Online appendix to *"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 Glasgow

Rainer Alexander Schüssler§
Technical University of Dortmund
University of Rostock

March 3, 2020

Abstract

This online supplementary appendix presents technical details of our proposed econometric methodology, simulation results not included in the paper, and additional results using both our benchmark data set and a data set with more time series observations.

Contents

*Department of Economics, University of Greifswald, Friedrich-Loeffler-Straße 70, 17489 Greifswald. E: joscha.beckmann@uni-greifswald.de

†Department of Economics, University of Strathclyde, 199 Cathedral Street, Glasgow, G4 0QU. E: gary.koop@strath.ac.uk

‡Adam Smith Business School, University of Glasgow, Glasgow G12 8QQ. E: dikorobilis@googlemail.com

§Department of Econometrics and Statistics, TU Dortmund, Vogelpothsweg 87, 44227 Dortmund. E: raineralexanderschuessler@gmail.com

1 Technical Appendix

1.1 Filtering

In this sub-section, we provide econometric details of our (TVP)-VARs. Filtered estimates can be obtained using the fact that the form of the state space model implies

$$\beta_t | y^{t-1}, \Sigma_{t-1} \sim N\left(\beta_{t|t-1}, \Omega_{t|t-1}\right),$$

where $t|t-1$ subscripts refer to estimates made of time-t quantities given information available at time $t-1$. Forecasts can be obtained using the fact that the predictive density is multivariate t:

$$y_t | y^{t-1} \sim t\left(\widehat{y}_{t|t-1}, x_t \Omega_{t|t-1} x_t^{'} + Q_{t|t-1}\right),$$

where $\widehat{y}_{t|t-1} = x_t \beta_{t|t-1}$. Standard Kalman filtering and Wishart matrix discounting formulas can be used to produce the quantities $\beta_{t|t-1}, \Omega_{t|t-1}$ and $Q_{t|t-1}$ as follows.

Predictive step

The Kalman filter provides, beginning with $\beta_{0|0} = 0$ (see below), simple updating formulas for producing $\beta_{t|t-1}$ and $\beta_{t|t}$ for $t = 1, ..., T$ which are standard and will not be reproduced here. Given these we can produce point forecasts as:

$$\widehat{y}_{t|t-1} = x_t \beta_{t|t-1}.$$

To produce $\Omega_{t|t-1}$ we use a discount factor approximation involving a discount factor λ and update as

$$\Omega_{t|t-1} = \frac{1}{\lambda} \Omega_{t-1|t-1}.$$

Note that such an approximation is well established (West and Harrison, 1997) and, for example, used by Koop and Korobilis (2013).

We select the discount factors in a data-adaptive fashion in real time. If $\lambda < 1$, the VAR coefficients are time varying and a lower value of λ is associated with more rapidly changing coefficients. If $\lambda = 1$, the special case of constant coefficients is obtained. An advantage of the discount factor approach is that we do not have to update the entire covariance matrix but instead only have to choose a single discount factor.

To retain conjugacy, Σ_t is modelled as Inverse Wishart (IW) with δn_{t-1} degrees of freedom and scale matrix S_{t-1},

$$\Sigma_{t|t-1} \sim IW(\delta n_{t-1}, S_{t-1}),$$

with the expected value

$$E\left(\Sigma_{t|t-1}\right) := Q_{t|t-1} = \frac{S_{t-1}}{\delta n_{t-1} + M - 1}.$$

Note that this density reflects the uncertainty about Σ_t and thus accounts for parameter uncertainty. Low values of δ are associated with increasingly rapid changes in the covariance matrix. Values near one are associated with slow adaptation, while $\delta = 1$ represents the case of a constant covariance matrix Σ.

Update step

The error e_t is obtained as the difference between the point forecast $\widehat{y}_{t|t-1}$ and the actual observation y_t

$$e_t = y_t - \widehat{y}_{t|t-1}.$$

The observational covariance matrix is updated as

$$\Sigma_{t|t} \sim IW(n_t, S_t)$$

with the scale

$$S_t = \left(k^{-1} S_{t-1} + e_t e_t^{'}\right)\left(I_M + F_t \Omega_{t|t-1} x_t^{'}\right),$$

where

$$k^{-1} = \frac{\delta\left(1 - M\right) + M}{\delta\left(2 - M\right) + M - 1},$$

using approximation results by Triantafyllopoulos (2011) exploiting the expectation invariance of the random walk process for $\Sigma_t : E\left(\Sigma_{t|t-1}\right) = E\left(\Sigma_{t-1|t-1}\right)$. As is common in the literature, the scale matrix is initialized as

$$S_0 = \begin{bmatrix} \widehat{u}_1^2 & & \\ & \ldots & \\ & & \widehat{u}_M^2 \end{bmatrix},$$

where $\widehat{u}_i^2, ..., \widehat{u}_M^2$ are the residuals from OLS estimation of a VAR over an initial training sample. The updated degrees of freedom are obtained as

$$n_t = \delta n_{t-1} + 1 \quad (n_t \rightarrow n = \frac{1}{1-\delta}).$$

It is a natural choice to initialize the degrees of freedom with

$$n_0 = \frac{1}{1-\delta}.$$

The expected observational covariance is obtained as

$$E\left(\Sigma_{t|t}\right) := Q_{t|t} = \frac{S_t}{n_t + M - 1}.$$

The time-t Kalman gain (KG_t) is obtained as

$$KG_t = \left(\Omega_{t|t-1} x_t'\right) \left(x_t \Omega_{t|t-1} x_t' + Q_{t|t}\right)^{-1}.$$

Given the Kalman gain, the coefficients and the system covariance are updated as

$$\beta_{t|t} = \beta_{t|t-1} + KG_t e_t$$

and

$$\Omega_{t|t} = \Omega_{t|t-1} + KG_t x_t \Omega_{t|t-1}.$$

1.2 Spike-and-slab interpretation of the prior

Here we provide an interpretation of our proposed prior structure in the main paper as a spike-and-slab prior. This is meant as an illustration of our prior structure from a different angle. Note that the notation introduced in this sub-section only applies locally and is not used elsewhere in the main text or the online appendix. Our starting point is the same type of time-varying parameter VAR with exogenous variables we consider in the main paper:

$$\begin{aligned} y_t &= x_t \beta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_t) \\ \beta_{t+1} &= \beta_t + u_t, \quad u_t \sim N(0, \Omega_t). \end{aligned}$$

As in the main paper, 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. Thus, we have, $k = M(1 + p \cdot M + N_x + N_{xx})$ elements in β_t.

The initial conditions for the time-varying VAR coefficients can be viewed as time $t = 0$ priors for the parameters β_t. For each coefficient in VAR equation i, $i = 1, ..., M$, and lag/predictor j, $j = 1, ..., k/M$, we use a variable selection prior of the form

$$\begin{aligned} \beta_{0,i,j} &\sim k_{i,j,t} N(0, V_{i,j}) + (1 - k_{i,j,t})\delta_0 \\ k_{i,j,t} &\sim DML, \end{aligned}$$

where δ_0 denotes the Dirac delta which assigns point mass at zero, and DML denotes dynamic model learning. Each indicator variable $k_{i,j,t}$ can take on a value of zero or one in each time period. When $k_{i,j,t} = 1$ the prior for $\beta_{0,i,j}$ is $N(0, V_{i,j})$ and when $k_{i,j,t} = 0$ the coefficient is exactly zero (and, hence, covariate j does not enter VAR equation i). Whether $k_{i,j,t}$ is one or zero is decided probabilistically via the DML procedure.[1] We make the time-dependency of the ks explicit here, using subscript t. To streamline notation, we do not use time-subscripts for the γs in the main text, although they are re-selected each period.

We choose $V_{i,j}$, which contains the prior variances for the included coefficients, using ideas from the Minnesota prior:

$$V_{i,j} = s_i^2 \times \begin{cases} \gamma_1 & \text{for intercepts} \\ \frac{\gamma_2}{r^2 s_i^2} & \text{for coefficients on lag } r \text{ of variable } i \text{ (own lag)} \\ \frac{\gamma_3}{r^2 s_k^2} & \text{for coefficients on lag } r \text{ of endogenous variable } k,\ k \neq i \\ \gamma_{(3+l)} & \text{for coefficients on the } l^{th} \text{ asset-specific exogenous variable} \\ \gamma_{(N_x+3+m)} & \text{for coefficients on the } m^{th} \text{ non asset-specific exogenous variable} \end{cases}$$

where $r = 1, ..., p$ indexes lag-length, $k = 1, ..., M$ indexes VAR equations, $l = 1, ..., N_x$

[1] It would be possible to treat $k_{i,j,t}$ as unknown parameters and include them in the Bayesian posterior. But these parameters are time-varying and directly drawing from them in an MCMC algorithm would be computationally burdensome. This motivates our use of DML which uses discounting methods to produce a computationally feasible approach. It is also worth noting that the selection indicators are updated online. That is, as new data becomes available the investor only needs to input the latest observation to update from $k_{i,j,t}$ to $k_{i,j,t+1}$.

indexes asset-specific predictors, $m = 1, ..., N_{xx}$ indexes non asset-specific exogenous predictors, while s_i^2 denotes the OLS estimate of the residual variance of a univariate AR(p) for variable i.

1.3 Dynamic asset allocation and evaluation of economic utility

1.3.1 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 $E_{t-1}(r_{i,t}) = int_{i,t-1} + y_{i,t}$.[2] 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 foreign bonds (with $1 - \sum_{i=1}^{9} w_{i,t}$ being the weight attached to the

[2] 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.

domestic bond):

$$\max_{w_t} \left\{ \mu_{p,t|t-1} = w_t^{'}\mu_{t|t-1} + (1 - w_t'\iota) r_f - \tau \left(\iota' \left| w_t - w_{t-1} \circ \frac{1 + r_t}{1 + r_{p,t}} \right| \right) \right\}$$

subject to

$$\left(\sigma_p^*\right)^2 = \underbrace{w_t' \frac{\delta n_{t-1}}{\delta n_{t-1} - 2} \left(x_{t-1} \Omega_{t|t-1} x_{t-1}^{'} + Q_{t|t-1} \right)}_{\text{estimate of the predictive covariance matrix}} w_t,$$

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 1.1 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).[3] 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.

