An Infinite Hidden Markov Model with GARCH for Short-Term Interest Rates

Li, Chenxing and Yang, Qiao

Center for Economics, Finance and Management Studies, Hunan University, School of Entrepreneurship and Management, ShanghaiTech University

4 January 2025

An Infinite Hidden Markov Model with GARCH for Short-Term Interest Rates *

Chenxing Li†, Qiao Yang‡

Janurary 2025

Abstract

This paper introduces a novel Bayesian time series model that combines the nonparametric features of an infinite hidden Markov model with the volatility persistence captured by the GARCH framework, to effectively model and forecast short-term interest rates. When applied to US 3-month Treasury bill rates, the GARCH-IHMM reveals both structural and persistent changes in volatility, thereby enhancing the accuracy of density forecasts compared to existing benchmark models. Out-of-sample evaluations demonstrate the superior performance of our model in density forecasts and in capturing volatility dynamics due to its adaptivity to different macroeconomic environments.

1 Introduction

Modeling and forecasting short-term interest rates are crucial in finance and economics, serving as key inputs for bond pricing, risk management, and macroeconomic forecasting (Gurkaynak et al., 2007; Duffee, 2013; Filipović, 2009). The time series dynamics of short-term rates have been extensively modeled. However, short-term interest rates are less suitable for traditional modeling approaches, as they are often

*Li is grateful for financial support from the Philosophy and Social Science Foundation of Hunan Province, China (Project No. 23YBA031). Yang is grateful for financial support from the start-up fund of ShanghaiTech and the Young Scientists Fund of NSFC (Project 72103137).

†Center for Economics, Finance and Management Studies, Hunan University, China. Email: lichenxing@hnu.edu.cn

‡School of Entrepreneurship and Management, ShanghaiTech University, China. Email: yangqiao@shanghaitech.edu.cn

subject to both transitory and persistent changes in the macroeconomic environment. In this paper, we integrate a generalized autoregressive conditional heteroskedasticity (GARCH) model with an infinite hidden Markov model (IHMM), allowing the GARCH component to capture persistent changes in volatility, and the IHMM accounts for Markovian changes in unknown conditional distribution over time, which is crucial for capturing potential complex economic shocks. Applying our model to the US 3-month Treasury bill rates, our model provides significant improvements in density forecasts compared to competitive benchmarks.

Short-term interest rates experience high and volatile periods and low and stable periods, suggesting switches among different states (Figure 1). Consequently, Markov switching (MS) models have been extensively applied for those rates. Parametric examples include Hamilton (1988), Albert and Chib (1993), Garcia and Perron (1996), and Pesaran et al. (2006). Maheu and Yang (2016) were the first to employ the Bayesian nonparametric MS extension, IHMM, to model and forecast short-term interest rates. They documented that more than two states are needed to continually capture sophisticated dynamics resulting from the ongoing monetary and economic shocks.

Meanwhile, much of the research aims to combine volatility models with MS. Early attempts include Cai (1994), Gray (1996), Durham (2003) and Hou and Suardi (2011), showing that the persistent volatility captured by GARCH is an important feature in short-term interest rates. And recent work by Jin et al. (2022) investigates model dynamics via model combination forecasting, documenting that both GARCH and IHMM play a key role in capturing volatility changes in short-term T-bill rates.

From the literature, it is evident that regime switching and volatility persistence are both essential for capturing short-rate dynamics over time, influencing the conditional mean, variance, and probably higher moments. Therefore, the existing application of the standalone IHMM model by Maheu and Yang (2016) is insufficient, which motivates the design of a new model capable of capturing both regime switches and persistent changes. This paper proposes a novel Bayesian semiparametric GARCH-infinite hidden Markov model (GARCH-IHMM) to capture the dynamics of the short-term interest rates. We allow the GARCH model to capture the persistent volatility changes, while the IHMM governs Markovian changes in the shape of conditional distributions.

Compared to the existing literature on time series modeling of interest rates, our model advances in several respects. With respect to Gray (1996) and Durham (2003), our model learns the number of latent states rather than pre-selecting it subjectively, and we specify MS and GARCH in a multiplicative form instead of making GARCH parameters state-dependent, which is arguably more restrictive and much harder to estimate. With respect to Maheu and Yang (2016), we integrate a GARCH component to capture the volatility persistence ignored by the IHMM. And with respect to Jin et al. (2022), we introduce a unified specification to achieve a similar goal with a simpler estimation at a faster speed.

Our paper also contributes to the literature on model specifications. The IHMM has been effectively used in various applications, and our model further enriches existing works.[1] Dufays (2016) and Shi and Song (2016) incorporate IHMM into GARCH by allowing the GARCH parameters to be state-dependent, whereas we let IHMM and GARCH contribute to the conditional volatility in a multiplicative form. Similar work by Jensen and Maheu (2013), Maheu and Shamsi Zamenjani (2021) and Li and Maheu (2023) attempts to combine various Bayesian nonparametric models with multivariate GARCH, while we investigate interest rates in a univariate setting.[2]

This paper is organized as follows. Section 2 illustrates the specification of the proposed GARCH-IHMM and the benchmark models. Section 3 illustrates the algorithms used to estimate and forecast the GARCH-IHMM. Section 4 describes the data. Section 5 discusses the in-sample estimation and out-of-sample forecast results. And Section 6 concludes.

2 Model Specification

Let r_t be the short rate at time t. The proposed univariate GARCH-IHMM model is specified as

$$r_t = \mu_{s_t} + \rho r_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_{s_t} h_t), \tag{1a}$$

$$h_t = \omega_0 + \omega_1 (r_{t-1} - \rho r_{t-2} - \omega_3)^2 + \omega_2 h_{t-1}, \tag{1b}$$

$$s_t | s_{t-1} \sim cat(\Pi_{s_{t-1}}),\ \Gamma \sim Stick1(\eta),\ \Pi \sim Stick2(\alpha, \Gamma), \tag{1c}$$

where $\Theta = \{\Theta_{s_t}\}_{s_t=1}^{\infty} = \left\{\mu_{s_t}, \sigma^2_{s_t}\right\}_{s_t=1}^{\infty}$ is the set of state-dependent parameters with an infinite number of states, and h_t is the GARCH volatility that is designated for smooth changes in conditional variance.

