Bayesian workflow for term structure modeling

Thomas Martins

April 15, 2023

Abstract

We apply the Bayesian workflow approach by [GVS+20] of model building, inference and comparison to the modeling of the term structure of interest rates. We describe novel procedures of Bayesian data analysis, such as prior predictive checking, computational diagnostics, and posterior predictive checkings. We propose versions of the Vasicek, CIR and Diebold-Li models for the term structure, elicit prior distributions for the parameters, estimate the models with both simulated and real data, and compare the models. Our results show that the Diebold-Li models have better predictive capabilities than those of the affine (Vasicek and CIR) models, however our main goal is to illustrate how the Bayesian workflow approach is carried out in a practical setting of financial econometrics.

Keywords: Term structure of interest rates, Bayesian data analysis, Bayesian inference, Financial econometrics, Time series modeling.

1 Introduction

The advances in computational methods over the last decades have allowed for possibilities in Bayesian modeling that were otherwise unfeasible. In order to tap into this newfound potential, modeling procedures that go beyond the regular Bayesian inference with conjugate priors must be formalized and developed. The Bayesian workflow proposed by [GVS+20] summarizes advances from the "Bayesian data analysis" approach, ranging from prior distribution elicitation and computation diagnostics until model comparison and prediction of future data.

The goal of this work is to show how the Bayesian workflow can be applied to a practical issue in financial econometrics, in this case term structure modeling. We propose several statistical models for the term structure of interest rates and carry out procedures such as prior and posterior predictive checking, estimation with simulated and real data, and convergence diagnostics. Guided by Bayesian reasoning, we analyse before and after estimation how consistent with domain knowledge our modeling is.

Section 2 details our methodology, stating and describing each step of our adaptation of the Bayesian workflow. Section 3 presents our results for each step in the context of term structure modeling. Section 4 contains concluding remarks.

2 Methodology

Our methodology consists of proposing a Bayesian workflow in the likes of the one by [GVS+20] for comparing between statistical models for the term structure of interest rates. Each step of the workflow is then carried out and their results are interpreted in the light of the domain knowledge for term structure modeling. The Bayesian workflow aims to be a step-by-step method based on the Bayesian data analysis approach, which goes beyond traditional Bayesian inference by also including model building and checking/comparison under Bayesian reasoning.

The steps of our proposed Bayesian workflow for term structure modeling consist of

1. Describing our data
2. Proposing models and priors for each parameter
3. Prior predictive checking

4. Estimation with simulated data
5. Computational diagnostics and assessing of estimation bias
6. Estimation with real data
7. Posterior predictive checking
8. Comparison of predictive abilities

The first step is describing the dataset we want to propose models for, and more generally yield curve data. The yield curve is the graphical representation of the term structure of interest rates at a given point in time, however it is common to refer to the term structure as the "yield curve". The main characteristic of yield curve data is how it is jointly determined by short and long interest rates. This makes term structure modeling different from e.g. modeling of stock prices, and, ideally, should be accounted for in our models. Empirically, the yield curve presents three patterns: flat, steep and inverted.

In the model proposing step, we look at some of the most common models for term structure modeling in financial econometrics. More specifically, we provide state-space specifications of the benchmark models by [Vas77], [CIR85] and [DL06]. We discuss the assumptions behind these models, and how reasonable they are in practice. We also describe the parameters of each model, and propose prior distributions for each parameter. We then move on to prior predictive checking, a procedure proposed by [GSV+19], in order to choose values for the hyperparameters of the priors. We use prior predictive checking to simulate data that is *a priori* consistent with our knowledge about the yield curve.

We then move on to estimating each of the models with fake, simulated data. This is important because, among other reasons, we can better assess issues such as adequate computation and estimation bias in a setting where the true parameters are known [GVS+20]. This becomes more difficult to do with real data, where we do not even know the true data-generating process, let alone parameter values. After estimating the models with simulated data, we analyse the computational diagnostics and parameter bias.

Being aware of computational issues and/or estimation bias, we can then move on to the estimation with real data. Seeing how our proposed model(s) reacts with the observed data is perhaps the most critical step in statistical modeling. However, in a setting where the models are complex and the direct, practical interpretation of the parameters might not be clear, we can propose other ways of evaluating models after fit. This is where procedures such as posterior predictive checking [GSV+19] come in, taking advantage of the fact that, with the previously observed data and updated distributions (the posteriors) for the parameters, we can sample from the posterior predictive distribution. This gives us hypothetical "to be observed" future data, which we can compare to the observed data with graphical plots in the style of [GSV+19]. This comparison is a good measure of model fit quality, as, if the model is good at describing our observed data, it makes sense that it can generate additional data that resemble the previously observed set. Besides posterior predictive checking, we present additional measures of predictive ability, such as a cross-validation procedure or an information criterion.

3 Steps of our Bayesian workflow

3.1 Describing the data

The data utilized in this work comes from NEFIN FEA-USP – Center for Research in Financial Economics of the Department of Economics, and consists of the spot rate curve for Brazil, calculated from One-Day Interbank Deposit Futures contracts (DI rate).[1] In our estimated models we shall use the yields with the following maturities: 2 months, 3 months, 6 months, 1 year, 3 years, 5 years, all with weekly frequency. Our utilized sample starts in 2012-01-08 and ends in 2020-05-24, totalling 438 observations for each maturity.

From the graphical visualization of our dataset, we can notice some features. The highest value observed for the rates is of around 16% for the 3 and 5 year maturities, and the lowest is of around 2% for the shortest maturities. If we imagine the observed data as a realization of our data-generating

[1] The data can be found at http://nefin.com.br/data/spot_rate_curve.html

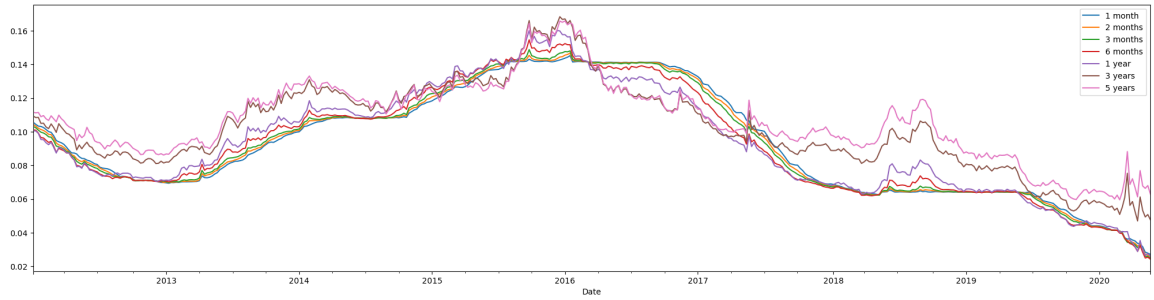

Figure 1: Term structure data for Brazil from 2012 to 2020

process, it would seem highly unlikely that the interest rates would ever go above 30% for any of the maturities. Also, one might argue that, given the characteristics of the Brazilian economy during our chosen time frame, interest rates below zero would also be highly unlikely. We shall keep this in mind when the prior predictive checking step comes.

3.2 Proposing models

As previously mentioned, the most relevant feature of yield curve data is that the term structure is jointly determined by long and short rates. It is then reasonable to expect that good models for term structure data take this into account. We look at some of the benchmark models in financial econometrics for term structure data, more specifically the ones by [Vas77], [CIR85], and [DL06]. As we have a panel of time series for rates with different maturities, we seek dynamic (as opposed to static) models for the term structure. We use a latent variable approach for all three models, specifying each model in a state space framework.

We start with the Vasicek model, which describes the rates for all of the maturities as a function of the short rate, which is defined as a spot rate with time to maturity tending towards zero. The law of motion for the short rate in the Vasicek model is an Ornstein-Uhlenbeck process, which is a continuous time analogous of an AR(1). The Vasicek model is an affine model for the term structure, meaning that the rates for each maturity are affine functions of the short rate [Pia10]. We can write the observed spot yields $Y_t(\tau)$ for maturities τ as an affine function of the short rate r_t:

$$\begin{aligned} dr_t &= \kappa(\mu - r_t)dt + \sigma dW_t, \\ Y_t(\tau) &= -\frac{A(b,a,\sigma,\tau) - B(b,\tau)r_t}{\tau} + \eta_t. \end{aligned}$$

The imposition of no-arbitrage is used to derive the coefficients A and B [LMN15]:

$$\begin{aligned} B(b,\tau) &= \frac{1-\exp\{-b\tau\}}{b}, \\ A(b,a,\sigma,\tau) &= \left(\frac{\sigma^2}{2b^2} - a\right)\tau + \frac{1-\exp\{-b\tau\}}{b}\left(a - \frac{\sigma^2}{b^2}\right) + \frac{\sigma^2}{4b^3}(1-\exp\{-2b\tau\}). \end{aligned}$$

.

The parameters of the standard Vasicek model are κ, μ and σ, but, following the approach by [JP10], we add a and b as the physical measure analogs of the risk-neutral parameters μ and κ. We also add observation noise η_t as this can break a possible stochastic singularity [JP10].

The proposed prior distributions for the parameters are

$$\begin{aligned} \kappa &\sim TN_{(0,\frac{2}{\Delta t})}(\mu_\kappa, \sigma_\kappa^2), \\ \mu &\sim TN_{(0,\infty)}(\mu_\mu, \sigma_\mu^2), \\ \sigma &\sim HalfCauchy(\beta_\sigma), \\ a &\sim N(\mu_a, \sigma_a^2), \\ b &\sim N(\mu_b, \sigma_b^2), \\ \eta_t &\sim N(0, \sigma_{obs}^2). \end{aligned}$$

We use truncated normal distributions for κ and μ as we want to impose stationarity in the short rate process (in the case of κ) and fix the mean of the process as a positive value (in the case of μ). We use the half-Cauchy prior for σ based on the recommendation by [PS12]. The values for the hyperparameters shall be chosen through the prior predictive checking procedure in the next subsection.

