 18–26.

BAKSHI, G., AND G. PANAYOTOV (2013): "Predictability of currency carry trades and asset pricing implications," *Journal of Financial Economics*, 110(1), 139–163.

BAŃBURA, M., D. GIANNONE, AND L. REICHLIN (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25(1), 71–92.

BARBERIS, N. (2000): "Investing for the long run when returns are predictable," *The Journal of Finance*, 55(1), 225–264.

Beckmann, J., and R. Schüssler (2016): "Forecasting exchange rates under parameter and model uncertainty," *Journal of International Money and Finance*, 60, 267–288.

Berg, K. A., and N. C. Mark (2015): "Third-country effects on the exchange rate," *Journal of International Economics*, 96(2), 227–243.

Berge, T. J. (2014): "Forecasting disconnected exchange rates," *Journal of Applied Econometrics*, 29(5), 713–735.

Burnside, C., M. Eichenbaum, I. Kleshchelski, and S. Rebelo (2010): "Do peso problems explain the returns to the carry trade?," *The Review of Financial Studies*, 24(3), 853–891.

Byrne, J. P., D. Korobilis, and P. J. Ribeiro (2016): "Exchange rate predictability in a changing world," *Journal of International Money and Finance*, 62, 1–24.

Canova, F. (1993): "Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model," *Journal of Economic Dynamics and Control*, 17(1-2), 233–261.

Carriero, A., G. Kapetanios, and M. Marcellino (2009): "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, 25(2), 400–417.

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

Chan, J. C., and E. Eisenstat (2018): "Bayesian model comparison for time-varying parameter VARs with stochastic volatility," *Journal of Applied Econometrics*.

Chib, S., F. Nardari, and N. Shephard (2006): "Analysis of high dimensional multivariate stochastic volatility models," *Journal of Econometrics*, 134(2), 341–371.

Dangl, T., and M. Halling (2012): "Predictive regressions with time-varying coefficients," *Journal of Financial Economics*, 106(1), 157–181.

Della Corte, P., L. Sarno, and I. Tsiakas (2008): "An economic evaluation of empirical exchange rate models," *The Review of Financial Studies*, 22(9), 3491–3530.

Della Corte, P., and I. Tsiakas (2012): "Statistical and economic methods for evaluating exchange rate predictability," *Handbook of exchange rates*, pp. 221–263.

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

Geweke, J., and G. Amisano (2012): "Prediction with misspecified models," *American Economic Review*, 102(3), 482–486.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior selection for vector autoregressions," *The Review of Economics and Statistics*, 97(2), 436–451.

Goetzmann, W., J. Ingersoll, M. Spiegel, and I. Welch (2007): "Portfolio performance manipulation and manipulation-proof performance measures," *The Review of Financial Studies*, 20(5), 1503–1546.

Han, Y. (2006): "Asset allocation with a high dimensional latent factor stochastic volatility model," *Review of Financial Studies*, 19(1), 237–271.

J.P. Morgan/Reuters (1996): "RiskMetrics™technical document," Discussion paper.

Koop, G., and D. Korobilis (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177(2), 185–198.

Kouwenberg, R., A. Markiewicz, R. Verhoeks, and R. C. J. Zwinkels (2017): "Model uncertainty and exchange rate forecasting," *Journal of Financial and Quantitative Analysis*, 52(01), 341–363.

Ledoit, O., and M. Wolf (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15(5), 850–859.

Li, J., I. Tsiakas, and W. Wang (2015): "Predicting exchange rates out of sample: Can economic fundamentals beat the random walk?," *Journal of Financial Econometrics*, 13(2), 293–341.

Lizardo, R. A., and A. V. Mollick (2010): "Oil price fluctuations and US dollar exchange rates," *Energy Economics*, 32(2), 399–408.

Markiewicz, A. (2012): "Model uncertainty and exchange rate volatility," *International Economic Review*, 53(3), 815–844.

Marquering, W., and M. Verbeek (2004): "The economic value of predicting stock index returns and volatility," *Journal of Financial and Quantitative Analysis*, 39(02), 407–429.

Maurer, T. A., and L. Pezzo (2018): "Importance of transaction costs for asset allocations in FX markets," *Available at SSRN: https://ssrn.com/abstract=3143970.*

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), 3–24.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012): "Carry trades and global foreign exchange volatility," *The Journal of Finance*, 67(2), 681–718.

Rossi, B. (2006): "Are exchange rates really random walks? Some evidence robust to parameter instability," *Macroeconomic Dynamics*, 10(1), 20–38.

——— (2013): "Exchange rate predictability," *Journal of Economic Literature*, 51(4), 1063–1119.

Rossi, B., and T. Sekhposyan (2011): "Understanding models' forecasting performance," *Journal of Econometrics*, 164(1), 158–172.

Sarno, L. (2005): "Viewpoint: Towards a solution to the puzzles in exchange rate economics: where do we stand?," *Canadian Journal of Economics*, 38(3), 673–708.

SARNO, L., AND G. VALENTE (2009): "Exchange rates and fundamentals: Footloose or evolving relationship?," *Journal of the European Economic Association*, 7(4), 786–830.

STOCK, J. H., AND M. W. WATSON (2007): "Why has U.S. inflation become harder to forecast?," *Journal of Money, Credit and Banking*, 39(s1), 3–33.

TRIANTAFYLLOPOULOS, K. (2011): "Time-varying vector autoregressive models with stochastic volatility," *Journal of Applied Statistics*, 38(2), 369–382.

WEST, K. D., H. J. EDISON, AND D. CHO (1993): "A utility-based comparison of some models of exchange rate volatility," *Journal of International Economics*, 35(1-2), 23–45.

WEST, M., AND J. HARRISON (1997): *Bayesian forecasting and dynamic models*. Springer, 2nd edn.

WRIGHT, J. H. (2008): "Bayesian Model Averaging and exchange rate forecasts," *Journal of Econometrics*, 146(2), 329–341.