Equation (1b) is an asymmetric GARCH structure, where $\theta_h = \{\rho, \omega_0, \omega_1, \omega_2, \omega_3\}$ is the set of GARCH parameters. Maheu and Yang (2016) finds that the autoregressive parameter ρ is likely non-state-dependent[3], so we categorize it as a GARCH

[1] Inflation dynamics (Song, 2014; Jochmann, 2015), short-term interest rates (Maheu and Yang, 2016), realized volatility (Jin and Maheu, 2016; Jin et al., 2019; Luo et al., 2022), macroeconomic forecasting (Hou, 2017; Yang, 2019), and model combination (Jin et al., 2022).

[2] Other relevant works include Jensen and Maheu (2010) and Li et al. (2024), which attempt to jointly model stochastic volatility with DPM and IHMM, respectively.

[3] We confirm this finding in a separate test, which is not included in this paper.

parameter, which greatly simplifies the estimation. Unlike conventional GARCH models, the feedback term $(r_{t-1} - \rho r_{t-2} - \omega_3)^2$ involves an additional parameter ω_3, which captures the potential leverage effect. For example, when $\omega_3 > \mu_{s_{t-1}}$, (1b) behaves similarly to the GJR-GARCH model.

Formula (1c) represents the Bayesian nonparametric IHMM component, which employs a hierarchical Dirichlet process prior. $\Gamma = (\gamma_1, \gamma_2, \dots)'$ is a probability vector, $\Pi_{s_{t-1}}$ is the s_{t-1}th row of the transition matrix Π, η and α are both concentration parameters for the hierarchical Dirichlet process. $Stick1(\eta)$ and $Stick2(\alpha, \Gamma)$ are stick-breaking processes that construct the hierarchical Dirichlet process and the IHMM (Sethuraman, 1994; Maheu and Yang, 2016), such that

$$\gamma_k = \tilde{\gamma}_k \prod_{l=1}^{k-1} (1 - \tilde{\gamma}_l), \qquad \tilde{\gamma}_k \sim Beta(1, \eta), \tag{2a}$$

$$\pi_{jk} = \tilde{\pi}_{jk} \prod_{l=1}^{k-1} (1 - \tilde{\pi}_{jl}), \qquad \tilde{\pi}_{jk} \sim Beta\left(\alpha\gamma_k, \alpha\left(1 - \sum_{l=1}^{k} \gamma_l\right)\right). \tag{2b}$$

The IHMM can be seen as a Bayesian nonparametric extension to the parametric MS model, whose number of states is finite and predefined while the IHMM allows the number of latent states to be learned from the data. The IHMM can also be seen as a dynamic extension of the Bayesian nonparametric DPM model whose mixture weight is static, wherein IHMM allows the mixture to change over time. As a type of mixture model, mixing over the conditional mean μ_{s_t} can generate skewness, and mixing over the conditional variance $\sigma^2_{s_t}$ can generate kurtosis. Combined with the GARCH component, which is adapted to the smooth change in conditional volatility, the proposed new model is able to approximate both the unknown shape and the unknown period-by-period evolution of the conditional distribution.

The base measure of the IHMM component is essentially the prior of the state-dependent parameters $\mu_{s_t} \sim N(b_0, B_0)$ and $\sigma^{-2}_{s_t} \sim Gamma\left(\frac{v_0}{2}, \frac{s_0}{2}\right)$. For non-state dependent parameters, we assume $\theta_h = (\omega_0, \omega_1, \omega_2, \omega_3, \rho)' \sim N(0, I)$. We further apply hierarchical priors to the base measure hyperparameters for robust estimation and forecast performance. (Song, 2014; Maheu and Yang, 2016; Li et al., 2024) Moreover, the concentration hyperparameters of the hierarchical Dirichlet prior can be estimated with Gamma hyperpriors. The hyperpriors and hierarchical priors are set as follows:

$$\eta \sim Gamma(5, 5), \qquad \alpha \sim Gamma(5, 5),$$
$$b_0 \sim N(0, 1), \quad B_0 \sim IG\left(\frac{3}{2}, \frac{1}{2}\right), \quad v_0 \sim Exp\left(\frac{1}{g_0}\right), \quad s_0 \sim Gamma(c_0, d_0).$$

The hierarchical priors for v_0 and s_0 are set to aim that $\mathrm{E}(\sigma^2_{s_t}) = \frac{v_0}{s_0} = 1$ for improved identification between the state-dependent volatility ($\sigma^2_{s_t}$) and the GARCH volatility (h_t). In this case, the $\sigma^2_{s_t}$ serves as a multiplier around 1 to scale or shrink h_t. One way to achieve this is to set $\frac{E(v_o)}{E(s_0)} = 1$ with linear approximation when $\mathrm{Var}(s_0)$ is small. We set $c_0 = d_0 = 6$ and $g_0 = 1$ so that $\frac{1}{g_0} = \frac{c_0}{d_0}$.

To compare the performance of the proposed new model, we consider the following benchmark models.

GARCH-DPM:

$$\begin{aligned} r_t &= \mu_{s_t} + \rho r_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_{s_t} h_t), \\ h_t &= \omega_0 + \omega_1 (r_{t-1} - \rho r_{t-2} - \omega_3)^2 + \omega_2 h_{t-1}, \\ s_t &\sim cat(\Gamma), \quad \Gamma \sim Stick1(\eta). \end{aligned}$$

This Bayesian semiparametric model replaces the IHMM with a DPM from our model. It is nested in the GARCH-IHMM as a special case where the mixture weights are static over time. The GARCH-DPM is a univariate version of the model introduced by Jensen and Maheu (2013) and Maheu and Shamsi Zamenjani (2021).

IHMM:

$$\begin{aligned} r_t &= \mu_{s_t} + \rho r_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_{s_t}), \\ s_t | s_{t-1} &\sim cat(\Pi_{s_{t-1}}),\ \Gamma \sim Stick1(\eta),\ \Pi \sim Stick2(\alpha, \Gamma). \end{aligned}$$

This Bayesian nonparametric model removes the GARCH component from the GARCH-IHMM and corresponds to the model of Maheu and Yang (2016).

