Dynamic Synthesis of Point Forecasts: A Cointegrated Bayesian Framework*

Zheng Fan†

November 27, 2025

Check the latest version HERE

Abstract

This paper develops a new Bayesian framework for forecast combination that embeds a Vector Error Correction Model (VECM) structure within a latent dynamic factor representation for model weights. By imposing cointegration across the latent weight structure, the framework establishes a structured yet flexible constraint on their long-run behavior while allowing for short-term deviations. The specification enables the weights to evolve around a stable equilibrium and to adjust rapidly toward a new equilibrium when the predictive environment changes. Monte Carlo studies and empirical applications to economic growth and inflation forecasting show that the proposed approach delivers substantial improvements in predictive performance over state-of-the-art models. The framework advances the literature by considering two key features: the ability of Bayesian Predictive Synthesis (McAlinn and West, 2019) to accommodate flexible, non-convex weights, and the capacity of Infinite Markov Pooling (Jin et al., 2022) to enable weights to transition across multiple equilibria.

Keywords: Bayesian Econometrics, Predictive Synthesis, Model Combination, VECM, Markov-switching, Shrinkage, Macroeconomic Forecasting

JEL Classification: C11, C32, C52, C53, E37

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

†PhD Candidate, Department of Economics, University of Melbourne, Level 4, FBE Building, 111 Barry Street, Carlton, VIC, Australia 3010, email: zhengfan2015@gmail.com

1 Introduction

Uncertainty is a central aspect of economic decision-making, shaping how households, firms, and policymakers form expectations and make intertemporal choices. In this context, accurate forecasting plays a substantial role in guiding monetary policy, investment strategies, and risk management. A single econometric model often fails to capture the full extent of uncertainty and structural complexity inherent in economic systems. This reality has motivated a shift toward forecast combination and predictive synthesis, where information is integrated from multiple agents or models to improve robustness.

This paper proposes a coherent Bayesian framework for forecast combination that embeds a Vector Error Correction Model (VECM) structure within a latent dynamic factor representation for model weights. The key idea is to impose cointegration across the weights so that they 'jump' around multiple long-run equilibria while remaining unrestricted to fluctuate in response to short-term information.[1] This construction provides a regularized mechanism that incorporates both local flexibility with global stability.

To accommodate both gradual evolution and abrupt shifts in the underlying predictive landscape, the framework incorporates a latent Markov-switching error correction mechanism of VECM. This specification explicitly links the error correction mechanism to a time-varying 'restoring force': a 'loose and slow' regime permits weights to explore local deviations (mimicking standard dynamic averaging), while a 'tight and fast' regime enforces rapid convergence. This permits time variation in the correction strength, where small adjustments produce local fluctuations while large adjustments signal transitions in the equilibrium configuration. Consequently, this state-dependent flexibility enables the model to detect shifts in the predictive environment and trigger a rapid re-alignment of the weights toward a new equilibrium.

A growing literature has developed around forecast combination, pooling, and synthesis methods (see Nonejad (2021); Martin et al. (2024) for recent reviews). Prominent approaches range from Bayesian model averaging (Hoeting et al., 1999) and static pooling

[1] Here, 'long run equilibrium' is used in the VECM sense, referring to cointegrating relationships among the model weights, which includes multiple possible weight combinations. It should not be confused with an economic steady state from structural growth theory.

(Hall and Mitchell, 2007; Geweke and Amisano, 2011) to more flexible dynamic pooling frameworks (Del Negro et al., 2016) and their Markov-switching extensions (Waggoner and Zha, 2012; Jin et al., 2022). More recently, the focus has shifted toward Bayesian predictive synthesis (BPS) (Johnson, 2017; McAlinn and West, 2019; McAlinn et al., 2020; Aastveit et al., 2023) and its generalization to Bayesian predictive decision synthesis (BPDS) (Tallman and West, 2024). These latter contributions highlight the importance of agent-based synthesis where weights are treated as dynamic regression coefficients rather than simple and convex probabilities.

This paper makes three primary contributions to the forecast combination literature. First, it introduces a novel formulation in which the model weight vector evolves according to a Vector Error Correction Model (VECM) within a latent dynamic factor structure. This imposes a cointegrating relationship that anchors weights to a long-run equilibrium while permitting short-run deviations. Second, the framework incorporates a latent state process on the adjustment coefficient, enabling a state-dependent 'restoring force'. This mechanism allows the model to distinguish between mild local fluctuations (weak correction) and structural shifts requiring rapid convergence (strong correction). Third, the proposed specification combines key features of the existing literature: it retains the flexibility of Bayesian Predictive Synthesis (McAlinn and West, 2019) to accommodate non-convex weights, while capturing the regime-switching dynamics of the Infinite Markov Pooling (Jin et al., 2022) to allow transitions across multiple equilibrium configurations, all within a coherent, VECM-driven framework.

The proposed framework formulates forecast combination as a latent variable problem, focusing on the dynamics of the model weights. Conceptually, the approach adopts the 'synthesis' perspective found in McAlinn and West (2019): it treats latent weights not merely as probabilities, but as dynamic coefficients in a synthesis function. This representation preserves the interpretability of the Bayesian Predictive Synthesis (BPS) framework, which views weights as portfolio allocations that reflect relative precision and hedging motives, while adopting a clearer regularization structure. This perspective also connects naturally to ensemble learning methods in the machine learning literature, particularly stacking techniques that combine multiple predictive models (e.g.,

Wolpert (1992); Breiman (1996); Dietterich (2000); Proscura and Zaytsev (2022); Muslim et al. (2023)).

The core novelty of this framework is the imposition of a Vector Error Correction Model (VECM) structure on the latent weight vector. This specification governs the weight evolution through a disciplined error-correction process, ensuring they respond promptly to changes in the relative predictive performance of the component models. The VECM introduces a cointegrating relation that anchors the weights to a long-run relationship, such as encouraging weights to sum to one, while still allowing short-run deviations. This design allows the weights to respond efficiently to new information without drifting in an unstructured manner, thereby improving predictive performance in settings with changes in the predictive environment.

One of the key challenge in standard VECMs is distinguishing long-run equilibrium relationships from short-run fluctuations (e.g., Juselius (2006); Garratt et al. (2008); Koop et al. (2011)). To help with identification, researchers often employ informative priors. In the context of forecast pooling, however, this framework proposes to pre-specify the equilibrium relationship rather than estimate it. By defining the cointegrating vector a priori, the model enforces a structured evolution where weights ultimately converge toward a coherent configuration, yet retain the flexibility to take negative values or temporarily depart from equilibrium, which is a flexibility that aligns with the desirable properties of predictive synthesis.

Estimating this structure requires the simultaneous inference of unobserved latent weights and their adjustment speeds. In this latent setting, standard shrinkage priors (Geweke, 1996; Strachan and Inder, 2004) can be problematic: excessive shrinkage effectively reduces weights to random noise, while insufficient regularization leads to unstable overshooting. To resolve this, the framework adopts a Markov-switching specification for the adjustment (error-correction) coefficients. Unlike approaches that assume smooth parameter evolution (Koop et al., 2011), this specification allows for discrete shifts in adjustment strength (see Jochmann and Koop (2015)). This enables the model to switch between a "loose" regime (weak correction) that permits local exploration, and a "tight" regime (strong correction) that enforces rapid re-alignment.

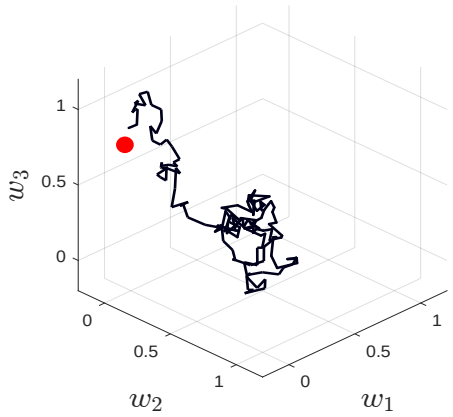
Eventually, the proposed method, VECM with Markov-switching (VECM-MS), captures the strengths of two state-of-the-art approaches into a single coherent structure. From Bayesian Predictive Synthesis (McAlinn and West, 2019), it retains the interpretation of weights as dynamic synthesis coefficients, allowing for flexible and hedging weights. From the Infinite Hidden Markov Pool (IMP) (Jin et al., 2022), it adapts the capacity to handle structural shifts via regime switching.

Although the Infinite Markov Pooling (IMP) moves among distinct clusters of model combinations, the proposed VECM-MS framework attains comparable flexibility through a continuous structure. In this formulation, the long-run equilibrium is not a single point but a unit complex, such as the weights summing to one, that can be satisfied by many different weight vectors. Both approaches reflect the idea that no single combination is appropriate for all periods. IMP captures this by switching across clusters, whereas the VECM-MS framework allows the weights to be 'corrected' towards another admissible combination. When the predictive environment shifts, the latent state process on the adjustment term permits rapid movement toward a new configuration while preserving the coherent equilibrium structure implied by the VECM-MS.

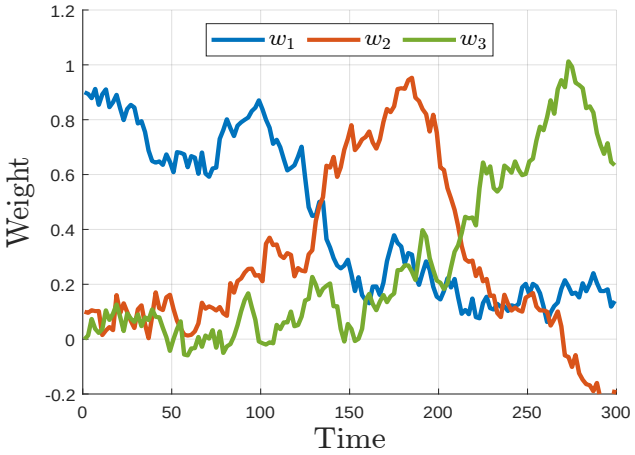
For illustration, Figure 1 contrasts the data-generating processes for 3 model weights under the three specifications. The left panel displays a 3D representation of the time evolution of the weight vector, and the right panel shows each of the time-varying weight trajectories. The top panel (Panel A) displays Bayesian Predictive Synthesis (BPS), where weights evolve as a smooth, drifting random walk. The middle panel (Panel B) illustrates the Infinite Hidden Markov Pooling (IMP) approach, characterized by discrete, abrupt jumps between static regimes strictly located on the unit simplex. The bottom panel (Panel C) demonstrates the proposed VECM-MS specification, which unifies these dynamics. As shown, the weights exhibit local fluctuations similar to those of BPS (reflecting the 'loose' regime), but are capable of rapid, jump-like realignment toward the cointegrating equilibrium (reflecting the 'tight' regime). This visualizes how the framework effectively balances the local flexibility of synthesis methods with the structural adaptability of regime-switching models.

To assess the proposed combination method, a series of Monte Carlo experiments is

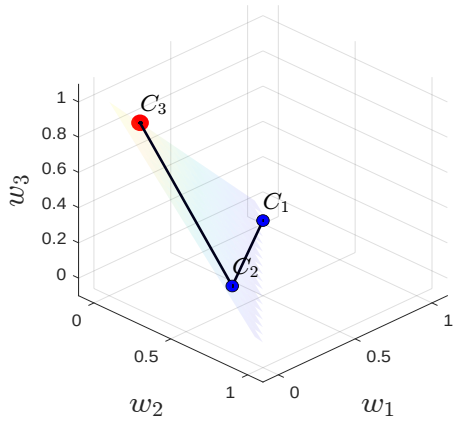
(A) BPS: 3D Path of Weights



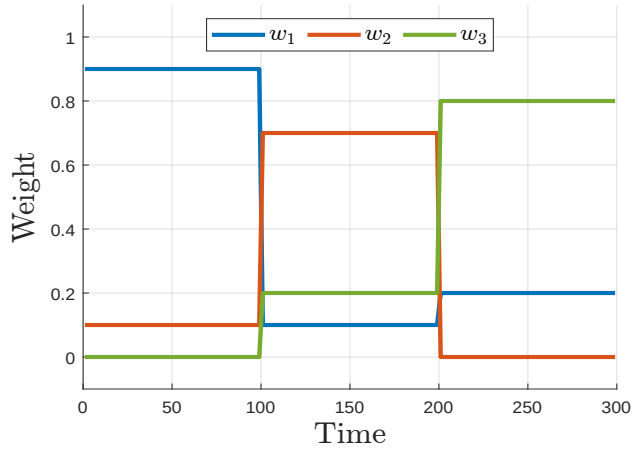
Evolution of Weights over Time



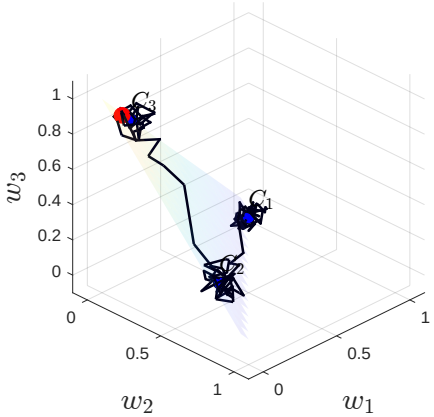
(B) IMP: 3D Path of Weights



Evolution of Weights over Time



(C) VECM-MS: 3D Path of Weights



Evolution of Weights over Time

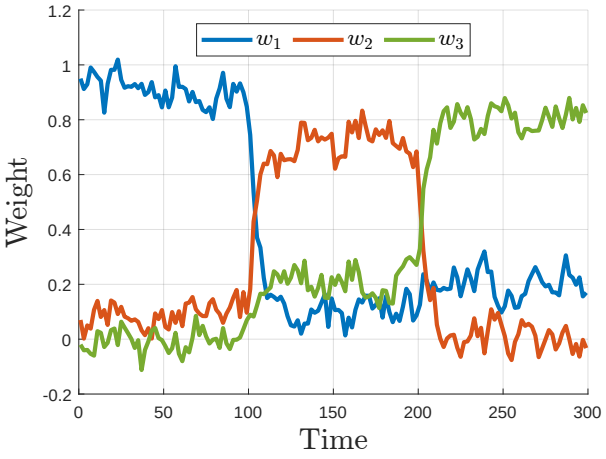


Figure 1: Comparison of illustrative data-generating processes across specifications. Panel (a) shows the smooth random walk of BPS. Panel (b) shows the discrete regime jumps of IMP. Panel (c) shows the proposed VECM-MS dynamics. Click on each panel to view the corresponding animation (external link).

conducted and consistently show that the framework outperforms state-of-the-art benchmark models, confirming its robustness and practical relevance. Building on these results, the method is applied to forecasting economic growth and inflation, where the incorporation of the VECM structure yields substantial improvements in predictive accuracy across multiple horizons.

The remainder of the paper is organized as follows. Section 2 outlines the proposed methodology. Section 3 details the Bayesian inference and forecasting procedures. Section 4 presents a Monte Carlo study assessing model performance. Sections 5 and 6 report empirical applications to economic growth and inflation forecasting, respectively. Section 7 concludes. Additional MCMC details and supplementary results are provided in the Appendix.

2 Econometric Model

Model combination techniques have evolved significantly: shifting from relying on the full set of candidate models toward selecting informative subsets (Madigan and Raftery, 1994; Onorante and Raftery, 2016), from static pooling schemes with fixed weights (Hall and Mitchell, 2007; Geweke and Amisano, 2011) to dynamic approaches that adapt over time (Koop and Korobilis, 2012; Del Negro et al., 2016), and from simple linear combinations to more flexible hierarchical frameworks (McAlinn and West, 2019; Jin et al., 2022).

Within a state space representation of forecast combination models, which provides a natural statistical foundation for dynamic systems, the Gaussian Dynamic Linear Model (DLM) is the most widely used specification (Harrison and Stevens, 1976; West and Harrison, 2006). In this setting, a linear relationship is assumed between the observables and latent states, with normally distributed innovations:

$$y_t = \omega_{0,t} + \boldsymbol{x}_t' \boldsymbol{\omega}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right) \tag{2.1}$$

$$\boldsymbol{\omega}_t = \boldsymbol{\alpha}_0 + A\boldsymbol{\omega}_{t-1} + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_\omega\right) \tag{2.2}$$

where the vector $\boldsymbol{x}_t$ is a $p \times 1$ collection of predictions from each base model (expert), and

the time varying model weights $\boldsymbol{\omega}_t$ are also represented as a $p \times 1$ vector. The intercept term $\omega_{0,t}$ evolves as a random walk, given by $\omega_{0,t} = \omega_{0,t-1} + u_{0,t}$, where $u_{0,t} \sim \mathcal{N}(0, \sigma^2_{\omega_0})$. The variance of error terms, σ^2_t and $\boldsymbol{\Sigma}_\omega$, are pre-determined or constant.

Recent work in macroeconomics and finance emphasizes the importance of time varying volatility. Stochastic volatility (SV) models have gained strong empirical support and are now common in applied research (e.g., Primiceri (2005); Justiniano and Primiceri (2008); Nakajima (2011); Chan and Eisenstat (2018); Chan et al. (2024)). To accommodate this feature, SV is incorporated into the measurement disturbance of the DLM (Aguilar and West, 2000) by specifying $\sigma^2_t = \exp(h_t)$, where

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + e_t, \qquad e_t \sim N\left(0, \sigma^2_h\right), \tag{2.3}$$

so that the log-volatility h_t evolves according to a stationary AR(1) process with $|\phi_h| < 1$. This formulation differs from the BPS model of McAlinn and West (2019), in which the variance of the error term is modeled using a beta-gamma random walk specification.

2.1 VECM Specification

To incorporate the VECM specification into the latent structure while preserving a clear economic interpretation, I follow the framework of Johansen (1995) and adopt a restricted intercept in the cointegration space. In this setting, the intercept affects the system only through the long-run equilibrium relationship such that the cointegrating relations have non-zero means. The specification is given by rewriting equation (2.2):

$$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}_0 + \Pi\boldsymbol{\omega}_{t-1} + \boldsymbol{u}_t \tag{2.4}$$

Time varying regularization of the long run equilibrium in the model weights is introduced through the decomposition of the VECM parameters, where $\Pi = A - I$ is commonly expressed as $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$, with both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being $p \times 1$ column vectors. The vector $\boldsymbol{\alpha}$, known as the adjustment or error correction coefficient, captures short run deviations, while $\boldsymbol{\beta}$, the cointegrating vector, defines the long run equilibrium relationship. The re-

stricted intercept is defined as $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}\mu$, implying that the constant term enters only through the cointegration component. Substituting this representation and rearranging yields

$$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}\left(\mu + \boldsymbol{\beta}'\boldsymbol{\omega}_{t-1}\right) + \boldsymbol{u}_t, \tag{2.5}$$

where $\Delta\boldsymbol{\omega}_t, \boldsymbol{\omega}_{t-1}, \boldsymbol{u}_t, \boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ are all $p \times 1$ column vectors. The vector $\boldsymbol{\beta}$ is pre-specified as a vector of -1's and μ as the scalar 1, such that the term $(1 + \boldsymbol{\beta}'\boldsymbol{\omega}_{t-1})$ explicitly measures the deviation of the summation of weights from unity. This design introduces long-run dynamics to the unit simplex. The short-run adjustment toward the long-run equilibrium is governed by $\boldsymbol{\alpha}$, which directly determines the evolution of the model weights. This simplifies equation (2.5) to

$$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}(1 - \sum_j \omega_{j,t-1}) + \boldsymbol{u}_t. \tag{2.6}$$

In summary, the complete VECM specification is given by

$$y_t = \omega_{0,t} + \boldsymbol{x}_t'\boldsymbol{\omega}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma_t^2), \tag{2.7}$$

$$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}(1 - \sum_j \omega_{j,t-1}) + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\omega). \tag{2.8}$$

One key challenge in this framework lies in how to appropriately treat $\boldsymbol{\alpha}$, which governs the speed of convergence to the long-run equilibrium. Although the structure of the transition equations resembles that of the standard VECM study, the underlying intuition and complexity differ substantially, especially when considering a latent environment.

The first fundamental difference is that in the standard VECM, $\boldsymbol{\omega}$ is treated as a known macroeconomic data input, whereas in my setting, it is modeled as a set of latent, time-varying model weights, introducing additional layers of complexity and model uncertainty. Following the conventional estimation may lack reliability.

Secondly, while shrinkage priors on $\boldsymbol{\alpha}$ are commonly used in the VECM literature to effectively eliminate the influence of non-cointegrated relationships, applying similar

shrinkage here would undesirably weaken or even eliminate the convergence mechanism to the long-run equilibrium. In the extreme case, leading to a standard random walk. Even though short-term large deviations may occur only occasionally, the magnitude of the correction is attenuated and dominated by the majority of periods in which the system remains near equilibrium and requires little or no adjustment. Conversely, failing to impose appropriate shrinkage on $\boldsymbol{\alpha}$ may allow its magnitude to influence the evolution of the latent weights, potentially resulting in excessive volatility or undesirable drift. This behavior can be interpreted as an over-adjustment in the updating mechanism, causing the weights to oscillate around the equilibrium.

Therefore, to capture the time-varying speed of convergence, I specify the adjustment coefficient $\boldsymbol{\alpha}_{s_t}$ as a Markov-switching process (e.g., Jochmann and Koop (2015)). This flexible specification is particularly valuable in the presence of structural shift, where traditional time-varying parameter models (typically evolving as random walks) often exhibit sluggish adaptation due to their inherent persistence.[2]

By permitting $\boldsymbol{\alpha}_{s_t}$ to switch instantaneously to a higher magnitude, the framework allows the weights to re-align rapidly with the long-run equilibrium following a shock. Intuitively, this implies that the "restoring force" of the error correction mechanism is state-dependent: it exerts fast jump during transitions (high values of $\boldsymbol{\alpha}_{s_t}$) to force realignment, while relaxing this constraint (low values of $\boldsymbol{\alpha}_{s_t}$) during stable periods to permit local fluctuation.

Thus, the primary model of interest, VECM-MS, can be written as

$$y_t = \omega_{0,t} + \boldsymbol{x}_t' \boldsymbol{\omega}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right), \tag{2.9}$$

$$\Delta \boldsymbol{\omega}_t = \boldsymbol{\alpha}_{s_t}(1 - \sum_j \omega_{j,t-1}) + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_\omega\right). \tag{2.10}$$

Here, the latent state vector $\boldsymbol{s} = (s_1, \ldots, s_T)$ governs the regime-switching behavior of the error-correction coefficients. Specifically, I assume a two-state process (though this is easily extendable): a *slow-adjustment state* ($s_t = 0$) and a *fast-adjustment state* ($s_t = 1$).

[2]Throughout this paper, 'structural shift' refers to regime shifts in the **predictive performance** of the model pool that necessitate a re-weighting. While often triggered by structural breaks in the underlying economy, our modeling focus is on the resulting discontinuity in the optimal weight vector.

In the slow-adjustment regime ($s_t = 0$), the prior specification for $\boldsymbol{\alpha}$ follows the conventional shrinkage approach found in the VECM literature (e.g., Koop et al. (2009)), suppressing large changes to prevent overfitting. In contrast, in the fast-adjustment regime ($s_t = 1$), $\boldsymbol{\alpha}$ is assigned a weakly informative prior with relatively large variance. This relaxation allows the adjustment coefficients to take on large magnitudes, enabling the system to rapidly correct deviations and drive the weights toward the long-run equilibrium—a solution space characterized not by a single point, but by infinitely many valid weight combinations.

Economically, the magnitude of the adjustment coefficient $\boldsymbol{\alpha}$ determines the system's 'restoring force'. A larger magnitude of $\boldsymbol{\alpha}$ implies a faster transition to equilibrium, while a smaller value indicates slower adjustment and more persistent deviations. The concept of *half-life* provides a convenient measure of this adjustment speed, representing the expected time for a deviation to decline by half. Accordingly, in the fast-adjustment state, where $\boldsymbol{\alpha}_{s_t}$ takes larger magnitudes, the implied half-life is shorter, indicating faster equilibrium reversion, whereas in the slow-adjustment state, smaller $\boldsymbol{\alpha}_{s_t}$ values correspond to slower correction dynamics.

The incorporation of the Markov-switching state vector $\boldsymbol{s}$ parallels the logic of the Infinite Hidden Markov Pool (IMP) (Jin et al., 2022): the Markov-switching adjustment term enables the model to transition swiftly across 'data-implied regimes,' thereby capturing abrupt structural changes on model weights. Crucially, however, the framework imposes minimal structure on the individual weights, enforcing only the long-run equilibrium relationship without restricting specific values. This design preserves the unrestricted support characteristic of Bayesian Predictive Synthesis (BPS) (McAlinn and West, 2019), allowing weights to take negative or large positive values if supported by the data. Consequently, unlike the static within-regime weights of Jin et al. (2022), this specification permits weights to evolve as random walks within each regime. This specification permits realistic fluctuations while maintaining a coherent long-run structure, avoiding the rigidity of fixed-weight assumptions while ensuring the system remains anchored to long-run equilibrium relationships.

