RESEARCH ARTICLE

WILEY

Model instability in predictive exchange rate regressions

Niko Hauzenberger[1,2] | Florian Huber[2]

[1] Department of Economics, WU Vienna University of Economics and Business, Vienna, Austria

[2] Salzburg Centre of European Union Studies (SCEUS), Paris Lodron University of Salzburg, Salzburg, Austria

Correspondence
Niko Hauzenberger, Department of Economics, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria.
Email: niko.hauzenberger@wu.ac.at

Abstract

In this paper we aim to improve existing empirical exchange rate models by accounting for uncertainty with respect to the underlying structural representation. Within a flexible Bayesian framework, our modeling approach assumes that different regimes are characterized by commonly used structural exchange rate models, with transitions across regimes being driven by a Markov process. We assume a time-varying transition probability matrix with transition probabilities depending on a measure of the monetary policy stance of the central bank at home and in the USA. We apply this model to a set of eight exchange rates against the US dollar. In a forecasting exercise, we show that model evidence varies over time, and a model approach that takes this empirical evidence seriously yields more accurate density forecasts for most currency pairs considered.

KEYWORDS

empirical exchange rate models, exchange rate fundamentals, Markov switching

1 | INTRODUCTION

Since the end of the Bretton Woods system in 1971, economists have been confronted with the challenging issue of designing empirical models of bilateral exchange rates, which are also useful for forecasting applications. In a seminal contribution, Meese and Rogoff (1983) provided some early evidence that exchange rates are difficult to predict, at least in the short run. Using a set of theoretical models in the spirit of Dornbusch (1976), Frankel (1979), and Hooper and Morton (1982), to guide the choice of covariates included in a forecasting regression, Meese and Rogoff (1983) found that a simple random walk benchmark is difficult to outperform for most major exchange rate pairs. One reason for the dismal performance of most empirical and structural models is that, within a standard asset pricing framework, the high persistence of the underlying fundamentals in light of a discount factor near unity translates into highly persistent exchange rates. As a consequence, a random walk appears to be a benchmark difficult to beat (see Engel & West, 2005).

Over the years, a plethora of alternative econometric techniques emerged that provide more sophisticated means for analyzing exchange rate data to successfully improve longer term predictions. The literature on unit roots and cointegration, for example, provided tools to explicitly discriminate between short-term movements of a given currency pair and its long-run behavior. Mark (1995), for instance, applied an error correction model to a set of four exchange rates against the US dollar. Within this error correction framework, the exchange rate is assumed to return to its long-run equilibrium value determined by a simple monetary model, with short-run fluctuations driven by lagged changes of the exchange rate and its fundamentals. The finding that exchange rates tend to be predictable in the medium and long run

sparked a series of related contributions that corroborate this result for different periods and currency pairs (Groen, 2000; Mark & Sul, 2001; Rapach & Wohar, 2002).

More recently, several studies emphasized the usefulness of accounting for nonlinearities in the underlying econometric models to provide more precise exchange rate predictions (see, for example, Byrne, Korobilis, & Ribeiro, 2016; Canova, 1993; Huber, 2016, 2017; Huber & Zörner, 2019; Mark, 2009; Sarno, Valente, & Wohar, 2004). The majority of this literature deals with the question on whether a given empirical model, that is loosely based on an underlying structural model, outperforms a set of competing models. In this context, introducing nonlinearities boils down to allow for time-varying error variances and/or time-varying regression coefficients for a certain structural model.

However, another key source of nonlinearities could stem from the fact that the underlying theoretical model changes over time, potentially jeopardizing the predictive fit of the econometric specification.[1] For instance, the recent success of Taylor rule-based models (see Engel & West, 2006; Molodtsova, Nikolsko-Rzhevskyy, & Papell, 2008, 2011; Molodtsova & Papell, 2009) can be attributed to the fact that involved central banks adopted a policy rule closely related to a Taylor rule. With short-term interest rates reaching the zero lower bound (ZLB) and central banks starting to implement unconventional monetary policy measures, the question arises whether a Taylor rule still proves to be an adequate exchange rate model. In fact, recent literature on nonlinear Taylor rules suggests that during the ZLB, Taylor rule-based models loose their momentum against simple random walk specifications (Byrne et al., 2016; Huber, 2017).

In this paper, we contribute to the literature by acknowledging this empirical evidence and propose a modeling framework capable of handling model instability over time in a flexible manner. This is achieved by proposing a Markov-switching (MS) regression model with each regime being characterized by different covariates arising from a set of structural exchange rate models. In contrast to the existing literature, which relies on dynamic Bayesian model averaging techniques, our approach is an integrated modeling device. In addition, the introduction of time-varying transition probabilities allows assessment of how the likelihood of a given structural model changes over time, depending on selected early-warning indicators. As signal variables, we adopt the (lagged) interest rates of the home country and the USA. This specification is motivated by the observation that Taylor rule fundamentals are good predictors in the

period before the global financial crisis (with policy rates being significantly larger than zero), but are known for their weak performance in the aftermath of the crisis (characterized by policy rates close to zero).

We assess the merits of the proposed approach using a forecasting exercise for eight different exchange rates against the US dollar. By considering the resulting regime allocation and the transition probabilities, we examine whether structural models indeed tend to change and how this is related to movements in policy rates. The findings indicate that allowing for time-varying probabilities is a key feature, pointing towards a strong relationship between policy rates and the underlying transition distribution of the Markov process. In terms of forecasting, we find that our proposed model improves upon the random walk for selected currencies, both in terms of point and density predictions. The improvements for point forecasts are, however, muted. Comparing different model features reveals that a model based on a larger set of fundamentals from various structural models is also competitive when combined with shrinkage priors and nonlinearities (in the form of MS).

The remainder of this paper is organized as follows. Section 2 discusses the four structural exchange rate models adopted, while Section 3 proposes the econometric framework. The empirical application is presented in Section 4. The final section summarizes and concludes the paper. A technical Appendix provides details on the estimation algorithm adopted.

2 | THEORETICAL EXCHANGE RATE MODELS

In this section, we briefly discuss the main theoretical underpinnings to be used to guide covariate inclusion in the empirical model as well as to structurally identify the different regimes considered in our nonlinear regression framework.

The point of departure for the discussion is a set of macroeconomic and financial quantities stored in an R-dimensional vector $\boldsymbol{X}_t$:

$$\boldsymbol{X}_t = \left(\underbrace{i_{t-1}, i_{t-1}^*, \pi_t, \pi_t^*, x_t, x_t^*, q_t}_{\mathbf{x}_{0t}}, \underbrace{m_t, m_t^*, y_t, y_t^*}_{\mathbf{x}_{1t}}, \underbrace{e_t, p_t, p_t^*}_{\mathbf{x}_{2t}}, \underbrace{i_t, i_t^*}_{\mathbf{x}_{3t}} \right)',$$

with i_{t-1} denoting the lagged short-term interest rate, π_t inflation, x_t output gap, m_t money supply, y_t income, p_t price level, while the real exchange rate is denoted by q_t and the nominal exchange rate by e_t.[2] The subsets of

[1]Recent contributions, dealing with this issue, are Wright (2008), Beckmann and Schüssler (2016), Beckmann, Koop, Korobilis, and Schüssler (2018), and Byrne, Korobilis, and Ribeiro (2018).

[2]Asterisks denote US quantities. Moreover, y_t, m_t, p_t, q_t and e_t are measured in logarithms. For simplicity, we suppress subset-specific intercepts.

X_t, X_{jt} (j=0,...,3), represent covariates associated with the different structural models that we describe next.

2.1 | A taxonomy of selected models of exchange rate determination

In the following, we provide a brief taxonomy of the theoretical models considered that determine the specific partitions of X_t.

- Our starting point is the model based on *Taylor rule fundamentals* (see, Molodtsova & Papell, 2009, for a recent forecasting study). This specification assumes that the set of predictors is given by X_{0t} and thus includes the lagged short-term interest rate, inflation and the output gap of both the home country and the USA, and the real exchange rate. This model has proved to be successful in terms of describing exchange rate movements, both in-sample (Engel & West, 2006) and out-of-sample (Molodtsova et al., 2008, 2011). However, one critical assumption of this specification is that the central bank at home and in the USA is actively pursuing a Taylor rule-based monetary policy strategy. Especially during the recent period at the ZLB, this assumption could be violated, effectively leading to an inferior model fit.
- The second model considered is the *long-run monetary model*. The monetary model assumes that the covariates are given by X_{1t} and include data on domestic and US money supply as well as cross-country differences in income for a given income elasticity. As mentioned by Rapach and Wohar (2002), the long-run monetary model simply states that the price level of the home country and the USA is determined by the money supply and the level of production. Assuming purchasing power parity (PPP) and uncovered interest rate parity (UIP), one is able to relate the change in the exchange rate to supply and demand for money.
- Third, we consider a model based on PPP. This model selects X_{2t}, leading to a regression model that includes domestic and US price indices. PPP originates from the law of one price, which in turn implies that the real exchange rate is supposed to revert to a long-run equilibrium level determined by relative prices. If this turns out to be true, the real exchange rate is a stationary process. However, Sarno (2005) highlighted substantial persistence in real exchange rates. The convergence towards PPP is thus slow in the long run and, real exchange rates typically display pronounced deviations from their PPP-implied fundamentals in the short run.

- Finally, we also augment our forecasting regression with the UIP model. By selecting X_{3t}, this model simply establishes a relationship between the change in the exchange rate and the interest rate differential between the home country and the USA. Following Chinn (2006), UIP implies a positive one-to-one relationship between the interest rate differential and changes in the exchange rate. A positive change in the interest differential may be potentially followed by an immediate and persistent appreciation in the short run, implying that UIP does not hold immediately.[3]

All these models have been shown to possess some merit in terms of predictive power. However, several recent studies find remarkable heterogeneity with respect to the fundamental model adopted (see, among others, Beckmann & Schüssler, 2016; Beckmann et al., 2018; Byrne et al., 2018; Wright, 2008).

3 | CONTROLLING FOR MODEL INSTABILITY IN EMPIRICAL EXCHANGE RATE MODELS

In this section, we propose a model that controls for dynamic model instability by specifying a nonlinear econometric framework. After summarizing the model structure in Section 3.1, we highlight the prior setup adopted in Section 3.2.

3.1 | An MS model specification

We now turn to describing the proposed MS model with time-varying transition probabilities (MS-TVP). The key feature of our proposed framework is that it allows for switching between the fundamentals implied by K competing theoretical exchange rate models. We assume that exchange rate returns Δe_t follow an MS-TVP model given by

$$\Delta e_t = X'_{S_t t-1} \beta_{S_t} + \eta_t, \quad (1)$$

where $S_t \in \{0,...,K-1\}$ follows a first-order Markov process, β_k represents a vector of dimension M_k that collects the state-specific coefficients of state $S_t=k$, while $\eta_t \sim \mathcal{N}(0, \sigma^2_{S_t})$ is a white noise shock with regime-specific variance $\sigma^2_{S_t}$. Note that each β_k may exhibit different dimensions. We depart from the traditional literature on

---

[3]See, for instance, Chinn (2006) and Engel (2014) who observe coefficients that are less than one or even negative.

MS models (see, among many others, Amisano & Fagan, 2013; Billio, Casarin, Ravazzolo, & Van Dijk, 2016; Casarin, Sartore, & Tronzano, 2018; Engel, 1994; Filardo, 1994; Hamilton, 1994; Huber & Fischer, 2018; Kaufmann, 2015) by assuming that the regimes are characterized by competing structural exchange rate models, implying that different fundamentals enter the predictive exchange rate regression at different points in time.

