Bayesian Analysis (2021) **16**, Number 2, pp. 459–487

Dynamic Regression Models for Time-Ordered Functional Data

Daniel R. Kowal*

Abstract. For time-ordered functional data, an important yet challenging task is to forecast functional observations with uncertainty quantification. Scalar predictors are often observed concurrently with functional data and provide valuable information about the dynamics of the functional time series. We develop a fully Bayesian framework for dynamic functional regression, which employs scalar predictors to model the time-evolution of functional data. Functional within-curve dependence is modeled using unknown basis functions, which are learned from the data. The unknown basis provides substantial dimension reduction, which is essential for scalable computing, and may incorporate prior knowledge such as smoothness or periodicity. The dynamics of the time-ordered functional data are specified using a time-varying parameter regression model in which the effects of the scalar predictors evolve over time. To guard against overfitting, we design shrinkage priors that regularize irrelevant predictors and shrink toward time-invariance. Simulation studies decisively confirm the utility of these modeling and prior choices. Posterior inference is available via a customized Gibbs sampler, which offers unrivaled scalability for Bayesian dynamic functional regression. The methodology is applied to model and forecast yield curves using macroeconomic predictors, and demonstrates exceptional forecasting accuracy and uncertainty quantification over the span of four decades.

Keywords: Bayesian methods, factor model, forecasting, shrinkage, yield curve.

1 Introduction

In business, science, and industrial applications, data are commonly measured over a continuous domain, such as time, space, or wavelength. *Functional data analysis* treats these data as a single realization of an underlying function, and seeks to model the associations among distinct functional observations. Functional data present significant challenges for modeling and prediction: the data are usually highly-correlated, high-dimensional, and often unevenly-spaced over the domain. We address the setting of *time-ordered* functional data, where the primary goal is to forecast future functional observations with uncertainty quantification. Examples of time-ordered functional data include yearly sea surface temperature as a function of time-of-year (Besse et al., 2000), daily pollution curves as a function of time-of-day (Damon and Guillas, 2002; Aue et al., 2015), yearly mortality rates as a function of age (Hyndman and Ullah, 2007), and yearly disease counts as a function of time-of-year (Kowal, 2019).

Forecasting and inference for functional time series data require adequate modeling of both the *functional* within-curve dependence and the *dynamic* between-curve dependence. Simultaneous identification of prominent functional features together with their

*Assistant Professor, Department of Statistics, Rice University, Houston, TX, daniel.kowal@rice.edu

 https://doi.org/10.1214/20-BA1213

temporal behaviors provides a recipe for understanding and predicting function-valued time series. To assist in this challenging task, *scalar predictors* are commonly recorded concurrently with functional data, and may provide information about the time evolution of the functional data. The utility of scalar predictors for modeling functional data has been demonstrated extensively in many instances of *function-on-scalars regression* (Ramsay and Silverman, 2005; Montagna et al., 2012; Morris, 2015; Chen et al., 2016; Barber et al., 2017; Fan and Reimherr, 2017; Kowal and Bourgeois, 2019). However, limited attention has been given to *dynamic* regression models for time-ordered functional data, in particular with the primary goal of forecasting functional data.

For scalar time series data, *time-varying parameter (TVP) regression models* have shown the ability to produce accurate forecasts with precise uncertainty quantification (Dangl and Halling, 2012; Korobilis, 2013; Belmonte et al., 2014; Kowal et al., 2019). In TVP regression, the scalar response is linearly associated with scalar predictors, such that the linear associations evolve dynamically over time. For financial and economic data in particular, the time-variation is essential: abrupt changes in regulations and policies, as well as gradual changes in market sentiments and the broader macroeconomy, may drive an evolving relationship between the predictors and the response (Dangl and Halling, 2012). TVP regression is a special case of a dynamic linear (or state space) model (West and Harrison, 1997), and benefits from the computational and inferential capabilities of these models.

Despite the impressive gains offered by TVP regression models, functional data present new and unique challenges. Model complexity—and risk of overparametrization—is greatly increased: each time-varying parameter is now a time-varying *function*. Function estimation implicitly requires regularization, usually via smoothness or sparsity, which must be combined with the simultaneous shrinkage of irrelevant predictors and extraneous time-variation. Equally important, the data dimensionality in the functional data setting presents a significant challenge for Bayesian computing, and therefore impedes access to the forecasting distribution.

To address these challenges, we develop a dynamic Bayesian model for functional regression and forecasting. We extract prominent functional features by modeling each functional observation using *unknown* basis functions, which are learned from the data. The unknown basis provides a lower-dimensional space in which to model the time-variation of the functional data and greatly improves computational scalability. The dynamics of the time-ordered functional observations are modeled using a TVP regression with autoregressive innovations, which incorporates both predictor-driven and latent dynamics. The resulting model allows the scalar predictors to inform the time-evolution of the functional response, specifically corresponding to the learned functional features, with time-variation in the key parameters for greater adaptability. Shrinkage priors are included to simultaneously (i) ensure smoothness of the basis functions, (ii) regularize irrelevant predictors, (iii) shrink toward time-invariance for model parsimony, and (iv) reduce sensitivity to the dimension of the unknown basis. Full posterior inference and multi-step forecasting distributions are available via an efficient Gibbs sampler. Simulation studies confirm the importance of these modeling choices, and in particular the model for the basis functions, the time-variation of model parameters, and the shrinkage priors.

The proposed methodology is applied to model and forecast U.S. yield curve data. For a given currency and level of risk of a debt, the yield curve describes the interest rate at a given time as a *function* of the length of the borrowing period, or time to maturity, and evolves over *time*. The yield curve is observed concurrently with financial and macroeconomic predictor variables, which may assist in constructing point and interval forecasts of the yield curve. While a variety of techniques are available for yield curve modeling, to the best of our knowledge there is no existing framework that simultaneously includes (i) a nonparametric regression model for unknown basis functions, (ii) dynamic regression coefficients, and (iii) Bayesian shrinkage priors for regularization. We evaluate the proposed framework for yield curve modeling and forecasting over the span of four decades, and demonstrate substantial gains in forecasting accuracy and uncertainty quantification relative to state-of-the-art and benchmark competitors.

The remainder of the paper is organized as follows: we introduce the model in Section 2; the unknown basis function model is in Section 3; the shrinkage priors are in Section 4; the MCMC algorithm is in Section 5; a simulation analysis is in Section 6; the yield curve forecasting comparison is in Section 7; and we conclude in Section 8. Supplementary files include `R` code and a document with additional details on the MCMC algorithm, the basis expansions, and the simulations (Kowal 2020).

2 Dynamic function-on-scalars regression

Let $\{Y_t\}_{t=1}^T$ be a time-ordered sequence of functions on a compact index set $\mathcal{T} \subset \mathbb{R}^D$ with $D \in \mathbb{Z}^+$. Each function is observed at a discrete set of points $\boldsymbol{\tau}_j \in \mathcal{T}, j = 1, \ldots, M$, which are fixed across all times t only for notational convenience. The observed data are noisy realizations of each Y_t at these observation points:

$$y_{j,t} = Y_t(\boldsymbol{\tau}_j) + \epsilon_{j,t}, \quad \mathbb{E}(\epsilon_{j,t}) = 0, \tag{1}$$

where $\epsilon_{j,t}$ represents observation or measurement error. Since the functions Y_t are typically modeled as smooth in $\boldsymbol{\tau}$, the errors $\epsilon_{j,t}$ account for non-smoothness of the observed data. Notably, omission of observation error for time-ordered functional data can induce misspecified dynamics for $\{Y_t\}$ (Kowal et al., 2017b). We assume conditionally Gaussian observation errors with time-varying variance, $[\epsilon_{j,t}|\sigma^2_{\epsilon_t}] \sim N(0, \sigma^2_{\epsilon_t})$, with alternative distributional assumptions in Lemma 2. The dynamic variance $\sigma^2_{\epsilon_t}$ allows the variability about Y_t to change over time, and is essential for constructing precise and accurate yield curve forecasting intervals (see Section 7).

Suppose that at each time t, the functional data $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{M,t})'$ are observed concurrently with time-ordered predictors, $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{p,t})'$. Note that we do *not* require $p < M$. Our goal is to leverage these dynamic predictors $\boldsymbol{x}_t$ to produce superior forecasting distributions for $\boldsymbol{y}_t$, and to construct inferential summaries of the associations between $\boldsymbol{x}_t$ and Y_t. For time-ordered data, it is likely that associations among variables change over time. Failure to incorporate such time-variation will produce inferior forecasts and unreliable inference. It is also expected that some predictors may not be useful in forecasting Y_t, while others may only provide predictive power for certain

features or subdomains of $\{Y_t\}$. Consequently, it is important to include mechanisms for regularization of irrelevant predictor variables.

Motivated by these considerations, we propose the following *dynamic function-on-scalars regression* (DFOSR) model:

$$Y_t(\boldsymbol{\tau}) = \sum_{k=1}^{K} f_k(\boldsymbol{\tau})\beta_{k,t}, \quad \boldsymbol{\tau} \in \mathcal{T}, \tag{2}$$

$$\beta_{k,t} = \mu_k + \sum_{j=1}^{p} x_{j,t}\alpha_{j,k,t} + \gamma_{k,t}, \quad k = 1, \ldots, K, \tag{3}$$

$$\alpha_{j,k,t} = \alpha_{j,k,t-1} + \omega_{j,k,t}, \quad j = 1, \ldots, p, \quad k = 1, \ldots, K, \tag{4}$$

$$\gamma_{k,t} = \phi_k \gamma_{k,t-1} + \eta_{k,t} \quad k = 1, \ldots, K \tag{5}$$

for $t = 1, \ldots, T$. The functions Y_t are decomposed into a linear combination of K *loading curves* $\{f_k(\boldsymbol{\tau})\}_{k=1}^K$ and *dynamic factors* $\{\beta_{k,t}\}_{k=1}^K$. The basis expansion (2) is a core component of many functional data models. In Section 3, we develop a nonparametric regression approach for modeling each f_k as unknown. As a result, the functional basis $\{f_k\}$ is data-adaptive, with each f_k learning an essential functional feature of $\{Y_t\}$.

The dynamic factors $\{\beta_{k,t}\}$ model the time dependence among the functions $\{Y_t\}$, and specifically correspond to the functional features defined by each f_k. Naturally, modeling the dynamics of $\{\beta_{k,t}\}$ is fundamental for constructing accurate forecasts of the functional time series data $\boldsymbol{y}_t$. We incorporate the dynamic predictors $\boldsymbol{x}_t$ into the model for $\beta_{k,t}$ via (3)–(4), which includes an intercept μ_k, dynamic regression coefficients $\{\alpha_{j,k,t}\}$, and an error term $\gamma_{k,t}$ for each factor k. Note that dynamics on the intercept μ_k are not needed, since these effects appear in $\gamma_{k,t}$. The regression coefficients $\{\alpha_{j,k,t}\}$ are time-varying for each predictor j and factor k in (4), which implies time-variation for the functional regression model. The regression model also features an autoregressive time series model for the error terms in (5), which accounts for time-variation in $\{Y_t\}$ that is not fully explained by the predictors $\boldsymbol{x}_t$.

