

A New Perspective of the Meese-Rogoff Puzzle: Application of Sparse Dynamic Shrinkage*

Zheng Fan^{†1}, Worapree Maneesoonthorn^{‡2}, and Yong Song^{§1}

¹Department of Economics, University of Melbourne

²Department of Econometrics and Business Statistics, Monash University

July 15, 2025

Abstract

We propose the Markov Switching Dynamic Shrinkage process (MSDSP), nesting the Dynamic Shrinkage Process (DSP) of [Kowal et al. \(2019\)](#). We revisit the Meese-Rogoff puzzle ([Meese and Rogoff, 1983a,b, 1988](#)) by applying the MSDSP to the economic models deemed inferior to the random walk model for exchange rate predictions. The flexibility of the MSDSP model captures the possibility of zero coefficients (sparsity), constant coefficient (dynamic shrinkage), as well as sudden and gradual parameter movements (structural change) in the time-varying parameter model setting. We also apply MSDSP in the context of Bayesian predictive synthesis (BPS) ([McAlinn and West, 2019](#)), where dynamic combination schemes exploit the information from the alternative economic models. Our analysis provide a new perspective to the Meese-Rogoff puzzle, illustrating that the economic models, enhanced with the parameter flexibility of the MSDSP, produce predictive distributions that are superior to the random walk model, even when stochastic volatility is considered.

Keywords: Bayesian Econometrics, Shrinkage Methods, Sparsity, Model Combination, Variable Selection, Exchange Rate Prediction

JEL Classification: C11, C14, C52, C53

*This research was supported by The University of Melbourne’s Research Computing Services and the Petascale Campus Initiative.

[†]Department of Economics, University of Melbourne.
Email: zhengf1@student.unimelb.edu.au.

[‡]Maneesoonthorn is supported by the ARC Discovery Grant DP200101414.

[§]Yong Song thanks for the support from ARC Discovery Grant DP230100959.

1 Introduction

Forecasting exchange rate is important for policy makers, investors and international traders. The seminal Meese–Rogoff Puzzle of [Meese and Rogoff \(1983a\)](#) claims that no model can systematically beat the random walk, with subsequent literature predominantly focusing on the point forecast. Thirty years on, [Rossi \(2013\)](#) confirmed such claim with an exhaustive empirical work. On the other hand, [Rossi \(2013\)](#) also inspires that

The efficient market hypothesis does not mean that exchange rates are unrelated to economic fundamentals, nor that exchange rates should fluctuate randomly around their past values.

This work motivates economic theoretic driven studies such as [Ferraro et al. \(2015\)](#); [Cheung et al. \(2019\)](#); [Engel et al. \(2019\)](#); [Candian and De Leo \(2023\)](#); [Neghab et al. \(2024\)](#) amongst others. A more recent and comprehensive survey is conducted in [Fang et al. \(2024\)](#), who thoroughly reviewed the literature in this area. They found no influential findings after [Rossi \(2013\)](#), with the exception of [Ferraro et al. \(2015\)](#), even though many researchers have attempted to address the puzzle.

Challenges in exchange rate forecasting include: (1) dynamic instability, (2) limited sample size, and (3) model uncertainty. There are other concerns such as data frequency choice and the discussion of big and small economy, but we focus on the aforementioned three points in this paper. Our novel econometric approach addresses these three challenges and indeed provide a dominating predictive outcome in many out-of-sample metrics than the random walk. The foundation of our econometric models is built upon established economic theories, with historical research on the relationship between exchange rates and economic fundamentals respected in our framework.

From the econometric modelling perspective, many research addressed dynamic instability. Existing popular method such as time-varying parameter (TVP) models [Wolff \(1987\)](#); [Canova \(1993\)](#); [Mumtaz and Sunder-Plassmann \(2013\)](#); [Byrne et al. \(2016\)](#) cater for gradual change of the parameters and regime-switching models [Engel \(1994\)](#); [Nikolsko-Rzhevskyy and Prodan \(2012\)](#) allow for sudden parameter change. These models are flexible and well-tested in time series forecasting. However, to our knowledge, no paper has seriously addressed

the importance of model parsimony in the presence of limited sample size for exchange rate forecasting. A data-driven balance between flexibility and parsimony, in our view, could shed lights on the Meese–Rogoff prediction puzzle.

To answer challenges (1) and (2), we consider the dynamic shrinkage process (DSP) of [Kowal et al. \(2019\)](#). The DSP is a process taking the form of a state space model, with the innovation taking a particular form such that the process could experience a decent duration of large negative values. See also [Hauzenberger et al. \(2024\)](#), who applied DSP to their global shrinkage factor. When used to model the log volatility of a dynamic parameter, the DSP allows for the variation of the time-varying parameter to approach zero, reducing its temporal variation. This approach shrinks the parameters towards the value in the previous period rather than shrinking them to zero. The same idea can be found in [Dufays et al. \(2021\)](#), but their model is computationally intensive because it requires a joint movements of the parameters at adjacent time points. In contrast, the DSP works as a clip by collecting adjacent parameters into a single value through the innovation’s volatility, which is a stark contrast to the standard TVP model that assumes constant volatility. Another closely related research is [Knaus and Frühwirth-Schnatter \(2023\)](#), who proposed dynamic triple gamma prior to model innovations in the dynamic linear model framework.

In addition to shrinkage of volatility, [Huber et al. \(2021\)](#) emphasized the empirical importance of sparsity. They pointed out that setting irrelevant parameters to zero may greatly improve prediction. Motivated by [West and Harrison \(2006\)](#), [Kalli and Griffin \(2014\)](#) developed a Normal Gamma Autoregressive process (NGAR) that can mask segments of a time series of coefficients to be numeric zero. Similar ideas include the dynamic slab and spike process in [Rockova and McAlinn \(2021\)](#) and the Markov mask of latent coefficients in [Uribe and Lopes \(2020\)](#) and [Bernardi et al. \(2023\)](#).¹ [Lopes et al. \(2022\)](#), instead, use a 4-component mixture prior to select between TVP and constant coefficient models, but they assume static mixtures. There is no dynamic framework to date that allow both sparsity and (potentially) multiple regimes of non-zero, but constant, coefficient in the TVP setting.

We bridge this gap by introducing the Markov Switching DSP (**MSDSP**), which allows for the time-varying coefficient to dynamically switch between the zero-state (achieving spar-

¹The idea of sparsity via masking can be traced back to [Nakajima and West \(2013\)](#).

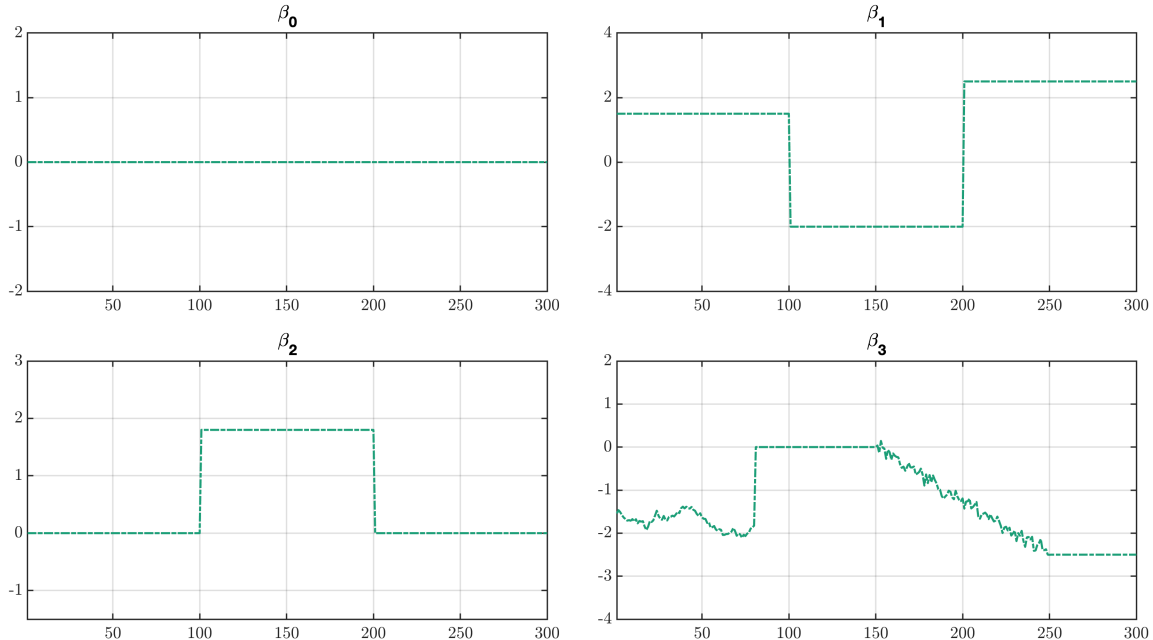


Figure 1: Four synthetic parameter paths

sity) and the DSP state (achieving shrinkage). To illustrate the flexibility of our proposed framework, we plot four synthetic parameter paths in Figure 1. The first parameter β_0 is a constant 0 which renders its corresponding covariate (or intercept) useless. The second parameter β_1 changes values twice, with all three values being none-zero. The third parameter β_2 is zero at the two ends of the illustrative period, and is only “activated” in the middle period. The last parameter β_3 experiences gradual change, becomes abruptly “inactivated” at zero, and then gradually move towards a fixed none-zero value. All these behaviors can be captured by our MSDSP framework.

In the exchange rate forecasting application, we follow the established benchmark models from Rossi (2013) and find that when we apply our proposed MSDSP method to the economic models, they substantially outperform the random walk across a range of metrics. In addition applying the MSDSP to enhance the individual models, we find that the MSDSP naturally lends itself to the Bayesian predictive synthesis (BPS). BPS is a novel and powerful model assembly method developed by McAlinn and West (2019). It takes care of potential model bias and allows unrestricted model weights (even negative) to optimally combine many sub-models or experts, see McAlinn et al. (2020a); Tallman and West (2024a)

for recent applications. In the BPS context, the MSDSP can be used directly on model weights in model assembly. These weights can move as in a state space model with parsimony (DSP-state) or simply shut down (zero-state) for a while and then reactivate (or not). We apply the MSDSP assembly method and compare it to some existing approaches. We find that the MSDSP assembly method works the best among multiple well-known assembly methods. With this, we answer to the aforementioned challenge (3).

The remainder of the paper is organized as follows. Section 2 outlines the construction of the MSDSP framework for TVP models, along with the Bayesian inference proposed for the the framework. A simulation study to assess the advocacy of the framework is provided in Section 3. In Section 4, focus is given to a detailed application of the MSDSP framework on existing economic models to address the Meese-Rogoff puzzle, while Section 5 provides an application of the MSDSP in model assembly for exchange rate forecasting. Section 6 concludes.

2 The Markov Switching Dynamic Shrinkage Process

The Markov Switching Dynamic Shrinkage Process (MSDSP) provides enhanced flexibility to the state space model that encapsulates the TVP specification in many macroeconomic models. The state-space model, tracing back to Kalman (1960), is a statistical framework for dynamic systems, with the most commonly used being the Gaussian Dynamic Linear Model (DLM) (Harrison and Stevens, 1976; West and Harrison, 2006), assuming linear relationship between the observables and the state, with normally distributed innovations. The TVP model for time series analysis belongs to this class, and can be specified as:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \quad (1)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, \Omega_t), \quad (2)$$

where the innovation ϵ_t 's variance σ_t^2 and $\boldsymbol{\omega}_t$'s covariance matrix Ω_t are pre-determined or a constant. The variable of interest y_t is a scalar, vector \mathbf{x}_t is the $p \times 1$ observed data including the intercept. The time-varying parameter $\boldsymbol{\beta}_t$ is a $p \times 1$ vector.