In the spirit of Belmonte & Koop, 2014 and Frühwirth-Schnatter, 2006, we introduce a selection matrix $\boldsymbol{D}_{S_t}$ that entails switching between K alternative model specifications:

$$\Delta e_t = \boldsymbol{X}'_{t-1} \boldsymbol{D}_{S_t} \boldsymbol{\beta} + \eta_t, \tag{2}$$

with $\boldsymbol{X}_{t-1}$ denoting an R-dimensional vector of the full set of economic fundamentals. We define $\boldsymbol{\beta} = (\boldsymbol{\beta}'_0, ..., \boldsymbol{\beta}'_{K-1})'$ as a stacked M-dimensional vector of regime-specific coefficients with $M = \sum_{j=0}^{K-1} M_j$, and $\boldsymbol{\sigma}^2 = (\sigma_0^2, ..., \sigma_{K-1}^2)'$ collecting the K state-specific variances.

The selection matrix $\boldsymbol{D}_k$ of state $S_t{=}k$ is an $R{\times}M$-dimensional matrix with binary indicators that select $\boldsymbol{\beta}_k$ and $\boldsymbol{X}_{kt-1}$ while zeroing out the elements in $\boldsymbol{\beta}$ and $\boldsymbol{X}_{t-1}$ associated with the remaining models. For instance, we effectively obtain the model based on Taylor rule fundamentals, characterized through $S_t{=}0$, by setting

$$\boldsymbol{D}_0 = \begin{pmatrix} \boldsymbol{I}_7 & ... & \boldsymbol{0}_{7\times 2} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}_{2\times 7} & ... & \boldsymbol{0}_{2\times 2} \end{pmatrix},$$

where $\boldsymbol{0}_{i\times j}$ is an $i{\times}j$-dimensional matrix of zeros. Multiplying $\boldsymbol{\beta}$ from the left with $\boldsymbol{D}_0$ yields

$$\boldsymbol{D}_0\boldsymbol{\beta} = (\boldsymbol{\beta}'_0, \boldsymbol{0}'_{5\times 1}, \boldsymbol{0}'_{3\times 1}, \boldsymbol{0}'_{2\times 1})'.$$

From this discussion, it is clear that the matrix $\boldsymbol{D}_{S_t}$ effectively controls the prevailing structural exchange rate model and the set of covariates to include in the state-specific regression. Note that an MS kitchen-sink regression is obtained by defining $\boldsymbol{D}_{S_t}$ in such a way that in each state all economic indicators are included for all t.

Time-varying transition probabilities

Assuming constant transition probabilities is a standard (and potentially restrictive) assumption in MS models. Both Amisano and Fagan (2013) and Kaufmann (2015) propose treating the transition distributions as being dependent on additional covariates. Here, and since our model features K regimes, we follow Kaufmann and parametrize the transition probabilities by a multinomial logit specification. Given the forecasting evidence provided in the literature quoted above, we assume transition probabilities to depend on a measure of the monetary policy stance such as the policy rate. This captures the notion that if policy rates approach the ZLB, a Taylor rule-based model might become inadequate and the likelihood of a regime shift could increase.

The multinomial likelihood reads

$$P(S_t = j | S_{t-1} = k, \boldsymbol{Z}_t, \boldsymbol{\gamma}) = p_{kj,t} = \frac{\exp\left(\boldsymbol{Z}'_t\boldsymbol{\gamma}_{kj}\right)}{1 + \sum_{l=1}^{K-1} \exp\left(\boldsymbol{Z}'_t\boldsymbol{\gamma}_{kl}\right)},$$

with $\boldsymbol{\gamma}_{kj}{=}(\gamma_{0,kj},...,\gamma_{N-1,kj})'$ being category-specific regression coefficients, collected in $\boldsymbol{\gamma}$ for all k and $j{=}1,...,K{-}1$. $\boldsymbol{\gamma}_{kj}$ determines the sensitivity of the transition probability that drives the transition from the kth to the jth state. Moreover, $\boldsymbol{Z}_t$ denotes an N-dimensional vector of covariates, defined as

$$\boldsymbol{Z}_t = (1, \boldsymbol{z}'_t, \mathcal{I}[S_{t-1} = 1], ..., \mathcal{I}[S_{t-1} = K{-}1])',$$

where $\boldsymbol{z}_t$ is a vector of early-warning indicators that determine the dynamics of the transition probabilities, while $\mathcal{I}(\bullet)$ denotes an indicator function that equals one if its argument is true. This implies that we capture a first-order Markov structure by including the previous states as additional regressors. Moreover, $\gamma_{0,kj}$ represents the intercept of the reference state $S_{t-1}{=}0$, and thus captures the corresponding time-invariant state persistence. Consistent with Amisano and Fagan (2013), we let the coefficients associated with $\boldsymbol{z}_t$ be regime invariant.

TABLE 1 Prior mean $\underline{\boldsymbol{\beta}}_k$ for each state

	Intercept	i_{t-1}	i^*_{t-1}	π_t	π^*_t	x_t	x^*_t	q_t	m_t	m^*_t	y_t	y^*_t	e_t	p_t	p^*_t	i_t	i^*_t
$\underline{\boldsymbol{\beta}}_0$	0	0	0	1.5	−1.5	0.5	−0.5	0									
$\underline{\boldsymbol{\beta}}_1$	0								1	−1	1	−1	−1				
$\underline{\boldsymbol{\beta}}_2$	0												−1	1	−1		
$\underline{\boldsymbol{\beta}}_3$	0															1	−1

It is worth noting that, if the coefficients of $\mathbf{z}_t$ are zero, we obtain a classic fixed transition probability MS model.

The specific choice of $\mathbf{z}_t$ proves to be an important modeling decision. As mentioned above, our goal is to include a measure of the (conventional) monetary policy stance to signal a potential transition from Taylor rule-based policy making to discretionary monetary policy actions such as quantitative easing (QE). In our case, we assume two early-warning indicators $\mathbf{z}_t = (\tilde{i}_{t-1}, \tilde{i}_{t-1}^*)'$, the demeaned, lagged interest rate at home and in the USA. Demeaning covariates ensures that the time-invariant part does not depend on the scale of $\mathbf{z}_t$. A covariate can always be rewritten as a linear combination of its time-varying component and its mean, with the latter determining the time-invariant average state persistence (Kaufmann, 2015).

3.2 | Prior specification and estimation strategy

Our approach is Bayesian, which implies that we have to carefully specify priors on the parameters of the model. Here, we follow George and McCulloch (1993, 1997) and specify a mixture of Gaussians prior on β_{ik}, the ith element of $\boldsymbol{\beta}_k$. The prior is centered on theoretically motivated restrictions in order to test whether these restrictions are consistent with the data. The prior mean is stored in an M_k-dimensional vector $\underline{\boldsymbol{\beta}}_k$ and summarized in Table 1. We assume a symmetric Taylor rule with homogeneous coefficients for the home country and the USA and do not consider interest smoothing (see Molodtsova & Papell, 2009, for a detailed discussion). For the remaining models we center them on values consistent with the implied long-run fundamental value.

Formally, this prior reads

$$\beta_{ik}|\delta_{ik} \sim \mathcal{N}\left(\underline{\beta}_{ik}, \tau^2_{ik,1}\right)\delta_{ik} + \mathcal{N}\left(\underline{\beta}_{ik}, \tau^2_{ik,0}\right)(1-\delta_{ik}), \quad (3)$$

where we let $\tau^2_{ik,0}$ and $\tau^2_{ik,1}$ be prior variances (with $\tau^2_{ik,1} \gg \tau^2_{ik,0}$), for $i=1,...M_k$, and $\underline{\beta}_{ik}$ denotes the ith element of $\underline{\boldsymbol{\beta}}_k$. The first mixture component is referred to as the "slab" component, introducing almost no prior information, while the second is called the "spike" component, tightly centered around the prior mean $\underline{\beta}_{ik}$. The indicator δ_{ik} serves to select the mixture component used. Following the semiautomatic approach of George, Sun, D., and Ni (2008) we scale the prior variances, $\tau^2_{ik,0}$ and $\tau^2_{ik,1}$, with the variances of the ordinary least square estimates of the underlying structural model in state $S_t=k$.

TABLE 2 Transformation of variables

Variable	Description	Transformation
EXR	Nominal exchange rate	log difference
IP	Industrial production	log
M	Money aggregate	log
3M-IR	3M Money market rate	—
CPI	Consumer price index	log
INF	Inflation	log differences of CPI
REXR	Real exchange rate	$\log(\text{EXR}) + \log(\text{CPI}*) - \log(\text{CPI})$
IP-GAP[a]	Output gap	HP filter

[a]$\lambda=14{,}400$ for monthly data.

This modeling approach constitutes a data-driven way of assessing whether coefficients should be pushed towards theoretically motivated restrictions or allowed to be closely related to the corresponding maximum likelihood estimate. Thus, if $\delta_{ik}=0$, the posterior estimate of β_{ik} is strongly pushed towards the prior restriction $\underline{\beta}_{ik}$ while, in the opposite case, only little prior information on β_{ik} is introduced.

In what follows, we store all regime-specific indicators in a vector $\boldsymbol{\delta}_k = (\delta_{1k}, ..., \delta_{M_k})'$ that corresponds to the block of $\boldsymbol{\beta}$ associated with the kth structural model. Each element of the latent variable $\boldsymbol{\delta}_k$ is a priori independently Bernoulli distributed:

$$p(\delta_{ik} = 1|S_t = k) = \underline{\omega}_{ik},$$
$$p(\delta_{ik} = 0|S_t = k) = 1-\underline{\omega}_{ik},$$

for hyperparameters $\underline{\omega}_{ik}$ chosen by the researcher. A reasonable choice is $\underline{\omega}_{ik} = 0.5$, for all i,k, implying an equal prior probability of introducing significant prior information or using a relatively loose prior.[4]

For the variances σ^2, we assume an independent inverse Gamma prior for each element σ_i^2 ($i = 0, ..., K-1$). More specifically, we set

$$\sigma_i^2 \sim \mathcal{G}^{-1}(a_0, A_0),$$

with a_0 and A_0 being scalars. The specific values for a_0 and A_0 are chosen to be weakly informative with hyperparameters $a_0=0.01$, $A_0=0.01$.

---

[4]When considering a weakly informative coefficient prior, we define $\underline{\omega}_{ik}$ as being one for including all state-specific coefficients with certainty.

TABLE 3 Sources of economic fundamentals

Country	Coverage	EXR	IP (2010 = 100)	M	3M-IR	CPI (2010 = 100)
Australia (AU)	1975:M06–2017:M09	IFS	OECD	M1, OECD	OECD	OECD
Canada (CA)	1973:M01–2017:M09	IFS	OECD	M1, OECD	OECD	OECD
Japan (JP)	1973:M01–2017:M03	IFS	OECD	M1, OECD	IFS	OECD
Norway (NO)	1979:M01–2017:M09	IFS	OECD	M1, OECD	OECD	OECD
South Korea (KR)	1981:M01–2017:M09	IFS	IFS	M1, OECD	IFS	OECD
Sweden (SE)	1973:M01–2017:M06	IFS	OECD	M3, OECD	IFS	OECD
Switzerland (CH)	1974:M01–2017:M09	IFS	OECD	M1, OECD	OECD	OECD
United Kingdom (UK)	1973:M01–2017:M02	IFS	OECD	M0, IMF/FRED	IFS	OECD
United States (USA)	1973:M01–2017:M09		OECD	M1, OECD	OECD	OECD

Note. All quantities are seasonally adjusted, except EXR and 3M-IR. IP of Australia and Switzerland are interpolated to monthly frequency with cubic spline.