1.3.2 Evaluation of economic utility

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

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

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,

[3] 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.

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

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

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 $W_0 = \$1$.

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:

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

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.

1.4 Fundamental exchange rate models

This section defines the fundamental exchange rate models which are used in the paper. One of these (UIP) is used in the main results in the body of th paper. The remainder are used in this online appendix.

1.4.1 Fama regression/UIP

The UIP condition is the fundamental parity condition for foreign exchange market efficiency under risk neutrality. This condition postulates that the difference in interest rates between two countries should equal the expected change in exchange rates between the countries' currencies (Engel, 2013):

$$E_t \Delta s_{t+1} = int_t - int_t^*,$$

where $\Delta s_{t+1} \equiv s_{t+1} - s_t$. $E_t \Delta s_{t+1}$ denotes the expected change (at time t for $t+1$) of log exchange rates, denominated as US dollar per foreign currency. int_t (int_t^*) is the one-period nominal interest rate US (foreign) securities. The following forecasting equation arises under the assumption that $E_t \Delta s_{t+1}$ equals Δs_{t+1}, where s_t denotes the log of realized exchange rates:

$$\Delta s_{t+1} = int_t - int_t^*.$$

We use $int_t - int_t^*$ as a predictor.

1.4.2 Purchasing power parity

Throughout the PPP literature, the real exchange rate is usually modelled as

$$q_t = s_t - p_t + p_t^*,$$

where q_t is the log of the real exchange rate and p_t (p_t^*) are the logs of the US (foreign) price levels (Rogoff, 1996). PPP postulates a constant real exchange rate, resulting in the price differential as the fundamental nominal exchange rate:

$$f_{PPP} = (p_t - p_t^*)$$

and rely on current deviations from this exchange rate as a predictor for Δs_{t+1}, that is, if PPP holds, we expect that $\Delta s_{t+1} = (f_{PPP} - s_t)$ holds. Thus, we use $f_{PPP} - s_t$ as a predictor.

1.4.3 Monetary fundamentals

The main feature of the monetary approach is that the exchange rate between two countries is determined via the relative development of money supply and industrial production (Dornbusch, 1976; Bilson, 1978). The underlying idea is that an increase in the relative money supply depreciates the US dollar, while the opposite holds for relative industrial production. A simplified version of the monetary approach adopted in previous studies (Mark and Sul, 2001) can be expressed as

$$f_{MON} = (m_t - m_t^*) - (ip_t - ip_t^*),$$

where $m_t - m_t^*$ denotes the (log) money supply and $ip_t - ip_t^*$ refers to (log) industrial production differentials. This implies $\Delta s_{t+1} = f_{MON} - s_t$ and we use $f_{MON} - s_t$ as a predictor.

1.4.4 Taylor rule fundamentals

The Taylor rule states that a central bank adjusts the short-run nominal interest rate in order to respond to inflation (π) and the output gap (ou). Postulating such Taylor rules for two countries and subtracting one from the other, an equation is derived with the interest rate differential on the left-hand side and the inflation and output gap on the right-hand side.[4] Provided that at least one of the two central banks also targets the PPP level of the exchange rate, the real exchange rate also appears on the right-hand side of the equation. The underlying idea is that both central banks follow a Taylor-rule

[4]The output gap is approximated as the deviation of industrial production from trend output which is calculated based on the Hodrick-Prescott filter with smoothing parameter $\lambda = 14,400$. For estimating the Hodrick-Prescott trend out of sample, we only use data that would have been available at the given point in time.

model and determine the interest rate differential which drives the exchange rate. We rely on a simple baseline specification with ad-hoc weights for inflation and output gap which also incorporates the real exchange rate:

$$\Delta s_{t+1} = 1.5(\pi_t - \pi_t^*) + 0.1(ou_t - ou_t^*) + 0.1q_t.$$

We use $1.5(\pi_t - \pi_t^*) + 0.1(ou_t - ou_t^*) + 0.1q_t$ as a predictor.

2 Data Appendix

Variable	Source
Consumer prices, seasonally adjusted	OECD
End-of-month dollar exchange rates	Datastream
Industrial production and GDP, seasonally adjusted	OCED
Money supply, seasonally adjusted	OCED
LIBOR and Eurodeposit interest rates	Datastream
FXDIS: Disagreement among exchange rate forecaster measured by standard deviation	Consensus Economics
10 Year Government Bonds	Datastream
CBOE Volatility Index (VIX)	Federal Reserve
FXVOL: J.P. Morgan G10 currency volatility index	Bloomberg
West Texas Intermediate Oil Price, denominated in US Dollar	Federal Reserve

3 Simulation experiment: model incompleteness

Model incompleteness refers to the situation that all of the entertained individual forecasting models are allowed to be false (Geweke and Amisano, 2011). Given the complexity of exchange rate dynamics, we do not assume that any of our entertained individual forecasting models reflects the true DGP. Rather, we consider a dynamic model learning mechanism that switches among differently specified model configurations to approximate the true DGP as closely as possible and acknowledge that the DGP changes through time. Our implementation of the model learning strategy involves only one parameter (α). It is thus parsimonious, limiting concerns about estimation error and can be interpreted as a shortcut to approximate complex nonlinear behavior in a timely manner. In this small simulation experiment we focus on the important aspect of our model learning strategy to timely detect the most appropriate DGP among a set of models in an incomplete model setting and changing DGPs.

We use a simulation setup considered in Billio, Casarin, Ravazzolo, and van Dijk (2013) and generate a random sample from the following autoregressive model with breaks:

$$y_t = 0.1 + 0.3\mathbb{I}_{(T_0,T]}(t) + \left(0.6 - 0.4\mathbb{I}_{(T_0,T]}(t)\right) y_{t-1} + \varepsilon_t,$$

for $t = 1, ..., T$, with $\varepsilon_t \sim N(0, \sigma^2), \sigma = 0.05, T_0 = 50$ and $T = 100$. $\mathbb{I}_{(z]}(A)$ takes the value 1 if $z \in A$ and 0 otherwise. $y_0 = 0.25$.

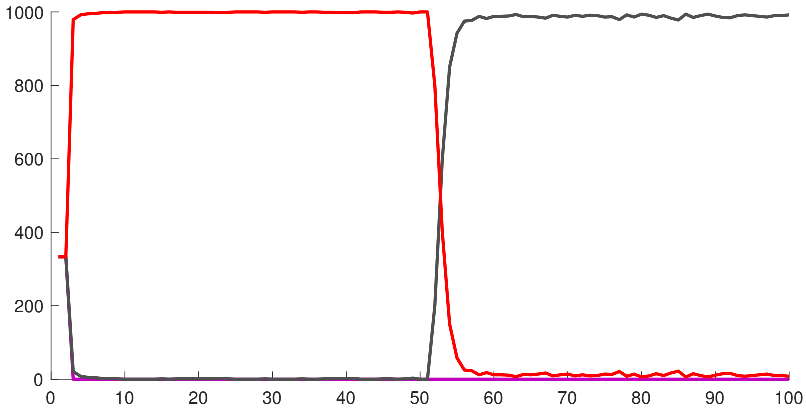
We apply our dynamic learning strategy to the following set of prediction models:

$$M_1 : y_{1t} = 0.1 + 0.6y_{1t-1} + \varepsilon_{1t}$$
$$M_2 : y_{2t} = 0.4 + 0.2y_{2t-1} + \varepsilon_{2t}$$
$$M_3 : y_{3t} = 0.9 + 0.1y_{3t-1} + \varepsilon_{3t}$$

with $\varepsilon_{it} \sim N(0, \sigma^2)$ independent for $i = 1, 2, 3$ and assume $y_{i0} = 0.25$, $i = 1, 2, 3$ and $\sigma = 0.05$. The model set is incomplete, but includes two models (M_1 and M_2) that are equivalent versions of the true model in the two parts of the sample.

We apply our dynamic model learning strategy to the simulated data. That is, we calculate the discounted predictive likelihood for each of the models (M_1, M_2 and M_3) and select the model (and value of the discount factor) which would have generated the highest product of predictive likelihoods until the given point in time. As we do for our application to exchange rate forecasting, we only consider information that would have been available at a certain point in time. Instead of excluding dynamic learning by setting $\alpha = 1$, we choose the same range of the discount factor as we do in our application to exchange rate forecasting: $\alpha \in \{0.50; 0.70; 0.80; 0.90; 0.99; 1\}$. We simulated $1,000$ runs and recorded how often each of the models was chosen at each point in time. Figure 1 presents the results. It shows that (i) in almost all cases the appropriate stochastic process was selected, (ii) the structural break was recognized quickly and that (iii) model 3 rightly played no role.

Figure 1: Dynamic Model Learning



The figure shows the number of simulation runs in which each of the three models was selected at each point in time by the dynamic model learning strategy. The red line represents model 1, the grey line model 2, and the purple line model 3.