The DFOSR has an equivalent representation which makes explicit the association between the factor-specific parameters in (3)–(5) and the functions Y_t. Let $\mathcal{GP}(c, C)$ denote a Gaussian process with mean function c and covariance function C. By substituting (3)–(5) into (2), the DFOSR model implies the dynamic functional regression model

$$Y_t(\boldsymbol{\tau}) = \tilde{\mu}(\boldsymbol{\tau}) + \sum_{j=1}^{p} x_{j,t}\tilde{\alpha}_{j,t}(\boldsymbol{\tau}) + \tilde{\gamma}_t(\boldsymbol{\tau}) \quad \boldsymbol{\tau} \in \mathcal{T}, \tag{6}$$

$$\tilde{\gamma}_t(\boldsymbol{\tau}) = \int \tilde{\phi}(\boldsymbol{\tau}, \boldsymbol{u})\tilde{\gamma}_{t-1}(\boldsymbol{u})\, d\boldsymbol{u} + \tilde{\eta}_t(\boldsymbol{\tau}), \quad \tilde{\eta}_t(\cdot) \overset{indep}{\sim} \mathcal{GP}(0, C_{\eta_t}), \tag{7}$$

$$\tilde{\alpha}_{j,t}(\boldsymbol{\tau}) = \tilde{\alpha}_{j,t-1}(\boldsymbol{\tau}) + \tilde{\omega}_t(\boldsymbol{\tau}), \quad \tilde{\omega}_{j,t}(\cdot) \overset{indep}{\sim} \mathcal{GP}(0, C_{\omega_{j,t}}), \tag{8}$$

where each functional term is a linear combination of the $\{f_k\}$, or explicitly, $\tilde{\mu}(\boldsymbol{\tau}) := \sum_k f_k(\boldsymbol{\tau})\mu_k$, $\tilde{\alpha}_{j,t}(\boldsymbol{\tau}) := \sum_k f_k(\boldsymbol{\tau})\alpha_{j,k,t}$, $\tilde{\gamma}_t(\boldsymbol{\tau}) := \sum_k f_k(\boldsymbol{\tau})\gamma_{k,t}$, $\tilde{\phi}(\boldsymbol{\tau}, \boldsymbol{u}) :=$

$\sum_k f_k(\boldsymbol{\tau}) f_k(\boldsymbol{u}) \phi_k$, $\tilde{\eta}_t(\boldsymbol{\tau}) := \sum_k f_k(\boldsymbol{\tau}) \eta_{k,t}$, $\tilde{\omega}_{j,t}(\boldsymbol{\tau}) := \sum_k f_k(\boldsymbol{\tau}) \omega_{j,k,t}$, with covariance functions $C_{\eta_t}(\boldsymbol{\tau}, \boldsymbol{u}) := \sum_k f_k(\boldsymbol{\tau}) f_k(\boldsymbol{u}) \sigma^2_{\eta_{k,t}}$ and $C_{\omega_{j,t}}(\boldsymbol{\tau}, \boldsymbol{u}) := \sum_k f_k(\boldsymbol{\tau}) f_k(\boldsymbol{u}) \sigma^2_{\omega_{j,k,t}}$.

The predictors $x_{j,t}$ are directly associated with the functional time series $Y_t(\boldsymbol{\tau})$ via the dynamic regression coefficient functions $\tilde{\alpha}_{j,t}(\boldsymbol{\tau}) := \sum_k f_k(\boldsymbol{\tau}) \alpha_{j,k,t}$. The error term $\tilde{\gamma}_t(\boldsymbol{\tau})$ models both *functional* dependence in $\boldsymbol{\tau}$ and *dynamic dependence* via a functional autoregressive model in (7), which has been shown to offer strong forecasting performance in the absence of predictors (Kowal et al., 2017b). Note that the auto- and cross-covariance functions of $Y_t(\boldsymbol{\tau})$ are available as a special case of Propositions 1 and 2 in Kowal (2019), and do *not* require separability in $\boldsymbol{\tau}$ and t.

The DFOSR model is highly flexible: the dynamic regression coefficients are permitted to change at every time t, while each functional feature f_k is associated with predictors $\boldsymbol{x}_t$ via (3) and has distinct idiosyncratic dynamics via $\gamma_{k,t}$ (5). However, without adequate regularization, there is a risk of overparametrization: it is unlikely that every predictor $x_{j,t}$ is associated with every factor $\beta_{k,t}$, and that this association changes substantially at each time t. Overparametrization is a particular concern for forecasting, and can lead to increases in both bias and variance. In Section 4, we introduce Bayesian shrinkage priors which simultaneously regularize against irrelevant predictors, unnecessary time-variation, and sensitivity to the dimension K of the unknown basis. These priors are represented as conditionally Gaussian distributions for the innovations $\omega_{j,k,t}$ and $\eta_{k,t}$, which preserves computational scalability of the state space model (2)–(5).

Note that our setting is similar to, but distinct from, longitudinal functional data analysis (Greven et al., 2011; Chen and Müller, 2012; Park and Staicu, 2015). Longitudinal functional data are time-ordered functional data, but typically include replicates of each functional time series (e.g., time-ordered functions are observed for each of many subjects).

3 Modeling the loading curves

The loading curves $\{f_k\}$ provide the foundation for functional modeling and forecasting of Y_t. The dynamics of the DFOSR model are determined by $\beta_{k,t}$ in (3)–(5), while each $\beta_{k,t}$ is mapped to the functional domain $\boldsymbol{\tau} \in \mathcal{T}$ via the corresponding $f_k(\boldsymbol{\tau})$ in (2). Clearly, any functional features that are not effectively captured by $\{f_k\}$ cannot be forecasted accurately by the DFOSR and inferentially cannot be linked to the predictors. Therefore, it is essential that the loadings $\{f_k\}$ adequately capture the functional dependence among the $\{Y_t\}$.

There is an unavoidable tradeoff between flexibility of the basis and parsimony. As the dimension K increases, the model accounts for more functional features, yet the TVP regression model in (3)–(5) becomes more heavily parametrized. Furthermore, computational scalability of the DFOSR model—which may be expressed as a state space model in $\{\alpha_{k,j,t}, \gamma_{k,t}\}$ (see Section 5)—is limited by K. Methods that use a fixed basis expansion in (2), such as splines (Laurini, 2014) or wavelets (Morris and Carroll, 2006), require a large K for adequate modeling flexibility, and therefore produce heavily

parametrized and computationally infeasible models when combined with the dynamics of (3)–(5). An appealing solution is to model the basis functions $\{f_k\}$ as unknown, which allows the curves $\{f_k\}$ to be optimized for each functional dataset. However, we must be careful to ensure that the uncertainty regarding the unknown $\{f_k\}$ is not ignored.

3.1 Priors and full conditional distributions for the loading curves

We propose a model for the loading curves $\{f_k\}$ that simultaneously (i) treats $\{f_k\}$ as unknown, which produces a data-adaptive basis and minimizes the number of necessary basis functions K; (ii) accounts for the inherent uncertainty in $\{f_k\}$; (iii) is scalable in the number of observation points, M; and (iv) is well-defined for $\mathcal{T} \subset \mathbb{R}^D$ with $D \in \mathbb{Z}^+$. Identifiability constraints are enforced on $\{f_k\}$ to preserve interpretability of the loading curves $\{f_k\}$ and factors $\{\beta_{k,t}\}$ in (2). These constraints, which are detailed in Section 3.2, are designed to simplify posterior sampling for the factor-specific parameters in (3)–(5), which greatly improves computational scalability.

A common approach in nonparametric regression is to represent each unknown function—here, each f_k—as a linear combination of known basis functions, and then model the corresponding unknown basis coefficients. Let $f_k(\boldsymbol{\tau}) = \boldsymbol{b}'(\boldsymbol{\tau})\boldsymbol{\psi}_k$, where $\boldsymbol{b}'(\boldsymbol{\tau}) := (b_1(\boldsymbol{\tau}), \ldots, b_{L_M}(\boldsymbol{\tau}))$ is an L_M-dimensional vector of known basis functions and $\boldsymbol{\psi}_k$ is an L_M-dimensional vector of unknown basis coefficients. We use low-rank thin plate splines (LR-TPS), which are smooth, flexible, and well-defined for $\mathcal{T} \subset \mathbb{R}^D$ with $D \in \mathbb{Z}^+$. However, modifications for other bases such as wavelets, Fourier expansions, and Gaussian processes are straightforward and follow the same construction as below. Given functional data observations $\boldsymbol{y}_t := (y_{1,t}, \ldots, y_{M,t})'$, we consolidate (1)–(2) to relate the unknown coefficients $\{\boldsymbol{\psi}_k\}$ to the observed data $\boldsymbol{y}_t$:

$$\boldsymbol{y}_t = \sum_{k=1}^{K} \boldsymbol{f}_k \beta_{k,t} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{indep}{\sim} N(\boldsymbol{0}, \sigma^2_{\epsilon_t} \boldsymbol{I}_M), \tag{9}$$

where $\boldsymbol{f}_k := (f_k(\boldsymbol{\tau}_1), \ldots, f_k(\boldsymbol{\tau}_M))' = \boldsymbol{B}\boldsymbol{\psi}_k$, $\boldsymbol{B} := (\boldsymbol{b}(\tau_1), \ldots, \boldsymbol{b}(\tau_M))'$ is the $M \times L_M$ basis matrix, and $\boldsymbol{\epsilon}_t := (\epsilon_{1,t}, \ldots, \epsilon_{M,T})'$.

To encourage smoothness of f_k, LR-TPS accompany the basis expansion with a roughness penalty. In the case of $D = 1$, the penalty is of the form $\mathcal{P}(f_k) = \int \{\ddot{f}_k(\tau)\}^2 d\tau$ for $\ddot{f}_k$ the second derivative of f_k; for $D > 1$, derivatives are replaced by gradients (see Wood, 2006 for details). For Bayesian implementations, the penalty corresponds to a prior distribution on the basis coefficients $\boldsymbol{\psi}_k$, or equivalently, the implied function f_k. Consider the conditionally Gaussian prior

$$[\boldsymbol{\psi}_k | \lambda_{f_k}] \overset{indep}{\sim} N(\boldsymbol{0}, \lambda_{f_k}^{-1} \boldsymbol{\Omega}_b^-), \quad k = 1, \ldots, K, \tag{10}$$

where $\lambda_{f_k} > 0$ is a prior precision parameter and $\boldsymbol{\Omega}_b$ is a penalty matrix with (ℓ, ℓ') entry $[\boldsymbol{\Omega}_b]_{\ell,\ell'} = \int \ddot{b}_\ell(\tau) \ddot{b}_{\ell'}(\tau) d\tau$. The implied log-posterior distribution for $\{f_k\}$ is then

$$-2 \log p(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_K | \{\boldsymbol{y}_t\}, -) \overset{c}{=} \sum_{t=1}^{T} \sigma^{-2}_{\epsilon_t} \left|\left| \boldsymbol{y}_t - \sum_{k=1}^{K} \boldsymbol{f}_k \beta_{k,t} \right|\right|^2 + \sum_{k=1}^{K} \lambda_{f_k} \int \{\ddot{f}_k(\tau)\}^2 d\tau, \tag{11}$$

which is a penalized least squares criterion, for which the precisions λ_{f_k} are smoothing parameters for each f_k (and $\stackrel{c}{=}$ denotes equality up to a constant). Details on the construction of $\boldsymbol{B}$ and $\boldsymbol{\Omega}_b$ and the choice of L_M for LR-TPS are provided in the supplement. The generalized inverse in (10) is standard for Bayesian splines, since $\boldsymbol{\Omega}_b$ has rank $L_M - 2$ and is not invertible. In our implementation (see the supplement), which diagonalizes $\boldsymbol{\Omega}_b$ and orthogonalizes $\boldsymbol{B}$, the improper prior is replaced with a proper prior to ensure propriety of the posterior.