The next model is the Cox-Ingersoll-Ross (CIR) model, which is also affine. It bears a great similarity to the Vasicek model but has one key difference: the law of motion for the short rate is a square-root process instead of the Ornstein-Uhlenbeck. This has two consequences in practice: it adds conditional heteroskedasticity and ensures that the short rate only assumes positive values. This is not intrinsically "better" or "worse" than the Vasicek model, as if these assumptions make sense or not depends on our data. The CIR model in state space form is

$$\begin{aligned}
dr_t &= \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t, \\
Y_t(\tau) &= -\frac{A(\tau) - B(\tau)r_t}{\tau} + \eta_t, \\
\eta_t &\sim N(0, \sigma^2_{obs}), \\
\gamma &= \sqrt{\kappa^2 + 2\sigma^2}, \\
g(\tau) &= 2\gamma + (\kappa + \gamma)(\exp\{\gamma\tau\} - 1), \\
B(\tau) &= 2(\exp\{\gamma\tau\} - 1)/g(\tau), \\
A(\tau) &= \frac{2\kappa\mu}{\sigma^2} \ln\left[\frac{2\gamma \exp\{(\kappa + \gamma)\tau/2\}}{g(\tau)}\right].
\end{aligned}$$

Our prior distributions for the CIR model are

$$\begin{aligned}
\kappa &\sim TN_{(0, \frac{2}{\Delta t})}(\mu_\kappa, \sigma^2_\kappa), \\
\mu &\sim TN_{(0,\infty)}(\mu_\mu, \sigma^2_\mu), \\
\sigma &\sim HalfCauchy(\beta_\sigma), \\
a &\sim N(\mu_a, \sigma^2_a), \\
b &\sim N(\mu_b, \sigma^2_b), \\
\eta_t &\sim N(0, \sigma^2_{obs}).
\end{aligned}$$

The derivation of the observation equation can be found in [LMN15] and the reasoning for the choice of our priors is very similar to the Vasicek case.

The last model is the dynamic adaptation of the three factor Nelson-Siegel model by [DL06]. The level, slope and curvature factors become time-varying. Each of the factors follow an AR(1), but we use continuous time specifications similarly to our Vasicek short rate. Our factors are also assumed to be independent of each other.

Our state space specification for the Diebold-Li model is

$$\begin{aligned}
dl_t &= \kappa_l(\mu_l - l_t)dt + \sigma_l dW_{1t}, \\
ds_t &= \kappa_s(\mu_s - s_t)dt + \sigma_s dW_{2t}, \\
dc_t &= \kappa_c(\mu_c - c_t)dt + \sigma_c dW_{3t}, \\
Y_t(\tau) &= l_t + s_t\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) + c_t\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \eta_t,
\end{aligned}$$

where λ is a decay parameter which control factor loadings. We propose two variants of the Diebold-Li model, one with λ as a static parameter and another with a time-varying λ following an AR(1) process. We assume the static specification of λ has prior distribution

$$\lambda \sim N(1.1, 0.09),$$

and the dynamic specification has priors

$$\begin{aligned}
\lambda_t &= \mu_\lambda + \exp\{\phi_\lambda \lambda_{t-1} + \varepsilon_t\}, \\
\varepsilon_t &\sim N(0, \frac{1}{\tau_\lambda}).
\end{aligned}$$

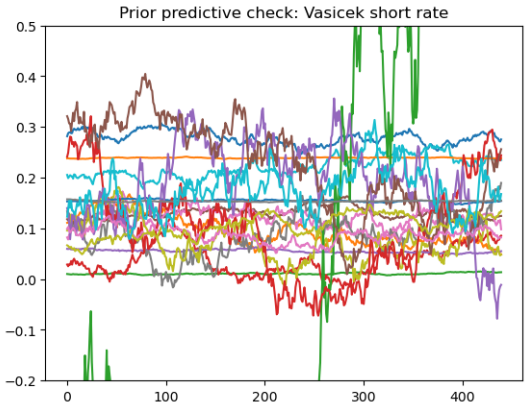

Figure 2: Prior predictive check: samples from Vasicek short rate

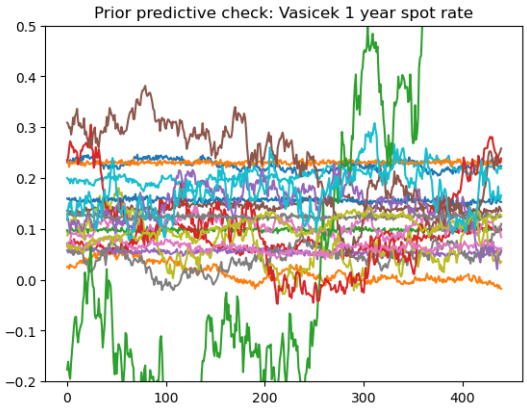

Figure 3: Prior predictive check: samples from Vasicek 1 year spot rate

The prior distributions for the remaining parameters are

$$\begin{aligned}
\kappa_{l,s,c} &\sim TN_{(0,\frac{2}{\Delta t})}(\theta_\kappa, \sigma_\kappa^2), \\
\mu_l &\sim TN_{(0,\infty)}(\theta_{\mu,l}, \sigma_{\mu,l}^2), \\
\mu_{s,c} &\sim TN_{(-\infty,0)}(\theta_{\mu,sc}, \sigma_{\mu,sc}^2), \\
\sigma_{l,s,c} &\sim HalfCauchy(\beta_\sigma), \\
\eta_t &\sim N(0, \sigma_{obs}^2),
\end{aligned}$$

chosen in a fashion similarly to the Vasicek and CIR priors.

3.3 Prior predictive checking

With the models and prior distributions at hand, we now look at the problem of choosing hyperparameter values. One tempting idea might be that of choosing "uninformative" priors, where we let the data "speak for itself". However, the analysis we did in subsection 3.1 shows how, given the nature of the data, some values are much more or less likely than others. One so-called "uninformative" prior might attribute equal likelihood to scenarios that are *a priori* highly or minimally likely. Thus, prior hyperparameters must be consistent with what we know *a priori* regarding our data.

By trial and error, we have found hyperparameter values for our priors that result in the samples for the factors and the 1 year spot rate shown in Figures 2 to 7.

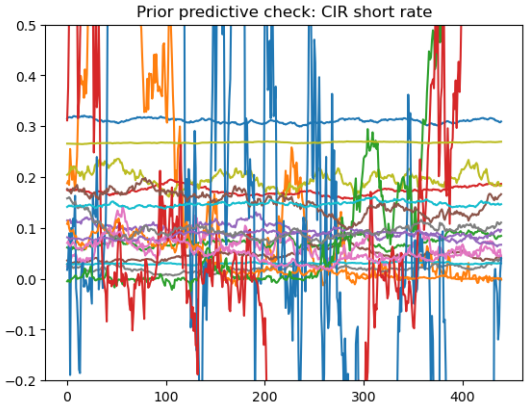

Figure 4: Prior predictive check: samples from CIR short rate

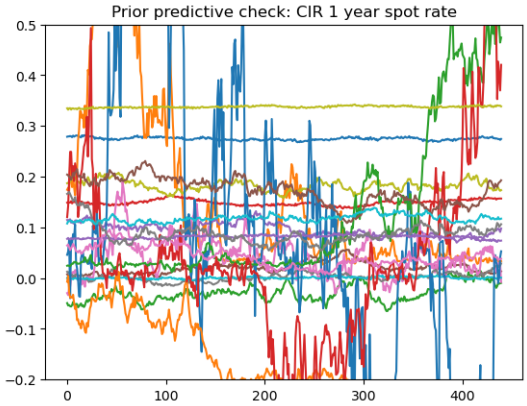

Figure 5: Prior predictive check: samples from CIR 1 year spot rate

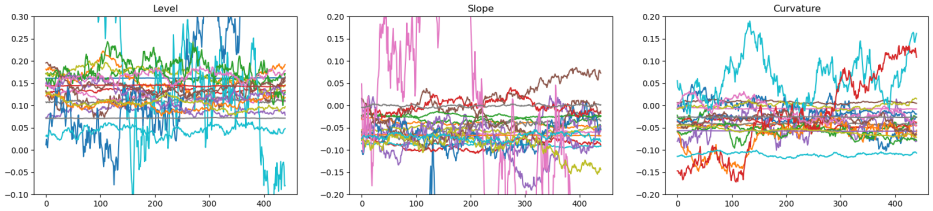

Figure 6: Prior predictive check: samples from Diebold-Li factors

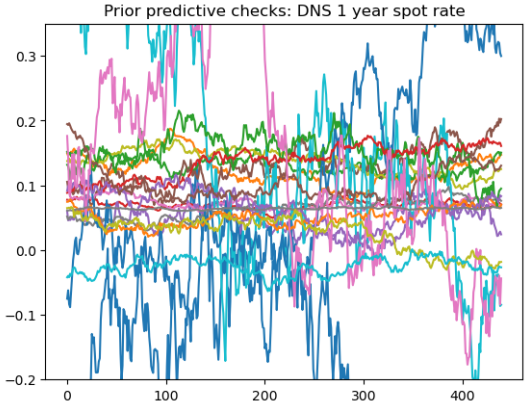

Figure 7: Prior predictive check: samples from Diebold-Li 1 year spot rate

The hyperparameter values for the samples above are

$$\begin{aligned}
\kappa &\sim TN_{(0,\frac{2}{\Delta t})}(0.7, 0.16),\\
\mu &\sim TN_{(0,\infty)}(0.08, 0.01),\\
\sigma &\sim HalfCauchy(0.03),\\
a &\sim N(0.1, 0.01),\\
b &\sim N(0.5, 0.25),\\
\sigma_{obs} &\sim HalfCauchy(0.008)
\end{aligned}$$

for the Vasicek model,

$$\begin{aligned}
\kappa &\sim TN_{(0,\frac{2}{\Delta t})}(0.7, 0.16),\\
\mu &\sim TN_{(0,\infty)}(0.08, 0.01),\\
\sigma &\sim HalfCauchy(0.1),\\
a &\sim N(0.1, 0.01),\\
b &\sim N(0.5, 1),\\
\sigma_{obs} &\sim HalfCauchy(0.008)
\end{aligned}$$

for the CIR model, and

$$\begin{aligned}
\kappa_l &\sim TN_{(0,\frac{2}{\Delta t})}(1, 1),\\
\kappa_s &\sim TN_{(0,\frac{2}{\Delta t})}(0.8, 0.25),\\
\kappa_c &\sim TN_{(0,\frac{2}{\Delta t})}(0.5, 0.16),\\
\mu_l &\sim TN_{(-0.5,\infty)}(0.12, 0.0025),\\
\mu_s &\sim TN_{(-\infty,0.5)}(-0.04, 0.0009),\\
\mu_c &\sim TN_{(-\infty,0.5)}(-0.03, 0.0016),\\
\sigma_{l,s,c} &\sim HalfCauchy(0.015),\\
\sigma_{obs} &\sim HalfCauchy(0.005),\\
\phi_\lambda &\sim U(-1, 1)\\
\mu_\lambda &\sim N(0.3, 0.25),\\
\tau_\lambda &\sim Exp(1)
\end{aligned}$$

for the Diebold-Li models.