Parametric GARCH models:

$$\begin{aligned} r_t &= \mu + \rho r_{t-1} + \sqrt{h_t}\epsilon_t, \\ h_t &= \omega_0 + \omega_1 (r_{t-1} - \rho r_{t-2} - \omega_3)^2 + \omega_2 h_{t-1}. \end{aligned}$$

This specification refers to four different specifications. One extension allows the shock term ϵ_t to follow either $N(0,1)$ or $t(\nu)$. Another extension is to set ω_3 to be the same as μ (symmetric) or different from μ (asymmetric). Therefore, there are four benchmark GARCH models, each corresponding to one particular combination: AGARCH-t (asymmetric GARCH with Student's t shocks), GARCH-t (symmetric GARCH with Student's t shocks), AGARCH-N (asymmetric GARCH with normal shocks), and GARCH-N (symmetric GARCH with normal shocks).

3 Estimation and Forecast Methods

A Markov chain Monte Carlo (MCMC) algorithm can be used to estimate the proposed and the benchmark models. For the proposed GARCH-IHMM, we partition the infinite-dimensional state space to K *active* states plus one *remaining* state, and apply the beam sampler (Van Gael et al., 2008) to estimate the number of active states K, the partitioned probability vector Γ, and the partitioned transition matrix Π. One way to see this algorithm is that it introduces new states when necessary and drops states without observations assigned. Each MCMC iteration consists of the following steps:

1. Sample the precision parameters $\eta, \alpha | K, s_{1:T}$ with the sampler from Fox et al. (2011).
2. Sample $\Gamma, \Pi | K, s_{1:T}, \eta, \alpha$ with conjugate results.
3. Sample the auxiliary slice variable $u_{1:T} | \Pi$ and update the number of active states K with the beam sampler from Van Gael et al. (2008).
4. Sample the latent states $s_{1:T} | r_{1:T}, \Theta_{1:K}, \theta_h, \Gamma, \Pi, u_{1:T}$ with the forward filter backward sampler (FFBS) from Chib (1996) with necessary adjustments.
5. Sample the state-dependent parameters $\Theta_{1:K} | r_{1:T}, s_{1:T}, \theta_h$ with conjugate results.
6. Sample the base measure hyperparameters $b_0, B_0, v_0, s_0 | \Theta_{s_t}$ with conjugate results.
7. Sample the GARCH parameters $\theta_h | r_{1:T}, s_{1:T}, \Theta_{1:K}$ with a block-move random-walk Metropolis-Hastings sampler.

The detailed sampling steps of the above algorithm can be found in Appendix A.

Log predictive likelihoods and log Bayes factors are used to compare out-of-sample forecast performance among models. Let $\Theta_{all,A}^{(i)}$ denote the set of all parameter draws for model $\mathcal{M}_A$ from the ith MCMC iteration out of M iterations. Predictive likelihood $p(r_{t+1} | r_{1:t}, \mathcal{M}_A)$ is evaluated in the following way:

$$p\left(r_{t+1} | r_{1:t}, \mathcal{M}_A\right) \approx \frac{1}{M} \sum_{i=1}^{M} p\left(r_{t+1} \middle| r_{1:t}, \Theta_{all,A}^{(i)}\right) p\left(\Theta_{all,A}^{(i)} \middle| r_{1:t}\right). \tag{3}$$

Then the log predictive likelihood of model $\mathcal{M}_A$ over the out-of-sample periods $t+1:T$ is computed as

$$\log PL_A = \log p\left(r_{t+1:T} | r_{1:t}, \mathcal{M}_A\right) = \sum_{l=t+1}^{T} \log p\left(r_l | r_{1:l-1}, \mathcal{M}_A\right), \tag{4}$$

and the log Bayes factor of model $\mathcal{M}_A$ against model $\mathcal{M}_B$ is $\log BF_{AB} = \log PL_A - \log PL_B$. A log Bayes factor greater than or equal to 5 is considered strong evidence supporting model $\mathcal{M}_A$ as suggested by Kass and Raftery (1995). Further, the log score difference (LSD) suggested by Pettenuzzo and Timmermann (2011) is calculated to measure the relative improvement in log predictive likelihoods for model $\mathcal{M}_A$ against model $\mathcal{M}_B$: $LSD_{AB} = \frac{\log BF_{AB}}{|\log PL_B|}$.

4 Data

We choose the 3-month US T-bill secondary market rates from the Federal Reserve Economic Data (FRED) database to represent the short-term interest rates. The data span from January 1948 to December 2023 at a monthly frequency with 912 observations. The quoted rates in percentage value are used to estimate each model.

Table 1 lists the descriptive statistics of the sample. The sample distribution is approximately symmetric with standard tails, and the short-rate process is highly autocorrelated, whereas the autocorrelation is significantly reduced after taking the first-order difference.

Figure 1 plots the quoted market rates over time. The short rates generally trend up and down, while after the 2008 crisis and during the COVID pandemic, the short interest rates are consistently low, suggesting potential regime switches.

5 Empirical Results

We employ the MCMC algorithm in Section 3 to estimate and forecast the proposed GARCH-IHMM model.

5.1 In-sample estimates

Table 2 presents the full-sample posterior estimates for selected models. Overall, the GARCH-IHMM model successfully captures the dynamic patterns that both the IHMM and the GARCH models identify in the data. For example, the volatility persistence ($\omega_1 + \omega_2$) of the GARCH-IHMM and that of the AGARCH-N are approximately the same, and the concentration parameters α, η along with the number of active states K of the GARCH-IHMM are very similar to those of the IHMM model.