For posterior inference, we design a Bayesian backfitting algorithm that iteratively draws from the full conditional distribution of each f_k. The iterative approach is useful for enforcing the identifiability constraints in Section 3.2, and leads to straightforward full conditional distributions: $[\boldsymbol{\psi}_k|\cdots] \sim N\big(\boldsymbol{Q}_{\psi_k}^{-1}\boldsymbol{\ell}_{\psi_k}, \boldsymbol{Q}_{\psi_k}^{-1}\big)$, where $\boldsymbol{Q}_{\psi_k} := (\boldsymbol{B}'\boldsymbol{B})\sum_{t=1}^{T}\big(\beta_{k,t}^2/\sigma_{\epsilon_t}^2\big) + \lambda_{f_k}\boldsymbol{\Omega}_b$ and $\boldsymbol{\ell}_{\psi_k} := \boldsymbol{B}'\sum_{t=1}^{T}\big\{(\beta_{k,t}/\sigma_{\epsilon_t}^2)\big(\boldsymbol{y}_t - \sum_{\ell\neq k}\boldsymbol{f}_\ell\beta_{\ell,t}\big)\big\}$. Note that the full conditional distribution of $\boldsymbol{f}_k = \boldsymbol{B}\boldsymbol{\psi}_k$ clearly depends on the dynamic factors $\{\beta_{k,t}\}$, which are learned simultaneously based on model (3)–(5). The prior precision λ_{f_k} is modeled on the standard deviation scale, $\lambda_{f_k}^{-1/2} \sim \text{Uniform}(0, 10^4)$ (Kowal et al., 2017a). By assigning a prior distribution to each λ_{f_k}, we allow the data to inform the degree of smoothness for each f_k—without needing a cross-validation step—while incorporating the uncertainty about λ_{f_k} into the posterior distribution.

Existing methods for learning $\{f_k\}$ generally fall into two categories: functional principal components analysis (FPCA) and Bayesian reduced rank models. In FPCA, the basis functions $\{f_k\}$ are estimated in advance of model-fitting (Goldsmith and Kitago, 2016). However, this approach implicitly requires a multi-step estimation procedure, whereby the FPCs are estimated separately from—and without utilizing—the remainder of the model in (3)–(4). Dynamic FCPA (Hörmann et al., 2015) incorporates temporal structure, but is primarily designed for dimension reduction rather than forecasting or regression. When pre-computed FPCs are treated as fixed and known, the ignored uncertainty is often nontrivial: Goldsmith et al. (2013) demonstrate that FPC-based methods can substantially underestimate total variability, even for densely-observed functional data. Bayesian reduced-rank functional data models such as Montagna et al. (2012) and Suarez et al. (2017) are similar to FPCA, yet incorporate the uncertainty of $\{f_k\}$ into the posterior distribution. Unfortunately, these approaches have limited computational scalability, and are not designed to incorporate the dynamic regression model (3)–(5) or produce forecasts of $\boldsymbol{y}_t$.

3.2 Simplifying the likelihood via identifiability constraints

We enforce identifiability constraints on the loading curves, $\{f_k\}$, which primarily serves two purposes. First, identifiability is necessary to interpret $\{f_k\}$ and the k-specific model parameters in (3) and (4). Second, our particular choice of constraints provides computational improvements for sampling the dynamic parameters in (3) and (4). We constrain $\boldsymbol{F}'\boldsymbol{F} = \boldsymbol{I}_K$, where $\boldsymbol{F} := (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_K)$ is the $M \times K$ matrix of loading curves evaluated at the observation points $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_M$ and $\boldsymbol{I}_K$ is the $K \times K$ identity matrix. This constraint, combined with a suitable ordering constraint on $k = 1, \ldots, K$ (see Section 4), is suffi-

cient for identifiability (up to sign changes, which in our experience are not problematic in the MCMC sampler).

While various identifiability constraints are available, the following result illustrates the utility of the proposed approach:

Lemma 1. *Under the identifiability constraint* $\boldsymbol{F}'\boldsymbol{F} = \boldsymbol{I}_K$ *and model* (9)*, the joint full conditional distribution for* $\{\beta_{k,t}\}$ *is proportional to*

$$p(\boldsymbol{\beta}|\boldsymbol{y}, -) \propto p(\boldsymbol{\beta})p(\boldsymbol{y}|-) \propto p(\boldsymbol{\beta}) \exp\Big\{-\sum_{t=1}^{T} \frac{1}{2\sigma^2_{\epsilon_t}}\big(||\boldsymbol{\beta}_t||^2 - 2\tilde{\boldsymbol{y}}_t'\boldsymbol{\beta}_t\big)\Big\} \propto p(\boldsymbol{\beta})\tilde{p}(\tilde{\boldsymbol{y}}|-), \quad (12)$$

where $\tilde{\boldsymbol{y}}_t' := (\tilde{y}_{1,t}, \ldots, \tilde{y}_{K,t})'$ *for* $\tilde{y}_{k,t} := \boldsymbol{f}_k'\boldsymbol{y}_t$ *and* $\tilde{p}(\tilde{\boldsymbol{y}}|-)$ *is the* working likelihood *defined by*

$$\tilde{y}_{k,t} = \beta_{k,t} + \tilde{\epsilon}_{k,t}, \quad \tilde{\epsilon}_{k,t} \overset{indep}{\sim} N(0, \sigma^2_{\epsilon_t}). \quad (13)$$

For sampling the dynamic parameters in (3)–(5), the full functional data likelihood under model (9) is unnecessary, and may be replaced by the simpler working likelihood under model (13). Consequently, all computations involving these parameters depend on M only via the projection $\tilde{y}_{k,t} = \boldsymbol{f}_k'\boldsymbol{y}_t$, which is a one-time cost per MCMC iteration. These simplifications facilitate the inclusion of complex dynamics in (3)–(4) without sacrificing computational feasibility.

Applying Lemma 1 to model (3)–(4), the TVP functional regression simplifies:

$$\tilde{y}_{k,t} = \mu_k + \begin{pmatrix}\boldsymbol{x}_t' & 1\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}_{k,t}\\ \gamma_{k,t}\end{pmatrix} + \tilde{\epsilon}_{k,t}, \quad (14)$$

$$\begin{pmatrix}\boldsymbol{\alpha}_{k,t}\\ \gamma_{k,t}\end{pmatrix} = \begin{pmatrix}\boldsymbol{I}_p & 0\\ 0 & \phi_k\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}_{k,t-1}\\ \gamma_{k,t-1}\end{pmatrix} + \begin{pmatrix}\boldsymbol{\omega}_{k,t}\\ \eta_{k,t}\end{pmatrix} \quad (15)$$

for $k = 1, \ldots, K$, where $\boldsymbol{\alpha}_{k,t} := (\alpha_{1,k,t}, \ldots, \alpha_{p,k,t})'$ and the errors $\tilde{\epsilon}_{k,t}$ and $(\boldsymbol{\omega}_{k,t}', \eta_{k,t})'$ are mutually independent and conditionally Gaussian (see Section 4). Model (14)–(15) is a *dynamic linear model* (West and Harrison, 1997) in the state variables $(\boldsymbol{\alpha}_{k,t}', \gamma_{k,t})'$. From Lemma 1, the joint full conditional distribution of the dynamic variables $\{\boldsymbol{\alpha}_{k,t}, \gamma_{k,t}\}_{k,t}$ using (14)–(15) is identical to that based on the functional data likelihood (9) with model (3)–(4). However, model (14)–(15) is much simpler, with a lower-dimensional data vector in (14) compared to (9), which offers massive computational improvements for sampling the state variables $\{\boldsymbol{\alpha}_{k,t}, \gamma_{k,t}\}_{k,t}$. Generalizations of the evolution equation (15), such as a non-diagonal evolution matrix, are straightforward within the proposed framework.

As an empirical illustration, Table 1 gives computation times for simulated data from Section 6 for the proposed DFOSR model compared to Kowal et al. (2017a) (referred to as DFOSR-NIG in Section 6). In particular, Kowal et al. (2017a) use a similar model for $\{f_k\}$, but do *not* use the identifiability constraint $\boldsymbol{F}'\boldsymbol{F} = \boldsymbol{I}_K$ and cannot apply Lemma 1. The improvements are substantial, particularly for the larger sample size.

This approach may be generalized to account for alternative dependence structures in the observation error, $\boldsymbol{\epsilon}_t$:

MCMC Algorithm	$T = 50, M = 20$	$T = 200, M = 100$
Proposed DFOSR	48 seconds	3 minutes
Kowal et al. (2017b)	15 minutes	74 minutes

Table 1: Computing times per 1000 MCMC iterations (implemented in `R` on a MacBook Pro, 2.7 GHz Intel Core i5). In all cases, $p = 15$ and $K = 6$.

Lemma 2. *For model (9) with the generalization* $[\boldsymbol{\epsilon}_t | \sigma^2_{\epsilon_t}, \boldsymbol{\Sigma}_\epsilon] \overset{indep}{\sim} N(\mathbf{0}, \sigma^2_{\epsilon_t}\boldsymbol{\Sigma}_\epsilon)$*, the simplification in Lemma 1 remains valid under the modified constraint* $\boldsymbol{F}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{F} = \boldsymbol{I}_K$.

Most importantly, Lemma 2 accommodates within-curve dependence of the errors $\epsilon_t(\boldsymbol{\tau}_j) := \epsilon_{j,t}$, such as autoregressive or non-smooth dependence in $\boldsymbol{\tau}$. Unlike the functions Y_t in (2), the error functions $\epsilon_t(\boldsymbol{\tau})$—and therefore the functional data realizations $y_t(\boldsymbol{\tau}_j) := y_{j,t}$—are not restricted to be low rank. However, the errors $\boldsymbol{\epsilon}_t$ remain conditionally independent across time $t = 1, \ldots, T$.

Proceeding under the setting of Lemma 1, we decompose the orthonormality constraint for each $\boldsymbol{f}_k$ into two sets of constraints: the linear constraints $\boldsymbol{f}_\ell'\boldsymbol{f}_k = 0$ for $\ell \neq k$ and the unit-norm constraint, $||\boldsymbol{f}_k||^2 = 1$. The sampler in Section 3.1 conditions on $\{\boldsymbol{f}_\ell\}_{\ell \neq k}$, so the linearity constraint is fixed for each $\boldsymbol{f}_k = \boldsymbol{B}\boldsymbol{\psi}_k$. Given the full conditional distribution $[\boldsymbol{\psi}_k | \cdots] \sim N\big(\boldsymbol{Q}_{\psi_k}^{-1}\boldsymbol{\ell}_{\psi_k}, \boldsymbol{Q}_{\psi_k}^{-1}\big)$ from Section 3.1, we enforce the linear orthogonality constraint by *conditioning* on $\boldsymbol{C}_k\boldsymbol{f}_k = \mathbf{0}$, where $\boldsymbol{C}_k = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{k-1}, \boldsymbol{f}_{k+1}, \ldots, \boldsymbol{f}_K)'$. Conditioning on the constraint is a natural Bayesian approach, and produces desirable optimality properties for constrained penalized regression (see Theorem 1 of Kowal et al., 2017a). Given an orthogonally-constrained sample $\boldsymbol{f}_k^* := \boldsymbol{B}\boldsymbol{\psi}_k^*$, which is obtained via Theorem 1 below, we rescale to enforce the unit-norm constraint: $\boldsymbol{f}_k = \boldsymbol{f}_k^*/||\boldsymbol{f}_k^*||$, and similarly rescale $\boldsymbol{\psi}_k^*$. This rescaling does not change the shape of the loading curve $\boldsymbol{f}_k$, and can be counterbalanced by an equivalent rescaling of the corresponding factor, i.e., $\beta_{k,t} \leftarrow \beta_{k,t}||\boldsymbol{f}_k^*||$. By applying this procedure iteratively for $k = 1, \ldots, K$, the constraint $\boldsymbol{F}'\boldsymbol{F} = \boldsymbol{I}_K$ is satisfied for every MCMC iteration, which is required to obtain the computational simplifications of Lemma 1. Alternative approaches for directly sampling $\boldsymbol{F}$ on the Stiefel manifold are promising, but require customized algorithms (Jauch et al., 2019).