4 Empirical Appendix

4.1 Point and interval forecasts

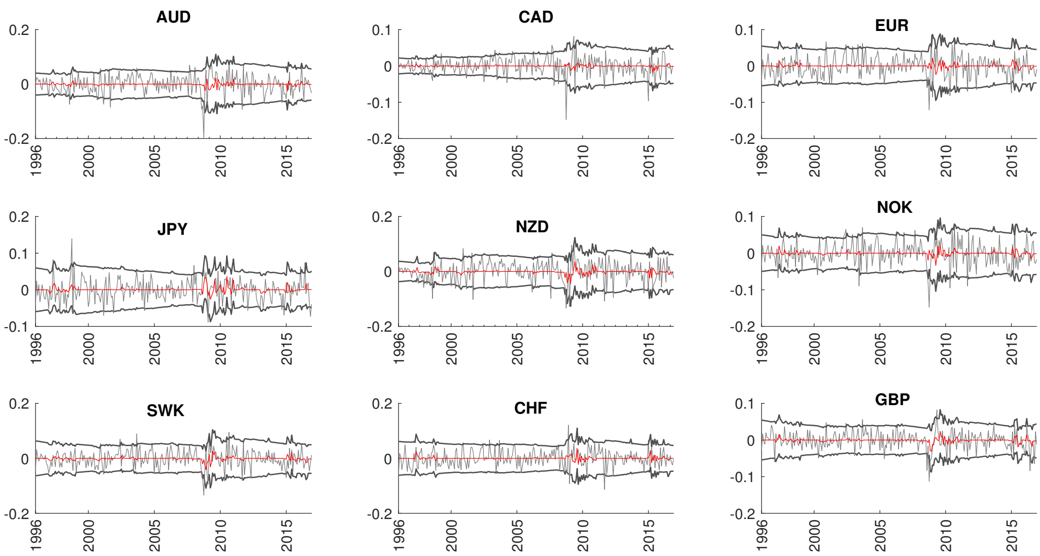
Bayesian methods provide the full predictive density, from which we can produce interval and point forecasts as a byproduct. Although our primary interest is on exploiting density forecasts for asset allocation, it is instructive to have a look at point and interval forecasts. In particular, the second column of Table 1 shows the empirical coverage rates (for a nominal coverage rate of 90%) for all currencies. These reveal good coverage properties, albeit very slightly too conservative.

The third column of Table 1 reports the ratio of mean squared forecasting errors relative to the simple random walk with constant volatility. Ratios below one indicate better point forecasting performance in terms of squared loss of the DML with UIP forecasts compared to those produced by the random walk. Our evidence on point forecasting is ambiguous with some ratios below and some above one. This finding once more shows how difficult it is to beat a simple random walk in terms of point forecasting accuracy. On the other hand, our previous results show that it is more fruitful to focus on density forecasts and exploit them for portfolio management.

Figure 2 plots point forecasts and credible intervals for each country in our sample along with the realizations. The predictive credible intervals show good coverage. This figure also illustrates, for every country, the importance of allowing for time-varying volatilities, particularly around the time of the subprime crisis where volatility forecast first increases significantly before gradually adjusting to the pre-crisis level.

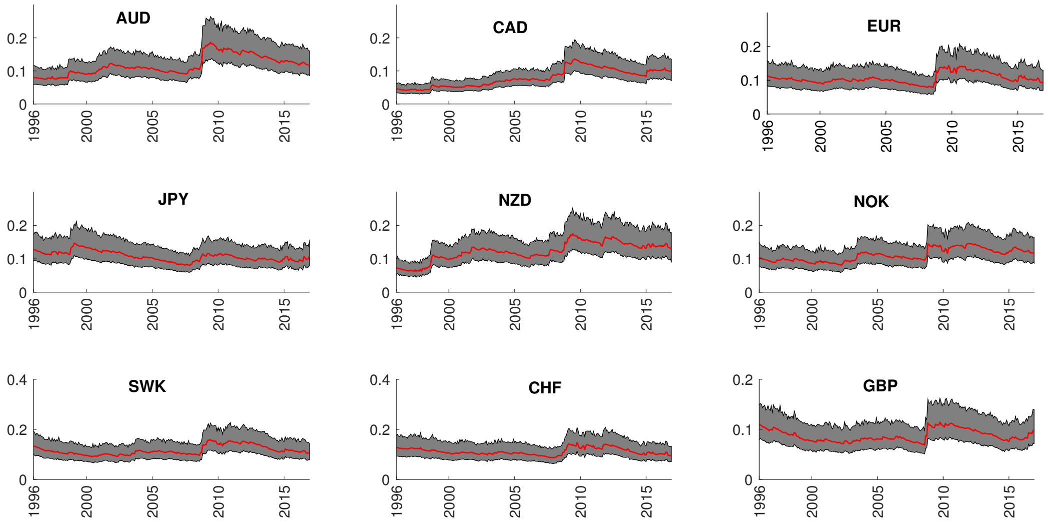
As we a adopt a Wishart matrix discounting (WMD) approach for the error covariance matrix, we are able to provide credibility intervals for our estimates of volatility and correlations. Figure 3 presents the point estimates of annualized volatility along with the 90% credibility intervals for the nine exchange rates. Figure 4 plots the point estimates of correlations along with the 90% credibility intervals for four selected exchange rates returns which display different patterns. The correlation between AUD and NZD increases to almost one at the end of the sample which reflects the well-established co-movements between these currencies. On the other hand, the intensity of the relationship between JPY and GBP strongly decreases, potentially due to country-specific drivers of the GBP exchange rate as a result of Brexit. Overall, these figures illustrate that this dimension of model flexibility is able to capture relevant global currency dynamics.

Figure 2: Interval and Point Forecasts



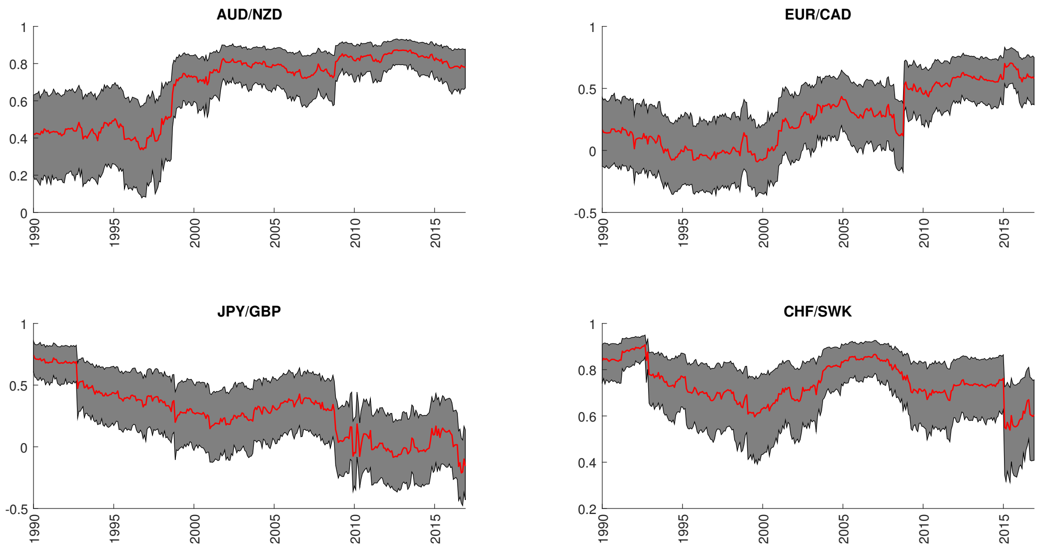
The figure shows the point forecast of exchange rate returns (red line) along with the 90% credibility intervals (dark grey) of the DML with UIP strategy. The realized exchange rate returns are indicated in light grey.

Figure 3: Volatility Forecasts and Credibility Intervals



The figure shows the point forecast of annualized volatility (red line) along with the 90% credibility intervals (dark grey) of the DML with UIP strategy.

Figure 4: Correlation Forecasts and Credibility Intervals



The figure shows the point forecast of correlations (red line) along with the 90% credibility intervals (dark grey) of the DML with UIP strategy.

Table 1: Evaluation of Interval and Point Forecasts

Currency	Nominal coverage: 90%	MSFE ratio
AUD	90%	1.03
CAD	88%	0.99
EUR	92%	1.06
JPY	93%	1.03
NZD	89%	0.98
NOK	91%	1.02
SWK	93%	0.99
CHF	93%	1.03
GBP	93%	0.97

The table summarizes the the coverage statistics of the interval forecasts and the relative point forecasting accuracy (MSFE ratio) of the DML with UIP model against the random walk.

4.2 Time-variation in performance

4.2.1 Evolution of wealth

Figure 5 compares the evolution of wealth for an investor who begins with one dollar and relies on DML with UIP to the wealth of an investor who uses a multivariate random walk with constant covariance to construct their portfolio. As is evident from the figure, the outperformance of DML with UIP is large, with the most striking gains around the time of the subprime crisis.

Figure 5: Evolution of Wealth

The figure depicts the evolution of wealth in the DML with UIP model and in the random walk model.

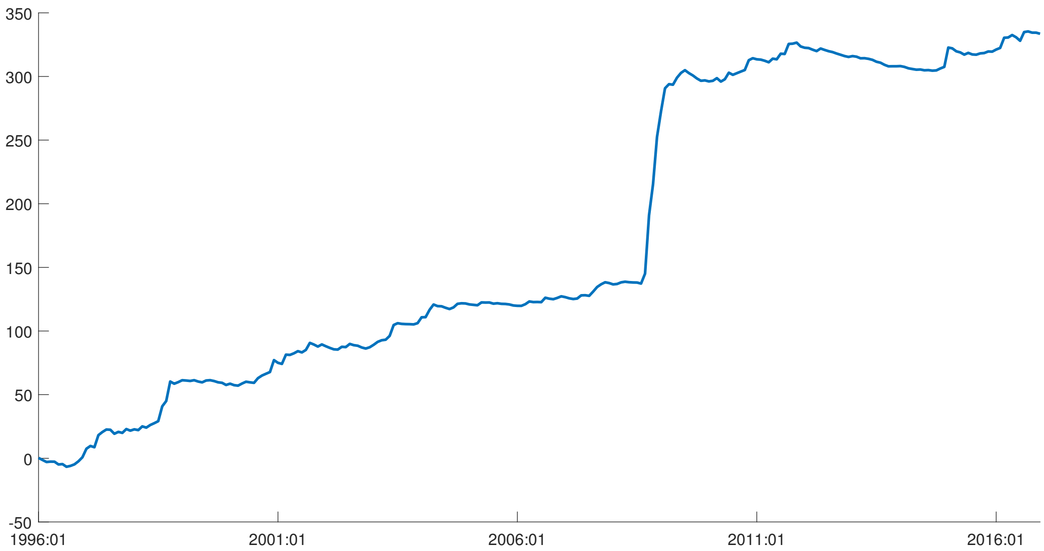
4.2.2 Cumulative differences in log predictive likelihoods

Figure 6 depicts the cumulative differences in predictive log likelihoods between DML with UIP and the random walk (with constant volatility). The out-performance of DML with UIP is most pronounced in the time of the subprime crisis.