3.4 Estimation with simulated data

For the construction of our fake, simulated data, we first choose "true" parameter values and plug those values in the model equations in order to generate fake data. For the simulated data, we choose $T = 440$ and discretization time step $dt = \frac{1}{52}$ in order to replicate the weekly frequency from our observed data. As we also want to assess how our model reacts to sampling uncertainty and learns from the data, we choose true parameter values different to those of the prior means. The chosen true values for the parameters in the simulated data are

$$\begin{aligned}
\kappa &= 0.4,\\
\mu &= 0.05,\\
\sigma &= 0.02,\\
a &= 0.2,\\
b &= 0.3,\\
\sigma_{obs} &= 0.003
\end{aligned}$$

parameter	mean	sd	hdi 0.5%	hdi 99.5%	mcse_{mean}	mcse_{sd}	ess_{bulk}	ess_{tail}	r_{hat}	median
a	0.204	0.002	0.199	0.208	0.000	0.000	1486.0	2443.0	1.0	0.204
b	0.287	0.005	0.275	0.299	0.000	0.000	1325.0	2263.0	1.0	0.287
kappa	0.483	0.256	0.001	1.114	0.006	0.004	1560.0	1152.0	1.0	0.473
mu	0.057	0.028	0.004	0.169	0.001	0.001	1627.0	1174.0	1.0	0.053
sigma	0.020	0.001	0.018	0.022	0.000	0.000	2328.0	2727.0	1.0	0.020
std_{obs}	0.003	0.000	0.003	0.003	0.000	0.000	3904.0	3099.0	1.0	0.003

Table 1: Summary for the Vasicek model with simulated data

for the Vasicek model,

$$\kappa = 0.4,$$
$$\mu = 0.05,$$
$$\sigma = 0.1,$$
$$a = 0.2,$$
$$b = 0.3,$$
$$\sigma_{obs} = 0.003$$

for the CIR model, and

$$\kappa_l = 0.8,$$
$$\kappa_s = 0.6,$$
$$\kappa_c = 0.4,$$
$$\mu_l = 0.11,$$
$$\mu_s = -0.03,$$
$$\mu_c = -0.03,$$
$$\sigma_l = 0.02,$$
$$\sigma_s = 0.03,$$
$$\sigma_c = 0.04,$$
$$\sigma_{obs} = 0.001,$$
$$\lambda = 1.2(static),$$
$$\phi_\lambda = -0.2,$$
$$\mu_\lambda = 0.2,$$
$$\tau_\lambda = 20(dynamic)$$

for the Diebold-Li models. We simulate data for the same maturities as in our observed data.

Our models are fit via Hamiltonian Monte Carlo (HMC) in Python language with the PyMC [SWF16] library. Some of the diagnostics are computed with the ArviZ package [KCHM19]. For each model, we simulate 2 chains with 2000 draws each. We use an Euler-Maruyama discretization scheme for the continuous time processes. We proceed to fit the models and report summary statistics, such as mean, median and standard deviations of posterior distributions. We also report computational diagnostics that shall be detailed in the following subsection.

3.5 Computational diagnostics and estimation bias

Some of the parameter estimates are biased in that they differ from e.g. the posterior mean by several standard deviations, with a few even being outside of the 99% highest density intervals (HDIs). Since our goal is not to obtain unbiased estimates of the parameters, but to illustrate the process of Bayesian data analysis for term structure modeling, we ignore this for now.

Beyond the "traditional" Bayesian computational diagnostics of trace and autocorrelation plots, and the old-fashioned Gelman-Rubin statistic ($\hat{R}$), we report some novel diagnostics. These include Monte Carlo standard error (MCSE), effective sample size (ESS) and the rank normalized version of $\hat{R}$ by [VGS+21], which are detailed below.

parameter	mean	sd	hdi 0.5%	hdi 99.5%	mcse_{mean}	mcse_{sd}	ess_{bulk}	ess_{tail}	r_{hat}	median
a	0.204	0.002	0.200	0.209	0.000	0.000	1040.0	1992.0	1.0	0.204
b	0.287	0.005	0.274	0.299	0.000	0.000	933.0	1769.0	1.0	0.287
kappa	0.472	0.261	0.000	1.112	0.006	0.004	1780.0	1649.0	1.0	0.462
sigma	0.100	0.004	0.089	0.113	0.000	0.000	2157.0	2504.0	1.0	0.100
mu	0.061	0.033	0.001	0.203	0.001	0.001	2063.0	1305.0	1.0	0.053
std_{obs}	0.003	0.000	0.003	0.003	0.000	0.000	4473.0	2869.0	1.0	0.003

Table 2: Summary for the CIR model with simulated data

parameter	mean	sd	hdi 0.5%	hdi 99.5%	mcse_{mean}	mcse_{sd}	ess_{bulk}	ess_{tail}	r_{hat}	median
lambda	1.220	0.031	1.142	1.297	0.002	0.002	197.0	424.0	1.02	1.222
kappa_l	0.580	0.359	0.000	1.501	0.009	0.006	1282.0	1288.0	1.00	0.540
mu_l	0.113	0.020	0.049	0.191	0.001	0.000	2004.0	1481.0	1.00	0.112
sigma_l	0.020	0.001	0.018	0.023	0.000	0.000	4275.0	3498.0	1.00	0.020
kappa_s	0.310	0.195	0.000	0.858	0.003	0.002	3104.0	2015.0	1.00	0.283
mu_s	-0.043	0.024	-0.109	0.019	0.000	0.000	6127.0	2687.0	1.00	-0.042
sigma_s	0.027	0.001	0.025	0.030	0.000	0.000	6304.0	3528.0	1.00	0.027
kappa_c	0.706	0.313	0.005	1.473	0.006	0.004	2275.0	1476.0	1.00	0.696
mu_c	-0.014	0.019	-0.079	0.038	0.000	0.000	3220.0	2270.0	1.00	-0.013
sigma_c	0.034	0.003	0.028	0.042	0.000	0.000	1032.0	2037.0	1.00	0.034
std_{obs}	0.001	0.000	0.001	0.001	0.000	0.000	3072.0	2873.0	1.00	0.001

Table 3: Summary for the Diebold-Li model (static λ) with simulated data

parameter	mean	sd	hdi 0.5%	hdi 99.5%	mcse_{mean}	mcse_{sd}	ess_{bulk}	ess_{tail}	r_{hat}	median
mu_{lamb}	0.209	0.035	0.126	0.306	0.002	0.001	425.0	915.0	1.01	0.209
phi_{lamb}	-0.159	0.084	-0.371	0.050	0.002	0.002	1471.0	2727.0	1.00	-0.160
tau_{lamb}	12.267	1.587	8.805	16.750	0.057	0.040	802.0	1715.0	1.00	12.177
kappa_l	0.570	0.359	0.000	1.504	0.009	0.006	1326.0	1632.0	1.00	0.527
mu_l	0.114	0.021	0.049	0.195	0.001	0.000	1862.0	1577.0	1.00	0.112
sigma_l	0.020	0.001	0.018	0.023	0.000	0.000	3263.0	3289.0	1.00	0.020
kappa_s	0.345	0.208	0.000	0.924	0.004	0.003	2569.0	2058.0	1.00	0.318
mu_s	-0.043	0.023	-0.107	0.014	0.000	0.000	5197.0	2916.0	1.00	-0.042
sigma_s	0.029	0.001	0.025	0.032	0.000	0.000	4573.0	3526.0	1.00	0.029
kappa_c	0.730	0.328	0.011	1.522	0.008	0.005	1707.0	1208.0	1.00	0.725
mu_c	-0.017	0.021	-0.097	0.033	0.000	0.000	2500.0	1635.0	1.00	-0.014
sigma_c	0.035	0.003	0.028	0.043	0.000	0.000	806.0	1643.0	1.01	0.035
std_{obs}	0.001	0.000	0.001	0.001	0.000	0.000	2365.0	3105.0	1.00	0.001

Table 4: Summary for the Diebold-Li model (dynamic λ) with simulated data

3.5.1 Monte Carlo standard error

It is possible to evaluate the precision of the average $\overline{\theta}$ of S independent draws as an estimate of $\mathrm{E}[\theta|y]$ through the Monte Carlo standard error (MCSE). The MCSE can be computed as

$$MCSE = \sqrt{\mathrm{Var}(\overline{\theta})} = \sqrt{\frac{\mathrm{Var}(\theta|y)}{S}},$$

and this can be generalized to the posterior expectation of any function $g(\theta)$. In this work the MCSEs are computed from the effective sample sizes through the procedure described in [VGS+21]. We report the mean and standard deviation estimates for the MCSE.

3.5.2 Effective sample size

The effective sample size (ESS) is a measure of information contained in each sampling chain. A higher sampling autocorrelation means higher sampling uncertainty and therefore smaller ESS. The ESS can be estimated as

$$\widehat{ESS} = \frac{S}{\hat{\tau}},$$
$$\hat{\tau} = -1 + 2\sum_{t'=0}^{K} \hat{P}_{t'},$$
$$\hat{P}_{t'} = \hat{\rho}_{2t'} + \hat{\rho}_{2t'+1},$$

where S is the number of samples and $\hat{\rho}_t$ are autocorrelations estimated at lag t via fast Fourier transform. K is the last integer for which $\hat{P}_K = \hat{\rho}_{2K} + \hat{\rho}_{2K+1}$ is positive.