More generally, we may condition on *any* linear constraints $\boldsymbol{C}_k\boldsymbol{f}_k = \mathbf{0}$ for each $k = 1, \ldots, K$. For example, the constraint $f_k(\boldsymbol{\tau}_1) = f_k(\boldsymbol{\tau}_M)$ ensures a periodic curve f_k, while the constraints $f_k(\boldsymbol{\tau}^*) = 0$ for $\boldsymbol{\tau}^* \in \mathcal{T}^*$ restrict the support of f_k to a subdomain of $\mathcal{T}^* \subset \mathcal{T}$. Incorporating constraints, when appropriate, reduces variability and enhances interpretability of the loading curves $\{f_k\}$. Within the proposed framework, conditioning on the linear constraints produces several important properties:

Theorem 1. *Let* $\boldsymbol{C}_k$ *be a* $J \times M$ *matrix of rank* J*. Denote* $\boldsymbol{\psi}_k^0 \sim N\big(\boldsymbol{Q}_{\psi_k}^{-1}\boldsymbol{\ell}_{\psi_k}, \boldsymbol{Q}_{\psi_k}^{-1}\big)$ *with* $\boldsymbol{f}_k^0 := \boldsymbol{B}\boldsymbol{\psi}_k^0$ *and define* $\boldsymbol{\psi}_k := \boldsymbol{\psi}_k^0 - \boldsymbol{Q}_{\psi_k}^{-1}\boldsymbol{B}'\boldsymbol{C}_k'\big(\boldsymbol{C}_k\boldsymbol{B}\boldsymbol{Q}_{\psi_k}^{-1}\boldsymbol{B}'\boldsymbol{C}_k'\big)^{-1}\boldsymbol{C}_k\boldsymbol{B}\boldsymbol{\psi}_k^0$ *with* $\boldsymbol{f}_k := \boldsymbol{B}\boldsymbol{\psi}_k$ *for* $k = 1, \ldots, K$*. The following properties hold:*

1. $\boldsymbol{f}_k \overset{d}{=} [\boldsymbol{f}_k^0 | \boldsymbol{C}_k\boldsymbol{f}_k^0 = \mathbf{0}]$;

2. $\mathbb{P}(\boldsymbol{C}_k \boldsymbol{f}_k = \boldsymbol{0}) = 1$ *for* $k = 1, \ldots, K$;

3. *If* $\boldsymbol{C}_k = \boldsymbol{C}$ *for all* k, *then* $\boldsymbol{C}\mathbb{E}(\boldsymbol{y}_t) = \mathbb{E}(\boldsymbol{C}\boldsymbol{y}_t) = \boldsymbol{C}\boldsymbol{Y}_t = \boldsymbol{0}$ *for all* $t = 1, \ldots, T$ *under model* (9), *where* $\boldsymbol{Y}_t = (Y_t(\boldsymbol{\tau}_1,) \ldots, Y_t(\boldsymbol{\tau}_M))'$.

Proof. Denote $\boldsymbol{f}_k^0 \sim N(\boldsymbol{\mu}_{f_k^0}, \boldsymbol{\Sigma}_{f_k^0})$ with $\boldsymbol{\mu}_{f_k^0} := \boldsymbol{B}\boldsymbol{Q}_{\psi_k}^{-1}\boldsymbol{\ell}_{\psi_k}$ and $\boldsymbol{\Sigma}_{f_k^0} := \boldsymbol{B}\boldsymbol{Q}_{\psi_k}^{-1}\boldsymbol{B}'$ the full conditional distribution of the (unconstrained) loading curve. Since $\boldsymbol{f}_k^0$ is Gaussian, so is $\boldsymbol{C}_k\boldsymbol{f}_k^0$, and consequently $[\boldsymbol{f}_k^0 | \boldsymbol{C}_k\boldsymbol{f}_k^0 = \boldsymbol{0}] \sim N(\boldsymbol{\mu}_{f_k}, \boldsymbol{\Sigma}_{f_k})$ with $\boldsymbol{\mu}_{f_k} := \boldsymbol{\mu}_{f_k^0} - \boldsymbol{\Sigma}_{f_k^0}\boldsymbol{C}_k'(\boldsymbol{C}_k\boldsymbol{\Sigma}_{f_k^0}\boldsymbol{C}_k')^{-1}\boldsymbol{C}_k\boldsymbol{\mu}_{f_k^0}$ and $\boldsymbol{\Sigma}_{f_k} := \boldsymbol{\Sigma}_{f_k^0} - \boldsymbol{\Sigma}_{f_k^0}\boldsymbol{C}_k'(\boldsymbol{C}_k\boldsymbol{\Sigma}_{f_k^0}\boldsymbol{C}_k')^{-1}\boldsymbol{\Sigma}_{f_k^0}$. Direct calculation shows that $\boldsymbol{f}_k$ has the same mean and covariance. Lastly, we simplify $\mathbb{E}(\boldsymbol{C}\boldsymbol{y}_t) = \boldsymbol{C}\mathbb{E}(\boldsymbol{y}_t) = \boldsymbol{C}\boldsymbol{Y}_t = \boldsymbol{C}\sum_{k=1}^K \boldsymbol{f}_k\beta_{k,t} = \sum_{k=1}^K(\boldsymbol{C}\boldsymbol{f}_k)\beta_{k,t} = 0$ for each $t = 1, \ldots, T$. □

Theorem 1 provides a recipe for generating constrained loading curves $\{f_k\}$ based on a draw from the unconstrained full conditional distribution. The resulting $\boldsymbol{f}_k$ is a sample from the requisite full conditional distribution restricted to the linear (orthogonality) constraint. Since the constraint holds almost surely, it follows that $\boldsymbol{C}_k\boldsymbol{f}_k = \boldsymbol{0}$ for each draw of $\boldsymbol{f}_k$, and therefore $\boldsymbol{f}_\ell'\boldsymbol{f}_k = 0$ for $\ell \neq k$ at every MCMC iteration.

The final property of Theorem 1 is broadly useful, since it provides a mechanism for applying constraints simultaneously to the loading curves $\{f_k\}$ and the expected functional data response, Y_t. For example, the proposed modeling framework is applicable for periodic functional data, such as daily pollution curves (Damon and Guillas, 2002; Aue et al., 2015), by constraining $f_k(\boldsymbol{\tau}_1) = f_k(\boldsymbol{\tau}_M)$, which guarantees that $Y_t(\boldsymbol{\tau}_1) = Y_t(\boldsymbol{\tau}_M)$ for all t. The unknown basis functions $\{f_k\}$ are still learned from the data, yet the posterior distribution properly accounts for the pre-specified structure implied by $\boldsymbol{C}\boldsymbol{f}_k = \boldsymbol{0}$ and consequently $\boldsymbol{C}\boldsymbol{Y}_t = \boldsymbol{0}$.

4 Shrinkage priors

While the DFOSR model provides substantial dynamic modeling flexibility, there is also a risk of overparametrization. The model may contain irrelevant predictors or unnecessary time-variation, and may be sensitive to the dimension K of the unknown basis. We introduce regularization via Bayesian shrinkage priors. In (non-functional) TVP regression, shrinkage priors offer improvements in prediction (Korobilis, 2013; Belmonte et al., 2014) and provide narrower posterior credible intervals (Kowal et al., 2019). Importantly, these effects are confirmed in the functional data setting for both simulated and real data (see Sections 6 and 7): the shrinkage priors improve forecast accuracy, point estimation accuracy, and calibration and precision of posterior uncertainty quantification.

Since many shrinkage priors may be expressed as conditionally Gaussian distributions (Polson and Scott, 2010, 2012; Kowal et al., 2019), we consider priors of the form

$$[\omega_{j,k,t} | \sigma^2_{\omega_{j,k,t}}] \overset{indep}{\sim} N(0, \sigma^2_{\omega_{j,k,t}}), \quad [\eta_{k,t} | \sigma^2_{\eta_{k,t}}] \overset{indep}{\sim} N(0, \sigma^2_{\eta_{k,t}}) \tag{16}$$

for all j, k, t. Importantly, the conditional Gaussianity in (16) maintains computational scalability of fully Bayesian inference for the DFOSR model (see Section 5). The priors on the variances $\sigma^2_{\omega_{j,k,t}}$ and $\sigma^2_{\eta_{k,t}}$ will determine the shrinkage effects for $\omega_{j,kt}$ and $\eta_{k,t}$.

The innovations $\omega_{j,k,t}$ control the time-variation in the regression coefficients $\alpha_{j,k,t}$: when $|\omega_{j,k,t}|$ is small, the change $|\alpha_{j,k,t} - \alpha_{j,k,t-1}|$ is small, so that the regression coefficients remain approximately constant at time t. In the absence of definitive evidence for time-variation of $\alpha_{j,k,t}$, the more parsimonious locally static model with $|\omega_{j,k,t}| = |\alpha_{j,k,t} - \alpha_{j,k,t-1}| \approx 0$ is preferred. Similarly, if predictor $x_{j,t}$ is irrelevant for factor $\beta_{k,t}$, the prior should globally shrink each $|\omega_{j,k,t}|$ and the initial state $|\alpha_{j,k,0}|$ to zero, which effectively removes $x_{j,t}$ from the model for $\beta_{k,t}$.

To achieve both local (in time t) and global (for factor k, predictor j) shrinkage of $\alpha_{j,k,t}$, we propose *nested horseshoe priors*, which extend the horseshoe prior of Carvalho et al. (2010) to the functional time series setting. By design, the horseshoe prior aggressively shrinks small (absolute) values toward zero, while large (absolute) values receive minimal shrinkage. The horseshoe prior has demonstrated broad success in applications, simulations, and theory (Polson and Scott, 2010; Datta and Ghosh, 2013; van der Pas et al., 2014). We apply shrinkage at multiple levels with the following hierarchy of half-Cauchy distributions:

$$\sigma_{\omega_{j,k,t}} \overset{ind}{\sim} C^+(0, \lambda_{j,k}), \quad \lambda_{j,k} \overset{ind}{\sim} C^+(0, \lambda_j), \quad \lambda_j \overset{ind}{\sim} C^+(0, \lambda_0),$$
$$\lambda_0 \overset{ind}{\sim} C^+(0, 1/\sqrt{T-1}). \tag{17}$$

First, $\sigma_{\omega_{j,k,t}} \approx 0$ implies that $|\omega_{j,k,t}| \approx 0$, so $\alpha_{j,k,t} \approx \alpha_{j,k,t-1}$ is locally constant. Each $\alpha_{j,k,t}$ for predictor j and factor k may vary at any time t, but the prior encourages most changes to be approximately negligible, which implies fewer effective parameters in the model. The shrinkage parameters $\lambda_{j,k}$ and λ_j are common for all times t, and provide factor- and predictor-specific shrinkage: for each predictor j, $\lambda_{j,k}$ allows some factors k to be nonzero, while λ_j operators as a group shrinkage parameter that may effectively remove predictor j from the model. Lastly, the global shrinkage parameter λ_0 controls the global level of sparsity, and is scaled by $1/\sqrt{T-1}$ following Piironen and Vehtari (2016). In the case of the non-dynamic FOSR and FOSR-AR models, we simply remove one level of the hierarchy: $\omega_{j,k,t} \sim N(0, \lambda^2_{j,k})$.