Contemporary macroeconomic or financial applications emphasized the time varying volatility. For example, stochastic volatility (SV) has been advocated and evidenced by numerous empirical works (Primiceri, 2005; Justiniano and Primiceri, 2008; Nakajima, 2011; Chan and Eisenstat, 2018; Chan et al., 2024). Here, we base our model on the DLM with SV specification of (Aguilar and West, 2000), who assumes (1)-(2), but with $\sigma_t^2 = \exp(g_t)$, and

$$g_t = \mu_g + \phi_g(g_{t-1} - \mu_g) + e_t, \quad e_t \sim N(0, \sigma_g^2), \quad (3)$$

The log volatility g_t follows an AR(1) process, with stationarity imposed by $|\phi_g| < 1$. In addition, we assume a diagonal matrix for Ω_t in (2) for simplicity.

2.1 TVP with MSDSP Flexibility

Our goal is to extend the TVP modelling framework to account for both sparsity and shrinkage of the coefficient to a non-zero value. To this end, we extend the work of Kowal et al. (2019), and introduce p two-state Markov switching processes to allow for the time-varying coefficients to “switch off” to zero. The model, in full, takes the form of

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_t + \epsilon_t, \quad \epsilon_t \sim N(0, \exp(g_t)) \quad (4)$$

$$g_t = \mu_g + \phi_g(g_{t-1} - \mu_g) + e_t, \quad e_t \sim N(0, \sigma_g^2) \quad (5)$$

$$\boldsymbol{\beta}_t = \mathbf{s}_t \circ \tilde{\boldsymbol{\beta}}_t + (\boldsymbol{\ell} - \mathbf{s}_t) \circ \mathbf{0} \quad (6)$$

$$\tilde{\beta}_{it} = \tilde{\beta}_{i,t-1} + \omega_{it} \quad \omega_{it} \sim N(0, \exp(h_{it})) \quad (7)$$

$$h_{i,t+1} = \mu_i + \phi_i(h_{it} - \mu_i) + \eta_{i,t+1}, \quad \eta_{i,t+1} \sim Z(\alpha_h, \beta_h, 0, 1) \quad (8)$$

$$P(s_{it} = k \mid s_{i,t-1} = j) = P_{jk}^i \quad (9)$$

for $i = 1, \dots, p$, $t = 1, \dots, T$ and $j, k \in \{0, 1\}$. The \circ in Equation (6) is the Hadamard (pointwise) product, and $\mathbf{0}$ is a $p \times 1$ vector of zeros. The introduction of the $p \times 1$ variable, $\mathbf{s}_t = (s_{1t}, \dots, s_{pt})$, allow for the elements of $\boldsymbol{\beta}_t$ to take value either the value of zero or the corresponding element of the “shadow” coefficient $\tilde{\boldsymbol{\beta}}_t$, based on the Markovian transition probability in (9). We assume that Markovian switches operate independently across the p

dimension, such that $\mathbf{s}_i \perp\!\!\!\perp \mathbf{s}_j$ for $j \neq i$ and $\mathbf{s}_i = (s_{i1}, \dots, s_{iT})$.

The element of the shadow coefficient $\tilde{\beta}_t, \tilde{\beta}_{it}$ for $i = 1, \dots, p$, follows the Dynamic Shrinkage Process (DSP) of [Kowal et al. \(2019\)](#). The DSP allows for the latent coefficient itself to exhibit SV structure through the modelling of the volatility of each innovation term ω_{it} in (7)-(8). [Kowal et al. \(2019\)](#) proposed the use of the Z-distribution as a prior in (8), with the feature of this particular distribution allowing for extreme negative values in the log volatility term h_{it} . The Z-distribution, introduced by [Barndorff-Nielsen et al. \(1982\)](#), is a flexible class of normal mixture that allows for extreme skewness and fat-tails. In the DSP context, we employ the mixture constructed using the logistic transformation of the beta variable with parameters α_h and β_h , with zero location and unit scale for identification. In its extremes, $\eta_{it} \rightarrow -\infty$ results in the log-variance $h_{j,t}$ approaches $-\infty$. In such circumstances, the volatility of $\tilde{\beta}_{it}$ approaches zero, rendering a constant coefficient environment with $\tilde{\beta}_{it} = \tilde{\beta}_{i,t-1}$.

Our proposed MSDSP specification offers additional parsimony and flexibility to the TVP model as follows. First, each element of β_t is governed by independent dynamic shrinkage and Markov switching processes, allowing for each covariate to behave independent through the sample period. The general framework also allows for the potential for the TVP to switch to zero for sparsity, to shrink to a constant value, or to revolve dynamically according to the state space structure. Our specification for sparsity and shrinkage is internally coherent, as also done in [Rockova and McAlinn \(2021\)](#), which is in contrast to the post-processing approaches of [Huber et al. \(2021\)](#); [Hahn and Carvalho \(2015\)](#), among many.

2.2 Nested and Related Models

The general MSDSP framework proposed here nests several existing models. The DSP proposed by [Kowal et al. \(2019\)](#) is its most obvious special case, which can be achieved when the transition probability $P_{jj}^i = 1$ for both $j \in \{0, 1\}$ and the two-state Markov switching process reduces to a one-state scenario. This coincide with the original TVP with DSP specification. In the case when $\alpha_h = \beta_h = 1/2$, and in the absence of dynamic structure, in the specification of the Z-distribution in (8), the model corresponds to a horse-shoe prior also discussed in [Kowal et al. \(2019\)](#). Of course, the original TVP model is also nested within our

framework, with $\beta_t = \tilde{\beta}_t$ and the latent time-varying parameter exhibit homoskedasticity. Lastly, if $\tilde{\beta}_{it}$ is a constant, the model assumes a dynamic stochastic search variable selection (SSVS) method for the corresponding variable \mathbf{x}_i .

A closely related framework is that of [Bernardi et al. \(2023\)](#), who employ one latent parameter process and one masking process, with roles similar to $\tilde{\beta}_{it}$ and s_{it} , respectively, in our framework. Their latent parameter process is a simple random walk (or some joint normal distribution) instead of DSP in this paper, so their latent model cannot afford short-term constant coefficient. The mask parameter is determined by an inverse logit function with input from another latent normal distribution, with the normality assumptions lending itself naturally the variational inference in their paper.

2.3 Bayesian Inference of the MSDSP Model

We use Bayesian computation to conduct inference on the MSDSP model. Using the Polya-gamma representation for the DSP ([Polson et al., 2013](#)) and the mixture technique for the SV specification ([Omori et al., 2007](#)), we seek to establish the augmented posterior:

$$p(\Psi \mid \mathbf{y}_{1:T}, \mathbf{x}_{1:T}), \tag{10}$$

where $\mathbf{y}_{1:T} = (y_1, \dots, y_T)^\top$, $\mathbf{x}_{1:T} = (\mathbf{x}_1, \dots, \mathbf{x}_T)$ with $\mathbf{x}_t = \{x_{it}\}_{i=1}^p$ and the model unknowns are collected in Ψ . The model unknowns include any auxiliary variables from the Polya-gamma representation and the SV mixture representation that is amendable to the Gibbs sampler in the Markov chain Monte Carlo (MCMC) sampling to estimate our posterior distribution. The elements of Ψ , along with their respective description and dimensions are summarized in [Table 1](#)

Within our MCMC algorithm, we utilize the techniques to sample the components related to the DSP component outlined in [Kowal et al. \(2019\)](#). Standard algorithms for Markov switching and SV inference are adopted. [Algorithm 1](#) outlines the steps for our MCMC inference, with full details of the algorithm and the prior distribution given in the Appendix. In all our inference, we construct $G = 4,000$ posterior samples after 30,000 burn-in draws, with every 5th draw retained through thinning. Simulation consistency inference can be

Notation	Description	Dimension
$\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1, \dots, \tilde{\boldsymbol{\beta}}_T)$ with $\tilde{\boldsymbol{\beta}}_t = \{\tilde{\beta}_{it}\}_{i=1}^p$	Shadow coefficient	$T \times p$
$\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_T)$ with $\mathbf{s}_t = \{s_{it}\}_{i=1}^p$	State indicator for each variable and time point	$T \times p$
$\mathbf{g} = (g_1, \dots, g_T)^\top$	Log volatility of the measurement equation	$T \times 1$
$\mu_g, \phi_g, \sigma_g^2$	Parameters of the log volatility process g_t	3
$P = \{P^i\}_{i=1}^p$ with $P^i \in \mathbb{R}^{2 \times 2}$	State transition probability matrices	$2p$
$H = (\mathbf{h}_1, \dots, \mathbf{h}_T)$ with $\mathbf{h}_t = \{h_{it}\}_{i=1}^p$	DSP-related log volatility	$T \times p$
$\boldsymbol{\mu} = \{\mu_i\}_{i=1}^p, \boldsymbol{\phi} = \{\phi_i\}_{i=1}^p$	Parameters associated with h_{it}	$2p$
$\boldsymbol{\xi}_\mu = \{\xi_{\mu,i}\}_{i=1}^p$	Auxiliary variable for sampling μ_i	p
$\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)$ with $\boldsymbol{\xi}_t = \{\xi_{it}\}_{i=1}^p$	Auxiliary variable for sampling h_{it}	$T \times p$
$\mathbf{r} = \{r_t\}_{t=1}^T$	Auxiliary variable for sampling g_t	$T \times 1$

Table 1: Summary of the elements of Ψ from the Markov Switching Dynamic Shrinkage Process model

carried out from our G this posterior samples. For instance, if the posterior expected value of $\boldsymbol{\beta}_t$, or simply $E(\boldsymbol{\beta}_t \mid \mathbf{y}_{1:T}, \mathbf{x}_{1:T})$, can be estimated by the posterior sample mean $\frac{1}{G} \sum_{g=1}^G \boldsymbol{\beta}_t^{(g)}$.

Algorithm 1 MCMC. After initialization, repeat the following steps:

- (a) Sample from $\tilde{\boldsymbol{\beta}} \mid \mathbf{y}, \mathbf{X}, \mathbf{s}, \mathbf{g}$.
- (b) Sample $\mathbf{s} \mid \mathbf{y}, \mathbf{X}, \tilde{\boldsymbol{\beta}}, \Sigma_\epsilon, P$ by using FFBS method.
- (c) Update $\boldsymbol{\beta}$ from Equation (6).
- (d) Sample $\mathbf{g} \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{r}, \mu_g, \phi_g, \sigma_g^2$.
- (e) Sample $\mu_g, \phi_g \mid \mathbf{g}, \sigma_g^2$ and then $\sigma_g^2 \mid \mathbf{g}, \mu_g, \phi_g$.
- (f) Sample $P \mid \mathbf{s}$.
- (g) Sample $H \mid \tilde{\boldsymbol{\beta}}, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\xi}$.
- (h) Sample $\boldsymbol{\mu} \mid H, \boldsymbol{\phi}, \boldsymbol{\xi}_\mu, \boldsymbol{\xi}$ and $\boldsymbol{\phi} \mid H, \boldsymbol{\mu}, \boldsymbol{\xi}$.
- (i) Sample $\boldsymbol{\xi}_\mu \mid H, \boldsymbol{\phi}, \boldsymbol{\xi}$.
- (j) Sample $\boldsymbol{\xi} \mid H, \boldsymbol{\mu}, \boldsymbol{\phi}$.
- (k) Sample $\mathbf{r} \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{g}$.

2.4 Direct h -step-ahead Forecasts

In the application of exchange rate forecasting, our primary focus is to construct the predictive distribution for future exchange rates. The model specified in Section 2.1 specifies a contemporaneous regression model in (4). Without an explicit dynamic for the covariate, x_t , constructing predictions for future time points is not feasible. In order to construct the h -step-ahead predictive distribution, we propose the direct h -step-ahead forecasting model

by adjusting the measurement equation (4) to

$$y_{t+h} = \mathbf{x}'_t \boldsymbol{\beta}_t + \epsilon_t, \quad (11)$$

with $\epsilon_t \sim N(0, \exp(g_t))$ and the remainder of the model components remain as defined in (5)-(9).