The prior distribution on the initial state is set to $p(S_0=k)=1/K$, for all k (Kaufmann, 2015). Finally, for the coefficients of the multinomial logit model, we adopt a weakly informative and symmetric prior across all states. That is,

$$\gamma_{kj} \sim \mathcal{N}(\boldsymbol{0}, \underline{\boldsymbol{V}}),$$

for all k and $j=1, ..., K-1$ with $\underline{\boldsymbol{V}} = \zeta \boldsymbol{I}_N$, and ζ denoting a scalar. In the empirical application we set $\zeta=100$.

In a Bayesian framework, we combine the likelihood with the prior to obtain the posterior distribution. In our case, the joint posterior density is intractable. Fortunately, the full conditional posterior distributions take simple forms, permitting Gibbs updating steps. The Markov chain Monte Carlo (MCMC) algorithm is described in more detail in the Appendix. In the empirical application, we repeat the algorithm 80,000 times, discard the first 30,000 draws as burn-in and define a thinning factor of 10, thus basing inference on 5,000 draws from the joint posterior.

Before proceeding to the empirical application, a brief word on identification is in order. Identification is necessary for structural interpretation of the states, but is not relevant if interest centers exclusively on the predictive density of the model (Frühwirth-Schnatter, 2001, 2006).[5] Recall that in the present model each regime is characterized by a different set of fundamentals, reflecting different theoretical exchange rate models. By exploiting the specific structure of the theoretical models, we impose inequality constraints on the coefficients by selecting the fundamentals for each regime. The only potential source of nonidentifiability occurs in the case of more than one state pointing towards a random walk. However, pushing coefficients in the direction of theoretical guided values is sufficient to disentangle regimes and fully identify the model. When considering the alternative specification, in which we always include all predictors, identification is certainly an issue. Moreover, each state is implicitly centered on a random walk a priori. In this case, we apply a permutation sampling step and solely focus on predictive densities.

4 | EMPIRICAL APPLICATION

This section starts by briefly describing the data set and forecasting design adopted in Section 4.1. We then discuss key in-sample features of the model in Section 4.2. Finally, Section 4.3 presents the main forecasting results, discussing the point and density forecasting performance of all models considered.

4.1 | Data, forecasting design, and competing models

In this paper, our aim is to forecast bilateral exchange rates for Australia, Canada, Japan, Norway, South Korea, Sweden, Switzerland and the UK relative to the US dollar. We collect monthly data on nominal exchange rates, industrial production, monetary aggregates, 3-month money market rates and consumer price indices for all countries under consideration (see Tables 2 and 3 for details).

In order to assess whether time-varying transition probabilities improve predictive accuracy, the proposed model framework is benchmarked with MS specifications featuring fixed transition probabilities (labeled MS-FT), as well as standard structural exchange rate models that are estimated under weakly informative

---

[5]MS models might suffer from identification problems due to the invariance of the likelihood with respect to permutations of the $K!$ possible labeling of the regimes, resulting in $K!$ modes.
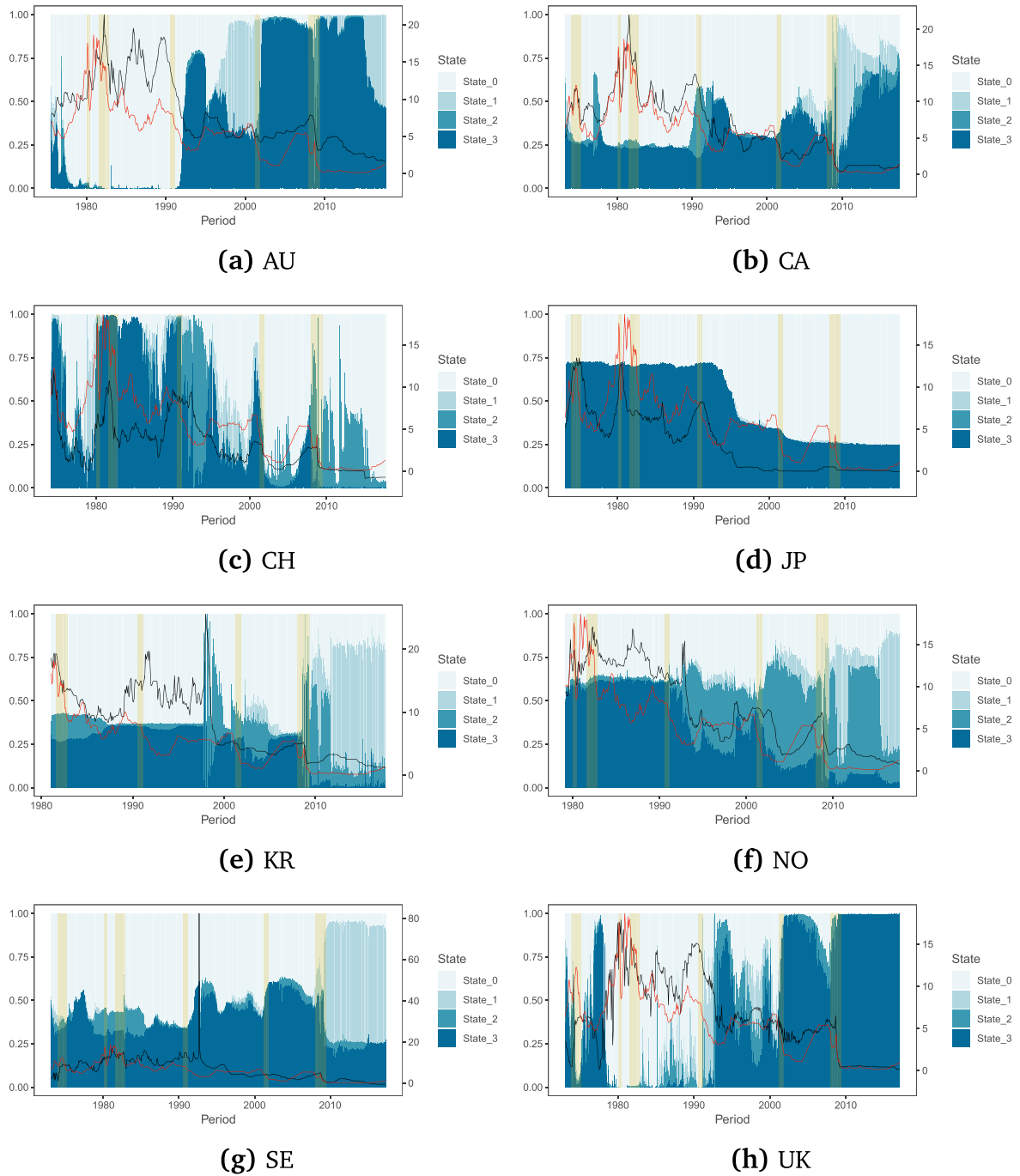
FIGURE 1 Filtered average posterior state probabilities for an SSVS prior and a common variance across states. State 0 indicates Taylor rule fundamentals, state 1 monetary fundamentals, state 2 PPP fundamentals, and state 3 UIP fundamentals. The vertical bars (yellow) indicate NBER recessions for the USA. The black solid line depicts the country-specific interest rate and the red solid line the interest rate for the USA. The left-axis shows the stacked probabilities and the right-scale the values of interest rates [Colour figure can be viewed at wileyonlinelibrary.com]

priors (labeled linear). These linear benchmarks are based on Taylor rule, monetary, PPP, and UIP fundamentals. The set of competing models can thus be divided into three overall classes: MS-TVP, MS-FT, and linear. Furthermore, we consider not only theoretically motivated MS-TVP and MS-FT specifications, but also models that include all macroeconomic indicators of X_{t-1} within each state (labeled kitchen-sink). For the

kitchen-sink regressions, we consider different numbers of states, ranging from two to four regimes. Moreover, to allow for state-specific shrinkage in kitchen-sink regressions, the SSVS prior described above is centered on zero and different state-specific indicators are estimated. To assess the role of allowing for heteroskedasticity in forecasting exchange rates, we also consider MS-TVP specifications with state-specific
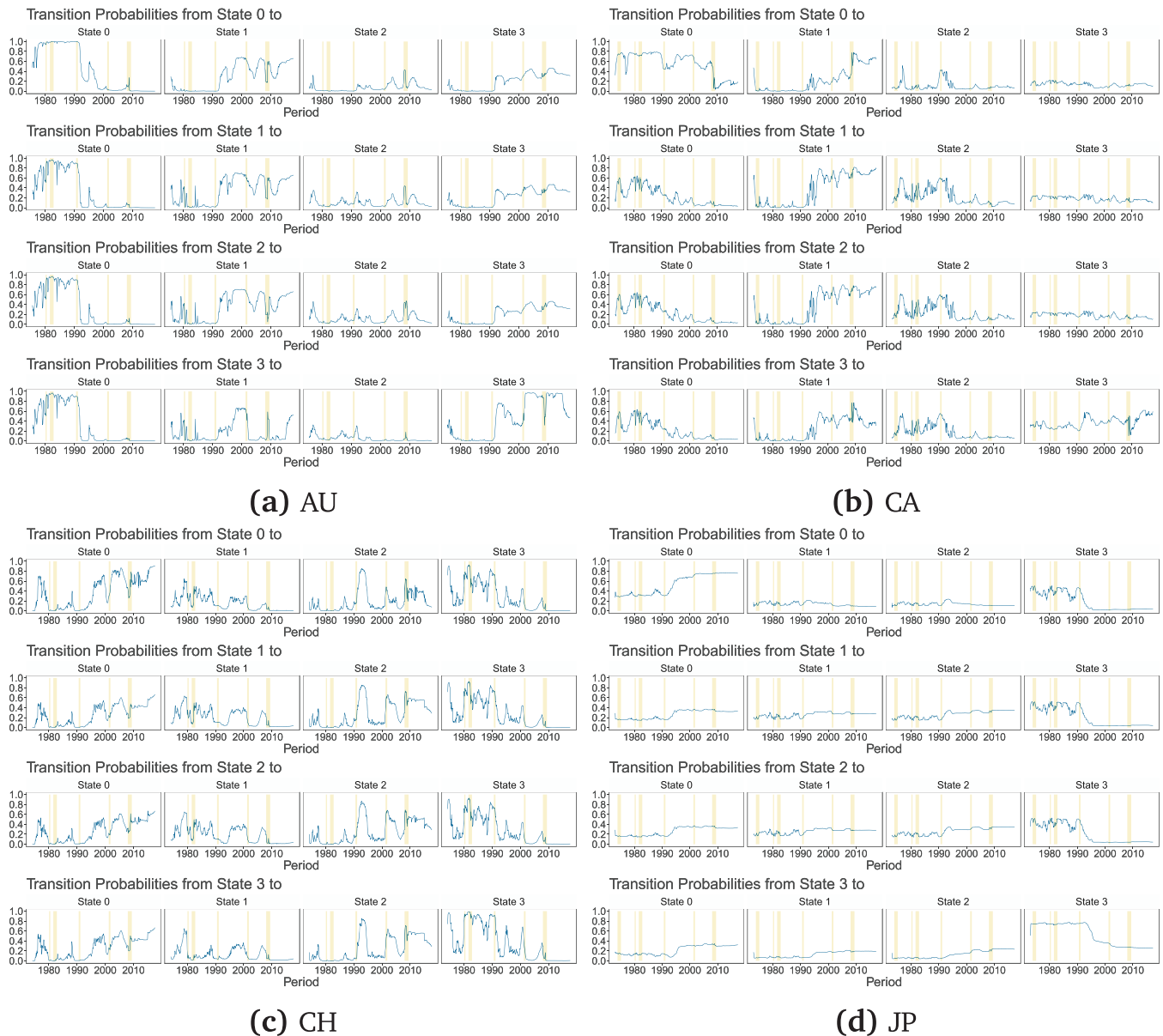
FIGURE 2 The blue line depicts the posterior mean of time-varying transition probabilities for each state with an SSVS prior and a common variance across states. State 0 indicates Taylor rule fundamentals, state 1 monetary fundamentals, state 2 PPP fundamentals, and state 3 interest rate fundamentals. The vertical bars (yellow) indicate NBER recessions for the USA [Colour figure can be viewed at wileyonlinelibrary.com]

variances. All models are then benchmarked to the random walk without drift.