To reduce sensitivity to the dimension K of the basis, we introduce *ordered* shrinkage across $k = 1, \ldots, K$ for both the innovations $\eta_{k,t}$ and the intercepts μ_k in (3). For this purpose, we use *multiplicative gamma process* (MGP) priors (Bhattacharya and Dunson, 2011), which assign increasing shrinkage toward zero for larger values of k. MGP priors are a popular choice among Bayesian models with unknown rank K, including factor models (Bhattacharya and Dunson, 2011) and functional regression models (Montagna et al., 2012; Kowal and Bourgeois, 2019). The MGP prior for the intercept terms is given by $\mu_k \sim N(0, \sigma^2_{\mu_k})$ with $\sigma^{-2}_{\mu_k} = \prod_{\ell \leq k} \delta_{\mu_\ell}$, for $\delta_{\mu_1} \sim \text{Gamma}(a_{\mu_1}, 1)$ and $\delta_{\mu_\ell} \sim \text{Gamma}(a_{\mu_2}, 1)$ when $\ell > 1$. Selecting $a_{\mu_1} > 0$ and $a_{\mu_2} \geq 2$ produces stochastic ordering among the implied variances $\sigma^2_{\mu_k}$ (Bhattacharya and Dunson, 2011; Durante, 2017), which also satisfies the ordering requirement for model identifiability. Similarly, for the innovations $\eta_{k,t} \sim N(0, \sigma^2_{\eta_{k,t}})$, let $\sigma^2_{\eta_{k,t}} = \sigma^2_{\eta_k}/\xi_{\eta_{kt}}$ with $\sigma^{-2}_{\eta_k} = \prod_{\ell \leq k} \delta_{\eta_\ell}$, $\delta_{\eta_1} \sim$

Gamma$(a_{\eta_1}, 1)$, $\delta_{\eta_\ell} \sim$ Gamma$(a_{\eta_2}, 1)$ for $\ell > 1$, and $\xi_{\eta_{kt}} \sim$ Gamma$(\nu_\eta/2, \nu_\eta/2)$. The rate of ordered shrinkage is determined by the data separately for $\{\mu_k\}$ and $\{\eta_{k,t}\}$ using the hyperpriors $a_{\mu_1}, a_{\mu_2}, a_{\eta_1}, a_{\eta_2} \sim$ Gamma$(2, 1)$. Finally, the hyperprior $\nu_\eta \sim$ Uniform$(2, 128)$ for the degrees of freedom parameter incorporates the possibility of heavy tails in the marginal distribution for $\eta_{k,t}$.

5 MCMC sampling algorithm

We develop an efficient Gibbs sampling algorithm for the DFOSR model based on four essential components: (i) the loading curve sampler for $\{f_k\}$ with the identifiability constraint $\boldsymbol{F}'\boldsymbol{F} = \boldsymbol{I}_K$; (ii) the projection-based simplification of model (9) from Lemma 1; (iii) a state space simulation smoother for the dynamic regression parameters in (3) and (4); and (iv) parameter expansions for the variance components in (16). For sparsely-observed functional data, in which the functions are observed only at a small subset of $\{\boldsymbol{\tau}_j\}_{j=1}^M$ at each time t, a sampling-based imputation step is included. An overview of the sampling algorithm is provided here, with details for the full sampling algorithm in the supplement.

Since model (14)–(15) is a dynamic linear model in the state variables $(\boldsymbol{\alpha}_{k,t}', \gamma_{k,t})'$, the dynamic parameters $\{\boldsymbol{\alpha}_{k,t}, \gamma_{k,t}\}_{t=1}^T$ are sampled *jointly* across all $t = 1, \ldots, T$ from their full conditional posterior distribution using an efficient simulation smoothing algorithm (Durbin and Koopman, 2002). These samplers are also valid for FOSR-AR with $\alpha_{j,k,t} = \alpha_{j,k}$. Note that the model (14)–(15) may be aggregated across $k = 1, \ldots, K$ to produce a jointly sampler with respect to k; in our experience, however, doing so increases computation time without improving MCMC efficiency. A single draw of all dynamic regression coefficients and autoregressive regression error terms $\{\boldsymbol{\alpha}_{k,t}, \gamma_{k,t}\}_{k,t}$ has computational complexity $\mathcal{O}(KTp^3)$. For a small to moderate number of predictors $p < 30$, the algorithm is efficient; for sufficiently small K, the sampler is nearly equivalent to the analogous non-functional TVP regression model.

In addition to the loading curve sampler for $\{f_k\}$ in Section 3 and the state space simulation sampler for $\{\boldsymbol{\alpha}_{k,t}, \gamma_{k,t}\}_{k,t}$ via (14)–(15), the Gibbs sampler proceeds by iteratively sampling the intercepts $\{\mu_k\}$, the autoregressive coefficients $\{\phi_k\}$, and the variance components $\sigma^2_{\epsilon_t}$, $\sigma^2_{\eta_{k,t}}$, and $\sigma^2_{\omega_{j,k,t}}$—as well as any relevant hyperparameters—from their full conditional distributions (see the supplement). Stationarity on $\gamma_{k,t}$ is imposed by the prior on the autoregressive coefficients: we set $\{(\phi_k + 1)/2\} \sim$ Beta$(5, 2)$, which ensures that $\mathbb{P}(|\phi_k| < 1) = 1$. This prior on ϕ_k is weakly informative and encourages positive autocorrelation: $\mathbb{P}(\phi_k > 0) \approx 0.9$ and $\mathbb{P}(\phi_k > 0.5) \approx 0.45$. For the yield curve data in Section 7, we estimate $\mathbb{P}(\phi_k > 0.5|\boldsymbol{y}) \approx 1$ for each k, so there is posterior learning which does not contradict the prior. Posterior inference is available for these quantities as well as the TVP regression functions $\tilde{\alpha}_{j,t}(\boldsymbol{\tau}) := \sum_{k=1}^K f_k(\boldsymbol{\tau})\alpha_{j,k,t}$ and the fitted curves $\hat{Y}_t(\boldsymbol{\tau}) := \sum_{k=1}^K f_k(\boldsymbol{\tau})\beta_{k,t}$.

To generate samples from the h-step forecasting distribution of $\boldsymbol{y}_{T+h}$, we leverage the state space construction in (14)–(15). Specifically, we draw from the h-step forecasting distribution of the state variables in (15), $[\boldsymbol{\alpha}_{k,T+h}, \gamma_{k,T+h}|\boldsymbol{y}, -]$, and subsequently sample $[\boldsymbol{y}_{T+h}|\boldsymbol{y}, -] \sim N(\sum_{k=1}^K \boldsymbol{f}_k\beta_{k,T+h}, \sigma^2_{\epsilon_{T+h}}\boldsymbol{I}_M)$, where $\beta_{k,T+h} = \mu_k +$

$\sum_{j=1}^{p} x_{j,T+h}\alpha_{j,k,T+h} + \gamma_{k,T+h}$. Note that $\sigma^2_{\epsilon_{T+h}}$ must be sampled from the forecasting distribution of the stochastic volatility model (see Section 7). If only point estimates are needed, we use $\hat{\boldsymbol{y}}_{T+h} = \sum_{k=1}^{K} \boldsymbol{f}_k \hat{\beta}_{k,T+h}$, where $\hat{\beta}_{k,T+h} = \hat{\mu}_k + \sum_{j=1}^{p} x_{j,T+h}\hat{\alpha}_{j,k,T+h} + \hat{\gamma}_{k,T+h}$ and $\hat{\alpha}_{j,k,T+h}$ and $\hat{\gamma}_{k,T+h}$ are the conditional expectations of $[\boldsymbol{\alpha}_{k,T+h}, \gamma_{k,T+h} | \boldsymbol{y}, -]$. The estimator $\hat{\boldsymbol{y}}_{T+h}$ reduces Monte Carlo error, and may be more accurate in some cases. These forecasting computations remain unchanged for various modifications of the DFOSR model, for example by using a known basis in place of $\{f_k\}$ or an alternative prior distribution for the innovations in (16).

6 Simulation study

The DFOSR model features several important components: (i) the model for $\{f_k\}$ from Section 3, (ii) the dynamic regression coefficients in (4), and (iii) the nested horseshoe and ordered shrinkage priors for the innovations (16) from Section 4. In concert, these features comprise the proposed DFOSR model (denoted DFOSR-HS in the subsequent figures). Naturally, these assumptions may be relaxed and modified to produce simpler models within the same modeling framework. It is therefore important to determine which of the DFOSR model features are necessary to produce reliable forecasts, estimates, and inference.

The following simulation study assesses the relative importance of these modeling choices. Specifically, we are interested in evaluating forecasts of the functional time series $\boldsymbol{y}_t$ and estimates of the dynamic regression coefficient functions $\tilde{\alpha}_{j,t}(\boldsymbol{\tau})$, as well as the empirical coverage and precision of the accompanying posterior forecasting and credible intervals. We consider simulation designs with *dynamic* and *non-dynamic* regression coefficients, and compare the proposed methods to state-of-the-art and benchmark alternatives for functional regression.

6.1 Simulation design

The synthetic data-generating process is based on model (1)–(4), but with some notable differences. There are two sources of sparsity in the regression: (i) some predictors are not associated with the functional response $Y_t(\boldsymbol{\tau})$ and (ii) some predictors are associated with $Y_t(\boldsymbol{\tau})$ exclusively via a small number of factors. Let $p_0 = 10$ regression coefficients be exactly zero (for all times t), and let $p_1 = 5$ be nonzero, resulting in $p = p_0 + p_1 = 15$ regression coefficients plus an intercept, which is fixed at $\mu^*_k := 1/k$. For each nonzero predictor $j = 1, \ldots, p_1 = 5$, uniformly sample p^*_j factors to be nonzero, where $p^*_j \sim$ Poisson(1) truncated to $[1, K^*]$. For dynamic regression coefficients, the nonzero factors k for predictor j are drawn from a Gaussian random walk with randomly selected jumps: $\alpha^*_{j,k,t} := Z_{k,0} + \sum_{s \leq t} Z_{k,s} I_{k,s}$ where $Z_{k,t} \sim N(0, 0.75^{k-1})$ and $I_{k,t} \sim$ Bernoulli(0.01), which results in time-varying yet locally constant regression coefficients $\alpha^*_{j,k,t}$. For non-dynamic regression coefficients, let $\alpha^*_{j,k,t} := \alpha^*_{j,k} \sim N(0, 0.75^{k-1})$. The regression errors are $\gamma^*_{k,t} = 0.8\gamma^*_{k,t-1} + \eta^*_{k,t}$ with $\eta^*_{k,t} \sim N(0, 0.75^{k-1}[1 - 0.8^2])$, which are autocorrelated yet stationary with unconditional variance 0.75^{k-1}. Finally, each time-ordered predictor

$\{x_{j,t}\}_{t=1}^{T}$ is simulated from a Gaussian AR(1) model with autoregressive coefficient 0.8, unconditional mean zero, and unconditional variance one for $j = 1, \ldots, p$.