4.2.3 Test statistics for Giacomini-Rossi Fluctuation test

The MSFE ratio is a measure of the global performance. It tells us whether the DML with UIP or the random walk have given more precise point forecasts in a mean squared error sense. However, we do not learn from this measure how the relative forecasting power has evolved over time. As we seek to shed some light on the evolution through time, we also provide a measure of local forecasting performance. A useful device for exploring time-variation in forecasting performance is the Fluctuation test by Giacomini and Rossi (2010). Figure 7 depicts (standardized) sequences of differences between the MSFE of the random walk and the MSFE of the DML with UIP model computed over rolling windows of 60 observations. Positive (negative) values of such differences indicate

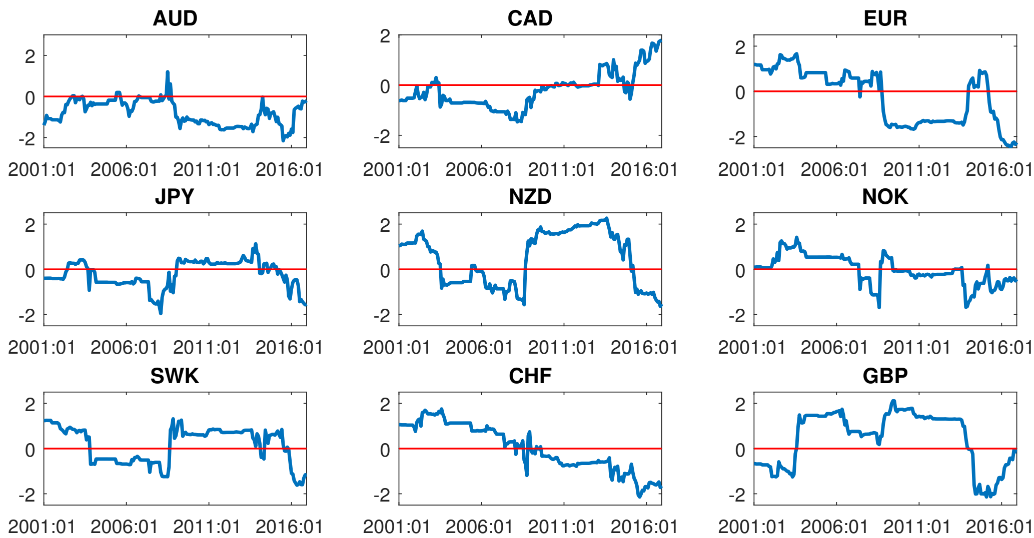
Figure 6: Cumulative Differences in Predictive Log Likelihoods



The figure shows the cumulative differences in predictive log likelihoods between DML with UIP and the random walk with constant volatility.

that DML with UIP forecasts better (worse) than the random walk. Figure 7 highlights that the relative forecasting performance is highly unstable across currencies and over time. This finding aligns with Rossi (2013). The standardized sequences of differences between the MSFE of the random walk and the MSFE of the DML with UIP provide the test statistics of the Fluctuation test. To carry out the Fluctuation test, i.e., testing the null hypothesis that the local relative MSFE equals zero at each point in time, requires computing critical values. Calculation of critical values in the Fluctuation test rests on the assumption that a rolling or fixed estimation window has been used for generating the out-of-sample forecasts. The out-of-sample forecasts in our setup were produced using exponential discounting. Hence, we cannot compute valid critical values for our application. This is, however, not a major concern since Figure 7 shows that the absolute values of the test statistics are greater than two only for few currencies at very few points in time. The null hypothesis of equal forecasting performance would thus essentially never be rejected at conventional significance levels.

Figure 7: Fluctuation Test Statistics



The figure shows Fluctuation test statistics through time. Positive values of the Fluctuation statistic imply that DML with UIP does better than the random walk.

4.3 Alternative sets of regressors

For our main results we did not include some of the traditional regressors used by exchange rate forecasters due to data revision concerns. But if we are willing to use final vintage data (as opposed to data that forecasters would have had in real time), we can extend our set of regressors to include purchasing power parity (PPP), the monetary model (MON) and an asymmetric Taylor Rule (ASYTAY). The Technical Appendix 1.4 provides details of what these are and how they are calculated.

Table 2 shows that including these fundamentals would not improve the performance of an investor's portfolio. Besides these conventional fundamentals, we also experimented with yield curve factors which are commonly used to exploit the terms structure of interest rates and the arising macroeconomic effects (Wright, 2011). This can be considered as an extension of the simple interest rate spread. However, in line with Berge (2014) including a level, slope and curvature factor does not improve our forecasts. The findings are available upon request.

Table 2: Alternative Set of Regressors

	Φ^{TC}	SR	SR^{TC}	PLL
DML with UIP	464*	1.12*	0.94*	22.01*
DML with PPP	217	0.90	0.71	22.00*
DML with MON	251	0.94	0.75	22.03*
DML with ASYTAY	263	0.95	0.76	22.02*
DML with ALL REGRESSORS	332	0.98	0.80	22.01*
DML	327	1.01*	0.82*	22.02*

The table summarizes the economic and statistical evaluation of our forecasts from different model configurations for the period from 1996:01*to* 2016:12. *We measure statistical significance for differences in performance fees* (Φ^{TC}) *and joint predictive log likelihoods (PLLs) using the (one-sided) Diebold and Mariano (1995) t-test using heteroskedasticity and autocorrelation robust (HAC) standard errors. We evaluate whether the Sharpe ratio before/after transaction costs* (SR/SR^{TC}) *of a model is different from that of the random walk (with constant volatility) benchmark based on 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.*

4.4 Alternative priors

We have experimented with many prior specifications that lie within our DML framework and, in particular, experimented with alternative choices of grids for the Minnesota shrinkage parameters. Results were robust. If we use a refined grid for the values of the hyperparameters, we find very slight forecast improvements (at the cost of increasing the computation time). This shows that our specification of grid points for the hyperparameters is sufficiently flexible to cover the model space. In this section, we discuss some alternative, more restrictive, prior specifications. Overall, we find that the rich shrinkage patterns we use pay off compared to more restrictive settings.

4.4.1 "Dense" prior structure

In this sub-section, we discuss a prior structure that represents a "dense" rather than a "sparse" modelling approach. We investigate how our results change when enforcing a "dense" prior rather than letting the data choose between a "dense" and a "sparse" structure. A dense prior is one where VAR lags and exogenous regressors cannot be removed from the model, instead only the degree of shrinkage intensity for each of the (blocks) of variables is selected (i.e. the prior shrinkage parameters cannot be set to be exactly zero as we do in our approach. We specify an alternative prior that features a dense structure as described in the following paragraph.

For γ_2 and γ_3, the shrinkage parameters for own and cross lags, we use grids of $\{0.0001; 0.01; 0.1\}$ and also the shrinkage parameter for UIP is estimated using a grid of $\{0.0001; 0.01; 0.1\}$. We do not take into account other exogenous variables in this setting.

Table 3 summarizes the results for this alternative selection of grid points for the shrinkage priors. It is evident that economic performance for the dense prior is inferior to the prior used for our main results, while the PLLs are similar. If the grid point 0.0001 is removed, performance deteriorates dramatically, with an annualized performance of −289 basis points and considerably lower PLL (21.70).

Table 3: "Dense" Prior Structure

	Φ^{TC}	SR	SR^{TC}	PLL
DML with UIP	285	1.00	0.78	22.00*
Type of restrictions: Model selection dynamics				
DML ($\alpha = 1$)	−321	0.29	0.21	21.71*
DML ($\alpha = 0.99$)	−163	0.53	0.34	21.72*
DML ($\alpha = 0.90$)	83	0.74	0.57	21.95*
DML ($\alpha = 0.80$)	189	0.92	0.71	22.00*
DML ($\alpha = 0.70$)	278	1.03	0.81	22.05*
DML ($\alpha = 0.50$)	231	0.96	0.74	22.03*

The table summarizes the economic and statistical evaluation of our forecasts from the DML and restricted versions thereof for the period from 1996*:*01 *to* 2016*:*12*. We measure statistical significance for differences in performance fees (Φ^{TC}) and joint predictive log likelihoods (PLLs) based on the (one-sided) Diebold and Mariano (1995) t-test using heteroskedasticity and autocorrelation robust (HAC) standard errors. We evaluate whether the Sharpe ratio before/after transaction costs (SR/SR^{TC}) 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.*

4.4.2 VAR with tight prior

We also explored a VAR for the nine exchange rates without any exogenous regressors and a very tight prior for the VAR coefficients. This setting is similar to Carriero, Kapetanios, and Marcellino (2009). For γ_2 and γ_3, the shrinkage parameters for own and cross lags, we use grids of $\{10^{-4}; 10^{-5}; 10^{-6}\}$. Although we find that for seven out of nine exchange rates the MSFE error is slightly lower than that of the random walk, results in terms of density forecasting accuracy and economic measures are inferior to our baseline setting: $PLL = 21.73$, $\Phi^{TC} = 164$, $SR = 0.64$ and $SR^{TC} = 0.62$.

4.4.3 Treating the exogenous variables as endogenous

We also investigate how the results change if the exogenous variables are not treated as such. Instead they are included as endogenous variables in the VAR. That is, instead of working with a 9 variable VAR with exogenous variables, we work with a 37 dimensional VAR involving the 9 exchange rates, 3 asset-specific variables, UIP, INT DIFF, STOCK GROWTH, (i.e. there are 3 such variables for each of 9 countries, hence this adds 27 variables to the VAR) and 1 non-asset specific variable, OIL. With this much larger VAR it is computationally infeasible to do a grid search over seven different prior shrinkage parameters. Accordingly, we employ the framework proposed by Koop and Korobilis (2013) which involves a single shrinkage parameter. We label this the KK-Minnesota-prior. The strategy of using a single shrinkage parameter for imposing shrinkage on all model parameters (except the intercept) is commonly used in the large Bayesian VAR literature; see Giannone, Lenza, and Primiceri (2015), Koop and Korobilis (2013) and Bańbura, Giannone, and Reichlin (2010). Following Koop and Korobilis (2013), the value of the single shrinkage parameter γ is adaptively (in each time period) selected from the grid $\gamma \in \{10^{-5}; 0.001; 0.005; 0.01; 0.05; 0.1\}$ and the shrinkage parameter of the intercepts a is set to 100 to be uninformative. The structure of this simpler version of the Minnesota prior is

$$\Omega_{0,i,jj} = \begin{cases} a \cdot s_i^2 & , a = 100 \ \text{ for INTERCEPTS} \\ \frac{\gamma}{r^2} & , \gamma \in \{10^{-5}; 0.001; 0.005; 0.01; 0.05; 0.1\} \text{ for } r = 1, ..., 6. \end{cases} .$$

Table 4 unambiguously conveys the message that the more restrictive structure of the Koop and Korobilis (2013) framework is clearly inferior in this exchange rate forecasting exercise compared to our proposed setting, both in statistical terms and even more so in economic terms. This highlights that allowing for different degrees of prior shrinkage on different blocks of parameters is empirically warranted.