The long-run equilibrium implies that the weights sum to one, consistent with their

interpretation as portfolio allocations on the unit simplex. This condition functions as a target rather than a hard constraint; temporary deviations are permitted and are governed by the state-dependent transition parameter $\boldsymbol{\alpha}_{s_t}$, which introduces flexibility in the adjustment toward equilibrium. In practice, shrinkage in $\boldsymbol{\alpha}_{s_t}$ can reduce the strength of the error-correction adjustment when appropriate, allowing the individual weights to behave approximately as random walks.

To further generalize the framework, the unit-sum condition can be relaxed by treating the cointegrating intercept, μ, as an estimable paramete. This specification, which I denote as VECM-MS-μ, is defined by:

$$y_t = \omega_{0,t} + \boldsymbol{x}_t'\boldsymbol{\omega}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right) \tag{2.11}$$

$$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}_{s_t}(\mu - \sum_j \omega_{j,t-1}) + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_\omega\right). \tag{2.12}$$

I call this VECM-MS-μ. In this specification, μ governs the long-run aggregate weights, while all other parameters are defined as before. This generalization is particularly advantageous for accommodating systematic biases in the model pool; for instance, in scenarios where all base models consistently underpredict or overpredict the outcome, estimating μ allows the framework to intrinsically correct for this deficiency by scaling the combined forecast.

2.2 Full Model

VECM-MS Model. To summarize, the full specification of the VECM-MS model is given by:

$$y_t = \omega_{0,t} + \boldsymbol{x}_t'\boldsymbol{\omega}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right), \tag{2.13}$$

$$\sigma_t^2 = \exp(h_t), \tag{2.14}$$

$$h_t = \mu_h + \phi_h\left(h_{t-1} - \mu_h\right) + e_t, \qquad e_t \sim \mathcal{N}\left(0, \sigma_h^2\right), \tag{2.15}$$

$$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}_{s_t}(1 - \sum_j \omega_{j,t-1}) + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_\omega\right), \tag{2.16}$$

$$P_{mk} = P\left(s_t = k \,|\, s_{t-1} = m\right). \tag{2.17}$$

for $j = 1, .., p$, $t = 1, ..., T$ and $m, k \in \{0, 1\}$. All other notations are defined as before.

Simple State-Space Model (SSM) Benchmark. A competing benchmark is the Bayesian Predictive Synthesis of McAlinn and West (2019) based on a simple state-space model (SSM), where I modify the inputs to use point forecasts and specify the error terms with stochastic volatility to provide a fair comparison with the proposed methods.

$$y_t = \omega_{0,t} + \boldsymbol{x}_t'\boldsymbol{\omega}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right), \tag{2.18}$$

$$\boldsymbol{\omega}_t = \boldsymbol{\omega}_{t-1} + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_\omega\right), \tag{2.19}$$

where all notations follow the previous definitions. This specification is equivalent to the VECM described earlier with $\boldsymbol{\alpha} = \mathbf{0}$, reducing the transition equation to a simple random walk.

Infinite Markov Simplification (IMS) Benchmark. As the proposed VECM framework also accommodates discrete regime shifts to capture structural changes, I include the Infinite Markov Pooling (IMP) model of Jin et al. (2022) as an additional benchmark. To ensure a fair comparison, all competing models differ only in how the model weights are updated. This specification is therefore theoretically distinct from the original IMP but

remains consistent with its underlying idea by adopting a simpler finite mixture approximation to represent an infinite mixture structure (see Rousseau and Mengersen (2011) for theoretical justification of finite approximations to infinite mixtures). In this simplified setting, the infinite state space is approximated by a large but finite number of regimes, allowing the data to automatically deactivate redundant states.

The model is given by

$$y_t = \omega_{0,t} + \boldsymbol{x}_t' \boldsymbol{\omega}_{s_t} + \varepsilon_t, \tag{2.20}$$

$$s_t \mid s_{t-1} \sim \text{Categorical}(\boldsymbol{\pi}_{s_{t-1}}), \tag{2.21}$$

$$\boldsymbol{\pi}_i \sim \text{Dirichlet}(\alpha_0^{\pi} \boldsymbol{\beta}^{\pi}), \tag{2.22}$$

$$\boldsymbol{\beta}^{\pi} \sim \text{Dirichlet}\left(\frac{\gamma}{K} \mathbf{1}_K\right), \tag{2.23}$$

$$\boldsymbol{\omega}_{s_t} \sim \text{Dirichlet}(\lambda \mathbf{1}_p), \tag{2.24}$$

In this specification, $\boldsymbol{x}_t$ represents the $p \times 1$ vector comprising the collection of p base forecasts. The regime-specific weight vector $\boldsymbol{\omega}_{s_t}$, constrained to the simplex, governs the relative contribution of each predictor in regime s_t. But note the regime indicator here is different from the one used for the error correction coefficient. The latent regime indicator $s_t \in \{1, \ldots, K\}$ evolves according to a finite-state Markov chain with transition probabilities $\boldsymbol{\pi}_i$, while the global measure $\boldsymbol{\beta}^{\pi}$ controls the overall frequency of regimes. All other components remain identical to the baseline specification; in particular, the intercept $\omega_{0,t}$ follows a random-walk evolution, and the error term ε_t is modeled with stochastic volatility.

When the truncation level K is set sufficiently high (e.g., $K = 30$), the posterior distribution naturally empties redundant states, yielding a practical finite approximation to the nonparametric IMP of Jin et al. (2022) (Ishwaran and Zarepour, 2002). While a 'sticky' extension can be readily implemented by adjusting the transition matrix priors to favor state persistence, which is the version used in all subsequent exercises. This IMS (with stickiness) benchmark not only provides a fair comparison with existing frameworks but also demonstrates that abrupt jumps alone may not be sufficient.

All Models Summary. The distinctions between the proposed VECM approach and the benchmarks lie solely in the specification of the transition equation (2.16). For clarity, all alternative formulations of the transition equation (2.16) are summarized in the following table 1.

Table 1: Transition Equations for Competing Model Specifications

Model	Transition Equation
SSM (benchmark)	$\Delta\boldsymbol{\omega}_t = \boldsymbol{u}_t$
VECM	$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}\left(1 - \sum_j \omega_{j,t-1}\right) + \boldsymbol{u}_t$
VECM-MS (key interest)	$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}_{s_t}\left(1 - \sum_j \omega_{j,t-1}\right) + \boldsymbol{u}_t$
VECM-MS-μ	$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}_{s_t}\left(\mu - \sum_j \omega_{j,t-1}\right) + \boldsymbol{u}_t$
IMS	$\boldsymbol{\omega}_{s_t} \sim \text{Dirichlet}(\lambda \mathbf{1}_p)$

Notes: The measurement equation is identical for all models: $y_t = \omega_{0,t} + \boldsymbol{x}_t'\boldsymbol{\omega}_t + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$ and $\boldsymbol{u}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\omega)$. SSM denotes a state space model with random walk weights and no equilibrium restriction. VECM incorporates a sum-to-one restriction on weights with constant adjustment speed $\boldsymbol{\alpha}$. VECM-MS allows for Markov-switching adjustment speeds, where s_t denotes the regime state. VECM-MS-μ additionally estimates the long-run sum μ, accommodating systematic deviations from unity. The IMS is implemented as a sticky, finite approximation to the nonparametric Infinite Markov Pool.

3 Bayesian Inference and Forecasting

I use Bayesian computation to conduct inference on the VECM and its variant models. Using the techniques widely used in VECM, state space model, and the mixture technique for the SV specification (Omori et al., 2007), I aim to infer the augmented posterior:

$$p(\Psi \mid \boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}), \tag{3.1}$$

where $\boldsymbol{y}_{1:T} = (y_1, \ldots, y_T)'$, $\boldsymbol{x}_{1:T} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ with $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{p,t})'$ and the model unknowns are collected in Ψ. The model unknowns include auxiliary variables and the stochastic volatility mixture representation, both of which are structured to facilitate estimation through Gibbs sampling within the Markov chain Monte Carlo framework. The elements of Ψ, along with their respective description and dimensions, are summarized

in Table 2 below.

Table 2: Summary of the elements of Ψ from the VECM-MS model

Notation	Description	Dimension
$\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_T)'$ with $\boldsymbol{\omega}_t = \{\omega_{i,t}\}_{i=1}^{p}$	Model weights	$T \times p$
$\boldsymbol{\Sigma}_\omega$	State disturbance covariance matrix	$p \times p$
$\boldsymbol{\omega}_0 = (\omega_{0,1}, \ldots, \omega_{0,T})'$	Bias correction term (intercept)	$T \times 1$
$\sigma^2_{\omega_0}$	Process variance of $\boldsymbol{\omega}_0$	1
$\boldsymbol{\alpha}$	Error correction coefficients	$p \times 1$
$\boldsymbol{s} = (s_1, \ldots, s_T)$	State indicator for error correction coefficients	$T \times 1$
$P \in \mathbb{R}^{2\times 2}$	State transition probability matrices	2
$h = (h_1, \ldots, h_T)'$	Log volatility of the measurement equation	$T \times 1$
$\mu_h, \phi_h, \sigma^2_h$	Parameters of the log volatility process h_t	3
$\boldsymbol{r} = \{r_t\}_{t=1}^{T}$	Auxiliary variable for sampling h_t	$T \times 1$

3.1 Priors

The prior choices are tailored to the VECM structure, while some other priors not explicitly discussed here follow standard specifications and are detailed in the Appendix.

Given that an specific structure has been imposed on β in the decomposition $\Pi = \boldsymbol{\alpha}\beta'$, I only need to specify a prior for $\boldsymbol{\alpha}$ conditional on the Markov states. I assume two states: a *slow-adjustment state* ($s_t = 0$) and a *fast-adjustment state* ($s_t = 1$) for slow and fast error correction. The slow-adjustment state imposes strong shrinkage on $\boldsymbol{\alpha}$ to mitigate potential overfitting without cointegration. The fast-adjustment state applies a weakly informative prior with a variance relatively larger than that in the shrinkage state, helping to avoid label switching and identification issues.

The prior for $\boldsymbol{\alpha}$ at slow-adjustment state as $s_t = 0$ inline with proceeding literatures, for example Geweke (1996); Strachan and Inder (2004); Villani (2005); Koop et al. (2009) etc, where it has form

$$\boldsymbol{\alpha} \mid s_t = 0, \beta, \tau, \nu, \Sigma_\omega \sim N(0, \nu(\beta' P_{1/\tau}\beta)^{-1}\boldsymbol{\Sigma}_\omega), \tag{3.2}$$

which is an informative prior specifically designed for the cointegrating space, where

the prior variance of $\boldsymbol{\alpha}$ depends on the identified cointegration relationships. In particular, τ and ν are hyperparameters restricted to the interval $(0, 1)$ that govern the degree of shrinkage. The matrix $P_{1/\tau}$ was originally introduced to impose shrinkage on $\boldsymbol{\alpha}$ toward the unknown $\boldsymbol{\beta}$, thereby linking the uncertainty in the adjustment coefficients to the strength of the unknown cointegration relationships. This formulation is particularly useful when the cointegration space is unknown, since it additionally penalizes large deviations in $\boldsymbol{\alpha}$ more heavily when the long run equilibrium is weakly identified. In the present setting, where the cointegrating vector is pre-specified, the calibration of $\boldsymbol{\alpha}$ can be more targeted, as it only needs to determine the strength of adjustment toward the imposed cointegrating relation rather than identify the relation itself.

Specifically, $P_{1/\tau} = (HH' + \tau H_\perp H_\perp')^{-1}$, where $H = H^g \left(H^{g\prime} H^g\right)^{-1/2}$ spans the same space as H^g. In this formulation, H^g corresponds to the cointegration vector $\boldsymbol{\beta}$, which in the current setting is fixed as a vector of negative ones. Although fixing $\boldsymbol{\beta}$ eliminates the usual identification issue in the decomposition $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$, the transformation of H^g into the semi-orthogonal matrix H ensures that HH' serves as a proper projection matrix. This construction preserves the desired shrinkage structure in $P_{1/\tau}$ and enhances numerical stability in the posterior simulation. The matrix $H_\perp$ denotes the orthogonal complement of the space spanned by H. Further technical details can be found in the methodological literature.

For the alternative fast-adjustment state $s_t = 1$, the prior variance of $\boldsymbol{\alpha}$ is set to the identity matrix to support larger coefficient values and enable faster convergence toward the long-run equilibrium after a temporary deviation. This is particularly useful when the weighting structure undergoes regime changes, as in Jin et al. (2022), as model weights would otherwise tend to recover slowly.

These two specifications can be interpreted in terms of adjustment half-lives. To build intuition, the half-life, defined as the time required for the deviation from the long run equilibrium to be reduced by half, is

$$h = \frac{\log(0.5)}{\log|1 - \alpha|}.$$

For a small adjustment coefficient, such as $\alpha = 0.1$, the half-life is approximately $h \approx \log(0.5)/\log(1-0.1) \approx 6.6$, indicating slow adjustment and persistent deviations from equilibrium; it takes more than six periods for the deviation to be halved. In contrast, a stronger adjustment, for instance $\alpha = 0.8$, yields $h \approx \log(0.5)/\log(1-0.8) \approx 0.3$, implying a much shorter half-life and a faster return to equilibrium, as the deviation is reduced by half within less than one period. Hence, a heavily shrunk α corresponds to sluggish adjustment, whereas a larger α reflects rapid correction toward the long run equilibrium.

3.2 Estimation

Within my MCMC algorithm, I follow the approach of Koop et al. (2009) for sampling components related to the VECM specification, while employing standard methods for Markov switching and stochastic volatility inference. The full sequence of steps is summarized in Algorithm 1, with complete algorithmic details and prior specifications provided in the Appendix.

For all posterior inference, I generate $G = 1{,}000$ retained draws after a burn-in period of 20,000 iterations, applying a thinning interval of 10 to reduce potential autocorrelation from MCMC. Simulation-based inference is then conducted using these posterior samples. For example, the posterior expectation of $\boldsymbol{\omega}_t$, denoted $E(\boldsymbol{\omega}_t \mid \boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T})$, is estimated as the posterior sample mean:

$$\widehat{E}(\boldsymbol{\omega}_t \mid \boldsymbol{y}_{1:T}, \boldsymbol{x}_{1:T}) = \frac{1}{G} \sum_{g=1}^{G} \boldsymbol{\omega}_t^{(g)}.$$

3.3 Predictive Scheme with Model Assembly

To design a two-stage sequential procedure for model assembly, the data are partitioned at two time points, T_0 and T_1, with $T_0 < T_1 < T$. The initial training sample, covering observations up to T_0, is used to estimate the base models and generate individual fore-

Algorithm 1 MCMC sampling scheme. After initialization, iterate the following steps:

(a) Sample $\boldsymbol{\omega} \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{h}, \boldsymbol{\omega}_0, \boldsymbol{s}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_\omega$.

(b) Sample $\boldsymbol{\Sigma}_\omega \mid \boldsymbol{\omega}, \boldsymbol{s}, \boldsymbol{\alpha}$.

(c) Sample $\boldsymbol{\omega}_0 \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\omega}, \boldsymbol{h}, \sigma^2_{\omega_0}$.

(d) Sample $\sigma^2_{\omega_0} \mid \boldsymbol{\omega}_0$.

(e) Sample $\boldsymbol{\alpha} \mid \boldsymbol{\omega}, \boldsymbol{s}, \boldsymbol{\Sigma}_\omega$.

(f) Sample $\boldsymbol{s} \mid \boldsymbol{\omega}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_\omega, P$ using the FFBS algorithm.

(g) Sample $P \mid \boldsymbol{s}$.

(h) Sample $\boldsymbol{h} \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\omega}, \boldsymbol{\omega}_0, \boldsymbol{r}, \mu_h, \phi_h, \sigma^2_h$.

(i) Sample $\mu_h, \phi_h \mid \boldsymbol{h}, \sigma^2_h$, then sample $\sigma^2_h \mid \boldsymbol{h}, \mu_h, \phi_h$.

(j) Sample $\boldsymbol{r} \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\omega}, \boldsymbol{\omega}_0, \boldsymbol{h}$.

casts. The period between T_0 and T_1 is then used to train the synthesis function and thus determine the structure of the model weights $\boldsymbol{\omega}_t$. Therefore, the rest after T_1 will be used to evaluate their out-of-sample performance. Since the weights, $\boldsymbol{w}_t$, cannot be inferred without input from the base models, this two-stage approach reduces computational cost at the expense of either a smaller testing sample or a smaller synthesis sample.

The choice of T_0 and T_1 necessarily involves researcher discretion, as one must balance the allocation of observations between the initial estimation stage and the subsequent synthesis stage. The detailed choices will be reported in the application. Conceptually, partitioning the sample into three subsets serves to separate the estimation of base models, the calibration of the combination scheme, and the evaluation of predictive performance. In contrast, the present framework updates the information set sequentially after each forecast, ensuring that newly available observations are incorporated in an efficient and timely manner.

3.3.1 Direct h-Step-Ahead Forecasting

In the macroeconomic forecasting applications, the primary objective is to construct predictive distributions for future macroeconomic outcome variables. Without explicitly modeling the dynamics of the covariates, it is not feasible to generate predictions for future time points iteratively. To address this, I propose a direct $h-$step ahead forecasting model by specifying the base model as

$$x_{t+h} = \boldsymbol{z}_t'\boldsymbol{\gamma} + \nu_{t+h}, \tag{3.3}$$

where the error term $\nu_{t+h} \sim N\left(0, \exp(g_t)\right)$ follows a stochastic volatility process. Estimation of the coefficient vector $\boldsymbol{\gamma}$ is conducted within a Bayesian linear regression framework, allowing for a coherent quantification of uncertainty.

The covariate vector $\boldsymbol{z}_t$ includes economic predictors relevant to each base model. Alternatively, each expert model may adopt its own set of predictors; for instance, one may use the unemployment rate to forecast inflation, another may rely on the interest rate, while others may use both. Each base model then produces a forecast $\hat{x}_{t+h}$, which serves as input to the second-stage synthesis model.

Motivated by extensive empirical evidence showing that models with time-varying parameters often outperform their constant-parameter counterparts (see, for example, Pesaran and Timmermann (2007); Giacomini and Rossi (2009)), I employ a rolling window of 100 observations to estimate $\boldsymbol{\gamma}$ in equation (3.3) using ordinary least squares. This approach is straightforward and empirically robust. In addition, I also consider more sophisticated statistical models that allow for time-varying coefficients to generate more reliable first-stage forecasts.

3.3.2 Combine Base Model Forecasts

The out-of-sample h-step-ahead predictions can be synthesized into the following model, which slightly modifies the initial measurement equation in terms of notation

$$y_{t+h} = \omega_{0,t} + \hat{\boldsymbol{x}}_{t+h}'\boldsymbol{\omega}_t + \epsilon_{t+h}, \tag{3.4}$$

where the superscript (h) denotes the h-period ahead forecasting, and all h-period forecasting are processed in parallel.

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

$$p(y_{T+h}|I_T) = \int p(y_{T+h} \mid \hat{\boldsymbol{x}}_{T+h}, \boldsymbol{\omega}_T, h_T) p(\Psi \mid \boldsymbol{y}_{1+h:T}, \hat{\boldsymbol{x}}_{1+h:T}) d\Psi, \tag{3.5}$$

where $I_T = (\boldsymbol{y}_{1:T}, \hat{\boldsymbol{x}}_{h+1:T}, \hat{\boldsymbol{x}}_{T+h})$ denotes the information available up to time T. By construction, the predictive distribution for all h, can be constructed using the covariate observed within the sample period. In addition, the implied time-varying parameter model structure, captured by $p(y_{T+h} \mid \hat{\boldsymbol{x}}_{T+h}, \boldsymbol{\omega}_T, h_T)$, reflects the specific dynamic relationship between the target variable and the $h-$ahead foreacasts 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 (3.4), the posterior distribution $p(\Psi \mid \boldsymbol{y}_{1+h:T}, \hat{\boldsymbol{x}}_{1+h:T})$ in (3.5) only involves inference for the latent states up to time T. In particular, I obtain G posterior draws of $(\boldsymbol{\alpha}_{s_T}, s_T, \boldsymbol{\omega}_T, \omega_{0,T}, \boldsymbol{h}_T)$. To generate the h-step-ahead forecast, I simulate forward from these draws using the dynamic specification in (2.7)–(2.8) to obtain G draws of $(\boldsymbol{\omega}_{T+h}, h_{T+h})$, which serve as inputs to the predictive density $p(y_{T+h} \mid \hat{\boldsymbol{x}}_{T+h}, \boldsymbol{\omega}_{T+h}, h_{T+h})$. This yields G simulated values of y_{T+h}, denoted $\{y_{T+h}^{(g)}\}_{g=1}^{G}$, from its conditional predictive distribution.

From these simulations, all predictive statistics can be computed. For example, the posterior mean forecast $E(y_{T+h} \mid I_T)$ is estimated by

$$\frac{1}{G} \sum_{g=1}^{G} y_{T+h}^{(g)}.$$

Similarly, the predictive density $p(y_{T+h} \mid I_T)$ can be approximated by

$$\frac{1}{G} \sum_{g=1}^{G} p\left(y_{T+h} \mid \hat{\boldsymbol{x}}_{T+h}, \boldsymbol{\omega}_{T+h}^{(g)}, h_{T+h}^{(g)}\right),$$

evaluated over a grid of plausible y_{T+h} values. Forecast performance is then assessed

using scoring rules designed for predictive distributions, as described in the following Section 3.4.

3.4 Forecasting Metrics

I consider **three** popular 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). All metrics are out-of-sample and calculated based on the Monte-Carlo estimate of (3.5), as described in Section 3.3.1. In what follows, I defined T_0 as the initial sample period and T_1 as the sample period for model assembly, and T denotes the overall sample size, with the number of h-period ahead periods evaluated being $T - h - T_1 + 1$.

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 $\mathcal{M}_B$:

$$\textbf{LPDR}(\mathcal{M}, \mathcal{M}_B) = \log(p(Y_{T_1+h:T} \mid I_{T_1}, \mathcal{M})) - \log(p(Y_{T_1+h:T} \mid I_{T_1}, \mathcal{M}_B)), \tag{3.6}$$

where

$$\log(p(Y_{T_1+h:T} \mid I_{T_1}, \mathcal{M})) = \sum_{t=T_1}^{T-h} \log(p(y_{t+h} \mid I_t, \mathcal{M}))$$

is the cumulative log score, with $\log(p(y_{t+h} \mid I_t, \mathcal{M}))$ 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)). The simplified dynamic latent factor model of McAlinn and West (2019) is adopted as the comparative benchmark ($\mathcal{M}_B$) for all subsequent analyses. The SSM serves as my $\mathcal{M}_B$ 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 behav-

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

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

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

$$\mathbf{CRPS}_{T_1+h:T}(\mathcal{M}) = \frac{1}{T - T_1 - h + 1} \sum_{t=T_1}^{T-h} \mathbf{CRPS}\left(p(y_{t+h} \mid I_t, \mathcal{M}), y_{t+h}\right). \tag{3.7}$$

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:

$$\mathbf{RMSFE}_{T_1+h:T} = \sqrt{\frac{1}{T - T_1 - h + 1} \sum_{t=T_1}^{T-h} \left(y_{t+h} - E(y_{t+h} \mid I_t, \mathcal{M})\right)^2}, \tag{3.8}$$

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

4 Monte Carlo Study

I present four sets of Monte Carlo experiments with different data-generating processes, each consisting of 100 simulation replications, to evaluate the effectiveness of the VECM framework in out-of-sample forecasting using the performance metrics described in Section 3.4. The experimental designs differ primarily in the data-generating processes (DGPs) governing the predictors ($\boldsymbol{x}$)[3] and the latent weights ($\boldsymbol{\omega}_t$). This comprehensive design demonstrates the flexibility of the VECM-MS specification in capturing complex time-varying dynamics and accommodating regime shifts in the weighting structure, thereby

[3]It is important to emphasize that in this synthesis framework, the 'predictors' do not represent raw economic variables, but rather the vector of forecasts produced by the base models. Formally, these correspond to $\hat{\boldsymbol{x}}_{t+h}$ in Equation (3.4).

validating its predictive gains.

4.1 Data Generating Process

Data are generated for four distinct experimental scenarios (two weight specification scenarios $\times$ two predictor structure types), each comprising 100 independent Monte Carlo replications. The underlying data-generating process (DGP) incorporates both continuous stochastic fluctuations and discrete structural shifts in model weights. The simulation procedure is structured as follows:

- **Sample Size and Model Pool:** I set the total sample size to $T = 300$. The set of predictors consists of $p = 3$ base forecasts representing the base models to be combined, augmented by a vector of ones for the intercept.

- **Weight Generation Process:** I evaluate the model under two distinct specifications for the weight evolution $\boldsymbol{w}_t = (w_{1,t}, w_{2,t}, w_{3,t})'$.

  Scenario 1 (Piecewise Static): In this baseline specification, the weights are drawn from a Dirichlet mixture with fixed parameters that shift abruptly across regimes. Specifically, the data generating process is defined as:

$$\boldsymbol{w}_t \sim \begin{cases} \text{Dirichlet}(9,1,0), & t = 1,\ldots,100, \\ \text{Dirichlet}(0,9,1), & t = 101,\ldots,200, \\ \text{Dirichlet}(1,0,9), & t = 201,\ldots,300. \end{cases}$$

  This specification enforces the unit-sum constraint at every time step while introducing sharp structural breaks in the weight composition. The intercept term (bias correction) is fixed at unity for all t.