For M equally-spaced points $\boldsymbol{\tau} \in [0, 1]$, the true loading curves are $f_1^*(\boldsymbol{\tau}) := 1/\sqrt{M}$ and for $k = 2, \ldots, K^* = 4$, f_k^* is an orthogonal polynomial of degree k. Given true factors $\beta_{k,t}^* := \mu_k^* + \sum_{j=1}^{p} x_{j,t}\alpha_{j,k,t}^* + \gamma_{k,t}^*$ and loading curves $f_k^*(\boldsymbol{\tau})$, the true curves are $Y_t^*(\boldsymbol{\tau}) := \sum_{k=1}^{K^*} f_k^*(\boldsymbol{\tau})\beta_{k,t}^*$ and the functional data are simulated from $y_t(\boldsymbol{\tau}) = Y_t^*(\boldsymbol{\tau}) + \sigma^*\epsilon_t^*(\boldsymbol{\tau})$, where $\epsilon_t^*(\boldsymbol{\tau}) \sim N(0, 1)$. After selecting a *root-signal-to-noise ratio* (RSNR), the observation error standard deviation is $\sigma^* := \sqrt{(TM-1)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{M}(Y_t^*(\boldsymbol{\tau}_j) - \bar{Y}^*)^2} \Big/ \text{RSNR}$ where $\bar{Y}^*$ is the sample mean of $\{Y_t^*(\boldsymbol{\tau}_j)\}_{j,t}$. We select RNSR = 5, which produces moderately noisy functional data, and use $T = 200$ time points with $M = 25$ functional observations points. Results for $M = 100$ are included in the supplement with similar conclusions.

6.2 Methods for comparison

We implement several competing methods within the dynamic Bayesian framework of (1)–(5), which in each case utilizes the same general MCMC approach of Section 5 with small modifications as needed. First, let DFOSR-NIG denote the DFOSR model with normal-inverse-gamma priors on the innovations, similar to Kowal et al. (2017a), instead of the nested horseshoe priors from Section 4. Next, replace the loading curves $\{f_k\}$ with functional principal components (FPCs) estimated *a priori* using Xiao et al. (2013), where K is selected to explain 90% of the variability in $\{\boldsymbol{y}_t\}_t$. For the FPC-based modifications, we consider both the nested horseshoe priors from Section 4 (DFPCA-HS) and the aforementioned normal-inverse-gamma prior (DFPCA-NIG). The comparative performance of these methods—DFOSR, DFOSR-NIG, DFPCA-HS, and DFPCA-NIG—illustrates the gains of including (i) shrinkage priors and (ii) a Bayesian model for $\{f_k\}$. Lastly, omit the dynamic regression coefficients, $\alpha_{j,k,t} = \alpha_{j,k}$, reducing to a function-on-scalars regression with autoregressive errors (FOSR-AR). The FOSR-AR model is identical to the full DFOSR with the exception of (4), and therefore isolates the utility of the time-variation of $\{\alpha_{j,k,t}\}$.

For benchmark comparisons, we also include non-Bayesian and non-dynamic regression models. First, we include a functional regression model which estimates the regression coefficient functions using generalized least squares (FOSR-LS; Reiss et al., 2010). FOSR-LS is similar to fitting separate linear regression models of $y_{j,t}$ on $\boldsymbol{x}_t$ for each $j = 1, \ldots, M$, but adds smoothness in $\boldsymbol{\tau}_j$ for the regression coefficients and incorporates covariance among the errors $\epsilon_{j,t}$ with respect to $j = 1, \ldots, M$. Next, we include a FOSR model which attaches a group lasso penalty to each regression function to provide variable selection (FOSR-Lasso). Both FOSR-LS and FOSR-Lasso are implemented using the `refund` package in `R` (Goldsmith et al., 2016). Note that these methods do not account for time-varying regression coefficients or autocorrelated errors (with respect to time).

6.3 Evaluation criteria

Forecasting ability is evaluated using root mean squared forecast errors, RMSFE $= \sqrt{\sum_{\ell=1}^{M}\{Y_{T+1}(\boldsymbol{\tau}_\ell) - \hat{Y}_{T+1}(\boldsymbol{\tau}_\ell)\}^2/M}$ for forecast $\hat{Y}_{T+1}$, and mean forecast interval widths, MFIW $= \sum_{\ell=1}^{M}\{\hat{y}^U_{T+1}(\boldsymbol{\tau}_\ell) - \hat{y}^L_{T+1}(\boldsymbol{\tau}_\ell)\}/M$ for forecast interval $(\hat{y}^L_{T+1}(\boldsymbol{\tau}_\ell), \hat{y}^U_{T+1}(\boldsymbol{\tau}_\ell))$, $\ell = 1, \ldots, M$. While RMSFE measures point forecast accuracy, MFIW assesses precision of the forecasting intervals. Narrower forecast intervals—and therefore smaller MFIW—are preferable among intervals that attain the nominal empirical coverage. The empirical coverage is $\sum_{\ell=1}^{M}\mathbb{I}\{\hat{y}^L_{T+1}(\boldsymbol{\tau}_\ell) \leq y_{T+1}(\boldsymbol{\tau}_\ell) \leq \hat{y}^U_{T+1}(\boldsymbol{\tau}_\ell)\}/M$. We consider 90% intervals for all simulations.

For the regression coefficient functions $\tilde{\alpha}_{j,t}(\boldsymbol{\tau})$, we similarly compute root mean squared errors, RMSE $= \sqrt{\sum_{j,t,\ell}(\tilde{\alpha}_{j,t}(\boldsymbol{\tau}_\ell) - \tilde{\alpha}^*_{j,t}(\boldsymbol{\tau}_\ell))^2/(pTM)}$, where $\tilde{\alpha}_{j,t}(\boldsymbol{\tau}_\ell)$ is the estimated regression coefficient for predictor j at time t and observation point $\boldsymbol{\tau}_\ell$ and $\tilde{\alpha}^*_{j,t}(\boldsymbol{\tau}_\ell) = \sum_{k=1}^{K^*} f^*_k(\boldsymbol{\tau}_\ell)\alpha^*_{j,k,t}$ is the true regression coefficient, and mean credible interval widths, MCIW $= \sum_{j,t,\ell}\{\tilde{\alpha}^U_{j,t}(\boldsymbol{\tau}_\ell) - \tilde{\alpha}^L_{j,t}(\boldsymbol{\tau}_\ell)\}/(pTM)$, for pointwise credible intervals $(\tilde{\alpha}^L_{j,t}(\boldsymbol{\tau}_\ell), \tilde{\alpha}^U_{j,t}(\boldsymbol{\tau}_\ell))$. Nominal empirical coverage for the intervals is also reported. We present RMSE, MCIW, and empirical coverage separately for the true nonzero coefficients (signals), $\{j : \tilde{\alpha}^*_{j,t}(\boldsymbol{\tau}_\ell) \neq 0 \text{ for some } t, \ell\}$, and the true zero coefficients (noise), $\{j : \tilde{\alpha}^*_{j,t}(\boldsymbol{\tau}_\ell) = 0 \text{ for all } t, \ell\}$.

6.4 Simulation results

The results from the dynamic regression setting are in Figures 1 and 2 for the forecasting and estimation comparisons, respectively. Similar figures for the non-dynamic regression setting, in which $\tilde{\alpha}^*_{j,t}(\boldsymbol{\tau}_\ell) = \tilde{\alpha}^*_j(\boldsymbol{\tau}_\ell)$ is time-invariant for all j, ℓ, are in Figures 3 and 4. The RMSFE and MFIW are reported in terms of the percent reduction relative to DFOSR-NIG, which provides standardization among the forecasting metrics. Specifically, for each method j we compute $100(\text{RMSFE}_j - \text{RMSFE}_0)/\text{RMSFE}_0$ and $100(\text{MFIW}_j - \text{MFIW}_0)/\text{MFIW}_0$ for each of 100 simulated datasets, where $j = 0$ denotes the reference model. The DFOSR-NIG is an appropriate reference model: it includes the model for unknown $\{f_k\}$ from Section 3 but does not include the nested horseshoe prior of Section 4. Methods with values above zero demonstrate improvements in forecasting performance relative to DFOSR-NIG.

Figure 1 shows the relative forecasting results for the dynamic regression simulations. The inclusion of the nested horseshoe priors (DFOSR-HS) provides more accurate point forecasts with narrower forecasting intervals compared to DFOSR-NIG, which demonstrates the utility of these shrinkage priors. At the same time, the performance of DFPCA-HS, which uses the same shrinkage priors but fixes f_k at the FPC estimates, is underwhelming, and is often outperformed by DFOSR-NIG. In concert, these results demonstrate that inclusion of *both* the model for $\{f_k\}$ and the nested horseshoe prior produces more accurate point forecasts and more precise forecast intervals. However, the time-variation in (4) is equally important: FOSR-AR omits these dynamics, and clearly produces less accurate point forecasts and forecast intervals with insufficient coverage.

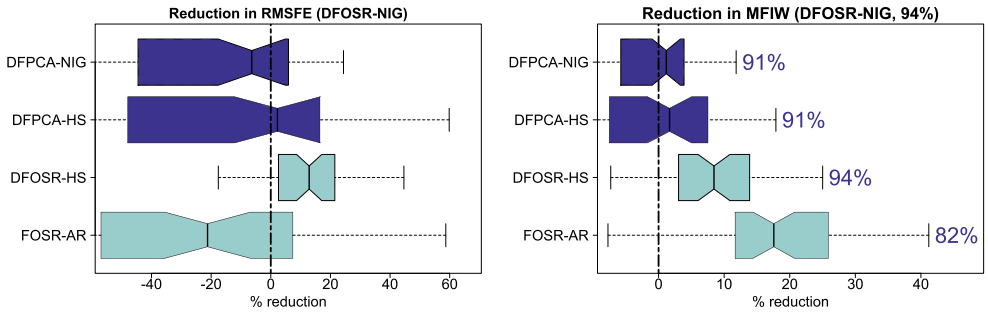

Figure 1: Forecasting in the dynamic regression model. Performance is evaluated relative to DFOSR-NIG. The proposed models are in light blue.

Figure 2 presents the results for the regression coefficient functions, again for the dynamic regression simulations. Among the true nonzero coefficients (signals), it is most notable that the proposed DFOSR-HS offers significantly more accurate estimation and narrower credible intervals which maintain the correct nominal coverage. By comparison, the methods that do not include *both* the time-varying parameters and the model for $\{f_k\}$ fail to attain the correct nominal coverage. As anticipated, the shrinkage priors are comparatively more important among true zeros, where the non-dynamic FOSR-AR is clearly the best while DFOSR-HS, DFPCA-HS, and FOSR-Lasso perform similarly.

The forecasting and estimation results for the non-dynamic regression simulations are in Figures 3 and 4. As expected, FOSR-AR performs exceptionally well, since it corresponds to the true data-generating process. Again, the proposed DFOSR-HS significantly outperforms the dynamic competitors in both forecasting and inference. Interestingly, even though FOSR-AR and the non-Bayesian methods FOSR-LS and FOSR-Lasso all (correctly) assume non-dynamic coefficients, the FOSR-AR provides substantially more accurate estimates for both signal and noise coefficients.

To characterize the sensitivity of the proposed DFOSR model to our choice of priors, we considered a variation of DFOSR-HS in which the autoregressive coefficients each received a flat prior restricted to stationarity, $\phi_k \sim \text{Uniform}(-1, 1)$, and the MGP hyperparameters were fixed at $a_{\mu_1} = a_{\eta_1} = 2$, $a_{\mu_2} = a_{\eta_2} = 3$, and $\nu_\eta = 3$ instead of using the hyperpriors from Section 4. The results are quite similar to those in Figures 1-4, with some attenuation of the gains offered by DFOSR-HS. Therefore, we recommend the weakly informative priors for the autoregressive coefficients ϕ_k and the MGP hyperpriors from Section 4.