We evaluate predictive accuracy by means of a recursive pseudo out-of-sample forecasting exercise. This implies choosing an initial estimation period that ranges from $t=1$ up to $t=T_0$, with the remaining periods used as a hold-out sample. In the present application, we estimate all models using data up to 2004:M12 and then proceed by computing h-step-ahead predictions for $t=T_0+1$. After obtaining draws from the corresponding predictive distributions, we consequently expand an initial estimation period by 1 month. This procedure is repeated until the end of the sample is reached.

To rank forecasts, we rely on cumulative squared forecast errors (CSFEs) to assess the quality of point forecasts. As point predictions, we take the posterior median of the predictive density. Turning to density forecasts, we follow Geweke and Amisano (2010) and rely on the log predictive score (LPS) to measure density forecasting accuracy. This has the advantage that, conditional on the proposed model and data, uncertainty surrounding the parameters and latent quantities is integrated out. After obtaining the LPS, we compute log predictive Bayes factors (LBFs) for the entire hold-out sample by computing the difference between the LPS of a given model and the LPS of the random walk.
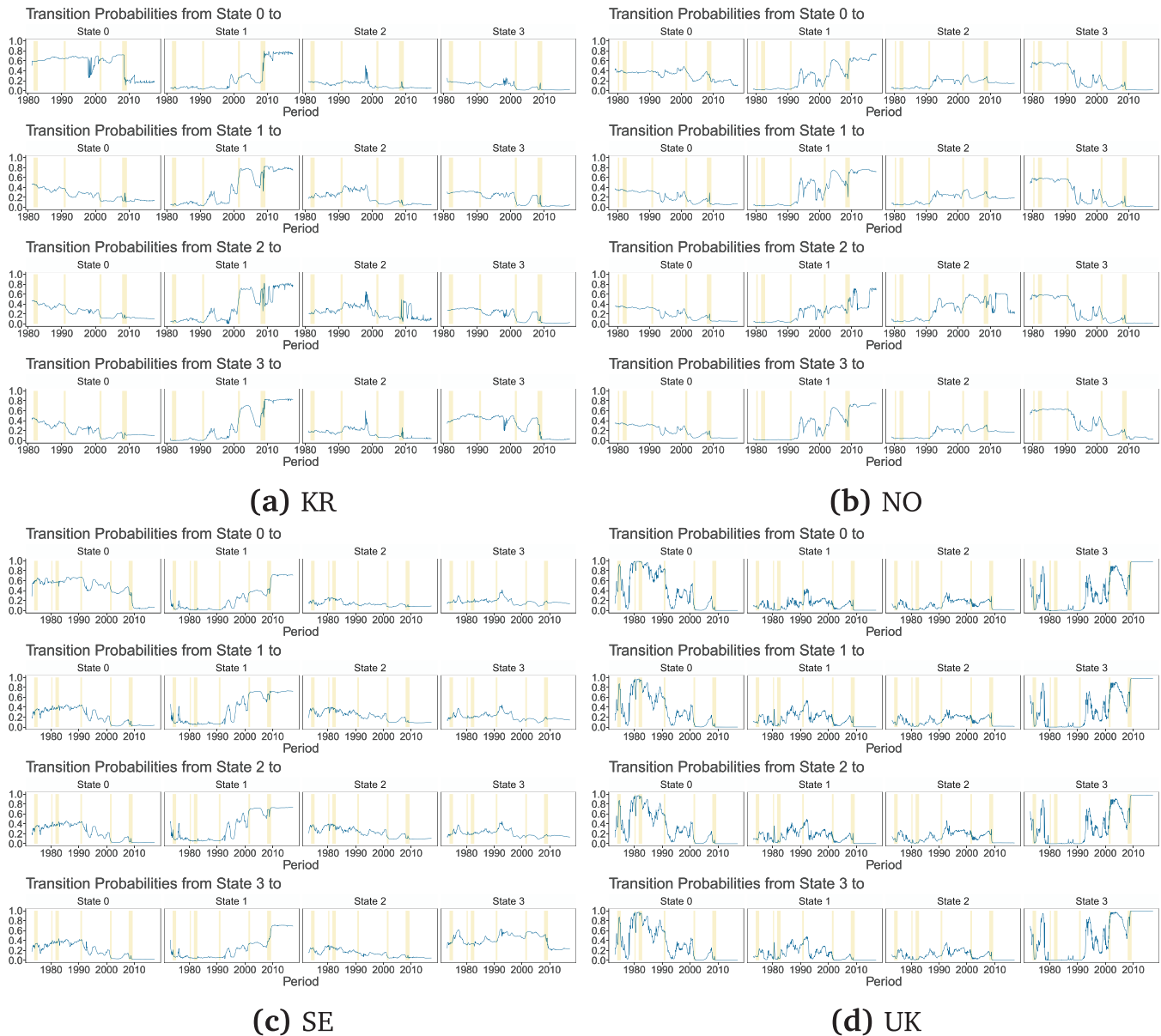
(a) KR

(b) NO

(c) SE

(d) UK

FIGURE 3 The blue line depicts the posterior mean of time-varying transition probabilities for each state with an SSVS prior and a common variance across states. State 0 indicates Taylor rule fundamentals, state 1 monetary fundamentals, state 2 PPP fundamentals, and state 3 interest rate fundamentals. The vertical bars (yellow) indicate NBER recessions for the USA [Colour figure can be viewed at wileyonlinelibrary.com]

4.2 | Some evidence for model instability

In this subsection we assess whether our proposed model signals significant shifts in the underlying structural representation. Figure 1 summarizes the mean of the filtered state probabilities for the eight exchange rates considered. In general, we observe that the regime dynamics across countries share one common feature. The models based on Taylor rule (state 0) and the UIP fundamentals (state 3) appear to dominate before the global financial crisis in 2008/2009. After that period, however, model evidence changes significantly for the majority of countries. More precisely, models based on monetary (state 1) and

PPP (state 2) fundamentals tend to receive more posterior support.

Compared to the remaining currencies, the Swiss franc (see Figure 1c) exhibits a somewhat higher regime-switching frequency. In addition, Figure 1 suggests that hitting the ZLB does shift filtered probabilities, pointing to regimes other than Taylor rule fundamentals (state 0) for countries such as Australia, Canada, South Korea, Sweden, and the UK.[6] Countries such as Japan and

---

[6]Note that Australia never hit the ZLB during the sample, but the USA did. Moreover, it is worth emphasizing that Japan already hit the ZLB in the midst of the 1990s.
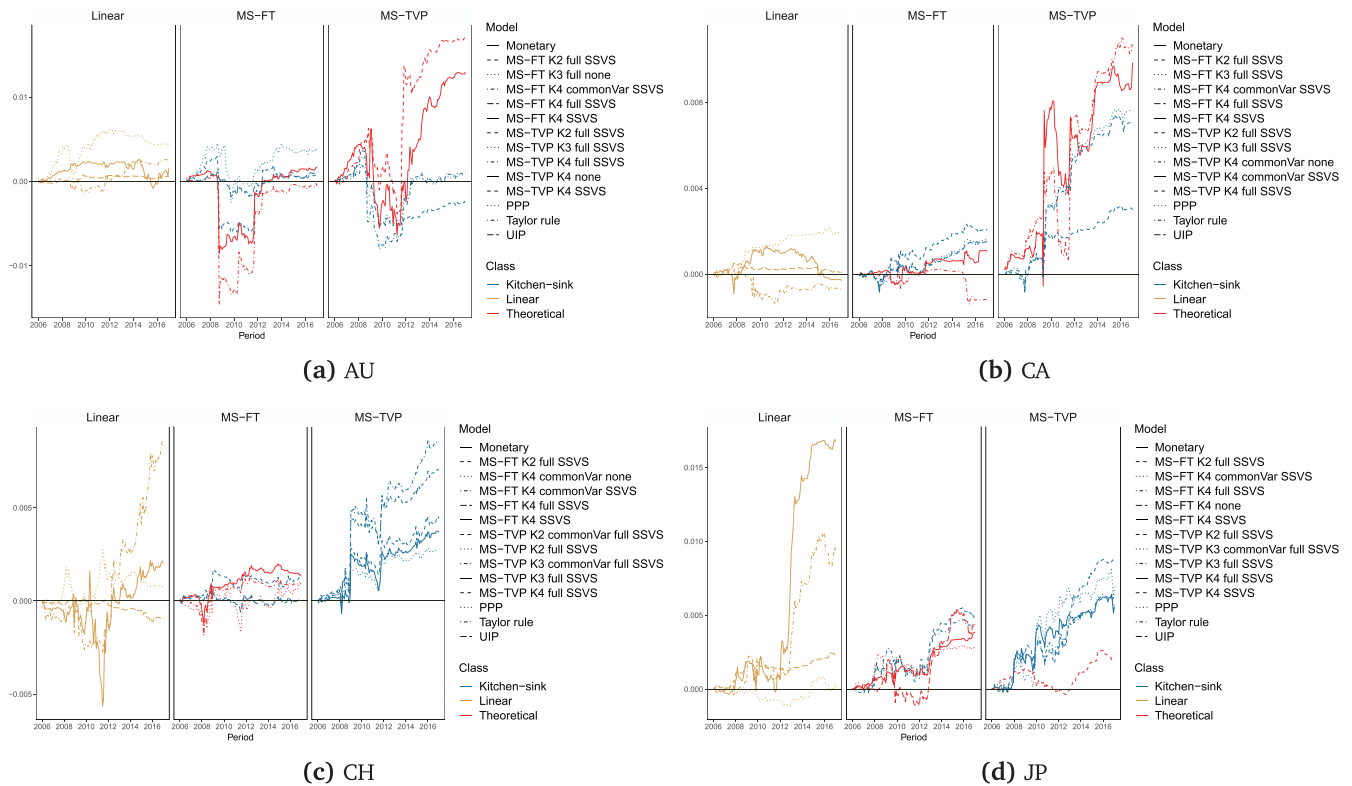
FIGURE 4 Difference to the random walk of one-step-ahead CSFEs for the full hold-out sample for Australia, Canada, Switzerland, and Japan. "Linear" specifies the linear univariate exchange rate regressions. For the Markov-switching models with time-varying transition probabilities (MS-TVP) and models with fixed transition probabilities (MS-FT), K[2–4] specifies the number of states. We evaluate all models with a common state variance (commonVar) and individual state variances, with both the theoretical state and the kitchen-sink (full) state specification. Moreover, we estimate all Markov-switching models with and without an SSVS prior. We consider the five best-performing MS-TVP and and five best MS-FT models according to the CSFEs at the end of the hold-out sample [Colour figure can be viewed at wileyonlinelibrary.com]

Switzerland, on the other hand, indicate an opposite dynamic, namely a shift of probabilities towards the Taylor rule state.