The out-of-sample h -step-ahead posterior predictive distribution, integrating out the uncertainty in the model unknowns, takes the form of

$$p(y_{T+h} | I_T) = \int p(y_{T+h} | x_T, \boldsymbol{\beta}_T, g_T) p(\Psi | \mathbf{y}_{h+1:T}, \mathbf{x}_{1:T-h}) d\Psi, \quad (12)$$

where $I_T = (\mathbf{y}_{1:T}, \mathbf{x}_{1:T})$ denotes the information set up to time T . By construction, the predictive distribution, even for $h > 1$, can be constructed using the covariate observed within the sample period. In addition, the implied TVP model structure, captured by $p(y_{T+h} | x_T, \boldsymbol{\beta}_T, g_T)$, reflects the specific dynamic relationship between the exchange rate and the h -lag covariate without the need to further assume the evolution of such relationship between periods T and $T + h$.

Note that when applying the model in (11), the posterior distribution $p(\Psi | \mathbf{y}_{h+1:T}, \mathbf{x}_{1:T-h})$ needed in (12) only include the inference of the latent quantities up to time point $T - h$. That is, we obtain G posterior draws of $(\mathbf{s}_{T-h}, \tilde{\boldsymbol{\beta}}_{T-h}, \mathbf{h}_{T-h}, g_{T-h})$. In order to produce the h -step-ahead prediction, we simulate forward using the dynamic specification in (5)-(9) to obtain G draws of $\boldsymbol{\beta}_T$ and g_T , upon which $p(y_{T+h} | x_T, \boldsymbol{\beta}_T, g_T)$ conditions on. By simulation, we obtain G draws of y_{T+h} , $\{y_{T+h}^{(g)}\}_{g=1}^G$, from its conditional predictive distribution. All predictive statistics in the application can be inferred from this framework. For example, the expected value $E(y_{T+h} | I_T)$ can be estimated by the sample mean $\frac{1}{G} \sum_{g=1}^G y_{T+h}^{(g)}$. The predictive density $p(y_{T+h} | I_T)$ can be estimated by the sample mean of the Gaussian conditional densities implied by (11) as $\frac{1}{G} \sum_{g=1}^G p(y_{T+h} | x_T, \boldsymbol{\beta}_T^{(g)}, g_T^{(g)})$, evaluated over the plausible grid value of y_{T+h} . We evaluate the prediction using five metrics geared to evaluate the predictive distribution. These metrics are defined and discussed in Section 4.2.

3 Simulation Study

We carried out two simulation studies to illustrate the effectiveness of the MSDSP framework. In particular, we highlight the ability of the MSDSP in flexibly capturing multiple regimes, with the posterior inference of these dynamic parameters efficient for both sudden and gradual shifts in the TVP.

3.1 Case 1: Abrupt Shifts in TVP

We consider a scenario with 5 explanatory variables and an intercept, for $T = 300$ periods. The explanatory variables $\mathbf{x}_t = (1, x_{1t}, \dots, x_{5t})$ are generated by following the structure from [George and McCulloch \(1997\)](#). First we generate six independent standard normal random variables, z_0, \dots, z_5 . We generate five correlated covariate variables by $x_{1t} = 0.4z_0 + z_1$, $x_{2t} = x_{1t} + 1.8z_2$, $x_{3t} = 0.4z_0 + z_3$, $x_{4t} = x_{3t} + 1.4z_4$, and $x_{5t} = x_{3t} + 1.4z_5$, resulting in the covariate vector \mathbf{x}_t that are independent across time t . The dependent variable y_t is generated by $y_t = \mathbf{x}'_t \boldsymbol{\beta}_t + \epsilon_t$, where the error term ϵ_t is independently drawn from a standard normal distribution. The coefficients, $\boldsymbol{\beta}$, are chosen such that some of the parameter values abruptly changes between different regimes of constant values or zero. In particular, β_0 is set to zero, rendering a zero intercept. β_2 and β_5 are both zero across all time points, rendering zero impact from the covariates x_{2t} and x_{5t} . β_1 has 3 regimes, with the parameter values being constant but non-zero in each regime. β_3 has 3 regimes, with the parameter taking non-zero constant values in the first two regimes before switching to a zero value in its last regime. β_4 also exhibit 3 regimes, but with the second of three regimes being the zero-state, while the other two being non-zero constants, implying that the impact x_{4t} is switched on and off. The dynamic of the true parameter values are shown in the green dashed line in the top and middle panels of [Figure 2](#)

[Figure 2](#) depict the comparison of posterior inference of the TVP between the DSP and the MSDSP models. The top panel shows the posterior means (solid yellow line) and point-wise 95% density intervals (gray shade) inferred from the DSP model. The middle panel depicts the posterior means (solid yellow line) and point-wise 95% density intervals (gray shade) inferred from the MSDSP model. The bottom panel plots the true Markov switching

state \mathbf{s}_t (green dashed line) and the posterior probability of being in the non-zero DSP-state (yellow solid line). In inference of both the DSP and the MSDSP models, we use identical uninformative prior specification for the parameters related to the DSP component for comparability.

From the bottom panel in Figure 2, the true state and the inferred state probability almost coincide, which indicates an almost perfect ability of the MSDSP model to switch from the zero state to the DSP state. Comparing the top and middle panels, we observe that when a coefficient’s true value is in the zero-state, the MSDSP will quickly adapts to that state. This is particularly evident in the inference of β_0, β_2 and β_5 , where the posterior uncertainty is much smaller for the MSDSP. Since the DSP is unable to shrink to absolute zero-state, the the posterior inference over these true zero-states still exhibit a larger degree of uncertainty, even when the posterior means are hovering around zero. Similarly, the zero-state episodes for β_3 and β_4 are well captured with much tighter posterior intervals when the MSDSP is used. For non-zero coefficient values, the MSDSP and DSP performs similarly, with both models able to capture the shifts of the coefficients to a non-zero constant value effectively.

3.2 Case 2: Gradual Shifts in TVP

For this case, we simulate three explanatory variables, $\mathbf{x}_t = (1, x_{1t}, \dots, x_{3t})$, from independent standard normal distributions with $T = 300$. As in Case 2, \mathbf{x}_t ’s are independent across time t . The coefficients, β , are given in Figure 1 and described in the Introduction. They are also shown as the green dashed lines in the top and middle panels in Figure 3. Again, the dependent variable y_t is generated by $y_t = \mathbf{x}_t' \beta_t + \epsilon_t$, where the error term ϵ_t is independently drawn from a standard normal distribution.

This collection of coefficients encompasses all possible scenarios that the dynamic TVP, and we compare the DSP and MSDSP models in this setting. We highlight that the difference between Case 1 and Case 2 is that Case 2 exemplifies the gradual change in parameter in the scheme for β_3 . As in Figure 2 for Case 1, Figure 3 depicts the posterior mean and point-wise 95% posterior interval of β_t , with those produced by the DSP model depicted in the top panel, and those from the MSDSP depicted in the middle panel. The bottom panel

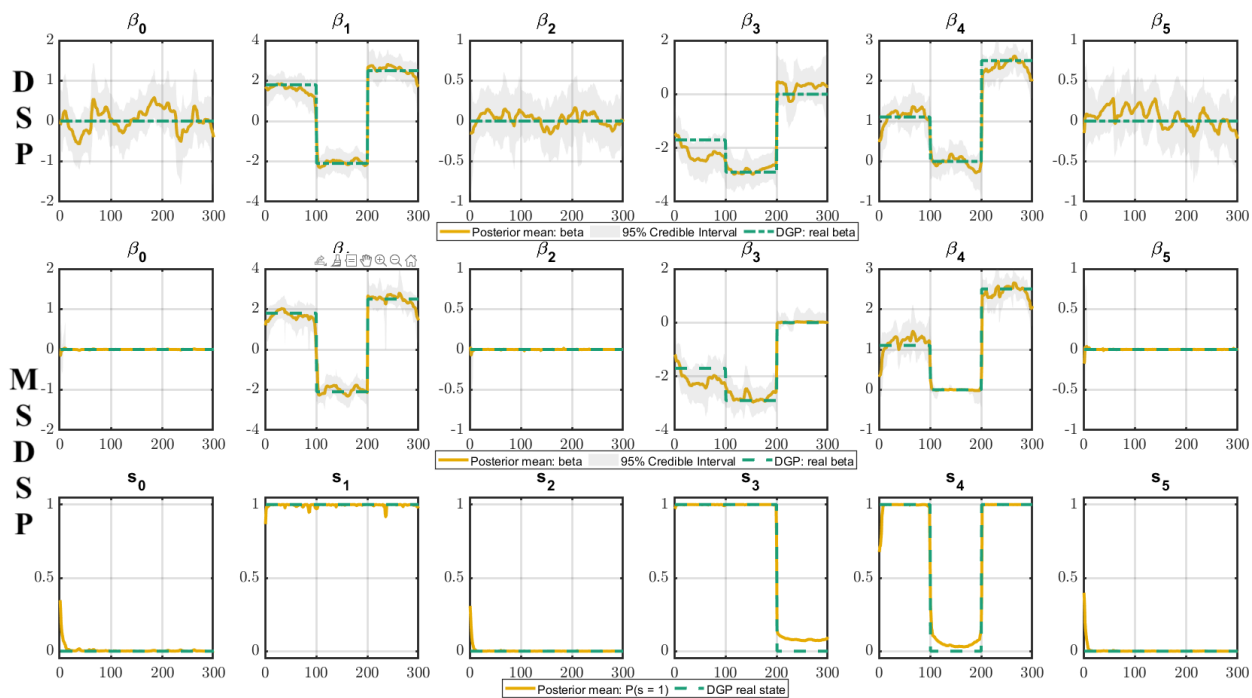


Figure 2: Posterior estimates from DSP and MSDSP models on DGP 1. The first two rows show the posterior mean and 95% credible intervals of β_t from DSP and MSDSP, respectively. The green dashed lines are the true values. The bottom row shows the posterior probabilities of being in the DSP-state in the MSDSP model. The green dashed lines are the indicators of the true DSP-state.