In contrast, the GARCH-DPM model displays a lower concentration parameter and identifies fewer active states compared to the IHMM. This suggests that the time-varying IHMM can approximate more complex distributions and dynamics than the static DPM model. Additionally, the 95% density interval of the GARCH parameters (ω_0, ω_1, ω_2) in the GARCH-DPM model differs significantly from those of the other two models.[4]

Figure 2 plots the posterior means of the time-varying parameters in the GARCH-IHMM. The state-dependent mean μ_{s_t} and standard deviation σ_{s_t} exhibit clear shifts across different states, effectively serving to capture complex economic shocks. In the lower panel, we observe that σ_{s_t} is close to one most of the time, indicating little deviation from the persistent volatility h_t, with notable exceptions. For example, σ_{s_t} spikes to 13 during the COVID-19 shock in March 2020 (the interest rate drops dramatically from 1.52% in February 2024 to 0.29% in one month) when h_t grew gradually. After 2008 during the quantitative easing, the state-dependent σ_{s_t} quickly reduced the impact of h_t, which dropped slowly during this period.

5.2 Out-of-sample forecasts

Out-of-sample forecast results are reported in Table 3. In terms of density forecasts, the proposed GARCH-IHMM achieves the highest log predictive likelihoods, while the IHMM performs the best among the benchmark models, consistent with the findings in Maheu and Yang (2016). However, the IHMM still significantly falls behind the GARCH-IHMM, with a log Bayes factor of 33.7646 and a log score difference of 16.89%. The GARCH-DPM performs similarly to the parametric GARCH models with Student t innovations, indicating that the DPM component captures the fat tails, but important distributional dynamics remain missing.

In terms of point forecasts, the IHMM model is the best among all models with the lowest root mean squared forecast error (RMSFE) of 0.3947, while the GARCH-IHMM is the second-best model (1.40% loss to IHMM). The static mixture GARCH-DPM and the parametric GARCH models perform approximately the same, with RMSFEs from 0.4058 to 0.4066. Switching to the GARCH-IHMM yields an improvement of 1.37% to 1.57%.

Figure 3 plots the out-of-sample cumulative log Bayes factors (CLBFs) of the GARCH-IHMM against the GARCH-DPM, IHMM, and GARCH-t. This measure provides a period-by-period comparison of forecast performance, illustrating how the log Bayes factors evolve over time to capture the ongoing predictive accuracy of each model. An upward-trending curve indicates that the GARCH-IHMM consistently outperforms the benchmark models, and vice versa.

A turning point can be seen roughly around 1980. Before that, the GARCH-DPM and GARCH-t generally performed on par with the GARCH-IHMM, while

[4] Full-sample posterior estimates for other benchmark models are available upon request.

both models quickly lost their predictive power after that, as indicated by their upward CLBF trajectory. In contrast, the GARCH-IHMM consistently outperforms the IHMM model before 1980, and then performs approximately the same afterwards with a flat CLBF curve. The design of the GARCH-IHMM leverages the strengths of both the IHMM and the GARCH components. It allows the component with greater predictive power to automatically take charge and beat the corresponding benchmarks, which is an excellent feature in case the macroeconomic condition transits back to the pre-1980 era in the future.

6 Conclusion

This paper introduces the GARCH-IHMM, a novel Bayesian semiparametric model that combines the IHMM with an asymmetric GARCH component, designed to effectively capture the complex dynamics of short-term interest rates. Applied to the US 3-month Treasury bill rates, the GARCH-IHMM exhibits significant advantages in density forecasts, outperforming the benchmark models and offering competitive point forecasts. These results indicate that the proposed GARCH-IHMM harnesses great predictive power for its excellent adaptivity.

References

Albert, J. H. and Chib, S. (1993). Bayes inference via gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business & Economic Statistics*, 11(1):1–15.

Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics*, 12(3):309–316.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75(1):79–97.

Dufays, A. (2016). Infinite-state Markov-switching for dynamic volatility. *Journal of Financial Econometrics*, 14(2):418–460.

Duffee, G. R. (2013). Forecasting interest rates. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2 of *Handbooks in Economics*, pages 385–426. Elsevier.

Durham, G. B. (2003). Likelihood-based specification analysis of continuous-time models of the short-term interest rate. *Journal of Financial Economics*, 70(3):463–487.

Filipović, D. (2009). Term-structure models: A review. *Foundations and Trends in Finance*, 3(1):1–119.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011). A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics*, pages 1020–1056.

Garcia, R. and Perron, P. (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics*, 78(1):111–125.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.

Gurkaynak, R. S., Sack, B., and Wright, J. H. (2007). The US treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304.

Hamilton, J. D. (1988). Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control*, 12:385–423.

Hou, A. J. and Suardi, S. (2011). Modelling and forecasting short-term interest rate volatility: A semiparametric approach. *Journal of Empirical Finance*, 18(4):692–710.

Hou, C. (2017). Infinite hidden Markov switching VARs with application to macroeconomic forecast. *International Journal of Forecasting*, 33(4):1025–1043.

Jensen, M. J. and Maheu, J. M. (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics*, 157(2):306–316.

Jensen, M. J. and Maheu, J. M. (2013). Bayesian semiparametric multivariate GARCH modeling. *Journal of Econometrics*, 176:3–17.

Jin, X. and Maheu, J. M. (2016). Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics*, 192(1):19–39.

Jin, X., Maheu, J. M., and Yang, Q. (2019). Bayesian parametric and semiparametric factor models for large realized covariance matrices. *Journal of Applied Econometrics*, 34(5):641–660.

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

Jochmann, M. (2015). Modeling U.S. inflation dynamics: A Bayesian nonparametric approach. *Econometric Reviews*, 34(5):537–558.

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

Li, C. and Maheu, J. (2023). Beyond conditional second moments: Does dynamic density modelling matter to portfolio allocation? Working Paper.

Li, C., Maheu, J. M., and Yang, Q. (2024). An infinite hidden markov model with stochastic volatility. *Journal of Forecasting*, 43:2187–2211.

Luo, J., Klein, T., Ji, Q., and Hou, C. (2022). Forecasting realized volatility of agricultural commodity futures with infinite hidden Markov HAR models. *International Journal of Forecasting*, 38(1):51–73.

Maheu, J. M. and Shamsi Zamenjani, A. (2021). Nonparametric dynamic conditional beta. *Journal of Financial Econometrics*, 19(4):583–613.

Maheu, J. M. and Yang, Q. (2016). An infinite hidden Markov model for short-term interest rates. *Journal of Empirical Finance*, 38:202–220.

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies*, 73(4):1057–1084.