Taking a closer look at the UK, Taylor rule fundamentals are the predominant regime, reflecting the fact that these quantities tend to describe exchange rates well in times when the primary policy rule of the Bank of England is the Taylor rule. After this period, transition probabilities change during the crisis of the European Monetary System. After the financial crisis, and upon hitting the ZLB, the model based on Taylor rule fundamentals receives only limited posterior support. It is noteworthy that after 2010 the short-term interest rate (both at home and in the USA) is stuck at zero (and almost constant). This implies that the model based on interest fundamentals closely mimic a random walk during this period, even without introducing shrinkage.

The transition probabilities, depicted in Figures 2 and 3, generally track the movements in filtered state probabilities, providing considerable evidence of time-varying transition distributions. Our findings thus suggest that a measure of the monetary policy stance at home and

in the USA tends to drive transitions between structural models. This is consistent with our conjecture that during the period of the ZLB using Taylor rule-based exchange rate models might be inappropriate, at least from an in-sample perspective.

4.3 | Forecasting results

In this section, interest centers on the predictive performance of our proposed MS-TVP specification. The discussion in the last section provided evidence in favor of time-varying transition probabilities for several exchange rate pairs. This suggests that parametrizing the transition distributions with additional covariates helps to avoid situations where the model gets stuck within a certain state. Amisano and Fagan (2013) and Kaufmann (2015) highlight this issue and point towards advantages of explaining the regime-switching behavior of the model as opposed to a model based on constant transition probabilities. However, the key question is whether this additional flexibility also improves predictive performance. We answer this question using both point and density forecasts.
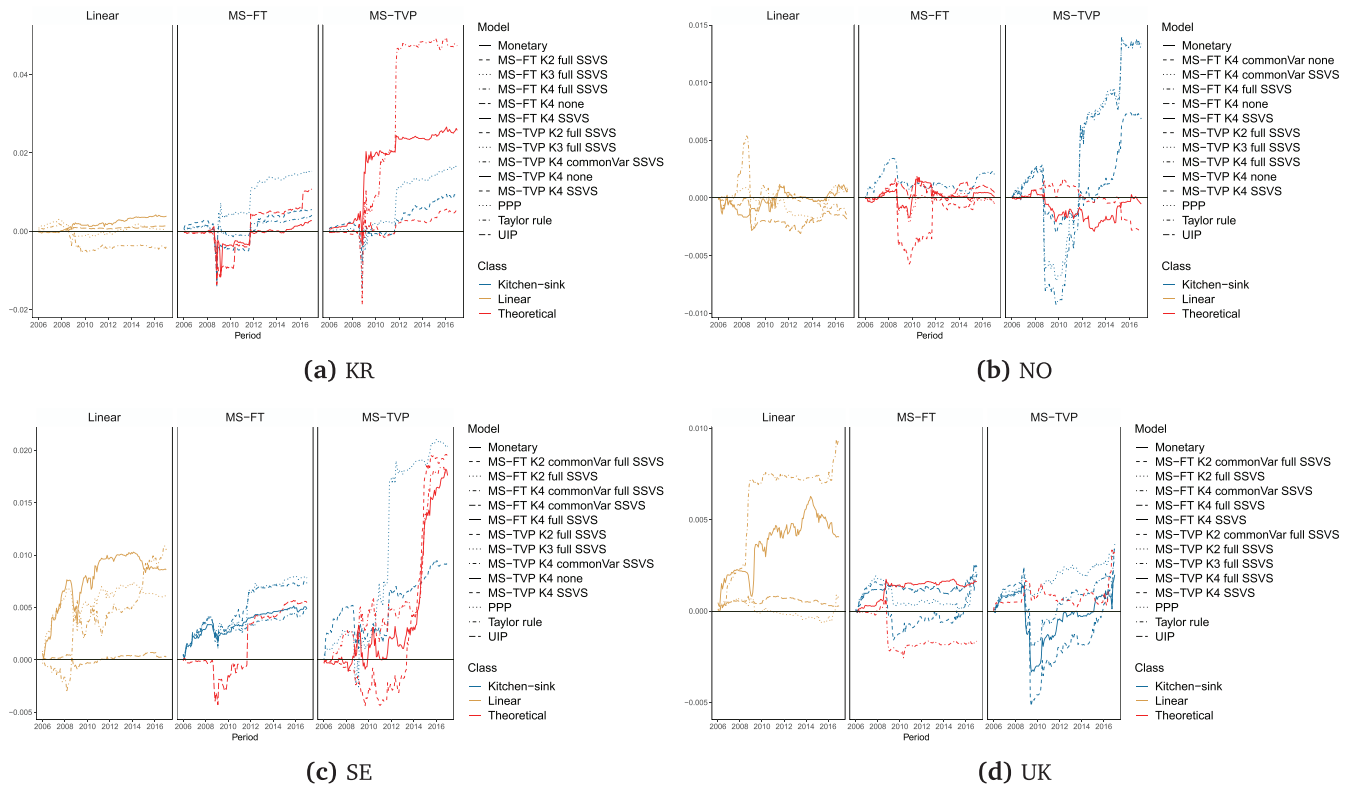
FIGURE 5 Difference to the random walk of one-step-ahead CSFEs for the full hold-out sample for South Korea, Norway, Sweden, and the UK. "Linear" specifies the linear univariate exchange rate regressions. For the Markov-switching model with time-varying transition probabilities (MS-TVP) and models with fixed transition probabilities (MS-FT), $K[2–4]$ specifies the number of states. We evaluate all models with a common state variance (commonVar) and individual state variances, with both the theoretical state and the kitchen-sink (full) state specification. Moreover, we estimate all Markov-switching models with and without an SSVS prior. We consider the five best-performing MS-TVP and and five best MS-FT models according to the CSFEs at the end of the hold-out sample [Colour figure can be viewed at wileyonlinelibrary.com]

4.3.1 | Point forecasts

Figures 4 and 5 present the evolution of CSFEs of one-step-ahead forecasts for the best-performing models across the set of model classes. We consider all linear predictive exchange rate regressions and the five best-performing MS-TVP and MS-FT models, according to the CSFEs at the end of the hold-out sample. The CSFEs of all models are shown as a difference between them and the CSFEs of the random walk benchmark. Thus values below zero indicate more accurate point forecasts than random walk predictions. Here, we focus on one-step-ahead forecasts, since we find that models that perform well at the one-step-ahead horizon also do well for $h>1$ periods ahead.[7] When considering density forecasts, we report the results for higher order predictions as well.

Turning to the actual results, we observe pronounced differences across countries. For instance, in Australia, Canada, Norway, Sweden, and the UK, modeling

nonlinearities pays off, in particular during periods of financial turmoil, outperforming forecasts of linear models and the random walk benchmark. Therefore, one interesting finding is that controlling for heteroskedasticity also tends to exert a positive effect on the point forecasting performance during volatile periods of the business cycle.

By contrast, for South Korea and Switzerland, the random walk appears to be hard to beat. In general, we observe CSFE differences that are close to zero when averaged over the full hold-out sample. This indicates that including more information does not necessarily translate into improved point predictions relative to a simple no-change forecast for these two economies. Again, we find some heterogeneity in relative forecasting performance over time.

Turning to the performance of the theoretically inspired MS-TVP specifications, we observe strong forecasting accuracy for Japan, Norway, South Korea, and the UK, at least for one specification (marked in red in Figures 6 and 7). On the other hand, it appears that kitchen-sink specifications dominate theoretically inspired regimes for Australia, Canada, Switzerland, and Sweden.

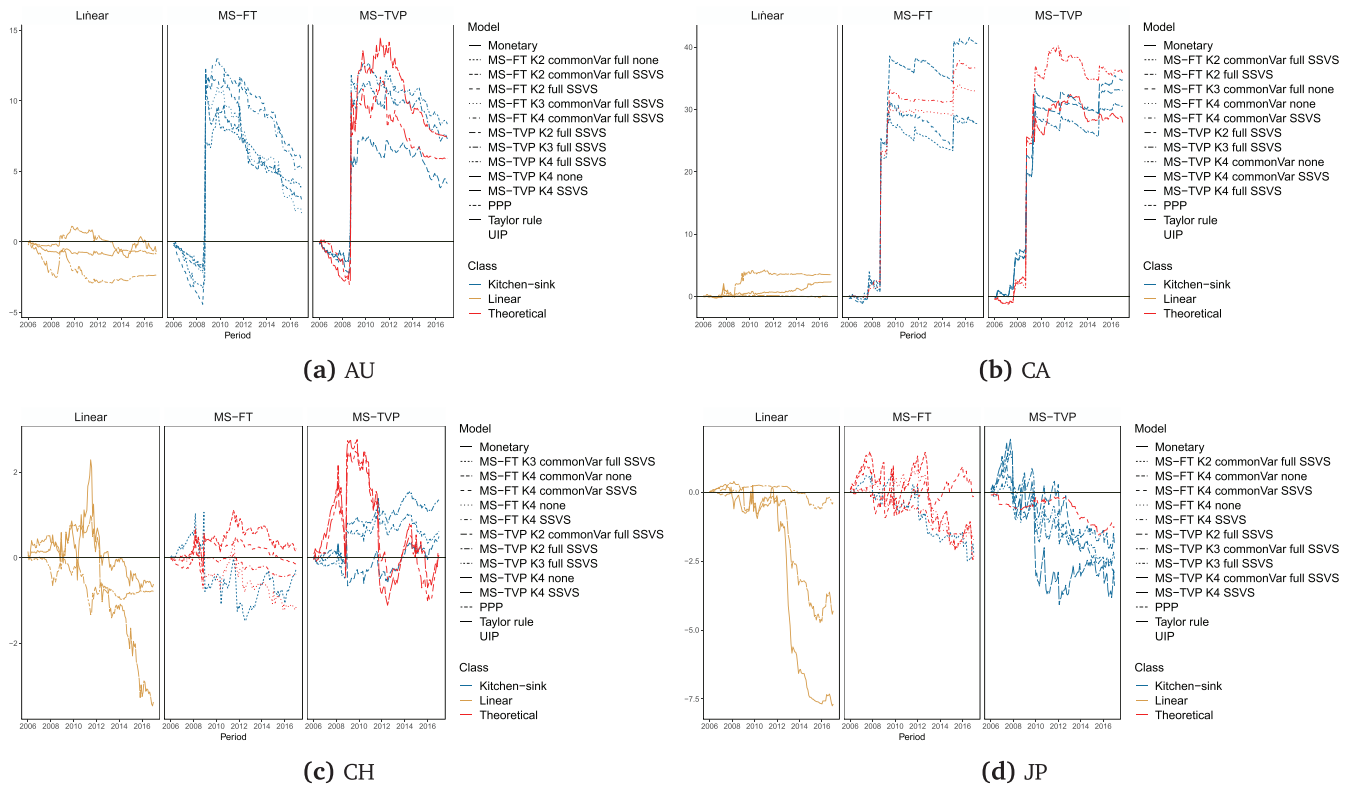[7]Additional results for $h(>1)$-step-ahead forecasts are provided upon request.