4.5 Time-varying coefficients

In the preceding section, all of our VARs involved constant coefficients (but had time-varying volatilities). Time-variation in VAR coefficients can easily be added, but leads to inferior forecasting performance. To show this, we present results using a DML with

Table 4: KK-Minnesota-Prior

	Φ^{TC}	SR	SR^{TC}	PLL
DML with UIP	-96	0.42	0.36	21.84
Type of restrictions: Model selection dynamics				
DML ($\alpha = 1$)	-201	0.33	0.30	21.63
DML ($\alpha = 0.99$)	-201	0.33	0.30	21.63
DML ($\alpha = 0.90$)	-287	0.29	0.23	21.79
DML ($\alpha = 0.80$)	-210	0.35	0.30	21.84
DML ($\alpha = 0.70$)	-116	0.43	0.38	21.86
DML ($\alpha = 0.50$)	-21	0.51	0.45	21.87

The table summarizes the economic and statistical evaluation of the KK-Minnesota-prior for the period from 1996:01 *to* 2016:12*. We measure statistical significance for differences in performance fees (Φ^{TC}) and joint predictive log likelihoods (PLLs) based on the (one-sided) Diebold and Mariano (1995) t-test using heteroskedasticity and autocorrelation robust (HAC) standard errors. We evaluate whether the Sharpe ratio before/after transaction costs (SR/SR^{TC}) 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.*

UIP specification identical to that used in the preceding section except that it sets $\lambda = 0.99$. Results are presented in Table 5. In comparison to the constant parameter case ($\lambda = 1$) in our main results, we find that using time-varying VAR coefficients is in general detrimental for forecasting performance, particularly when evaluating forecasts in terms of the economic performance measures. We find strong evidence that allowing for abrupt switching between different models for handling the evolving relationship between exchange rates and fundamentals as highlighted by Sarno and Valente (2009). But allowing for gradual change in parameters is not a useful addition. An exception is the specification "DML without own/cross lags but with ALL REGRESSORS". In this case time-varying parameters do not turn out to be detrimental. It appears that, in specifications that involve estimation of many parameters for the VAR lags, time-variation in parameters leads to lower performance. This finding aligns with the econometric literature with respect to time-varying VAR parameters in medium-size VARs (Chan and Eisenstat, 2018; Koop and Korobilis, 2013).

4.6 Additional results on portfolio performance

In this sub-section, we explore in greater detail the portfolio performance implied by our flexible DML with UIP model and the portfolio performance based on the random walk.

Table 5: TVP-VAR

	Φ^{TC}	SR	SR^{TC}	PLL
DML with UIP	189	0.85	0.70	22.00*
Alternative sets of regressors				
DML with INT_DIFF	114	0.77	0.62	21.98*
DML with STOCK_GROWTH	32	0.71	0.55	21.99
DML with OIL	−20	0.64	0.49	21.95
DML with ALL REGRESSORS	202	0.83	0.69	21.99*
DML with NO REGRESSORS	42	0.71	0.55	22.00*
Type of restrictions: VAR lags				
DML without own lags ($\gamma_2 = 0$) and NO REGRESSORS	−111	0.45	0.38	21.89*
DML without cross lags ($\gamma_3 = 0$) and NO REGRESSORS	−107	0.55	0.39	21.82**
DML without own/cross lags ($\gamma_2 = \gamma_3 = 0$) and NO REGRESSORS	5	0.54	0.53	21.72**
DML without own/cross lags ($\gamma_2 = \gamma_3 = 0$) but with ALL REGRESSORS	289	0.81	0.75	22.00*
Type of restrictions: Model selection dynamics				
DML ($\alpha = 1$)	−325	0.37	0.20	21.65
DML ($\alpha = 0.99$)	−247	0.44	0.27	21.67
DML ($\alpha = 0.90$)	0	0.63	0.51	21.95
DML ($\alpha = 0.80$)	66	0.70	0.56	21.97
DML ($\alpha = 0.70$)	42	0.71	0.56	22.00*
DML ($\alpha = 0.50$)	−29	0.65	0.48	21.98

The table summarizes the economic and statistical evaluation of our forecasts from the TVP-VAR for the period from 1996:01 *to* 2016:12. *We measure statistical significance for differences in performance fees (Φ^{TC}) and joint predictive log likelihoods (PLLs) based on the (one-sided) Diebold and Mariano (1995) t-test using heteroskedasticity and autocorrelation robust (HAC) standard errors. Restrictions on α correspond to the specification DML with NO REGRESSORS. We evaluate whether the Sharpe ratio (SR) 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.*

4.6.1 Statistics of the portfolio results

Table 6 compares descriptive statistics of the portfolio performance based on the DML with UIP model and the random walk. The mean return (measured at a monthly frequency) is almost twice as high based on the DLM with UIP model than based on the random walk. The annualized volatility of portfolio returns is 10.81% ($3.12\% \times \sqrt{12}$) and is hence only slightly higher than the target portfolio volatility of 10%. Skewness of portfolio returns is substantially higher based on DLM with UIP, while the kurtosis is considerably lower than in the random walk case. Altogether, the portfolio characteristics of the DML with UIP model are clearly superior to those of the random walk. In addition, the characteristics of the portfolio returns based on the DML with UIP strategy are also more favourable for risk management and diversification purposes: the correlation of the returns to equities (proxied by S&P500 returns) is even negative and the first-order autocorrelation of returns and squared returns is lower than in case of the portfolio returns based on the random walk.

Table 6: Statistics of Portfolio Results

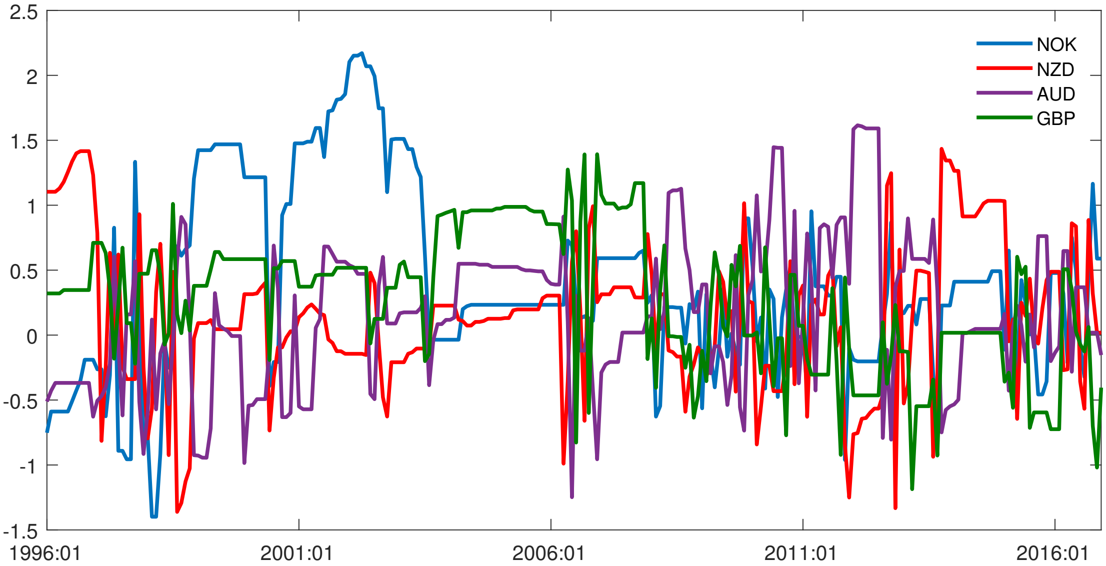
	DML with UIP	Random walk
Mean return (in %)	1.22	0.69
Volatility (in %)	3.12	3.50
Skewness	−0.15	−0.94
Kurtosis	3.18	5.18
Positive returns (>0 in %)	68	62
First-order autocorrelation of returns	0.06	0.12
First-order autocorrelation of squared returns	0.06	0.23
Correlation to S&P 500 returns	−0.05	0.22

The table summarizes the portfolio results from the DML with UIP model and the random walk for the period from 1996:01 *to* 2016:12.

4.6.2 Evolution of portfolio weights

It is of interest how the portfolio weights have evolved through time. Figure 8 and Figure 9 depict the evolution of portfolio weights for the bonds of the high-interest-rate countries (NOK, NZD, AUD, GBP) and for the bonds of the low-interest-rate countries (CHF, EUR, JPY, USD), respectively. The figures show that there is considerable portfolio rebalancing over time. However, the implied portfolio weights are not excessive and are hence implementable by an investor without imposing additional restrictions on portfolio weights. Remember that our portfolio optimization exercises take into account transaction costs ex ante and the results in Section 4.6.1 reveal that expected and realized volatility of portfolio returns are close together. Figure 10 shows the average portfolio weight for the bonds of low-interest-rate currencies and high-interest-rate currencies. The pattern is not surprising: the average portfolio weights of the high-interest-rate currencies are all positive, while the average portfolio weight of the low-interest-rate currencies JPY and CHF is negative and only marginally positive for EUR. The largest average short position has been the CHF, while the weight for the JPY is only moderately negative. Interestingly, those patterns are in line with the findings by Ackermann, Pohl, and Schmedders (2016). The clearly positive net position of the USD (+1.06) was to be expected as the USD is the domestic currency in our portfolio setup and therefore short-term USD is the risk-free asset. This means that, on average, long and short positions in foreign bonds have been approximately zero.

