Christoph Berninger, David Rügamer, Almond Stöcker

A Bayesian Time–Varying Autoregressive Model for Improved Short– and Long–Term Prediction

Working Paper Number 22, 2020
Center for Quantitative Risk Analysis (CEQURA)
Department of Statistics
University of Munich

http://www.cequra.uni-muenchen.de

A Bayesian Time-Varying Autoregressive Model for Improved Short- and Long-Term Prediction

Christoph Berninger[1], Almond Stöcker[2], David Rügamer[1]

[1]Department of Statistics, LMU München
[2] School of Business and Economics, Humboldt University of Berlin

April 21, 2020

Abstract

Motivated by the application to German interest rates, we propose a time-varying autoregressive model for short and long term prediction of time series that exhibit a temporary non-stationary behaviour but are assumed to mean revert in the long run. We use a Bayesian formulation to incorporate prior assumptions on the mean reverting process in the model and thereby regularize predictions in the far future. We use MCMC-based inference by deriving all full conditional distributions and employ the Gibbs Sampler to sample from the posterior (predictive) distribution. By combining data-driven short term predictions with long term distribution assumptions our model excels in predictive performance in comparison to the Gauss2++ model used in the insurance industry for the given application.

1. Introduction

To forecast an univariate time series the first model of choice is often a linear model. A very basic example of this model class in the context of time series analysis is the autoregressive model of order 1 (AR(1)), which is defined as follows:

$$x_t = \alpha + \beta x_{t-1} + \epsilon_t \tag{1}$$

where x_t represents the observed variable at time point t and α and β are real valued constants, while $|\beta| < 1$ is assumed to ensure stationarity. The innovation process ϵ_t can be, e.g., a Gaussian white noise process, i.e. $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.
Linearity is often an excessively strict assumption in practice and many time series exhibit features that can not be captured by a linear model [Hamilton, 1989]. In the last decades a lot of research has been conducted to introduce different types of nonlinear models. A bilinear model is an example of a nonlinear model type, which assumes a nonlinear relationship between the covariates and response variable (see, e.g., [Granger et al., 1978] and [Subba Rao and Gabr, 1984]), although not often used in macroeconomic applications [Morley, 2009]. A more typical approach is to allow one (or more) parameters of a linear model to change over time. This comprises the regime switching and time-varying parameter models.

The first approaches to regime switching models were conducted by [Quandt, 1958], who considered a switching regression model extending a linear regression model by allowing the parameters to switch according to a random variable. [Bacon and Watts, 1971] introduced a smooth transition model, which implements a smooth transition from one regime to another without a sudden jump. [Goldfeld and Quandt, 1973] introduce the Markov switching regression model and use a discrete latent Markov process to determine the current regime. These models were adapted to time series models by Lim and Tong [1980] and Chan and Tong [1986] introducing the threshold autoregressive model (TAR) and the smooth transition autoregressive model (STAR), respectively. [Hamilton, 1989] introduced the Markov switching autoregressive model for applications in economics. These are amongst the most famous regime switching models used in macroeconomics and have been investigated thoroughly together with different variants in the literature [Haggan and Ozaki, 1981], [Teräsvirta, 1994], [Jansen and Teräsvirta, 1996]. [Lanne and Saikkonen, 2002] used a TAR-model, which only allows regime changes for the constant parameter α and applied it to strongly autocorrelated time series data.

In contrast to regime switching models, which allow the parameters to take a finite number of states, time-varying parameter models allow one (or more) of the parameters in a linear model to be driven by its own continuous process [Morley, 2009]. For example, if the parameter vector (α, β, σ^2) of the linear AR(1) model becomes a stochastic process, this results in a time-varying autoregressive model of order 1 (TV-AR(1))

$$x_t = \alpha_t + \beta_t x_{t-1} + \epsilon_t \tag{2}$$

with $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$. Certain distribution assumptions for the underlying stochastic process of the parameter vector (α_t, β_t, σ_t) are made in practice to complete the TV-AR(1) model specification [Teräsvirta et al., 2010]. Similar to the TAR model in [Lanne and Saikkonen, 2002] the time variation of the TV-AR(1) model can be restricted to the constant parameter α_t, resulting in a time-varying constant autoregressive model of order 1 (TVC-AR(1)):

$$x_t = \alpha_t + \beta x_{t-1} + \epsilon_t. \tag{3}$$

If $|\beta| < 1$ and the latent process of α_t is mean reverting, the model is stationary. But due to random shifts in the mean reversion level – because of the time-varying constant parameter – realizations of the model can resemble those of a (close to) random walk process, when restricting to a limited time window.

Strong autocorrelation and (close to) random walk behaviour of actual stationary time series is exactly the time-series feature we will address in this work. Using a linear (near) integrated process to model these time series might not account for characteristics valid according to economic theory. As Lanne and Saikkonen point out, an impulse response function would imply a very slow mean reversion inconsistent with properties of many economic variables [Lanne and Saikkonen, 2002]. Also, the behaviour in the very long horizon might be unrealistic. A (near) integrated process, e.g. applied to interest rates or unemployment rates, might lead – due to its large variance – to extreme values in the long run never observed in the past. Furthermore, estimating the model parameters of a near integrated but stationary process might include large estimation errors if the sample size is not very large. We therefore consider a nonlinear model, which allows for a time varying mean reversion level. Specifically, we propose a Bayesian TVC-AR(1) model, which is still stationary but has linear properties similar to an integrated or nearly integrated process due to a stochastic mean reversion level. Furthermore, the Bayesian approach allows us to regularize the long run distribution of the time series without affecting the short-term distributions adversely.
The novelty of our approach lies in the proposed Bayesian framework that allows (1) a model with linear properties in accordance with economic theory, (2) the possibility to regularize the long run distribution by using prior assumptions and (3) if applied e.g. to interest rates an improved forecasting performance in the short horizon compared to commonly used linear models in practice. Moreover, we place particular emphasis on the interpretability of the model structure and prior parameters allowing to include expert knowledge or assumptions in accordance with economic theory about the long run behaviour of a time series into the model in a sound mathematical way.

The paper is arranged as follows. Section 2 specifies the Bayesian TVC-AR(1), including the derivation of required full conditional posterior distributions and the application of the Gibbs Sampler for statistical inference. In Section 3 we discuss an application of our model to interest rate data and compare the forecasting performance as well as the long run distribution of our nonlinear model with the dynamic Nelson-Siegel Model and the Gauss2++ Model, which is the industry standard in the insurance industry.

2. A Bayesian TVC-AR(1) model for long run regularisation

In this Section we introduce the Bayesian TVC-AR(1) (BTVC-AR(1)) model. The model incorporates assumptions about the long term behaviour of the time series and thereby regularises the process in the long horizon. At the same time, the model is mainly driven by the given data in the short run and thus fosters a good short- and long-term prediction.

2.1. The BTVC-AR(1) model

The BTVC-AR(1) model is defined as follows:

$$x_t = \alpha_t + \beta x_{t-1} + \epsilon_t, \tag{4}$$

where β represents the mean reversion speed and $|\beta| < 1$ to secure stationarity. ϵ_t is assumed to be a Gaussian white noise process, i.e. $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. We further specify α_t as a Gaussian mean reverting process specified by its unconditional expectation $\boldsymbol{\theta}$ and covariance structure $\boldsymbol{\Sigma}$, i.e.

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_t) \sim \mathcal{N}_t(\boldsymbol{\theta}, \boldsymbol{\Sigma}). \tag{5}$$

The Bayesian approach considers the parameters of the model (4) as random variables. For $\beta|\sigma^2$ a truncated normal distribution with lower bound $l(\sigma^2)$ and upper bound $u(\sigma^2)$ is assumed as a prior, i.e.

$$\beta|\sigma^2 \sim \mathcal{N}(\mu_\beta, \sigma^2_\beta, l(\sigma^2), u(\sigma^2)).$$

The prior distribution for σ^2 is an inverse gamma distribution with shape and scale parameter, a and b, i.e.

$$\sigma^2 \sim \mathcal{IG}(a, b).$$

These two prior distributions are conjugate priors for the model (4) and therefore allow an analytical derivation of the corresponding full conditional distributions. Additionally, the truncation parameters $l(\sigma^2)$ and $u(\sigma^2)$ of the β-prior depend on σ^2 to control the unconditional variance of the model. This is further elaborated in Section 2.2.
The parameters $\boldsymbol{\theta}$ and covariance structure $\boldsymbol{\Sigma}$ might be assumed fixed or subject to further prior distributions. In the latter case the specified distribution of $\boldsymbol{\alpha}$ in (5) holds only conditional on $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$. A suitable choice for these hyperparameters is discussed in Section 2.2 in the case of an AR-covariance structure.