FIGURE 6 Cumulative one-step-ahead LBFs (random walk benchmark) for the full hold-out sample for Australia, Canada, Switzerland, and Japan. "Linear" specifies the linear univariate exchange rate regressions. For the Markov-switching models with time-varying transition probabilities (MS-TVP) and models with fixed transition probabilities (MS-FT), $K[2–4]$ specifies the number of states. We evaluate all models with a common state variance (commonVar) and individual state variances, with both the theoretical state and the kitchen-sink (full) state specification. Moreover, we estimate all Markov-switching models with and without an SSVS prior. We consider the five best-performing MS-TVP and and five best MS-FT models according to cumulative LBFs at the end of the hold-out [Colour figure can be viewed at wileyonlinelibrary.com]

In particular, this holds true for Switzerland as, indicated by the absence of a red-colored line in Figure 6c for MS-TVPs.

When comparing MS-TVP to MS-FT models, we observe that MS-FT specifications appear to be more robust over time in terms of CSFEs. This can be seen by noting that the forecast errors of the MS-FTs feature fewer outliers. In general, we find a better performance of MS-TVP models for Australia, Japan, and Norway, at least for one model, compared to MS-FT models, which display a remarkable performance for the UK and Canada.

Accuracy differences across simple linear specifications appear to be diverse. For the UK, Sweden, and Japan, we observe an inferior predictive performance for at least two linear models. In particular, models based on monetary fundamentals exhibit a weak forecast performance relative to the remaining models under scrutiny. For Australia, Canada, and South Korea, all linear models do well and show a similar point forecast performance as the random walk. When focusing on Taylor rule fundamentals, the linear regression performs well at the beginning of the hold-out sample (see, e.g.,

the results for the UK), but exhibits a systematic accuracy loss in periods after the financial crisis. This can be explained by the fact that Tayor rule-based models build on the assumption that both central banks' monetary policy might be well described by a Taylor rule. However, after the financial crisis, interest rates hit the ZLB and central banks increasingly adopted nonstandard policy measures. This, in turn, leads to a deteriorating performance of this model class, effectively confirming findings reported in the recent literature (see, e.g., Byrne et al., 2016; Molodtsova & Papell, 2012).

4.3.2 | Density forecasts

Tables 4 and 5 depict a summary of all models' LBFs for all currency pairs considered. Values larger than zero point towards outperformance of a model relative to the random walk, while negative values signal a weaker predictive performance when benchmarked against the random walk. To provide a dynamic picture of LBFs over time, Figures 6 and 7 show the LBFs of all linear
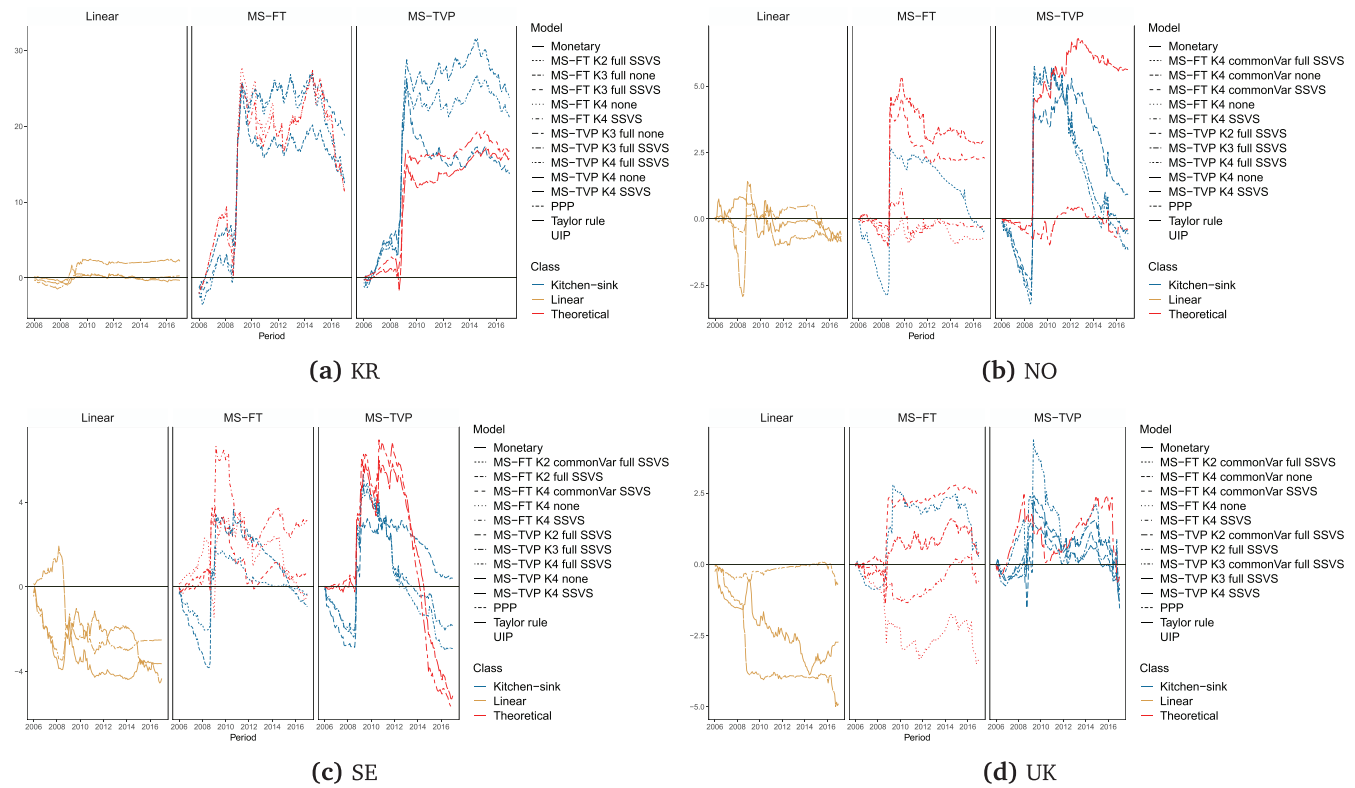
FIGURE 7 Cumulative one-step-ahead LBFs (random walk benchmark) for the full hold-out sample for South Korea, Norway, Sweden, and the UK. "Linear" specifies the linear univariate exchange rate regressions. For the Markov-switching models with time-varying transition probabilities (MS-TVP) and models with fixed transition probabilities (MS-FT), $K[2–4]$ specifies the number of states. We evaluate all models with a common state variance (commonVar) and individual state variances, with both the theoretical state and the kitchen-sink (full) state specification. Moreover, we estimate all Markov-switching models with and without an SSVS prior. We consider the five best-performing MS-TVP and and five best MS-FT models' cumulative LBFs at the end of the hold-out [Colour figure can be viewed at wileyonlinelibrary.com]

predictive exchange rate regressions and the five best-performing MS-TVP and MS-FT models.

In general, both tables attest to nonlinear specifications' good predictive power, while linear models display a somewhat weaker forecast performance. Furthermore, our results suggest that nonlinear models that perform well in terms of point predictions also exhibit high predictive capabilities in terms of density forecasts. However, we find that predictive performance evolves differently for CSFEs and the LBFs. Density forecasts strengthen the argument in favor of nonlinear models, as the performance gains of the MS models are more sizable in periods of high exchange rate volatility, whereas accuracy losses in tranquil periods are rather muted.

Although we do not observe a single dominant nonlinear model across forecast horizons and countries, Tables 4 and 5 suggest that at least one nonlinear specification outperforms the random walk and the linear competitors. One exception proves to be Japan, for which the best specification is either the random walk (for one-step-ahead predictions) or the linear PPP model (for longer horizons).

For all forecast horizons considered, nonlinear specifications perform well for Australia, Canada, South

Korea, Sweden, and the UK. Specifically, the MS-TVP kitchen-sink specification with four states coupled with state-specific variances, constituting the most flexible specification, has good predictive power for Australia, Canada, and South Korea. For Australia, this model is the single best-performing model across forecast horizons and, for South Korea, it is the best for three- and 12-step-ahead forecasts while being the second-best specification in terms of one-step-ahead forecasts. Moreover, as shown in Figures 6 and 7, the poor performance of linear exchange rate models is even more pronounced for density forecasts. In particular, linear structural regressions display a sharp decline in predictive power after the financial crisis, corroborating findings in Molodtsova and Papell (2012) and Byrne et al. (2016).

Turning to the question of whether allowing for heteroskedasticity pays off in terms of density forecasting, we find substantial evidence that this additional flexibility proves to be important. The gains in predictive accuracy of nonlinear models can mainly be attributed to the more flexible variance specification of MS models. This can also be seen by comparing MS specifications with a common variance across states with MS models

TABLE 4 Cumulative one-, three-, and 12-step-ahead LBFs (random walk benchmark) at the end of the full hold-out sample summarized for Australia, Canada, Switzerland, and Japan. Values highlighted green (red) are larger (smaller) than zero, indicating a better (weaker) performance compared to the random walk. Best model in bold [Colour table can be viewed at wileyonlinelibrary.com]

