UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Comunicaciones en Estadística
Diciembre 2023, Vol. 16, No. 2, pp. 68–88

Bayesian Time-Varying Modeling of Time Series Data

Modelado Bayesiano Variable en el Tiempo de Datos de Series Temporales

Juan Sosa[a]
jcsosam@unal.edu.co

Abstract

In this paper, we illustrate in-depth several Bayesian models for time series data. To do so, we consider a dataset about weekly change series of the U.S. 3-year Treasury constant maturity interest rate from March 18, 1988 to September 10, 1999. We consider two fully Bayesian approaches: One of them static, based on an retrogressive model of order 3 as in Prado and West (2010), and other two time-varying models, based on the dynamic framework given in West and Harrison (1999). One of these dynamic models is specially design for detecting outliers. Both alternatives are sensible ones, but due to the nature of this dataset, the dynamic modeling approach is more appealing since it gives a complete characterization of the response.

***Keywords*:** Autoregressive models; Bayesian inference; time series data; time-varying modeling..

Resumen

En este documento, ilustramos en profundidad varios modelos Bayesianos para datos de series temporales. Para ello, consideramos un conjunto de datos sobre la serie de cambios semanales de la tasa de interés de la madurez constante de los bonos del Tesoro de EE. UU. a 3 años, desde el 18 de marzo de 1988 hasta el 10 de septiembre de 1999. Consideramos dos enfoques completamente Bayesianos: uno de ellos estático, basado en un modelo autorregresivo de orden 3 como en Prado and West (2010), y otros dos modelos de variación temporal, basados en el marco dinámico de West and Harrison (1999). Uno de estos modelos dinámicos está especialmente diseñado para detectar valores atípicos. Ambas alternativas son sensatas, pero debido a la naturaleza de este conjunto de datos, la alternativa de modelamiento dinámico es más atractiva ya que proporciona una caracterización

[a]Departamento de Estadística, Universidad Nacional de Colombia

completa de la respuesta.

***Palabras clave*:** Modelos autorregresivos; inferencia Bayesiana; datos de series temporales; modelado de variación temporal..

1. Introduction

The aim of our study is to demonstrate the effectiveness of two robust Bayesian methodologies for modeling time series data. One approach employs a static model, specifically a third-order autoregressive model as described in Prado and West (2010). The other utilizes a time-varying framework outlined in West and Harrison (1999), which offers dynamic capabilities. Notably, one of these dynamic models is tailored for outlier detection, enhancing its utility in anomaly identification. While both approaches are valid, the dynamic model is particularly compelling for our dataset due to its ability to provide a comprehensive characterization of the response variable throughout time.

On the one hand, Bayesian autoregressive models represent a powerful tool for analyzing time series data within a probabilistic framework. These models extend the classical autoregressive models by incorporating Bayesian principles, allowing for the inclusion of prior beliefs and uncertainty in model parameters. The key idea behind Bayesian autoregressive models is to model each observation in a time series as a function of its past observations, where the dependencies are captured by autoregressive terms. Unlike traditional frequentist approaches, Bayesian methods offer flexibility in handling complex data patterns and incorporating prior information, making them particularly useful when dealing with small or sparse datasets. By leveraging Bayesian inference techniques such as Markov chain Monte Carlo (MCMC) sampling, these models can provide not only point estimates but also full posterior distributions of parameters, enabling more nuanced interpretation and uncertainty quantification in time series analysis.

On the hand, Bayesian time-varying autoregressive (TVAR) models, as developed within the framework of West and Harrison (1999), offer a sophisticated approach to modeling time series data by allowing parameters to evolve over time. Unlike traditional autoregressive models where coefficients remain constant, TVAR models recognize that relationships in time series data can change dynamically. West and Harrison's framework integrates Bayesian inference to estimate time-varying parameters, enabling the modeling of non-stationary processes and capturing shifts in relationships between variables over different time periods. This flexibility is crucial in applications such as econometrics, where economic relationships can evolve due to changing policies or market conditions. Again, by employing techniques like stochastic search variable selection (SSVS) and Markov chain Monte Carlo (MCMC) methods, Bayesian TVAR models can identify significant time-varying patterns and provide probabilistic forecasts that account for uncertainty in both past observations and future predictions. This makes them a powerful tool for researchers and practitioners seeking to understand and predict complex temporal

dynamics in various fields including economics, social sciences, and environmental studies.

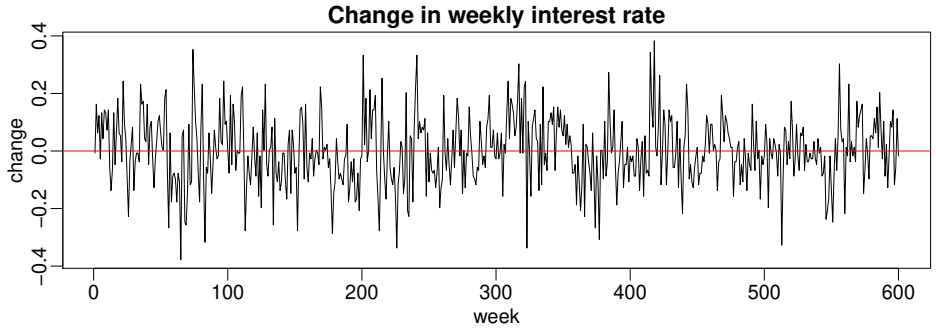

Figure 1: Time series of the detrended weekly change series of the U.S. 3-year Treasury constant maturity interest rate from March 18, 1988 to September 10, 1999.

Here, we illustrate our modeling strategies using the detrended weekly change series of the U.S. 3-year Treasury constant maturity interest rate from March 18, 1988 to September 10, 1999. The corresponding plot of the data y_t after subtracting the mean is shown in Figure 1. We have $T = 600$ equally spaced measurements in total. This series does not seem to exhibit any trend and appear to be stationary (which is confirm below), so we decide to not differentiate the data again.

This paper is structured as follows. In Section 2, we fit an autoregressive model of order 3 to out time series data. In Section 3, we go beyond our previous approach and consider a regular time-varying model. In Section 4, we extend the time-varying model in order to detect outliers in the series. Finally, in Section 7, we discuss our main findings.

2. Autoregressive model

Firstly, we consider the model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} \mathsf{N}(0, v) \tag{1}$$

which corresponds to a AR model with order $p = 3$. We fit a Bayesian AR(3) model to the data using a prior of the form $p(\boldsymbol{\phi}, v) \propto 1/v$ with $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)'$, and the conditional likelihood

$$p(\boldsymbol{y} \mid y_{1:p}) = \prod_{t=p+1}^{T} N(y_t \mid \mathbf{f}_t'\boldsymbol{\phi}, v) = \mathsf{N}(\boldsymbol{y} \mid \mathbf{F}'\boldsymbol{\phi}, v\mathbf{I}_n) \tag{2}$$

where $\boldsymbol{y} = (y_{p+1}, \ldots, y_T)'$ (the first three observations being conditioned upon for initial values), $\mathbf{f}_t = (y_{t-1}, y_{t-2}, y_{t-3})'$, $t = (p+1) : T$, $\mathbf{F} = [\mathbf{f}_{p+1}, \ldots, \mathbf{f}_T]$, and $\mathbf{I}_n$ is

the $n \times n$ identity matrix with $n = T - p$. Under this conditions, we are dealing with a linear model and we can apply standard theory (Prado and West, 2010, pp. 19-22).

Combining likelihood 2 with the reference prior $p(\boldsymbol{\phi}, v)$, we are able to obtain samples from the posterior distribution of $(\boldsymbol{\phi}, v)$ using direct sampling as follows: first, sample v from a $\mathsf{IG}((n-p)/2, (n-p)s^2/2)$, and then for each v in the previous step sample $\boldsymbol{\phi}$ from a $\mathsf{N}(\hat{\boldsymbol{\phi}}_{\text{mle}}, v(\mathbf{F}\mathbf{F}')^{-1})$, where $(n-p)s^2$ is the residual sum of squares (which leads to $\hat{v}_{\text{mle}} = 0.013$ the MLE of v), $\hat{\boldsymbol{\phi}}_{\text{mle}} = (\mathbf{F}\mathbf{F}')^{-1}\mathbf{F}\boldsymbol{y} = (0.227, 0.006, 0.113)'$ is the MLE of $\boldsymbol{\phi}$, and $n = T - p = 597$. Posterior summaries for 4,000 samples from $p(\boldsymbol{\phi}, v \mid \boldsymbol{y})$ are provided in Table 1. We see that the MLEs and the posterior estimates of $\boldsymbol{\phi}$ and v are very similar (this is expected since we are considering a non-informative prior).