Besides the mean and standard deviation estimates of ESS, [VGS+21] propose two other versions: Bulk-ESS and Tail-ESS. Bulk-ESS is obtained through rank normalization of the draws in the chain, that is, replacing drawn parameter values with rank normalized values, i.e., normal scores for pooled draws from all chains. Tail-ESS is the minimum of the effective sample sizes of the 5% and 95% quantiles. While Bulk-ESS is useful for assessing problems due to trend behavior of chains, Tail-ESS allows for diagnosing issues related to the scale of chains.

3.5.3 Rank normalized $\hat{R}$

The Gelman-Rubin statistic, or $\hat{R}$, is one of the most common diagnostics for MCMC convergence. We do not use the original Gelman-Rubin statistic, but the rank normalized version from [VGS+21]. The Gelman-Rubin statistic is defined as

$$\hat{R} = \frac{\hat{V}}{W},$$

where W is the within-chain variance and $\hat{V}$ is the posterior variance estimate for the rank normalized pooled traces calculated via the same procedure from Bulk-ESS. If convergence has been achieved, the within-chain and between-chain (pooled) variances should be the same, and the $\hat{R}$ should be equal to 1.0. Values above that, e.g., 1.1 indicate one or more chains have not converged.

3.6 Estimation with real data

We estimate the models with the observed dataset described in subsection 3.1 with the priors specified in subsections 3.2 and 3.3 and the same estimation procedure employed in subsection 3.4. As, with the real world dataset we do not know what the true data-generating process is, we cannot assess estimation bias. We provide tables with the same summary statistics and computational diagnostics as we did with the simulated data. We report the KDE and trace plots for the posterior distributions of the parameters and the estimated latent factors in Figures 8 to 15. We also provide autocorrelation plots from Figures 16 to 19.

We also notice that the autocorrelation plots show the persistence of autocorrelation for the λ parameter of the Diebold-Li models. The autocorrelation seems more persistent for the parameters of the dynamic λ than for the static λ model. The autocorrelation is also persistent for the σ parameters in one of the chains for the dynamic λ model.

parameter	mean	sd	hdi 0.5%	hdi 99.5%	mcse_{mean}	mcse_{sd}	ess_{bulk}	ess_{tail}	r_{hat}	median
a	0.137	0.001	0.133	0.141	0.000	0.000	3339.0	3205.0	1.0	0.137
b	0.181	0.006	0.166	0.197	0.000	0.000	3113.0	3254.0	1.0	0.181
kappa	0.084	0.056	0.000	0.247	0.001	0.001	1835.0	1398.0	1.0	0.078
mu	0.047	0.041	0.000	0.204	0.001	0.001	2452.0	1934.0	1.0	0.038
sigma	0.012	0.001	0.010	0.014	0.000	0.000	1156.0	1837.0	1.0	0.012
std_{obs}	0.006	0.000	0.006	0.006	0.000	0.000	5951.0	3127.0	1.0	0.006

Table 5: Summary for the Vasicek model with real data

parameter	mean	sd	hdi 0.5%	hdi 99.5%	mcse_{mean}	mcse_{sd}	ess_{bulk}	ess_{tail}	r_{hat}	median
a	0.138	0.001	0.134	0.142	0.000	0.000	4523.0	3349.0	1.0	0.138
b	0.179	0.006	0.163	0.195	0.000	0.000	4165.0	3256.0	1.0	0.179
kappa	0.107	0.059	0.000	0.262	0.001	0.001	1808.0	1797.0	1.0	0.105
sigma	0.040	0.002	0.034	0.046	0.000	0.000	989.0	2138.0	1.0	0.040
mu	0.028	0.029	0.000	0.143	0.001	0.000	2549.0	1676.0	1.0	0.021
std_{obs}	0.006	0.000	0.006	0.006	0.000	0.000	4729.0	2668.0	1.0	0.006

Table 6: Summary for the CIR model with real data

parameter	mean	sd	hdi 0.5%	hdi 99.5%	mcse_{mean}	mcse_{sd}	ess_{bulk}	ess_{tail}	r_{hat}	median
lambda	1.396	0.014	1.360	1.432	0.001	0.000	381.0	665.0	1.0	1.396
kappa_{l}	0.590	0.371	0.000	1.572	0.008	0.006	1684.0	1735.0	1.0	0.547
mu_{l}	0.109	0.023	0.031	0.191	0.000	0.000	2648.0	2114.0	1.0	0.109
sigma_{l}	0.025	0.001	0.023	0.028	0.000	0.000	5475.0	3363.0	1.0	0.025
kappa_{s}	0.540	0.288	0.001	1.276	0.006	0.004	2331.0	1662.0	1.0	0.524
mu_{s}	-0.033	0.019	-0.097	0.019	0.000	0.000	5084.0	2780.0	1.0	-0.031
sigma_{s}	0.027	0.001	0.025	0.030	0.000	0.000	4849.0	3303.0	1.0	0.027
kappa_{c}	0.330	0.204	0.000	0.878	0.004	0.003	2316.0	2258.0	1.0	0.305
mu_{c}	-0.041	0.030	-0.122	0.049	0.000	0.000	5453.0	2336.0	1.0	-0.041
sigma_{c}	0.034	0.002	0.030	0.039	0.000	0.000	2108.0	2878.0	1.0	0.034
std_{obs}	0.001	0.000	0.001	0.001	0.000	0.000	2468.0	3018.0	1.0	0.001

Table 7: Summary for the Diebold-Li model (static λ) with real data

parameter	mean	sd	hdi 0.5%	hdi 99.5%	mcse_{mean}	mcse_{sd}	ess_{bulk}	ess_{tail}	r_{hat}	median
mu_{lamb}	0.300	0.039	0.202	0.392	0.005	0.004	61.0	137.0	1.04	0.301
phi_{lamb}	0.946	0.018	0.898	0.987	0.001	0.000	861.0	1406.0	1.01	0.947
tau_{lamb}	27.039	3.452	19.269	37.282	0.280	0.199	154.0	389.0	1.02	26.885
kappa_{l}	0.350	0.248	0.000	1.054	0.005	0.004	1487.0	1193.0	1.00	0.307
mu_{l}	0.112	0.027	0.024	0.201	0.001	0.000	2836.0	1876.0	1.00	0.112
sigma_{l}	0.018	0.001	0.016	0.021	0.000	0.000	16.0	28.0	1.08	0.018
kappa_{s}	0.300	0.185	0.000	0.814	0.004	0.003	1722.0	1544.0	1.00	0.273
mu_{s}	-0.044	0.021	-0.109	0.006	0.000	0.000	4507.0	2490.0	1.00	-0.042
sigma_{s}	0.019	0.001	0.016	0.021	0.000	0.000	14.0	32.0	1.09	0.019
kappa_{c}	0.440	0.245	0.001	1.060	0.004	0.003	2557.0	1869.0	1.00	0.420
mu_{c}	-0.040	0.029	-0.123	0.040	0.000	0.000	4931.0	2403.0	1.00	-0.041
sigma_{c}	0.044	0.003	0.038	0.052	0.001	0.001	12.0	20.0	1.11	0.044
std_{obs}	0.001	0.000	0.001	0.001	0.000	0.000	824.0	1774.0	1.00	0.001