Using these priors the defined model can also be seen as a Bayesian version of the TVC-AR(1) Model. While the framework is flexible to assume various covariance structures for $\boldsymbol{\alpha}$, we here present insightful arguments when assuming an AR-covariance structure and thereby demonstrate the properties of our framework.

2.2. Further prior assumptions

The goal of our Bayesian approach is to regularize the distribution in the long forecasting horizon without affecting the short term distribution adversely. Assumptions about the long term mean and the long term variance can be incorporated into the model via the prior distributions of the parameters of the latent $\boldsymbol{\alpha}$-process.

Long term mean. In order to regularize the model's long run mean, we define $\boldsymbol{\theta}$ to be a fixed value $\vartheta \cdot \mathbf{1}$. Whereas the local long run mean at time point t of the BTVC-AR(1) Model amounts to

$$\mu_t^{loc} = \frac{\alpha_t}{1-\beta},$$

setting the unconditional expectation of $\boldsymbol{\alpha}$ to $\vartheta \cdot \mathbf{1}$ leads to a long run mean of $\frac{\vartheta}{1-\beta}$. We here assume the data to be centered around a prior specified expected long run mean. By setting $\boldsymbol{\theta} = \mathbf{0}$, i.e., $\vartheta = 0$, this long run mean is reached in expectation after reshifting the centralized data.

Long term variance. If we assume that $\boldsymbol{\alpha}$ has an AR-covariance structure, $\boldsymbol{\Sigma}$ is specified by the parameters ρ and τ^2 representing the correlation of two successive time steps and the conditional variance, respectively. The long run variance of the model is then given by

$$Var(x_t) = \frac{\sigma^2}{(1-\beta^2)} + \frac{\tau^2(1+\rho\beta)}{(1-\rho\beta)(1-\beta^2)(1-\rho^2)} \tag{6}$$

For given ρ, β and σ^2 and a prior assumption, the variance of x_t can be solved for τ^2. While our model itself has two sources of variance (the residual term of the AR1 model and the latent $\boldsymbol{\alpha}$-process), we can find a suitable value for one of the variance terms, here τ^2, by specifying a realistic value of the long term variance $Var(x_{\mathcal{T}})$ for a pre-defined horizon $\mathcal{T}$. In doing so, equation (6) can be solved for τ^2. Let denote the solution by $\tilde{\tau}^2$. By setting the conditional prior distribution of τ^2 to

$$\tau^2|\rho, \beta, \sigma^2 \sim \delta_{\tilde{\tau}^2} \tag{7}$$

where δ denotes a degenerated distribution with point mass 1 at $\tilde{\tau}^2$, the prior specified long term variance $Var(x_{\mathcal{T}})$ is reached.

Apart from regularizing the long run distribution, we want to capture the (close to) random walk behaviour of a (nearly) integrated time series. This behaviour of an actually stationary process can be a hint of a missing variable [Lanne and Saikkonen, 2002], which we assume leads to changes of the mean reversion level in a linear AR(1) model. We therefore apply the TVC-AR(1) model, which incorporates all these missing fundamental information into the latent $\boldsymbol{\alpha}$-process. The time-varying constant leads to a time-varying mean reversion level and the slower the time varying mean reversion level changes the more pronounced the (close to) random walk behaviour will be. The speed of mean reversion of the $\boldsymbol{\alpha}$-process is determined by ρ, which represents the correlation between two successive time steps of the latent $\boldsymbol{\alpha}$-process. We therefore choose a normal distribution truncated below by -1 and from above by 1 as a prior for ρ, i.e.

$$\rho \sim \mathcal{N}(\mu_\rho, \sigma_\rho, -1, 1) \tag{8}$$

If ρ is close to 1, the α-process has a very weak mean reversion itself leading to slow changes of the original process' mean reversion level and to a pronounced (close to) random walk behaviour.

Another aspect related to the (close to) random walk behaviour is the variance of the time varying mean reversion level. If we set τ^2 to 0 the $\boldsymbol{\alpha}$-process reduces to a constant process and the resulting model is a linear AR(1) model with variance

$$\frac{\sigma^2}{(1-\beta^2)}.$$

Therefore, the second term in (6) can be regarded as an additional variance source coming from the stochastic $\boldsymbol{\alpha}$. The greater the proportion of the second term the more variance of the original process is explained by the time-varying mean reversion level and the more pronounced the (close to) random walk behaviour – under the assumption of ρ being close to 1. A prior assumption about the amount of variance coming from each of the two stochastic sources can be incorporated through the conditional prior $\beta|\sigma^2$ by imposing a minimum and a maximum share, φ_{min} and φ_{max}, of a target long run variance. Based on (6) these shares can be defined by

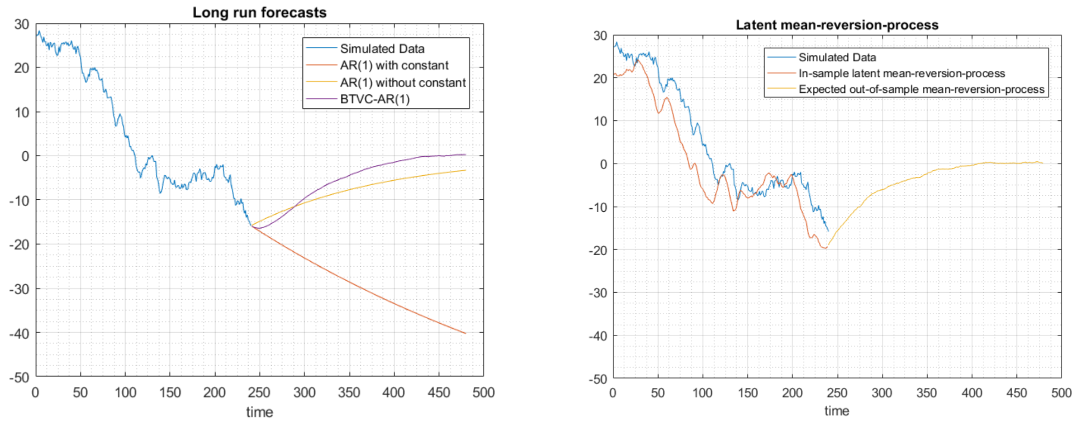
$$\varphi_{min} = \frac{\frac{\sigma^2}{(1-l(\sigma^2)^2)}}{Var(x_\mathcal{T})} \qquad \varphi_{max} = \frac{\frac{\sigma^2}{(1-u(\sigma^2)^2)}}{Var(x_\mathcal{T})},$$

where the denominator is the pre-defined long term variance and the nominator corresponds to the first term of (6) evaluated at the two truncation limits.

2.3. Model Characteristics

The time-varying α parameter in the BTVC-AR(1) Model allows to model stationary time series, which exhibit a strong autocorrelation and a (close to) random walk behaviour. In figure (1) a linear AR(1) Model and the BTVC-AR(1) Model have been exemplary fitted to a simulated stationary time series, which shows exactly such a behaviour. We estimate two AR(1) Models, one by imposing no restrictions or one by setting the constant parameter of the model to 0 to regularize the long run mean. In the left graphic the "historic" time series can be seen as well as the expected future development according to the three models. The mean reversion level of the unrestricted AR(1) Model lies far away from the historic domain. The AR(1) Model with the parameter restriction forces the time series back to the pre-specified mean reversion level.

The BTVC-AR(1) Model also has a long run mean of 0, but as the mean reversion is time-varying it allows the time series to follow the current trend for a couple of time steps before mean reverting to 0. The average latent mean reversion level extracted during the simulation process is visualised in the right picture. It can be seen that it lies below the last observation having caused the downturn in the last time steps. The mean reversion level then slowly reverts back to 0 in expectation. Because of this time-varying mean reversion level the BTVC-AR(1) Model can better account for longer deviations from the long run mean, which leads to a (close to) random walk behaviour. It further accounts for local trends leading to good short-term predictions and allows at the same time to incorporate long run assumptions about the time series in a sound mathematical approach.