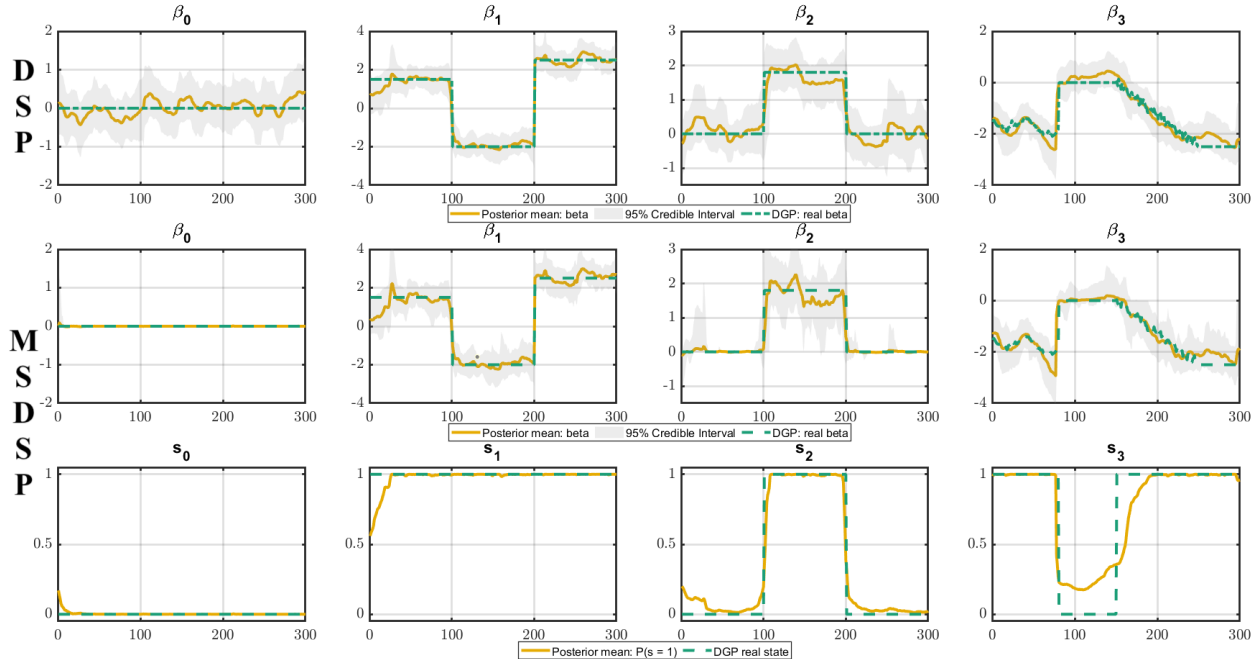


Figure 3: Posterior estimates from DSP and MSDSP models on DGP 2. The first two rows show the posterior mean and 95% credible intervals of β_t from DSP and MSDSP, respectively. The green dashed lines are the true values. The bottom row shows the posterior probabilities of being in the DSP-state in the MSDSP model. The green dashed lines are the indicators of the true DSP-state.

depicts the posterior summary for the Markov switching variable s_t . Consistent with what we observed in Case 1, when the coefficient enters the zero-state, the MSDSP model is able to capture this scenario with a much higher degree of certainty. When the parameters are non-zero, the MSDSP and DSP models perform similarly, including when the shift in the parameter is gradual, as in the case of β_3 .

4 Addressing the Meese-Rogoff Puzzle with MSDSP

We employ the flexible MSDSP model to offer a novel econometric perspective on the Meese-Rogoff puzzle. In particular, we seek to establish if the model's ability to achieve sparsity and incorporate dynamic parameters that can potentially shrink toward a constant makes economic models competitive with random walk predictions. To set notation, let e_t denote the logarithm of spot change rate between a country to the US dollar at time t , and we construct the response variable as $y_t = e_t - e_{t-1}$ as the growth rate of exchange rate at time

t . We use the direct quotation method, hence a positive y_t means that the domestic currency depreciates against the US dollar.

The Meese–Rogoff Puzzle (Meese and Rogoff, 1983a,b, 1988) claimed that the random walk model is a competitive model that is difficult to beat for exchange rate prediction. Therefore, we view the random walk as the benchmark model and augment it with a SV component to reflect the contemporary development in the empirical macroeconomic modeling literature. Specifically,

$$y_t = \mu + \epsilon_t, \quad \epsilon_t \sim N(0, \exp(g_t)), \quad (13)$$

$$g_t = \mu_g + \phi_g(g_{t-1} - \mu_g) + e_t, \quad e_t \sim N(0, \sigma_g^2). \quad (14)$$

If we restrict the $\mathbf{x}_t = 1$ in the measurement equation, (4), and assume a constant intercept, the MSDSP model reduces to the random walk model. We use both the random walk model with non-zero drift (**RW-drift-SV**) and the driftless random walk where $\mu = 0$ (**RW-SV**). As pointed out by Rossi (2013), the RW-SV is the toughest benchmark to beat. Since both RW-drift-SV and RW-SV models contains no external covariates, we apply the indirect method to recursively construct the h -period ahead forecast from these models.

4.1 Competing Economic Models with MSDSP

We respect the existing literature, as summarized by Rossi (2013), to include four different models driven by economic theory, along with three further predictive models based on commodity prices (oil, gold, and copper) inspired by Ferraro et al. (2015). All seven competing economic models can be written in the form of a linear model in (11), but with varying set of covariates. We summarize the specification of the covariate dictated by each of the economic model in Table 2.

The adoption of MSDSP (and its nested DSP) in the time-varying coefficient in these models allow for the coefficient to move over time, as well as allow for stochastic volatility in the predictive distribution of exchange rate changes. Both features in the econometric modelling of these economics theories have been tested in the contemporaneous macroeconomic literature. We employ both the MSDSP and the DSP specifications on all seven competing

Model	Acronym	\mathbf{x}_t
Interest Rate Parity	IRP	$(1, (r_t - r_t^*))$ $r_t - r_t^* =$ difference in log interest rates
Taylor Rule	TR	$(1, (r_{t-1} - r_{t-1}^*), [1.5(\pi_t - \pi_t^*) + 0.5(x_t - x_t^*)])$ $\pi_t - \pi_t^* =$ difference in inflation rates $x_t - x_t^* =$ difference in output gaps
Purchasing Power Parity	PPP	$(1, (p_t - p_t^* - e_t))$ $p_t - p_t^* =$ difference in log price levels $e_t =$ exchange rate at time t
Monetary Model	MON	$(1, [(m_t - m_t^*) - (g_t - g_t^*) - e_t])$ $m_t - m_t^* =$ difference in log money supply $g_t - g_t^* =$ difference in log real output
Commodity Models	Oil Gold Copper	$(1, (p_t^{oil} - p_{t-1}^{oil}))$ $(1, (p_t^{gold} - p_{t-1}^{gold}))$ $(1, (p_t^{copper} - p_{t-1}^{copper}))$ p_t^x denotes the log price of commodity ‘ x ’ at time t

Table 2: Summary of the competing economic models. The model specification follows (4)-(9), with the set of \mathbf{x}_t changing according to the economic model framework. All quantities denoted with ‘*’ corresponds to those of the U.S.

economic models listed in Table 2. Our analysis provide a new perspective to the Meese-Rogoff puzzle, with the analysis allowing for additional flexibility in the economic models to allow them to adapt to changing economic environments.

The two random walk models, RW-SV and RW-drift-SV, along with the constant coefficient versions of the seven competing economic models, serve as benchmarking models in our analysis. We refer to the constant coefficient economic models simply as “linear” models. For the linear economic models, we consider expanding window as well as moving window of 100 months and 50 months to capture potential structural instability brutally. All models, including the RW-SV, RW-drift-SV and the linear benchmark models, are inferred in the Bayesian framework. Hence, we take account the parameter uncertainties and are able to provide a full profile of the predictive distribution, including densities and point forecasts.

We evaluate our predictive distributions over the forecast horizon $h = 1, \dots, 12$. Since we employ monthly data, these horizons corresponds to from 1 month to 1 year. All predictions are out-of-sample by using expanding as delineated in Section 4.4, unless stated otherwise. For completeness, we also report the contemporaneous forecasts conducted in Ferraro et al.

(2015), which is equivalent to predicting y_{t+h} by assuming x_{t+h} exists, in the appendix. ²

4.2 Forecasting Metrics

We consider **five** metrics to evaluate the predictive distributions of the competing models: log predictive likelihood (Bayes factor), continuous ranked probability score (CRPS), root mean squared forecast error (RMSFE), tail coverage rate, and model confidence set (MCS). All metrics are out-of-sample and calculated based on the Monte-Carlo estimate of (12), as described in Section 2.4. In what follows, we define T_0 as the initial sample period, and T denotes the overall sample size, with the number of h -period ahead periods evaluated being $T - h - T_0 + 1$.

First, the predictive distribution is evaluated by the log predictive density ratio (LPDR), which evaluates the predictive performance of model \mathcal{M} relative to the benchmark model \mathcal{M}_B :

$$\mathbf{LPDR}(\mathcal{M}, \mathcal{M}_B) = \log(p(Y_{T_0+h:T} | I_{T_0}, \mathcal{M})) - \log(p(Y_{T_0+h:T} | I_{T_0}, \mathcal{M}_B)). \quad (15)$$

Here,

$$\log(p(Y_{T_0+h:T} | I_{T_0}, \mathcal{M})) = \sum_{t=T_0}^{T-h} \log(p(y_{t+h} | I_t, \mathcal{M}))$$

is the cumulative log score, with $\log(p(y_{t+h} | I_t, \mathcal{M}))$ denoting the log predictive density of the h -period ahead prediction constructed based on information set up to time t . The LPDR is promoted by Geweke and Amisano (2010), and can be viewed as the log predictive Bayes factor of model \mathcal{M} relative to the benchmark \mathcal{M}_B . A value that is larger than 3 means that model \mathcal{M} is strongly preferred to the benchmark model \mathcal{M}_B (see (Kass and Raftery, 1995)). With the RW-SV being identified by Rossi (2013) as being the toughest model to beat, the RW-SV model serves as our \mathcal{M}_B for all subsequent analyses.

The **continuous ranked probability score** (CRPS) assesses the full predictive distribution prioritizing the sharpness of the distribution and is thus less sensitive to tail behaviors

²This approach is a pseudo prediction since it ignores simultaneity. It is not a fair comparison to the random walk model, but can provide information for comparison between economic models by stripping off predictor's uncertainty.

(Gneiting and Raftery, 2007). It is defined, for model \mathcal{M} , as

$$\mathbf{CRPS}_{t+h}(p(y_{t+h} | I_t, \mathcal{M}), y_{t+h}) = \int_{-\infty}^{\infty} (\mathbf{1}(y_{t+h} \leq r) - F(r | I_t, \mathcal{M}))^2 dr,$$

where $F(r | I_t, \mathcal{M}) = \int_{-\infty}^r p(\tilde{y}_{t+h} | I_t, \mathcal{M}) d\tilde{y}_{t+h}$ is the CDF for y_{t+h} and $\mathbf{1}(\cdot)$ is the indicator function which returns one when the condition in the parenthesis is true. We report the average CRPS

$$\mathbf{CRPS}_{T_0+h:T}(\mathcal{M}) = \frac{1}{T - T_0 - h + 1} \sum_{t=T_0}^{T-h} \mathbf{CRPS}(p(y_{t+h} | I_t, \mathcal{M}), y_{t+h}). \quad (16)$$

for model comparison. A smaller value means better performance.

For point forecast evaluation, the **root mean squared forecast error** (RMSFE) is one standard approach, and computed as follows:

$$\text{RMSFE}_{T_0+h:T} = \sqrt{\frac{1}{T - T_0 - h + 1} \sum_{t=T_0}^{T-h} (y_{t+h} - E(y_{t+h} | I_t, \mathcal{M}))^2}, \quad (17)$$

where the conditional mean $E(y_{t+h} | I_t, \mathcal{M})$ is calculated by the sample average of the corresponding set of $\{y_{t+h}^{(g)}\}_{g=1}^G$.

To evaluate the tail behavior, we report the **coverage rate** (CR) at different tail quantiles. It is the ratio of the number of observations falling into a specified range to the total number of observations. For example, to assess the predictive 5% quantile, the coverage rate will be the number of observations that are below the predictive 5% quantile value to the total number of observations. Obviously, a model with a ratio closer to 5% would be favored in this example. In general, we define, for the lower-tailed $\alpha\%$ quantile prediction,

$$\mathbf{CR}(\alpha, \mathcal{M}) = \frac{1}{T - T_0 - h + 1} \sum_{t=T_0}^{T-h} \mathbf{1}(y_{t+h} \leq q_{t+h}^{\alpha, \mathcal{M}}), \quad (18)$$

where $q_{t+h}^{\alpha, \mathcal{M}}$ is the predictive quantile from the predictive distribution, such that $Pr(y_{t+h} \leq q_{t+h}^{\alpha, \mathcal{M}} | I_t, \mathcal{M}) = \alpha$. Analogously, it is straightforward to calculate the upper-tailed quantile coverage rate, with the “ \leq ” sign replaced with the “ \geq ” sign in all calculations above.