Figure 8: Portfolio Weights for the Bonds of the High-Interest-Rate Countries



This figure shows the evolution of the portfolio weights for the bonds of the high-interest-rate currencies based on the DML with UIP model.

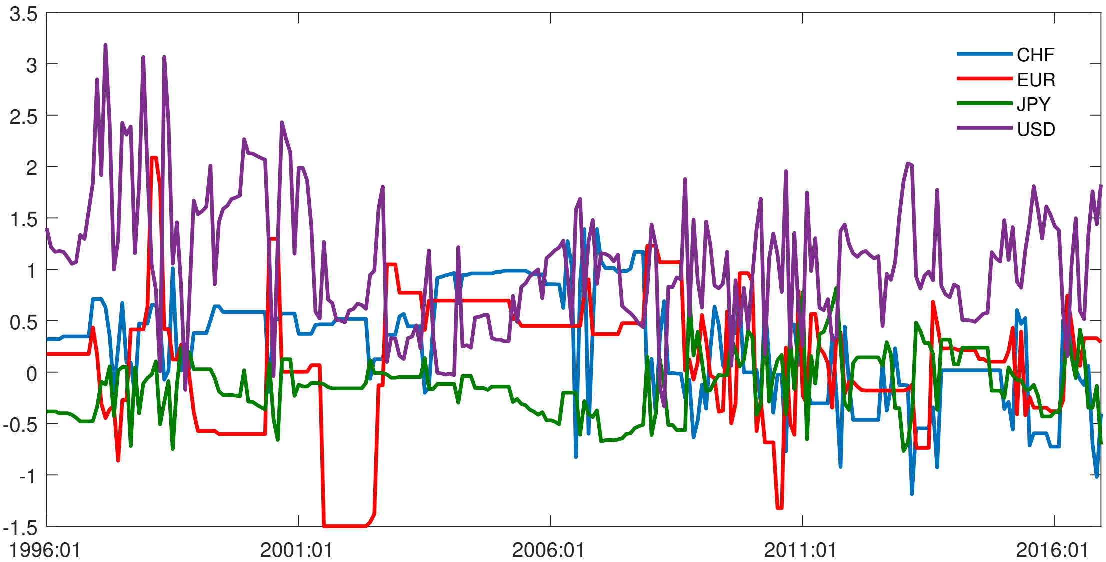
4.6.3 Restrictions on portfolio weights

Table 7 summarizes the effect of restrictions on portfolio weights on economic utility and the Sharpe ratio. Restricting the portfolio weights to $[-1; 1]$ leads to even slightly better portfolio performance than in the case where the portfolio weights are left unrestricted. This is good news from a risk-management perspective since excessive portfolio weights are not required to achieve high utility gains. However, severe restrictions on the portfolio weights are clearly detrimental for portfolio performance.

4.6.4 Global Harvest Index as benchmark

We consider the Deutsche Bank Global Currency Harvest Index as an additional benchmark strategy. This index can be seen as a proxy for carry trade returns as a style strategy. Figure 11 shows that the wealth path generated by our random walk model and the evolution of the Global Currency Harvest Index are broadly similar. The correlation between the returns is 0.60. This result is not surprising, given that our

Figure 9: Portfolio Weights for the Bonds of the Low-Interest-Rate Countries



This figure shows the evolution of the portfolio weights for the bonds of the low-interest-rate currencies based on the DML with UIP model.

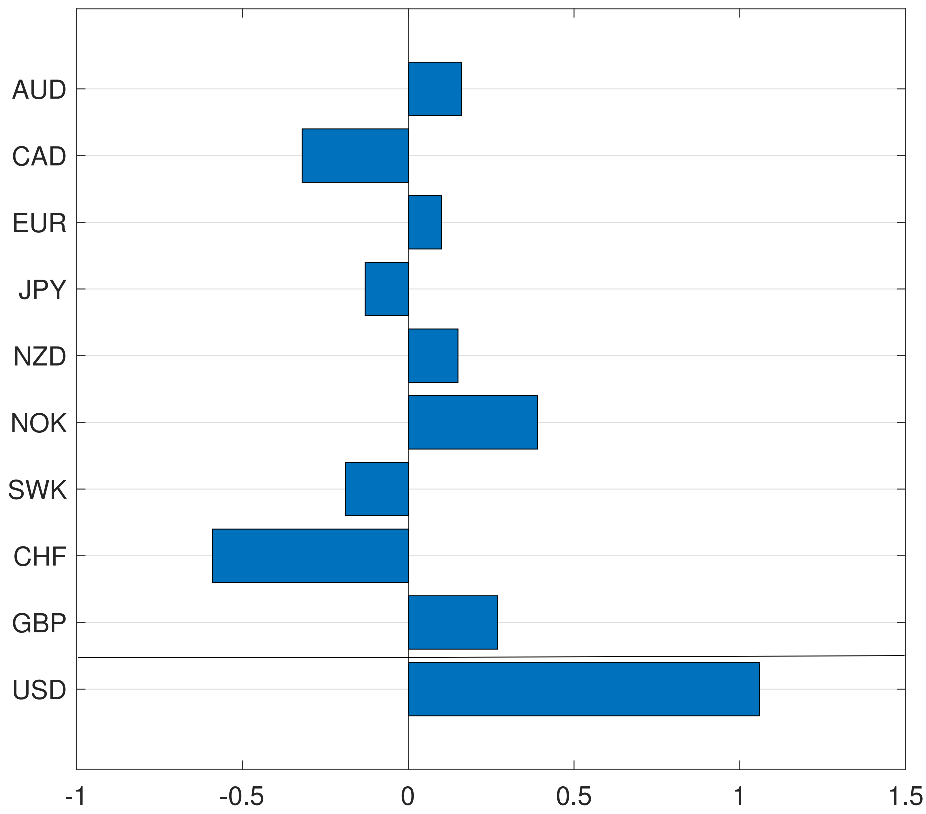
Table 7: Restrictions on Portfolio Weights

	DML with UIP		
Weight restriction	Φ^{TC}	SR	SR^{TC}
No restriction	464*	1.12**	0.93**
$[-1; 1]$	468*	1.13**	0.95**
$[-0.5; 0.5]$	305	0.97	0.82
$[-0.25; 0.25]$	-8	0.67	0.57
$[-0.1; 0.1]$	-320	0.32	0.25
Equal weights	-602	-0.11	-0.11

The table summarizes the effect of restrictions on the portfolio weights on economic utility and the Sharpe ratio. One star indicates significance at 10% level; two stars indicate significance at the 5% level.

random walk strategy and the strategy underlying the Global Currency Harvest Index are both carry trade strategies, only differing with respect to implementation details. The Global Currency Harvest Index does not start before 2000:09 and there are some missing data observations which we imputed by (linear) interpolation. The findings are essentially similar to the original carry trade strategy we consider and therefore leaves

Figure 10: Average Portfolio Weights



This figure shows the average portfolio weights for the bonds based on the DML with UIP model.

our results unchanged.

4.6.5 Portfolio performance when removing one currency

To assess the sensitivity of the portfolio performance, we compute the Sharpe ratios when we respectively remove one currency from the set of currencies and set the respective portfolio weight to 0. Table 8 shows that there is not one particular currency that drives the results. Not surprisingly, enforcing dollar neutrality leads, in relative terms, to the largest decrease in the Sharpe ratio.

4.6.6 Results for single currencies

We also analyze the case where only one foreign bond is considered for investment in addition to the risk-less USD bond (from the perspective of a US investor). Table 9 reports

Figure 11: Evolution of Wealth Relative to Carry Trade Strategies

The figure depicts the evolution of wealth for the DML with UIP strategy, the random walk model and the Global Currency Harvest Index.

the results and once again sends the story that there does not emerge one particular currency that leads to attractive portfolio results and reinforces our finding that market timing in a large set of currencies is key for economic utility gains. In Table 9, we also report the ratio of univariate predictive log scores relative to the random walk (Pred Log Score ratio) and, once more, find that forecasting gains stem from taking into account cross dynamics in the multivariate approach rather than from particular currencies. We stress that, given our multivariate approach, the reported results for the univariate case are somewhat artificial for the Sharpe ratios and predictive log scores.

4.7 Additional robustness checks

In this sub-section, we briefly mention a couple of additional specifications we considered.

Spillover effects

The first of these investigated whether spillover effects involving macroeconomic fundamentals might be important. Such third-country effects have been discussed in

Table 8: Removing Single Currencies

Removed currency	SR	SR^{TC}
AUD	1.01*	0.84*
CAD	1.09*	0.93*
EUR	1.08*	0.91*
JPY	1.06*	0.88*
NZD	0.97*	0.80*
NOK	1.10*	0.93*
SWK	0.99*	0.83*
CHF	1.09*	0.93*
GBP	0.97*	0.80*
USD	0.90	0.71

The table reports the Sharpe ratios for the DML with UIP model if one particular is left out in each row. 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.*

Table 9: Single Currencies

Currency	SR	SR^{TC}	*Pred Log Score ratio*
AUD	0.33	0.29	1.0220
CAD	0.17	0.12	1.0401
EUR	0.26	0.21	1.0076
JPY	0.32	0.27	1.0104
NZD	0.38	0.34	1.0366
NOK	0.02	0.01	1.0156
SWK	0.22	0.19	1.0117
CHF	0.04	0.01	1.0174
GBP	0.20	0.16	1.0216

The first two columns of the table summarize the Sharpe ratios (before and after transaction costs) for the DML with UIP model if only one currency is considered for investment in addition to the USD. In the third column, we report the ratio of predictive log scores (Log Pred Score ratio) for the DML with UIP model against the random walk for each country.

Berg and Mark (2015). For instance, instead of including only the UIP for the UK in the equation for the UK currency (as we do), we can also include the UIPs for all the other currencies as well. If we do this, results are not noticeably affected. Our VAR specification allows spillovers between the exchange rates for different countries. This kind of spillover we have found to improve forecasts. Adding spillovers involving macroeconomic fundamentals results in no additional benefits.