(a) (b)

Figure 1: A comparison of a linear AR(1) model with no restritions for the constant parameter, a linear AR(1) model with restrictions to the constant parameter and a BTVC-AR(1) model applied on a simulated time series.

2.4. Bayesian Inference

The main parameters of interest in the BTVC-AR(1) model are $\widetilde{\boldsymbol{\alpha}}$, β and σ^2 with $\widetilde{\boldsymbol{\alpha}}$ extending $\boldsymbol{\alpha}$ by future time points up to the modelling horizon h, i.e.

$$\widetilde{\boldsymbol{\alpha}} = (\alpha_1, ..., \alpha_t, ..., \alpha_{t+h}).$$

This extension is necessary to sample from the predictive posterior distribution of the parameters and generate forecasts. The prior distribution of $\widetilde{\boldsymbol{\alpha}}$ incorporates the same assumptions as the prior distribution of $\boldsymbol{\alpha}$, i.e.

$$\widetilde{\boldsymbol{\alpha}} \sim \mathcal{N}_{t+h}(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\Sigma}})$$

where

$$\widetilde{\boldsymbol{\theta}} = \vartheta \mathbf{1} \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Pi_{t+1}} & \dots & \boldsymbol{\Pi_{t+h}} \\ \boldsymbol{\Pi}_{\boldsymbol{t+1}}^{\top} & \sigma_\alpha^2 & & \\ \vdots & & \ddots & \\ \boldsymbol{\Pi}_{\boldsymbol{t+h}}^{\top} & & & \sigma_\alpha^2 \end{pmatrix}$$

with $\boldsymbol{\Pi_{t+j}} = \{\text{Cov}(\alpha_{t+j}, \alpha_1), \dots, \text{Cov}(\alpha_{t+j}, \alpha_{t+j-1})\}^{\top}$, i.e. the vector of covariances of α_{t+j} and all previous time points $1, \dots, (t+j-1)$. For these time points the same covariance parameterization as for the prior of $\boldsymbol{\alpha}$, e.g., an autoregressive assumption, is assumed for consistency. σ_α^2 represents the unconditional variance of the latent $\boldsymbol{\alpha}$-process.

The goal of Bayesian inference is to find the joint posterior distribution, $p(\widetilde{\boldsymbol{\alpha}}, \beta, \sigma^2|x)$, conditional on the observed data $x = (x_1, ..., x_t)$. If the full conditional distribution of all parameters is known, the Gibbs sampler can be used to draw samples from this joint posterior distribution and inference can be based on Monte Carlo approximation (see e.g., [Chib, 2001]). Assuming (conditional) independence of the different model components, it holds

$$p(\widetilde{\boldsymbol{\alpha}}|\beta, \sigma^2, x) \propto p(x|\widetilde{\boldsymbol{\alpha}}, \beta, \sigma^2) \cdot p(\widetilde{\boldsymbol{\alpha}}) = \mathcal{L}(\widetilde{\boldsymbol{\alpha}}, \beta, \sigma^2) \cdot p(\widetilde{\boldsymbol{\alpha}}). \tag{9}$$

Due to the conditional independence induced by the Markov assumption in the AR(1) model the likelihood of the parameters is given by

$$\mathcal{L}(\widetilde{\boldsymbol{\alpha}}, \beta, \sigma^2) = p(x|\alpha, \beta, \sigma^2) = \prod_{j=0}^{t-1} \mathcal{G}(x_{t-j}|\alpha_{t-j} + \beta x_{t-j-1}, \sigma^2). \tag{10}$$

where $\mathcal{G}(\cdot|\mu, \widetilde{\sigma}^2)$ denotes the density function of a normal distribution with expectation μ and variance $\widetilde{\sigma}^2$. With (9) and (10) and the prior distributions specified in Section 2.1 the full conditional distributions of $\widetilde{\boldsymbol{\alpha}}$, β and σ^2 can be derived analytically. Under the assumption that $\widetilde{\boldsymbol{\theta}} = \mathbf{0}$ as specified in Section 2.2 to regularize the long run mean, the full conditional distribution of $\widetilde{\boldsymbol{\alpha}}$ is given by

$$\widetilde{\boldsymbol{\alpha}}|\beta, \sigma^2, x \sim \mathcal{N}_t(\widetilde{\boldsymbol{\mu}}_{\boldsymbol{post}}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}}).$$

with

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{post}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}} \widetilde{\boldsymbol{\delta}} \frac{1}{\sigma^2} \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}} = \left(\tilde{\boldsymbol{\Sigma}}^{-1} + \frac{1}{\sigma^2} \begin{pmatrix} \boldsymbol{I_t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right)^{-1}.$$

$\widetilde{\boldsymbol{\delta}}$ is hereby defined by

$$\widetilde{\boldsymbol{\delta}} = (x_2 - \beta x_1, \dots, x_t - \beta x_{t-1}, 0, \dots, 0).$$

As $\widetilde{\boldsymbol{\delta}}$ incorporates data information up to time point (vector entry) t, is 0 afterwards and $Cov(\alpha_{t+j}, \alpha_t) \longrightarrow 0$ with increasing j, the mean of the full conditional distribution tends to 0, corresponding to the unconditional mean of the prior distribution. The covariance structure of the full conditional distribution behaves analogously. Therefore,

the distribution of $\alpha_{t+j} \mid x, \beta, \sigma^2$ tends to the prior distribution. This means that the prior distribution of $\boldsymbol{\alpha}$ effectively regularizes the distribution of x in the long horizon towards the pre-specified mean of the latent process.
Note that the derivations are independent of the specific choice of $\boldsymbol{\Sigma}$. If prior distributions for the parameters of the $\tilde{\boldsymbol{\Sigma}}$ are employed, we need to further condition on the hyperparameters for the full conditional distribution of $\widetilde{\boldsymbol{\alpha}}$.
Under the assumption as specified in Section 2.2, i.e. $\tilde{\boldsymbol{\Sigma}}$ has an AR-covariance structure with parameters ρ and τ^2 with prior distributions (8) and (7) respectively, the full conditional distribution of ρ is given by

$$\rho|\boldsymbol{\alpha}, \tau^2 \sim \mathcal{N}(\mu_{\rho,post}, \sigma^2_{\rho,post}, -1, 1)$$

where $\sigma^2_{\rho,post} = \left(\frac{\sum_{j=0}^{t-1} \alpha^2_{t-j-1}}{\tau^2} + \sigma_\rho^{-2}\right)^{-1}$ and $\mu_{\rho,post} = \left(\frac{\sum_{j=0}^{t-1} \alpha_{t-j}\alpha_{t-j-1}}{\tau^2} + \frac{\mu_\rho}{\sigma^2_\rho}\right)\sigma^2_{\rho,post}$.

As the prior distributions of τ^2 is a degenerated distribution, the corresponding full conditional distribution is directly given.

The full conditional distribution of β is given by

$$\beta|x, \boldsymbol{\alpha}, \sigma^2 \sim \mathcal{N}(\mu_{\beta,post}, \sigma^2_{\beta,post}, l(\sigma^2), u(\sigma^2))$$

where $\sigma^2_{\beta,post} = \left(\frac{\sum_{j=0}^{t-1} x^2_{t-j-1}}{\sigma^2} + \sigma_\beta^{-2}\right)^{-1}$ and $\mu_{\beta,post} = \left(\frac{\sum_{j=0}^{t-1} \check{d}_{t-j} x_{t-j-1}}{\sigma^2} + \frac{\mu_\beta}{\sigma^2_\beta}\right)\sigma^2_{\beta,post}$.
$\check{d}_{t-j}$ is defined by

$$\check{d}_{t-j} := x_{t-j} - \alpha_{t-j}.$$

The full conditional distribution of σ^2 is given by an inverse gamma distribution with parameters $\tilde{a} = t/2 + a$ and $\tilde{b} = \frac{\sum_{j=0}^{t-1} \epsilon^2_{t-j}}{2} + b$, i.e.

$$\sigma^2|\widetilde{\boldsymbol{\alpha}}, \beta, x \sim \mathcal{IG}(\tilde{a}, \tilde{b}).$$

A more detailed derivation of the full conditional distributions can be found in Appendix A.2-A.5.