Specification			Log predictive Bayes factor					
States	Shrinkage	Variance	1-step	3-step	12-step	1-step	3-step	12-step
			AU			CA		
MS-TVP theoretical								
K = 4	None	Common	−16.208	−11.860	−7.476	35.802	28.588	22.695
	None	State-specific	7.449	−0.068	−0.953	1.911	6.161	28.652
	SSVS	Common	−6.908	−3.737	2.113	27.912	24.035	27.860
	SSVS	State-specific	5.904	−0.597	4.272	3.577	23.006	10.700
MS-FT theoretical								
K = 4	None	Common	−1.941	−2.809	−1.232	32.992	29.238	29.081
	None	State-specific	1.693	1.070	−2.181	8.760	7.406	1.590
	SSVS	Common	1.696	0.602	0.386	36.737	28.763	30.674
	SSVS	State-specific	1.261	0.144	−2.762	7.856	9.720	7.995
MS-TVP kitchen-sink								
K = 2	SSVS	Common	−22.236	−12.630	−26.266	−42.757	−38.942	−43.791
	SSVS	State-specific	4.142	5.954	2.386	33.067	33.136	31.270
K = 3	SSVS	Common	−19.408	−19.043	−24.598	−47.925	−47.945	−50.501
	SSVS	State-specific	7.306	6.406	5.417	30.543	31.877	24.891
K = 4	SSVS	Common	−23.884	−29.087	−29.255	−50.257	−57.303	−53.021
	SSVS	State-specific	**8.342**	**7.407**	**6.263**	34.765	35.336	29.015
MS-FT kitchen-sink								
K = 2	SSVS	Common	2.911	0.250	0.201	27.733	27.812	24.748
	SSVS	State-specific	5.765	4.801	3.528	**40.572**	**39.997**	**37.250**
K = 3	SSVS	Common	1.962	3.487	1.319	−31.072	−25.389	−23.545
	SSVS	State-specific	1.856	1.061	0.368	25.606	24.982	22.641
K = 4	SSVS	Common	5.192	3.376	3.624	24.224	27.691	19.228
	SSVS	State-specific	−16.540	−17.012	−17.715	15.123	14.930	12.654
Linear								
	Taylor rule		−0.866	−1.661	−0.592	3.496	3.324	−0.164
	Monetary		−0.709	−0.052	−1.914	2.379	1.594	0.007
	PPP		−2.362	−2.061	−3.797	0.050	0.347	−1.133
	UIP		−0.389	−0.286	−0.580	−0.479	0.478	0.207
			CH			JP		
MS-TVP theoretical								
K = 4	None	Common	−6.457	−4.135	−2.080	−11.153	−10.498	−6.965
	None	State-specific	0.081	−3.537	**0.655**	−5.688	−7.339	−1.672
	SSVS	Common	−4.213	−6.261	−0.866	−5.443	−4.734	−0.660
	SSVS	State-specific	0.176	−2.195	−1.363	−1.258	−3.183	−0.842
MS-FT theoretical								
K = 4	None	Common	0.210	−1.905	−2.192	−2.310	−3.072	−0.849
	None	State-specific	−1.142	−1.636	−1.787	−1.599	−1.536	−3.688
	SSVS	Common	−0.136	−0.471	−2.668	−0.151	−1.152	−1.759
	SSVS	State-specific	−0.385	−0.431	−3.329	−1.722	−0.936	−1.896
MS-TVP kitchen-sink								
K = 2	SSVS	Common	0.497	−2.012	−2.175	−9.272	−2.967	−1.222
	SSVS	State-specific	**1.353**	**0.408**	−0.569	−2.964	−2.992	−0.679
K = 3	SSVS	Common	−0.213	−2.987	−4.000	−1.906	−2.174	0.331
	SSVS	State-specific	0.619	−0.520	−1.308	−2.411	−2.367	−1.370
K = 4	SSVS	Common	−1.158	−5.658	−5.427	−3.400	−3.803	−1.843
	SSVS	State-specific	−0.543	−1.922	−2.322	−3.669	−3.662	−3.095
MS-FT kitchen-sink								
K = 2	SSVS	Common	−2.365	−2.906	−2.920	−2.230	−3.487	−2.038
	SSVS	State-specific	−5.686	−6.616	−6.344	−6.240	−8.560	−8.543
K = 3	SSVS	Common	−0.256	−3.726	−4.255	−3.233	−4.880	−3.300
	SSVS	State-specific	−6.504	−9.566	−10.908	−12.533	−13.159	−13.005
K = 4	SSVS	Common	−1.772	−3.870	−4.908	−8.006	−7.445	−6.180
	SSVS	State-specific	−27.363	−27.798	−27.909	−27.305	−26.146	−25.993
Linear								
	Taylor rule		−3.356	−2.715	−4.107	−4.302	−3.359	−3.218
	Monetary		−0.617	−2.009	−3.243	−7.684	−6.439	−1.597
	PPP		−0.787	−1.344	−1.349	−0.369	**0.193**	**0.769**
	UIP		0.090	0.002	0.035	−0.507	−0.548	−0.559

that feature state-specific variances. For these models, we observe a slight accuracy premium relative to their homoskedastic counterparts. Allowing for state-specific variances thus appears to be an important ingredient of a successful forecasting model. However, this increased flexibility comes at a cost. Specifically, we observe that during normal periods relative predictive accuracy declines steadily for several MS models (see, e.g., 7c). This observation confirms recent evidence provided in Abbate and Marcellino (2018).

TABLE 5 Cumulative one-, three-, and 12-step-ahead LBFs (random walk benchmark) at the end of the full hold-out sample summarized for South Korea, Norway, Sweden, and the UK. Values highlighted green (red) are larger (smaller) than zero, indicating a better (weaker) performance compared to the random walk. Best model in bold [Colour table can be viewed at wileyonlinelibrary.com]

Specification			Log predictive Bayes factor					
States	Shrinkage	Variance	1-step	3-step	12-step	1-step	3-step	12-step
			KR			NO		
MS-TVP theoretical								
K = 4	None	Common	−91.338	−73.679	−40.795	−3.684	−2.929	0.518
	None	State-specific	15.820	8.041	5.458	**5.623**	−0.229	0.430
	SSVS	Common	−48.414	−51.446	−43.495	−4.393	−5.983	**2.784**
	SSVS	State-specific	16.783	12.134	10.815	−0.411	**2.961**	0.638
MS-FT theoretical								
K = 4	None	Common	7.901	−30.386	4.123	2.898	0.733	1.624
	None	State-specific	11.263	12.679	8.562	−0.737	−1.887	6.043
	SSVS	Common	−5.469	0.267	2.419	2.278	2.075	1.958
	SSVS	State-specific	11.321	13.302	13.524	−0.273	0.219	1.010
MS-TVP kitchen-sink								
K = 2	SSVS	Common	−39.401	−84.104	−51.234	−4.853	−5.579	7.278
	SSVS	State-specific	12.215	9.430	14.090	0.908	0.514	1.961
K = 3	SSVS	Common	−114.149	−156.993	−126.614	−12.394	−10.818	14.418
	SSVS	State-specific	**23.788**	22.774	25.185	−0.577	−0.994	3.072
K = 4	SSVS	Common	−187.724	−190.133	−209.206	−16.165	−17.181	19.057
	SSVS	State-specific	21.200	**25.655**	**26.937**	−1.168	−1.968	3.554
MS-FT kitchen-sink								
K = 2	SSVS	Common	−57.328	−73.540	−42.487	−0.957	−1.397	1.684
	SSVS	State-specific	12.749	11.300	15.030	−1.861	−1.944	5.495
K = 3	SSVS	Common	−28.715	−72.196	−64.311	−1.758	−0.587	8.577
	SSVS	State-specific	18.781	16.662	17.812	−11.315	−12.182	14.188
K = 4	SSVS	Common	−50.802	3.179	5.573	−0.430	−0.716	2.885
	SSVS	State-specific	5.515	3.421	3.481	−21.445	−22.107	23.478
Linear								
Taylor rule			2.200	0.357	0.169	−0.601	0.605	3.410
Monetary			−0.324	−1.030	−1.239	−0.875	−0.115	2.687
PPP			0.289	−1.328	−0.407	−0.656	−0.071	1.127
UIP			−0.276	−0.083	−0.190	0.042	−0.087	0.354
			SE			UK		
MS-TVP theoretical								
K = 4	None	Common	−8.982	−9.025	−2.870	−5.982	−1.321	5.162
	None	State-specific	−5.147	−0.361	−2.487	−3.696	0.197	3.626
	SSVS	Common	−6.630	−5.702	**1.461**	−1.987	1.101	**6.036**
	SSVS	State-specific	−5.582	1.893	−0.237	−0.432	1.954	4.312
MS-FT theoretical								
K = 4	None	Common	−1.387	−1.828	−2.744	0.430	2.637	0.912
	None	State-specific	−0.370	1.454	−2.202	−3.299	−1.581	0.467
	SSVS	Common	0.694	1.635	0.271	**2.522**	**3.532**	1.887
	SSVS	State-specific	**3.139**	**3.433**	0.806	−0.595	0.148	0.734
MS-TVP kitchen-sink								
K = 2	SSVS	Common	−15.887	−25.834	−10.620	−1.545	0.370	0.034
	SSVS	State-specific	0.412	−0.519	−0.114	−1.090	2.290	2.428
K = 3	SSVS	Common	−16.742	−19.861	−13.622	−0.896	1.430	0.443
	SSVS	State-specific	−1.821	−0.810	−2.402	−1.363	1.245	0.352
K = 4	SSVS	Common	−20.343	−24.384	−21.758	−1.955	0.359	0.078
	SSVS	State-specific	−2.911	−2.124	−3.380	−2.086	−0.007	1.260
MS-FT kitchen-sink								
K = 2	SSVS	Common	−0.319	−1.792	−3.581	0.199	2.924	1.127
	SSVS	State-specific	−0.941	−3.452	−6.323	−5.674	−4.547	9.718
K = 3	SSVS	Common	−9.635	−4.012	−3.461	−5.519	−1.261	3.492
	SSVS	State-specific	−9.374	−10.639	−14.715	−14.169	−12.835	14.602
K = 4	SSVS	Common	−2.274	−2.466	−4.302	−8.277	−5.785	6.437
	SSVS	State-specific	−23.645	−25.141	−25.895	−26.066	−25.454	26.217
Linear								
Taylor rule			−4.328	−4.140	−3.620	−4.893	−3.199	0.890
Monetary			−3.632	−4.209	−7.265	−2.725	−1.626	2.589
PPP			−2.508	−2.435	−4.737	−0.652	0.335	0.748
UIP			0.275	−0.025	−0.139	−0.388	−0.373	0.194

Contrasting MS-TVP with MS-FT models, Figures 6 and 7 show that MS-TVP models yield more precise density forecasts for Australia, Norway, South Korea, and Switzerland, while yielding an almost equivalent performance for Canada. Moreover, with time-varying transition probabilities, theoretically motivated specifications play a more important role than for MS-FT models. This points towards potential accuracy premia obtained by allowing for time-varying transition probabilities and thus more precise inference surrounding the regime allocation.

Theoretically motivated MS-TVP specifications exhibit good forecast performance for Australia, Canada, and

Norway. Considering the results for Sweden, theoretically motivated MS-TVPs perform well during prolonged periods of high exchange rate volatility. In periods of low volatility, however, this specification is slightly outperformed by competing specifications. By contrast, we find that for South Korea MS-TVP kitchen-sink regressions improve upon our proposed MS-TVP models that allow for switching across structural exchange rate regressions. For the remaining countries in Figures 6 and 7, no clear pattern emerges when comparing both model types. For example, using structural MS-TVP specifications yields strong increases in predictive power during the financial crisis but a weaker performance afterwards. For kitchen-sink MS-TVP regressions, we find no gain during the crisis but also no subsequent loss in the aftermath of the crisis.

Finally, we assess whether using shrinkage priors on the coefficients improves forecasts. Tables 4 and 5 indicate that shrinkage generally translates into better results in pairwise comparisons with the corresponding nonshrinkage counterpart. This observation is not consistent across models, countries, and forecast horizons considered. In particular, using a kitchen-sink regression without shrinkage leads to poor forecast performance, as already shown by Wright (2008) and Li, Tsiakas, and Wang (2015). Turning to theoretically motivated MS models provides limited evidence on the usefulness of shrinkage priors. We conjecture that this stems from the fact that adopting MS specifications with theoretically defined regimes already introduces a certain amount of regularization that helps avoid overfitting.

5 | CONCLUDING REMARKS

In this paper, we propose a Bayesian nonlinear time series model to forecast exchange rates. Our framework allows for dynamically switching between selected theoretical exchange rate models that are used to guide the specific choice of covariates included. As an additional novelty, we assume that the transition probabilities vary over time and depend on a measure of the monetary policy stance at home and in the USA. This feature enables us to capture breaks in the policy rule of the central bank that, in turn, could impact the prevailing structural exchange rate model adopted. For instance, our framework entails dynamically switching between models if short-term interest rates hit the zero lower bound.

We use this framework to predict eight exchange rates vis-à-vis the US dollar. Considering the transition probabilities, we find considerable evidence of time variation. The filtered probabilities indicate that, especially after interest rates approach zero, model evidence shifts in favor of models other than the Taylor rule-based models, highlighting the necessity to control for model uncertainty. To assess whether this feature also translates into predictive accuracy gains, we conduct a forecasting exercise. We find that results appear to be rather mixed, with point forecasts being only slightly better than those obtained from standard models. For density predictions, by contrast, we observe pronounced accuracy increases for selected exchange rates.