Table 8: Summary for the Diebold-Li model (dynamic λ) with real data

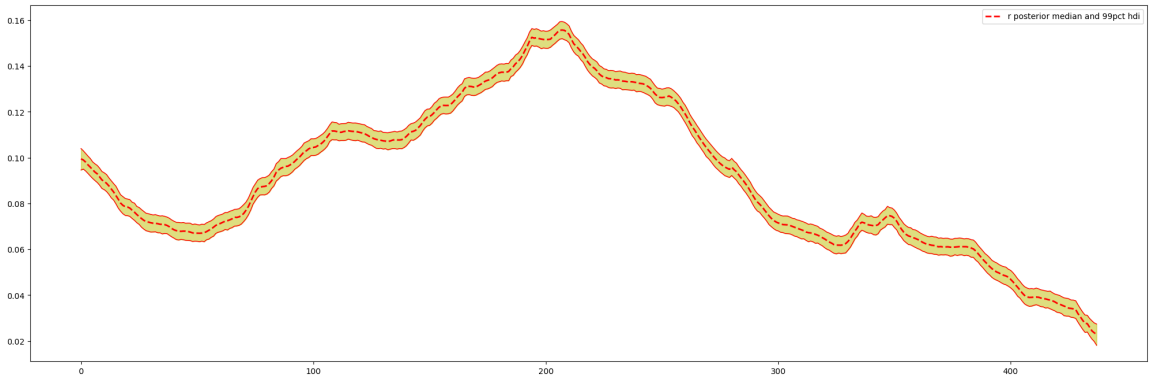

Figure 8: Estimated short rate for the Vasicek model

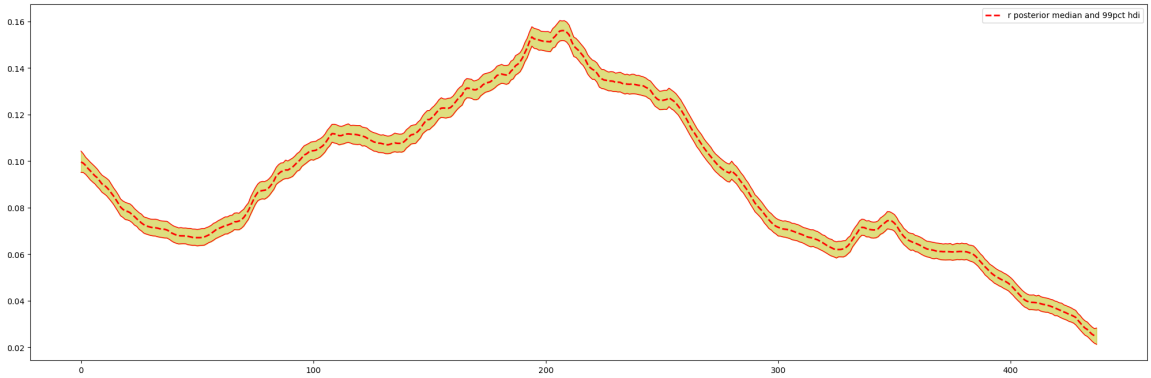

Figure 9: Estimated short rate for the CIR model

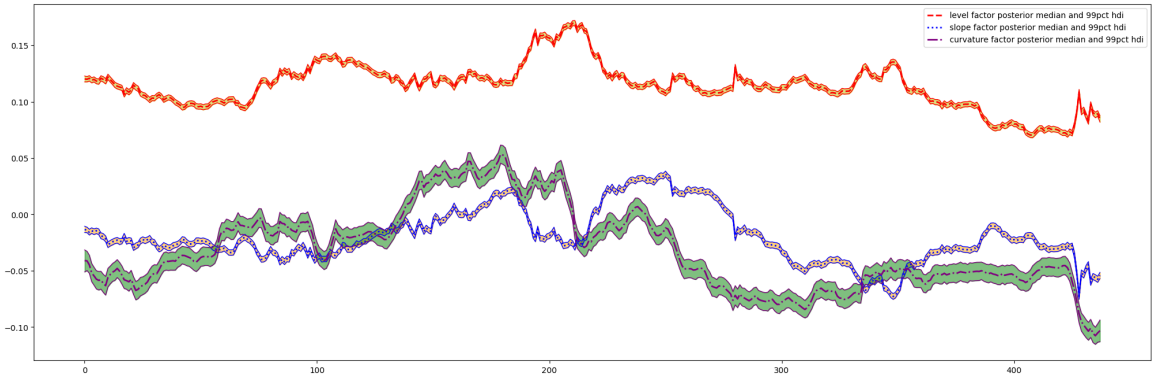

Figure 10: Estimated factors for the Diebold-Li model (static λ)

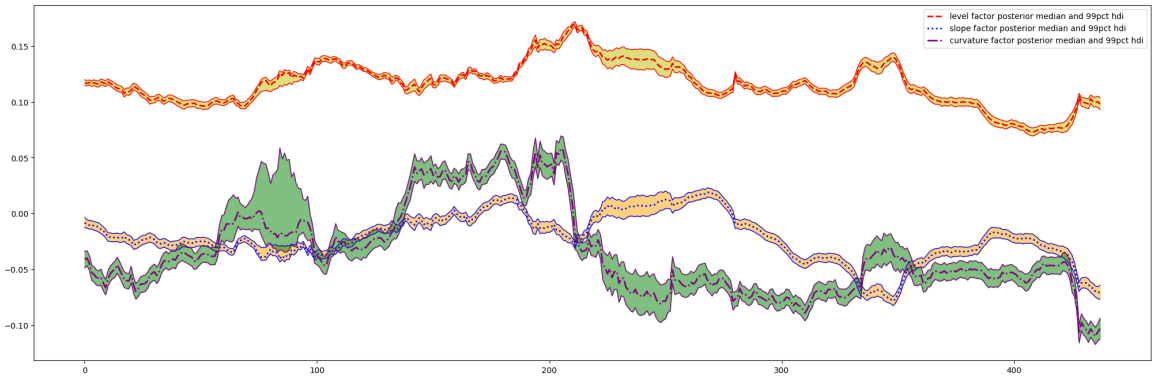

Figure 11: Estimated factors for the Diebold-Li model (dynamic λ)