2.5. Markov Chain Monte Carlo Inference

To conduct inference, we use a Markov Chain Monte Carlo approach. More specifically, as distributions of all full conditionals are known, we use the Gibbs sampler [see, e.g., Gelman et al., 2013]. To this end, a sample from the posterior distribution is drawn by iteratively sampling from the conditional distribution of each parameter (vector). In the BTVC-AR(1) model we use the Gibbs sampler in a first step to draw from the joint posterior distribution $p(\widetilde{\boldsymbol{\alpha}}, \beta, \sigma^2 \mid x)$.
In the following we assume again an AR-covariance structure for $\boldsymbol{\Sigma}$ determined by the

parameters ρ and τ^2 with prior distributions as specified in Section 2.2. Starting with an initial sample $(\alpha^{(0)}, \beta^{(0)}, (\sigma^2)^{(0)}, \rho^{(0)}, (\tau^2)^{(0)})$ we first draw a sample of $\widetilde{\boldsymbol{\alpha}}$ values from its full conditional distribution. We proceed with a sample for ρ, σ^2 and β drawn from their full conditional distribution respectively. Finally, τ^2 is set according to (6) such that a prior specified long run variance is met. A detailed description of the algorithm can be found in the Appendix A.6.

After a burn-in period the parameter set $(\widetilde{\boldsymbol{\alpha}}^{(m)}, \beta^{(m)}, (\sigma^2)^{(m)})$ is approximately distributed according to the joint posterior distribution $p(\widetilde{\boldsymbol{\alpha}}, \beta, \sigma^2 \mid x)$. In a second and final step, we use these sample to generate paths of the x-process

$$x_{t+j}^{(m)} = \alpha_{t+j}^{(m)} + \beta^{(m)} x_{t+j-1}^{(m)} + \epsilon_{t+j}, \qquad j > 0.$$

3. Application To Interest Rate Data

We now apply the BTVC-AR(1) Model to the first principal component (PC) of a principal component analysis (PCA) on interest rate data to predict the term structure of interest rates and compare it to the *2-Additive-Factor Gaussian (Gauss2++) Model* [see, e.g., Brigo and Mercurio, 2007] and the *dynamic Nelson-Siegel Model* [Diebold and Li, 2006] with respect to the forecasting performance and the long run distribution.

The Gauss2++ model is a popular short-rate model in the insurance industry, used, e.g., to classify certified pension contracts into risk classes. Because its mean reversion level is calibrated to external interest rate forecasts, it generates realistic interest rates in the very long horizon, which is a necessary model feature for insurance companies, as they are obliged to calculate risk measures and performance scenarios for specific insurance contracts for up to 40 years [European Union, 2017]. Nevertheless, Diebold and Li point out that short-rate models perform poorly in forecasting [Diebold and Li, 2006]. Their dynamic Nelson-Siegel Model in fact shows a better forecasting performance than the Gauss2++ Model in the short horizon, but can produce unrealistic interest rates in the very long horizon. Our model, which we call the *BTVC-AR(1)-Factor Model* in the following, combines both: a good forecasting performance in the short horizon and realistic interest rates in the very long horizon. It further accounts for the strong autocorrelation and the (close to) random walk behaviour of interest rates.

3.1. Data

We use data of the German term structure of interest rates estimated by the Deutsche Bundesbank from prices of German government bonds. The exact estimation procedure can be found in [Schich, 1997]. The time span ranges from September 1997 to August 2016. Figure (2) shows the monthly evolution of the interest rate curves.

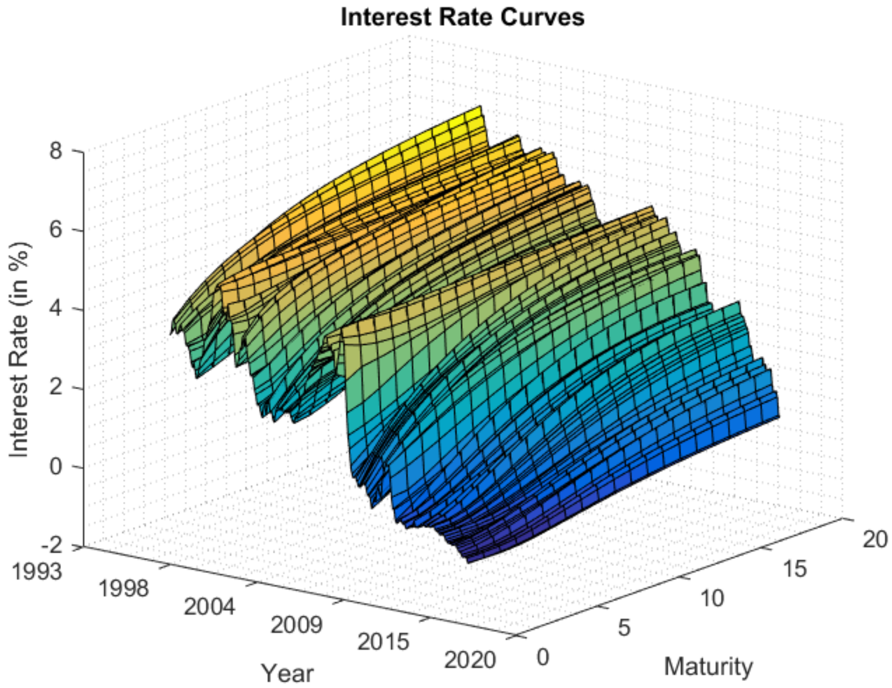


Figure 2: Time series of the term structure of German government bond yields.

In the last ten to fifteen years a decrease of the interest rates can be observed. Each maturity represents a dimension in the data set. We use PCA to reduce the dimension of the data set for the following reason. According to Litterman and Scheinkman [1991] a three factor model can explain for each interest rate with a specific maturity a minimum of 96% of the variability in the data.. We here extract these (principle) factors but only use the first two to facilitate a fair comparison with the Gauss2++ model, which is a two factor model. Furthermore, the first two PCs already account for more than 99% of the variability in the given data. Figure (3) shows the loadings and the time series of the two extracted PCs.

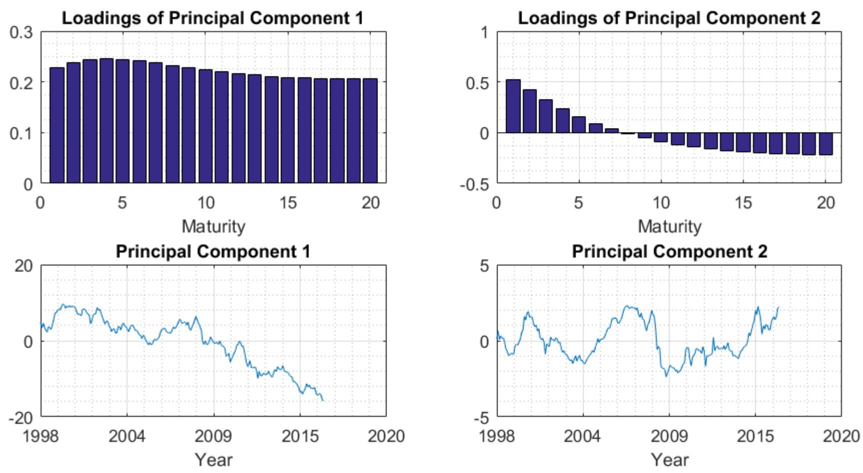


Figure 3: The scores and the loadings of the first two PCs.

The loadings of the first PC are similar for all 20 maturities, while the loadings of the second PC are positive for short and negative for long maturities. The first and the second PC are therefore often interpreted as level and slope of the term structure, respectively.

The decrease of the interest rates in the last years is also visible in the level factor showing a downward trend. There is an ongoing discussion in the literature about mean reversion of interest rates. Economic theory predominantly assumes that interest rates are (in the long run) mean reverting. But statistical evidence is not so clear [van den End, 2011]. The mainstream literature says that unit roots can not be rejected, which would imply that interest rates are not mean reverting [Siklos and Wohar, 1997], [Rose, 1988], [Stock and Watson, 1988], [Campbell and Shiller, 1991]. More recent literature investigates the unit root hypothesis by fractional integrated techniques that apply differencing to time series by an order smaller than or greater than one [Baum et al., 2000], [Gil-Alana, 2004]. These studies find that shocks to interest rates have a long memory, which explains their (close to) random walk behaviour.

3.2. Estimation of model parameters

In this Subsection the estimation of the BTVC-AR(1)-Factor Model and the two benchmark models is described.