Parameter	ϕ_1	ϕ_2	ϕ_3	v
Mean	0.227	0.004	0.113	0.013
SD	0.041	0.042	0.041	0.001
Q2.5 %	0.146	-0.080	0.031	0.012
Q97.5 %	0.308	0.086	0.192	0.014

Table 1: Posterior summaries of $\boldsymbol{\phi}$ and v.

Now, we fix $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}_{\text{bayes}} = (0.227, 0.004, 0.113)'$ to explore the model based on this posterior estimate. The corresponding characteristic polynomial is $\Phi(u) = 1 - 0.227u - 0.004u^2 - 0.113u^3$ leads to one pairs of complex roots and one real root, whose corresponding modulus and wavelength (r_i, λ_i) (in order of decreasing modulus) are: $(0.575, \text{NA})$ and $(0.443, 3.182)$. Note that the largest modulus corresponds to the real root and that the high frequency term is capturing short run

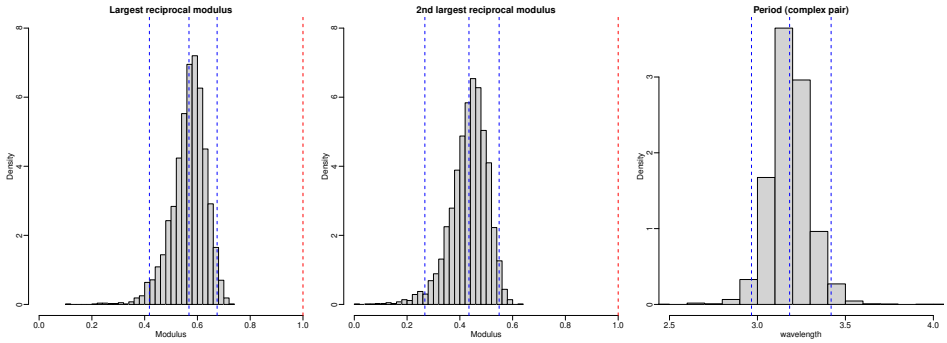

Figure 2: Left and middle: histogram of the posterior samples of the largest reciprocal characteristic modulus (left) and histogram of the posterior samples of the second largest reciprocal characteristic modulus (right). The posterior mean (blue), 2.5 % and 97.5 % quantiles (also blue), and the number 1 (red) are shown in each histogram. Right: histogram of the posterior samples of the period associated to the complex reciprocal pair. The posterior mean and 2.5 % and 97.5 % quantiles are also shown in blue.

oscillations of low magnitude rather than meaningful cyclical components.

Then, we consider posterior inference about the modulus and the period of the reciprocal roots. As in the previous paragraph, almost every posterior sample (3,996 out of 4,000) of $\boldsymbol{\phi}$ gives one pair of complex roots and one real root; in each one of those cases the modulus of the real root happens to be greater than the modulus associated with the complex pair. In summary, the largest modulus corresponds to the modulus of the real root whereas the de second largest modulus corresponds to the modulus of the complex pair.

The left and middle panels of Figure 2 display the posterior distribution of the largest and second largest reciprocal characteristic modulus plotted in the same scale. These distributions are centered at 0.568 and 0.434, respectively, which is consistent with what we get by fixing $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}_{\text{bayes}}$. Moreover, we see that the

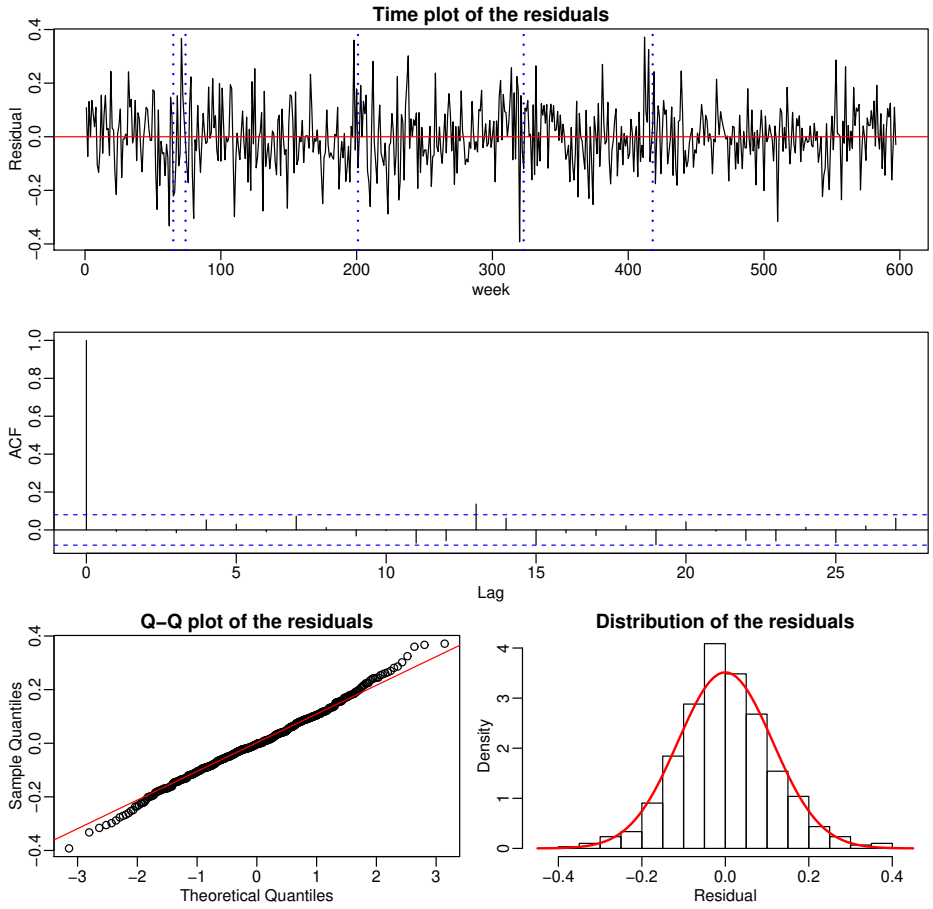

Figure 3: Residual analysis: Time series plot (top), ACF (middle), Q-Q plot (bottom left), and histogram (bottom right) of the residuals obtained from fitting model 1.

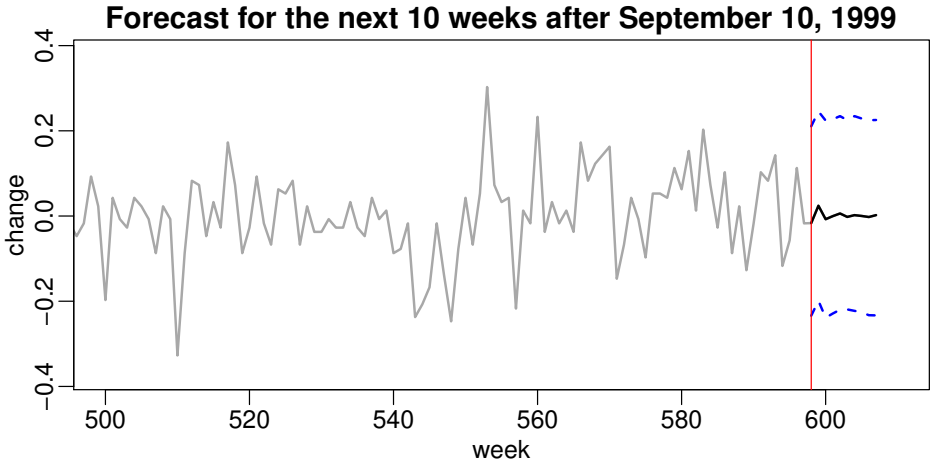

Figure 4: Forecasts (thick black) of the change of the U.S. 3-year Treasury constant maturity interest rate after the week of September 10, 1999 (dotted line in red) along with 95 % credible intervals (dashed lines in blue).

posterior probability that the largest reciprocal characteristic root is less than 1 is 100 % (no value was above the unity which corresponds to the red line in the histogram); the same probability for the second largest reciprocal characteristic root is also 100 %. These quantities confirm that the process is stationary. Finally, for completeness we include in the last panel of Figure 2 the posterior distribution of the period associated to the complex reciprocal pair. As expected it is centered at 3.183, which corresponds to the period associated to the complex pair of reciprocal roots above.

Again, we fix $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}_{\text{bayes}} = (0.227, 0.004, 0.113)'$ but this time to perform a residual analysis of the model. Figure 3 shows a residual analysis of the residuals after fitting model 1. Note that in general the series of the residuals does not exhibit either a persistent trend or a seasonal pattern, but around certain times, in particular around weeks 65, 74, 201, 323, 418 (lines in blue), where residuals take the largest values, we recognize an explicit pattern which clearly suggests that the model is not able to detect outliers effectively. Even though, the ACF does not reveal any significant correlation in any lag of the series of the residuals, looking at the Q-Q plot and the histogram of the residuals, because of the heavy tails we suspect that the residuals could not correspond to a exactly i.i.d. sample from a normal distribution.