Lastly, we construct the **Model Confidence Set** (MCS, (Hansen et al., 2011)), and

executed in (Bernardi and Catania, 2018)) to select a list of ‘superior’ models. The MCS controls for multiple comparisons to avoid inflated Type I error and is robust to model misspecification, especially when evaluating out-of-sample predictive performance. We implement the MCS using three different out-of-sample loss functions for MCS: Squared Forecast Errors (SFE), negative Log Score, and Continuous Ranked Probability Score (CRPS). The MCS produces a set of “competitive” models within the given confidence set along with their respective rankings, with uncompetitive models discarded. We choose a 95% model confidence set in our application.

4.3 Data Description

With the United Kingdom and the USA being key international financial market players, we present our application using the United Kingdom’s British Pound (GBP/USD) exchange rate and discuss the results comprehensively. We also analyze the cases of the Canadian dollar (CAD/USD) and Japanese Yen (JPY/USD) exchange rates, but their results are provided in the appendix due to the volume of results from the variety of models investigated.

Following Rossi (2013) and Aristidou et al. (2022b,a), we collected data from the following primary sources: FRED (Federal Reserve Economic Data), the IMF (International Monetary Fund), the OECD (Organisation for Economic Cooperation and Development), and the Bank of England. We employ the end-of-month nominal exchange rate for GBP/USD. As noted by Rossi (2013), end-of-month data generally exhibits weaker predictive power, providing a lower bound for forecasting accuracy. The monthly total industrial production is used as a proxy for real GDP, while the real potential GDP is derived by using the one-sided HP-filter. The other times series include the Producer Price Index (PPI), monetary aggregates (M1), and commodity prices for crude oil, gold, and copper. All variables are standardized to have a mean of 0 and a variance of 1. Detailed data sources and series codes are provided in section A of the supplementary material.

Due to the availability of commodity prices and interest rates, the longest common sample period spans from January 1990 to June 2017, totaling $T=330$ months. The out-of-sample period is from Nov 2000 to June 2017, comprising 200 observations, leaving the initial sample period $T_0 = 130$ observations.

4.4 Analysis of Predictive Performance

We first summarize that the economic models with the MSDSP flexibility perform the best in every out-of-sample evaluation metric. It consistently demonstrates superior forecasting performance compared to the random walks, producing higher LPDR, lower CRPS, lower RMSFE, more accurate coverage rate, as well as always attaining high ranks in the MCS assessment. At the same time, our results are also consistent with the literature, in that the random walk models are superior to the linear economic models (expanding or rolling window) in general. While there are a few exceptions to this general conclusion, but they are not systemic do not alter our overall conclusions that the added flexibility from the MSDSP prior indeed lead to better predictive performance of the economic models. Our observations also extend to the cases of Canadian dollar and Japanese Yen, with their detailed results provided in sections D and E of the supplementary material, respectively. In what follows, we discuss the GBP/USD results in more details.

4.4.1 Density Forecasts

Table 3 reports the LPDR of the competing models, evaluated relative to the RW-SV benchmark, at selected horizons of 1, 3, 6 and 12 months. In each row, the bolded statistic indicated the best predictive score, with positive numbers indicating that particular model outperforms the benchmark. It is evident that the MSDSP version of each economic model is always the best performing, with the MSDSP column always bolded. The results for other forecast horizons are provided in section C of the supplementary material and are consistent with those reported in Table 3.

Importantly, the first three columns under “Linear” always produce negative LPDR, indicating that the basic economic models are always worse than the random walk, an observation that is consistent with the Meese–Rogoff Puzzle. Once the MSDSP is implemented, however, the results are overthrown. This is the decisive evidence that MSDSP can discover the nonlinearity and dynamic instability to improve exchange rate prediction. In addition, the predictive LPDR from the DSP column are all negative as well, highlighting the value added from the Markov switching process that we proposed, suggesting that the DSP itself is

not parsimonious enough to produce a competitive density prediction relative to the RW-SV.

The above interpretation of results in Table 3 is convincing, and more than sufficient to provide evidence for the predictive dominance of our MSDSP framework. Convincingly, the MSDSP outperforms the RW-SV at all horizon and with all economic models. Based on the magnitude of the LPDR itself, we can also infer the degree of the superiority of the predictive performance. For example, when $h = 1$, the best performing model is the MSDSP-PPP, which is the time varying parameter PPP model with an MSDSP prior. The LPDR is larger than 3 and hence it is strongly supported by data against the random walk model RW-SV. In fact, the MSDSP-PPP model strongly dominates the random walk at all horizons, providing a strong signal that MSDSP captures important data dynamic features. These observations are also confirmed by the analysis of CRPS metric, reported in Table 4, with relative differences in this metric also confirming the superiority of the MSDSP models relative to the RW-SV, as well as its nested alternative the DSP specification.

From the angle of tail prediction, we report the 1-period ahead coverage rate for top and bottom 2.5% and 10%, respectively, in Table 6³. The MSDSP does not perform ideally in bottom 2.5% and top 10%, but it is much better than the RW-SV in all other quantile regions. The MSDSP models also generally produces more accurate coverages than their linear counterparts. We observe that the RW-SV tends to provide tail predictions that produce under-coverage. For example, in the top panel (Top 2.5%) of Table 6, the coverage rate from MSDSP is very consistent to 2.5%, but the random walk only covers 1.5%. In the third panel (Top 10%), the coverage rate from MSDSP is about 8%, but the random walk benchmark is only 4.5%. In summary, the MSDSP still performs better than the RW-SV.

4.4.2 Point Forecasts

The root mean squared forecast error (RMSFE) is the most common evaluation prediction metric for point forecasts, especially in the frequentist approach. The RW-SV model usually has its home advantage in this metric (Rossi, 2013). We are excited to reveal that the MSDSP performs extremely well with point forecasts according to the RMSFE. Table 5 shows the ratio of RMSFE of the competing models relative to that of the RW-SV benchmark model.

³Due to page limitation, other horizon's results are available upon request.

A value less than 1 means that the corresponding model produces smaller RMSFE statistic than the RW-SV benchmark.

With only one exception, the MSDSP model produces smaller RMSFE statistics compared to the RW-SV model. The only exception is the linear economic model using copper prices, based on a 50-period rolling window scheme. Even in this case, the MSDSP method is still better than the RW-SV, and is only slightly worse than the linear counterpart. As in the case of density forecast, the MSDSP dominates the RW-SV benchmark, but the RW-SV is better than linear or DSP economic models in most cases. We also observe that the linear models based on commodity prices are also capable of performing better than the RW-SV model in point forecasting, although this cannot be generalized to density predictions.

We carried out pairwise Diebold-Mariano (DM) tests for all the models against the RW-SV. Unfortunately, we did not achieve many significant result despite the MSDSP column in Table 5 indicating the direction for better performance. We also performed the Wilcoxon signed-rank test, which is a joint test based on the DM statistics from for all pairs of MSDSP models and the RW-SV benchmark ($7 \times 12 = 84$). The joint test reports the p-value of 0.0000, indicating that there is indeed a significant difference between the RMSFE of the MSDSP models compared to the RW-SV. We also performed the same Wilcoxon signed-rank test for each linear economic model with its MSDSP counterparts (12 statistics/horizons each). For all variants of the model, the test reports p-values less than 0.001, with exception of the model that involve gold prices returning a p-value of 0.017. This reflects a strong evidence that the superior performance of the MSDSP is supported by the data, even for point forecast.

4.4.3 Model Confidence Set and Rankings

Table 7 reports the 95% Model Confidence Sets for $h = 1$ and $h = 2$ months ahead forecasts based on three evaluation criteria: Squared Forecast Error (SFE), negative Log Score, and the Continuous Ranked Probability Score (CRPS). Each column report integer rank of each of the model included in the 95% model confidence set, for the corresponding selection criterion and forecasting horizon. Symbol “-” indicates that the model is eliminated from the confidence set, which implies that the models are not ranked.

The MSDSP is always the number 1 choice in the confidence sets. With only two ex-

ceptions, the MSDSP is always among top-10 choices. Considering there are 7 MSDSP economic models, this is a striking result. Amazingly, at horizon $h = 1$, all three criteria admit that MSDSP method dominate any other models, since the seven MSDSP economic models ranks the top 7 in each of the model confidence sets. Although the RW-SV and RW-drift-SV models are not rejected and are retained in the confidence sets, their highest ranking is 8.

Once again, we observe that the economic models, along with the flexibility of the MSDSP, beats the RW-SV benchmark. Our MCS results also confirm prior discussions on the DSP model performance being inferior, and are often eliminated from the confidence set, highlighting the added value of Markov switching mechanism in exchange rate prediction. We also confirm the superiority of the RW-SV and RW-drift-SV compared to the linear economics models, with the benchmark models' rankings consistently higher than those of the linear models.

5 Model Assembly with MSDSP

Since there has been weak consensus on the dominant model choice in exchange rate forecasting, a natural path is to combine the predictions via model assembly. Researchers applied the Bayesian Model Averaging (BMA) ([Wright, 2008](#); [Della Corte et al., 2009](#)) or dynamic Bayesian Model Averaging (DBMA) ([Časta, 2024](#)) in exchange rate forecasting, combining many weak models to form a strong prediction. We do not intend to exhaustively explore the model combination method due the space limitation and our focus on MSDSP. We consider each prediction from a small model can be viewed as an expert's opinion, and propose a new model assembly method that naturally adapts to the MSDSP setting based on dynamic Bayesian predictive synthesis (DBPS) from [McAlinn and West \(2019\)](#). The adaptation of the MSDSP allows for the predictive synthesis to dynamically adapt, with shrinkage of the combination parameters to a constant or a switch to no contribution in a single framework. We investigate if the use of MSDSP in this setting also leads to improved predictive performance in exchange rates.

5.1 Dynamic Bayesian Predictive Synthesis

We build our method on the dynamic Bayesian predictive synthesis from [McAlinn and West \(2019\)](#). There is a burgeoning literature as [Johnson \(2017\)](#); [McAlinn et al. \(2020b\)](#); [Aastveit et al. \(2023\)](#); [Tallman and West \(2024b\)](#). The DBPS of [McAlinn and West \(2019\)](#) can be expressed as follows

$$y_t = \omega_{t,0} + \sum_{j=1}^L \omega_{t,j} \hat{y}_{t,j} + e_t, \quad e_t \sim N(0, v_t), \quad (19)$$

$$\omega_t = \omega_{t-1} + w_t, \quad w_t \sim N(0, v_t W_t), \quad (20)$$

where v_t is defined via a standard beta-gamma random walk volatility model ([Prado and West, 2010](#)) and W_t follows a standard single discount factor specification ([West and Harrison, 1997](#)). Each $\hat{y}_{t,j}$ is the prediction from model (or expert) j . The dynamic BPS considers distribution input, hence $\hat{y}_{t,j}$ is treated as random. The vector $\omega_t = (\omega_{t,0}, \omega_{t,1}, \dots, \omega_{t,L})'$ is the time-varying intercept and weights, with L here denoting the number of models or experts. The intercept $\omega_{t,0}$ captures the systemic time-varying bias. There is no restrictions to the model weights ω_t . The absence of any restrictions provides a great deal of flexibility. For example, it allows a “bad” model to provide “good” results, as a negative weight is also permitted. It can also take advantage of a systemically upward biased model by giving it a discount, resulting the sum of model weights being less than unity.

The conditional posterior distribution of each model’s prediction $\hat{y}_t = (\hat{y}_{t,1}, \dots, \hat{y}_{t,L})'$ is updated as is given by:

$$p(\hat{y}_t | Y_{1:t}, \omega_t, M_{1:L}) \propto \alpha(y_t | \omega_t, \hat{y}_t) \prod_{j=1}^L p(\hat{y}_{t,j} | y_{1:t-1}, M_j)$$

where $\alpha(y_t | \omega_t, \hat{y}_t)$ is the density implied by (19), which is named as *synthesis function*. The predictors $\hat{y}_{t,j}$ are sampled from this conditional kernel. This sampling approach can be considered as tilting the originally standalone predictive distribution of model j by the observations via the synthesis function and associated interactions with other models.