3.2.1. Modelling interest rates with the BTVC-AR(1)-Factor model

The factors of our BTVC-AR(1)-Factor Model are the first two PCs extracted by a PCA and interpreted as level and the slope of the interest rate curve. The level factor shows a (close to) random walk behaviour, which can not be adequately captured by a stationary linear model. Following the economic theory view that interest rates (and therefore also the level) are mean reverting (in the long run), we use therefore the BTVC-AR(1) Model for this PC to account for the (close to) random walk behaviour as well as to regularize the level of the interest rate curve in the long horizon. The slope factor is more stable over time. As an augmented Dickey Fuller test suggests that the existence of a unit root can be rejected, a linear AR(1) model is used for this factor. By modelling the level and the slope factor interest rate forecasts $\hat{r}_t(\tau)$ with maturity τ can be calculated via

$$\hat{r}_t(\tau) = \mu(\tau) + \xi_1(\tau)\hat{l}_t + \xi_2(\tau)\hat{s}_t, \tag{11}$$

where $\hat{l}_t$ and $\hat{s}_t$ denote the forecasts of the level and the slope factor, respectively. $\xi_1(\tau)$ and $\xi_2(\tau)$ denote the loading of the first and second PC for maturity τ. Before applying the PCA the data has been centered and therefore $\mu(\tau)$ is the mean interest rate of the data set for maturity τ. We now specify the prior assumptions of the BTVC-AR(1) model for the level factor.

Latent AR1 constant $\boldsymbol{\alpha}$. For this application we assume an AR-covariance structure for the $\boldsymbol{\alpha}$-process of the BTVC-AR(1) Model with the parameters ρ and τ^2 representing the correlation of two successive time points and the conditional variance, respectively. The unconditional mean of the $\boldsymbol{\alpha}$-process is set to 0, which implies the assumption that the long run mean of the level factor is 0. Because we also assume that the slope factor is a centered process this means that the long run interest rate curve converges in expectation to the average interest rate curve of the dataset.

Autocorrelation parameter $\boldsymbol{\rho}$. As specified in Section 2.2 for ρ we assume a truncated normal distribution with the parameters $\mu_\rho = 0.98$ and $\sigma_\rho^2 = 0.01^2$ with lower truncation -1 and and upper truncation 1 as a hyper prior, i.e.

$$\rho \sim \mathcal{N}(0.98, 0.01, -1, 1)$$

The truncation ensures the stationarity of the process. Using this hyper-prior we incorporate the assumption of a weak mean reverting α-process into the model and therefore allow the mean reversion level of the level factor to deviate from the long run mean for longer periods. This yields the (close to) random walk behaviour present in (our) interest rate data.

Variance of the latent process. According to Section 2.2 the parameter τ^2 is set in each iteration of the Gibbs Sampler such that the long run variance of the level factor amounts to a pre-specified value. We here use the value 120, which is inferred from a quantile of the unconditional distribution. By giving consideration of the rather unusual market situation of extremely low interest rates we make the assumption that the last observation is equal to the 7.5%-quantile. As we know that due to the model's assumptions the unconditional distribution is normal with a mean of 0, the corresponding unconditional variance can be easily calculated.

Slope parameter of the AR1-model. For β we assume that $\mu_\beta = 0.95$ and $\sigma_\beta^2 = 0.1^2$ representing a weak mean reversion to the time-varying mean reversion level. Furthermore as β regularizes the amount of variance that stems from the stochastic α-process,

the lower and upper bound of its truncated normal prior is set in each sampling iteration such that $90\% - 95\%$ of the variance comes from the stochasticity of the α-process. This elaborate choice incorporates the assumption that the (close to) random walk behaviour stems from fundamental changes of the mean reversion level and that most of the deviation of the level factor from its long run mean stems from these changes.
Residual variance. For the prior distribution of σ^2 the shape and scale parameter a and b are set to 0.5 and 2 respectively, representing an uninformative prior.

By specifying the parameters of the prior (and hyper-prior) distributions the full conditional distributions can be analytically derived as described in Section 2.4 and paths of the level factor can be generated using the Gibbs Sampler as explained in Section 2.5. Forecasts of the level factor are then represented by the average of the simulated paths.

The linear AR(1) model for the slope factor is given by

$$s_t = c + \gamma s_{t-1} + \delta_t,$$

where γ is a real valued constant between -1 and 1 and δ_t is a Gaussian white noise process, i.e. $\delta_t \sim \mathcal{N}(0, \breve{\sigma}^2)$. The constant parameter c is set to 0. The other parameters are estimated by a standard ordinary least squares approach.

3.2.2. Modeling interest rates with the Gauss2++ model

The Gauss2++ Model is a popular interest rate model in the insurance industry used for pricing interest rate derivatives as well as for risk management and forecasting purposes. The model assumes that the short-rate $r(t)$, which is the interest rate with an infinitesimal small maturity, is given by the sum of two latent processes $x(t)$ and $y(t)$, and a deterministic function $\varphi(t)$:

$$r(t) = x(t) + y(t) + \varphi(t).$$

The latent processes are modelled by dependent Ornstein-Uhlenbeck processes, which are the continuous version of a linear AR(1) process. Interest rates with longer maturities are then derived from the short-rate via pricing the corresponding zero-coupon bonds, which is analytically possible due to the model's distributional assumptions.
The estimation process is materially different from the one of the other two models as it does not use historic data but calibrates the model to current future market assumptions (implicitly) provided by the current interest rate curve, interest rate derivatives as well as interest rate forecasts. By applying the downhill simplex algorithm the parameters of the model are chosen in such a way that forward rates – implicitly given by the current interest rate curve – and swaption prices are met in expectation. The data has been extracted from Bloomberg. Additionally the mean reversion level of the two latent factors are analytically set such that two interest rate forecasts with a maturity of 3 months and 10 years, which are published by the OECD, are met in expectation. This approach is in line with the standard calibration procedure in the insurance industry.

3.2.3. Modeling interest rates with the dynamic Nelson-Siegel model

The dynamic Nelson-Siegel model of Diebold and Li applies specific time series models to extracted latent factors. Diebold and Li tested several time series models on the level, slope and curvature factors of the Nelson-Siegel interest rate curve and compared the forecasting performance [Diebold and Li, 2006]. In this paper we follow one of their approaches, in which they apply a PCA on interest rate data and use an univariate linear AR(1) process for each of the first three PCs. Because of comparison reasons to the other two two-factor models in this paper, we just use the first two PCs. The parameters of the AR(1) model are estimated by the ordinary least squares method.

3.3. Backtest

We now compare the forecasting performance of the BTVC-AR(1)-Factor Model, the Gauss2++ Model and the dynamic Nelson-Siegel Model and analyse their long run distributions of the 10-year interest rate.

3.3.1. Comparison of the forecasting performance

For the out-of-sample backtest we apply an expanding window approach. The data of the first 10 years of the observations are used to estimate the parameters of all three models as described in the Section 3.2. We then forecast the interest rates for the maturities of 1, 3, 5 and 10 years (representing the interest rate curve) for the horizons of 1, 6 and 12 months. We expand the training sample by one month and repeat the procedure again. This is done until 12 months before the last observation in the data set. To evaluate the forecasting performance the error between the predicted interest rate $\hat{r}_\tau(t)$ and the actual interest rate $r_\tau(t)$ with the maturity τ is calculated, i.e.

$$error_\tau(t) = r_\tau(t) - \hat{r}_\tau(t)$$

Table (1)-(3) in the appendix A.7 show the mean and the standard deviation of this error for each model. In addition, the root mean squared error

$$RMSE(\tau) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(r_\tau(k) - \hat{r}_\tau(k))^2}. \tag{12}$$

for the given deviation is calculated, where N is the number of forecasts conducted in the backtest.

Except for the 1-month ahead forecast, the Gauss2++ Model shows the highest RMSE for all maturities. For example, the 6-month ahead forecast of the 10-year interest rate of the Gauss2++ Model has a RMSE, which is approximately twice as high as the RMSE of the other two models. For the the 12-month ahead forecast it is more than three times as high. This supports the statement of Diebold and Li that short-rate models perform poorly in forecasting [Diebold and Li, 2006]. However, it should be mentioned that the

performance of the Gauss2++ Model highly depends on the interest rate forecasts used in the calibration process. Regarding the predominant negative mean suggests that the OECD forecasts have been too optimistic in the past.
The results of the BTVC-AR(1)-Factor Model and the dynamic Nelson-Siegel Model are more similar. For the forecasting horizon of 1-month the BTVC-AR(1)-Factor Model shows a slightly lower RMSE except for the 10-year interest rate. For the 6-month and 12-month forecasting horizons the dynamic Nelson-Siegel Model shows a better performance. Note that the dynamic Nelson-Siegel model anticipated the downward trend present in the last years, which might have been beneficial in terms of the forecasting performance in the past, but also produces unrealistic interest rates in the long horizon. In contrast the BTVC-AR(1)-Factor Model forces the model to mean revert to a pre-specified level to regularize the interest rates in the long horizon. The fact that the RMSE error is similar to the dynamic Nelson-Siegel Model suggests that this does not affect the forecasting performance in the short horizon much.