7 Forecasting yield curves using macroeconomic variables

The yield curve describes the time-varying term structure of interest rates: at each time t, the yield curve $Y_t(\boldsymbol{\tau})$ characterizes how interest rates vary over the length of

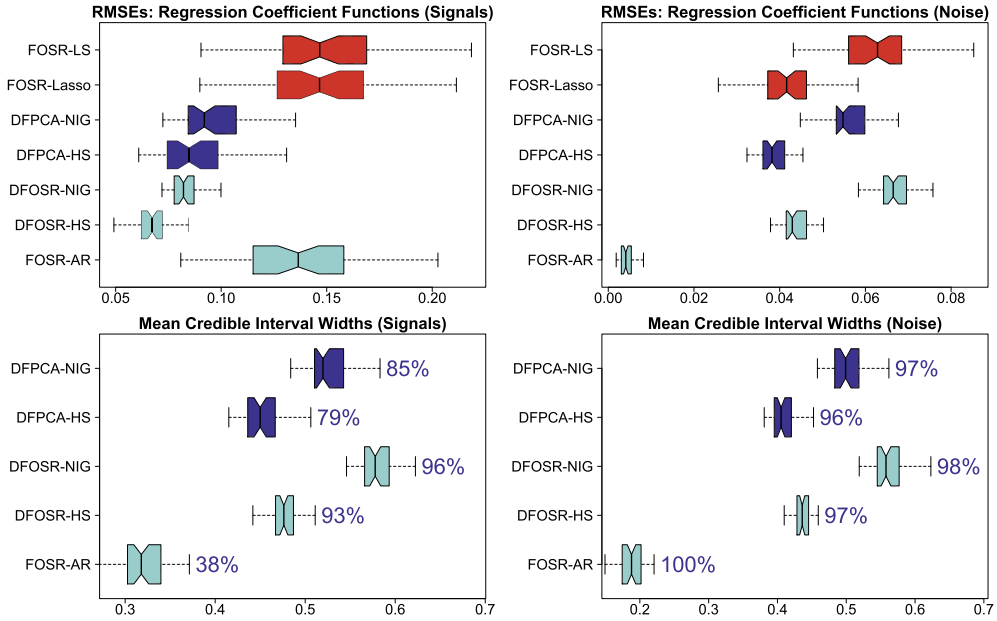

Figure 2: Root mean squared errors (top) and mean credible intervals widths with empirical coverage (bottom) for $\tilde{\alpha}_{j,t}(\boldsymbol{\tau})$ in the dynamic setting separated into true nonzeros (left) and true zeros (right). The proposed models are in light blue. DFOSR-HS offers clear gains for estimation and inference on the signals, while competing methods fail to attain the correct nominal coverage.

the borrowing period, or maturity, $\boldsymbol{\tau}$. Yield curves are an essential component in many economic and financial applications: they provide valuable information about economic and monetary conditions, inflation expectations, and business cycles, and are used to price fixed-income securities and construct forward curves (Bolder et al., 2004). Connections between yield curves and macroeconomic variables are of particular interest, both for understanding the interplay of various components of the economy (Diebold et al., 2006; Aguiar-Conraria et al., 2012) and for producing improved forecasts of key economic and financial variables (Mönch, 2008; Koop, 2013).

We are interested in leveraging macroeconomic variables to produce more accurate yield curve forecasts with reliable uncertainty quantification. There is accumulating evidence to support the inclusion of macroeconomic variables in a yield curve forecasting model (Coroneo et al., 2016; Altavilla et al., 2017). Among the most successful forecasting approaches, a noteworthy feature is the use of time-varying parameters for modeling the associations between macroeconomic variables and the yield curve (Bianchi et al., 2009; Mumtaz and Surico, 2009; Byrne et al., 2017). However, all of the models referenced above—with or without time-varying parameters—rely on the Nelson-Siegel basis expansion (Nelson and Siegel, 1987). The Nelson-Siegel approach uses a similar construction as (2), but with $K = 3$ parametric functions:

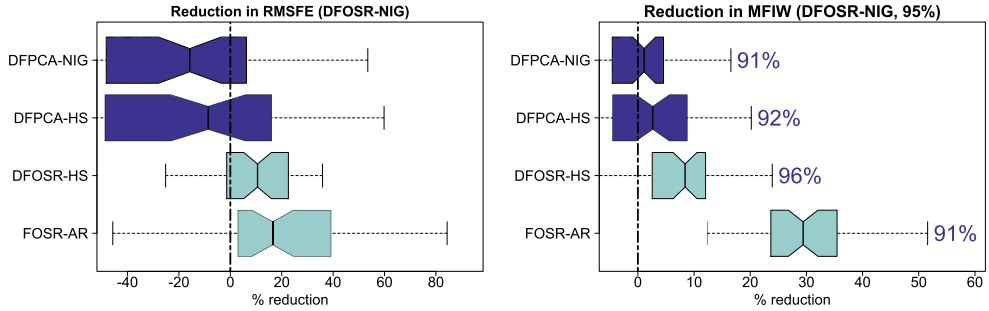

Figure 3: Forecasting in the non-dynamic regression model. Performance is evaluated relative to DFOSR-NIG. The proposed models are in light blue. The FOSR-AR performs the best, while the proposed DFOSR-HS performs the best among the dynamic models.

the *level* $f_1(\boldsymbol{\tau};\lambda) = 1$, the *slope* $f_2(\boldsymbol{\tau};\lambda) = \{1-\exp(-\tau\lambda)\}/(\tau\lambda)$, and the *curvature* $f_3(\boldsymbol{\tau};\lambda) = \{1-\exp(-\tau\lambda)\}/(\tau\lambda) - \exp(-\tau\lambda)$. Although some approaches estimate the parameter $\lambda > 0$ (Laurini and Hotta, 2010; Cruz-Marcelo et al., 2011), it is more common to fix λ *a priori*, such as $\lambda = 0.0609$ (Diebold and Li, 2006; Bianchi et al., 2009; Mumtaz and Surico, 2009; Coroneo et al., 2016; Byrne et al., 2017; Altavilla et al., 2017).

The Nelson-Siegel basis provides a convenient simplification of the functional dependence, but faces several limitations. First, it is clearly inadequate for other functional datasets, many of which have no known parametric structure. Second, the common choice of $\lambda = 0.0609$ may be inappropriate for other term structure data or non-US yield curves, yet relaxation of this assumption is challenging due to the resulting nonlinearities in the model. A more subtle difficulty is that the parametric biases of the Nelson-Siegel basis may be undetected or understated among methods that are evaluated using the popular yield curve dataset from Gürkaynak et al. (2007). Gürkaynak et al. (2007) employ a Svensson model for smoothing (Svensson, 1994), which is an augmented Nelson-Siegel model with one additional nonlinear term. Consequently, the Nelson-Siegel-based models that use this dataset (Aguiar-Conraria et al., 2012; Altavilla et al., 2017; Byrne et al., 2017) are likely to show inflated performance relative to other yield curve datasets that do not pre-smooth using Nelson-Siegel or Svensson models.

The DFOSR model offers an appealing alternative: the basis functions $\{f_k\}$ are modeled nonparametrically, while time-varying regression coefficients are included for the macroeconomic variables $\boldsymbol{x}_t$ via (4). Existing nonparametric regression approaches for yield curve modeling include Hays et al. (2012) and Jungbacker et al. (2013), which similarly model $\{f_k\}$ as unknown and incorporate a state space framework for $\{\beta_{k,t}\}$, yet do not provide the uncertainty quantification, TVP regression, and shrinkage capabilities of the DFOSR model. The shrinkage priors of the DFOSR model are particularly important, since it is *a priori* unclear whether any individual macroeconomic variable is important for forecasting and which, if any, corresponding regression coefficients are time-varying.

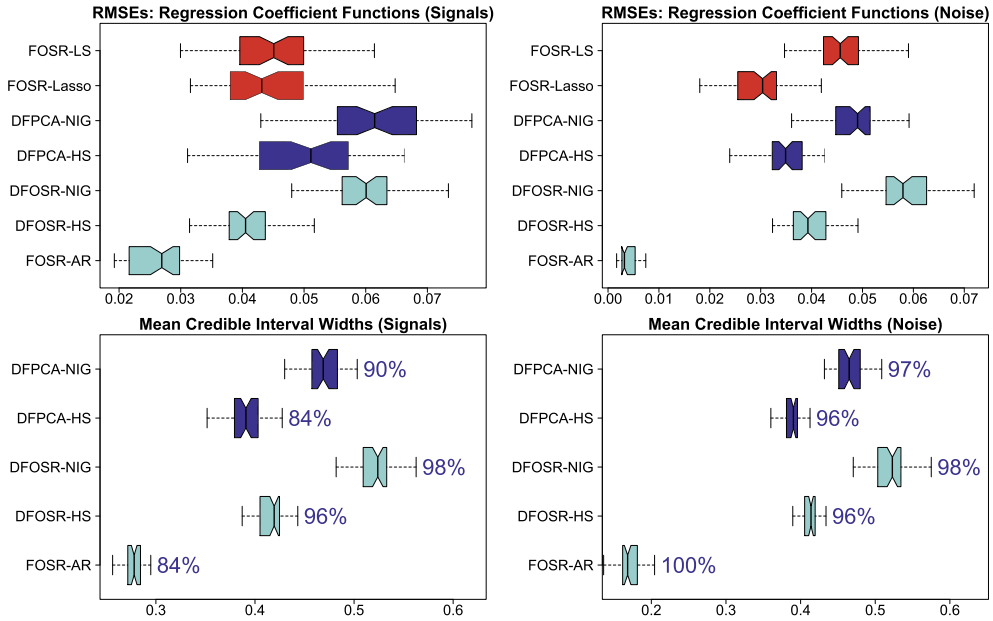

Figure 4: Root mean squared errors (top) and mean credible intervals widths with empirical coverage (bottom) for $\tilde{\alpha}_{j,t}(\boldsymbol{\tau})$ in the non-dynamic setting separated into true nonzeros (left) and true zeros (right). The proposed models are in light blue. FOSR-AR demonstrates the most accurate estimation, while the proposed DFOSR-HS provides the best credible intervals for nonzero regression coefficients (signals).

More generally, it is of interest to assess (i) whether including macroeconomic predictors can improve forecasting performance and (ii) whether alternative models for $\{f_k\}$—such as FPCs or the Nelson-Siegel basis—are competitive. The benefits of including macroeconomic predictors may depend on whether or not the model allows for time-varying regression coefficients. Therefore, we consider two special cases of the DFOSR model: FOSR-AR, which omits the time-variation in (4), and the functional dynamic linear model (FDLM) of Kowal et al. (2017a) with predictors omitted entirely. In each case, we use $K = 4$ factors, although results are similar for $K = 6$. We include two variations of each of the FOSR-AR and the FDLM by modifying the model for $\{f_k\}$: first, using the Nelson-Siegel basis with (NS-X) and without (NS) predictors, and second, using the FPC basis with (FPC-X) and without (FPC) predictors. All competing models may be written in the form of (1)–(5), with forecast estimates and intervals computed using the same MCMC algorithm of Section 5.

For each model, we include a stochastic volatility model for $\sigma^2_{\epsilon_t}$ to incorporate volatility clustering:

$$h_t = \mu_h + \phi_h(h_{t-1} + \mu_h) + \nu_t, \quad \nu_t \overset{iid}{\sim} N(0, \sigma^2_\nu), \tag{18}$$

where $h_t = \log \sigma^2_{\epsilon_t}$ is the log-volatility. Sampling the log-volatility proceeds using a Gaussian mixture approximation similar to Kim et al. (1998) and Omori et al. (2007).

7.1 Forecasting design

We consider 40 years of monthly data from 1970-2009 using the unsmoothed Fama and Bliss (1987) US government bond yields provided by Van Dijk et al. (2014). These yield curve data are available for maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months ($M = 17$), and importantly are *not* pre-smoothed using the Nelson-Siegel or Svensson model. To accompany the yield curve data, we include the macroeconomic variables in Table 2, which are provided by the Federal Reserve and retrieved from the FRED data download website. We consider two sets of predictors: a small set of $p = 3$ variables and medium set of $p = 12$ predictors, which are transformed according to Table 2 and lagged by one month for forecasting. These variable sets are similar to those used in Koop (2013). While it is possible to incorporate concurrent (non-lagged) predictors in the DFOSR model, this requires augmentation of the dynamics in (15) and produces a more complex state space model with accompanying computational challenges.