Conditional on a draw of \hat{y}_t , standard MCMC can be applied. To ensure reliability and

consistency, both inference and prediction procedures are implemented by using the code provided on McAlinn’s website. The hyper-parameters of the DBPS such as the discount factor are set the same as [McAlinn and West \(2019\)](#). We have also performed robustness checks on these hyper-parameters and results are available upon request.

5.2 Application of MSDSP and DSP

The idea of DBPS is very appealing to practitioners. But we do not want to increase additional computational cost on simulating \hat{y}_t . Hence, instead of inputting predictive distributions, we only input one prediction value of \hat{y}_t such as the predictive mean and treat it as the data. This would free the cost of computation by make the assembly a two-stage task. In stage 1, we compute the prediction from each model. In stage 2, assemble these models by the synthesis function.

Our approach is closer to the pooling method such as static pooling ([Hall and Mitchell, 2007](#); [Geweke and Amisano, 2011](#)) and dynamic pooling ([Waggoner and Zha, 2012](#); [Del Negro et al., 2016](#)). Some other variants include two-state Markov process pooling by [Waggoner and Zha \(2012\)](#) and the infinite-state Markov process pooling by [Jin et al. \(2022\)](#). There are two key differences. First, we use the MSDSP as the weight dynamics, so that sparsity is explicit. Second, we use the DBPS idea to incorporate the bias term in the synthesis function. Our method is also close to ensemble learning via stacking in machine learning ([Wolpert, 1992](#); [Breiman, 1996](#); [Dietterich, 2000](#); [Proscura and Zaytsev, 2022](#); [Muslim et al., 2023](#)).

The specifics are expressed as follows.

$$y_t = \omega_{t,0} + \sum_{j=1}^L \omega_{t,j} \hat{y}_{t,j} + e_t, \quad e_t \sim N(0, e^{v_t}), \quad (21)$$

$$v_t = \mu_v + \phi_v(v_t - \mu_v) + e_t^v, \quad e_t^v \sim N(0, \sigma_v^2) \quad (22)$$

where $\omega_{t,0}$ is the time-varying intercept and $\omega_{t,j}$, for $j = 1, \dots, L$, is the time-varying weight for j -th model’s prediction as in DBPS.⁴ The difference in (21) is that we treat $\hat{y}_{t,j}$ as data

⁴We double use the notation ω , which also exists in (7), to represent the model weights. Since Equation (7) is not explicit in this section, we hope readers can understand that these two ω ’s are different.

in MSDSP assembly. In this paper, we use out-of-sample predictive mean as $\hat{y}_{t,j}$ while other options such as the median is also feasible. One can also input multiple predictive values from one model to exploit various information content in our framework, because the computation is parallelizable and the sparsity is explicit. The volatility of y_t is set as the standard SV in (22). The dynamic of each $\omega_{t,i}$, for $i = 0, 1, \dots, L$, is set as an independent MSDSP process in (6)-(9).

For simplicity, we set $L = 7$ base models as the linear economic models in Section 4.1. For prediction at horizon h , simply replace $\hat{y}_{t,j}$ by $\hat{y}_{t+h,j}$ and y_t by y_{t+h} , respectively. As long as the information set used to produce these predictions is up to time t , these are out-of-sample prediction. In addition to the MSDSP assembly, we also apply its nested alternative, the DSP, in model assembly, along with the standard BPS framework described in Section 5.1, for comparison. Note that no MSDSP or DSP methods are used in the construction of our base models. The ideas applying “MSDSP-in-MSDSP”, that is ensembling MSDSP models with MSDSP weights are left for future research.

5.3 Predictive Scheme with Model Assembly

Because of the the two-stage procedure for model assembly, we cut the data at two time points: T_0 and T_1 with $T_0 < T_1 < T$. The initial training sample uses the information up to time T_0 , which is the same as in Section 4. The initial training sample serves to construct the predictions from the base models in this section. The period between T_0 and T_1 is used to train the synthesis function, and thus the structure of the model weights ω_t It cannot be inferred without input from the base models, so the two-stage method wins us some computational time, at the cost of a smaller testing sample. We choose T_0 and T_1 such that $T - T_1 = 100$ and $T_1 - T_0 = 100$, and keep T_0 the same as in Section 4. This idea of breaking the sample to three subset is analogous to the training, validation and testing sampling scheme in the machine learning literature, but we keep updating the data after each prediction and hence can exploit the information as much as possible.

5.4 Model Assembly Results

Table 8 reports the log predictive likelihood (LPL), Continuous Ranked Probability Score (CRPS), and Root Mean Squared Forecast Error (RMSFE) for horizons $h = 1, \dots, 12$ months. These values are based on the last 100 samples, adjusted by the forecasting horizons. The methods include the MSDSP and DSP assembly, along with the standard DBPS of [McAlinn and West \(2019\)](#). The RW-SV predictive metrics over the corresponding sample period are also reported for comparison. Note that all base models are the linear economic models with static parameters for both versions of model assembly and DBPS.

Amongst the model assembly method, the MSDSP performs better than the DSP assembly and the standard DPBS in all evaluation metrics at all horizon. Thus, we show here that the MSDSP assembly can easily dominate the state-of-the-art dynamic Bayesian predictive synthesis. Without the need to deploy heavy computation for the distributional input, the MSDSP assembly is scalable. Its Markov switching structure explicate sparsity to improve prediction, which is evidenced by the dominance of MSDSP assembly over the DSP assembly.

The MSDSP assembly is slightly worse than the random walk at short horizons but tends to improve at longer horizons. For example, the point forecast (RMSFE) from MSDSP is better than the random walk when $h > 9$, which is empirically plausible (see also [Mark, 1995](#); [Rapach and Wohar, 2002](#); [Molodtsova and Papell, 2009](#)). A better model pool might improve the MSDSP assembly, but it is not our focus in this paper.

6 Discussion and Conclusion

We propose the novel Markov Switching Dynamic Shrinkage Process method to address dynamic instability, model parsimony and sparsity, and use the proposed framework to address the renowned Meese–Rogoff puzzle, which states that the random walk is highly competitive in exchange rate forecasting. We consider adopting the MSDSP in two settings: directly incorporating the flexible structure into existing economic models; and applying the flexibility in model assembly of standard linear economic models.

In our application, we show that adopting the MSDSP directly to enhance the economic

model structure leads to significant improvements in the predictive performance of the economic models relative to the benchmark random walk. The decisive evidence illustrate that the random walk model, even when adjusted to include stochastic volatility, is systematically inferior to the economic models improved with MSDSP priors for both point and density forecasts. This evidence, in itself, provide a new perspective on the Meese-Rogoff puzzle, highlighting the fact that the models resulting from economic theories possess predictive power when flexibility and sparsity is taken into account. Existing results literature have evaluated these models in its most stringent form via the constant coefficient linear model, under the assumption that the respective economic relationship is static, which, in the current economic environment, is unrealistic.

Furthermore, we provide a natural framework for model assembly by using the MSDSP toolkit. With a very primitive linear constant coefficient model pool, the MSDSP assembly consistently beat the state-of-the-art dynamic Bayesian predictive synthesis, while being scalable computationally. The MSDSP model assembly produces similar predictive performance as the benchmark random walk. Compared to the individual model improvement with MSDSP, there is still further improvements that can be investigated with future research, including the inclusion of more sophisticated base model and the incorporation of predictive distribution (rather than point) in the construction of predictive synthesis.

7 Tables

Table 3: UK: Comparative Forecast Performance (1,3,6,12-period-ahead forecasting) Using Log Predictive Density Ratio (LPDR) with reference: Random Walk with Stochastic Volatility with No Drift. Out-of-sample period: 2000/11 - 2017/6, in total 200 periods.

Macro Models	Linear			MSDSP	DSP
	Expanding Window	100-rolling Window	50-rolling Window		
h=1: 1-month ahead. RW-SV: benchmark. RW-drift-SV: -0.3883					
IRP	-29.6791	-5.8994	-5.8353	3.4615	-9.3826
TR	-30.5840	-7.1152	-6.9684	2.5229	-13.8571
PPP	-29.0611	-6.6741	-6.6981	3.7681	-8.6213
MON	-27.8990	-4.0341	-3.9542	2.7240	-7.6936
Oil	-26.6876	-2.0256	-1.9684	2.2365	-6.2400
Gold	-29.2406	-5.0844	-5.1214	2.9484	-10.0253
Copper	-25.9588	-0.5934	-0.6708	2.6266	-4.5026
h=3: 3-month ahead. RW-SV: benchmark. RW-drift-SV: -0.0616					
IRP	-29.1939	-5.2789	-5.1703	3.3508	-8.7248
TR	-29.7365	-6.6652	-6.6531	3.3151	-14.1063
PPP	-25.0914	-3.7528	-3.6766	5.2003	-7.4086
MON	-25.7211	-3.5646	-3.6080	3.8906	-4.9224
Oil	-28.1046	-3.9143	-3.8716	3.2649	-7.2334
Gold	-28.7496	-5.3689	-5.3944	3.0078	-9.0592
Copper	-27.4407	-2.8235	-2.7777	3.4050	-6.4234
h=6: 6-month ahead. RW-SV: benchmark. RW-drift-SV: -0.2911					
IRP	-28.2834	-5.5577	-5.5701	3.3833	-8.5100
TR	-28.2732	-5.5316	-5.5212	3.4639	-12.2628
PPP	-25.1515	-4.1046	-4.0514	3.1787	-9.3380
MON	-25.5214	-4.3304	-4.3226	2.9245	-8.7579
Oil	-25.0861	-1.8254	-1.8836	3.0826	-5.7115
Gold	-27.5627	-4.6564	-4.7751	3.6933	-8.3479
Copper	-27.2085	-3.8037	-3.9387	3.2561	-5.9737
h=12: 12-month ahead. RW-SV: benchmark. RW-drift-SV: -0.5118					
IRP	-28.7304	-4.9605	-4.8939	3.3245	-8.2378
TR	-28.8694	-5.2189	-5.2878	5.9622	-13.3841
PPP	-24.0625	-4.1664	-4.0931	4.0214	-9.6608
MON	-24.4873	-4.0269	-3.9948	3.7719	-9.6037
Oil	-28.1024	-4.2732	-4.2203	3.0321	-9.9155
Gold	-27.6593	-3.6843	-3.7151	3.0996	-6.4277
Copper	-28.6770	-5.4533	-5.4238	3.0763	-8.6516

Note: All LPDR values are computed as the summation over all OOS periods, and recorded as the distance to RW-SV without drift (benchmark values are -272.0914, -272.9213, -272.6078 and -272.7673 for each h). Bold numbers indicate the best-performing model in each row (higher values indicate better predictive performance), and the boxed number indicates the best-performing model in each block. Model abbreviations: IRP = Interest Rate Parity, TR = Taylor Rule, PPP = Purchasing Power Parity, MON = Monetary Model.