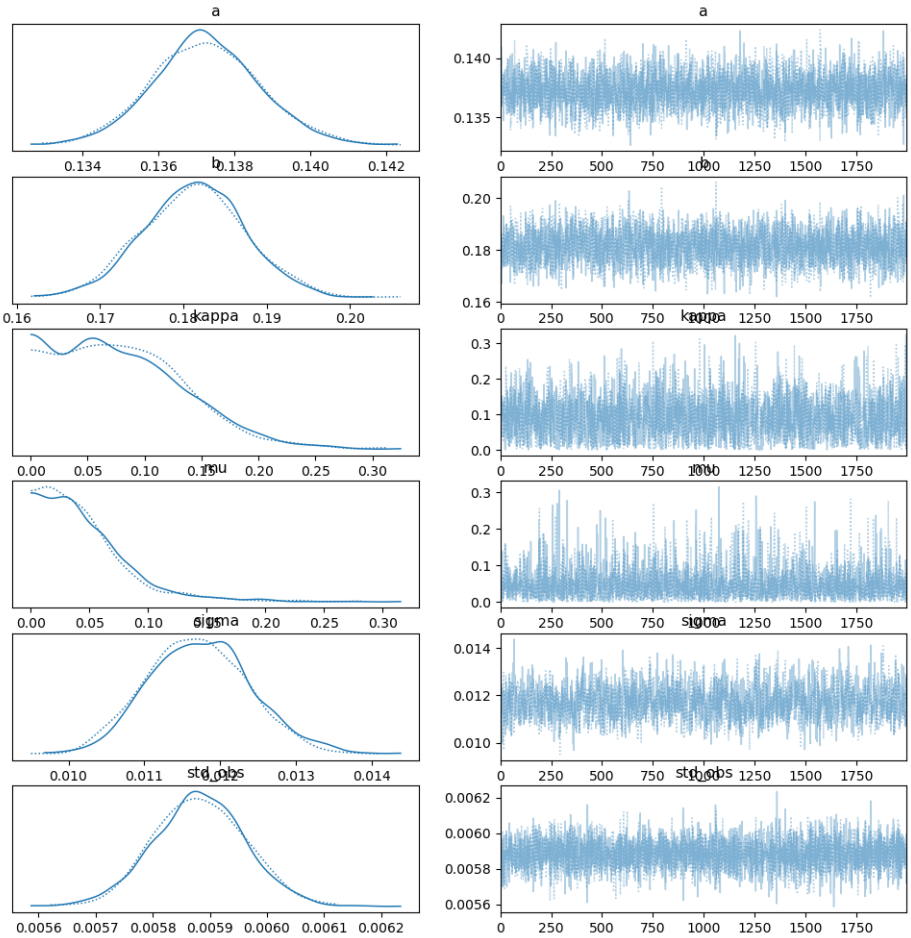

Figure 12: Trace and KDE plots for the Vasicek model

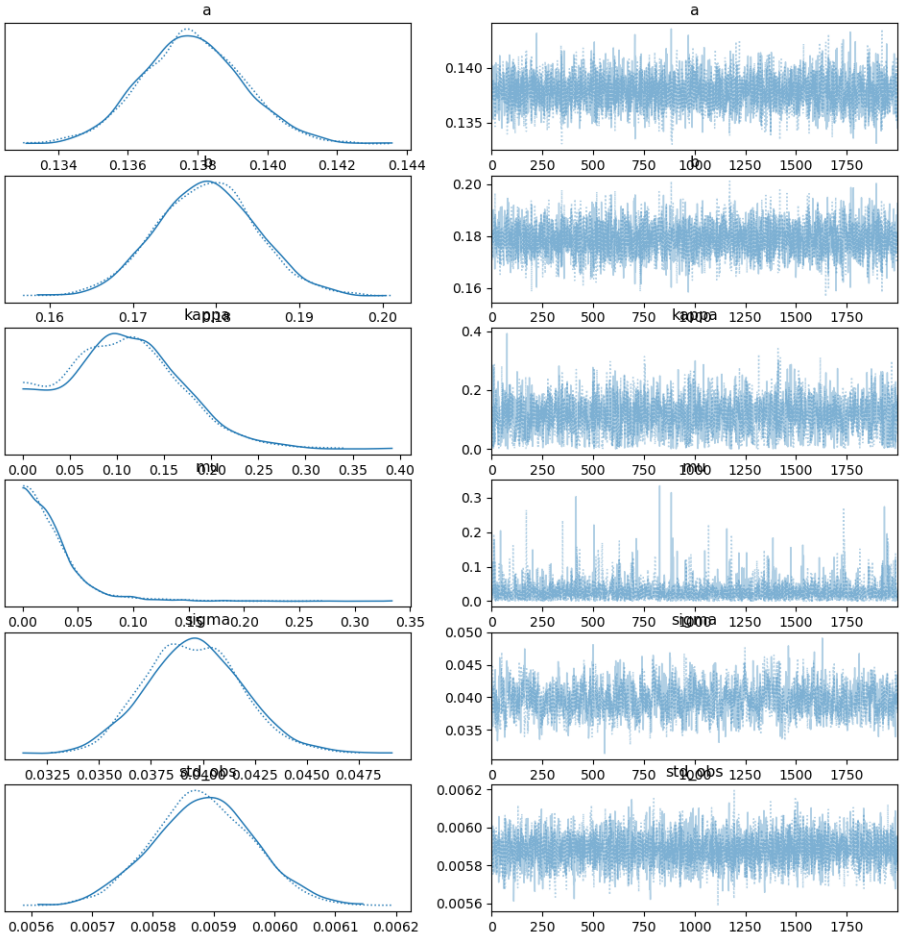

Figure 13: Trace and KDE plots for the CIR model

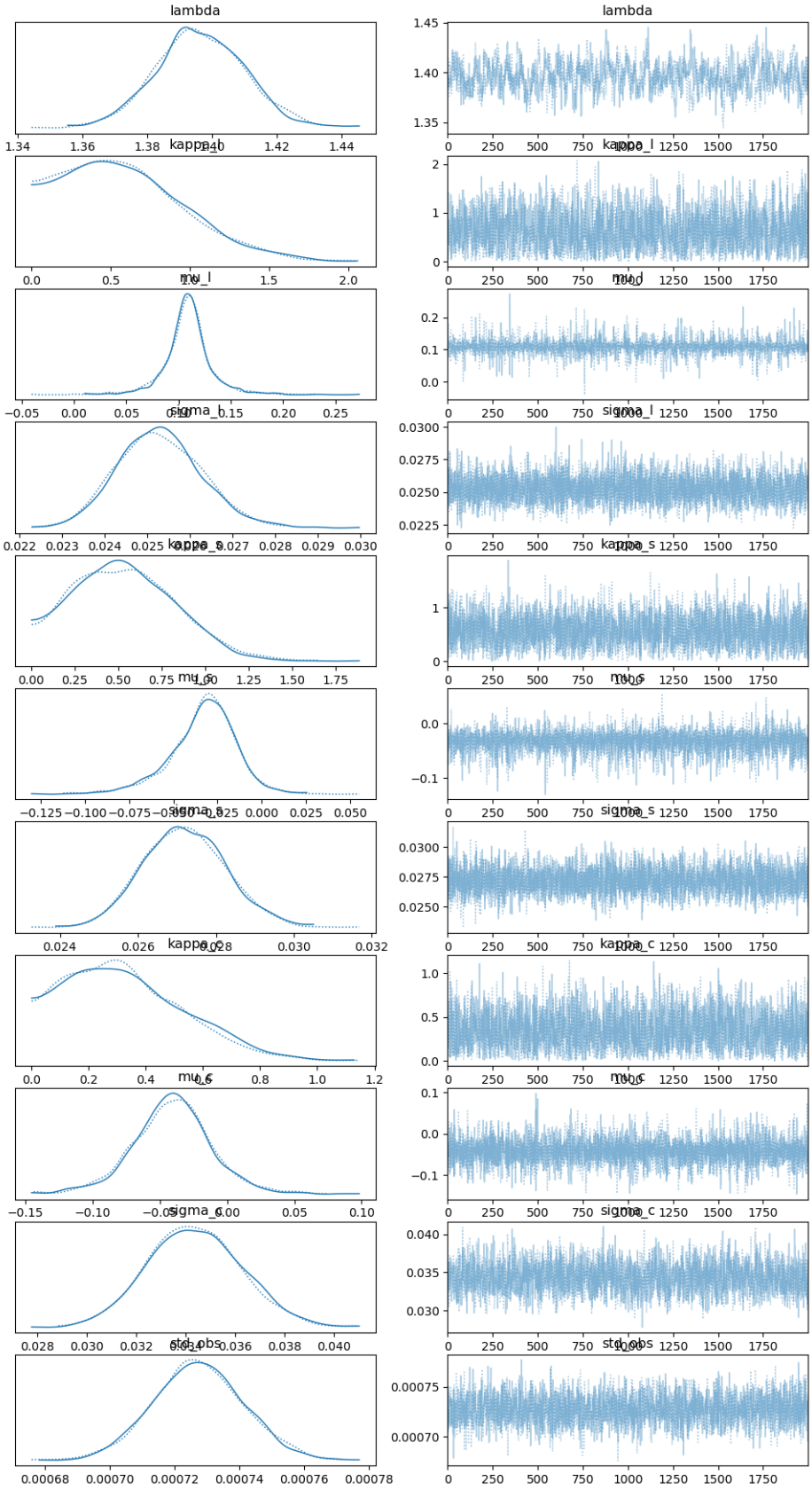

Figure 14: Trace and KDE plots for the Diebold-Li model (static λ)

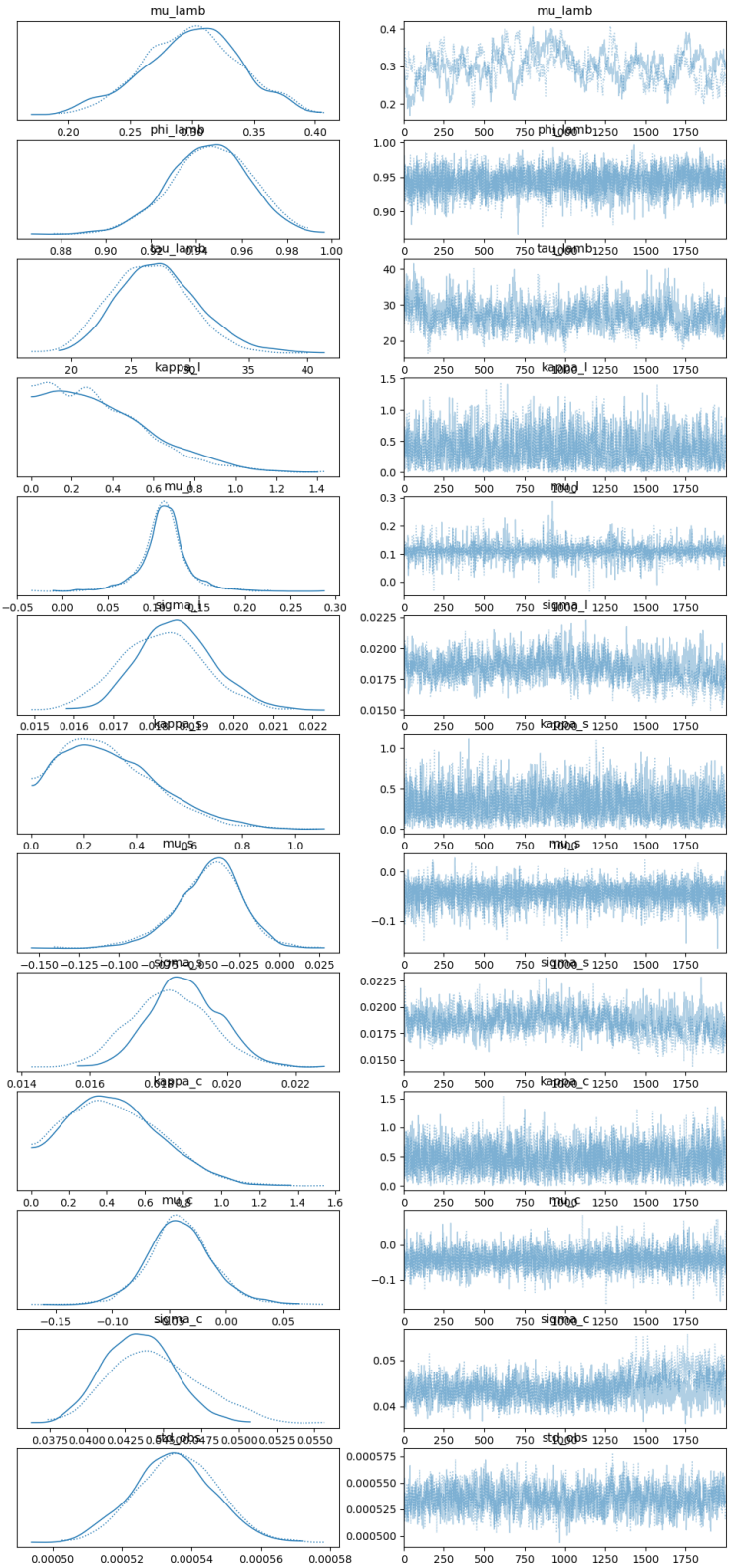

Figure 15: Trace and KDE plots for the Diebold-Li model (dynamic λ)