  Scenario 2 (Dynamic Evolution): The second specification introduces persistence, allowing for smooth evolution within regimes and rapid adjustment during transitions. I define three regime-specific attractors (means), denoted by $\boldsymbol{m}_s$, with transi-

tions occurring at $t = 100$ and $t = 200$:

$$\boldsymbol{m}_1 = (0.9, 0.1, 0.0)', \quad \boldsymbol{m}_2 = (0.0, 0.9, 0.1)', \quad \boldsymbol{m}_3 = (0.1, 0.0, 0.9)'.$$

The dynamic mean of the process, $\boldsymbol{\mu}_t$, follows a partial adjustment mechanism:

$$\boldsymbol{\mu}_t = \begin{cases} (1-\rho)\boldsymbol{w}_{t-1} + \rho\, \boldsymbol{m}_s, & \text{within regime (smooth evolution),} \\ (1-\phi)\boldsymbol{w}_{t-1} + \phi\, \boldsymbol{m}_s, & \text{at regime transition (rapid shift),} \end{cases}$$

Here, $\rho = 0.1$ governs within-regime persistence, while $\phi = 0.5$ controls the adjustment speed at structural breaks. The weights follow $\boldsymbol{w}_t \sim \text{Dirichlet}(\kappa\, \boldsymbol{\mu}_t)$, where $\kappa = 80$ determines the dispersion around the dynamic mean.

Figure 2 presents illustrative trajectories of the simulated model weights under the two distinct specifications. **Scenario 1** corresponds to the baseline specification, in which the weights remain constant within each time block and exhibit abrupt shifts at $t = 100$ and $t = 200$. This design directly replicates the piecewise-static data-generating process described in Jin et al. (2022). In contrast, **Scenario 2** allows the weights to fluctuate stochastically within each block while retaining sharp transitions at the same structural breakpoints. This introduces within-regime variability, effectively creating a hybrid DGP that embeds the regime-switching structure of Jin et al. (2022) within the flexible synthesis framework of McAlinn and West (2019). Crucially, this dynamic mirrors the specific behavior that the proposed VECM-MS specification is designed to capture.

- **Predictor Process:** The vector of base forecasts, $\boldsymbol{x}_t \in \mathbb{R}^p$, is generated according to two different types of design to capture varying degrees of temporal and cross-sectional dependence.

  Type A: Moderate Correlation (VAR Process). In this specification, the base forecasts are derived from a set of economic indicators, $\boldsymbol{z}_t$, which evolve according to a

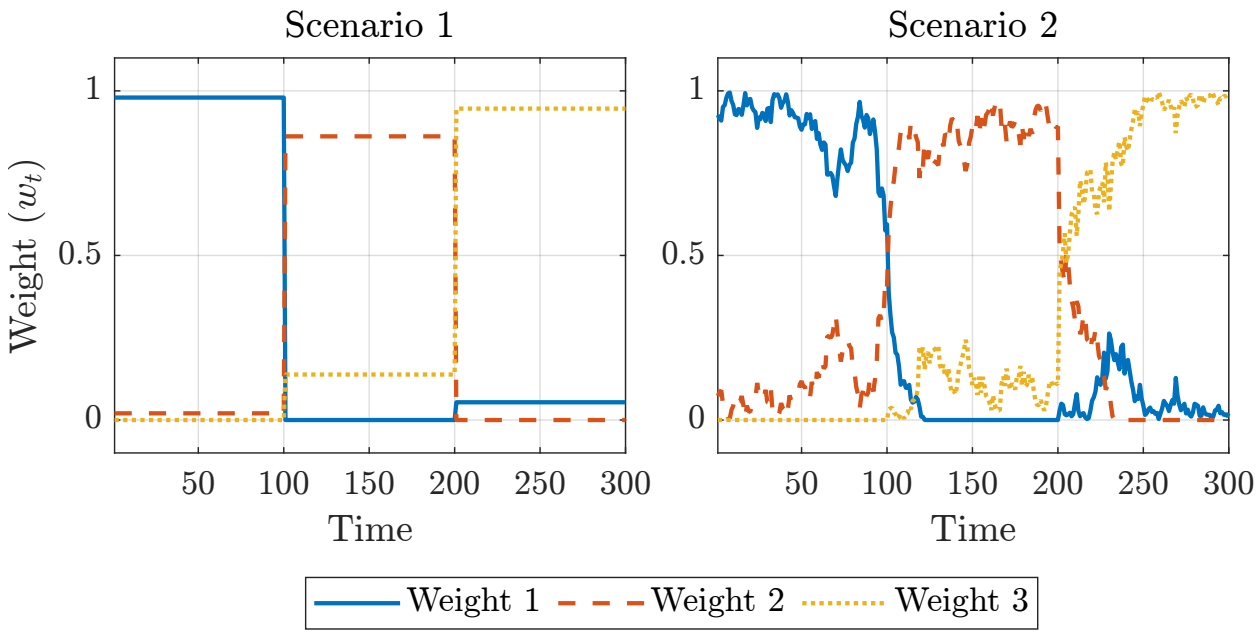


Figure 2: Simulated weight trajectories under different specifications. **Scenario 1** exhibits stepwise regime shifts with constant weights within each block. **Scenario 2** introduces within-regime fluctuations while preserving the regime shifts.

stationary VAR(1) process:

$$\begin{aligned} \boldsymbol{z}_1 &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_z), \\ \boldsymbol{\eta}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_z), \\ \boldsymbol{z}_t &= \boldsymbol{A}_z \boldsymbol{z}_{t-1} + \boldsymbol{\eta}_t. \end{aligned}$$

The vector of base forecasts, $\boldsymbol{x}_t$, is then constructed as a linear transformation of these latent indicators:

$$\boldsymbol{x}_t = \boldsymbol{z}_t' \boldsymbol{\gamma},$$

where $\boldsymbol{\gamma}$ represents the mapping from economic fundamentals to model forecasts, and I assume they are simply drawn from a normal distribution. The autoregressive

coefficients $\boldsymbol{A}_z$ and error covariance $\boldsymbol{\Sigma}_z$ are specified as:

$$\boldsymbol{A}_z = \begin{bmatrix} 0.5 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.2 \\ 0.1 & 0.3 & 0.5 \end{bmatrix}, \quad \boldsymbol{\Sigma}_z = \begin{bmatrix} 0.6 & 0.1 & 0.2 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.2 & 0.3 \end{bmatrix}.$$

This VAR structure induces persistent temporal dependence (autocorrelation ≈ 0.7) and cross-sectional correlation among the predictors, mimicking realistic economic forecasting series.

Type B: High Correlation (Latent Factor Process). The second specification considers a scenario where base model forecasts are highly correlated when models employ similar information sets or methods. Here, the predictors are driven by a single common latent factor f_t:

$$\begin{aligned} \boldsymbol{x}_t &= f_t \mathbf{1}_p + \boldsymbol{\eta}_t, \\ f_t &= \phi f_{t-1} + \varepsilon_t, \qquad\qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2), \end{aligned}$$

where $\mathbf{1}_p$ is a vector of ones. The idiosyncratic components $\boldsymbol{\eta}_t$ follow a multivariate normal distribution with an equicorrelated covariance structure:

$$\boldsymbol{\eta}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_f^2 \left[(1-\rho)\boldsymbol{I}_p + \rho \mathbf{1}_p \mathbf{1}_p'\right]\right).$$

We set the parameters to $\sigma_\epsilon = 1$, $\sigma_f = 0.5$, $\phi = 0.8$, and $\rho = 0.5$. This design generates strong co-movement among the predictors, resulting in pairwise correlations of approximately 0.95.

- **Outcome Process:** The target variable y_t is generated as a linear combination of the base forecasts, weighted by the time-varying latent weights:

$$y_t = w_0 + \sum_{j=1}^{p} x_{j,t} w_{j,t} + e_t, \quad e_t \sim \mathcal{N}(0, \sigma_e^2),$$

where the intercept is fixed at $w_0 = 1$. The variance of the error term, σ_e^2, is calibrated based on the variance of the latent signal to achieve a Signal-to-Noise Ratio (SNR) of 10.

Overall, the data-generating processes are constructed to reflect a wide range of patterns in dynamic weight evolution, guided by my general forecasting intuition. By assessing performance under these deliberately agnostic weight dynamics, which span smooth persistence as well as abrupt changes, the experiment evaluates the ability of the framework to adapt to shifts in model relevance across different environments. Table 3 summarises the four different DGP specifications considered across the 4×100 Monte Carlo replications used to assess predictive performance.

4.2 Statistical Inference

As an illustrative example, I use one of the simulated datasets from the first set (DGP 1) of Monte Carlo replications to demonstrate how the model operates in terms of its statistical inference, with particular emphasis on inference for the VECM-MS specification. The simulation framework follows the same estimation procedure described earlier and applied later in the empirical analysis. The algorithm includes 20,000 total iterations, discard the first 16,000 as burn-in, and retain every 5th draw from the remaining samples for posterior inference.

For comparison, estimation results for the benchmark model (SSM) are also included. Figure 3 illustrates that both the VECM-MS and SSM models successfully track the major temporal variations in the true $\boldsymbol{\omega}_t$. While the differences between the two are subtle, the VECM-MS specification tends to respond slightly more sharply to weight shifts, while yielding relatively smoother trajectories during stable periods. Overall, both methods effectively capture the general evolution of $\boldsymbol{\omega}_t$.

The difference in performance is more apparent in Figure 4. The sum of model weights remains close to unity on average for both models, consistent with the data-generating process. Nevertheless, temporary deviations occur during volatile periods, with the SSM exhibiting larger swings and wider uncertainty bands compared to the VECM-MS, par-

Table 3: Summary of Monte Carlo Data Generating Processes (DGPs)

Panel A: Moderate Correlated Base Forecasts		
Components	**DGP 1 (rng 1 - 100)**	**DGP 2 (rng 101 - 200)**
Predictor Structure	Type A (VAR)	
Predictor (Base forecasts)	$\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_z), \quad \boldsymbol{z}_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_z)$ $\boldsymbol{z}_t = \boldsymbol{A}_z \boldsymbol{z}_{t-1} + \boldsymbol{\eta}_t$ $\boldsymbol{x}_t = \boldsymbol{z}_t' \gamma$	
Parameters	$\boldsymbol{A}_z = \begin{bmatrix} 0.5 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.2 \\ 0.1 & 0.3 & 0.5 \end{bmatrix}, \quad \boldsymbol{\Sigma}_z = \begin{bmatrix} 0.6 & 0.1 & 0.2 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.2 & 0.3 \end{bmatrix}$ $\gamma \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_3)$	
Weight Specification (ω_t)	**Scenario 1** (Piecewise Static)	**Scenario 2** (Dynamic Evolution)

Panel B: High Correlated Base Forecasts		
Components	**DGP 3 (rng 201 - 300)**	**DGP 4 (rng 301 - 400)**
Predictor Structure	Type B (Common Factor)	
Predictor (Base forecasts)	$\boldsymbol{x}_t = f_t \mathbf{1}_p + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_\eta\right)$ $f_t = \phi f_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$	
Parameters	$\boldsymbol{\Sigma}_\eta = \sigma_f^2\big((1-\rho)\boldsymbol{I}_p + \rho\, \mathbf{1}_p \mathbf{1}_p'\big)$ $\phi = 0.8, \quad \rho = 0.5, \quad \sigma_f = 0.5, \quad \sigma_\epsilon = 1$	
Weight Specification (ω_t)	**Scenario 1** (Piecewise Static)	**Scenario 2** (Dynamic Evolution)

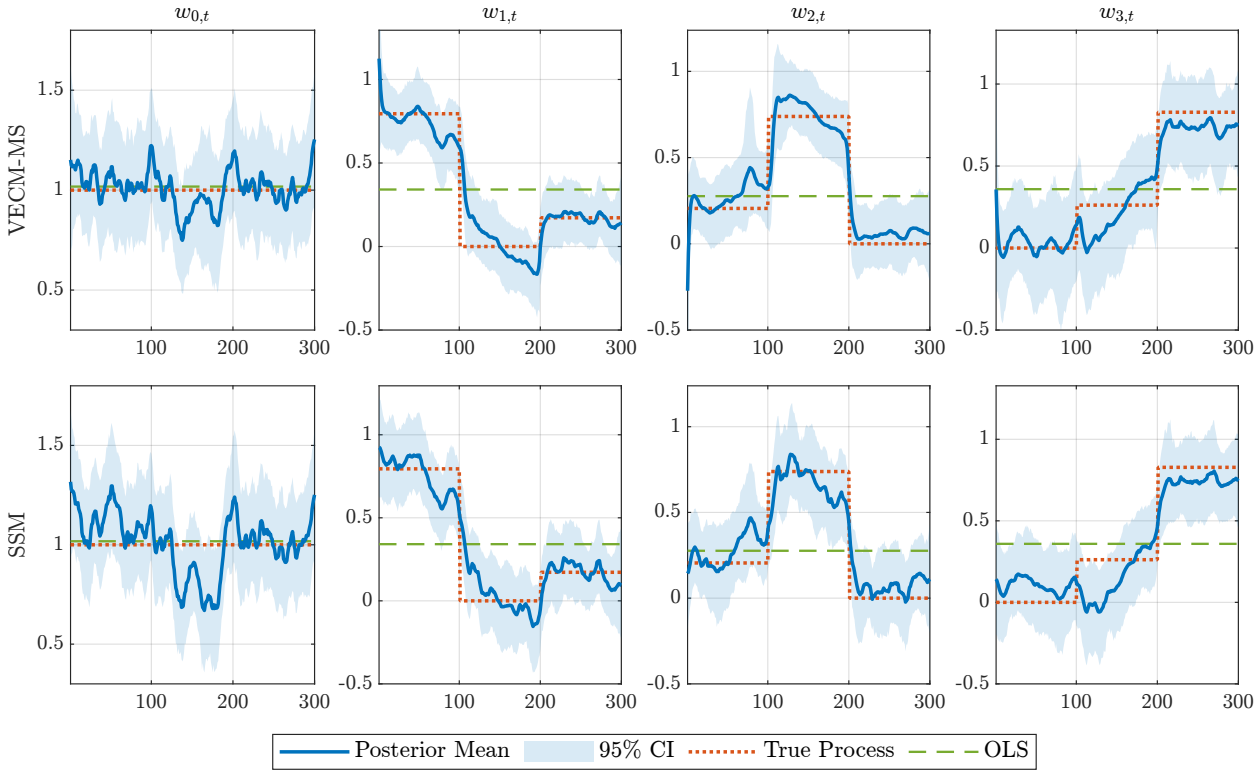


Figure 3: Posterior estimates from VECM-MS (top row) and SSM (bottom row) models under the DGP1-rng(64) design. Each panel reports the posterior mean (solid blue), 95% credible interval (shaded area), true parameter values (red dotted), and corresponding OLS estimates (green dashed) for $\omega_{0,t}$ and the elements of $\boldsymbol{\omega}_t$.

ticularly in the first 200 time periods where it deviates substantially from the true value. In contrast, the VECM-MS stays much closer to unity throughout.

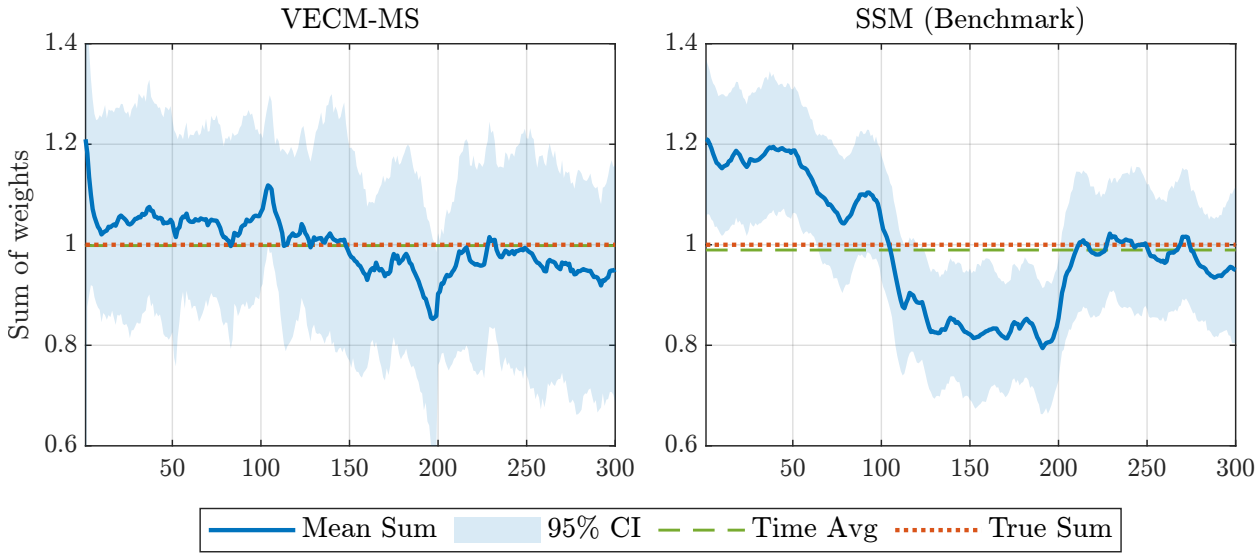


Figure 4: Posterior estimates of the cross-sectional sum of $\boldsymbol{\omega}_t$ (excluding ω_0) for the VECM-MS (left) and SSM (right) models under the DGP-rng(64) design. The solid blue line denotes the posterior mean, with the shaded area representing the 95% credible interval. The red dotted line marks the true sum of weights, and the green dashed line shows the time-average of the posterior mean.

The substantial difference in the sum of weights can be explained by observing Figure 5, which shows that the VECM-MS accurately assigns high posterior probabilities to periods of large deviations in the error correction coefficients, thereby effectively identifying the timing of low- and fast-adjustment states. This allows the aggregate weights to respond rapidly and adjust toward the long-run equilibrium, whereas the SSM lacks an explicit correction mechanism.

For example, at the first 100 time periods in Figure 4, without VECM-MS, the sum of weights consistently deviates from the true value. In contrast, when large deviation parameters are included at $t = 1$, the aggregate weights are immediately pushed back toward the long-run equilibrium. A similar phenomenon can be observed around $t \approx 100$ and $t \approx 200$. During structural changes, the estimated quantity begins to diverge, but the VECM promptly intervenes and, through the recovered state $s_t = 1$, introduces a large $\boldsymbol{\alpha}$ that immediately corrects the values back to the long-run equilibrium. On the contrary, the SSM lacks this adjustment mechanism.

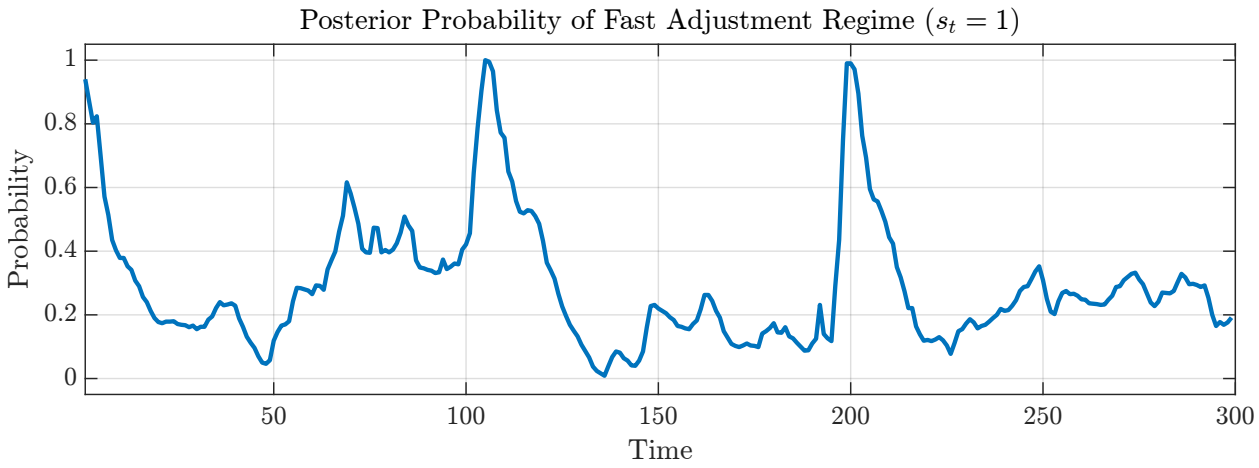


Figure 5: Posterior probability of being in the recovery state ($s_t = 1$) from the VECM-MS model under the DGP-rng(64) design.

This posterior inference for both s_t and $\sum_j \omega_{j,t}$ highlights the value of incorporating a Markov-switching mechanism into the error-correction coefficients, particularly in this latent structure, where the goal is for the VECM to play an active role while ensuring a smooth and reliable estimation of the state variable $\boldsymbol{\omega}_t$. In simple terms, the correction model is not always active in correcting the deviation; it intervenes only when the system deviates from its equilibrium and a correction becomes necessary.

4.3 Forecasting Performance

The statistical inference supports the validity of the extended VECM models, but the more interesting aspect lies in their out-of-sample performance. Each model is evaluated in parallel over the out-of-sample (OOS) period from $t = 51$ to $t = 300$, with the simulation repeated 400 times as part of a Monte Carlo study. For brevity, results based on DGPs 1, 2, 3 and 4 are reported in the main text, while some extra cases are placed to the Appendix, as all DGPs exhibit broadly similar patterns.

Performance is assessed using one-step-ahead forecasts; specifically, $\boldsymbol{x}_{t+1}$ is assumed to be known when predicting y_{t+1}, and the predictions are compared with the observed values. This coincides with the one-step-ahead forecast described in equation 3.4. The corresponding log predictive density ratio (LPDR), relative RMSFE improvements (%), and relative CRPS improvements (%) over the benchmark are presented in Figures 6, 7,

and 8.

I also evaluated the method proposed by Jin et al. (2022) using a large finite mixture model to approximate their infinite hidden Markov specification. Stochastic volatility was included to maintain consistency across experiments. Results from both the Monte Carlo simulations and the empirical application indicate that this approach does not yield significant improvement over the benchmark BPS model and can be even worse. This finding suggests that incorporating either abrupt regime changes or random-walk dynamics alone offers limited benefit, whereas combining the two elements generates complementary effects that enhance overall performance.

LPDR Figure 6 presents the cumulative Log Predictive Density Ratios (LPDR) relative to the SSM benchmark across the four Data Generating Processes (rows) and four model specifications (columns). The solid lines represent the mean cumulative LPDR across 100 replications, while the shaded regions denote the 95% credible intervals.

The results for the standard VECM (leftmost column) reveal a heterogeneous performance profile: while it outperforms the benchmark by 3.2 and 2.9 units under moderate-correlation scenarios of DGP 1 and 2, its performance falters to approximately 1.2 units under the high-correlation scenarios of DGP 3 and 4. Notably, the 95% credible intervals among all cases frequently encompass zero, suggesting that the static VECM struggles to distinguish itself statistically from the random walk benchmark in complex identification settings.