3.3.2. Comparison of the distribution in the long run

We further investigate the interest rate distribution in the very long horizon. This is especially important for insurance companies as risk measures and performance scenarios for their products have to be calculated for up to 40 years [European Union, 2017]. We therefore fit all three models on all data points up to the last observation date of the data set. We then simulate paths of the 10-year interest rate and visualize the distribution in 40 years.

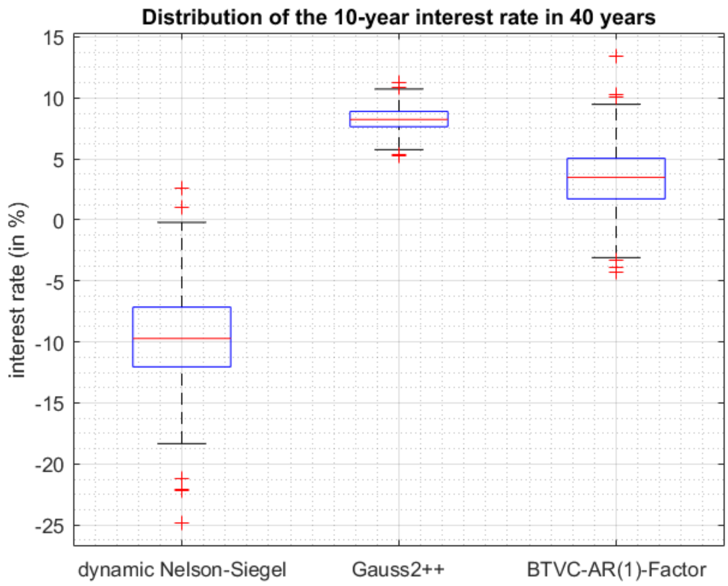


Figure 4: Comparison of the distributions of the 10-year interest rate in 40 years modelled by the dynnmic Nelson-Siegel Model, the Gauss2++ Model and the BTVC-AR(1)-Factor Model.

The median of the dynamic Nelson-Siegel Model amounts to approximately -10%. A value that is not realistic for the 10-year interest rate. In comparison, the distribution of the BTVC-AR(1)-Factor model and the Gauss2++ model seem to be more realistic as the range of their distributions is (mainly) positive between 0% and 10%. It can be observed that the standard deviation of the Gauss2++ Model is much smaller than of the BTVC-AR(1)-Factor Model and as the median is quite high negative values are not reached by this model. This is due to the fact that the Gauss2++ Model assumes a stronger mean reversion than historic data would suggest. The (close to) random walk behaviour is better captured by the BTVC-AR(1)-Factor Model leading to a prediction range which fits historic observations quite well. This is due to the regularization of the mean and the standard deviation of the BTVC-AR(1)-Factor Model induced by appropriate prior assumptions, which represents the main difference to other interest rate models.

4. Conclusion

In this paper we introduced a new Bayesian framework for the TVC-AR(1) Model particularly suitable for nearly integrated time series which can not be estimated by a linear model consistent with economic theory or historic observations. In these cases a (close to) random walk behaviour can be an indication for a missing variable, for which we account for by the usage of a non-linear model. The time-varying constant of the BTVC-AR(1) allows a stochastic mean reversion level leading to realizations, which exhibit a random walk behaviour although being stationary and do not have an exploding long run variance. Additionally, with the Bayesian approach it is possible to incorporate prior assumption about the long run distribution into the model without affecting the short-term predictions adversely. This gives the possibility to include expert knowledge or well known economic facts about the long term behaviour of the time series into the model that is otherwise fully data-driven in the short term forecast.

We apply the proposed approach to interest rate data. We find that the BTVC-AR(1)-Factor Model, which applies a BTVC-AR(1) Model to the first PC of a PCA, shows a similar forecasting performance as the dynamic Nelson-Siegel Model in the short horizon but in contrast produces realistic interest rates in the very long horizon and also yields better forecasts compared to the Gauss2++ Model, the industry standard interest rate model in the insurance industry.

The presented framework allows for many different specifications and is flexible regarding the assumed covariance structure of the latent α process in the model. In this paper we propose an AR-covariance structure and explain how model parameters can be inferred in this special case. Investigating other covariance structures could further improve the forecasting performance in the short horizon while still regularizing the distribution in the long run.

References

David W Bacon and Donald G Watts. Estimating the transition between two intersecting straight lines. *Biometrika*, 58(3):525–534, 1971.

Christopher F Baum, Basma Bekdache, et al. A re-evaluation of empirical tests of the fisher hypothesis. *Working Papers in Economics*, page 148, 2000.

Damiano Brigo and Fabio Mercurio. *Interest rate models – theory and practice: with smile, inflation and credit.* Springer Science & Business Media, 2007.

John Y Campbell and Robert J Shiller. Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58(3):495–514, 1991.

Kung Sik Chan and Howell Tong. On estimating thresholds in autoregressive models. *Journal of time series analysis*, 7(3):179–190, 1986.

Siddhartha Chib. Markov chain monte carlo methods: computation and inference. In *Handbook of econometrics*, volume 5, pages 3569–3649. Elsevier, 2001.

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

European Union. Commission delegated regulation (eu) 2017/653 of 8 march 2017 supplementing regulation (eu) no 1286/2014 of the european parliament and of the council on key information documents for packaged retail and insurance-based investment products (priips) by laying down regulatory technical standards with regard to the presentation, content, review and revision of key information documents and the conditions for fulfilling the requirement to provide such documents. *Official Journal of the European Union*, 60, 2017.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis.* Chapman and Hall/CRC, 2013.

Luis A Gil-Alana. Long memory in the us interest rate. *International Review of Financial Analysis*, 13(3):265–276, 2004.

Stephen Goldfeld and Richard Quandt. The estimation of structural shifts by switching regressions. In *Annals of Economic and Social Measurement, Volume 2, number 4*, pages 475–485. NBER, 1973.

Clive William John Granger, GRANGER CWJ, and ANDERSEN AP. An introduction to bilinear time series models. 1978.

Valérie Haggan and Tohru Ozaki. Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, 68(1):189–196, 1981.

James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384, 1989.

Eilev S Jansen and Timo Teräsvirta. Testing parameter constancy and super exogeneity in econometric equations. *Oxford Bulletin of Economics and Statistics*, 58(4):735–763, 1996.

Markku Lanne and Pentti Saikkonen. Threshold autoregressions for strongly autocorrelated time series. *Journal of Business & Economic Statistics*, 20(2):282–289, 2002.

KS Lim and H Tong. Threshold autoregressions, limit cycles, and data. *Journal of the Royal Statistical Sociaty, B*, 42:245–92, 1980.

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

James C Morley. Nonlinear time series in macroeconomics. *Encyclopedia of Complexity and System Science, forthcoming*, 2009.

Richard E Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the american statistical association*, 53 (284):873–880, 1958.

Andrew K Rose. Is the real interest rate stable? *The Journal of Finance*, 43(5):1095–1112, 1988.

Sebastian Schich. Schätzung der deutschen zinsstrukturkurve. *Bundesbank Series 1 Discussion Paper*, 1997.

Pierre L Siklos and Mark E Wohar. Convergence in interest rates and inflation rates across countries and over time. *Review of International Economics*, 5(1):129–141, 1997.

James H Stock and Mark W Watson. Testing for common trends. *Journal of the American statistical Association*, 83(404):1097–1107, 1988.

T Subba Rao and MM Gabr. An introduction to bispectral analysis and bilinear. 1984.

Timo Teräsvirta. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the american Statistical association*, 89(425):208–218, 1994.

Timo Teräsvirta, Dag Tjøstheim, Clive William John Granger, et al. *Modelling nonlinear economic time series*. Oxford University Press Oxford, 2010.