Alternative measure of portfolio performance

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 do not have to assume a particular utility function. The results compared to the reported quadratic utility case are very similar and available upon request.

Alternative risk aversion

It is of interest whether the economic utility gains can also be achieved by investors with higher risk aversion. To explore this issue, we also considered the risk aversion coefficient $\theta = 6$ instead of $\theta = 2$. For this case, we found even larger utility gains than in our baseline setting, achieving an annualized performance fee of 525 basis points, that is $\Phi^{TC} = 525$, for the DML with UIP model (464 basis points in the base case) .

Specific degrees of time variation for different blocks of coefficients

As discussed previously, we have found that working with a constant coefficient VAR by setting $\lambda = 1$ leads to improved forecast performance over $\lambda = 0.99$. But these specifications assume the same λ applies to all the VAR coefficients. It is theoretically possible that, by allowing for different degrees of shrinkage for different blocks of coefficients, forecast performance can be improved. In practice, we have done extensive experimentation and have not found any forecast improvements by doing so.

Alternative grid for the decay factor α

We also considered a more refined grid for choosing α, namely $\alpha \in \{0.40 : 0.01 : 1.00\}$. In this case, $\alpha = 0.73$ is selected from the data over the entire period and, hence, is quite similar to our benchmark results ($\alpha = 0.70$). The results for the refined grid are almost exactly the same as in our base case.

4.8 The backward elimination approach

We explore how the backward elimination (BE) approach applied by Kouwenberg, Markiewicz, Verhoeks, and Zwinkels (2017) for exchange rate forecasting works in our setting, that is, for the sample from 1986:01 to 2016:12 and the G10 currencies. As potential predictors, we consider UIP, INT DIFF, STOCK GROWTH, OIL and the first own lag of each exchange rate in an AR1 model. Each of these predictors along with an intercept defines an exchange rate model, leaving us with 5 different univariate forecasting models. We slightly extend the procedure used by Kouwenberg, Markiewicz, Verhoeks, and Zwinkels (2017) to allow for density forecasts rather than point forecasts only. To this end, we use Bayesian predictive regressions with diffuse normal-inverse gamma priors. Apart from that, we follow their procedure for forecasting exchange rates using backward elimination:

Step 1: In-sample estimation of the models

We estimate the 5 exchange rate models individually using data in the in-sample estimation period $[t_0, t_1]$ from time t_0 to t_1, where initially $t_0 = 1986{:}01$ and $t_1 = 1993{:}12$. Hence, the first estimation sample contains 96 monthly observations. Subsequently, the 5 exchange rate models are re-estimated using an expanding window in each cycle.

Step 2: Generating pseudo out-of-sample forecasts with individual models

At the end of the in-sample estimation period $[t_0, t_1]$, we generate one-step-ahead forecasts $E^i_{t_1}(\Delta s_{t_1+1})$ for time $t_1 + 1$, with $i = 1, 2, ..., 5$ indicating the exchange rate model. Steps 1 and 2 are now repeated another 23 times, with t_1 moved forward 1 period in every iteration, until we have created 24 out-of-sample forecasts from $t_2 = 1994{:}01$ until $t_3 = 1995{:}12$. We refer to $[t_2, t_3]$ as the out-of-sample forecast evaluation period. The univariate predictive regressions in steps 1 and 2 are Bayesian regressions with diffuse normal-inverse gamma priors. Hence, in terms of point forecasting, our results are essentially the same as in non-Bayesian regressions.

Step 3: Model selection based on forecast errors and backward elimination

In this step, we compare the forecast errors of the 5 models in the evaluation period $[t_2, t_3]$ and select the models that generate the most accurate point predictions of the exchange rate. We do so by regressing the realized exchange rate returns in $[t_2, t_3]$ on the 5 model forecast series $E^i_{t_1}(\Delta s_{t_1+1})$, using a (non-Bayesian) multivariate regression model in this step:

$$\Delta s_t = \alpha + \sum_{i=1}^{N} \beta_{i,t} E^i_{t-1}(\Delta s_t) + \epsilon_t$$

for $t = t_2, t_2 + 1, \ldots, t_3$, and $E^i_{t-1}(\Delta s_t)$ is the one-step-ahead model (i)-based forecast at time $t - 1$. In this step, we estimate the weights, $\beta_{i,t}$, that should be attributed to each model-based forecast. We first include all 5 forecasting models in the regression equation. Second, we eliminate insignificant variables one by one (stepwise), effectively setting the estimated coefficients to 0 $\left(\widehat{\beta}_{i,t} = 0\right)$, using a p-value of 5% as the cutoff for significance.

Convergence is reached when either all remaining forecasts have significant coefficients or a random walk model with drift α remains. Given that the evaluation period $[t_2, t_3]$ covers 24 monthly observations in the first cycle, this procedure is initially applied to 24 observations and subsequently to an expanding window.

Step 4: Generate one out-of-sample combination forecast

We create one out-of-sample combination forecast for time $t_3 + 1$ by using the estimated $\widehat{\beta}_{i,t}$ coefficients from the final regression model in step 3:

$$E_t^{\mathrm{BE}}\left(\Delta s_{t+1}\right) = \sum_{i=1}^{N} \widehat{\beta}_{i,t} E_t^i\left(\Delta s_{t+1}\right),$$

where E_t^{BE} indicates the one-period-ahead forecast based on the BE weights at time $t = t_3$. Several of the $\widehat{\beta}_{i,t}$ coefficients may be equal to 0, depending on the outcome of the BE procedure. Steps $1, 2, 3$, and 4 are now repeated, each time adding one more observation to the in-sample estimation window $[t_0, t_1]$ by increasing t_1 by 1, and adding one more observation to the out-of-sample model selection period $[t_2, t_3]$ by increasing t_3 by 1. Eventually, the BE procedure makes 252 out-of-sample forecasts of the exchange rate, covering the period from 1996:01 to 2016:12. We use these 252 forecasts to evaluate the performance of the BE rule. We leave the obtained $\widehat{\beta}_{i,t}$ coefficients unrestricted for generating the point forecasts and adjust the weights in the case of density prediction to ensure that the combination provides a predictive density.

We evaluate the backward elimination rule in terms of statistical and economic forecasting gains. Further, we compute how often each model was selected by the backward elimination rule. Table 10 reports the relative frequency each model was selected and the MSFE ratio against the random walk. It is evident that each of the models is either never or rarely selected. Hence, given such a high degree of shrinkage, the overall forecasts are typically closer to the random walk forecasts than the point forecasts of our DML with UIP strategy or other DML strategies.

Comparing the density forecasts of our DML with UIP strategy and the BE procedure, we recall that our DML with UIP strategy is a multivariate approach and we obtain the average joint predictive log likelikood (PLL). In case of the univariate BE procedure, we cannot compute the average joint predictive log likelihood. Instead, we compute the average sum of individual log predictive likelihoods. Note that the average is computed over the forecast evaluation period and the sum is computed over the currencies. The ratio

Table 10: Model Selection Frequencies and Point Forecasting Accuracy

Model	AUD	CAD	EUR	JPY	NZD	NOK	SWK	CHF	GBP
AR1	0%	7%	0%	0%	0%	2%	0%	0%	0%
UIP	1%	3%	0%	0%	0%	2%	11%	0%	7%
INT DIFF	0%	1%	0%	0%	0%	2%	0%	0%	0%
STOCK GROWTH	1%	0%	0%	0%	12%	0%	0%	0%	0%
OIL	0%	0%	0%	0%	4%	0%	3%	0%	0%
MSFE ratio	1.0127	1.0147	1.0000	1.0000	1.0066	0.9966	1.0168	1.0000	1.0308

This table reports how often the 5 *individual forecasting models were selected by the backward model elimination rule (rows* $2-6$*). The model selection rule was applied in* 252 *consecutive months from* 1996:01 *to* 2016:12. The last row of the table reports the relative point forecasting accuracy (MSFE ratio) of the approach against the random walk (without drift).

between the average joint predictive log likelihood of our DML with UIP strategy and the average sum of individual predictive log likelihoods of the BE procedure is 1.21 in case of constant volatilities in the BE procedure. If we consider time-varying volatilities in the BE procedure (using a decay factor of 0.97 as we have in the DML with UIP strategy), the ratio is 1.18. These large gains relative to the BE rule in terms of predictive log likelihoods point to the importance of a multivariate forecasting approach for exchange rates with respect to density forecasting accuracy.

In addition to the statistical performance, we also evaluate the BE rule in economic terms. To do so, we use a strategy that has also been applied in Kouwenberg, Markiewicz, Verhoeks, and Zwinkels (2017). We calculate the returns of an investment strategy that buys (sells) one unit of the foreign currency vis-à-vis the US dollar when the model predicts an appreciation (depreciation) of the foreign currency. Here, we calculate the return (r) to this strategy without interest rate differentials, hence, capturing only the foreign exchange rate returns. The (raw) foreign-exchange return is given by

$$r = \frac{E_{t-1}\left(\Delta s_i\right)}{|E_{t-1}\left(\Delta s_t\right)|} \Delta s_t.$$

We subsequently form equal-weighted portfolios of the nine foreign currencies. The annualized Sharpe ratio of this strategy is 0.08 before transaction costs.

We stress that Kouwenberg, Markiewicz, Verhoeks, and Zwinkels (2017) consider a broader set of regressors and use quarterly data. These difference to our setting might offer an explanation why the backward elimination procedure struggles to differentiate from a strategy based on the random walk in our setup.

4.9 Results for the long sample

In this sub-section, we report some additional key results for our long sample period which starts in 1973:01 and for which we compute out-of-sample results from 1990:01 to 2016:12. Due to data availability we do not consider the inclusion of exogenous regressors for this sample period.

Table 11 summarizes the results. As is the case for the short sample, DML substantially outperforms the multivariate random walk (i.e. DML without own/cross lags) both in terms of PLLs and economic criteria. Here again, fast model switching is found to be crucially important for the accuracy of density forecasts and portfolio allocation. The optimal decay factor α is found to be 0.80 over the entire evaluation period and is thus comparable to the optimal decay factor of the short period ($\alpha = 0.70$). As for our short sample, we consider the G10 currencies.