We perform the forecasting by successively simulating future values y_{T+h} according to a normal distribution with mean $\sum_{j=1}^{p} \phi_j^{(m)} y_{T+h-j}$ and variance $v^{(m)}$ over $h = 1 : 10$, where $\phi_j^{(m)}$ and $v^{(m)}$ correspond to the m-th posterior sample of ϕ_j and v, respectively, and substituting sampled values as regressors for the future (Prado and West, 2010, p. 44). Figure 4 exhibits the forecasts for the next 10 weeks. We highlight how the predicted values go towards zero relatively quick.

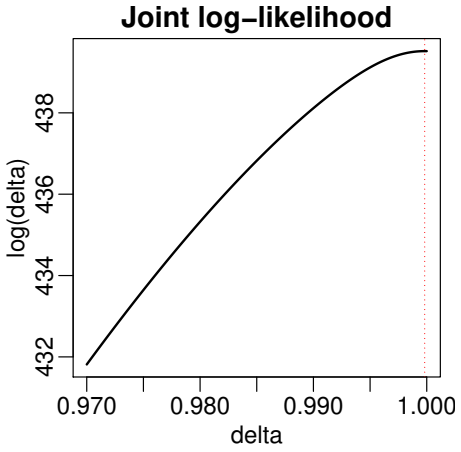

Figure 5: Log-likelihood function of δ. The value of δ that maximizes this function (dotted line in red) in a grid of 1,000 points from 0.97 to 1 is $\delta = 0.9998$.

We conclude that model (1) is not fully appropriate to describe these data because the model is not characterizing important features of the series, such as the mean evolution of the data and the volatility of the variability. This is the case because model (1) is not able to identify additive outliers; as a consequence the fitted values do not accurately describe the mean behavior of the series (which explains why the 95 % credible interval of the forecast in Figure 4 is that wide) and the variance of the residuals turns out to be a under estimate of the true variability involved in the data (which also justifies the patterns around the times with more volatility and the heavy tails detected in the residual analysis).

3. Time-varying autoregressive model

Now we consider a time-varying autoregressive model of order $p = 3$ given by

$$\begin{aligned} y_t &= \phi_{t,1} y_{t-1} + \phi_{t,2} y_{t-2} + \phi_{t,3} y_{t-3} + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} \mathsf{N}(0, v_t), \\ \boldsymbol{\phi}_t &= \boldsymbol{\phi}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \overset{\text{iid}}{\sim} \mathsf{N}(\mathbf{0}_3, \mathbf{W}_t(\delta)), \end{aligned} \tag{3}$$

where $\boldsymbol{\phi}_t = (\phi_{t,1}, \phi_{t,2}, \phi_{t,3})'$ and δ is a discount factor with $\delta \in (0, 1]$. We start by noticing that model (3) is a TVAR model, see West and Harrison (1999, Sec. 9.6) and Prado and West (2010, Ch. 5) for details, which can be written in standard DLM notation as:

$$\begin{aligned} y_t &= \mathbf{F}_t' \boldsymbol{\phi}_t + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} \mathsf{N}(0, v_t), \\ \boldsymbol{\phi}_t &= \mathbf{G}_t \boldsymbol{\phi}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \overset{\text{iid}}{\sim} \mathsf{N}(\mathbf{0}_3, \mathbf{W}_t(\delta)), \end{aligned} \tag{4}$$

where

$$\mathbf{F}_t = (y_{t-1}, y_{t-2}, y_{t-3})', \qquad \boldsymbol{\phi}_t = (\phi_{t,1}, \phi_{t,2}, \phi_{t,3})',$$
$$\mathbf{G}_t \equiv \mathbf{G} = \mathbf{I}_3, \qquad \mathbf{W}_t(\delta) \equiv \mathbf{W}_t = \frac{1-\delta}{\delta}\,\mathbf{P}_t,$$

with $\mathbf{I}_3$ the 3×3 identity matrix and $\mathbf{P}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' = \mathbf{C}_{t-1}$ (since $\mathbf{G}_t = \mathbf{I}_3$ for all t). Here and in what follows we use the standard DLM notation used by West and Harrison (1999), in particular that corresponding to the Kalman Filter (West and Harrison, 1999, Theorem 4.1, p. 103 and Theorem 4.3, p. 109) and the discount factors (Prado and West, 2010, Sec. 4.3.6, p. 130 and Sec. 4.3.7, p.131).

We assume that $v_t \equiv v$ for all t with v unknown and therefore we implement the Kalman Filter according to the formulae given in West and Harrison (1999, Sec. 4.6, p. 111). First of all, we select the appropriate value of δ before performing any posterior inference about the parameters of the model. We choose the value of δ

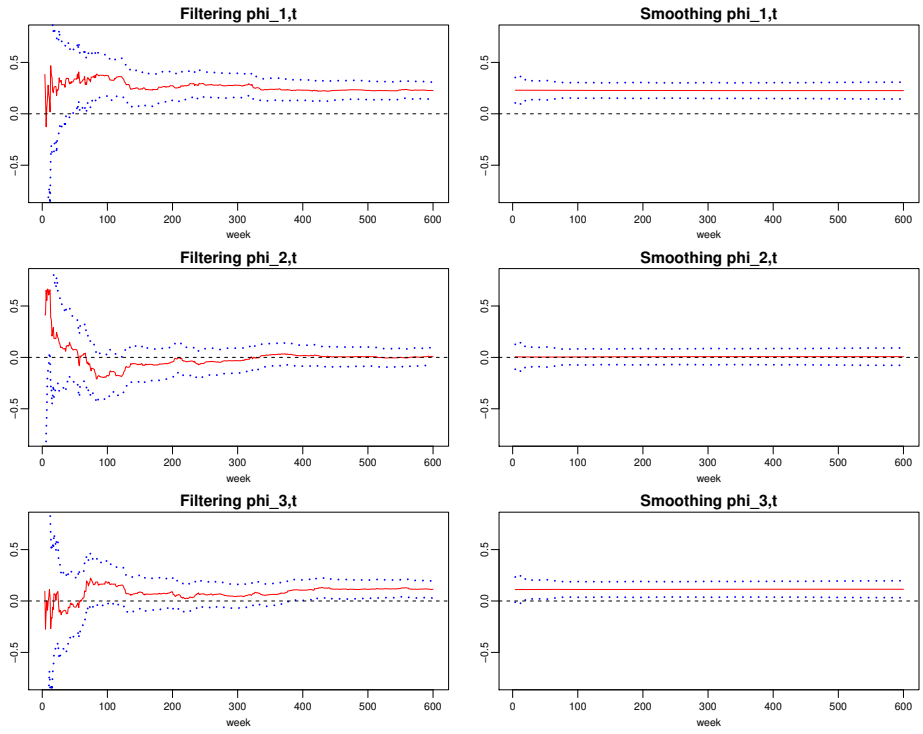

Figure 6: Posterior mean (solid line in red) along with the correspondent 95 % posterior bands (dotted lines in blue) of the filtering distribution $(\phi_{t,j}|\mathcal{D}_t)$ (left column) and the smoothing distribution $(\phi_{t,j}|\mathcal{D}_T)$ (right column).

that maximizes the log-likelihood function

$$\log(\delta) = \sum_{t=1}^{T} \log p(y_t \mid \mathcal{D}_{t-1}) \tag{5}$$

where $p(y_t \mid \mathcal{D}_{t-1}) = T_{n_t-1}(f_t, Q_t)$. Figure 5 shows this log-likelihood function in a grid of 1,000 point from 0.97 to 1. The value of δ that maximizes this function is $\delta = 0.9998$. Alternatively, we can choose that value of δ that minimizes the MSE; this method leads to the same value. Recalling that $\mathbf{W}_t$ measures how quickly the value of the current information $\mathcal{D}_t$ decays through time, since $\delta \approx 1$, then $\mathbf{W}_t \approx \mathbf{0}$ for all t, and therefore the system model is globally true (reliable).