Jan Willem van den End. Statistical evidence on the mean reversion of interest rates. *De Nederlandsche Bank Working Paper*, 2011.

A. Appendix

A.1. Rewriting the likelihood of the parameters

By defining $\breve{d}_{t-j} := x_{t-j} - \alpha_{t-j}$, the likelihood of the parameters can be reformulated as follows:

$$\begin{aligned}
\mathcal{L}(\beta, \alpha, \sigma^2) &= \prod_{j=0}^{t-1} \mathcal{G}(x_{t-j}|\alpha_{t-j} + \beta x_{t-j-1}, \sigma^2) \\
&= \prod_{j=0}^{t-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_{t-j} - \alpha_{t-j} - \beta x_{t-j-1})^2}{2\sigma^2}\right) \\
&= \prod_{j=0}^{t-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\breve{d}_{t-j} - \beta x_{t-j-1})^2}{2\sigma^2}\right) \\
&= \prod_{j=0}^{t-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\breve{d}_{t-j}^2 - 2\beta x_{t-j-1}\breve{d}_{t-j} + \beta^2 x_{t-j-1}^2}{2\sigma^2}\right) \\
&\propto \prod_{j=0}^{t-1} \exp\left(-\frac{1}{2\sigma^2}\left\{-2\beta x_{t-j-1}\breve{d}_{t-j} + \beta^2 x_{t-j-1}^2\right\}\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}\left\{-2\beta \sum_{j=0}^{t-1} \breve{d}_{t-j} x_{t-j-1} + \beta^2 \sum_{j=0}^{t-1} x_{t-j-1}^2\right\}\right).
\end{aligned}$$

A.2. The full conditional distribution of $\widetilde{\alpha}$

The prior distribution of $\widetilde{\boldsymbol{\alpha}}$ is a centered Gaussian process with a specific covariance structure, i.e.

$$\widetilde{\boldsymbol{\alpha}} = (\alpha_1, ..., \alpha_t, ..., \alpha_{t+h}) \sim \mathcal{N}_t(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}})$$

The following derivations will be independent of the specific choice of $\widetilde{\boldsymbol{\Sigma}}$. We further define $Var(\alpha_s) = \tau^2 \ \forall s \geq 1$. By defining

$$\delta_j = x_{j+1} - \beta x_j$$

as well as $\boldsymbol{\delta} = (\delta_0, \ldots, \delta_{t-1})^\top$ and the fact that

$$\mathcal{G}(x_t|\alpha_t + \beta x_{t-1}, \sigma^2) = \mathcal{G}(\alpha_t|\delta_{t-1}, \sigma^2)$$

allows a straightforward derivation of the full conditional of α:

$$\begin{aligned}
p(\widetilde{\boldsymbol{\alpha}}|\beta,\sigma^2,\boldsymbol{x}) &\propto p(\boldsymbol{x}|\widetilde{\boldsymbol{\alpha}},\beta,\sigma^2)p(\widetilde{\boldsymbol{\alpha}}|\beta,\sigma^2) \\
&\propto p(\boldsymbol{x}|\boldsymbol{\alpha},\beta,\sigma^2)p(\widetilde{\boldsymbol{\alpha}}) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\alpha}-\boldsymbol{\delta})^\top(\boldsymbol{\alpha}-\boldsymbol{\delta})\right)\cdot\exp\left(-\frac{1}{2}\widetilde{\boldsymbol{\alpha}}^\top\widetilde{\boldsymbol{\Sigma}}^{\boldsymbol{-1}}\widetilde{\boldsymbol{\alpha}}\right) \\
&\propto \exp\left(-\frac{1}{2}(\widetilde{\boldsymbol{\alpha}}^\top\widetilde{\boldsymbol{\Sigma}}^{\boldsymbol{-1}}_{\boldsymbol{post}}\widetilde{\boldsymbol{\alpha}}-2\widetilde{\boldsymbol{\alpha}}^\top\widetilde{\boldsymbol{\Sigma}}^{\boldsymbol{-1}}_{\boldsymbol{post}}\underbrace{\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}}\widetilde{\boldsymbol{\delta}}_{\boldsymbol{0}}\frac{1}{\sigma^2}}_{=:\widetilde{\boldsymbol{\mu}}_{\boldsymbol{post}}})\right)
\end{aligned}$$

with $\widetilde{\boldsymbol{\Sigma}}^{\boldsymbol{-1}}_{\boldsymbol{post}}=\widetilde{\boldsymbol{\Sigma}}^{\boldsymbol{-1}}+\frac{1}{\sigma^2}\begin{pmatrix}\boldsymbol{I_t} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}$ and $\widetilde{\boldsymbol{\delta}}_{\boldsymbol{0}}=(\boldsymbol{\delta}^\top,\boldsymbol{0})^\top$.

This is the kernel of a multivariate Gaussian distribution with covariance $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}}$ and mean vector $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{post}}$, i.e

$$\widetilde{\boldsymbol{\alpha}}\mid\beta,\sigma^2,\boldsymbol{x}\sim\mathcal{N}(\widetilde{\boldsymbol{\mu}}_{\boldsymbol{post}},\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{post}}).$$

A.3. The Full conditional Distribution of ρ

If an AR-covariance structure is assumed for $\widetilde{\boldsymbol{\alpha}}$ the latent α-process can be written in the following form

$$\alpha_t=\rho\alpha_{t-1}+\eta_t,$$

where ρ determines the correlation between two successive time steps and η_t is a Gaussian white noise process, i.e. $\eta_t\sim\mathcal{N}(0,\tau^2)$.
The full conditional distribution of ρ can be therefore derived as follows:

$$p(\rho|\tau^2,\boldsymbol{\alpha})\propto\mathcal{L}(\rho,\tau^2)\cdot p(\rho)=\prod_{j=0}^{t-1}\mathcal{G}(\alpha_{t-j}|\rho\alpha_{t-j-1},\tau^2)\cdot p(\rho). \tag{13}$$

The likelihood $\mathcal{L}(\cdot)$ in the above equation can be reformulated as

$$\mathcal{L}\propto\exp\left(-\frac{1}{2\tau^2}\left\{-2\rho\left[\sum_{j=0}^{t-1}\alpha_{t-j}\alpha_{t-j-1}\right]+\rho^2\left[\sum_{j=0}^{t-1}\alpha^2_{t-j-1}\right]\right\}\right).$$

The calculation is similar to the one in appendix A.1. Defining the two terms in square brackets as η and χ, respectively, we get

$$\mathcal{L}\propto\exp\left(-\frac{1}{2\tau^2}\left\{-2\rho\eta+\rho^2\chi\right\}\right).$$

Plugging this into (13) and using a normal prior with parameters μ_ρ,σ^2_ρ for ρ, we have

$$\begin{aligned}
p(\rho|\tau^2,\alpha) &\propto \exp\left(-\frac{1}{2}\left\{\frac{\rho^2\chi}{\tau^2}-2\frac{\rho\eta}{\tau^2}\right\}\right)\exp\left(-\frac{1}{2}\left\{\frac{\rho^2}{\sigma^2_\rho}-2\frac{\rho\mu_\rho}{\sigma^2_\rho}\right\}\right) \\
&\propto \exp\left(-\frac{1}{2}\left\{\rho^2\cdot\left(\frac{\chi}{\tau^2}+\sigma_\rho^{-2}\right)-2\rho\left(\frac{\eta}{\tau^2}+\frac{\mu_\rho}{\sigma^2_\rho}\right)\right\}\right)
\end{aligned}$$

and thus $\rho|\tau^2, \alpha \sim \mathcal{N}(\mu_{\rho,post}, \sigma^2_{\rho,post})$ with

$$\sigma^2_{\rho,post} = \left(\frac{\chi}{\tau^2} + \sigma_\rho^{-2}\right)^{-1}$$

and

$$\mu_{\rho,post} = \left(\frac{\eta}{\tau^2} + \frac{\mu_\rho}{\sigma_\rho^2}\right) \sigma^2_{\rho,post}.$$

If a truncated normal prior is used, the truncation is transferred to the full conditional distribution.