The introduction of the Markov-switching error correction mechanism (VECM-MS, second column) largely improves the performance. This specification delivers robust predictive gain, increasing the terminal LPDR by a factor of five compared to the static version, namely from 3.2 to 16.9 under DGP 1, and from 2.9 to 11.0 under DGP 2. Even under the high-correlation structures of DGP 3 and 4, where the standard model stagnated, the VECM-MS achieves substantial gains of 6.9 and 6.5, respectively. In addition, the credible intervals for the VECM-MS specification (shaded blue) are tighter and almost exclusively exclude zero across all DGPs. This confirms that the ability to switch between 'slow' and 'fast-adjustment' states allows the model to filter out noise while reacting aggressively to

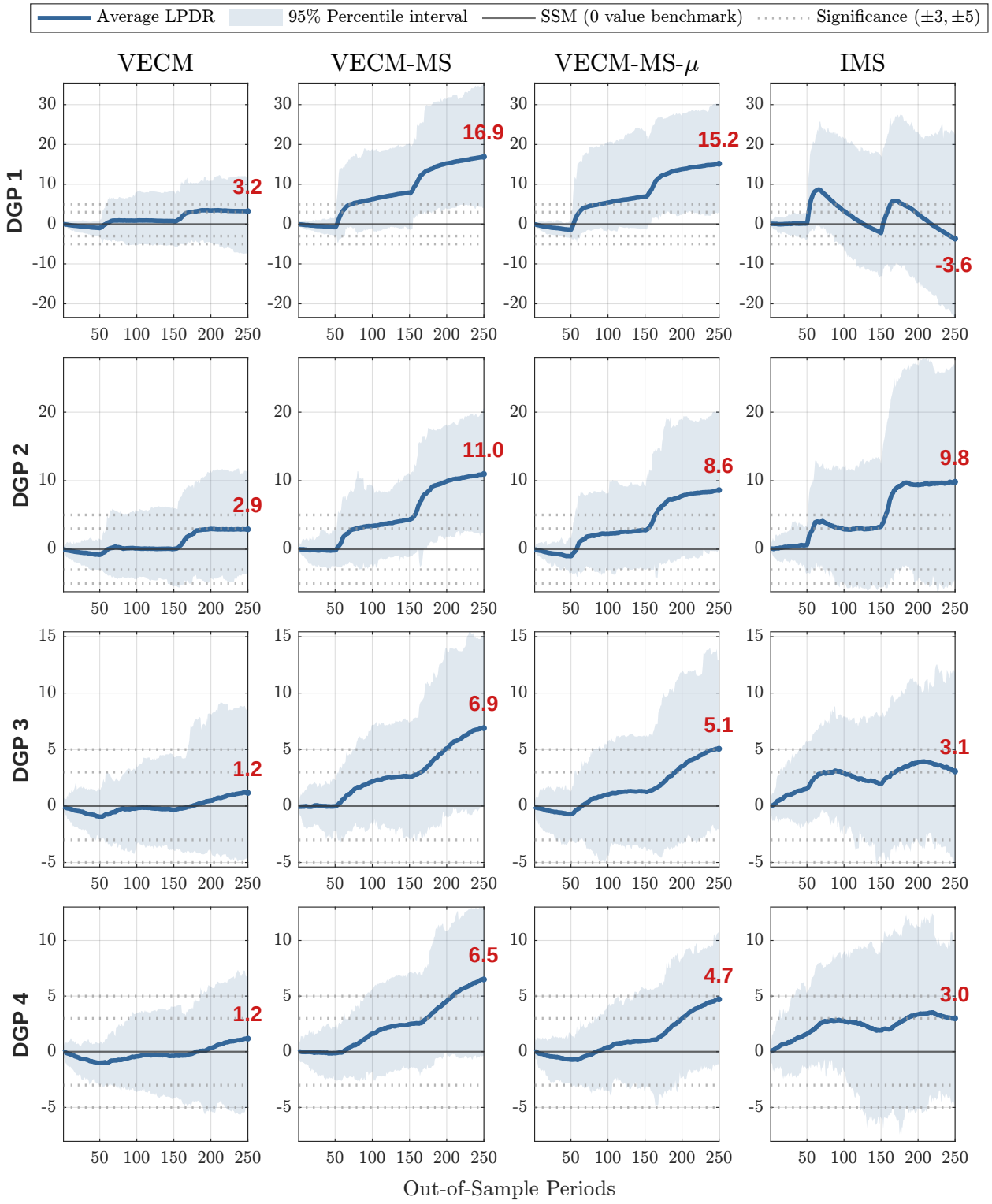


Figure 6: Monte Carlo evaluation of LPDR relative to the benchmark SSM. Each row corresponds to a different simulation group (DGP 1,2,3,4), and column panels within each row show results for different model specifications.

structural changes.

The trajectory of the LPDR gains in the VECM-MS column reveals strong accelerations around periods after $h = 50$ and $h = 150$. These intervals correspond to the structural changes in the data-generating process. The steep positive slope during these windows confirms that the 'fast-adjustment' regime is successfully detecting the break and rapidly re-anchoring the weights, whereas the standard VECM adapts slowly.

The VECM-MS-μ (third column) exhibits a similar qualitative pattern to the VECM-MS but with slightly lower efficiency (e.g., 15.2 vs 16.9 in DGP 1), suggesting that while estimating the intercept provides flexibility, the strict unit-sum anchor of the VECM-MS offers better regularization under this specific DGP.

It is also worth noting the evolution of the LPDR over time. Apart from the two sharp jumps, the performance gains during more static periods are relatively modest, though still positive. This pattern is consistent with the intuition that when forecasts are already close to the long-run equilibrium, the incremental benefit of the VECM specification is limited. The most striking improvements occur during episodes of structural change, when model weights experience substantial shifts. In these situations, models without cointegration adjustment tend to drift away from equilibrium, while the VECM specification activates its error-correction mechanism, pulling the forecasts' weights back toward the correct equilibrium path. This dynamic highlights a key feature where the VECM contributes not only marginal improvements in stable periods but, more notably, delivers substantial gains when the forecasting environment becomes unstable or when structural breaks lead to temporary misspecification.

Finally, the IMS (rightmost column) displays a volatile performance profile. While it achieves positive predictive gains under DGP 2, 3, and 4 (9.8, 3.1, and 3.0, respectively), it yields a negative cumulative LPDR of -3.6 under DGP 1, failing to outperform the simple random-walk benchmark. Even in scenarios where the IMS improves upon the benchmark, its gains are consistently smaller than those of the VECM-MS. Moreover, the percentile intervals for the IMS are notably diffuse, reflecting substantial estimation uncertainty and inconsistent performance across replications. This highlights a critical methodological insight: regime-switching alone is insufficient for robust forecasting. Rather, it is

the integration of regime-switching dynamics with the cointegrating VECM anchor that is required to secure consistent, high-precision predictive gains.

RMSFE Figure 7 summarizes the relative RMSFE improvements (%) over the SSM benchmark across four Data Generating Processes (rows) and four model specifications (columns). Each histogram displays the distribution of out-of-sample performance gains at the final forecast horizon across 100 independent Monte Carlo replications. The red dashed line marks the zero-improvement threshold over the benchmark. The numerical labels in the top corners quantify performance relative to the benchmark: red values (top-left) indicate the count of simulations where the model underperformed, while green values (top-right) indicate the count where it outperformed (out of 100 replications).

The leftmost column shows that the standard VECM offers mixed performance. While the majority of simulations yield positive gains (e.g., 86% in DGP 1 and 78% in DGP 2), the magnitude of improvement is generally modest, typically clustering between 0% and 3%. Furthermore, a non-negligible fraction of replications (e.g., 39% in DGP 3 and 39% in DGP 4) perform worse than the benchmark, resulting in negative improvement values. This dispersion underscores the sensitivity of the static VECM to structural changes, where a fixed adjustment speed may fail to capture rapid shifts.

VECM with Markov-switching adjustment dynamics (VECM-MS, second column) significantly enhances robustness. This specification delivers consistent improvements, with the vast majority of probability mass concentrated in the positive and higher percentage region. Specifically, the model outperforms the benchmark in 93% of simulations under DGP 1 and 95% under DGP 2. Even in the more challenging high-correlation environments (DGP 3 and 4), the VECM-MS achieves positive gains in 97% and 95% of replications, respectively. This rightward shift in the distribution highlights the model's ability to consistently reduce forecast error by adapting its adjustment speed to the new equilibrium of another weighting combination.

The third column panel presents the VECM specification that also estimates the mean of the cointegrating relation μ. This version also performs better overall, with most simulations yielding positive gains relative to the benchmark. A few simulations, however, ex-

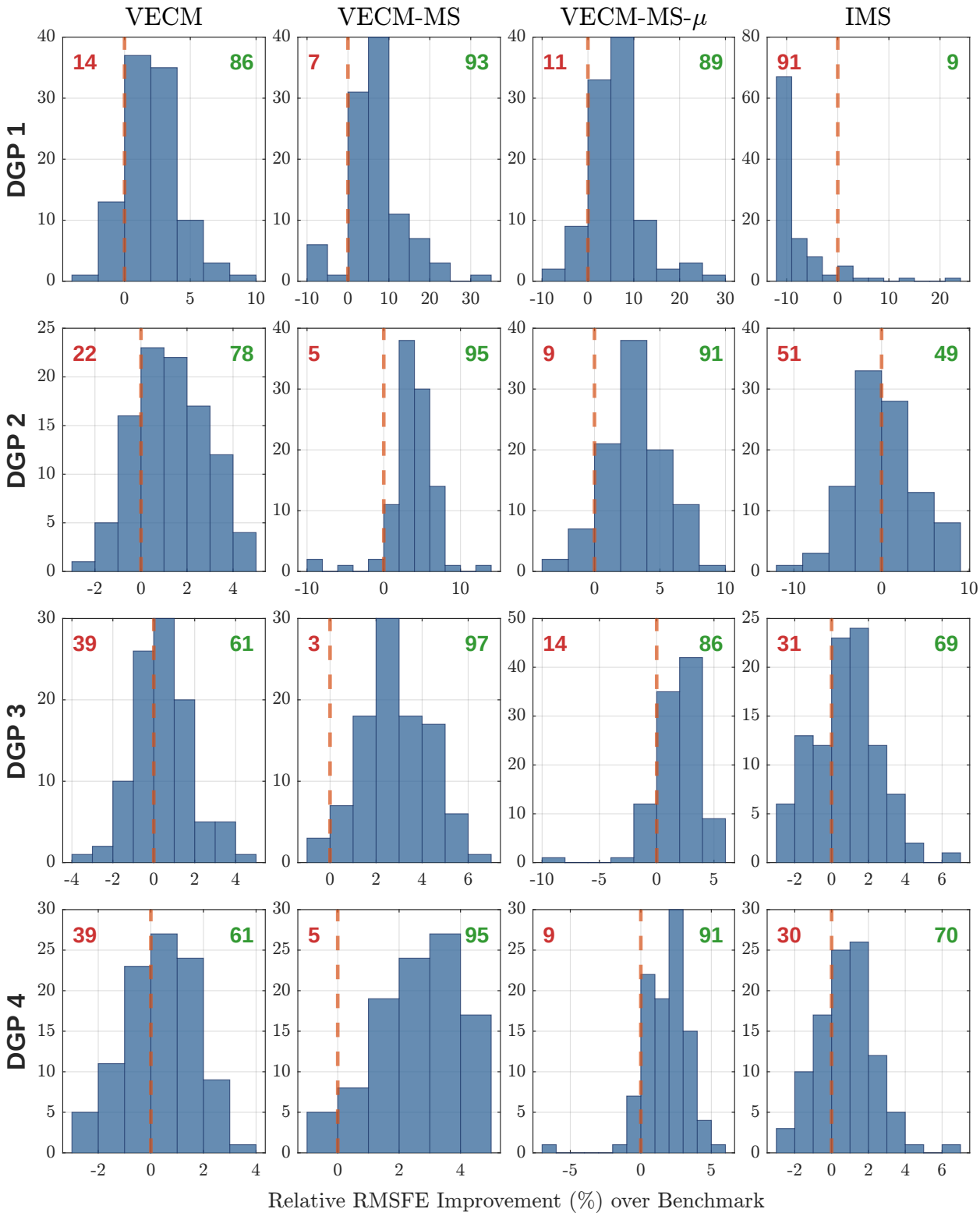


Figure 7: Monte Carlo evaluation of predictive performance relative to the SSM benchmark. The histograms display the distribution of relative RMSFE improvements (%) across 100 replications (each DGP). Panels are organized by Data Generating Process (Rows) and Model Specification (Columns). The numerical labels in the top corners indicate the number of simulations where the model underperformed (Red, left) or outperformed (Green, right) the benchmark.

hibit slightly worse performance, which may reflect the presence of outliers or increased uncertainty associated with estimating the long-run mean.

The third column presents the specification with an estimated cointegrating intercept (μ). While still superior to the standard VECM (e.g., 89% positive share in DGP 1 vs 86%), it exhibits slightly higher variance than the restricted VECM-MS. For instance, in DGP 3, the success rate drops to 86% compared to 97% for the restricted model. This suggests that while estimating μ provides flexibility, the additional parameter uncertainty can occasionally offset the gains from regime switching.

Finally, the rightmost column displays the performance of the IMS. The results reveal significant instability: under DGP 1 and 2, a substantial portion of simulations (91% and 51%, respectively) perform worse than the benchmark, indicated by the heavy left tails. While the IMS shows improvement in DGP 4 (70% positive), its performance is consistently dominated by the VECM-MS specification. This reinforces the finding that regime-switching flexibility alone is insufficient without the stabilizing anchor of cointegration.

CRPS Figure 8 presents the distribution of relative CRPS improvements (%) over the SSM benchmark across the four data-generating processes (DGPs) and four model specifications. The histograms illustrate the density forecasting performance at the final forecast horizon across 100 simulation replications, with values to the right of the red dashed line indicating superior probabilistic accuracy.

The leftmost column shows that the standard VECM offers inconsistent gains in density forecasting. Under the moderate-correlation scenarios (DGP 1 and 2), the model outperforms the benchmark in 75% and 70% of simulations, respectively. However, in the high-correlation environments (DGP 3 and 4), the performance degrades significantly: the win rate drops to 54% and 62%, with distributions heavily centered around zero. This indicates that without regime-switching capabilities, the standard VECM frequently fails to offer better uncertainty quantification than the random walk benchmark.

In contrast, including Markov-switching adjustment coefficients (VECM-MS, second column) yields robust and consistent improvements. This specification achieves win rates of 92% and 95% under DGP 1 and 2, respectively. More importantly, it maintains this

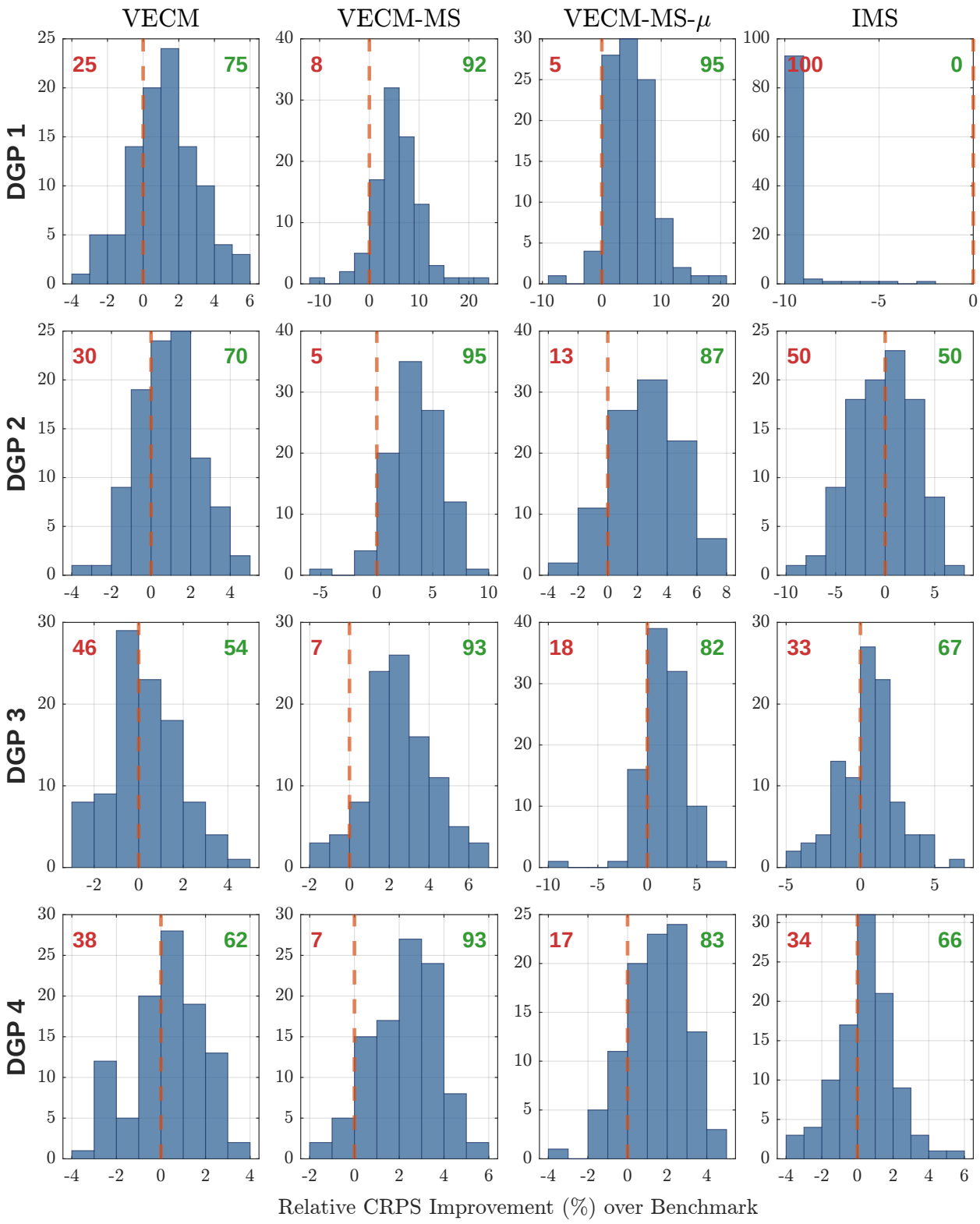


Figure 8: Monte Carlo evaluation of predictive performance relative to the SSM benchmark. The histograms display the distribution of relative CRPS improvements (%) across 100 replications (each DGP). Panels are organized by Data Generating Process (Rows) and Model Specification (Columns). The numerical labels in the top corners indicate the number of simulations where the model underperformed (Red, left) or outperformed (Green, right) the benchmark.

dominance even in the challenging high-correlation settings, maintaining positive gains in 93% of replications for both DGP 3 and 4. The distributions are visibly shifted to the right, confirming that the state-dependent adjustment mechanism not only improves point forecasts but also enhances the calibration of the predictive density.

The VECM-MS-μ (third column) performs comparably to the VECM-MS, achieving improvement rates between 82% and 95%. Finally, the IMS (rightmost column) exhibits substantial weakness. Under DGP 1, all the simulations perform significantly worse than the benchmark, clustering entirely in the lower tail (representing relative losses exceeding 10%). Similarly, under DGP 2, performance is essentially random, with a 50% better rate. This reinforces the conclusion that the unconstrained regime-switching of the IMS leads to poor density calibration compared to the cointegrated VECM framework, particularly in environments where the optimal weights indicate a stable long-run equilibrium.

Overall, the Monte Carlo results reveal a clear pattern in model performance. Regime switching on its own, as in the IMS, is insufficient and can be unstable. The standard VECM and random walk specification is more stable but reacts slowly to changes in the predictive environment. The VECM-MS specification maintains a balance between these extremes. By allowing the adjustment coefficient to switch across latent states, it enables the weights to shift among different long-run vector structures while still permitting local fluctuations, leading to consistently superior performance in both point and density forecasts.

5 Application to Economic Growth

In this section, I consider empirical applications of forecasting economic growth. The base models are specified with yield spread, unemployment, and inflation as predictors, where Harvey (1989) supports the use of yield spread as a predictor and. Across horizons up to 12 steps ahead, the VECM structure consistently improves predictive accuracy relative to the benchmark models, highlighting the value of embedding long-run equilibrium information within the pooling framework.

5.1 Data

Economic growth is measured by the growth rate of U.S. industrial production. To forecast growth, I employ three macroeconomic predictors: the term spread (10-year Treasury yield minus the federal funds rate), the inflation rate measured by the Producer Price Index, and the unemployment rate. All series are obtained from FRED at a monthly frequency, recorded as end-of-month observations. The sample spans from 1962M12 to 2025M7, yielding $T = 752$ months.

The raw series are collected in non-seasonally adjusted form. To account for seasonality, I apply a one-sided moving average filter, relying exclusively on past observations to prevent look-ahead bias, which is constructed by averaging only past observations within a fixed window. For a monthly series $\{\tilde{y}_t\}$ and a window length m, the one sided moving average at time t is $y_t = \frac{1}{m}\sum_{j=0}^{m-1} \tilde{y}_{t-j}$. This only uses information from months $t, t-1, \ldots, t-m+1$ only. In practice, I set $m = 12$ to smooth out the seasonal pattern. The observations from 1961M12 to 1962M11 supply the required initial lags for computing $y_{1962\text{M}12}$ so that no future values are used at any point in the construction.

5.2 Base Models: Linear

As the primary focus of this paper is to develop and improve the forecast synthesis framework, I begin with a simple setting where all base models are evaluated using linear models.

$$\text{Growth}_{t+h} = \gamma_0 + \gamma_{1:3}\,\text{Unem}_{t:t-2} + \nu_{t+h}, \tag{5.1}$$

$$\text{Growth}_{t+h} = \gamma_0 + \gamma_1\,\text{Spread}_t + \nu_{t+h}, \tag{5.2}$$

$$\text{Growth}_{t+h} = \gamma_0 + \gamma_1\,\text{Inflation}_t + \nu_{t+h}, \tag{5.3}$$

where Growth_{t+h} denotes the h-month-ahead growth rate. Model (5.1) uses unemployment over the current and previous three months as predictors. Model (5.2) specifies the term spread as the sole predictor. Model (5.3) relies only on the current inflation rate. In all cases, the error term ν_{t+h} captures unexplained variation and is modeled with stochas-

tic volatility, adopting the same variance specification as in Equation (2.3) to allow for time-varying uncertainty.

Although more advanced approaches to growth forecasting exist, such as factor models, large scale VARs, and Bayesian TVP VARs, and although more sophisticated methods will be employed in the next section for comparison, the focus here is on showing that forecast combination under the proposed VECM structure can deliver meaningful gains even when the underlying inputs are simple benchmarks. These inputs are used primarily to illustrate how the dynamic latent factor BPS formulation adjusts for time-varying biases, reallocates weights according to the imposed cointegration, captures more structured dependencies across models, and improves predictive performance across forecast horizons.

All base models are estimated within a rolling-window framework using window lengths of 100 periods, which allows coefficients to evolve over time and capture potential structural changes in the data, and are computationally efficient. The initial training sample for the base models is set to $T_0 = 153 - h$, ensuring that a total of 50 years (600 months) of out-of-sample point forecasts for all periods are generated for evaluating model performance.

In addition, I also consider a more complex base model beyond the simple linear specification. In Section 5.3, I apply the state-of-the-art dynamic shrinkage process and its Markov-switching extension as the base model to investigate the improvements of the VECM over the SSM when the base models are more appropriately specified.

5.2.1 Forecasts Combination

Forecast pooling combines the point forecasts from the above-mentioned single base models. I modify equation (3.4) as

$$\text{Growth}_{t+h} = \omega_{0,t} + \hat{\boldsymbol{x}}'_{t+h}\boldsymbol{\omega}_t + \epsilon_{t+h}, \tag{5.4}$$

where $\hat{\boldsymbol{x}}_{t+h}$ is the vector of predictive values $\widehat{\text{Growth}}_{t+h}$ obtained from Model (5.1), Model (5.2), and Model (5.3). That is, at time t, each base model has already produced an h-step-ahead

forecast for π^h_{t+h}, and these forecasts serve as inputs to the synthesis equation. The coefficients ω_t then determine how the individual forecasts are combined, while $\omega_{0,t}$ captures any systematic bias not explained by the base model predictions.

I set $T_1 = 252$, which yields a total of roughly 42 years (500 months) of data for evaluating model performance. For an h-step-ahead forecast, the number of evaluation observations decreases to $T - h - T_1 + 1$. In this application, I consider $h = 1, 2, 3, 6, 12$ months ahead forecast.

5.2.2 Out-of-Sample Forecasting

The out-of-sample performance is evaluated based on the metrics mentioned in Section 3.4.

Overall, the proposed VECM model and its variants outperform the benchmark SSM across most evaluation metrics, with only a few exceptions: the LPDR of VECM at $h = 2$, and the RMSFE and CRPS of VECM-MS and VECM-MS-μ at $h = 12$. They also deliver substantial improvements over IMS, indicating that capturing abrupt jumps alone is insufficient to model realistic fluctuations. Except for these specific cases, the results overwhelmingly favor the proposed approach. For all horizons with $h < 12$, the model combination methods consistently outperform the base models across all metrics considered.

Figure 9 and Figure 10 present the cumulative log predictive density ratio (LPDR) relative to the benchmark state-space model (SSM) for one-period-ahead and three-period-ahead forecasts, respectively. Across both horizons, the proposed VECM-MS specifications (and other form of VECM) generally deliver higher cumulative LPDR values than both the benchmarks, indicating improved density forecasting performance.

In both cases, there are periods when the VECM models perform similarly to the SSM, followed by sharp upward shifts that coincide with episodes of structural change (e.g., the early 2000s recession, the global financial crisis, and the onset of COVID-19). This illustrates the benefit of incorporating the VECM-MS structure in the model assembly, as the error-correction mechanism enables deviated weights to adjust back toward the equilibrium more rapidly when the predictive portfolio shifts, leading to notable improvements

in forecasting performance. While the models do not respond immediately to economic shocks, they response much faster than the benchmark, which explains why the LPDR increases slightly after the onset of each event. This finding is consistent with the results in Figure 6 from the Monte Carlo study (Section 4.3), where density gains primarily arise slightly after periods of structural change.