Even though $\delta = 0.9998$ is a value extremely close to one, we decide to perform posterior inference over the parameters of model (4) as a first approach to understand the dynamic features of the system. To this end, we implement a modified version of the Kalman Filter and the smoothing equations since v is unknown but constant for all t (Prado and West, 2010, Sec. 4.3.2, p. 126). Figure 6 displays the posterior mean along with the correspondent 95 % posterior bands of the filtering

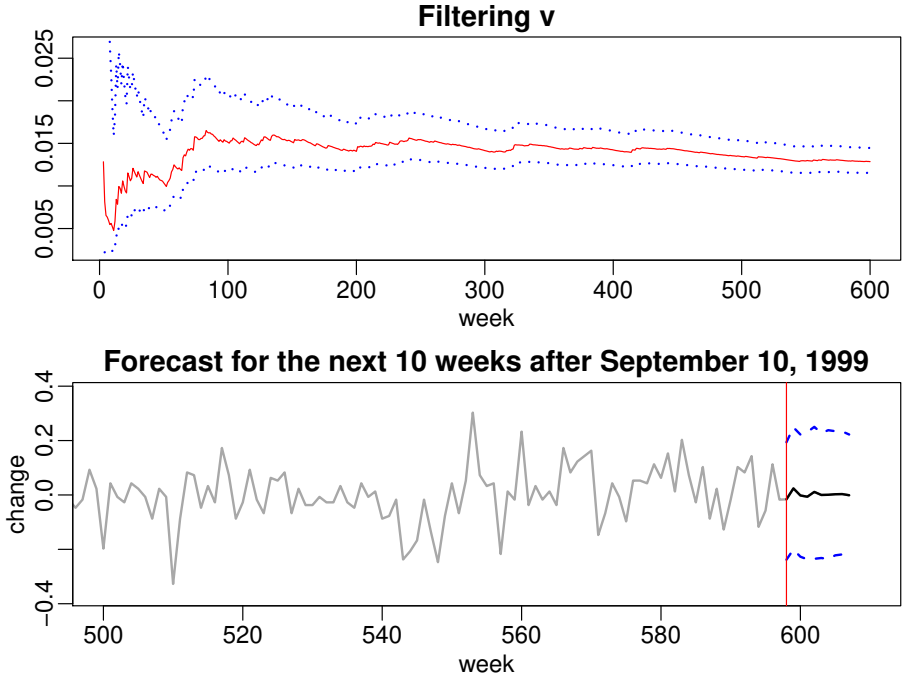

Figure 7: Top panel: posterior mean (solid line in red) along with the correspondent 95 % posterior bands (dotted lines in blue) of the filtering distribution $(v \mid \mathcal{D}_t)$. Bottom panel: forecasts (thick black) of the change of the U.S. 3-year Treasury constant maturity interest rate after the week of September 10, 1999 (dotted line in red) along with 95 % credible intervals (dashed lines in blue).

distribution $(\phi_{t,j} \mid \mathcal{D}_t)$ and the smoothing distribution $(\phi_{t,j} \mid \mathcal{D}_T)$. It takes about 100 weeks to the filtering distribution $(\phi_{t,j} \mid \mathcal{D}_t)$ for each $j = 1, 2, 3$ to get informed about the magnitude of the dynamic parameters, which stabilizes around the correspondent Bayes estimate $\hat{\phi}_{j\,\text{bayes}}$ in model (1) (see part a. for details). From a retrospective point of view, the smoothing distribution of the dynamic coefficients captures this feature immediately and remains almost constant for all t. Note that all the graphs in Figure 6 are given in the same scale. For completeness, Figure 7 shows the posterior mean of the filtering distribution $(v \mid \mathcal{D}_t)$ (which stabilizes around the correspondent Bayes estimate $\hat{v}_{\text{bayes}}$ in model (1) and remains almost constant from week 100 onwards) and the forecasts for the next 10 weeks (which is consistent with the predictions obtained in part a.iii.). We leave the discussion of how to get the 10-step ahead forecasts, the description of the smoothing distribution $(\mu_t \mid \mathcal{D}_T)$, and residual analysis for part b.ii., where we consider two discount factors, one for each layer of model (4).

Now, we assume that v_t is unknown and use a discount factor $\beta \in (0, 1]$ to specify this time-varying observational variance. In this way, we modify the Kalman Filter accordingly and use the formulae to update n_t and d_t given in Prado and West (2010, Sec. 4.3.7, p. 131). Once again, we select the appropriate value of δ and β before performing any posterior inference about the parameters of the model. We

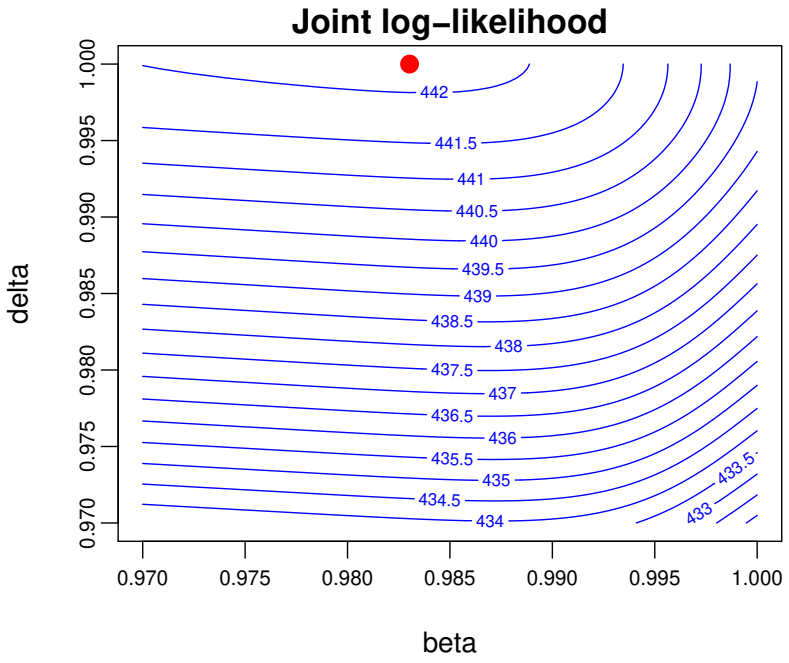

Figure 8: Log-likelihood function of δ and β. The values of δ and β that maximize this function (dotted line in red) in a bidimensional grid of $1,000^2$ points from 0.97 to 1 are $\delta = 0.9998$ and $\beta = 0.9830$.

choose the value of δ and β that maximizes the log-likelihood function given in (5) but this time as a function of both δ and β, i.e.,

$$(\alpha, \beta) = \arg\max_{(\delta,\beta)} \sum_{t=1}^{T} \log p(y_t \mid \mathcal{D}_{t-1}).$$

Figure 8 shows this log-likelihood function in a bidimensional grid of $1,000^2$ points from 0.97 to 1. The values of δ and β that maximize this function are $\delta = 0.9998$ and $\beta = 0.9830$. This is consistent with the maximization when considering a single discount factor in the system level of model 4.

Even though, the MSE criteria (not shown here) suggests a value for β around 0.9700 we follow a conservative approach by setting $\beta = 0.9830$. We now perform posterior inference over the parameters of model (4) with $\delta = 0.9998$ and $\beta = 0.9830$. Figure 9 displays the posterior mean along with the correspondent 95 % posterior bands of the filtering distribution $(\phi_{t,j} \mid \mathcal{D}_t)$ and the smoothing distribution $(\phi_{t,j} \mid \mathcal{D}_T)$. As expected (since $\delta \approx 1$ in both scenarios), Figure 6 and Figure 9 are fairly similar, but the confidence bands in the latter are wider. This is consistent with fact that now we are allowing v_t to be time-varying in the obser-

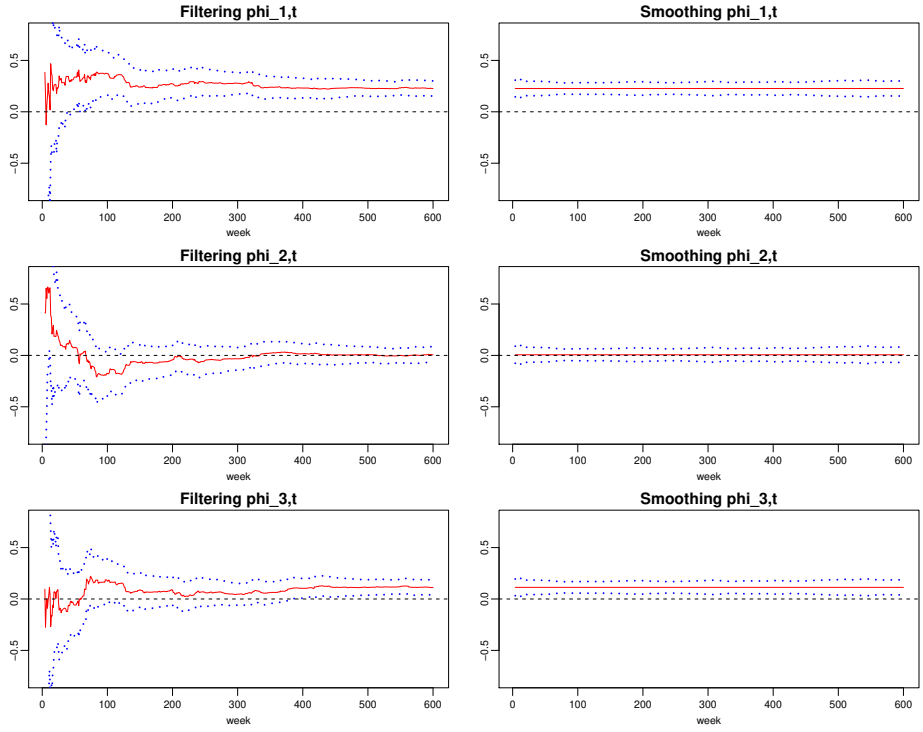

Figure 9: Posterior mean (solid line in red) along with the correspondent 95 % posterior bands (dotted lines in blue) of the filtering distribution $(\phi_{t,j} \mid \mathcal{D}_t)$ (left column) and the smoothing distribution $(\phi_{t,j} \mid \mathcal{D}_T)$ (right column).