Pettenuzzo, D. and Timmermann, A. (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics*, 164(1):60–78.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, pages 639–650.

Shi, S. and Song, Y. (2016). Identifying spoeculative bubbles using an infinite hidden Markov model. *Journal of Financial Econometrics*, 14(1):159–184.

Song, Y. (2014). Modelling regime switching and structural breaks with an infinite hidden Markov model. *Journal of Applied Econometrics*, 29(5):825–842.

Van Gael, J., Saatci, Y., Teh, Y. W., and Ghahramani, Z. (2008). Beam sampling for the infinite hidden Markov model. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1088–1095. ACM.

Yang, Q. (2019). Stock returns and real growth: A Bayesian nonparametric approach. *Journal of Empirical Finance*, 53:53–69.

Table 1: Descriptive Statistics for the Short Rates

Panel A: Key moments for the short rates

Obs	Mean	Median	Std.Dev	Skew	Ex.Kurt	Min	Max
912	3.9707	3.6500	3.0679	0.9661	1.2069	0.0100	16.3000

Panel B: Autocorrelation for the short rate process

Lags	1	2	3	4	5	10	20
ACF	**0.9914**	**0.9772**	**0.9640**	**0.9517**	**0.9395**	**0.8860**	**0.7439**

Panel C: Autocorrelation for the I(1) process

Lags	1	2	3	4	5	10	20
ACF	**0.3427**	-0.0555	-0.0604	-0.0042	0.0474	**0.0832**	**-0.2188**

1. Data is the quoted rates in percentage, ranging from January 1948 to December 2023 at a monthly frequency.
2. Bold ACF indicates significance at 5% level.

Table 2: Full-Sample Posterior Estimates for Selected Models

GARCH-IHMM:

$$r_t = \mu_{s_t} + \rho r_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_{s_t}^2 h_t),$$
$$h_t = \omega_0 + \omega_1 (r_{t-1} - \rho r_{t-2} - \omega_3)^2 + \omega_2 h_{t-1},$$
$$s_t | s_{t-1} \sim cat(\Pi_{s_{t-1}}),\ \Gamma \sim Stick1(\eta),\ \Pi \sim Stick2(\alpha, \Gamma).$$

GARCH-DPM:

$$r_t = \mu_{s_t} + \rho r_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_{s_t}^2 h_t),$$
$$h_t = \omega_0 + \omega_1 (r_{t-1} - \rho r_{t-2} - \omega_3)^2 + \omega_2 h_{t-1},$$
$$s_t | s_{t-1} \sim cat(\Gamma), \quad \Gamma \sim Stick1(\eta).$$

IHMM:

$$r_t = \mu_{s_t} + \rho r_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_{s_t}^2),$$
$$s_t | s_{t-1} \sim cat(\Pi_{s_{t-1}}),\ \Gamma \sim Stick1(\eta),\ \Pi \sim Stick2(\alpha, \Gamma).$$

AGARCH-N:

$$r_t = \mu + \rho r_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, h_t),$$
$$h_t = \omega_0 + \omega_1 (r_{t-1} - \rho r_{t-2} - \omega_3)^2 + \omega_2 h_{t-1}.$$

	GARCH-IHMM		GARCH-DPM		IHMM		AGARCH-N	
	Mean	0.95DI	Mean	0.95DI	Mean	0.95DI	Mean	0.95DI
μ							0.0075	(-0.0029,0.0183)
ω_0	0.0008	(0.0002, 0.0017)	0.0000	(0.0000,0.0001)			0.0006	(0.0003,0.0010)
ω_1	0.1566	(0.0507, 0.2622)	0.3564	(0.2971,0.4037)			0.2572	(0.2205,0.2968)
ω_2	0.7761	(0.7021, 0.8437)	0.6075	(0.5574,0.6680)			0.7403	(0.7014,0.7766)
ω_3	0.0334	(-0.0162, 0.0871)	-0.0022	(-0.0117,0.0058)			0.0226	(0.0094,0.0354)
ρ	0.9949	(0.9874, 1.0019)	1.0040	(1.0013,1.0071)	1.0007	(0.9935, 1.0057)	1.0067	(1.0013,1.0124)
α	1.7692	(1.0198, 2.7393)			1.4563	(0.8512, 2.2537)		
η	1.4568	(0.6773, 2.5435)	0.7989	(0.3412,1.4741)	1.6188	(0.7699, 2.7919)		
K	8.0957	(6.0000,12.0000)	5.1573	(4.0000,8.0000)	9.3216	(6.0000,13.0000)		

1. Full sample consists of 912 observations, ranging from January 1948 to December 2023 at a monthly frequency.
2. “Mean” denotes the posterior average of the corresponding parameter.
3. “0.95DI” denotes the posterior 95% density interval of the corresponding parameter.

Table 3: Out-of-Sample Forecast Results

	LPL	LBF	LSD	RMSFE	(%)
GARCH-IHMM	233.6668	—	—	0.4003	—
GARCH-DPM	188.1377	45.5291	24.20%	0.4056	1.32%
AGARCH-t	187.1061	46.5607	24.88%	0.4063	1.49%
GARCH-t	187.9625	45.7043	24.32%	0.4065	1.53%
AGARCH-N	50.7593	182.9075	360.34%	0.4058	1.37%
GARCH-N	57.5381	176.1287	306.11%	0.4066	1.57%
IHMM	199.9023	33.7646	16.89%	0.3947	-1.40%
Rolling-window AR	-150.2158	383.8826	255.55%	0.4172	4.22%

1. Initial training set spans from January 1948 to December 1953, and the out-of-sample period spans from January 1954 to December 2023 (840 out-of-sample periods).

2. LPL denotes log predictive likelihoods (the higher the better). LBF denotes log Bayes factors for the proposed GARCH-IHMM against the corresponding model, where an LBF of 6 and above suggests significant evidence for GARCH-IHMM. LSD denotes the log score difference, which is the percentage improvement of the LPL for the GARCH-IHMM from the corresponding benchmark model with the base of the benchmark.

3. RMSFE denotes root mean squared forecast error for predictive means (the lower the better), and the following (%) denotes the percentage improvement of the RMSFE for the GARCH-IHMM from the corresponding benchmark model with the base of the benchmark.