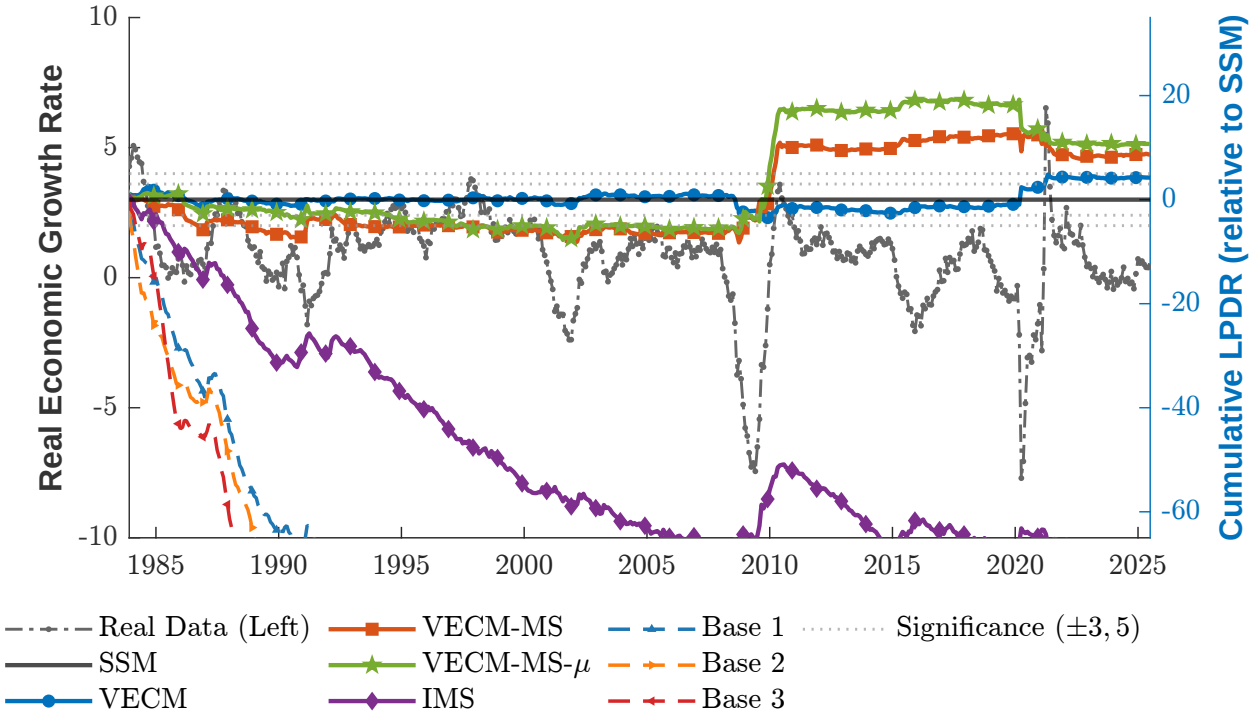


Figure 9: Comparative Forecast Performance for 1-month-ahead Forecasting Using Log Predictive Density Ratio (LPDR) with reference: SSM. Out-of-sample period: 1983/M12 - 2025/M7, in total 500 months (41.7 years).

All remaining evaluations for multi-period-ahead forecasts are reported in Table 4. The proposed VECM specifications generally outperform the benchmark SSM across all metrics and forecast horizons, with only a few exceptions noted earlier. Overall, the results indicate a consistent improvement in predictive performance for all VECM variants relative to the benchmark. At $h = 3$, the gains in RMSFE and CRPS are somewhat weaker than those in LPDR but strengthen again at $h = 6$. All VECM-based models outperform the individual base models, except at $h = 12$, where the LPDR remains positive, while RMSFE and CRPS become negative for both VECM-MS and VECM-MS-μ.

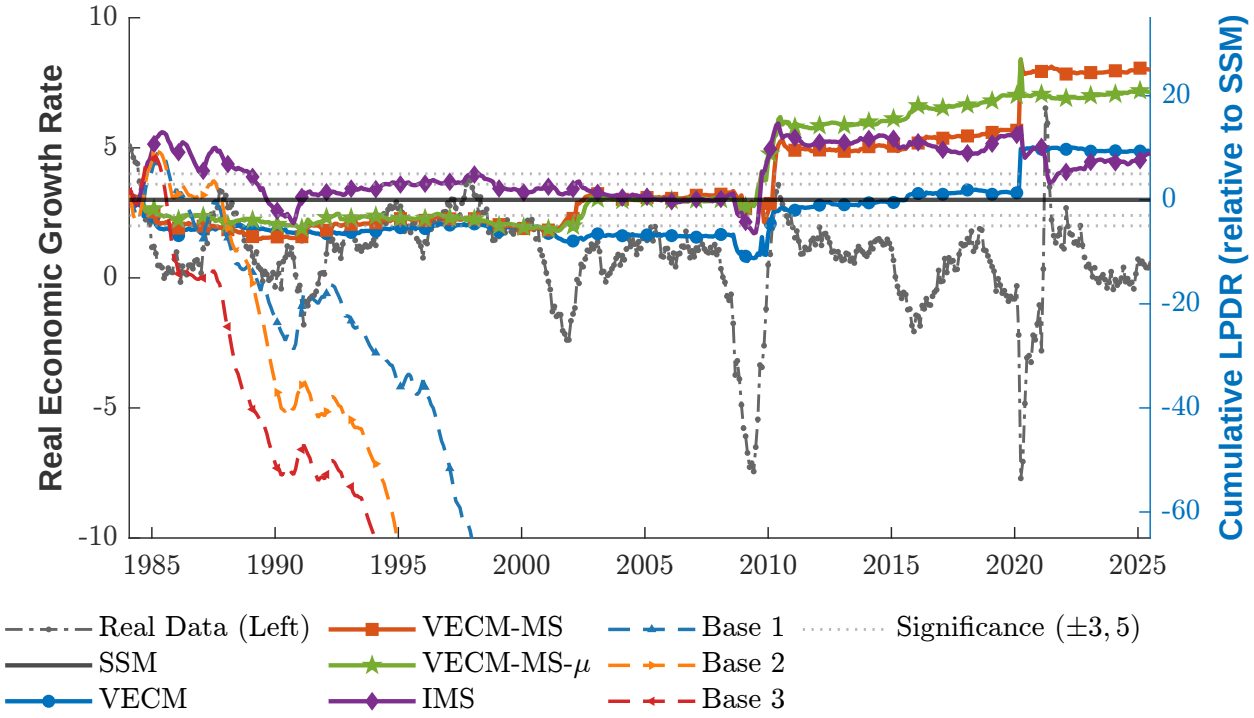


Figure 10: Comparative Forecast Performance for 3-month-ahead Forecasting Using Log Predictive Density Ratio (LPDR) with reference: SSM. Out-of-sample period: 1984/M2 - 2025/M7, in total 498 months (41.5 years).

This decline for $h = 12$ arises from the inclusion of the Markov-switching error-correction coefficients, which enable faster regime transitions. Generally, quicker adjustment leads to better performance; however, at this particular horizon, coinciding with the global financial crisis, the growth rate exhibits an abrupt collapse followed by a sharp surge. The rapid state transition in the VECM-MS captures the temporary downturn more quickly but also amplifies the subsequent departure to the spike, resulting in large deviations from the true values. In contrast, the standard VECM and SSM, being slower to adjust, smooth over these extreme fluctuations and therefore perform relatively better in this period.

As illustrated in Figure 11, there is a notable lag in model responsiveness for such longer period ahead forecasting. Specifically, while the realised growth rate reached its trough of -7.44% in June 2009, neither model captured this downturn immediately. By February 2010, the VECM-MS eventually responded, dropping to -6.99%, whereas the SSM exhibited significant inertia, declining only to -1.49%. Crucially, by this time, the realised growth rate had already rebounded to 0.9%, resulting in a substantial deviation for the VECM-MS. This divergence persisted into June 2010; the SSM finally fell to -7.44% and the VECM-MS deepened to -8.09%, despite the realised growth having already recovered to 3.58%

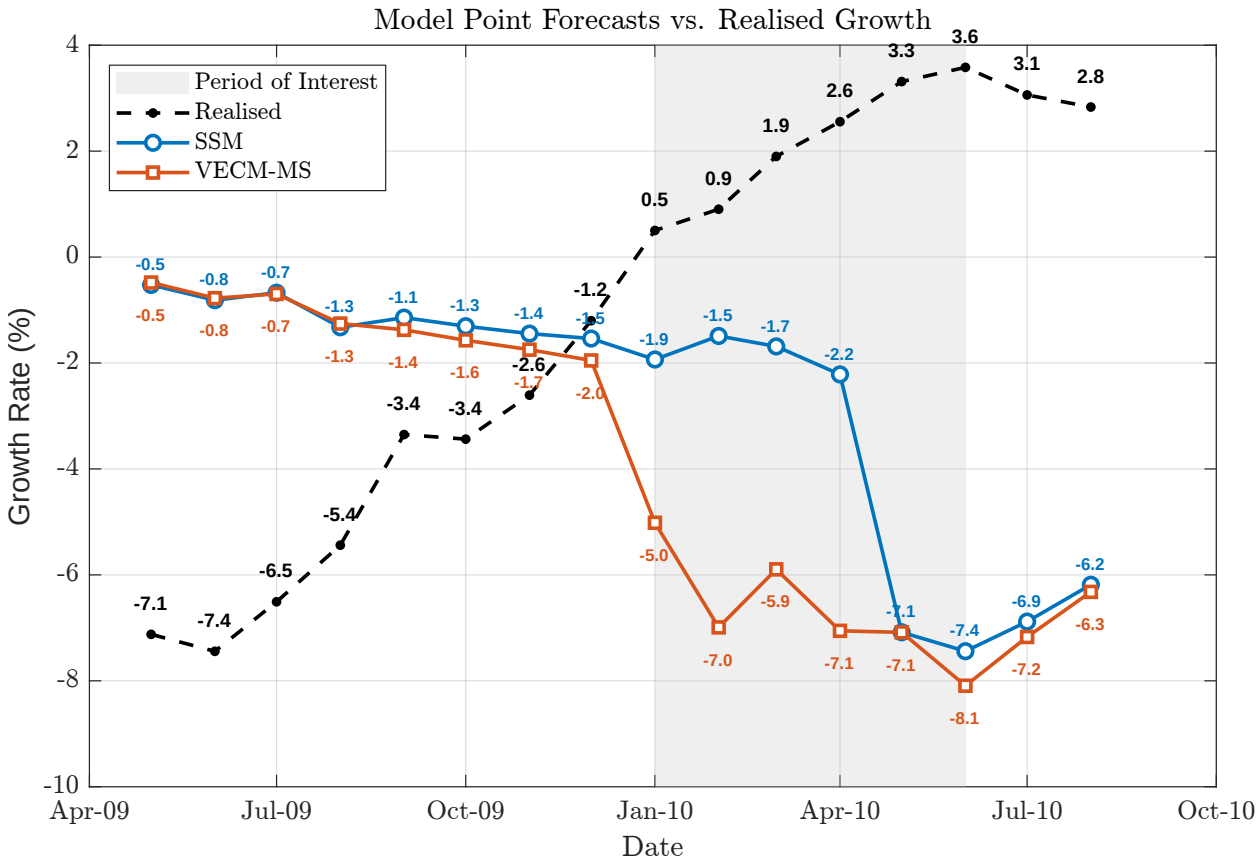


Figure 11: Comparison of 12-period Ahead Forecasts during the GFC Recovery (2009–2010). The figure highlights the lagged response of the VECM-MS and SSM models relative to the realised growth rate.

Table 4: Economic Growth Forecasting Performance Comparison Across Different Horizons

Model	Metric	Forecast Horizon (h months)				
		$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$
SSM - Benchmark		-	-	-	-	-
VECM	RMSFE	2.14%	1.40%	1.13%	1.53%	0.51%
	(CRPS)	(1.16%)	(1.73%)	(0.27%)	(1.70%)	(0.62%)
	{LPDR}	{4.24}	{-2.42}	{9.35}	{10.24}	{7.86}
VECM-MS	RMSFE	7.32%	1.59%	0.45%	3.75%	-4.82%
	(CRPS)	(2.96%)	(0.64%)	(0.85%)	(4.32%)	(-2.01%)
	{LPDR}	{8.62}	{3.39}	{24.94}	{6.85}	{0.30}
VECM-MS-μ	RMSFE	3.00%	0.53%	1.26%	4.10%	-6.64%
	(CRPS)	(1.55%)	(0.05%)	(0.34%)	(2.65%)	(-2.13%)
	{LPDR}	{10.68}	{10.44}	{20.63}	{0.58}	{14.26}
IMS	RMSFE	-9.87%	-8.49%	-10.65%	-3.67%	-5.34%
	(CRPS)	(-12.81%)	(-13.20%)	(-14.95%)	(-13.01%)	(-18.67%)
	{LPDR}	{-70.95}	{-12.06}	{8.93}	{-232.58}	{-771.98}
Base 1	RMSFE	-91.7%	-47.9%	-32.9%	-9.36%	-2.83%
	(CRPS)	(-93.0%)	(-57.6%)	(-37.0%)	(-13.0%)	(-11.96%)
	{LPDR}	{-378.3}	{-248.5}	{-144.2}	{-79.1}	{-131.4}
Base 2	RMSFE	-151.3%	-89.2%	-59.6%	-12.2%	6.89%
	(CRPS)	(-142.4%)	(-85.7%)	(-55.2%)	(-10.9%)	(2.40%)
	{LPDR}	{-429.3}	{-262.1}	{-163.6}	{-67.3}	{-84.2}
Base 3	RMSFE	-153.9%	-91.8%	-64.3%	-18.5%	5.15%
	(CRPS)	(-148.6%)	(-84.1%)	(-54.7%)	(-14.8%)	(0.63%)
	{LPDR}	{-414.0}	{-257.4}	{-145.5}	{-38.9}	{-33.9}

Notes: RMSFE = Root Mean Square Forecast Error; CRPS = Continuous Ranked Probability Score; LPDR = Log Predictive Density Ratio relative to benchmark SSM. Lower RMSFE and CRPS values indicate superior forecasting performance, but here both RMSFE and CRPS has been calculated as the percentage improvement relative to benchmark SSM. Higher LPDR values indicate better performance relative to the benchmark SSM. Bold values denote the optimal value within each column. Statistical significance is assessed using the Diebold-Mariano test with mean squared error loss function: ** denotes significance at the 5% level and * denotes significance at the 10% level, indicating significantly better performance than the benchmark SSM. Base models 1, 2, and 3 are linear specifications evaluated using equations (5.1), (5.2), and (5.3).

5.3 Base models: DSP and MSDSP

Forecasting involves several sources of uncertainty, including parameter uncertainty, model uncertainty, and structural instability. Parameter uncertainty is naturally handled through Bayesian posterior sampling, and model uncertainty can be addressed through model combination methods. Structural instability is often treated using time-varying parameter approaches. Rolling window method provides only a partial solution. To more effectively accommodate structural changes, I consider more flexible state space specifications. In particular, I examine two parallel scenarios: (1) all base models are estimated under the Dynamic Shrinkage Process (DSP) of Kowal et al. (2019), and (2) all base models are estimated under the extended Markov Switching Dynamic Shrinkage Process (MSDSP) of Fan et al. (2025).

The DSP provides a flexible formulation for time varying parameter models by allowing each coefficient to evolve smoothly over time while applying adaptive shrinkage. When the data offer limited evidence for movement, the shrinkage mechanism pulls the change of coefficient toward zero, thereby reducing unnecessary fluctuations. This yields a parsimonious representation of gradual structural change.

The MSDSP extends the DSP by introducing a latent regime-switching mechanism that governs how each coefficient behaves. These regimes are driven by a Markov switching process, enabling the potential for the time-varying parameter to switch to zero for sparsity, to shrink to a constant value, or to revolve dynamically according to the state space structure. The MSDSP therefore accommodates a wider range of behaviour, including sparsity, regime shifts, and persistent dynamics, while retaining the shrinkage principles of the DSP.

Both methods are applied to the same economic fundamentals specified in equations (5.1), (5.2), and (5.3), with the difference arising solely from the coefficient specification. I primarily focus on the MSDSP, since it nests several existing models, and the DSP is its

most immediate special case.

$$\text{Growth}_{t+h} = \gamma_{0,t} + \gamma_{1:3,t}\,\text{Unem}_{t:t-2} + \nu_{t+h}, \tag{5.5}$$

$$\text{Growth}_{t+h} = \gamma_{0,t} + \gamma_{1,t}\,\text{Spread}_{t} + \nu_{t+h}, \tag{5.6}$$

$$\text{Growth}_{t+h} = \gamma_{0,t} + \gamma_{1,t}\,\text{Inflation}_{t} + \nu_{t+h}, \tag{5.7}$$

Here, the γ's are no longer static but time-varying, following a structure in which the $p \times 1$ vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)'$ contains $i = 1, \ldots, p$ coefficients, with p determined by the number of covariates in equations (5.5), (5.6), and (5.7).

$$\boldsymbol{\gamma}_t = \boldsymbol{s}_t^{\gamma} \circ \tilde{\boldsymbol{\gamma}}_t + (\boldsymbol{\ell} - \boldsymbol{s}_t^{\gamma}) \circ \mathbf{0} \tag{5.8}$$

$$\tilde{\gamma}_{it} = \tilde{\gamma}_{i,t-1} + \omega_{it}^{\gamma} \qquad \omega_{it}^{\gamma} \sim N\left(0, \exp(h_{it}^{\gamma})\right) \tag{5.9}$$

$$h_{i,t+1}^{\gamma} = \mu_i^{\gamma} + \phi_i^{\gamma}(h_{it}^{\gamma} - \mu_i^{\gamma}) + \eta_{i,t+1}, \qquad \eta_{i,t+1} \sim Z(\alpha_h, \beta_h, 0, 1) \tag{5.10}$$

$$P_{jk}^{i,\gamma} = P(s_{it}^{\gamma} = k \mid s_{i,t-1}^{\gamma} = j) \tag{5.11}$$

Let $t = 1, \ldots, T$ and $j, k \in \{0, 1\}$. The operator $\circ$ in Equation (5.8) denotes the Hadamard (elementwise) product, and $\mathbf{0}$ is a $p \times 1$ vector of zeros.

The introduction of the $p \times 1$ variable $\boldsymbol{s}_t^{\gamma} = (s_{1t}^{\gamma}, \ldots, s_{pt}^{\gamma})$ allows the elements of $\boldsymbol{\gamma}_t$ to take either the value zero or the corresponding element of the "shadow" coefficient $\tilde{\boldsymbol{\gamma}}_t$, depending on the Markovian transition probabilities specified in (5.11). We assume that these Markovian switches operate independently across the p dimensions, so that $\boldsymbol{s}_i^{\gamma} \perp\!\!\!\perp \boldsymbol{s}_j^{\gamma}$ for $j \neq i$, where $\boldsymbol{s}_i^{\gamma} = (s_{i1}^{\gamma}, \ldots, s_{iT}^{\gamma})$. Finally, the Z-distribution in (5.10) generalizes the horseshoe prior when $\alpha_h = \beta_h = 1/2$, as discussed in Kowal et al. (2019).

The dynamic shrinkage–only specification can be obtained by restricting $\boldsymbol{s}_t^{\gamma} = \boldsymbol{\ell}$, or equivalently by setting $P_{jj}^{i,\gamma} = 1$ for both $j \in \{0, 1\}$. Since the primary objective of this paper is to examine how forecast synthesis performs when the base models are relatively well specified, the detailed specification of the base models will not be emphasized too much. For extra details, see Kowal et al. (2019); Fan et al. (2025).

5.3.1 Forecasts Combination

The synthesis step remains the same as in the linear model case described in Section 5.2.1. The key difference is that $\hat{\boldsymbol{x}}_{t+h}$ in equation (5.4) now represents the vector of predictive values $\widehat{\text{Growth}}_{t+h}$ obtained from Model (5.5), Model (5.6), and Model (5.7), each estimated under both DSP and MSDSP frameworks. The evaluation sample spans 400 months, with $T_1 = 352$ marking the beginning of the forecast evaluation period.

5.3.2 Out-of-Sample Forecasting

The out-of-sample performance based on more sophisticated statistical base models shows a broadly similar pattern to that obtained with linear base models. Table 5 reports results using DSP base models, and Table 6 reports results using MSDSP base models

Overall, the forecasting results improve relative to the simple SSM framework and also IMS across horizons and evaluation metrics, with only a few exceptions. Most notably, the main specification of interest, VECM-MS, delivers substantially better performance in nearly all cases, with the only exception occurring for the DSP input at $h = 6$ in terms of the LPDR in Table 5. Aside from this isolated case, it consistently outperforms the benchmarks across metrics and horizons. Other VECM variants are not uniformly dominant but still yield meaningful gains in most settings.

Figure 12 illustrates this exception in more detail by comparing the SSM and VECM-MS at horizon $h = 6$ during the GFC. The VECM-MS produces more accurate point forecasts and tracks the realised data more closely, yet both models exhibit sizeable deviations from the true values. A crucial difference is that the VECM-MS generates a narrower predictive volatility. In most scenarios, this combination of tighter volatility and accurate point prediction would yield a higher LPDR. Here, however, the point forecast errors at $h = 6$ remain relatively large, and the reduced predictive variance leads to stronger penalisation in the LPDR measure.

It is also worth noting that model combination methods are not guaranteed to outperform every individual base model, particularly when the underlying models are already sophisticated and well specified. In such settings, structural instability often plays a more

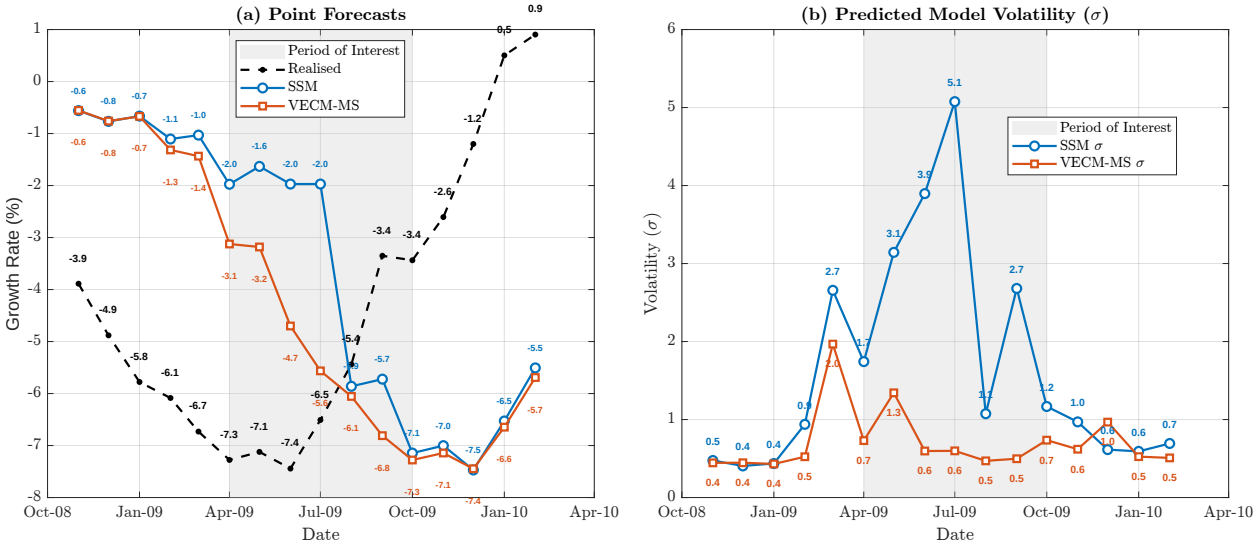


Figure 12: Comparison of point forecasts and predictive volatility for SSM and VECM-MS at horizon $h = 6$ during the GFC.

prominent role than model uncertainty in determining predictive performance. For instance, in Table 5 at $h = 1$, all DSP base models outperform the combined methods in terms of RMSFE and CRPS, though not in LPDR. Similarly, in Table 6 at $h = 12$, MSDSP base model 3 achieves a lower RMSFE than all assembly models. Nonetheless, the proposed approach demonstrates clear improvements over existing benchmark combination methods overall.

Table 5: Economic Growth Forecasting Performance Comparison Across Different Horizons

Model	Metric	Forecast Horizon (h months)				
		$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$
SSM - Benchmark		-	-	-	-	-
VECM	RMSFE	1.46%	1.00%	-0.32%	1.40%	0.36%
	(CRPS)	(1.10%)	(0.49%)	(-0.04%)	(0.78%)	(0.47%)
	{LPDR}	{4.12}	{1.65}	{3.19}	{-6.26}	{16.15}
VECM-MS	RMSFE	4.25%**	0.17%	1.85%*	1.94%	0.20%
	(CRPS)	(3.59%)	(0.55%)	(2.46%)	(0.69%)	(1.67%)
	{LPDR}	{14.95}	{0.93}	{15.07}	{-6.17}	{76.73}
VECM-MS-μ	RMSFE	1.81%	3.24%	1.47%	2.91%	0.21%
	(CRPS)	(0.92%)	(-3.75%)	(0.72%)	(0.44%)	(-0.51%)
	{LPDR}	{3.58}	{0.11}	{10.40}	{-22.69}	{-41.98}
IMS	RMSFE	-10.86%	-21.71%	-32.18%	-44.72%	-17.54%
	(CRPS)	(-5.66%)	(-22.54%)	(-25.50%)	(-27.62%)	(-16.75%)
	{LPDR}	{-57.99}	{-24.65}	{27.58}	{-10.16}	{-12.28}
Base 1	RMSFE	14.4%	-1.22%	-1.93%	3.45%	-7.63%
	(CRPS)	(13.9%)	(6.10%)	(0.22%)	(-18.1%)	(-25.0%)
	{LPDR}	{-131.0}	{-360.7}	{-608.9}	{-1190.8}	{-1768.8}
Base 2	RMSFE	20.7%	6.27%	-1.68%	-7.17%	-11.0%
	(CRPS)	(18.2%)	(8.00%)	(2.17%)	(-16.5%)	(-20.1%)
	{LPDR}	{-65.1}	{-273.2}	{-453.3}	{-930.8}	{-1357.8}
Base 3	RMSFE	17.4%	-17.5%	-104.2%	-24.5%	-35.5%
	(CRPS)	(19.9%)	(3.44%)	(-12.7%)	(-30.3%)	(-25.5%)
	{LPDR}	{-24.4}	{-118.2}	{-249.4}	{-708.6}	{-730.8}

Notes: RMSFE = Root Mean Square Forecast Error; CRPS = Continuous Ranked Probability Score; LPDR = Log Predictive Density Ratio relative to benchmark SSM. Lower RMSFE and CRPS values indicate superior forecasting performance, but here both RMSFE and CRPS has been calculated as the percentage improvement relative to benchmark SSM. Higher LPDR values indicate better performance relative to the benchmark SSM. Bold values denote the optimal value within each column. Statistical significance is assessed using the Diebold-Mariano test with mean squared error loss function: ** denotes significance at the 5% level and * denotes significance at the 10% level, indicating significantly better performance than the benchmark SSM. Base models 1, 2, and 3 are DSP specifications evaluated using equations (5.5), (5.6), and (5.7).