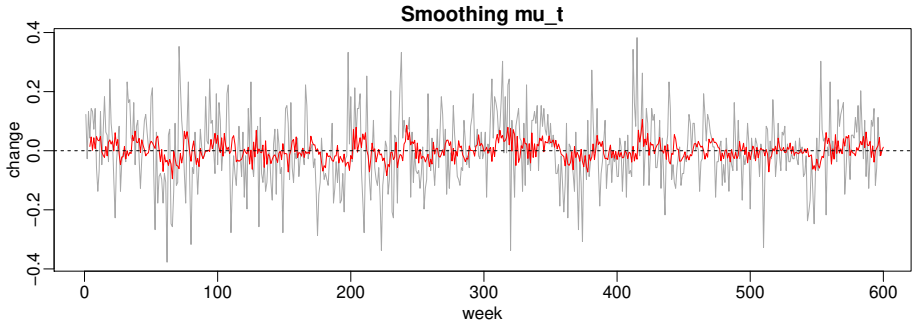

Figure 10: Time series of the detrended weekly change series of the U.S. 3-year Treasury constant maturity interest rate from March 18, 1988 to September 10, 1999 (solid line in gray) and posterior mean (solid line in red) of the smoothing distribution $(\mu_t \mid \mathcal{D}_T)$.

vation level of model (4). On the other hand, Figure 10 shows the posterior mean of the smoothing distribution $(\mu_t \mid \mathcal{D}_T)$. We see that the model does a reasonable job in describing the process underlying the series. However, since the model is not designed to capture additive outliers, it results in large residuals for some of the observations (see Figure 12).

The first panel in Figure 11 exhibits the posterior mean of the filtering distribution $(v \mid \mathcal{D}_t)$. As expected (since now we are considering a time-varying observational variance), the posterior mean of this distributions is more dynamic than the one presented in Figure 7, and is not clearly stabilized around a specific value. Note also that de the confidence bands are wider than before. This time-varying observational variance allows the model to capture more signals from the process underlying the series than a constant variance in the observational equation is capable of. Recall that the moments of the h-step ahead forecast predictive distribution $(y_{t+h} \mid \mathcal{D}_t)$ are

$$f_t(h) = \mathbf{F}'_{t+h}\mathbf{a}_t(h), \quad q_t(h) = \mathbf{F}'_{t+h}\mathbf{R}_t(h)\mathbf{F}_{t+h} + v_{t+h},$$

for $h = 1, 2 \ldots$ with initial values $\mathbf{a}_t(0) = \mathbf{m}_t$ and $\mathbf{R}_t(0) = \mathbf{C}_t$ (see Prado and West (2010, Sec. 4.3.4, p. 128 for details). Given the fact that these calculations require that the future values of the variance components v_{t+h} and $\mathbf{W}_{t+h}$ be known or estimated up to the forecast horizon, we adopt the strategy proposed by West and Harrison (1999, p. 199), in which we assume a conditional constant variance

$$\mathbb{V}ar\left(\boldsymbol{\omega}_{t+h} \mid \mathcal{D}_t\right) = \mathbf{W}_t(h) = \mathbf{W}_{t+1}.$$

Thus, this the step-ahead forecast distributions will be based on the addition of evolution errors with same variance matrix $\mathbf{W}_{t+1}$ for all h. Having this in mind, the second panel in Figure 11 shows the forecasts for the next 10 weeks. We notice that the confidence bands are narrower than before and therefore these predictions are more accurate than the ones obtained in parts a.iii. and b.i.

Finally, Figure 12 shows a residual analysis of the residuals after fitting model 3 with time-varying observational variance. Note that in general we are having the

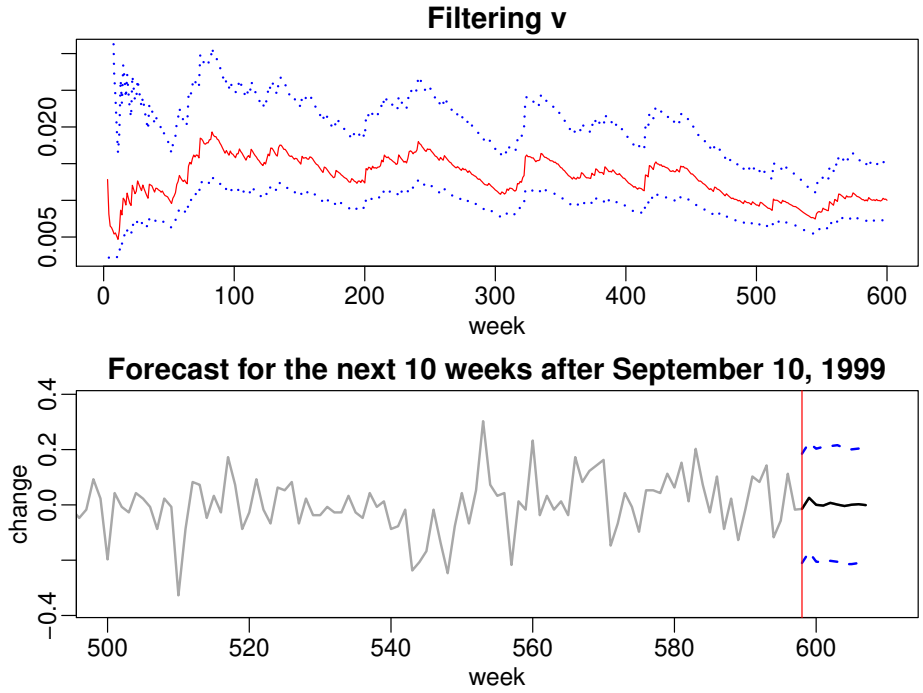

Figure 11: Top panel: posterior mean (solid line in red) along with the correspondent 95 % posterior bands (dotted lines in blue) of the filtering distribution $(v \mid \mathcal{D}_t)$. Bottom panel: forecasts (thick black) of the change of the U.S. 3-year Treasury constant maturity interest rate after the week of September 10, 1999 (dotted line in red) along with 95 % credible intervals (dashed lines in blue).

same issues than before with model (1), because around weeks 65, 74, 201, 323, 418 (lines in blue), where residuals take the largest values, we still recognize am explicit pattern which clearly suggests that the model is not able to detect outliers effectively.

4. Time-varying model for detecting outliers

Now we consider the following model for detecting outliers (Tsay, 2010, p. 561):

$$\begin{aligned} y_t &= \gamma_t \alpha_t + x_t + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} \mathsf{N}(0, 0.01) \\ x_t &= \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \eta_t, \quad \eta_t \overset{\text{iid}}{\sim} \mathsf{N}(0, w), \end{aligned} \tag{6}$$

where $\gamma_t \overset{\text{iid}}{\sim} \mathsf{Ber}(0.2)$, and $\alpha_t \overset{\text{iid}}{\sim} \mathsf{N}(0, 0.1)$. This model allows additive outliers to occur at every time point with the probability of being an outlier for each observation equal to 0.2. We have $T = 600$ observations in the series and $N+N+3+1 = 2N+4$ parameters, namely $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_T)'$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)'$, $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)'$, and w.