Table 4: UK: Comparative Forecast Performance (1,3,6,12-period-ahead forecasting) Using Relative Continuous Ranked Probability Score (CRPS) - overall score. Out-of-sample period: 2000/11 - 2017/6, in total 200 periods

Macro Models	Linear			MSDSP	DSP
	Expanding Window	100-rolling Window	50-rolling Window		
h=1: 1-month ahead. RW-SV: benchmark. RW-drift-SV: +0.0015					
IRP	0.0063	0.0067	0.0071	-0.0095	0.0396
TR	0.0115	0.0113	0.0102	-0.0087	0.0408
PPP	0.0106	0.0115	0.0113	-0.0061	0.0277
MON	0.0052	0.0054	0.0037	-0.0060	0.0213
Oil	-0.0054	-0.0040	-0.0050	-0.0078	0.0125
Gold	0.0061	0.0061	0.0051	-0.0055	0.0377
Copper	-0.0063	-0.0053	-0.0039	-0.0058	0.0154
h=3: 3-month ahead. RW-SV: benchmark. RW-drift-SV: +0.0023					
IRP	0.0042	0.0062	0.0061	-0.0080	0.0247
TR	0.0064	0.0093	0.0087	-0.0067	0.0351
PPP	0.0025	0.0038	0.0025	-0.0108	0.0221
MON	0.0043	0.0039	0.0027	-0.0083	0.0197
Oil	0.0007	0.0013	0.0004	-0.0092	0.0258
Gold	0.0071	0.0068	0.0069	-0.0022	0.0287
Copper	-0.0030	-0.0027	-0.0023	-0.0068	0.0200
h=6: 6-month ahead. RW-SV: benchmark. RW-drift-SV: -0.0003					
IRP	0.0042	0.0043	0.0043	-0.0084	0.0430
TR	0.0046	0.0029	0.0045	-0.0101	0.0375
PPP	-0.0004	-0.0008	-0.0002	-0.0096	0.0589
MON	0.0015	0.0017	0.0008	-0.0098	0.0276
Oil	-0.0041	-0.0047	-0.0033	-0.0107	0.0102
Gold	0.0015	0.0015	0.0023	-0.0101	0.0318
Copper	0.0016	0.0022	0.0022	-0.0082	0.0099
h=12: 12-month ahead. RW-SV: benchmark. RW-drift-SV: -0.0024					
IRP	0.0021	0.0042	0.0033	-0.0095	0.0192
TR	0.0027	0.0031	0.0024	-0.0144	0.0392
PPP	0.0018	0.0016	0.0020	-0.0087	0.0285
MON	0.0014	0.0009	-0.0019	-0.0098	0.0280
Oil	0.0008	0.0008	-0.0004	-0.0101	0.0187
Gold	-0.0009	-0.0012	-0.0015	-0.0097	0.0155
Copper	0.0046	0.0051	0.0049	-0.0063	0.0245

Note: CRPS values are averaged across all OOS periods and recorded as the distance to RW-SV without drift (benchmark values are 0.5226, 0.5246, 0.5264 and 0.5267 for each h). Bold numbers indicate the best-performing model in each row (Lower CRPS values indicate better predictive performance), and the boxed number indicates the best-performing model in each block. Model abbreviations: IRP = Interest Rate Parity, TR = Taylor Rule, PPP = Purchasing Power Parity, MON = Monetary Model.

Table 5: UK: Comparative Forecast Performance (1,3,6,12-period-ahead forecasting) Using Relative Root Mean Squared Forecast Error (RMSFE). Out-of-sample period: 2000/11 - 2017/6, in total 200 periods.

Macro Models	Linear			MSDSP	DSP
	Expanding Window	100-rolling Window	50-rolling Window		
h=1: 1-month ahead. RW-SV: benchmark. RW-drift-SV: 1.0017					
IRP	1.0166	1.0167	1.0163	0.9841	1.0778
TR	1.0311	1.0316	1.0310	0.9880	1.0687
PPP	1.0161	1.0162	1.0163	0.9904	1.0419
MON	1.0000	1.0002	1.0001	0.9891	1.0291
Oil	0.9903	0.9899	0.9900	0.9870	1.0246
Gold	1.0054	1.0052	1.0055	0.9911	1.0634
Copper	0.9866	0.9866	0.9863	0.9898	1.0163
h=3: 3-month ahead. RW-SV: benchmark. RW-drift-SV: 1.0049					
IRP	1.0152	1.0151	1.0151	0.9861	1.0427
TR	1.0246	1.0248	1.0243	0.9879	1.0587
PPP	1.0032	1.0032	1.0031	0.9743	1.0296
MON	1.0016	1.0015	1.0016	0.9858	1.0260
Oil	1.0009	1.0005	1.0005	0.9835	1.0454
Gold	1.0107	1.0107	1.0105	0.9980	1.0488
Copper	0.9954	0.9957	0.9952	0.9893	1.0308
h=6: 6-month ahead. RW-SV: benchmark. RW-drift-SV: 1.0002					
IRP	1.0086	1.0084	1.0081	0.9850	1.1073
TR	1.0114	1.0117	1.0119	0.9841	1.0622
PPP	0.9996	0.9996	0.9995	0.9841	1.1772
MON	1.0010	1.0011	1.0009	0.9821	1.0465
Oil	0.9919	0.9919	0.9919	0.9817	1.0017
Gold	1.0028	1.0032	1.0031	0.9827	1.0637
Copper	0.9976	0.9975	0.9974	0.9860	1.0083
h=12: 12-month ahead. RW-SV: benchmark. RW-drift-SV: 0.9961					
IRP	1.0034	1.0033	1.0034	0.9821	1.0270
TR	1.0164	1.0164	1.0167	0.9687	1.0666
PPP	1.0008	1.0010	1.0008	0.9843	1.0452
MON	1.0000	1.0003	1.0001	0.9851	1.0413
Oil	1.0017	1.0017	1.0018	0.9822	1.0307
Gold	0.9906	0.9903	0.9906	0.9832	1.0168
Copper	1.0057	1.0060	1.0059	0.9901	1.0387

Note: RMSFE values are averaged across all OOS periods, and recorded as the ratio to RW-SV without drift (benchmark values are 0.9546, 0.9581, 0.9609 and 0.9609 for each h). Bold numbers indicate the best-performing model in each row (Lower RMSFE values indicate better predictive performance), and the boxed number indicates the best-performing model in each block. Model abbreviations: IRP = Interest Rate Parity, TR = Taylor Rule, PPP = Purchasing Power Parity, MON = Monetary Model.

Table 6: UK: Comparative Forecast Performance (1-period ahead forecasting) Using Coverage Rate across Categories. Out-of-sample period: 2000/11- 2017/6, in total 200 periods.

Macro Models	Linear			MSDSP	DSP
	Expanding Window	100-rolling Window	50-rolling Window		
Top 2.5%. RW-SV:1.5%. RW-drift-SV: 1.5%					
IRP	3.0%	3.0%	3.0%	2.5%	2.0%
TR	3.5%	4.0%	3.5%	2.5%	2.0%
PPP	3.0%	3.0%	2.5%	2.5%	2.5%
MON	2.0%	2.0%	2.0%	3.0%	2.5%
Oil	2.5%	2.0%	2.0%	2.5%	2.0%
Gold	2.5%	2.5%	2.0%	2.5%	2.0%
Copper	2.0%	2.5%	2.5%	2.5%	2.0%
Average	2.64%	2.71%	2.50%	2.57%	2.14%
Bottom 2.5%. RW-SV: 0.5%. RW-drift-SV: 0.5%					
IRP	1.0%	1.0%	1.0%	1.0%	1.0%
TR	1.0%	1.0%	1.5%	1.0%	1.0%
PPP	1.0%	1.0%	1.0%	1.0%	1.0%
MON	0.5%	1.0%	1.0%	1.0%	1.0%
Oil	0.5%	0.5%	1.0%	1.0%	1.0%
Gold	0.5%	0.5%	1.0%	1.0%	0.5%
Copper	1.5%	1.5%	1.0%	0.5%	1.0%
Average	0.86%	0.93%	1.07%	0.93%	0.93%
Top 10%. RW-SV:4.5%. RW-drift-SV: 5.5%					
IRP	8.0%	7.0%	7.5%	8.0%	7.0%
TR	8.0%	8.0%	8.0%	8.0%	6.5%
PPP	8.0%	8.0%	8.5%	8.5%	8.5%
MON	8.5%	8.5%	8.5%	8.0%	9.0%
Oil	6.5%	6.5%	7.0%	8.0%	6.5%
Gold	8.5%	8.5%	8.5%	8.0%	9.5%
Copper	7.0%	7.5%	7.0%	8.0%	9.0%
Average	7.79%	7.71%	7.86%	8.07%	8.00%
Bottom 10%. RW-SV: 3.5%. RW-drift-SV: 2.5%					
IRP	7.0%	7.0%	7.5%	10.5%	10.0%
TR	8.0%	8.5%	8.5%	10.5%	9.0%
PPP	7.5%	6.5%	6.5%	11.0%	11.5%
MON	7.0%	7.5%	7.5%	10.5%	10.0%
Oil	9.5%	8.5%	8.5%	10.0%	9.0%
Gold	6.5%	7.0%	7.0%	10.0%	9.0%
Copper	8.5%	8.0%	8.0%	10.0%	8.0%
Average	7.71%	7.57%	7.71%	10.36%	9.36%

Note: Coverage rate values are averaged across all OOS periods. Model abbreviations: IRP = Interest Rate Parity, TR = Taylor Rule, PPP = Purchasing Power Parity, MON = Monetary Model.

Table 7: UK: 95% Model Confidence Set for 1,2-period-ahead forecasting

Estimation Methods	Macro Models	1. SFE		2. Log Score		3. CRPS	
		h=1	h=2	h=1	h=2	h=1	h=2
RW-SV		14	11	8	8	14	9
RW-drift-SV		18	10	9	9	15	8
Linear Expanding Window	IRP	28	30	-	-	19	24
	TR	31	25	-	-	26	20
	PPP	23	24	-	-	24	17
	MON	15	14	-	-	-	-
	Oil	12	21	-	-	9	21
	Gold	20	16	-	-	17	15
	Copper	10	7	-	-	10	12
Linear 100-rolling Window	IRP	26	28	22	23	20	28
	TR	32	27	24	20	28	25
	PPP	25	22	20	12	27	26
	MON	17	15	18	18	-	-
	Oil	13	20	13	22	11	22
	Gold	19	17	15	15	18	14
	Copper	8	8	10	11	12	7
Linear 50-rolling Window	IRP	27	29	21	24	21	27
	TR	30	26	23	19	23	23
	PPP	24	23	19	14	25	19
	MON	16	13	17	17	-	16
	Oil	11	19	12	21	8	18
	Gold	21	18	16	16	16	13
	Copper	9	6	11	10	13	10
MSDSP	IRP	5	2	6	1	6	3
	TR	4	4	4	2	4	2
	PPP	7	12	1	7	7	11
	MON	6	9	7	5	5	6
	Oil	3	3	6	6	3	4
	Gold	2	1	3	3	1	1
	Copper	1	5	2	4	2	5
DSP	IRP	34	-	-	-	31	-
	TR	-	32	-	25	-	29
	PPP	-	-	-	-	32	-
	MON	29	31	-	13	29	30
	Oil	-	34	-	-	-	32
	Gold	33	33	-	-	30	31
	Copper	22	-	14	-	22	-

Note: The MCS results are based on the three loss functions considered. Rank M reflects a model's performance relative to others according to the selected loss function, with a lower rank indicating better performance (i.e., rank 1 corresponds to the best model). A dash (-) signifies that the model was eliminated in the procedure. Model abbreviations: IRP = Interest Rate Parity, TR = Taylor Rule, PPP = Purchasing Power Parity, MON = Monetary Model.

Table 8: UK: Comparative Forecast Performance. Out-of-sample period: 2009/03 - 2017/6, in total 100 periods.