Table 6: Economic Growth Forecasting Performance Comparison Across Different Horizons

Model	Metric	Forecast Horizon (h months)				
		$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$
SSM - Benchmark		-	-	-	-	-
VECM	RMSFE	2.46%	1.40%	0.78%	-0.94%	0.78%
	(CRPS)	(1.30%)	(0.20%)	(0.09%)	(0.48%)	(0.64%)
	{LPDR}	{2.22}	{1.12}	{7.17}	{8.74}	{1.97}
VECM-MS	RMSFE	7.93%**	4.31%*	1.64%	4.63%	1.67%
	(CRPS)	(3.95%)	(1.98%)	(0.81%)	(2.62%)	(1.12%)
	{LPDR}	{2.78}	{5.91}	{6.94}	{22.52}	{30.88}
VECM-MS-μ	RMSFE	5.06%**	4.77%*	3.47%	3.37%	1.70%*
	(CRPS)	(4.01%)	(2.24%)	(1.08%)	(1.26%)	(0.96%)
	{LPDR}	{4.48}	{0.88}	{-8.00}	{-1.99}	{-17.77}
IMS	RMSFE	7.45%	3.44%	-2.83%	-9.94%	-11.23%
	(CRPS)	(3.61%)	(-4.87%)	(-9.23%)	(-12.15%)	(-18.76%)
	{LPDR}	{-27.67}	{23.10}	{46.99}	{-42.06}	{-283.92}
Base 1	RMSFE	-35.7%	-38.6%	-35.9%	-43.7%	-7.74%
	(CRPS)	(-67.7%)	(-66.9%)	(-58.1%)	(-68.1%)	(-26.3%)
	{LPDR}	{-291.5}	{-124.1}	{-89.8}	{-166.7}	{-262.5}
Base 2	RMSFE	-165.3%	-51.9%	-15.2%	4.41%	-36.8%
	(CRPS)	(-277.6%)	(-95.5%)	(-19.9%)	(-8.97%)	(-77.0%)
	{LPDR}	{-264.4}	{-126.8}	{-69.8}	{-142.2}	{-408.1}
Base 3	RMSFE	-25.7%	-3.14%	1.64%	10.9%	12.2%
	(CRPS)	(-49.1%)	(-19.8%)	(-4.09%)	(4.09%)	(-6.31%)
	{LPDR}	{-323.5}	{-298.4}	{-276.1}	{-205.7}	{-358.5}

Notes: RMSFE = Root Mean Square Forecast Error; CRPS = Continuous Ranked Probability Score; LPDR = Log Predictive Density Ratio relative to benchmark SSM. Lower RMSFE and CRPS values indicate superior forecasting performance, but here both RMSFE and CRPS has been calculated as the percentage improvement relative to benchmark SSM. Higher LPDR values indicate better performance relative to the benchmark SSM. Bold values denote the optimal value within each column. Statistical significance is assessed using the Diebold-Mariano test with mean squared error loss function: ** denotes significance at the 5% level and * denotes significance at the 10% level, indicating significantly better performance than the benchmark SSM. Base models 1, 2, and 3 are MSDSP specifications evaluated using equations (5.5), (5.6), and (5.7).

6 Application to Inflation Forecasting

The second application is inflation forecasting, which is central to economic decision-making: it informs portfolio allocation, price and wage setting, household financial planning, and the conduct of monetary policy

6.0.1 Data and Base Models

The dataset consists of the chain-type price index (GDPCTPI) as a measure of the price level, the federal funds rate as the measure of the interest rate, and the unemployment rate. All series are publicly available from the FRED database. The sample period spans from 1954Q3 to 2025Q1, yielding a total of $T = 283$ quarters, as the federal funds rate begins in 1954Q3. To construct forecasts up to twelve periods ahead, the price index series is used starting from 1951Q3. Since both the literature and the available data employ seasonally adjusted series, I follow this convention rather than applying the adjustment ourselves, in contrast to the economic growth application in Section 5.

Inflation is constructed following Stock and Watson (1999) so that the forecast target aligns exactly with the decision horizon. The h quarter inflation rate is defined as $\pi_t^h = 400/h \log(P_t/P_{t-h})$, where P_t denotes the price level and π_t^h is reported at an annual rate. The one quarter case yields $\pi_t \equiv \pi_t^1 = 400 \log(P_t/P_{t-1})$, which represents quarterly inflation at an annual rate.

The base forecasting models follow McAlinn and West (2019) and consider four specifications:

$$\pi_{t+h}^h = \gamma_0 + \gamma_1 \pi_t^h + \nu_{t+h}, \tag{6.1}$$

$$\pi_{t+h}^h = \gamma_0 + \boldsymbol{\gamma}_{1:3} \pi_{t-1:t-3}^h + \boldsymbol{\gamma}_{4:6} r_{t-1:t-3} + \boldsymbol{\gamma}_{7:9} \mathrm{Unem}_{t-1:t-3} + \nu_{t+h}, \tag{6.2}$$

$$\pi_{t+h}^h = \gamma_0 + \boldsymbol{\gamma}_{1:3} \pi_{t-1:t-3}^h + \nu_{t+h}, \tag{6.3}$$

$$\pi_{t+h}^h = \gamma_0 + \gamma_1 \pi_t^h + \gamma_2 r_{t-1} + \gamma_3 \mathrm{Unem}_{t-1} + \nu_{t+h}, \tag{6.4}$$

where π_t^h denotes the h-quarter inflation rate at an annualized percentage rate, r_t is the federal funds rate, and Unem_t is the unemployment rate. Model (6.1) is an autoregres-

sive model with the current inflation rate as the sole predictor. Model (6.2) augments lagged inflation with lagged interest rates and unemployment over the past three quarters. Model (6.3) is a purely autoregressive specification with three lags of inflation. Model (6.4) includes the current inflation rate along with the most recent lag of the interest rate and unemployment rate. The error term ν_{t+h} captures unexplained variation and is modeled with stochastic volatility to allow for time-varying uncertainty, adopting the same log variance specification as that of the synthesis function in Equation (2.3).

All base models are estimated in a rolling-window framework with a window length of 100 quarters, allowing the coefficients to evolve over time and capture potential structural changes in the data. The initial training sample for the base models is set to $T_0 = 84 - h$, ensuring that a total of 50 years (200 quarters) of out-of-sample point forecasts for all periods are generated for evaluation.

6.0.2 Forecasts Combination

To combine the point forecasts from each base model, I modify equation (3.4) as

$$\pi^h_{t+h} = \omega_{0,t} + \boldsymbol{x}_t^{(h)\prime}\boldsymbol{\omega}_t + \epsilon_{t+h}, \tag{6.5}$$

where $\boldsymbol{x}_t^{(h)}$ is the vector of predictive values $\hat{\pi}^h_{t+h}$ obtained from Model (6.1), Model (6.2), Model (6.3), and Model (6.4). That is, at time t, each base model has already produced an h-step-ahead forecast for π^h_{t+h}, and these forecasts serve as inputs to the synthesis equation. The coefficients $\boldsymbol{\omega}_t$ then determine how the individual forecasts are combined, while $\omega_{0,t}$ captures any systematic bias not explained by the base model predictions.

I set $T_1 = 139$, which yields a total of 36 years (144 quarters) of data for evaluating model performance. For an h-step-ahead forecast, the number of out-of-sample observations is $T - h - T_1 + 1$. In this application, I consider $h = 1, 2, 4, 6, 8, 12$ quarters ahead forecast.

6.0.3 Out-of-Sample Forecasting

The out-of-sample performance is evaluated based on the metrics mentioned in section 3.4.

Overall, I observe a clear dominance of the proposed VECM model (and its variants) over the benchmark SSM across all evaluation metrics, with only two exceptions: the CRPS value of VECM-MS at $h = 1$ and the CRPS of VECM-MS-μ at $h = 12$. When it comes to IMS, the forecasting performance of VECM can still be substantially improved over IMS in terms of RMSFE and CRPS, except for the LPDR when $h > 1$.

Figure 13 displays the cumulative log predictive density ratio (LPDR) for one-period-ahead inflation forecasting relative to the benchmark state-space model (SSM). The LPDR serves as a dynamic measure of relative density forecast performance, where positive values signify improvements over the benchmark model.

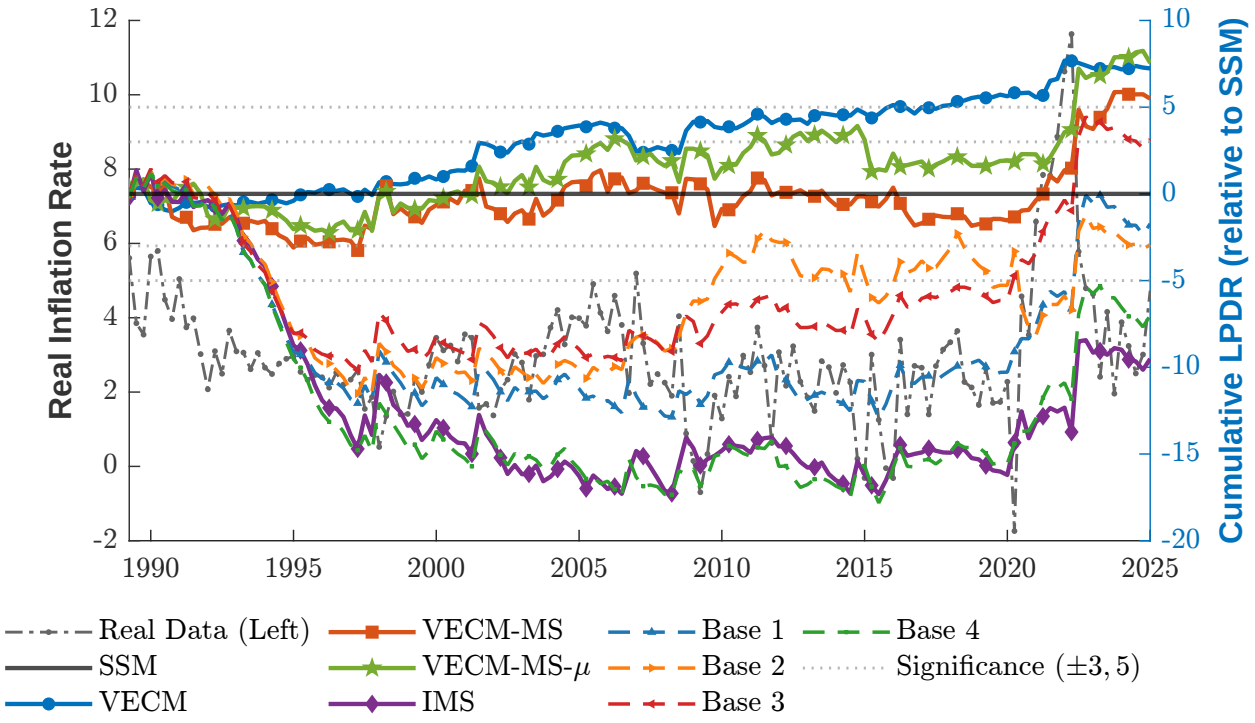


Figure 13: Comparative Forecast Performance for 1-quarter-ahead Forecasting Using Log Predictive Density Ratio (LPDR) with reference: SSM. Out-of-sample period: 1989/Q2 - 2025/Q1, in total 144 quarters (36 years).

The proposed VECM specifications demonstrate substantial improvements in forecasting accuracy. The standard VECM consistently generates positive cumulative gains,

while the extended variants: VECM-MS and VECM-MS-μ also enhance predictive performance over the benchmark. These improvements are particularly pronounced during periods of heightened macroeconomic uncertainty, such as the post-2020 inflation surge.

In contrast, the individual base models display persistently negative LPDR values over most of the sample period. Only the base model 3 attains positive performance relative to the SSM after 2022, yet its improvement remains inferior to that of the VECM. This evidence underscores the limited ability of individual models to generate accurate density forecasts when used in isolation.

These findings underscore the superior adaptability of the VECM framework, particularly in environments characterized by regime shifts. The framework's ability to maintain robust performance across varying economic conditions highlights its value for practical forecasting applications.

However, the gains from adopting the VECM specification are not immediately evident in this example. Intuitively, if forecasts are already anchored to the long-run equilibrium, I should expect to see small improvements due to the slow error correction. In such cases, the short-term adjustment mechanism enables the model to capture deviations and revert to equilibrium more quickly than models without VECM. The advantages are especially pronounced at longer forecast horizons. For instance, Figure 14 shows that the VECM-based methods outperform the benchmark in a step-like function, particularly during episodes after abrupt inflation jumps (e.g., the early 2000s recession, the global financial crisis, and the onset of COVID-19), which coincide with the evidence observed from Figure 6 in the Monte Carlo study Section 4.3 and also Figure 9 and Figure 10 in the Economic growth Section 5.2.2

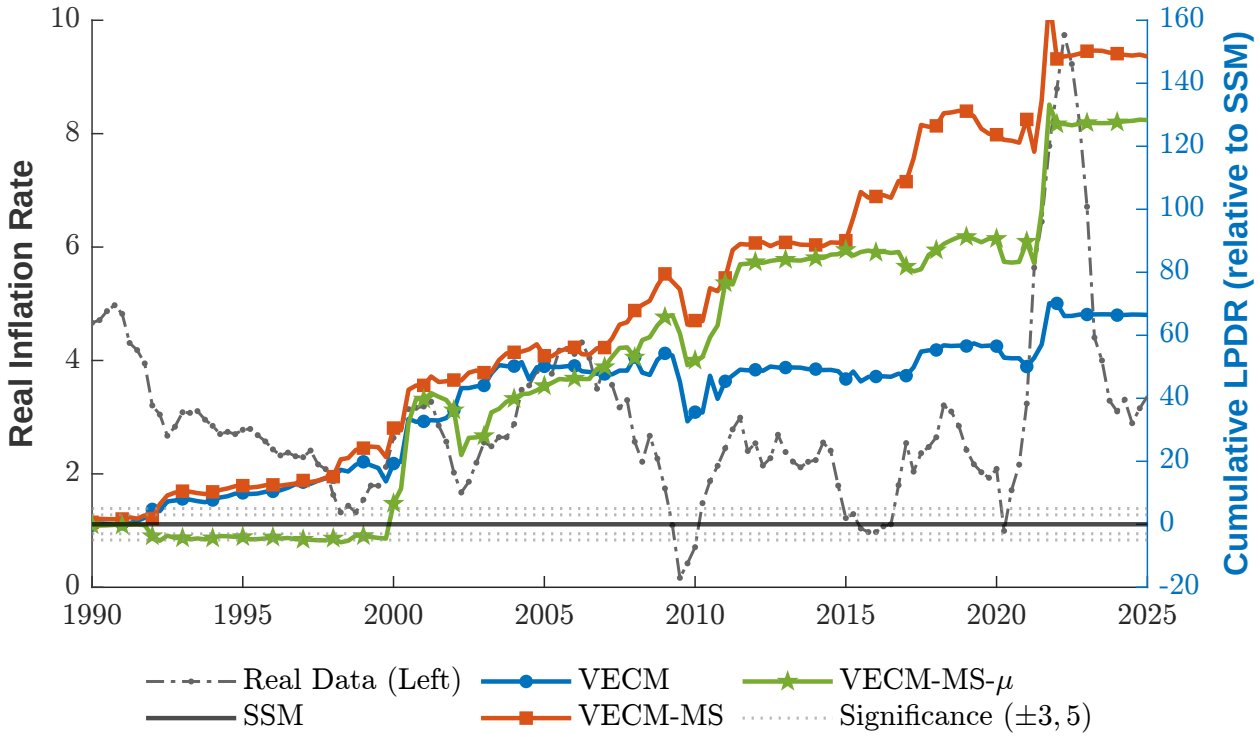


Figure 14: Comparative Forecast Performance for 4-quarter-ahead Forecasting Using Log Predictive Density Ratio (LPDR) with reference: SSM. Out-of-sample period: 1990/Q1 - 2025/Q1, in total 141 quarters (35 years +). IMS and base models are excluded due to their dominant LPDR value.

The results for all other multi-period-ahead forecasts are reported in Table 7. Our proposed VECM methods outperform the benchmark SSM across all metrics and horizons, with the exception of the two cases noted earlier. At every horizon, the VECM specifications exceed the benchmark SSM by more than 5 units in terms of LPDR. Both RMSFE and CRPS are demonstrating substantially positive improvements, indicating better predictive accuracy. Although only a few entries are marked as statistically significant, the overall consistency demonstrates substantial gains from the proposed model combination approach.

It is also important to note that across all horizons, the benchmark SSM does not consistently outperform the base models, which again echoes the trade-off between structural instability and model uncertainty. For example, at $h = 1$, the SSM exhibits larger RMSFE and CRPS than base models 1, 3, and 4, and a smaller log score than base model 3. While adopting VECM-MS substantially improves RMSFE and LPDR, the CRPS does not display comparable gains. Similar patterns emerge at other horizons, where the SSM fails to dominate the base models in most cases. By contrast, VECM-MS generally outperforms all base models in terms of RMSFE, except at $h = 2$ and $h = 4$.

Although the VECM is significantly better than the SSM in terms of the log predictive Bayes factor, none of the model combination methods achieve higher log scores than the individual base models for $h > 2$. This pattern is consistent with the evidence in **?**, who show that gains in density forecast performance depend on the complementarities among agent models and tend to be more pronounced at shorter horizons, with limited room for improvement when the individual models are already well calibrated.

The estimated error variance of the combination methods is noticeably smaller than that of the individual base models. This reflects the fact that combining forecasts often stabilises the predictive signal by averaging across multiple sources of information. The resulting aggregate forecast tends to be smoother than any single model and therefore leaves less residual variation for the error term. Although this leads to more accurate point forecasts, the reduced residual variability can limit the ability of the combination methods to capture long-term outliers arising from cumulative uncertainty.

Table 7: Inflation Forecasting Performance Comparison Across Different Horizons

Model	Metric	Forecast Horizon (h quarters)					
		$h=1$	$h=2$	$h=4$	$h=6$	$h=8$	$h=12$
SSM - Benchmark		-	-	-	-	-	-
VECM	RMSFE	4.39%**	6.59%	1.92%	6.34%	4.38%	0.28%
	(CRPS)	(**3.34%**)	(3.28%)	(0.75%)	(6.56%)	(1.71%)	(0.60%)
	{LPDR}	{7.23}	{6.93}	{53.68}	{124.28}	{44.36}	{21.08}
VECM-MS	RMSFE	**8.79%****	**14.23%**	2.09%	**11.13%**	0.78%	**0.35%**
	(CRPS)	(-0.82%)	(5.96%)	(0.34%)	(**9.55%**)	(0.97%)	(**2.77%**)
	{LPDR}	{5.49}	{20.55}	{**126.94**}	{**219.90**}	{**158.78**}	{**274.51**}
VECM-MS-μ	RMSFE	7.99%*	**10.45%****	3.66%	**10.33%**	5.21%	0.04%
	(CRPS)	(0.91%)	(**7.73%**)	(**1.28%**)	(4.16%)	(**2.19%**)	(-0.09%)
	{LPDR}	{**7.55**}	{**34.07**}	{109.39}	{69.69}	{77.53}	{27.76}
IMS	RMSFE	-10.03%	-4.96%	-17.90%	2.16%	-1.72%	-6.40%
	(CRPS)	(-5.26%)	(6.21%)	(-1.22%)	(6.99%)	(1.52%)	(-2.72%)
	{LPDR}	{-9.52}	{+}	{+}	{+}	{+}	{+}
Base 1	RMSFE	3.50%	21.16%	6.82%	6.47%	-1.29%	-16.05%
	(CRPS)	(3.50%)	(10.54%)	(26.52%)	(16.46%)	(5.82%)	(-7.17%)
	{LPDR}	{ -1.72 }	{+}	{+}	{+}	{+}	{+}
Base 2	RMSFE	-37.6%	-6.69%	-5.89%	-0.65%	-6.63%	-42.26%
	(CRPS)	(4.14%)	(-6.23%)	(18.45%)	(10.58%)	(-3.04%)	(-84.21%)
	{LPDR}	{-2.93}	{+}	{+}	{+}	{+}	{+}
Base 3	RMSFE	5.42%	20.48%	12.6%	8.75%	0.09%	-9.68%
	(CRPS)	(5.57%)	(5.94%)	(25.45%)	(21.84%)	(8.27%)	(-3.41%)
	{LPDR}	{3.15}	{+}	{+}	{+}	{+}	{+}
Base 4	RMSFE	2.34%	21.00%	4.49%	5.66	-4.18%	-18.46
	(CRPS)	(1.35%)	(2.92%)	(16.87%)	(13.05%)	(0.82%)	(-15.9%)
	{LPDR}	{-7.00}	{+}	{+}	{+}	{+}	{+}

Notes: RMSFE = Root Mean Square Forecast Error; CRPS = Continuous Ranked Probability Score; LPDR = Log Predictive Density Ratio relative to benchmark SSM. Lower RMSFE and CRPS values indicate superior forecasting performance, but here both RMSFE and CRPS has been calculated as the percentage improvement relative to benchmark SSM. Higher LPDR values indicate better performance relative to the benchmark SSM. Bold values denote the optimal value within each column among model assembly methods. Statistical significance is assessed using the Diebold-Mariano test with mean squared error loss function: ** denotes significance at the 5% level and * denotes significance at the 10% level, indicating significantly better performance than the benchmark SSM. '+' indicates dominating LPDR due to different model specifications.

7 Conclusion

This study establishes that imposing cointegrating restrictions on latent forecast weights offers a simple solution to the stability-flexibility trade-off in forecast combination. By treating the weight vector as a system anchored to a long-run equilibrium, the proposed framework resolves the tendency of standard synthesis methods to overfit short-term noise. The key methodological insight is that regularization need not come at the expense of flexibility; rather, a well-defined equilibrium attractor provides the necessary regularization to allow for unconstrained, non-convex weighting strategies in the short run.

Monte Carlo evidence confirms that the proposed method yields substantial improvements in predictive accuracy, particularly in environments where the optimal weight structure evolves dynamically. In empirical applications, the framework demonstrates consistent predictive gains over state-of-the-art benchmarks.

Notably, the Markov-switching specification of the adjustment coefficients plays an essential role in this performance. By allowing the error-correction mechanism to switch between slow-adjustment and fast-adjustment states, the model can distinguish between transient noise and fundamental shifts in model performance. This enables the weights to rapidly recalibrate when the predictive environment changes, without requiring explicit modeling of the underlying structural changes.

These findings highlight the value of imposing cointegrating relationships within forecast combination methods. The framework establishes a generalized structure that is broadly applicable and can be naturally extended to settings involving time-varying cointegrating vectors, alternative prior specifications, and policy evaluation contexts.

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*, 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.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.

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.

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.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple Classifier Systems*, pages 1–15, Berlin, Heidelberg. Springer Berlin Heidelberg.

Fan, Z., Maneesoonthorn, W., and Song, Y. (2025). A new perspective of the meese-rogoff puzzle: Application of sparse dynamic shrinkage.

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.

Garratt, A., Lee, K., Mise, E., and Shields, K. (2008). Real-time representations of the output gap. *The review of economics and statistics*, 90(4):792–804.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, 75(1):121–146.

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.

Giacomini, R. and Rossi, B. (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies*, 76(2):669–705.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378.

Hall, S. G. and Mitchell, J. (2007). Combining density forecasts. *International journal of forecasting*, 23(1):1–13.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 38(3):205–228.

Harvey, C. R. (1989). Forecasts of economic growth from the bond and stock markets. *Financial Analysts Journal*, 45(5):38–45.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. *Statistical Science*, 14(4):382 – 417.