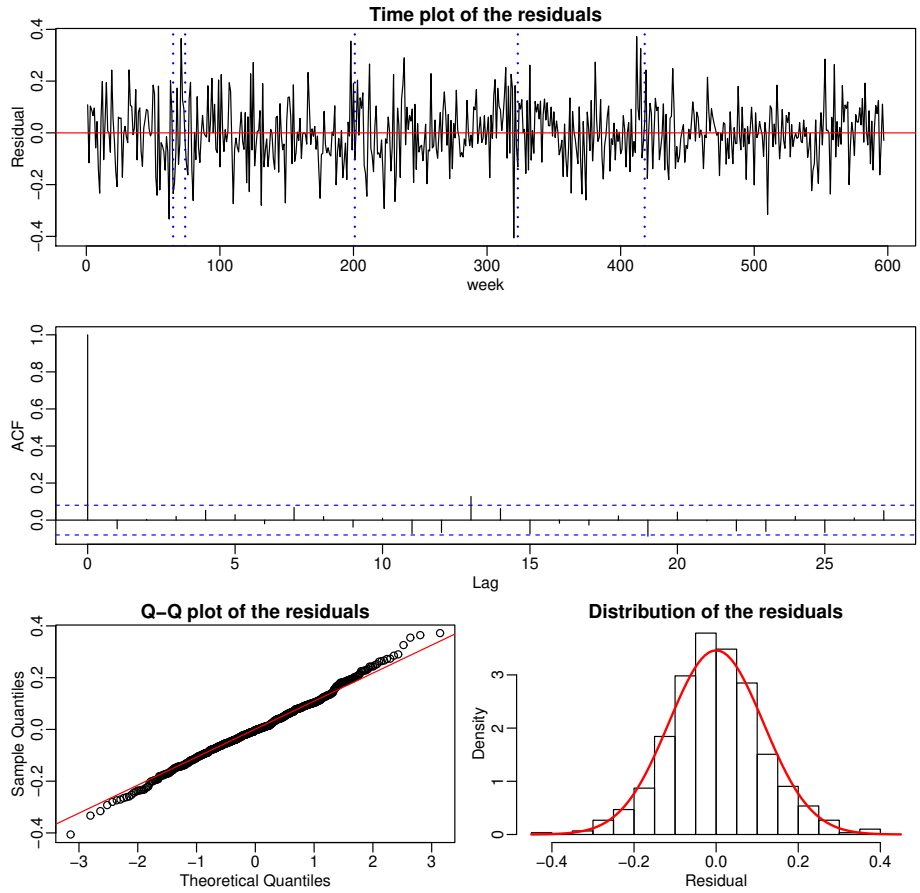

Figure 12: Residual analysis: Time series plot (top), ACF (middle), Q-Q plot (bottom left), and histogram (bottom right) of the residuals obtained from fitting model 1.

The parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ are introduced by using the idea of data augmentation with γ_t denoting the presence of absence of an additive outlier at time t, and α_t is the magnitude of the outlier at time t when it is present (Tsay, 2010, p. 561). This very same model has also been consider extensively in Sosa and Buitrago (2022, Ch. 3).

We want to obtain samples from the joint posterior $p(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{\phi}, w \mid \boldsymbol{y})$, where $\boldsymbol{x} = (x_1, \ldots, x_T)'$, and $\boldsymbol{y} = (y_1, \ldots, y_T)'$. We assume that the prior distributions are $p(\boldsymbol{\phi}) = \mathsf{N}(\boldsymbol{\phi} \mid \mathbf{0}_3, 0.25\mathbf{I}_3)$ and $p(w) = \mathsf{IG}(w \mid a_0, b_0)$ which are conjugate priors. To implement a MCMC algorithm for model 6, we consider the joint posterior distribution expressed as:

$$p(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{\phi}, w \mid \boldsymbol{y}) = p(\boldsymbol{y} \mid \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{x})\, p(\boldsymbol{x} \mid \boldsymbol{\phi}, w)\, p(\boldsymbol{\gamma})\, p(\boldsymbol{\alpha})\, p(\boldsymbol{\phi})\, p(w). \tag{7}$$

This expression is particularly useful for deriving the full conditionals (see the

Appendix for details).

Note that model (6) is a non-Gaussian DLM. However, conditional on $\boldsymbol{\gamma}$, this model can be written as a DLM as follows:

$$\begin{aligned} y_t &= \mathbf{F}'\boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim \mathsf{N}(0, v_{\gamma_t}), \\ \boldsymbol{\theta}_t &= \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathsf{N}(\mathbf{0}_3, \mathbf{W}_t) \end{aligned} \tag{8}$$

with

$$\mathbf{F} = (1,0,0)', \quad \boldsymbol{\theta}_t = (x_t, x_{t-1}, x_{t-2})', \quad v_{\gamma_t} = \begin{cases} 0.01, & \gamma_t = 0; \\ 0.11, & \gamma_t = 1, \end{cases}$$

and

$$\mathbf{G} = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \boldsymbol{\omega}_t = (\eta_t, 0, 0)', \quad \mathbf{W}_t \equiv \mathbf{W} = \begin{pmatrix} w & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus, the model $\{\mathbf{F}, \mathbf{G}, v_{\gamma_t}, \mathbf{W}\}$ has a standard normal DLM structure conditional on the latent parameters γ_t. We use this fact to develop the second step of the MCMC algorithm described below. There we implement a FFBS algorithm (Prado and West, 2010, p. 137) to get samples from $p(\boldsymbol{\theta}_{1:T} \mid \boldsymbol{\gamma}, \boldsymbol{\phi}, w)$, $t = 1:T$ (notice that getting samples from this distribution automatically allow us to draw samples from $p(\boldsymbol{x} \mid \boldsymbol{\gamma}, \boldsymbol{\phi}, w)$).

Now, we describe the general form of the MCMC implemented in the next part of the problem to obtain samples from the joint posterior given in (7). We derive the full conditionals used in the algorithm looking at the dependencies in the full posterior distribution 7 (see the Appendix for details). Let $\theta^{(m)}$ denote the state of parameter θ in the m-th iteration of the algorithm. The algorithm is as follows:

1. Initialize: sample $\gamma_t^{(0)} \sim \mathsf{Ber}(0.2)$, $t = 1:T$, $\boldsymbol{\phi}^{(0)} \sim \mathsf{N}(\boldsymbol{\phi} \mid \mathbf{0}_3, \mathbf{I}_3)$, and $w^{(0)} \sim \mathsf{IG}(w \mid a_0, b_0)$.

2. Update $\boldsymbol{\gamma}^{(m-1)}, \boldsymbol{\alpha}^{(m-1)}, \boldsymbol{\theta}_t^{(m-1)}, \boldsymbol{\phi}^{(m-1)}, w^{(m-1)}$ as follows:

   a) FFBS step: sample $\boldsymbol{\theta}_t^{(m)} \sim p(\boldsymbol{\theta}_{1:T} \mid \boldsymbol{\gamma}^{(m-1)}, \boldsymbol{\phi}^{(m-1)}, w^{(m-1)}, \boldsymbol{y})$, $t = 1:T$, via the following steps:

      i) Use the DLM filtering equations to compute $\mathbf{a}_t$, $\mathbf{R}_t$, $\mathbf{m}_t$, and $\mathbf{C}_t$, for $t = 1:T$.

      ii) At time $t = T$ sample $\boldsymbol{\theta}_T^{(m)} \sim \mathsf{N}(\boldsymbol{\theta}_T \mid \mathbf{m}_T, \mathbf{C}_T)$.

      iii) For $t = (T-1):0$ sample $\boldsymbol{\theta}_t^{(m)} \sim \mathsf{N}(\boldsymbol{\theta}_t \mid \mathbf{m}_t^*, \mathbf{C}_t^*)$, where

$$\mathbf{m}_t^* = \mathbf{m}_t + \mathbf{B}_t(\boldsymbol{\theta}_{t+1} - \mathbf{a}_{t+1}), \quad \mathbf{C}_t^* = \mathbf{C}_t - \mathbf{B}_t\mathbf{R}_{t+1}\mathbf{B}_t',$$

      with $\mathbf{B}_t = \mathbf{C}_t\mathbf{G}_{t+1}'\mathbf{R}_{t+1}^{-1}$.