4. The window width for the rolling-window AR model is a fixed 60 months (5 years).

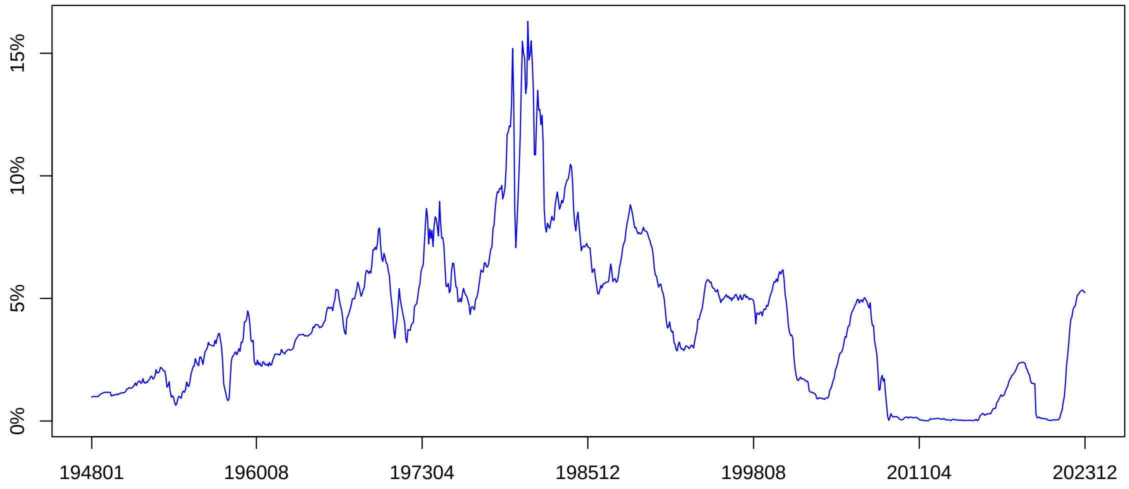

Figure 1: Quoted Market 3-Month T-bill Rates over Time

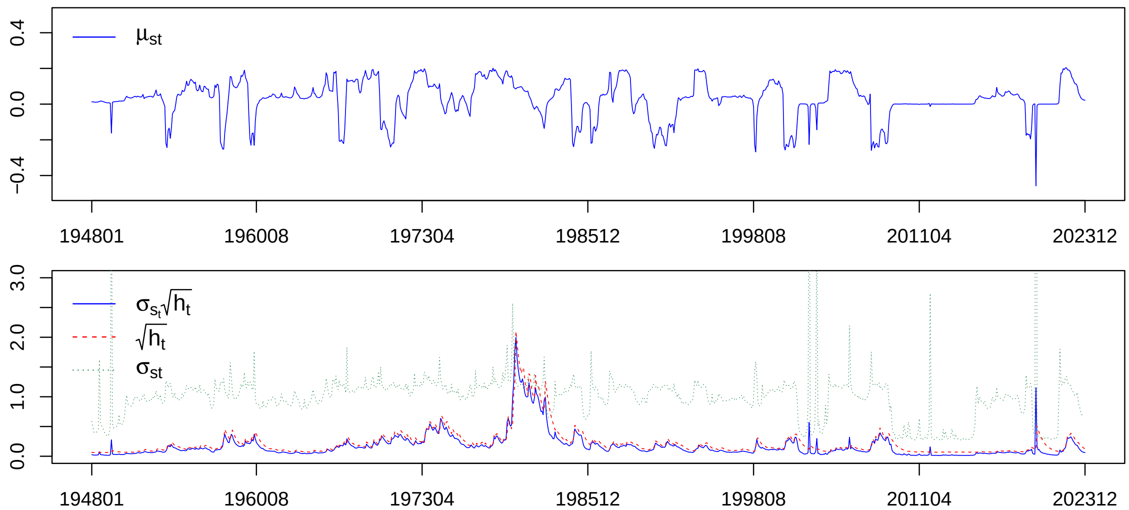

Figure 2: Posterior Means of Time-Varying Parameters

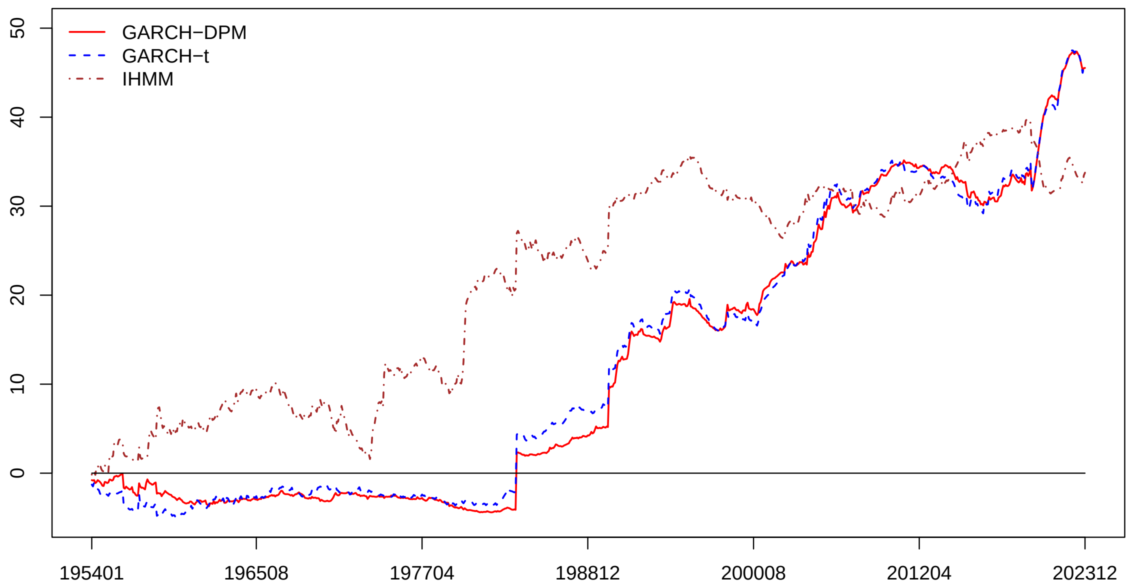

Figure 3: Out-of-Sample Cumulative Log Bayes Factors

A Detailed Sampling Algorithm for GARCH-IHMM

Recall that the infinitely dimensional state space is partitioned into K active states plus one remaining state. Then, Γ and Π are partitioned into $\Gamma = (\gamma_1, \ldots, \gamma_K, \gamma_R)'$ and $\Pi_j = (\pi_{j1}, \ldots, \pi_{jK}, \pi_{jR})$, where $\gamma_R = \sum_{k=K+1}^{\infty} \gamma_k = 1 - \sum_{k=1}^{K} \gamma_k$ and $\pi_{jR} = \sum_{k=K+1}^{\infty} \pi_{jk} = 1 - \sum_{k=1}^{K} \pi_{jk}$. The sampling steps are:

1. Sample $c_{1:K}|s_{1:T}, \Gamma, \alpha$. $c_{1:K}$ is essential for sampling Γ, and it counts balls of different colors in the "oracle" urn. Fox et al. (2011) proposes simulating c_k from the hierarchical Pòlya urn scheme instead of sampling it.

   (a) Count the number of each transition type, n_{jk}, for the number of times switching from state j to state k.

   (b) Simulate an auxiliary trail variable $x_i \sim Bernoulli\left(\frac{\alpha\gamma_k}{i-1+\alpha\gamma_k}\right)$, for $i = 1, \ldots, n_{jk}$. If the trial is successful, an "oracle" urn step is involved at the ith step towards n_{jk}, and we increase the corresponding "oracle" counts, o_{jk}, by one.

   (c) $c_k = \sum_{j=1}^{K} o_{jk}$.

2. Sample η. Following Fox et al. (2011); Maheu and Yang (2016), assume a prior $\eta \sim Gamma\,(a_1, b_1)$, and let $c = \sum_{j=1}^{K} c_j$,

   (a) $\nu \sim Bernoulli\left(\frac{c}{c+\eta}\right)$

   (b) $\lambda \sim Beta\,(\eta + 1, c)$

   (c) $\eta \sim Gamma\,(a_1 + K - \nu, b_1 - \log\lambda)$

3. Sample α. Following Fox et al. (2011), assume a prior $\alpha \sim Gamma\,(a_2, b_2)$, and let $n_j = \sum_{k=1}^{K} n_{jk}$,

   (a) $\nu_j \sim Bernoulli\left(\frac{n_j}{n_j+\alpha}\right)$

   (b) $\lambda_j \sim Beta\,(\alpha + 1, n_j)$

   (c) $\alpha \sim Gamma\left(a_2 + c - \sum_{j=1}^{K} \nu_j, b_2 - \sum_{j=1}^{K} \log(\lambda_j)\right)$

4. Sample $\Gamma|c_{1:K}, \eta$. Given the "oracle" urn counts $c_{1:K}$ and the property of the Dirichlet process, the conjugate posterior is

$$\Gamma|c_{1:K}, \eta \sim Dir\,(c_1, \ldots, c_K, \eta) \tag{5}$$

5. Sample $\Pi | n_{1:K,1:K}, \Gamma, \alpha$. Similarly, the conjugate posterior of Π_j is

$$\Pi_j | n_{j,1:K}, \Gamma, \alpha \sim Dir\left(\alpha\gamma_1 + n_{j1}, \ldots, \alpha\gamma_K + n_{jK}, \alpha\gamma_R\right) \tag{6}$$

6. Sample $u_{1:T} | \Gamma, \Pi$. The auxiliary slice variable $U = \{u_t\}_{t=1}^T$ is drawn by $u_1 \sim U\left(0, \gamma_{s_1}\right)$ and $u_t \sim U\left(0, \pi_{s_{t-1}s_t}\right)$.

7. Update K. Similar to the DPM model, if K does not meet the condition

$$\min \{u_t\}_{t=1}^T > \max \{\pi_{jR}\}_{j=1}^K \tag{7}$$

   then K needs to be increased by 1 ($K' = K + 1$) and all of the corresponding parameters need to be drawn from the base measure. In addition, since a new active state is introduced, Γ and Π also need to be updated accordingly:

   (a) $\Theta_{K'} \sim H$;

   (b) Draw $v \sim Beta\left(1, \eta\right)$, then update $\Gamma = \left(\gamma_1, \ldots, \gamma_K, \gamma_{K'}, \gamma_R\right)'$, where $\gamma_{K'} = v\gamma_R$ and $\gamma_R = \left(1 - v\right)\gamma_R$;

   (c) Draw $v_j \sim Beta\left(\alpha\gamma_{K'}, \alpha\gamma_R\right)$, update $\Pi_j = \left(\pi_{j1}, \ldots, \pi_{jK}, \pi_{jK'}, \pi_{jR}\right)$ for $j = 1, \ldots, K$, where $\pi_{jK'} = v\pi_{jR}$ and $\pi_{jR} = \left(1 - v\right)\pi_{jR}$;

   (d) Draw the K'th row of Π, $\Pi_{K'}$, by $\Pi_{K'} \sim Dir\left(\alpha\gamma_1, \ldots, \alpha\gamma_K, \alpha\gamma_{K'}, \alpha\gamma_R\right)$.

   Repeat the above steps until inequality (7) holds.