Assembly	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
LPL												
MSDSP	-142.3786	-144.8047	-142.3443	-144.7816	-141.3632	-141.8206	-143.1669	-142.9444	-144.5245	-142.6884	-142.1529	-142.9249
DBPS	-156.3835	-162.8345	-154.9650	-153.9961	-149.9553	-148.0839	-147.8373	-144.2275	-149.1719	-147.4700	-142.3262	-145.5031
DSP	-151.1876	-158.5468	-148.4923	-151.4041	-151.7629	-152.0837	-152.2112	-146.9985	-156.2246	-149.0994	-146.9559	-150.3118
RW-SV	-140.0417	-139.8219	-139.8489	-139.8884	-139.5621	-139.9888	-140.0941	-140.2194	-140.4038	-140.1973	-140.3494	-140.3048
CRPS												
MSDSP	0.5692	0.5659	0.5500	0.5730	0.5523	0.5598	0.5716	0.5767	0.5638	0.5482	0.5506	0.5514
DBPS	0.7306	0.7509	0.7048	0.7786	0.6978	0.6953	0.6938	0.6486	0.6909	0.6217	0.6626	0.6129
DSP	0.6492	0.6853	0.6122	0.6364	0.6338	0.6345	0.6434	0.6320	0.6786	0.6034	0.5972	0.6178
RW-SV	0.5422	0.5427	0.5413	0.5404	0.5481	0.5481	0.5411	0.5496	0.5485	0.5572	0.5475	0.5438
RMSFE												
MSDSP	1.0041	1.0208	0.9894	1.0302	0.9860	0.9945	1.0182	1.0224	1.0104	0.9918	0.9872	0.9886
DBPS	1.0928	1.1784	1.0936	1.1284	1.0411	1.0622	1.0753	1.0499	1.0244	1.0187	0.9895	1.0051
DSP	1.1289	1.2143	1.1061	1.1277	1.1113	1.1017	1.1414	1.1466	1.2314	1.0810	1.0432	1.0879
RW-SV	0.9841	0.9867	0.9824	0.9805	0.9962	0.9920	0.9826	0.9958	0.9902	1.0080	0.9917	0.9848

Note: All LPL values are computed as the summation over all OOS periods. CRPS and RMSFE values are averaged across all OOS periods. Bold numbers indicate the best-performing model in each row (Higher LPDR values or Lower CRPS and RMSFE values indicate better predictive performance). Model abbreviations: IRP = Interest Rate Parity, TR = Taylor Rule, PPP = Purchasing Power Parity, MON = Monetary Model. These single models are the forecast results from the previous section using Linear models.

References

- Aastveit, K. A., Cross, J. L., and van Dijk, H. K. (2023). Quantifying time-varying forecast uncertainty and risk for the real price of oil. *Journal of business & economic statistics: a publication of the American Statistical Association*, 41(2):523–537.
- Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, 18(3):338–357.
- Aristidou, C., Lee, K., and Shields, K. (2022a). Fundamentals, regimes and exchange rate forecasts: Insights from a meta exchange rate model. *Journal of International Money and Finance*, 123(102601):102601.
- Aristidou, C., Lee, K., and Shields, K. (2022b). A meta model analysis of exchange rate determination. In *Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling*, Advances in econometrics, pages 199–215. Emerald Publishing Limited.
- Barndorff-Nielsen, O., Kent, J., and Sørensen, M. (1982). Normal Variance-Mean mixtures and z distributions. *International statistical review = Revue internationale de statistique*, 50(2):145–159.
- Bernardi, M., Bianchi, D., and Bianco, N. (2023). Dynamic variable selection in high-dimensional predictive regressions. *arXiv preprint arXiv:2304.07096*.
- Bernardi, M. and Catania, L. (2018). The model confidence set package for R. *International Journal of Computational Economics and Econometrics*, 8(2):144–158.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.
- Byrne, J. P., Korobilis, D., and Ribeiro, P. J. (2016). Exchange rate predictability in a changing world. *Journal of International Money and Finance*, 62:1–24.
- Candian, G. and De Leo, P. (2023). Imperfect exchange rate expectations. *Review of Economics and Statistics*, pages 1–46.

- 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.
- Časta, M. (2024). Forecasting nominal exchange rates using a dynamic model averaging framework. *Heliyon*, 10(20).
- Chan, J. C. and Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter vars with stochastic volatility. *Journal of applied econometrics*, 33(4):509–532.
- Chan, J. C., Koop, G., and Yu, X. (2024). Large order-invariant bayesian vars with stochastic volatility. *Journal of Business & Economic Statistics*, 42(2):825–837.
- Cheung, Y.-W., Chinn, M. D., Pascual, A. G., and Zhang, Y. (2019). Exchange rate prediction redux: New models, new data, new currencies. *Journal of International Money and Finance*, 95:332–362.
- Del Negro, M., Hasegawa, R. B., and Schorfheide, F. (2016). Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics*, 192(2):391–405.
- Della Corte, P., Sarno, L., and Tsiakas, I. (2009). An economic evaluation of empirical exchange rate models. *The review of financial studies*, 22(9):3491–3530.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple Classifier Systems*, pages 1–15, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Dufays, A., Li, Z., Rombouts, J. V., and Song, Y. (2021). Sparse change-point var models. *Journal of Applied Econometrics*, 36(6):703–727.
- Engel, C. (1994). Can the markov switching model forecast exchange rates? *Journal of international economics*, 36(1-2):151–165.
- Engel, C., Lee, D., Liu, C., Liu, C., and Wu, S. P. Y. (2019). The uncovered interest parity puzzle, exchange rate forecasting, and taylor rules. *Journal of International Money and Finance*, 95:317–331.

- Fang, S., Wei, Y., and Wang, S. (2024). 30 years of exchange rate analysis and forecasting: a bibliometric review. *Journal of Economic Surveys*, 38(3):973–1007.
- Ferraro, D., Rogoff, K., and Rossi, B. (2015). Can oil prices forecast exchange rates? an empirical analysis of the relationship between commodity prices and exchange rates. *Journal of International Money and Finance*, 54:116–141.
- George, E. I. and McCulloch, R. E. (1997). Approaches for bayesian variable selection. *Statistica sinica*, pages 339–373.
- Geweke, J. and Amisano, G. (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International journal of forecasting*, 26(2):216–230.
- Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164(1):130–141.
- Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378.
- Hahn, P. R. and Carvalho, C. M. (2015). Decoupling shrinkage and selection in bayesian linear models: A posterior summary perspective. *Journal of the American Statistical Association*, 110(509):435–448.
- Hall, S. G. and Mitchell, J. (2007). Combining density forecasts. *International journal of forecasting*, 23(1):1–13.
- Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica: journal of the Econometric Society*, 79(2):453–497.
- Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 38(3):205–228.
- Hauzenberger, N., Huber, F., and Koop, G. (2024). Dynamic shrinkage priors for large time-varying parameter regressions using scalable markov chain monte carlo methods. *Studies in Nonlinear Dynamics & Econometrics*, 28(2):201–225.

- Huber, F., Koop, G., and Onorante, L. (2021). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics*, 39(3):669–683.
- Jin, X., Maheu, J. M., and Yang, Q. (2022). Infinite markov pooling of predictive distributions. *Journal of econometrics*, 228(2):302–321.
- Johnson, M. C. (2017). *Bayesian predictive synthesis: Forecast calibration and combination*. PhD thesis, Duke University.
- Justiniano, A. and Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review*, 98(3):604–641.
- Kalli, M. and Griffin, J. E. (2014). Time-varying sparsity in dynamic regression models. *Journal of Econometrics*, 178(2):779–793.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.
- Knaus, P. and Frühwirth-Schnatter, S. (2023). The dynamic triple gamma prior as a shrinkage process prior for time-varying parameter models. *arXiv preprint arXiv:2312.10487*.
- Kowal, D. R., Matteson, D. S., and Ruppert, D. (2019). Dynamic shrinkage processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(4):781–804.
- Lopes, H. F., McCulloch, R. E., and Tsay, R. S. (2022). Parsimony inducing priors for large scale state–space models. *Journal of Econometrics*, 230(1):39–61.
- Mark, N. C. (1995). Exchange rates and fundamentals: Evidence on long-horizon predictability. *The American Economic Review*, pages 201–218.
- McAlinn, K., Aastveit, K. A., Nakajima, J., and West, M. (2020a). Multivariate bayesian predictive synthesis in macroeconomic forecasting. *Journal of the American Statistical Association*, 115(531):1092–1110.

- McAlinn, K., Aastveit, K. A., Nakajima, J., and West, M. (2020b). Multivariate bayesian predictive synthesis in macroeconomic forecasting. *Journal of the American Statistical Association*, 115(531):1092–1110.
- McAlinn, K. and West, M. (2019). Dynamic bayesian predictive synthesis in time series forecasting. *Journal of econometrics*, 210(1):155–169.
- Meese, R. A. and Rogoff, K. (1983a). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14(1-2):3–24.
- Meese, R. A. and Rogoff, K. (1983b). The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification? *Exchange Rates and International Macroeconomics*, pages 67–112.
- Meese, R. A. and Rogoff, K. (1988). Was it real? the exchange rate-interest differential relation over the modern floating-rate period. *The Journal of Finance*, 43(4):933–948.
- Molodtsova, T. and Papell, D. H. (2009). Out-of-sample exchange rate predictability with taylor rule fundamentals. *Journal of international economics*, 77(2):167–180.
- Mumtaz, H. and Sunder-Plassmann, L. (2013). Time-varying dynamics of the real exchange rate: An empirical analysis. *Journal of applied econometrics*, 28(3):498–525.
- Muslim, M. A., Nikmah, T. L., Pertiwi, D. A. A., Subhan, Jumanto, Dasril, Y., and Iswanto (2023). New model combination meta-learner to improve accuracy prediction p2p lending with stacking ensemble learning. *Intelligent Systems with Applications*, 18:200204.
- Nakajima, J. (2011). Time-varying parameter var model with stochastic volatility: An overview of methodology and empirical applications. *Monetary and Economic Studies*, 29:107–142.
- Nakajima, J. and West, M. (2013). Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics*, 31(2):151–164.

- Neghab, D. P., Cevik, M., Wahab, M., and Basar, A. (2024). Explaining exchange rate forecasts with macroeconomic fundamentals using interpretive machine learning. *Computational Economics*, pages 1–43.
- Nikolsko-Rzhevskyy, A. and Prodan, R. (2012). Markov switching and exchange rate predictability. *International journal of forecasting*, 28(2):353–365.
- Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449.
- Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American statistical Association*, 108(504):1339–1349.
- Prado, R. and West, M. (2010). *Time series: Modeling, computation, and inference*. Chapman & Hall/CRC Texts in Statistical Science. Chapman & Hall/CRC, Philadelphia, PA.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of economic studies*, 72(3):821–852.
- Proscura, P. and Zaytsev, A. (2022). Effective training-time stacking for ensembling of deep neural networks.
- Rapach, D. E. and Wohar, M. E. (2002). Testing the monetary model of exchange rate determination: new evidence from a century of data. *Journal of International Economics*, 58(2):359–385.
- Rockova, V. and McAlinn, K. (2021). Dynamic variable selection with spike-and-slab process priors. *Bayesian Analysis*, 16(1):233–269.
- Rossi, B. (2013). Exchange rate predictability. *Journal of economic literature*, 51(4):1063–1119.
- Tallman, E. and West, M. (2024a). Bayesian predictive decision synthesis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(2):340–363.

- Tallman, E. and West, M. (2024b). Bayesian predictive decision synthesis. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 86(2):340–363.
- Uribe, P. W. and Lopes, H. F. (2020). Dynamic sparsity on dynamic regression models. *arXiv preprint arXiv:2009.14131*.
- Waggoner, D. F. and Zha, T. (2012). Confronting model misspecification in macroeconomics. *Journal of Econometrics*, 171(2):167–184.
- West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*. Springer Series in Statistics. Springer, New York, NY, 2 edition.
- West, M. and Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.
- Wolff, C. C. (1987). Time-varying parameters and the out-of-sample forecasting performance of structural exchange rate models. *Journal of Business & Economic Statistics*, 5(1):87–97.
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2):241–259.
- Wright, J. H. (2008). Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics*, 146(2):329–341.