DATA AVAILABILITY STATEMENT

The data sets as well as the estimation codes are available from the corresponding author upon request.

REFERENCES

Abbate, A., & Marcellino, M. (2018). Point, interval and density forecasts of exchange rates with time varying parameter models. *Journal of the Royal Statistical Society: Series A*, 181(1), 155–179.

Amisano, G., & Fagan, G. (2013). Money growth and inflation: A regime switching approach. *Journal of International Money and Finance*, 33, 118–145.

Beckmann, J., Koop, G., Korobilis, D., & Schüssler, R. (2018). Exchange Rate Predictability and Dynamic Bayesian Learning. (*Working Paper*). Leigh-on-Sea, UK: Essex Finance Centre.

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

Belmonte, M., & Koop, G. (2014). Model switching and model averaging in time-varying parameter regression models, *Bayesian model comparison*. Bingley, UK:Emerald Group, pp. 45–69.

Billio, M., Casarin, R., Ravazzolo, F., & Van Dijk, H. K. (2016). Interconnections between eurozone and US booms and busts using a Bayesian panel Markov-switching VAR model. *Journal of Applied Econometrics*, 31(7), 1352–1370.

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

Byrne, J. P., Korobilis, D., & Ribeiro, P. J. (2018). On the sources of uncertainty in exchange rate predictability. *International Economic Review*, 59(1), 329–357.

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.

Casarin, R., Sartore, D., & Tronzano, M. (2018). A Bayesian Markov-switching correlation model for contagion analysis on exchange rate markets. *Journal of Business and Economic Statistics*, 36(1), 101–114.

Chinn, M. D. (2006). The (partial) rehabilitation of interest rate parity in the floating rate era: Longer horizons, alternative expectations, and emerging markets. *Journal of International Money and Finance*, 25(1), 7–21.

Dornbusch, R. (1976). Expectations and exchange rate dynamics. *Journal of Political Economy*, 84(6), 1161–1176.

Engel, C. (1994). Can the Markov switching model forecast exchange rates? *Journal of International Economics*, 36(1–2), 151–165.

Engel, C. (2014). Exchange rates and interest parity. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.), *Handbook of international economics*, Vol. 4. Amsterdam, Netherlands:Elsevier, pp. 453–522.

Engel, C., & West, K. D. (2005). Exchange rates and fundamentals. *Journal of Political Economy*, 113(3), 485–517.

Engel, C., & West, K. D. (2006). Taylor rules and the Deutschmark–Dollar real exchange rate. *Journal of Money, Credit and Banking*, 38(5), 1175–1194.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics*, 12(3), 299–308.

Frankel, J. A. (1979). On the mark: A theory of floating exchange rates based on real interest differentials. *American Economic Review*, 69(4), 610–622.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96(453), 194–209.

Frühwirth-Schnatter, S. (2006). *Finite mixture and markov switching models*. Berlin, Germany:Springer Science and Business Media.

Frühwirth-Schnatter, S., & Frühwirth, R. (2010). *Data augmentation and MCMC for binary and multinomial logit models*, pp. 111–132.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423), 881–889.

George, E. I., & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7(2), 339–373.

George, E. I., Sun, D., & Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142(1), 553–580.

Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26(2), 216–230.

Groen, J. J. (2000). The monetary exchange rate model as a long-run phenomenon. *Journal of International Economics*, 52(2), 299–319.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ:Princeton University Press.

Hooper, P., & Morton, J. (1982). Fluctuations in the dollar: A model of nominal and real exchange rate determination. *Journal of International Money and Finance*, 1(2), 39–56.

Huber, F. (2016). Forecasting exchange rates using multivariate threshold models. *BE Journal of Macroeconomics*, 16(1), 193–210.

Huber, F. (2017). Structural breaks in Taylor rule based exchange rate models: Evidence from threshold time varying parameter models. *Economics Letters*, 150, 48–52.

Huber, F., & Fischer, M. M. (2018). A Markov switching factor-augmented VAR model for analyzing US business cycles and monetary policy. *Oxford Bulletin of Economics and Statistics*, 80(3), 575–604.

Huber, F., & Zörner, T. O. (2019). Threshold cointegration in international exchange rates: A Bayesian approach. *International Journal of Forecasting*, 35(2), 458–473.

Kaufmann, S. (2015). K-state switching models with time-varying transition distributions: Does loan growth signal stronger effects of variables on inflation? *Journal of Econometrics*, 187(1), 82–94.

Kim, C. J., & Nelson, C. R. (1999). *State-space models with regime switching: Classical and gibbs-sampling approaches with applications*. Cambridge, MA:MIT Press.

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

Mark, N. C. (1995). Exchange rates and fundamentals: Evidence on long-horizon predictability. *American Economic Review*, 85(1), 201–218.

Mark, N. C. (2009). Changing monetary policy rules, learning, and real exchange rate dynamics. *Journal of Money, Credit and Banking*, 41(6), 1047–1070.

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

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14(1), 3–24.

Molodtsova, T., Nikolsko-Rzhevskyy, A., & Papell, D. H. (2008). Taylor rules with real-time data: A tale of two countries and one exchange rate. *Journal of Monetary Economics*, 55, 63–79.

Molodtsova, T., Nikolsko-Rzhevskyy, A., & Papell, D. H. (2011). Taylor rules and the euro. *Journal of Money, Credit and Banking*, 43(2/3), 535–552.

Molodtsova, T., & Papell, D. H. (2009). Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics*, 77(2), 167–180.

Molodtsova, T., & Papell, D. H. (2012). Taylor rule exchange rate forecasting during the financial crisis. *NBER International Seminar on Macroeconomics*, 9(1), 55–97.

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using pólya–Gamma latent variables. *Journal of the American Statistical Association*, 108(504), 1339–1349.

Rapach, D. E., & 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.

Sarno, L. (2005). Towards a solution to the puzzles in exchange rate economics: Where do we stand? *Canadian Journal of Economics/Revue canadienne d'é,conomique*, 38(3), 673–708.

Sarno, L., Valente, G., & Wohar, M. E. (2004). Monetary fundamentals and exchange rate dynamics under different nominal regimes. *Economic Inquiry*, 42(2), 179–193.

Wright, J. H. (2008). Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics*, 146(2), 329–341.

AUTHOR BIOGRAPHIES

Niko Hauzenberger is a PhD student at the Vienna University of Economics and Business. He is currently working on a project titled 'High-Dimensional Statistical Learning: Application and Methods' funded by the Austrian Science Fund (FWF) and the Austrian Academy of Sciences (ÖAW) for Zukunftskolleg ZK35. His dissertation and research interests focus on Bayesian inference on highdimensional models using non-linear econometric techniques. He holds a Master degree in Economics and got a scholarship for outstanding academic performance from the Austrian Federal Ministry of Education, Science and Research.

Florian Huber is Professor of Economics at the Salzburg Centre of European Union Studies (SCEUS), Paris Lodron University of Salzburg. His research interests focus primarily on macroeconometric modeling and forecasting, time series analysis, and Bayesian econometrics. He has published in journals such as the Journal of Business & Economic Statistics, European Economic Review, Journal of Applied Econometrics, Macroeconomic Dynamics, Journal of Economic Dynamics and Control, and the Oxford Bulletin of Economics and Statistics, inter alia.

APPENDIX A

MCMC ALGORITHM

Let $\tilde{S}_T = (S_1, \ldots, S_T)'$ denote the full history of states and $\tilde{Z}_T$ be a $T \times N$-dimensional matrix of stacked covariates with the tth row given by Z_t'. After specifying appropriate starting values, the Gibbs sampler iterates through the following steps:

1. Sample parameters of the measurement equation $\theta = (\beta, \sigma^2, \delta)$ from $p(\theta|\tilde{S}_T, \Delta\tilde{e}_T)$, with $\Delta\tilde{e}_T = (\Delta e_1, \ldots \Delta e_T)$.
   a. Conditional on the exchange rate data $\tilde{e}_T$ and allocation of the states $\tilde{S}_T$, sampling β and σ^2 can be done in a standard way by drawing the coefficients β_k for $k = 0, \ldots, K-1$ in a block from a multivariate Normal distribution and the variances independently for each state from an inverse Gamma distribution.
   b. Conditional on β_k of state $S_t = k$, one is able to sample the elements of δ_k from a Bernoulli conditional posterior distribution.
2. For sampling the unknown states $p(\tilde{S}_T|\Delta\tilde{e}_T, \theta, \tilde{Z}_T, \gamma)$ we adapt the filtering algorithm put forth by Kim and Nelson (1999).
3. Following Polson, Scott, and Windle (2013), we sample multinomial coefficients from $p(\gamma|\tilde{S}_T, \tilde{Z}_T)$ to construct the time-varying transition probabilities.
4. In case of label switching, we implement an additional permutation step outlined in Frühwirth-Schnatter (2001) ensuring an equal probability of each mode appearing in the posterior distribution.

For the second step, we follow Kim and Nelson (1999) and sample $\tilde{S}_T$ in a block using a multimove Gibbs sampler. This implies simulating $\tilde{S}_T$ in block from the following joint conditional distribution:

$$p(\tilde{S}_T|\Delta\tilde{e}_T, \theta, \tilde{Z}_T, \gamma),$$

and applying a forward filtering and backward smoothing algorithm (Frühwirth-Schnatter, 2006).

Unfortunately, for step 3, there is no closed form of the multinomial logit posterior distribution. Frühwirth-Schnatter and Frühwirth (2010) therefore introduced two auxiliary layers for estimating state-specific utilities. Polson et al. (2013) took a different approach by representing the likelihood of multinomial logit model as a scale mixture of Gaussians with a Pólya Gamma mixing distribution. In a hierarchical form (by introducing a single layer of latent variables) this strategy implies that the multinomial coefficients are drawn from a set of Gaussians, and the auxiliary variables are sampled from a Pólya Gamma distribution. This approach has the advantage of fast convergence, simple implementation, and no need for an additional layer for approximating the error distribution.

Following Polson et al. (2013), conditional on the latent states $\tilde{S}_T$ and the auxiliary variables ψ_{jt} for $t = 1, \ldots, T$ sampled from a $\mathcal{PG}(\psi_{jt}|1, 0)$ distribution, the posterior quantities are given by

$$\bar{V}_j = \left(\tilde{Z}_T'\Psi_j\tilde{Z}_T + \underline{V}^{-1}\right)^{-1}$$
$$\bar{\gamma}_{kj} = \bar{V}_j\left(\tilde{Z}_T'\tilde{\kappa}_j\right)$$

where γ_{k0} is zero for $k=1,\ldots,K-1$ for reasons of identification, $\Psi_j=\mathrm{diag}(\psi_{j1},\ldots,\psi_{jT})$, and

$$\tilde{\kappa}_j = (\mathcal{I}[S_t = j]-0.5)-\Psi_j \boldsymbol{C}\boldsymbol{e}_j.$$

$\boldsymbol{C}$ is a $T\times K-1$-dimensional matrix, with elements being defined as

$$C_{tl} = \log\left(\sum_{k \neq l} \exp(\boldsymbol{Z}_t'\boldsymbol{\gamma}_{kj})\right),$$

and $\boldsymbol{e}_j$ represents the unit vector with a one at the jth position.