Ishwaran, H. and Zarepour, M. (2002). Exact and approximate sum representations for the dirichlet process. *Canadian Journal of Statistics*, 30(2):269–283.

Jin, X., Maheu, J. M., and Yang, Q. (2022). Infinite markov pooling of predictive distributions. *Journal of econometrics*, 228(2):302–321.

Jochmann, M. and Koop, G. (2015). Regime-switching cointegration. *Studies in Nonlinear Dynamics & Econometrics*, 19(1):35–48.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Johnson, M. C. (2017). *Bayesian Predictive Synthesis: Forecast Calibration and Combination*. PhD thesis, Duke University.

Juselius, K. (2006). *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press.

Justiniano, A. and Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review*, 98(3):604–641.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Koop, G. and Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, 53(3):867–886.

Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2011). Bayesian inference in a time varying cointegration model. *Journal of Econometrics*, 165(2):210–220.

Koop, G., León-González, R., and Strachan, R. W. (2009). Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric reviews*, 29(2):224–242.

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.

Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of the American Statistical Association*, 89(428):1535–1546.

Martin, G. M., Frazier, D. T., Maneesoonthorn, W., Loaiza-Maya, R., Huber, F., Koop, G., Maheu, J., Nibbering, D., and Panagiotelis, A. (2024). Bayesian forecasting in economics and finance: A modern review. *International journal of forecasting*, 40(2):811–839.

McAlinn, K., Aastveit, K. A., Nakajima, J., and West, M. (2020). 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. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14(1-2):3–24.

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.

Nonejad, N. (2021). An overview of dynamic model averaging techniques in time-series econometrics. *Journal of Economic Surveys*, 35(2):566–614.

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.

Onorante, L. and Raftery, A. E. (2016). Dynamic model averaging in large model spaces using dynamic occam s window. *European Economic Review*, 81:2–14.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161.

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.

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

Rousseau, J. and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(5):689–710.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44(2):293–335.

Strachan, R. W. and Inder, B. (2004). Bayesian analysis of the error correction model. *Journal of Econometrics*, 123(2):307–325.

Tallman, E. and West, M. (2024). Bayesian predictive decision synthesis. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 86(2):340–363.

Villani, M. (2005). Bayesian reference analysis of cointegration. *Econometric Theory*, 21(2):326–357.

Waggoner, D. F. and Zha, T. (2012). Confronting model misspecification in macroeconomics. *Journal of Econometrics*, 171(2):167–184.

West, M. and Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2):241–259.

Appendix

A1 Extra Technical Details

This section provides additional details on the MCMC sampling procedure.

For brevity, the description focuses on one variant of the model, the VECM-$\boldsymbol{\alpha}_{s_t}$ specification; the remaining cases follow analogously:

$$y_t = \omega_{0,t} + \boldsymbol{x}_t'\boldsymbol{\omega}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right), \tag{A1.1}$$

$$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}_{s_t}(1 - \sum_j \omega_{j,t-1}) + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_\omega\right). \tag{A1.2}$$

where $\sigma_t^2 = \exp(h_t)$

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + e_t, \qquad e_t \sim N\left(0, \sigma_h^2\right), \tag{A1.3}$$

For estimation purpose, equation (A1.2) can be write back to the expression as following

$$\boldsymbol{\omega}_t = \boldsymbol{\alpha}_0 + A\boldsymbol{\omega}_{t-1} + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_\omega\right) \tag{A1.4}$$

where $A_{s_t} = \boldsymbol{\alpha}_{s_t}\beta' + I_p$ and $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}_{s_t}\mu = \boldsymbol{\alpha}_{s_t}$ as μ was defined to be 1.

A1.1 MCMC Sampling Algorithm

(a) Sample $\boldsymbol{\omega} \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{h}, \boldsymbol{\omega}_0, \boldsymbol{s}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_\omega$.

Conditional on the latent states $\boldsymbol{s}$ and the regime-specific adjustment coefficients $\boldsymbol{\alpha}_{s_t}$, the evolution of the model weights $\boldsymbol{\omega}_t$ follows

$$\boldsymbol{\omega}_t = \boldsymbol{\alpha}_{s_t} + A_{s_t}\boldsymbol{\omega}_{t-1} + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_\omega\right), \tag{A1.5}$$

where $A_{s_t} = \boldsymbol{\alpha}_{s_t}\beta' + I_p$. The observation equation is given by

$$y_t = \omega_{0,t} + \boldsymbol{x}_t'\boldsymbol{\omega}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right). \tag{A1.6}$$

Define $y_{-0,t} = y_t - \omega_{0,t}$, I have

$$y_{-0,t} = \boldsymbol{x}_t'\boldsymbol{\omega}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right). \tag{A1.7}$$

Stacking all time periods, we can express the joint system in SUR form as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\omega} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}\left(\mathbf{0}, \Omega_1\right), \tag{A1.8}$$

where $\Omega_1 = diag(\sigma_1^2, \ldots, \sigma_T^2)$. $\boldsymbol{y}, \boldsymbol{\epsilon}$ ara $T \times 1$, $\boldsymbol{X}$ takes SUR form with dimension $T \times Tp$, and $\boldsymbol{w}$ has dimension $pT \times 1$. The transition equation can be written as

$$D\boldsymbol{w} = \text{vec}(\boldsymbol{\alpha}) + G_1\boldsymbol{\omega}_0 + \boldsymbol{u}, \qquad \boldsymbol{u} \sim \mathcal{N}\left(\mathbf{0}, \Omega_2\right), \tag{A1.9}$$

where $G_1\boldsymbol{\omega}_0$ correspond to the prior for $\boldsymbol{\omega}_1$, $\Omega_2 = diag(\boldsymbol{\Sigma}_\omega, \ldots, \boldsymbol{\Sigma}_\omega)$ and D is a block diagonal matrix with dimension $Tp \times Tp$. $\boldsymbol{\alpha}$ is p by T matrix that includes $\boldsymbol{\alpha}_{s_t}$ from $t = 1$ to T.

$$D = \begin{pmatrix} I_p & 0 & 0 & \ldots & \ldots & 0 \\ -A_{s_t} & I_p & 0 & \ldots & \ldots & 0 \\ 0 & -A_{s_t} & I_p & \ldots & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & I_p & 0 \\ 0 & \ldots & \ldots & \ldots & -A_{s_t} & I_p \end{pmatrix} \quad \boldsymbol{w} = \begin{pmatrix} \boldsymbol{\omega}_1 \\ \boldsymbol{\omega}_2 \\ \boldsymbol{\omega}_3 \\ \vdots \\ \boldsymbol{\omega}_{T-1} \\ \boldsymbol{\omega}_T \end{pmatrix} \quad \boldsymbol{\omega}_t = \begin{pmatrix} \omega_{t,j=1} \\ \omega_{t,j=2} \\ \vdots \\ \omega_{t,j=p} \end{pmatrix} \quad G_1 = \begin{pmatrix} I_p \\ 0_p \\ 0_p \\ \vdots \\ 0_p \\ 0_p \end{pmatrix}$$

The full conditional distribution of $\boldsymbol{\omega}$ is Gaussian:

$$\boldsymbol{\omega} \mid \cdot \sim \mathcal{N}\left(\boldsymbol{Q}^{-1}\boldsymbol{\ell}, \boldsymbol{Q}^{-1}\right), \tag{A1.10}$$

where

$$\boldsymbol{Q} = \boldsymbol{X}'\Omega_1^{-1}\boldsymbol{X} + \boldsymbol{D}'\Omega_2^{-1}\boldsymbol{D}, \tag{A1.11}$$

$$\boldsymbol{\ell} = \boldsymbol{X}'\Omega_1^{-1}\boldsymbol{y} + \boldsymbol{D}'\Omega_2^{-1}(G_1\boldsymbol{\omega}_0 + \text{vec}(\boldsymbol{\alpha})). \tag{A1.12}$$

To sample $\boldsymbol{\omega}$, compute the Cholesky factor $\boldsymbol{L}$ of $\boldsymbol{Q}$ such that $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}'$, and draw

$$\boldsymbol{\omega} = \boldsymbol{Q}^{-1}\boldsymbol{\ell} + \boldsymbol{L}'^{-1}\boldsymbol{z}, \quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \tag{A1.13}$$

then reshape $\boldsymbol{\omega}$ into a $T \times p$ matrix.

(b) Sample $\boldsymbol{\Sigma}_\omega \mid \boldsymbol{\omega}, \boldsymbol{s}, \boldsymbol{\alpha}$.

Let $\boldsymbol{y}_\omega = (\Delta\boldsymbol{\omega}_2, \ldots, \Delta\boldsymbol{\omega}_T)'$ and $\boldsymbol{x}_\omega = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_{T-1})'$ denote the $T-1 \times p$ matrices of first differences and lagged model weights, respectively. Conditional on the latent regime sequence $\boldsymbol{s}$ and the adjustment coefficients $\boldsymbol{\alpha}_{s_t}$, the error term in the state equation can be written as

$$\boldsymbol{u}_t = \Delta\boldsymbol{\omega}_t - \boldsymbol{\beta}\boldsymbol{\alpha}_{s_t}'\boldsymbol{\omega}_{t-1} - \boldsymbol{\alpha}_{s_t}, \tag{A1.14}$$

where μ_t is the scalar mean term and $\boldsymbol{\alpha}_{s_t}$ varies with the regime index s_t.

Stacking over time gives the residual matrix

$$\boldsymbol{U} = \boldsymbol{y}_\omega - \boldsymbol{x}_\omega\boldsymbol{\beta}\boldsymbol{A}_\alpha' - \boldsymbol{A}_\mu', \tag{A1.15}$$

where $\boldsymbol{A}_\alpha = [\boldsymbol{\alpha}_{s_1}, \ldots, \boldsymbol{\alpha}_{s_T}]$ and $\boldsymbol{A}_\mu = [\boldsymbol{\alpha}_{s_1}, \ldots, \boldsymbol{\alpha}_{s_T}]$ are $p \times T$ matrices.

Given the prior specificed in the main paper, the posterior distribution for $\boldsymbol{\Sigma}_\omega$ is inverse Wishart:

$$\boldsymbol{\Sigma}_\omega \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega} \sim IW(\boldsymbol{S}, \nu_\Sigma), \tag{A1.16}$$

with parameters

$$\boldsymbol{S} = \boldsymbol{U}'\boldsymbol{U} + \nu^{-1}\boldsymbol{\alpha}_{s_t=0}\big(\boldsymbol{\beta}' P_{1/\tau}\boldsymbol{\beta}\big)\boldsymbol{\alpha}'_{s_t=0}, \tag{A1.17}$$

$$\nu_\Sigma = T + 1. \tag{A1.18}$$

(c) Sample $\boldsymbol{\omega}_0 \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\omega}, \boldsymbol{h}, \sigma^2_{\omega_0}$.

The measurement equation (A1.1) can be rewritten as

$$y_t = \omega_{0,t} + \boldsymbol{x}'_t\boldsymbol{\omega}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right) \tag{A1.19}$$

$$y_t - \boldsymbol{x}'_t\boldsymbol{\omega}_t = \omega_{0,t} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right) \tag{A1.20}$$

Define the measurement residual $y^*_t = y_t - \boldsymbol{x}'_t\boldsymbol{\omega}_t$

$$y^*_t = \omega_{0,t} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right) \tag{A1.21}$$

Stack over time

$$\boldsymbol{y}^* = \boldsymbol{\omega}_0 + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \Omega_1) \tag{A1.22}$$

where the intercept sequence, $\boldsymbol{\omega}_0 = (\omega_{0,1}, \ldots, \omega_{0,T})'$, is modeled as a random walk,

$$\omega_{0,t} = \omega_{0,t-1} + u_{0,t}, \qquad u_{0,t} \sim \mathcal{N}(0, \sigma^2_{\omega_0}). \tag{A1.23}$$

Conditional on $\sigma^2_{\omega_0}$, the prior precision matrix is constructed using the first-difference operator

$$\boldsymbol{D} = I_T - \text{diag}(\mathbf{1}_{T-1}, -1),$$

where the first off-dignal are -1, and

$$\boldsymbol{\Omega}_h^{-1} = \text{diag}\left(\frac{1}{V_h}, \frac{1}{\sigma^2_{\omega_0}}, \ldots, \frac{1}{\sigma^2_{\omega_0}}\right),$$

where V_h is the prior variance for the initial state.

The conditional posterior of $\boldsymbol{\omega}_0$ is Gaussian:

$$\boldsymbol{\omega}_0 \mid \cdot \sim \mathcal{N}\left(\boldsymbol{Q}^{-1}\boldsymbol{\ell},\, \boldsymbol{Q}^{-1}\right), \tag{A1.24}$$

where

$$\boldsymbol{Q} = \boldsymbol{D}'\boldsymbol{\Omega}_h^{-1}\boldsymbol{D} + \Omega_1^{-1}, \tag{A1.25}$$

$$\boldsymbol{\ell} = \Omega_1^{-1}\boldsymbol{y}^*. \tag{A1.26}$$

Sampling proceeds via the Cholesky decomposition $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}'$:

$$\boldsymbol{\omega}_0 = \boldsymbol{Q}^{-1}\boldsymbol{\ell} + \boldsymbol{L}'^{-1}\boldsymbol{z}, \qquad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_T).$$

(d) Sample $\sigma^2_{\omega_0} \mid \boldsymbol{\omega}_0$.

Given the random walk prior for $\boldsymbol{\omega}_0$,

$$\omega_{0,t} = \omega_{0,t-1} + u_{0,t}, \qquad u_{0,t} \sim \mathcal{N}(0, \sigma^2_{\omega_0}),$$

the conditional posterior of $\sigma^2_{\omega_0}$ follows from the inverse-gamma conjugacy. Let $d_{\omega_0} = \Delta\boldsymbol{\omega}_0 = (\omega_{0,2} - \omega_{0,1}, \ldots, \omega_{0,T} - \omega_{0,T-1})'$ and compute the sum of squared deviations

$$SSE = \sum_{t=2}^{T} (\omega_{0,t} - \omega_{0,t-1})^2.$$

Assuming the prior $\sigma^2_{\omega_0} \sim IG(v_0/2, s_0/2)$, the posterior is

$$\sigma^2_{\omega_0} \mid \boldsymbol{\omega}_0 \sim IG\left(\frac{v_0 + T - 1}{2},\, \frac{s_0 + SSE}{2}\right). \tag{A1.27}$$

A sample from this distribution is obtained via

$$\sigma^2_{\omega_0} = \frac{1}{\mathrm{G}\left(\frac{v_0+T-1}{2}, \frac{2}{s_0+SSE}\right)}.$$

Where IG denotes inverse-Gamma and G denotes Gamma distributions.

(e) Sample $\boldsymbol{\alpha} \mid \boldsymbol{\omega}, \boldsymbol{s}, \boldsymbol{\Sigma}_\omega$.

Conditional on the latent regime sequence $\boldsymbol{s}$ and the covariance matrix $\boldsymbol{\Sigma}_\omega$, the regime-specific adjustment coefficients $\boldsymbol{\alpha}_{s_t}$ are sampled separately for each regime. Recall the state equation

$$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}_{s_t}(1 + \beta'\boldsymbol{\omega}_{t-1}) + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_\omega\right), \tag{A1.28}$$

where $\boldsymbol{\alpha}_{s_t}$ captures the adjustment strength within regime $s_t \in \{0, 1\}$.

Let $\boldsymbol{y} = (\Delta\boldsymbol{\omega}_1, \ldots, \Delta\boldsymbol{\omega}_T)'$ and define the regressor vector

$$\boldsymbol{c}_t = 1 + \beta'\boldsymbol{\omega}_{t-1}, \quad \text{so that} \quad \boldsymbol{y}_t = \boldsymbol{\alpha}_{s_t}\boldsymbol{c}_t + \boldsymbol{u}_t.$$

Stacking observations yields regime-specific partitions $\{\boldsymbol{y}_t, \boldsymbol{c}_t \mid s_t = 0\}$ and $\{\boldsymbol{y}_t, \boldsymbol{c}_t \mid s_t = 1\}$ corresponding to time indices where $s_t = 0$ and $s_t = 1$, respectively.

The prior has been outlined in the main paper, and thus, for each regime $j \in \{0, 1\}$, the conditional posterior of $\boldsymbol{\alpha}_j$ is Gaussian:

$$\boldsymbol{\alpha}_j \mid \cdot \sim \mathcal{N}\left(\boldsymbol{Q}_j^{-1}\boldsymbol{m}_j,\, \boldsymbol{Q}_j^{-1}\right), \tag{A1.29}$$

where

$$\boldsymbol{Q}_j = \boldsymbol{c}_j'\boldsymbol{c}_j\boldsymbol{\Sigma}_\omega^{-1} + \boldsymbol{P}_j^{-1}, \tag{A1.30}$$

$$\boldsymbol{m}_j = \boldsymbol{\Sigma}_\omega^{-1}\boldsymbol{y}_j'\boldsymbol{c}_j. \tag{A1.31}$$

where $\boldsymbol{P}_j$ is the regime specific prior variance.

Draws are obtained via the Cholesky decomposition $\boldsymbol{Q}_j = \boldsymbol{L}_j\boldsymbol{L}_j'$:

$$\boldsymbol{\alpha}_j = \boldsymbol{Q}_j^{-1}\boldsymbol{m}_j + \boldsymbol{L}_j'^{-1}\boldsymbol{z}_j, \qquad \boldsymbol{z}_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_p).$$

The two regime-specific estimates $\boldsymbol{\alpha}_{s_t=0}$ and $\boldsymbol{\alpha}_{s_t=1}$ are then combined as

$$\boldsymbol{\alpha} = [\boldsymbol{\alpha}_{s_t=0}\ \boldsymbol{\alpha}_{s_t=1}].$$

(f) Sample $s \mid \omega, \alpha, \Sigma_\omega, P$ using the FFBS algorithm.

The latent regime indicator $s_t \in \{0, 1\}$ governs the adjustment coefficient $\boldsymbol{\alpha}_{s_t}$ in the state equation

$$\Delta\boldsymbol{\omega}_t = \boldsymbol{\alpha}_{s_t}(1 + \beta'\boldsymbol{\omega}_{t-1}) + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\omega). \tag{A1.32}$$

Let $\boldsymbol{y}_t = \Delta\boldsymbol{\omega}_t$ and define $c_t = (1 + \beta'\boldsymbol{\omega}_{t-1})$ so that

$$\boldsymbol{y}_t = \boldsymbol{\alpha}_{s_t} c_t + \boldsymbol{u}_t.$$

Conditional on $\boldsymbol{\alpha}_0$, $\boldsymbol{\alpha}_1$, and $\boldsymbol{\Sigma}_\omega$, the conditional likelihood of $\boldsymbol{y}_t$ given $s_t = j$ is

$$p(\boldsymbol{y}_t \mid s_t = j, \boldsymbol{\alpha}_j, \boldsymbol{\Sigma}_\omega) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}_\omega|^{-1/2} \exp\left[-\tfrac{1}{2}(\boldsymbol{y}_t - \boldsymbol{\alpha}_j c_t)'\boldsymbol{\Sigma}_\omega^{-1}(\boldsymbol{y}_t - \boldsymbol{\alpha}_j c_t)\right]. \tag{A1.33}$$

Let P denote the 2×2 transition probability matrix with elements $P_{ij} = \Pr(s_t = j \mid s_{t-1} = i)$. The joint distribution of $(\boldsymbol{y}_{1:T}, \boldsymbol{s}_{1:T})$ is

$$p(\boldsymbol{y}_{1:T}, \boldsymbol{s}_{1:T}) = p(s_1)\, p(\boldsymbol{y}_1 \mid s_1) \prod_{t=2}^{T} p(s_t \mid s_{t-1})\, p(\boldsymbol{y}_t \mid s_t), \tag{A1.34}$$

and the conditional posterior of the latent states is proportional to

$$p(\boldsymbol{s}_{1:T} \mid \boldsymbol{y}_{1:T}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_\omega) \propto p(s_1)\, p(\boldsymbol{y}_1 \mid s_1) \prod_{t=2}^{T} P_{s_{t-1}, s_t}\, p(\boldsymbol{y}_t \mid s_t), \tag{A1.35}$$

where $p(\boldsymbol{y}_t \mid s_t)$ given by (A1.33). Sampling $\boldsymbol{s}_{1:T}$ from (A1.35) is done efficiently via the Forward-Filtering Backward-Sampling (FFBS) algorithm.

Forward filtering: Define the filtered probabilities $\Pr(s_t = j \mid \boldsymbol{y}_{1:t})$ recursively. Starting

with a uniform prior $\Pr(s_1 = j) = 1/2$, the filtering step proceeds as:

$$\Pr(s_t = j \mid \boldsymbol{y}_{1:t}) \propto p(\boldsymbol{y}_t \mid s_t = j) \sum_{i=0}^{1} P_{ij} \Pr(s_{t-1} = i \mid \boldsymbol{y}_{1:t-1}), \tag{A1.36}$$

where normalization ensures that $\sum_j \Pr(s_t = j \mid \boldsymbol{y}_{1:t}) = 1$. To avoid numerical underflow, the recursion is implemented in the log domain:

$$\log a_t(j) = \log p(\boldsymbol{y}_t \mid s_t = j) + \log \sum_{i=0}^{1} \exp\big(\log a_{t-1}(i) + \log P_{ij}\big), \tag{A1.37}$$

followed by normalization $\log a_t(j) \leftarrow \log a_t(j) - \log \sum_j \exp(\log a_t(j))$.

Backward sampling: The backward step samples $\boldsymbol{s}_{1:T}$ recursively from the smoothed posterior $p(\boldsymbol{s}_{1:T} \mid \boldsymbol{y}_{1:T})$. Starting with the last period,

$$\Pr(s_T = j \mid \boldsymbol{y}_{1:T}) \propto \exp(\log a_T(j)),$$

we draw s_T accordingly. Then for $t = T-1, \ldots, 1$, we sample s_t from

$$\Pr(s_t = j \mid s_{t+1}, \boldsymbol{y}_{1:T}) \propto \exp\left[\log a_t(j) + \log P_{j,\, s_{t+1}}\right], \tag{A1.38}$$

which is normalized so that the probabilities sum to one.

The FFBS procedure produces a full draw of the regime sequence $\boldsymbol{s}$ from its exact conditional posterior distribution.

(g) Sample $P \mid \boldsymbol{s}$.

Given the latent regime sequence $\boldsymbol{s}$, the transition counts are updated by

$$N_{gh} = \sum_{t=2}^{T} \mathbb{I}(s_{t-1} = g, s_t = h), \qquad g, h \in \{0, 1\},$$

and the posterior of the transition probability matrix P is Dirichlet:

$$P_h \mid \boldsymbol{s} \sim \text{Dirichlet}(N_{g1} + \bar{P}_{g1},\, N_{g2} + \bar{P}_{g2}), \tag{A1.39}$$

where $\bar{P}$ denotes the prior counts.

(h) Sample $\boldsymbol{h} \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\omega}, \boldsymbol{\omega}_0, \boldsymbol{r}, \mu_h, \phi_h, \sigma_h^2$.

From equation (A1.3), the variance of error term follows a stochastic volatility process,

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + e_t, \qquad e_t \sim N\left(0, \sigma_h^2\right), \tag{A1.40}$$

Given all other variables, the joint density of $\boldsymbol{h} = (h_1, \ldots, h_T)'$ follows a multivariate normal distribution. Since the latent volatility process exhibits a first-order autoregressive structure, its distribution can be represented using a tridiagonal precision matrix. The matrix form of the AR(1) process is

$$\boldsymbol{D}_{\phi_h}\boldsymbol{h} = \gamma + X_1\phi_h h_0 + \boldsymbol{e} \tag{A1.41}$$

where $\boldsymbol{D}_{\phi_h}$ is a first difference matrix as similarly as defined in equation (A1.23) where the -1 entries are replaced by $-\phi_h$, and $X_1 = (1, 0, \ldots, 0)'$. As in Omori et al. (2007), the measurement equation in matrix form can be rewritten as

$$\tilde{\boldsymbol{y}} = \boldsymbol{h} + \tilde{\boldsymbol{z}} \tag{A1.42}$$

where $\tilde{\boldsymbol{y}}$ is the stacked vector of $\log\left((y_t - \omega_{0,t} - \boldsymbol{x}_t'\boldsymbol{\omega}_t)^2 + c\right)$ over time, and $\tilde{\boldsymbol{z}} \sim N(\boldsymbol{m}_r, \boldsymbol{\Sigma}_r)$, where $\tilde{\boldsymbol{z}} = (\log(z_1^2), \ldots, \log(z_T^2))$ with $z_t \sim N(0,1)$. An auxiliary mixture normal (described in the next step) is used, where $\boldsymbol{m}_r = (m_{r_1}, \ldots, m_{r_T})$ and $\boldsymbol{\Sigma}_r = (\sigma_{r_1}^2, \ldots, \sigma_{r_T}^2)$. Therefore, the posterior for $\boldsymbol{h}$ is

$$\boldsymbol{h}|\cdot \sim N(\boldsymbol{Q}_h^{-1}\boldsymbol{\ell}_h, \boldsymbol{Q}_h^{-1})$$

where $\boldsymbol{Q}_h^{-1} = \boldsymbol{\Sigma}_r^{-1} + \boldsymbol{D}'\sigma_h^2\boldsymbol{D}$, and $\boldsymbol{\ell}_h = \boldsymbol{\Sigma}_r(\tilde{\boldsymbol{y}} - \boldsymbol{m}_r) + \boldsymbol{D}'\sigma_h^2\left(\gamma + X_1 \cdot \phi_h \cdot h_0\right)$. Finally, the initial value can be sampled from $h_0 \mid h_1, \mu_h, \phi_h, \sigma_h^2 \sim N\left(\mu_h + \phi_h\left(h_1 - \mu_h\right), \sigma_h^2\right)$.