A.4. The Full Conditional Distribution of β

Analogously to (9) and (10) we have

$$p(\beta|\boldsymbol{\alpha}, \sigma^2, x) \propto \mathcal{L}(\beta, \boldsymbol{\alpha}, \sigma^2) \cdot p(\beta) = \prod_{j=0}^{t-1} \mathcal{G}(x_{t-j}|\alpha_{t-j} + \beta x_{t-j-1}, \sigma^2) \cdot p(\beta). \tag{14}$$

By defining $\breve{d}_{t-j} := x_{t-j} - \alpha_{t-j}$ and as $\mathcal{G}(x_{t-j}|\alpha_{t-j} + \beta x_{t-j-1}, \sigma^2) = \mathcal{G}(\beta x_{t-j-1}|\breve{d}_{t-j}, \sigma^2)$ the likelihood $\mathcal{L}(\cdot)$ in the above equation can be reformulated as

$$\mathcal{L} \propto \exp\left(-\frac{1}{2\sigma^2}\left\{-2\beta\left[\sum_{j=0}^{t-1} \breve{d}_{t-j} x_{t-j-1}\right] + \beta^2 \left[\sum_{j=0}^{t-1} x^2_{t-j-1}\right]\right\}\right).$$

You can find a more detailed calculation in A.1. Defining the two terms in square brackets as η and χ, respectively, we get

$$\mathcal{L} \propto \exp\left(-\frac{1}{2\sigma^2}\left\{-2\beta\eta + \beta^2\chi\right\}\right).$$

Plugging this into (14) and using a normal prior with parameters $\mu_\beta, \sigma^2_\beta$ for β, we have

$$\begin{aligned} p(\beta|\alpha, \sigma^2, x) &\propto \exp\left(-\frac{1}{2}\left\{\frac{\beta^2\chi}{\sigma^2} - 2\frac{\beta\eta}{\sigma^2}\right\}\right) \exp\left(-\frac{1}{2}\left\{\frac{\beta^2}{\sigma^2_\beta} - 2\frac{\beta\mu_\beta}{\sigma^2_\beta}\right\}\right) \\ &\propto \exp\left(-\frac{1}{2}\left\{\beta^2 \cdot \left(\frac{\chi}{\sigma^2} + \sigma_\beta^{-2}\right) - 2\beta\left(\frac{\eta}{\sigma^2} + \frac{\mu_\beta}{\sigma^2_\beta}\right)\right\}\right) \end{aligned}$$

and thus $\beta|x, \alpha, \sigma^2 \sim \mathcal{N}(\mu_{\beta,post}, \sigma^2_{\beta,post})$ with

$$\sigma^2_{\beta,post} = \left(\frac{\chi}{\sigma^2} + \sigma_\beta^{-2}\right)^{-1}$$

and

$$\mu_{\beta,post} = \left(\frac{\eta}{\sigma^2} + \frac{\mu_\beta}{\sigma^2_\beta}\right) \sigma^2_{\beta,post}.$$

If a truncated normal prior is used, the truncation is transferred to the full conditional distribution.

A.5. The Full Conditional Distribution of σ^2

In this Section we derive the full conditional distribution of σ^2. As before

$$p(\sigma^2|\boldsymbol{\alpha}, \beta, x) \propto \prod_{j=0}^{t-1} \mathcal{G}(x_{t-j}|\alpha_{t-j} + \beta x_{t-j-1}, \sigma^2) \cdot p(\sigma^2),$$

which is equal to

$$(\sigma^2)^{-(t/2)} \exp\left(-\frac{1}{2\sigma^2}\sum_{j=0}^{t-1}\epsilon_{t-j}^2\right) \cdot p(\sigma^2) =: (\sigma^2)^{-(t/2)} \exp\left(-\frac{1}{2\sigma^2}\kappa\right) \cdot p(\sigma^2).$$

By using an inverse gamma distribution with shape and scale parameters a, b, or short $\mathcal{IG}(a, b)$, for the prior of σ^2 we get

$$(\sigma^2)^{-(t/2)} \exp\left(-\frac{1}{2\sigma^2}\kappa\right) \cdot (\sigma^2)^{-(a+1)} \exp(-b/\sigma^2)$$

and thus also an inverse gamma distribution as posterior with parameters $\tilde{a} = t/2 + a$ and $\tilde{b} = \kappa/2 + b$, i.e.

$$\sigma^2|\boldsymbol{\alpha}, \beta, x \sim \mathcal{IG}(\tilde{a}, \tilde{b}).$$

A.6. Gibbs Sampler algorithm

Algorithm 1

1: Simulate from the full conditional distribution of $\widetilde{\boldsymbol{\alpha}}$ by using the latest sample of the other parameters:

- $\widetilde{\boldsymbol{\alpha}}^{\boldsymbol{(m+1)}} \sim p(\widetilde{\boldsymbol{\alpha}} \mid \rho^{(m)}, (\tau^2)^{(m)}, \beta^{(m)}, (\sigma^2)^{(m)}, x)$

2: Simulate from the full conditional distribution of ρ, the hyper parameter of $\widetilde{\boldsymbol{\alpha}}$, by using the latest sample of $\boldsymbol{\alpha}$ and τ^2:

- $\rho^{(m+1)} \sim p(\rho \mid \boldsymbol{\alpha}^{\boldsymbol{(m+1)}}, (\tau^2)^{(m)}, x)$

3: Simulate from the full conditional distribution of σ^2 by using the latest sample of $\boldsymbol{\alpha}$ and β^2:

- $(\sigma^2)^{(m+1)} \sim p((\sigma^2)^{(m+1)} \mid \boldsymbol{\alpha}^{(m+1)}, \beta^{(m)}, x)$

4: Simulate from the full conditional distribution of $\beta|\sigma^2$ by using the latest sample of $\boldsymbol{\alpha}$ and σ^2:

- $\beta^{(m+1)} \sim p(\beta \mid \boldsymbol{\alpha}^{(m+1)}, (\sigma^2)^{(m+1)}, x)$

5: Set $(\tau^2)^{(m+1)}$ according to (6) such that a prior specified long run variance is met.

A.7. Backtest Results

Maturity	Mean	Std. Dev.	RMSE
The BTVC-AR(1)-Factor Model			
1 year	-0.0126	0.2561	0.0651
3 year	-0.0302	0.2311	0.0538
5 year	-0.0505	0.2433	0.0611
10 year	-0.0466	0.2383	0.0584
The Gauss2++ Model			
1 year	-0.0808	0.2361	0.0618
3 year	-0.1037	0.2252	0.0610
5 year	-0.1203	0.2139	0.0598
10 year	-0.1429	0.2130	0.0654
The dynamic Nelson-Siegel Model			
1 year	-0.0290	0.2615	0.0685
3 year	-0.0462	0.2311	0.0550
5 year	-0.0653	0.2410	0.0617
10 year	-0.0589	0.2340	0.0577

Table 1: Results of the out-of-sample 1-month ahead forecasting.

Maturity	Mean	Std. Dev.	RMSE
The BTVC-AR(1)-Factor Model			
1 year	-0.2320	0.8105	0.7040
3 year	-0.2462	0.7538	0.6230
5 year	-0.2615	0.6954	0.5470
10 year	-0.2388	0.5996	0.4129
The Gauss2++ Model			
1 year	-0.4094	0.8105	0.8184
3 year	-0.5402	0.6768	0.7457
5 year	-0.6098	0.6090	0.7393
10 year	-0.6545	0.5824	0.7693
The dynamic Nelson-Siegel Model			
1 year	-0.2900	0.7857	0.6951
3 year	-0.3022	0.7045	0.5825
5 year	-0.3130	0.6380	0.5008
10 year	-0.2812	0.5446	0.3727

Table 2: Results of the out-of-sample 6-month ahead forecasting.

Maturity	Mean	Std. Dev.	RMSE
The BTVC-AR(1)-Factor Model			
1 year	-0.5570	1.0310	1.3623
3 year	-0.5628	0.8727	1.0705
5 year	-0.5750	0.7757	0.9262
10 year	-0.5379	0.6865	0.7557
The Gauss2++ Model			
1 year	-0.9047	1.0709	1.9546
3 year	-1.1531	0.7939	1,9541
5 year	-1.2745	0.7255	2.1458
10 year	-1.3345	0.8060	2.4246
The dynamic Nelson-Siegel Model			
1 year	-0.6004	0.9961	1.3424
3 year	-0.6024	0.8218	1.0316
5 year	-0.6098	0.7096	0.8702
10 year	-0.5657	0.6024	0.6793

Table 3: Results of the out-of-sample 12-month ahead forecasting.