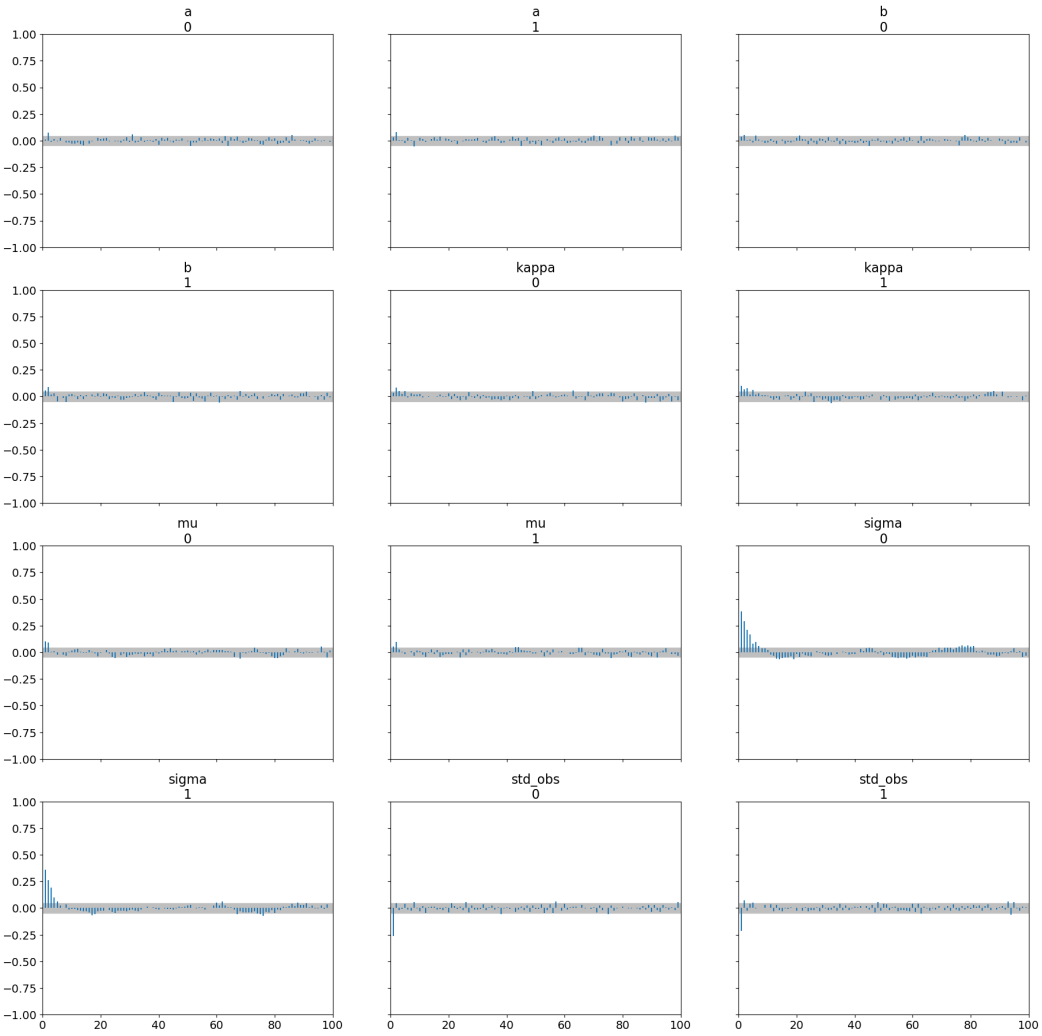

Figure 16: Autocorrelation plots for the Vasicek model

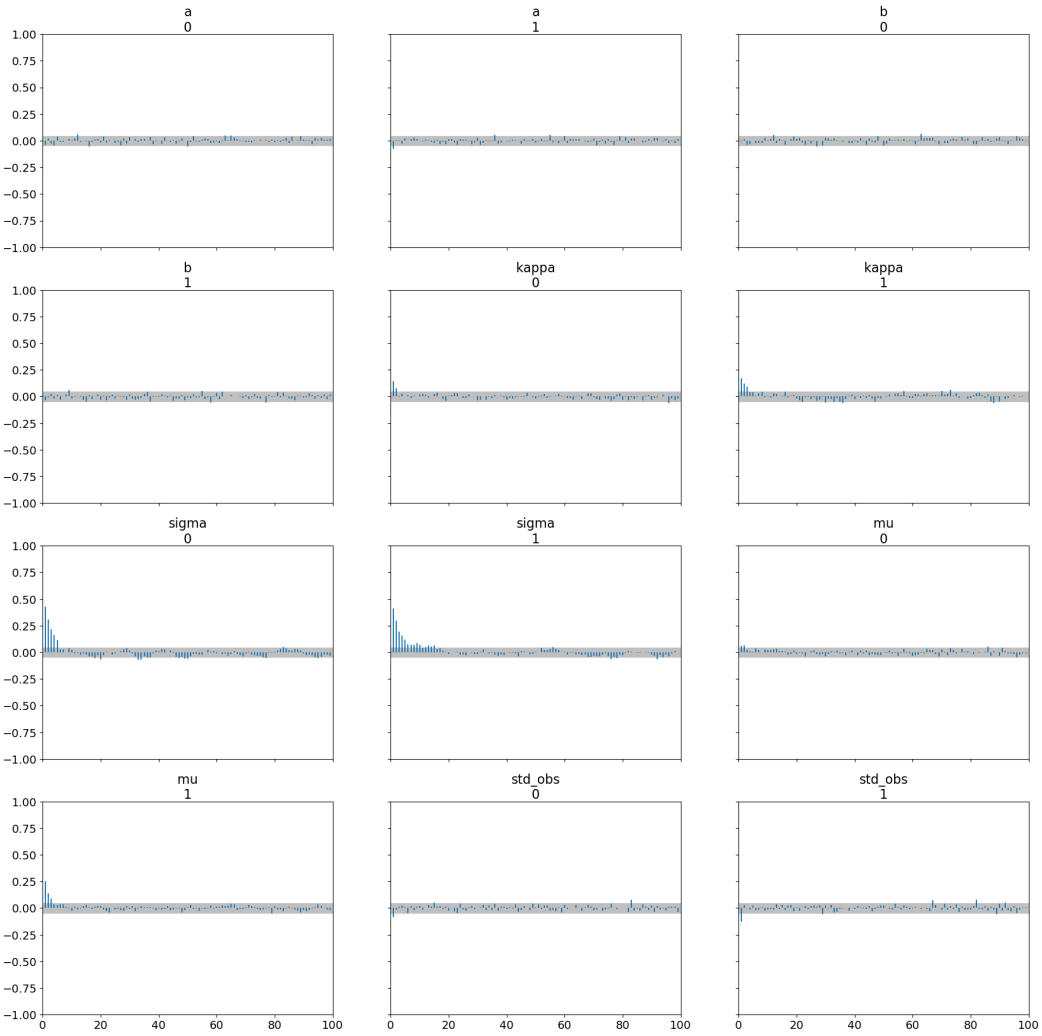

Figure 17: Autocorrelation plots for the CIR model

Figure 18: Autocorrelation plots for the Diebold-Li model (static λ)

Figure 19: Autocorrelation plots for the Diebold-Li model (dynamic λ)

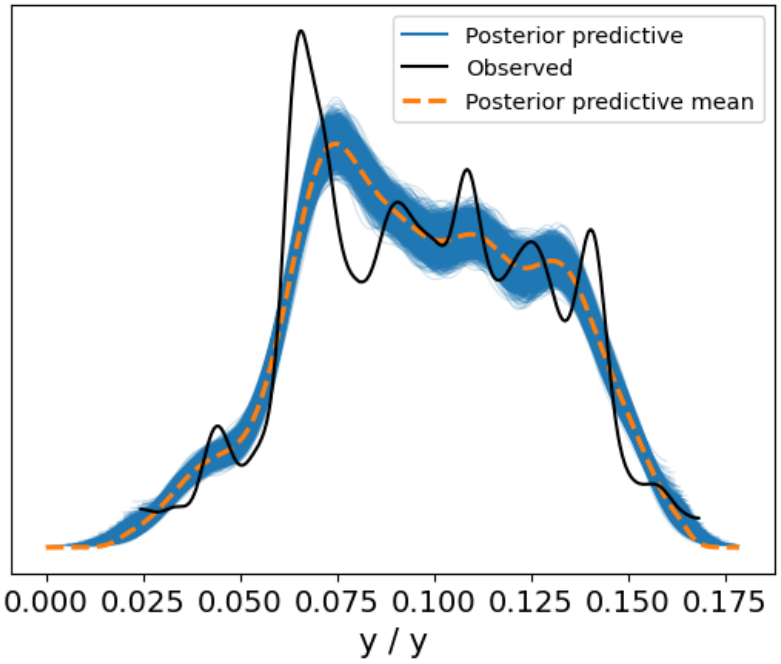

Figure 20: Posterior predictive check for the Vasicek model

3.7 Evaluating predictive accuracy

Our first measure of model predictive accuracy are posterior predictive checks. This graphical procedure is suggested by [GSV+19] and consists of simulating data from the model posterior predictive distribution and plotting the KDE of each simulation together with the KDE for the observed data. If the model has a good fit to the observed data, then it is possible to simulate data from the posterior predictive that closely resembles the observed data. The authors recognize this procedure is mostly qualitative.

Our posterior predictive checks, shown in Figures 20 to 23, indicate that the Diebold-Li models are able to produce data that better resembles the observed data rather than the affine (Vasicek and CIR) models.

Besides that, we also present the results of a cross validation procedure, leave-one-out cross validation (LOO-CV), and an information criterion, the Watanabe-Akaike information criterion (WAIC) [VGG17]. The WAIC and LOO-CV take into acount the posterior distribution of the Bayesian model, unlike, e.g., AIC or BIC. Both measures require estimating the expected log pointwise predictive density (elpd). [VGG17] devise a new procedure for estimating the LOO-CV elpd, via Pareto smoothed importance sampling (PSIS). The PSIS-LOO elpd is computed as

$$\widehat{elpd}_{PSIS-LOO} = \sum_{i=1}^{n} \ln\left(\frac{\sum_{s=1}^{S} w_i^s p(y_i|\theta^s)}{\sum_{s=1}^{S} w_i^s}\right),$$

where w_i^s are weights obtained from the PSIS procedure and θ^s are samples from the posterior, for $s = 1, \ldots, S$ samples in total. Per [VGG17], the WAIC for the elpd is estimated as

$$\widehat{elpd}_{WAIC} = \widehat{lpd} - \hat{p}_{WAIC},$$

where $\widehat{lpd}$ is the computed log pointwise predictive density estimated as

$$\widehat{lpd} = \sum_{i=1}^{n} \ln\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i|\theta^s)\right),$$

for the n data points, and the estimated effective number of parameters $\hat{p}_{WAIC}$ as

$$\hat{p}_{WAIC} = \sum_{i=1}^{n} V_{s=1}^{S}(\ln p(y_i|\theta^s)),$$

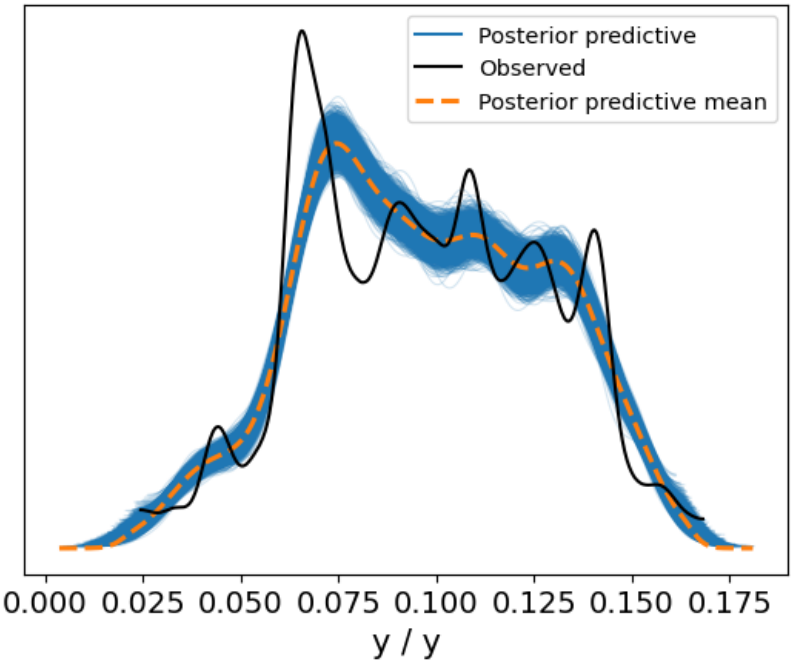

Figure 21: Posterior predictive check for the CIR model

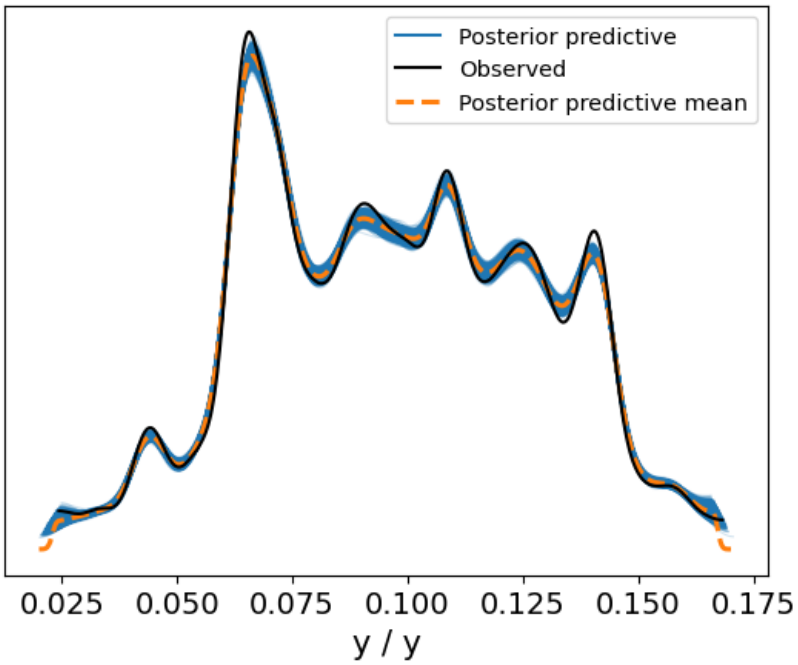

Figure 22: Posterior predictive check for the Diebold-Li model (static λ)