b) For $t = 1 : T$ sample $\alpha_t^{(m)} \sim p(\alpha_t \mid \boldsymbol{\theta}_t^{(m)}, \gamma_t^{(m-1)}, \boldsymbol{y})$, where $p(\alpha_t \mid \boldsymbol{\theta}_t, \gamma_t, \boldsymbol{y}) = \mathsf{N}(\alpha_t \mid \mu_t, \sigma_t^2)$ with

$$\sigma_t^2 = \frac{1}{10(10\gamma_t + 1)}, \quad \mu_t = 100\sigma_t^2\gamma_t(y_t - \theta_{t,1}).$$

c) For $t = 1 : T$ sample $\gamma_t^{(m)} \sim p(\gamma_t \mid \boldsymbol{\theta}_t^{(m)}, \boldsymbol{y})$, where $p(\gamma_t \mid \boldsymbol{\theta}_t, \boldsymbol{y}) = \mathsf{Ber}(\gamma_t \mid \pi_t)$ with $\pi_t = \mathsf{P}[\gamma_t = 1 \mid \boldsymbol{\theta}_t, \boldsymbol{y}]$, i.e., γ_t is set to 1 or 0 with probabilities defined in terms of the odds ratio

$$\frac{\pi_t}{1 - \pi_t} = \frac{1}{4\sqrt{11}} \exp\left\{-\frac{1}{2}(y_t - \theta_{t,1})^2(1/0.11 - 1/0.01)\right\}$$

d) Sample $\boldsymbol{\phi}^{(m)} \sim p(\boldsymbol{\phi} \mid \boldsymbol{\theta}_{1:T}^{(m)}, w^{(m-1)}, \boldsymbol{y})$, where $p(\boldsymbol{\phi} \mid \boldsymbol{\theta}_{1:T}, w, \boldsymbol{y}) = \mathsf{N}(\phi \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\Sigma} = \left(\frac{\sum_{t=1}^{T} \boldsymbol{\theta}_{t,1}\boldsymbol{\theta}'_{t-1}}{w} + 4\mathbf{I}_3\right)^{-1}, \quad \boldsymbol{\mu} = \boldsymbol{\Sigma}\left(\frac{\sum_{t=1}^{T} \theta_{t,1}\boldsymbol{\theta}_{t-1}}{w}\right).$$

e) Sample $w^{(m)} \sim p(w \mid \boldsymbol{\theta}_{1:T}^{(m)}, \boldsymbol{\phi}^{(m)}, \boldsymbol{y})$, where $p(w \mid \boldsymbol{\theta}_{1:T}, \boldsymbol{\phi}, \boldsymbol{y}) = \mathsf{IG}(w \mid a_w, b_w)$ with

$$a_w = a_0 + \frac{T}{2}, \quad b_w = b_0 + \frac{1}{2}\sum_{t=1}^{T}(\theta_{t,1} - \boldsymbol{\phi}'\boldsymbol{\theta}_t)^2.$$

3. Cycle until achieve convergence.

Now, we implement the MCMC described in part i. In order to run the algorithm we must pick appropriate values for a_0 and b_0. These hyperparameters control the prior distribution of the scale parameter w, which directly impacts the evolution of the system parameters in the conditional DLM 8. Recalling that the MLE of the variance of the error in the AR(3) model 1 is $v_{\text{mle}} = 0.0128$ (which is certainly an overestimation of w), we weakly concentrate the prior distribution around this value by using $a_0 = 3$ and $b_0 = 0.0257$: doing so we get that $\mathbb{E}(w \mid a_0, b_0) = v_{\text{mle}}$ with $cv = 100\,\%$, e.i., a weak concentration of w around v_{mle}.

We run the algorithm for 4,000 iterations but discard results of the first 1,000 iterations (burn-in period). In order to have approximately independent draws, we select one sample from each 15 iterations, and a total of 201 samples are selected altogether. Those 201 samples are approximately independently and identically distributed according to the related posterior distribution and form the basis of posterior inference.

Before making any inference based on the MCMC samples we determine if there are any indications that the chains is not stationary. The chains achieve convergence quickly and there does not seem to be any evidence that the chain has not achieved

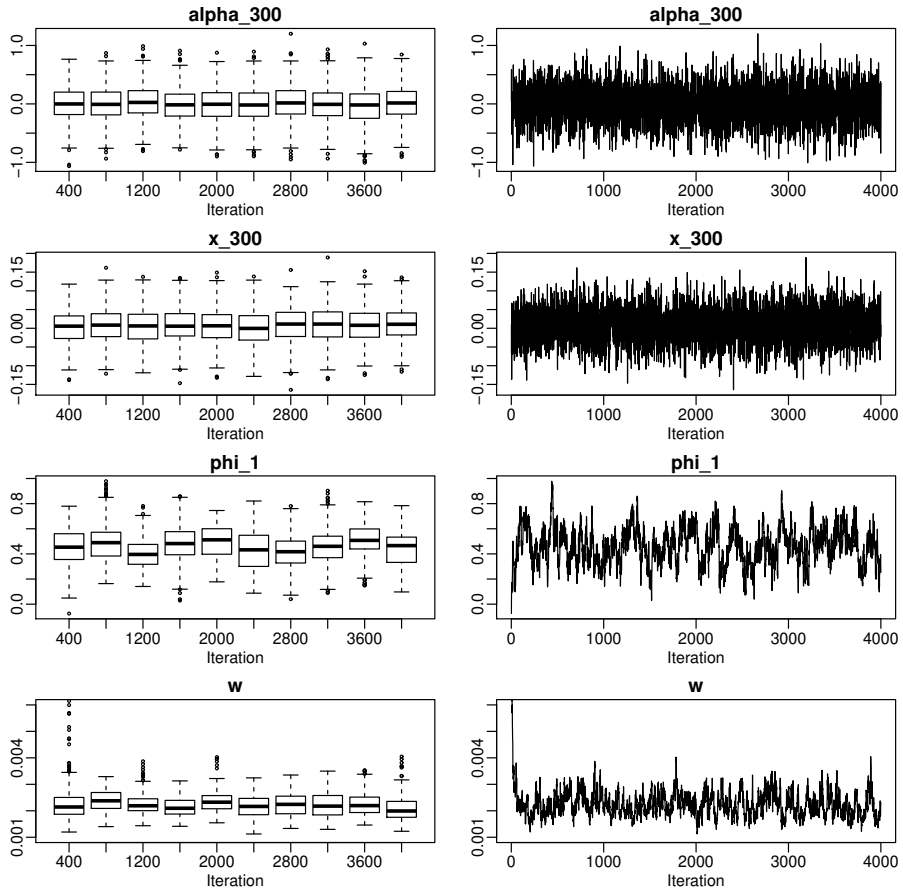

Figure 13: Stationary plots of some parameters in the model.

stationarity. Stationary plots of some parameters in the model are shown in Figure 13.

In what follows we summarize the posterior distributions of the parameters of the model. Table 2 display Monte Carlo summaries to the posterior densities of $\boldsymbol{\phi}$ and w. This estimates are roughly consistent with the ones obtained with models

Parameter	ϕ_1	ϕ_2	ϕ_3	w
Mean	0.4711	0.0710	0.1540	0.0022
SD	0.1408	0.1751	0.1227	0.0004
Q2.5 %	0.2074	-0.2426	-0.0727	0.0015
Q97.5 %	0.7288	0.3817	0.3846	0.0031

Table 2: Posterior summaries of $\boldsymbol{\phi}$ and w.

Week	65	74	201	323	418
Probability	0.8209	0.9154	0.8060	0.9403	0.9055
Magnitude	-0.2288	0.2745	0.2535	-0.3087	0.2989

Table 3: Posterior probability of being an outlier along with the posterior mean of outlier magnitude at weeks 65, 74, 201, 323, 418.

(1) and (4). Now we center our attention in Figure 14. Panel (a) summarizes the CDLM mean at the observation x_t. Panel (b) shows the time plot of the posterior probability of each observation being an additive outlier, and panel (c) exhibits the posterior mean of outlier magnitude. There we summarize the posterior densities $\mathsf{P}[\gamma_t = 1 \mid \boldsymbol{y}]$ and $\mathsf{P}[\alpha_t \mid \gamma_t = 1, \boldsymbol{y}]$, for t=1:T. From the probability plot we identify 17 observations having posterior probability of being an outlier greater than 0.5, and 5 of those observations with this probability greater than 0.8, namely, observation at weeks 65, 74, 201, 323, 418. Magnitudes and corresponding probabilities of being an outlier of these time points are given in Table 3. Tsay (2010, p. 564) also classifies these points as outliers (employing a approach with slightly differences) and highlight times $t = 323$ (May 20, 1994) and $t = 201$ (January 17, 1992). At the former there was a 0.6 % drop in the weekly interest rate

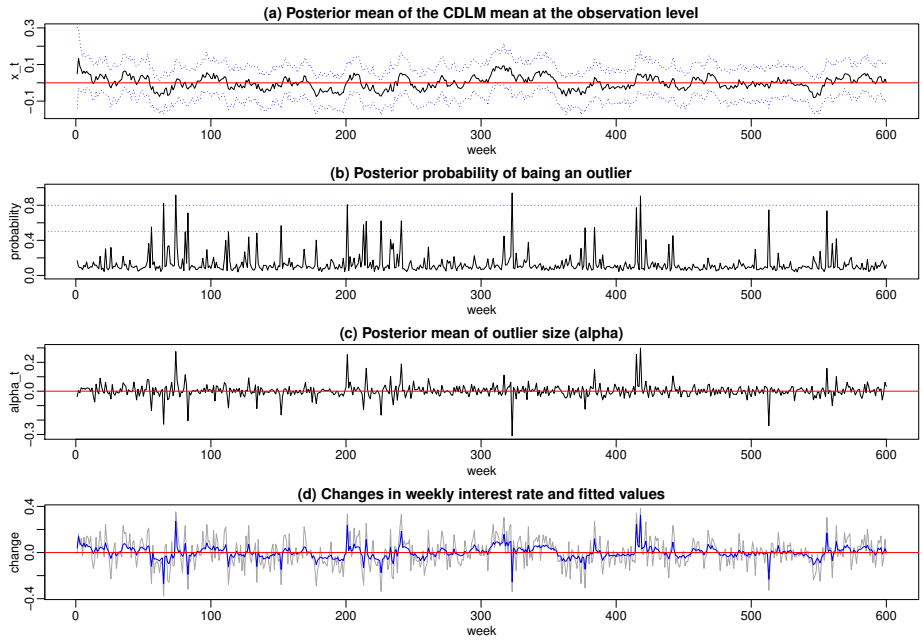

Figure 14: Time plots of weekly change series of the U.S. 3-year Treasury constant maturity interest rate from March 18, 1988 to September 10, 1999: (a) posterior mean of the CDLM mean at the observation x_t; (b) the posterior probability of being an outlier (the red dotted line corresponds to 0.5 and the blue one to 0.8); (c) the posterior mean of outlier size; and (d) the data (gray) and fitted values (blue).

within two weeks, and at the later there was a jump of about 0.35 % in the weekly interest rate. In addition, the bottom panel of Figure 14 displays the original data in gray and the fitted values in blue. We see that model 6 reasonably captures the "volatility" that models 1 and 3 do not.