(i) Sample $\mu_h, \phi_h \mid \boldsymbol{h}, \sigma_h^2$, then sample $\sigma_h^2 \mid \boldsymbol{h}, \mu_h, \phi_h$.

For sampling $\mu_h, \phi_h, \sigma_h^2$, it can be very helpful to rewrite the mean reverting AR(1) process of the stochastic volatility as a conditional regression model as following

$$h_t = \gamma + \phi_h h_{t-1} + e_t \tag{A1.43}$$

where $\gamma = \mu_h(1 - \phi_h)$. Recall $\boldsymbol{h}$ was defined as $\boldsymbol{h} = (h_1, h_2, \ldots, h_T)'$, and stack the equations over time into a matrix representation

$$\boldsymbol{h} = X_h\beta_h + e, \qquad\qquad e \sim N_T(0, \sigma_h^2 I_T) \tag{A1.44}$$

where $X_h = (1_T, \boldsymbol{h}_{0:T-1})$ and $\beta_h = (\gamma, \phi_h)'$. We define flat priors as $p(\beta_h) \sim \mathcal{N}_2(0, \text{diag}(b_h^1, b_h^2))$ by imposing relative large numbers on the variance of the prior probability. This yields posterior distribution

$$\beta_h|\sigma_h^2, \boldsymbol{h} \sim \mathcal{N}_2(\boldsymbol{Q}_{\beta_h}^{-1}\boldsymbol{\ell}_{\beta_h}, \boldsymbol{Q}_{\beta_h}^{-1})$$

where $\boldsymbol{Q}_{\beta_h} = X_h'\sigma_h^{-2}X_h + \text{diag}(b_h^1, b_h^2)^{-1}$, and $\boldsymbol{\ell}_{\beta_h} = X_h'\sigma_h^{-2}\boldsymbol{h}$. Then we can recover ϕ_h and μ_h as how β_h was defined via $\phi_h = \beta_h(2)$ and $\mu_h = \beta_h(1)/(1 - \phi_h)$.

More standard for σ_h^2, where a standard prior is given $\mathcal{IG}(v_h/2, s_h/2)$, thus the posterior can be derived

$$\sigma_h^2|X_h, \boldsymbol{h}, \beta_h \sim \mathcal{IG}\left(\bar{v}_h/2, \bar{s}_h/2\right)$$

where $\bar{v}_h = v_h + T$, and $\bar{s}_h = s_h + (X_h\beta_h - \boldsymbol{h})'(X_h\beta_h - \boldsymbol{h})$

(j) Sample $\boldsymbol{r} \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\omega}, \omega_0, \boldsymbol{h}$.

We proceed exactly as Omori et al. (2007), who approximate the log chi-squared distribution with a mixture of ten normal distributions. Observing that $\tilde{\boldsymbol{y}} - \boldsymbol{h} = \tilde{\boldsymbol{z}}$ from Equation (A1.42) with $\tilde{\boldsymbol{z}} \sim N(\boldsymbol{m}_r, \boldsymbol{\Sigma}_r)$, the posterior probabilities can be easily derived for each $\mathbb{P}(r_t = k \mid \cdot)$ for $k \in \{1, \ldots, 10\}$ and $t \in \{1, \ldots, T\}$ according to

$$\mathbb{P}(r_t = k \mid \cdot) \propto \mathbb{P}(r_t = k) \frac{1}{\sigma_k} \exp\left\{ -\frac{\left(\log(z_t^2) - m_k\right)^2}{2\sigma_k^2} \right\}$$

where $\mathbb{P}(r_t = k)$ denotes the prior mixture weight for the k th component in the Gaussian mixture approximation, and $(\cdot)$ includes all data and model parameters except r_t. In our implementation, sampling r_t is performed independently for each t and is computationally straightforward due to the use of inverse transform sampling.

A1.2 IMS Sampling

I set the hyperparameters to $\alpha_0^{\pi} = 15$, $\gamma = 15$, and $\lambda = 1$, and choose $K = 30$ to ensure that the number of regimes is sufficiently large to capture all potentially relevant states while allowing the posterior to automatically empty redundant ones. The estimation algorithm proceeds as follows.

(1) Sample regimes s via Forward–Backward recursion. Given the model specified in equations (2.20)–(2.24), the posterior distribution of the latent regime sequence $\boldsymbol{s}_{1:T} = (s_1, \ldots, s_T)$ is derived as

$$p(\boldsymbol{s}_{1:T} \mid y_{1:T}, \Theta) \propto p(y_{1:T} \mid \boldsymbol{s}_{1:T}, \Theta)\, p(\boldsymbol{s}_{1:T} \mid \Theta),$$

where $\Theta = \{\boldsymbol{\omega}_{1:S}, \boldsymbol{\pi}_{1:S}, \omega_{0,t}, \sigma_t^2\}$ denotes the set of model parameters. The conditional independence structure of the hidden Markov model implies

$$p(y_{1:T} \mid \boldsymbol{s}_{1:T}, \Theta) = \prod_{t=1}^{T} p(y_t \mid s_t, \Theta), \quad p(\boldsymbol{s}_{1:T} \mid \Theta) = p(s_1) \prod_{t=2}^{T} \pi_{s_{t-1}, s_t}.$$

Direct sampling from this joint posterior is computationally infeasible, so the *Forward–Filtering Backward–Sampling* (FFBS) algorithm is employed to efficiently sample from $p(\boldsymbol{s}_{1:T} \mid y_{1:T}, \Theta)$.

Forward recursion. Define the forward filtering probability

$$\alpha_t(j) = p(y_{1:t}, s_t = j \mid \Theta),$$

which satisfies the recursion

$$\alpha_t(j) = p(y_t \mid s_t = j, \Theta) \sum_{i=1}^{S} \alpha_{t-1}(i)\, \pi_{ij}, \tag{A1.45}$$

initialized with

$$\alpha_1(j) = p(s_1 = j)\, p(y_1 \mid s_1 = j, \Theta).$$

The observation likelihood for regime j is given by

$$p(y_t \mid s_t = j, \Theta) = \mathcal{N}\left(y_t;\, \omega_{0,t} + \boldsymbol{x}_t' \boldsymbol{\omega}_j, \sigma_t^2\right).$$

The recursion in (A1.45) integrates over all possible previous states, accumulating the total probability of observing $y_{1:t}$ and ending in regime j.

Backward sampling. After completing the forward pass, the regime sequence $\boldsymbol{s}_{1:T}$ is drawn backward in time using the conditional distributions

$$p(s_T = j \mid y_{1:T}, \Theta) \propto \alpha_T(j),$$
$$p(s_t = i \mid s_{t+1} = j, y_{1:T}, \Theta) \propto \alpha_t(i)\, \pi_{ij}, \quad t = T-1, \ldots, 1.$$

This backward step reconstructs one posterior draw of $\boldsymbol{s}_{1:T}$ conditional on all parameters and data.

(2) Sample regime-specific weights $\boldsymbol{w}_{s_t}$.

For each regime s, conditional on $\{y_t : s_t = s\}$:

$$p(\boldsymbol{w}_s \mid y_t, s_t = s) \propto \Big[\prod_{t:s_t=s} \mathcal{N}(y_t \mid \boldsymbol{x}_t' \boldsymbol{w}_s, e^{h_t}) \Big] \text{Dirichlet}(\lambda \mathbf{1}_K).$$

The draw is obtained using a Metropolis–Hastings step restricted to the simplex. This updating step differs fundamentally from the Dirichlet process mixture approach of Jin et al. (2022), where the weighting vectors are conditionally independent of y_t given the state indicators, and are updated via a Dirichlet posterior. In contrast, the present framework samples $\boldsymbol{w}_s$ directly from the likelihood, allowing the weights to adjust to the observed data within each regime.

(3) Sample transition probabilities and global weights.

Let $n_{ij} = \sum_{t=2}^{T} \mathbb{I}(s_{t-1} = i, s_t = j)$. Then I have the following

$$\boldsymbol{\pi}_i \sim \text{Dirichlet}(\alpha_0^{\pi} \boldsymbol{\beta}^{\pi} + n_{i,1:K}), \tag{A1.46}$$

$$\boldsymbol{\beta}^{\pi} \sim \text{Dirichlet}\left(\frac{\gamma}{K} + \alpha_0^{\pi} \sum_i \boldsymbol{\pi}_i \right). \tag{A1.47}$$

For the sticky specification, one can impose that $\boldsymbol{\beta}^{\pi}$ is a self looping transition matrix.

(4) Sample $\{h_t\}$ via the mixture SV smoother (same as VECM case).

(5) Sample $(\mu_h, \phi_h, \sigma_h^2)$ from their conjugate conditionals (same as VECM case).

A2 Application to Exchange Rate Forecasting

Forecasting exchange rates is of critical importance to policymakers, investors, and international traders. The seminal study by Meese and Rogoff (1983), known as the Meese–Rogoff puzzle, showed that no structural or econometric model could consistently outperform a random walk in out-of-sample forecasts. Subsequent research has largely concentrated on point forecasting, and three decades later, Rossi (2013) reaffirmed this conclusion through an extensive empirical analysis. These two influential contributions have

shaped much of the literature that followed. Despite considerable advances in Bayesian predictive synthesis, the use of model combination methods in exchange rate forecasting remains relatively underexplored. In the frequentist domain, Aristidou et al. (2022b,a) adopt meta-model approaches to address model uncertainty.

A2.0.1 Data and Base Models

Following Fan et al. (2025), I focus on the monthly exchange rate between the British pound (GBP) and the United States dollar (USD). The primary data sources are FRED (Federal Reserve Economic Data), the IMF (International Monetary Fund), the OECD (Organisation for Economic Co-operation and Development), and the Bank of England. The analysis uses the end-of-month nominal GBP/USD exchange rate. The sample period spans from January 1990 to June 2017, yielding a total of $T = 330$ observations. Further details on the data are documented in the online appendix of Fan et al. (2025).

The base forecasting models follow the literature and consider 3 simplified and modified specifications:

$$S_{t+h} = \gamma_0 + \gamma_1 S_t + \gamma_2 (r_t - r_t^*) + \nu_{t+h}, \tag{A2.1}$$

$$S_{t+h} = \gamma_0 + \gamma_1 S_t + \gamma_2 (p_t - p_t^* - S_t) + \nu_{t+h}, \tag{A2.2}$$

$$S_{t+h} = \gamma_0 + \gamma_1 S_t + \gamma_2 (p_t^{oil} - p_{t-1}^{oil}) + \nu_{t+h}, \tag{A2.3}$$

where S_t is the log exchange rate between GBP and USD, and r_t and p_t denote the log interest rate and log price level for the UK, respectively. Variables with a superscript "$*$" refer to the corresponding U.S. counterparts. The term p_t^{oil} represents the log price of crude oil at time t. The first two equations correspond to the Interest Rate Parity and Purchasing Power Parity conditions, as summarized by Rossi (2013). In addition, the last equation represent commodity-based models for oil, inspired by Ferraro et al. (2015).

Same as inflation, all base models are estimated in a rolling-window framework with a window length of 100 months, allowing the coefficients to evolve over time and capture potential structural changes in the data. The initial training sample for the base models is set to $T_0 = 80 - h$, leading to around 21 years (250 months) of out-of-sample point

forecasts for all periods being generated for synthesis learning and evaluation.

A2.0.2 Predictive Synthesis

To combine the point forecasts from each base model, the synthesis function follows the same structure as in equation (3.4) for the inflation forecasting application:

$$S_{t+h} = \omega_{0,t} + \boldsymbol{x}_t^{(h)\prime}\boldsymbol{\omega}_t + \epsilon_{t+h}, \tag{A2.4}$$

where $\boldsymbol{x}_t^{(h)}$ is the vector of predictive values $\hat{S}_{t+h}$ obtained from Model (A2.1), Model (A2.2), Model (A2.3).

The estimation, prediction, and evaluation procedures are identical to those used for inflation forecasting, as described in Section 6.0.2. I set $T_1 = 234$, which provides 8 years (96 months) of data for evaluating the performance of the synthesis function, and the conclusion does not change if I choose a different evaluating size. In this application, I consider forecast horizons of $h = 1, 2, 3, 6, 12, 18, 24, 30$ and 36 months ahead. The number of evaluation observations decreases with h according to $T - h - T_1 + 1$.

A2.0.3 Out-of-Sample Forecasting

The out-of-sample performance is evaluated based on the metrics mentioned in section 3.4. Here I first report the LPDR of one-month period ahead forecasting in the plot.

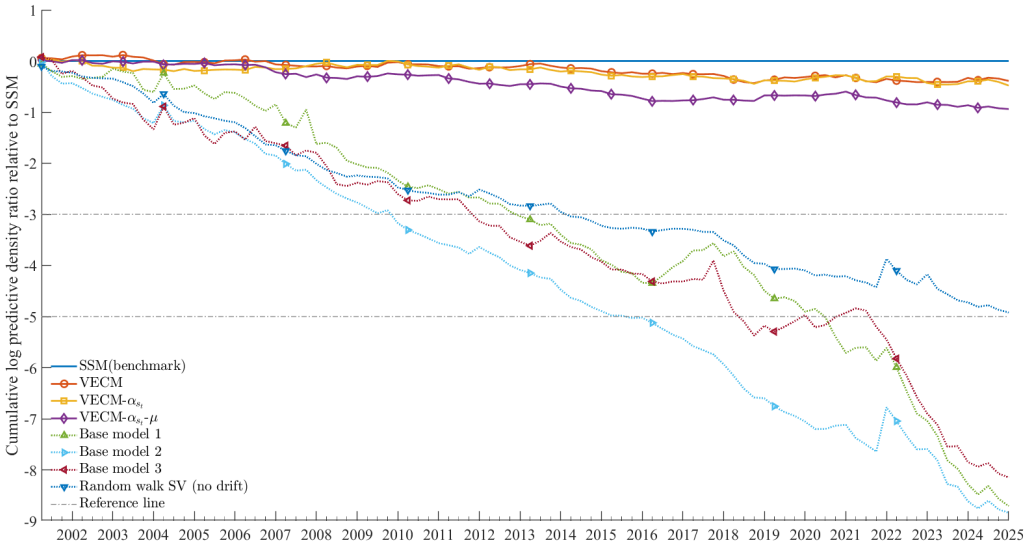


Figure 15: Comparative Forecast Performance for 1-quarter-ahead Forecasting Using Log Predictive Density Ratio (LPDR) with reference: SSM. Out-of-sample period: 2009/M7 - 2017/M6, in total 96 months (8 years).

The results may at first appear surprising, as the proposed model does not deliver substantial improvements over the benchmark SSM, and the out-of-sample performance is quite similar. This outcome, however, is not unexpected, since the cointegrating relationships among the base models are not guaranteed to hold in all scenarios. Rather, the finding underscores an important property of the framework: when the VECM structure is absent, the proposed model naturally collapses to a simple BPS specification, thereby providing a lower bound for performance. The exchange rate application offers a clear illustration of this feature. This point is revisited later, with supporting evidence from the posterior inference following the presentation of the full forecasting results.

Table 8 reports the evaluation metrics across forecast horizons. Up to $h = 6$, differences among the four pooling methods are negligible, and all pooling methods outperform the base models while remaining comparable to RW, with higher log scores. From $h = 12$ onward, the proposed VECM specifications, particularly VECM-$\boldsymbol{\alpha}_{s_t}$, consistently dominate the benchmark SSM across all metrics. RMSFE improves notably at $h = 12$ and $h = 18$, and the LPDR exceeds thresholds of 3 and 5. Unlike the SSM, which does not consistently improve on RW in terms of RMSFE, the VECM achieves clear gains. Beyond

$h = 6$, the LPDR of the base models and RW often surpasses that of the pooled methods, reflecting the same mechanism observed in inflation forecasting: the estimated error variance of the pooled forecasts is substantially smaller than that of the individual base models. At longer horizons, this can lead to predictive likelihoods that underestimate the accumulated uncertainty from outliers.

A2.1 Investigation

To further investigate why the VECM begins to outperform the SSM after $h = 6$, I analyze the posterior inference for the cases $h = 1$ and $h = 12$. In particular, attention is given to the posterior aggregate weights and the posterior probability of being in the recovery state, as illustrated in Figure 16. Together, these results highlight the mechanism through

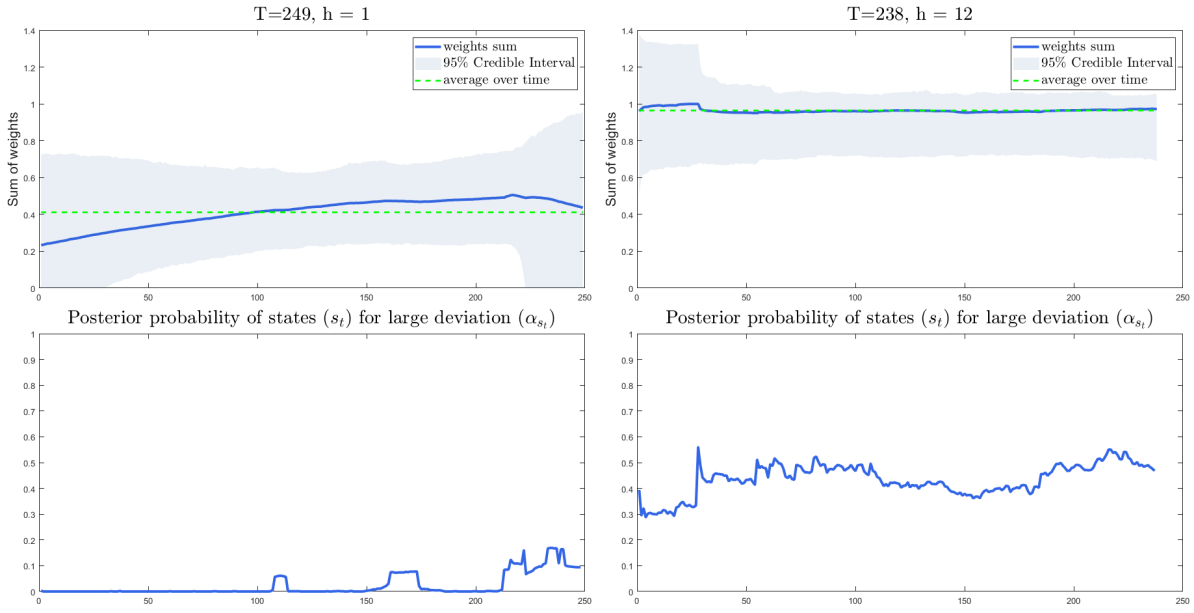


Figure 16: Posterior analysis of the VECM-$\boldsymbol{\alpha}_{s_t}$ specification for one-step ($h = 1$) and twelve-step ($h = 12$) ahead forecasts. The left column shows results for $h = 1$, while the right column corresponds to $h = 12$. The top panels display the sum of weights with 95% credible intervals and their time averages. The bottom panels report the posterior probability of the large-deviation state (recovery state) α_{s_t}.

which the VECM specification gains an advantage over the SSM at longer horizons. At short horizons, the aggregate weights are diffuse, with no long-run equilibrium to enforce, and the recovery state is rarely activated. As a result, $\boldsymbol{\alpha}$ remains predominantly in the shrinkage state, causing the VECM structure to collapse and behave similarly to

Table 8: Exchange Rate Forecasting Performance Comparison Across Different Horizons

Model	Metric	Forecast Horizon (h months)								
		$h=1$	$h=2$	$h=3$	$h=6$	$h=12$	$h=18$	$h=24$	$h=30$	$h=36$
SSM	RMSFE	0.0253	0.0333	0.0396	0.0579	0.0807	0.0930	0.1083	0.1227	0.1208
	(CRPS)	(0.0141)	(0.0186)	(0.0229)	(0.0344)	(0.0465)	(0.0472)	(0.0498)	(0.0485)	(0.0411)
	{LPDR}	{0}	{0}	{0}	{0}	{0}	{0}	{0}	{0}	{0}
VECM	RMSFE	0.0253	0.0334	0.0394	0.0580	0.0806	**0.0927**	**0.1078**	0.1228	0.1207
	(CRPS)	(0.0141)	(0.0187)	(0.0227)	(0.0345)	(0.0464)	(0.0470)	(**0.0494**)	(0.0484)	(0.0410)
	{LPDR}	{-0.39}	{0.48}	{0.62}	{0.52}	{2.71}	{3.59}	{4.27}	{10.26}	{10.61}
VECM-MS	RMSFE	0.0252	0.0334	0.0394	0.0580	**0.0803***	**0.0927***	0.1079	**0.1224**	**0.1204**
	(CRPS)	(0.0140)	(0.0186)	(0.0227)	(0.0344)	(**0.0460**)	(**0.0469**)	(0.0496)	(0.0483)	(**0.0407**)
	{LPDR}	{-0.48}	{0.37}	{0.94}	{-0.34}	{**4.82**}	{**9.21**}	{**15.67**}	{**40.13**}	{**23.01**}
VECM-MS-μ	RMSFE	0.0254	0.0334	0.0395	0.0581	0.0807	0.0928	0.1079	**0.1224**	0.1207
	(CRPS)	(0.0141)	(0.0187)	(0.0228)	(0.0345)	(0.0465)	(0.0470)	(0.0495)	(**0.0481**)	(0.0409)
	{LPDR}	{-0.94}	{0.17}	{0.32}	{1.6}	{ 2.31}	{4.28}	{14.15}	{17.95}	{15.18}
Base1: IRP	RMSFE	0.0267	0.0388	0.0532	0.1003	0.1094	0.0981	0.1024	0.1241	0.1235
	(CRPS)	(0.0139)	(0.0215)	(0.0290)	(0.0531)	(0.0690)	(0.0667)	(0.0898)	(0.1228)	(0.1715)
	{LPDR}	{-8.71}	{-15.12}	{-11.06}	{+}	{+}	{+}	{+}	{+}	{+}
Base2: PPP	RMSFE	0.0258	0.0358	0.0464	0.0810	0.0866	0.0904	0.1075	0.1071	0.1074
	(CRPS)	(0.0139)	(0.0208)	(0.0276)	(0.0520)	(0.0747)	(0.0725)	(0.0919)	(0.4704)	(0.0830)
	{LPDR}	{-8.84}	{-15.76}	{-12.50}	{+}	{+}	{+}	{+}	{+}	{+}
Base3: Oil	RMSFE	0.0271	0.0403	0.0555	0.1030	0.1082	0.1025	0.1157	0.1150	0.1174
	(CRPS)	(0.0139)	(0.0222)	(0.0303)	(0.0560)	(0.0668)	(0.0691)	(0.0852)	(0.0960)	(0.1071)
	{LPDR}	{-8.15}	{-19.54}	{-19.34}	{+}	{+}	{+}	{+}	{+}	{+}
RW	RMSFE	0.0253	0.0334	0.0395	0.0579	0.0809	0.0929	0.1081	0.1230	0.1213
	(CRPS)	(0.0135)	(0.0190)	(0.0230)	(0.0335)	(0.0460)	(0.0554)	(0.0695)	(0.0782)	(0.0812)
	{LPDR}	{-4.92}	{-11.72}	{-7.42}	{+}	{+}	{+}	{+}	{+}	{+}

Notes: RMSFE = Root Mean Square Forecast Error; CRPS = Continuous Ranked Probability Score; LPDR = Log Predictive Density Ratio relative to benchmark SSM. Lower RMSFE and CRPS values indicate superior forecasting performance, while higher LPDR values indicate better performance relative to the benchmark SSM. Bold values denote the optimal performance within each column among model assembly methods (top panel) and baseline models (bottom panel). Statistical significance is assessed using the Diebold-Mariano test with mean squared error loss function: ** denotes significance at the 5% level and * denotes significance at the 10% level, indicating significantly better performance than the benchmark SSM. '+' indicates dominating LPDR due to different model specifications. RW is a random walk with stochastic volatility without drift, specifically $S_{t+h} = S_t + \nu_{t+h}$, where ν_{t+h} follows stochastic volatility.

the benchmark SSM. By contrast, at longer horizons, the aggregate weights concentrate around the long-run equilibrium, and the recovery state becomes more prominent. This allows the VECM to capture short-term deviations from the long-run equilibrium and thereby improve forecast accuracy.

Another reason why the performance of the VECM may improve for horizons $h > 6$ can be traced back to its fundamental definition. The VECM framework is designed to capture both short-run dynamics and long-run equilibrium relationships among variables. At longer horizons, the long-run equilibrium component becomes more influential, allowing the model to better exploit the cointegration structure and thereby generate more accurate forecasts.