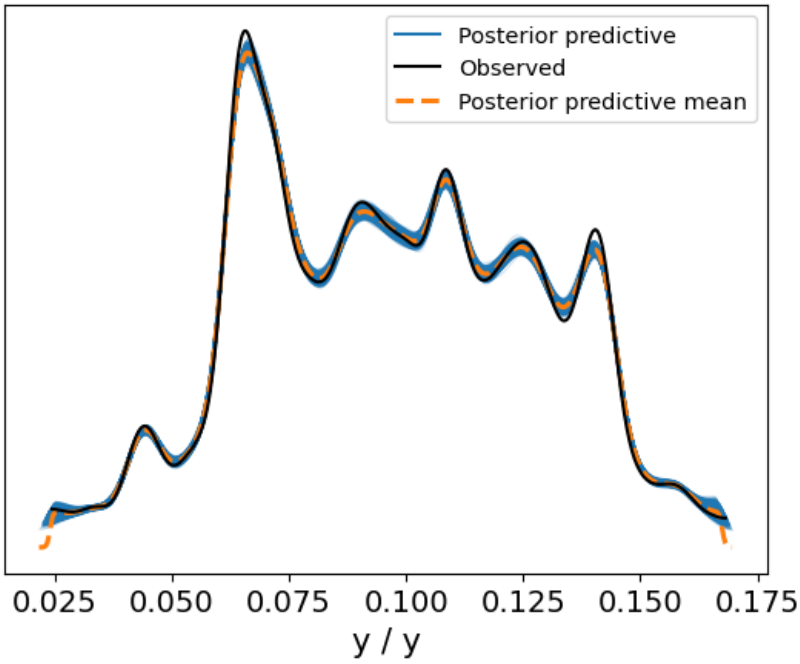

Figure 23: Posterior predictive check for the Diebold-Li model (dynamic λ)

where $V_{s=1}^{S}$ is the sample variance. The LOO-CV and WAIC are calculated for the models and presented in Figures 24 and 25. On the horizontal axis of the plots is the calculated value for the LOO-CV/WAIC on the logarithmic scale. It is worth of notice that we use the Pareto smoothed importance sampling (PSIS) version of the LOO-CV and that both LOO-CV and WAIC are asymptotically convergent.[2] The Diebold-Li models both rank above the affine models, with the dynamic λ model performing the best.

4 Concluding remarks

As pointed out by [Sim10], many economists see Bayesian inference as just a set of practical tools, and not as a different way of statistical reasoning. Bayesian data analysis provides a novel approach for tackling each of the steps in statistical modeling of empirical data. Our process of model building, inference and comparison shows how problems in financial econometrics, such as term structure modeling, can benefit from the Bayesian workflow approach.

Our main results that the Diebold-Li outperforms the Vasicek and CIR models are themselves hardly surprising. We are just using novel tools in Bayesian computation developed over the last years to confirm the canonical result [LS91] obtained more than three decades ago that three factors suffice at describing the term structure. However, this was not the goal of this work. Our main objective was to show how the Bayesian workflow of [GVS+20] can be applied to financial econometrics.

As for further directions, now that we have shown how Bayesian data analysis can be applied to term structure modeling, the same approach can be used for evaluating other models, such as multifactor affine models [DK96] an arbitrage-free version of the Diebold-Li model [CDR11], or a term structure model with macroeconomic variables [RW03]. The Bayesian data analysis can even lead us to think how would our results change with yield curve data for another countries. For a term structure like that of Japan, we could hardly rule out negative rates, as an example. We would have to adapt our models accordingly, and the Bayesian workflow might help us do just that.

[2]We also note the ArviZ Python package used for calculating LOO-CV and WAIC has given out warnings that the estimated shape parameter of Pareto distribution of the PSIS-LOO is greater than 0.7 for one or more samples, for the two Diebold-Li, and the the posterior variance of log predictive densities for the WAIC exceeds 0.4 for one or more samples.

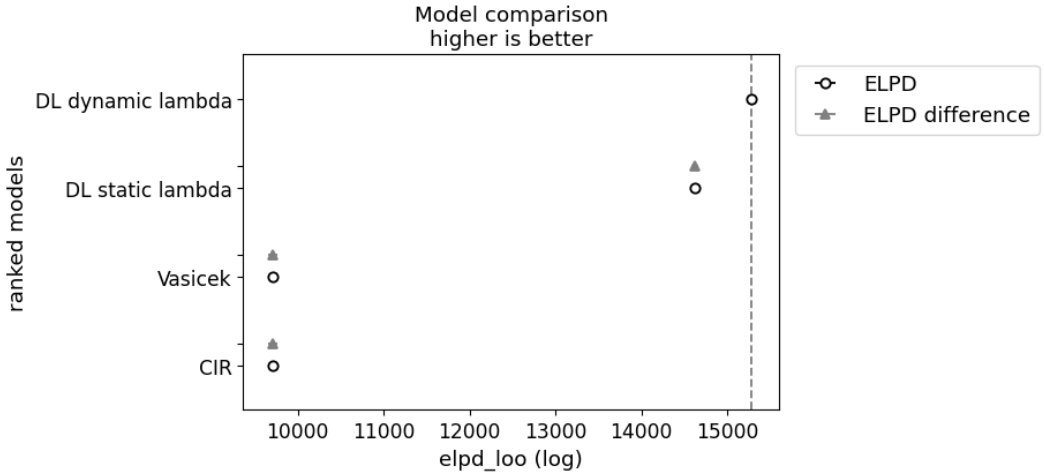

Figure 24: Leave-one-out cross validation

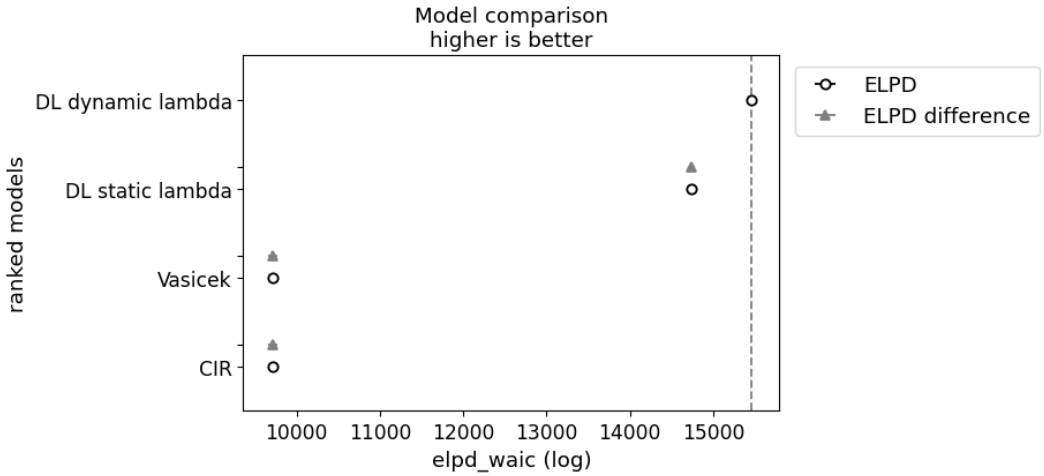

Figure 25: Watanabe-Akaike information criterion

References

[CDR11] Jens HE Christensen, Francis X Diebold, and Glenn D Rudebusch. The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics*, 164(1):4–20, 2011.

[CIR85] John C Cox, Jonathan E Ingersoll, and Stephen A Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.

[DK96] Darrell Duffie and Rui Kan. A yield-factor model of interest rates. *Mathematical finance*, 6(4):379–406, 1996.

[DL06] Francis X Diebold and Canlin Li. Forecasting the term structure of government bond yields. *Journal of econometrics*, 130(2):337–364, 2006.

[GSV+19] Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):389–402, 2019.

[GVS+20] Andrew Gelman, Aki Vehtari, Daniel Simpson, Charles C Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. Bayesian workflow. *arXiv preprint arXiv:2011.01808*, 2020.

[JP10] Michael Johannes and Nicholas Polson. MCMC methods for continuous-time financial econometrics. In *Handbook of Financial Econometrics: Applications*, pages 1–72. Elsevier, 2010.

[KCHM19] Ravin Kumar, Colin Carroll, Ari Hartikainen, and Osvaldo Martin. ArviZ a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software*, 4(33):1143, 2019.

[LMN15] E. Lindström, H. Madsen, and J.N. Nielsen. *Statistics for Finance*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2015.

[LS91] Robert Litterman and Jose Scheinkman. Common factors affecting bond returns. *Journal of fixed income*, 1(1):54–61, 1991.

[Pia10] Monika Piazzesi. Affine term structure models. In *Handbook of financial econometrics: Tools and Techniques*, pages 691–766. Elsevier, 2010.

[PS12] Nicholas G Polson and James G Scott. On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4):887–902, 2012.

[RW03] G Rudebusch and Tao Wu. A no-arbitrage model of the term structure and the macroeconomy. *manuscript, Federal Reserve Bank of San Francisco*, 2003.

[Sim10] Chris Sims. Understanding non-Bayesians. *Unpublished chapter, Department of Economics, Princeton University*, 2010.

[SWF16] John Salvatier, Thomas V Wiecki, and Christopher Fonnesbeck. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55, 2016.

[Vas77] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188, 1977.

[VGG17] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, 27(5):1413–1432, 2017.

[VGS+21] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian analysis*, 1(1):1–28, 2021.