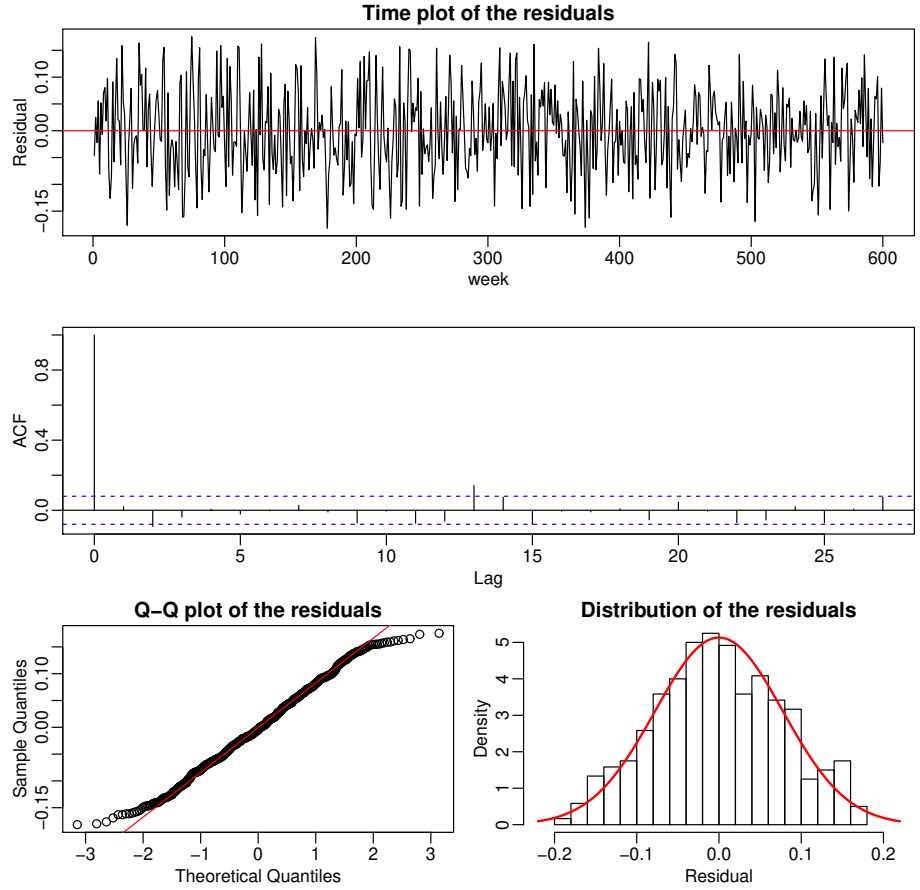

Figure 15: Residual analysis: Time series plot (top), ACF (middle), Q-Q plot (bottom left), and histogram (bottom right) of the residuals obtained from fitting model 6.

Finally, Figure 15 shows a residual analysis of the residuals after fitting model 6. Note that the time series of the residuals does not exhibit a strong trend, seasonality, or other suspicious pattern. Furthermore, the ACF does not reveal any significant correlation in any lag of the series of the residuals which strongly suggests that the series correspond to an independent sample. Then, looking at the Q-Q plot and the histogram of the residuals, it seems that the residuals indeed correspond to a i.i.d. sample from a normal distribution (note the improvement in the tails in comparison to model 1 for example). Thus, we are confident about the good performance of model 6 to correctly characterize the data.

5. Discussion

There exists a wide array of time series data types, each addressed in extensive statistical literature that covers most common scenarios. The dynamic strategy discussed here offers a highly flexible method for characterizing temporal settings, particularly those where the evolution of the response variable changes over time. This approach provides a comprehensive description of how variables interact and change throughout the course of the data series, making it particularly suitable for modeling complex systems and phenomena that exhibit varying dynamics over time.

The fully Bayesian approach discussed in this context enables a comprehensive characterization of the time series process while also facilitating accurate predictions. In addition to this approach, alternative strategies encompass dynamic frameworks incorporating dependencies on covariates, as well as Bayesian nonparametric methods. These methods offer diverse approaches to modeling time series data, accommodating various complexities and nuances present in real-world datasets.

Statements and Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this art

Recibido: Noviembre 1 de 2023
Aceptado: Diciembre 17 de 2023

Referencias

G. Petris, S. Petrone, and P. Campagnoli. *Dynamic Linear Models with R*. Use R! Springer, 2009. ISBN 9780387772387. URL `https://books.google.com/books?id=VCt3zVq8TO8C`.

R. Prado and M. West. *Time Series: Modeling, Computation, and Inference*. Chapman and Hall/CRC, 2010. ISBN 9781420093360.

J. Sosa and L. Buitrago. Illustrating advantages and challenges of bayesian statistical modelling: An empirical perspective. *Model Assisted Statistics and Applications*, 17(3):175–187, 2022.

R. Tsay. *Analysis of Financial Time Series*. CourseSmart. Wiley, 2010. ISBN 9781118017098. URL `https://books.google.com/books?id=OKUGARAXKMwC`.

M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics. Springer New York, 1999. ISBN 9780387947259. URL `https://books.google.com/books?id=jcl8lD75fkYC`.

A. Notation

The cardinality of a set A is denoted by $|A|$. If P is a logical proposition, then $\mathbf{1}\{\mathrm{P}\} = 1$ if P is true, and $\mathbf{1}\{\mathrm{P}\} = 0$ if P is false. $\lfloor x \rfloor$ denotes the floor of x, whereas $[n]$ denotes the set of all integers from 1 to n, i.e., $\{1, \ldots, n\}$. The Gamma function is given by $\Gamma(x) = \int_0^\infty u^{x-1}\, e^{-u}\, \mathrm{d}u$. Matrices and vectors with entries consisting of subscripted variables are denoted by a boldfaced version of the letter for that variable. For example, $\boldsymbol{x} = (x_1, \ldots, x_n)$ denotes an $n \times 1$ column vector with entries $x_1, \ldots, x_n$. We use $\mathbf{0}$ and $\mathbf{1}$ to denote the column vector with all entries equal to 0 and 1, respectively, and $\mathbf{I}$ to denote the identity matrix. A subindex in this context refers to the corresponding dimension; for instance, $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. The transpose of a vector $\boldsymbol{x}$ is denoted by $\boldsymbol{x}^\mathsf{T}$; analogously for matrices. Moreover, if $\mathbf{X}$ is a square matrix, we use $\mathrm{tr}(\mathbf{X})$ to denote its trace and $\mathbf{X}^{-1}$ to denote its inverse. The norm of $\boldsymbol{x}$, given by $\sqrt{\boldsymbol{x}^\mathsf{T}\boldsymbol{x}}$, is denoted by $\|\boldsymbol{x}\|$.

Now, we present the form of some standard probability distributions used in this article:

- Multivariate normal:

  A $d \times 1$ random vector $\boldsymbol{X} = (X_1 \ldots, X_d)$ has a multivariate Normal distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, denoted by $\boldsymbol{X} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathsf{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its density function is

  $$p(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-d/2}\, |\boldsymbol{\Sigma}|^{-1/2}\, \exp\left\{-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

- Inverse Gamma:

  A random variable X has an Inverse Gamma distribution with parameters $\alpha, \beta > 0$, denoted by $X \mid \alpha, \beta \sim \mathsf{IG}(\alpha, \beta)$, if its density function is

  $$p(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{-(\alpha+1)}\, \exp\{-\beta/x\}, \quad x > 0\,.$$