Figure 12 illustrates the high frequency of model change when the decay factor is chosen from the data. The vertical axis plots the model numbers from 1 to 32 against time for two cases. The set of models begins with model number 1 which is the multivariate random walk without drift and ends with model number 32 which is one of the most flexible models (i.e. the VAR model with an intercept, own lags with shrinkage parameter $\gamma_2 = 0.9$ and cross lags with shrinkage parameter $\gamma_3 = 0.9$). The two lines in Figure 12 are for DML (with α selected in a time-varying manner) and DML ($\alpha = 1$). In our flexible specification where the decay factor is chosen from the data, model change occurs much more frequently than in the case when there is no discounting of forecasting performance ($\alpha = 1$). Many different models are selected over time. The individual specification which is picked most frequently is the multivariate random walk (in approximately half of the cases). This)prominent role of the multivariate random walk reinforces the story that sparsity is a key aspect. Figure 13 shows which blocks of variables are included at each point in time (coloured diamonds). Blank spaces in the graph depict the time-varying sparsity induced by DML, that is, periods where a block of variables is not selected. Typically, we observe persistence in the selection of a block of variables.

Figure 14 compares the evolution of wealth for an investor who begins with one dollar and relies on DML to the wealth of an investor who uses a multivariate random walk with constant covariance to construct their portfolio. As for the short sample, the outperformance of DML is large, with the most striking gains around the time of the subprime crisis.

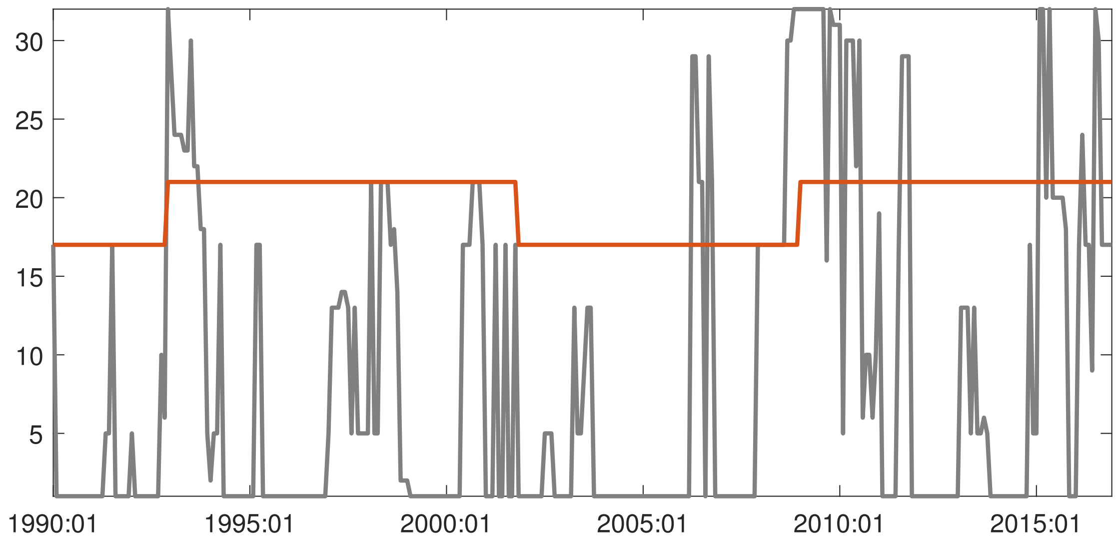
Overall, all key findings for the short sample period also apply for the long sample.

Table 11: Evaluation of Forecasting Results

	Φ^{TC}	SR	SR^{TC}	PLL
DML	485^{**}	1.08^{**}	0.92^{**}	22.05
Type of restrictions: VAR lags				
DML without own lags ($\gamma_2 = 0$)	365^{*}	0.82^{*}	0.72^{*}	21.86
DML without cross lags ($\gamma_3 = 0$)	278	0.80	0.66	21.78
DML without own/cross lags ($\gamma_2 = \gamma_3 = 0$)	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 ($\alpha = 0.90$)	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 ($\alpha = 0.50$)	409^{*}	1.04^{**}	0.84^{**}	22.05

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 (Φ^{TC}) and joint predictive log likelihoods (PLLs) based on the (one-sided) Diebold and Mariano (1995) t-test using heteroskedasticity and autocorrelation robust (HAC) standard errors. We evaluate whether the Sharpe ratio before/after transaction costs (SR/SR^{TC}) 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.*

Figure 12: Frequency of Model Change

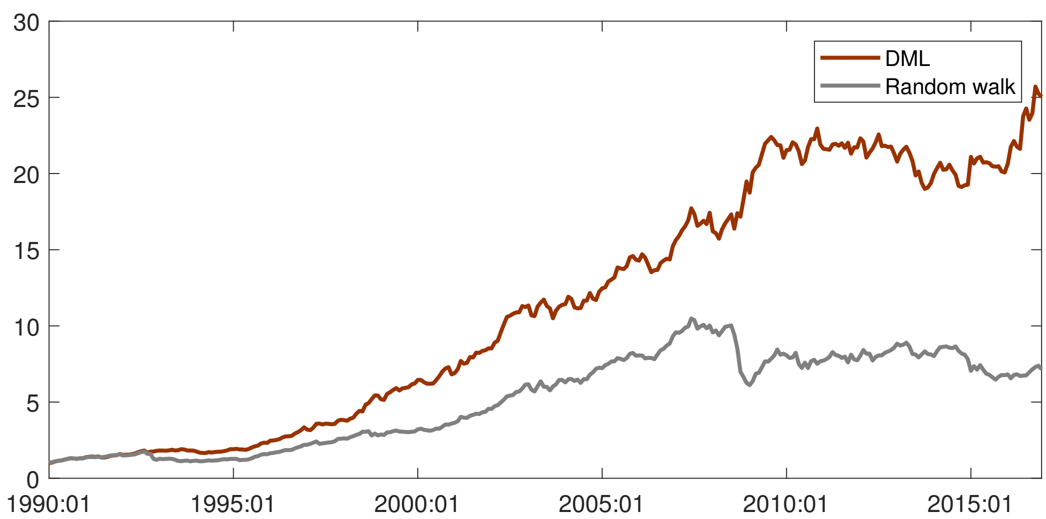


The figure displays the frequency of model change over time using the long sample. The vertical axis represents the model configurations $1, ..., 32$*. The red line depicts the evolution of the selected model configuration for* $\alpha = 1$*. The grey line shows the evolution of the selected model configuration when is dynamically chosen from the grid of values* $\alpha \in \{0.50; 0.70; 0.80; 0.90; 0.99; 1\}$*.*

Figure 13: Inclusion of Blocks of Variables

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

Figure 14: Evolution of Wealth



The figure depicts the evolution of wealth in the DML model and the random walk model.

References

Ackermann, F., W. Pohl, and K. Schmedders (2016): "Optimal and naive diversification in currency markets," *Management Science*, 63(10), 3347–3360, doi: https://doi.org/10.1287/mnsc.2016.2497.

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.

Bańbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25(1), 71–92, doi: https://doi.org/10.1002/jae.2733.

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, doi: https://doi.org/10.1002/jae.2350.

Billio, M., R. Casarin, F. Ravazzolo, and H. van Dijk (2013): "Time-varying combinations of predictive densities using nonlinear filtering," *Journal of Econometrics*, 177(2), 213–232, doi: https://doi.org/10.1016/j.jeconom.2013.04.009.

Bilson, J. F. O. (1978): "The Current Experience with Floating Exchange Rates: An Appraisal of the Monetary Approach," *American Economic Review*, 68(2), 392–397.

Carriero, A., G. Kapetanios, and M. Marcellino (2009): "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, 25(2), 400–417, doi: https://doi.org/10.1016/j.ijforecast.2009.01.007.

Chan, J. C., and E. Eisenstat (2018): "Bayesian model comparison for time-varying parameter VARs with stochastic volatility," *Journal of Applied Econometrics*, doi: https://doi.org/10.1002/jae.2617.

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.

Dornbusch, R. (1976): "Expectations and Exchange Rate Dynamics," *Journal of Political Economy*, 84(6), 1161–1176.

Engel, C. (2013): "Exchange Rates and Interest Parity," NBER Working Papers 19336, National Bureau of Economic Research, Inc.

Geweke, J., and G. Amisano (2011): "Optimal prediction pools," *Journal of Econometrics*, 164(1), 130–141.

Giacomini, R., and B. Rossi (2010): "Forecast comparisons in unstable environments," *Journal of Applied Econometrics*, 25(4), 595–620, doi: https://doi.org/10.1002/jae.1177.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior selection for vector autoregressions," *The Review of Economics and Statistics*, 97(2), 436–451, doi: https://doi.org/10.1162/REST_a_00483.

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.

Koop, G., and D. Korobilis (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177(2), 185–198, doi: https://doi.org/10.1016/j.jeconom.2013.04.007.

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, doi: https://doi.org/10.1017/S0022109017000011.

Ledoit, O., and M. Wolf (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15(5), 850–859, doi: https://doi.org/10.1016/j.jempfin.2008.03.002.

Mark, N. C., and D. Sul (2001): "Nominal exchange rates and monetary fundamentals: evidence from a small post-Bretton Woods panel," *Journal of International Economics*, 53(1), 29–52.

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.*

Rogoff, K. (1996): "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, 34(2), 647–668.

Rossi, B. (2013): "Exchange rate predictability," *Journal of Economic Literature*, 51(4), 1063–1119, doi: https://doi.org/10.1257/jel.51.4.1063.

Sarno, L., and G. Valente (2009): "Exchange rates and fundamentals: Footloose or evolving relationship?," *Journal of the European Economic Association*, 7(4), 786–830, doi: https://10.1162/JEEA.2009.7.4.786.

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. (2011): "Term premia and inflation uncertainty: Empirical evidence from an international panel dataset," *American Economic Review*, 101(4), 1514–34.