8. The forward filter for $s_{1:T} | r_{1:T}, u_{1:T}, \Gamma, \Pi, \Theta, \theta_h$. Iterating the following steps forward from 1 to T:

   (a) The prediction step for the initial state s_1 is as follows:

$$p(s_1 = k | u_1, \Gamma) \propto \mathbb{1}\left(u_1 < \gamma_k\right), \qquad k = 1, \ldots, K \tag{8}$$

   for the following states $s_{2:T}$:

$$p(s_t = k | r_{1:t-1}, u_{1:t}, \Pi, \Theta, \theta_h) \propto \sum_{j=1}^{K} \mathbb{1}\left(u_t < \pi_{jk}\right) p\left(s_{t-1} = j | r_{1:t-1}, u_{1:t-1}, \Pi, \Theta, \theta_h\right) \tag{9}$$

   (b) The update step for $s_{1:T}$:

$$p\left(s_t = k | r_{1:t}, u_{1:t}, \Pi, \Theta, \theta_h\right) \propto p\left(r_t | r_{t-1}, \Theta_k, \theta_h\right) p\left(s_t = k | r_{1:t-1}, u_{1:t}, \Pi, \Theta, \theta_h\right) \tag{10}$$

9. The backward sampler for $s_{1:T}|r_{1:T}, u_{1:T}, \Pi, \Theta, \theta_h$. Sample the states $s_{1:T}$ using the previously filtered values backward from T to 1:

   (a) for the terminal state s_T directly from $p\left(s_T|r_{1:T}, u_{1:T}, \Pi, \Theta, \theta_h\right)$

   (b) for the rest states,

$$p\left(s_t = k|s_{t+1} = j, r_{1:t}, u_{1:t+1}, \Pi, \Theta, \theta_h\right) \propto \mathbb{1}\left(u_{t+1} < \pi_{kj}\right) p\left(s_t = k|r_{1:t}, u_{1:t}, \Pi, \Theta, \theta_h\right) \tag{11}$$

10. Sample $\Theta|r_{1:T}, s_{1:T}, \theta_h$. Assume conjugate priors $\mu_{s_t} \sim N\left(b_0, B_0\right)$ and $\sigma_{s_t}^{-2} \sim Gamma\left(\frac{v_0}{2}, \frac{s_0}{2}\right)$. Define $Y_k \equiv \left\{h_t^{-1/2} r_t | s_t = k\right\}_{t=2}^{T}$ and $X_k \equiv \left\{h_t^{-1/2} | s_t = k\right\}_{t=2}^{T}$. The linear model is now

$$Y_k = X_k \mu_k + \epsilon_k, \quad \epsilon_k \sim N\left(0, \sigma_k^2\right) \tag{12}$$

The posteriors are

$$p\left(\mu_k|Y_k, \sigma_k^2, h_{1:T}\right) \propto \prod_{t:s_t=k} p\left(Y_t|\mu_k, \sigma_k^2, h_t\right) p\left(\mu_k\right) \tag{13}$$

$$\sim N\left(M_\mu, V_\mu\right) \tag{14}$$

where

$$M_\mu = V_\mu \left(\sum_{t:s_t=k} h_t^{-1} \sigma_k^{-2} r_t + B_0^{-1} b_0\right) \tag{15}$$

$$V_\mu = \left(\sum_{t:s_t=k} h_t^{-1} \sigma_k^{-2} + B_0^{-1}\right)^{-1} \tag{16}$$

and

$$p\left(\sigma_k^{-2}|Y_k, \mu_k, h_{1:T}\right) \propto \prod_{t:s_t=k} p\left(r_t|\mu_k, \sigma_k^{-2}, h_t\right) p\left(\sigma_k^{-2}\right) \tag{17}$$

$$\sim Gamma\left(\frac{\bar{v}}{2}, \frac{\bar{s}}{2}\right) \tag{18}$$

where

$$\bar{v} = T_k + v_0 = \sum_{t=1}^{T} \mathbb{1}\left(s_t = k\right) + v_0 \tag{19}$$

$$\bar{s} = \sum_{t:s_t=k} h_t^{-1} \left(r_t - \mu_k\right)^2 + s_0 \tag{20}$$

11. Sample hierarchical priors.

    (a) Sample $b_0|\mu_{1:K}, B_0 \sim N\left(\mu_b, \sigma_b^2\right)$, where the prior is $b_0 \sim N(b_b, B_b)$ and

$$\mu_b = \sigma_b^2 \left(B_0^{-1} \sum_{k=1}^{K} \mu_k + B_b^{-1} b_b \right) \tag{21}$$

$$\sigma_b^2 = \left(K B_0^{-1} + B_b^{-1} \right)^{-1} \tag{22}$$

    (b) Sample $B_0|\boldsymbol{\mu}_{1:K}, b_0 \sim IG\left(\frac{\bar{v}_B}{2}, \frac{\bar{s}_B}{2}\right)$, where the prior is $B_0 \sim IG(\frac{v_B}{2}, \frac{s_B}{2})$ and

$$\bar{v}_B = K + v_B \tag{23}$$

$$\bar{s}_B = \sum_{k=1}^{K} \left(\mu_k - b_0\right)^2 + s_B \tag{24}$$

    (c) Sample $v_0|\sigma_{1:K}^2, s_0$, where the prior is $v_0 \sim Exp(g_0)$. There is no easily applicable conjugate prior for v_0 so a Metropolis-Hastings step needs to be applied. Implement a Gamma proposal following Maheu and Yang (2016):

$$v_0'|v_0 \sim Gamma\left(\tau, \frac{\tau}{\nu}\right) \tag{25}$$

    and the acceptance rate is

$$\min\left\{1, \frac{p\left(v_0'|\sigma_{1:K}^2, s_0, g_0\right)/q\left(v_0'|v_0\right)}{p\left(v_0|\sigma_{1:K}^2, s_0, g_0\right)/q\left(v_0|v_0'\right)}\right\} \tag{26}$$

    (d) Sample $s_0|\sigma_{1:K}^2, v_0 \sim Gamma\left(c_s, d_s\right)$, where the prior is $s_0 \sim Gamma(c_0, d_0)$ and

$$c_s = \frac{K v_0}{2} + c_0 \tag{27}$$

$$d_s = \frac{\sum_{k=1}^{K} \sigma_k^{-2}}{2} + d_0 \tag{28}$$

12. Sample the GARCH parameters $\theta_h = (\omega_0, \omega_1, \omega_2, \omega_3, \rho)'|r_{1:T}, s_{1:T}, \Theta$. With normal prior $\theta_h \sim N\left(0, I\right)$, the posterior is

$$p\left(\theta_h|r_{1:T}, s_{1:T}, \Theta\right) \propto \prod_{t=1}^{T} p\left(r_t|\Theta_{s_t}, h_t\right) p\left(\theta_h\right). \tag{29}$$

    And then apply a random-walk Metropolis-Hastings algorithm.