Variable	Code	Description
FEDFUNDS‡	0	Effective Federal Funds Rate
CPIAUCSL‡	1	Consumer Price Index
GDP‡	1	Gross Domestic Product
INDPRO†	1	Industrial Production Index
M2SL†	1	M2 Money Stock
TCU†	0	Capacity Utilization: Total Industry
UNRATE†	0	Civilian Unemployment Rate
M1SL†	1	M1 Money Stock
PAYEMS†	1	All Employees: Total Nonfarm
HOUST†	1	Housing Starts: Total New Privately Owned Housing Units Started
AHETPI†	1	Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private
TWEXBMTH†	1	Trade weighted U.S. Dollar Index: Broad

Table 2: Predictor variables included in the small (‡) and medium (†) sized models. The code denotes the transformation: 0 = no transformation, 1 = year-over-year growth rate. The variable names correspond to the FRED series.

7.2 Forecasting results

Point and 95% interval forecasts were computed for each month from 1990-2009, totaling 227 months for evaluation. Methods are compared using RMSFE and MFIW defined in Section 6, which are reported relative to the FDLM. We select the FDLM as the baseline because it incorporates some, but not all, of the proposed features in the DFOSR: it includes the model for $\{f_k\}$ from Section 3, but omits predictors and consequently does not include the time-variation in (4) or the shrinkage priors of Section 4.

To illustrate the challenge of the forecasting exercise, Figure 5 provides examples of the DFOSR yield curve forecasts in 2006 and 2008. There is considerable month-to-month time-variation in both the yield curve level and shape over the previous 24

months. Nonetheless, the DFOSR point and interval forecasts are accurate, and the interval forecasts are sufficiently precise that they include the realized yield curves while excluding many of the previous yield curve observations.

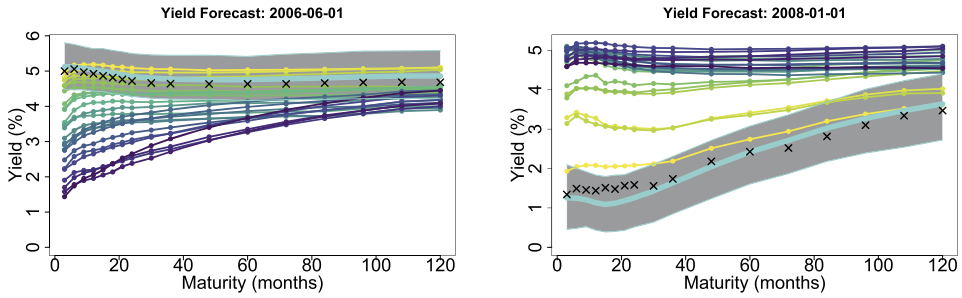

Figure 5: One-month-ahead yield forecast (cyan line) with 95% forecast intervals (gray bands) on June 1, 2006 (left) and January 1, 2008 (right) for $p = 3$ predictors. The crosses indicate the realized values $\boldsymbol{y}_{T+1}$, while the connected points give the previous 24 months of yield curves (from dark to light, oldest to most recent).

Figure 6 displays the forecasting results for the small set of predictors ($p = 3$) in Table 2. All models achieve the correct nominal empirical coverage, yet each method is overconservative. Comparing methods with and without the macroeconomic variables, we identify a small decrease in forecast interval width, which suggests that predictors provide slightly more precision in the forecasting intervals. Most notable, however, is the performance of the DFOSR, which provides substantial reductions in forecasting interval widths relative to all competitors. In particular, it is clear from comparing the FDLM, the FOSR-AR, and the DFOSR that (i) including the macroeconomic predictors produces narrower forecasting intervals, and (ii) this effect is heightened for time-varying regression coefficients. The DFOSR also provides the most accurate point forecast, albeit with smaller improvements relative to competing methods.

The forecasting results for the medium set of predictors ($p = 12$) are in Figure 7. The larger set of predictors is more challenging for the TVP regressions, especially in the functional data setting. Nonetheless, the DFOSR again performs exceptionally well: the model for $\{f_k\}$ from Section 3 provides substantial gains in forecasting performance relative to the Nelson-Siegel and FPC bases, while the time-variation of the DFOSR model offers consistent improvements relative to the non-dynamic FOSR-AR model.

To further illuminate the differences between dynamic and non-dynamic regression coefficient functions, we plot the posterior expectations of $\tilde{\alpha}_{j,t}(\tau)$ for each predictor j as a function of maturity τ across times t from 1970-2009 in Figure 8. We include the posterior expectations under both the DFOSR and FOSR-AR models for comparison. Many of the regression coefficient functions exhibit substantial time-variation in both shape and magnitude, which cannot be modeled by non-dynamic regression coefficient functions. The improvements in forecasting performance of the DFOSR relative to the FOSR-AR suggest that the time-variation in the regression coefficient functions is an important feature of the model.

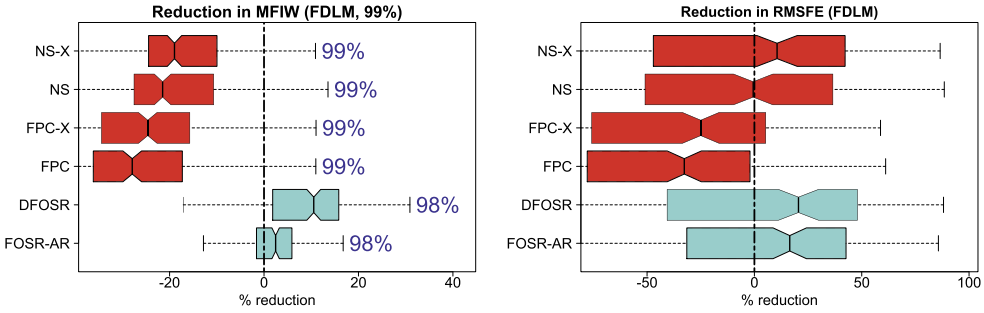

Figure 6: For $p = 3$ predictors, the percent reduction in mean forecast interval widths (left) and root mean squared forecast error (right) relative to the FDLM of Kowal et al. (2017a) excluding predictors. The proposed models are in light blue. The empirical coverage of the 95% forecast intervals is also given.

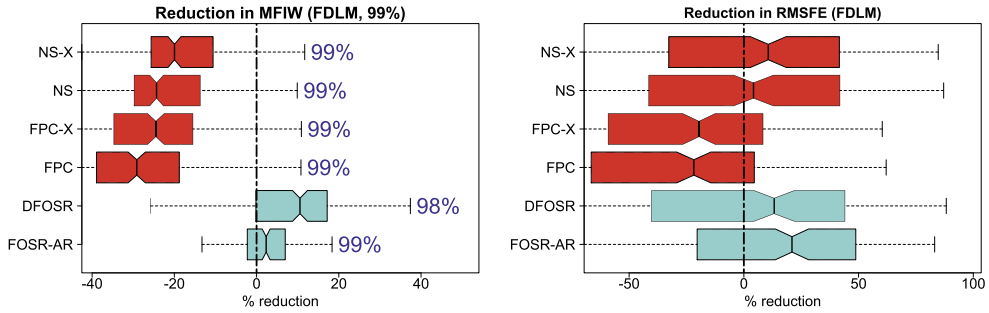

Figure 7: For $p = 12$ predictors, the percent reduction in mean forecast interval widths (left) and root mean squared forecast error (right) relative to the FDLM Kowal et al. (2017a) excluding predictors. The proposed models are in light blue. The empirical coverage of the 95% forecast intervals is also given.

8 Discussion

The proposed *dynamic function-on-scalars regression* (DFOSR) model features several essential components: (i) a nonparametric regression model for the unknown basis functions, (ii) time-variation in the regression parameters, and (iii) shrinkage priors that regularize irrelevant predictors and shrink toward time-invariance. These model features are synthesized within a fully Bayesian framework, with posterior inference available via an efficient Gibbs sampling algorithm. Simulation studies demonstrate the utility of these modeling choices, in particular for constructing precise forecasting intervals and estimating time-varying regression coefficient functions. The proposed model is evaluated empirically for modeling and forecasting yield curves using macroeconomic predictor variables. Compared to models that exclude time-variation in the regression coefficients and models that exclude the macroeconomic predictors altogether, the DFOSR fore-

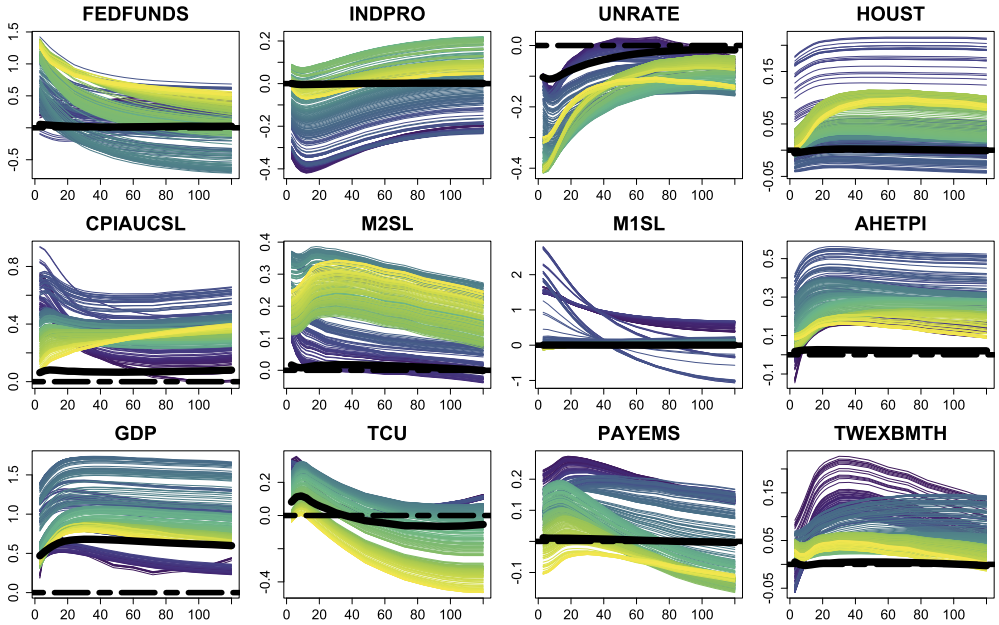

Figure 8: Posterior expectations of $\tilde{\alpha}_{j,t}(\tau)$ for each predictor j as a function of maturity τ (in months) across times t from 1970 (dark) to 2009 (light). The dashed line indicates zero and the solid black line is the posterior mean of $\tilde{\alpha}_j(\tau)$ under the FOSR-AR model with time-invariant parameters. For many of the variables, there is notable time-variation in both magnitude and shape.

casts are more accurate, and in particular provide narrower forecasting intervals that attain the correct nominal coverage.

The DFOSR model is a natural extension of existing functional regression and functional time series models. When $\boldsymbol{x}_t = \mathbf{0}$ for all t, i.e., there are no predictors, the DFOSR model is a reduced-rank functional factor model with autocorrelated factors, which is useful for modeling and forecasting functional time series data (Hays et al., 2012; Aue et al., 2015; Kowal et al., 2017a). When the regression coefficients $\alpha_{j,k,t} = \alpha_{j,k}$ are non-dynamic and the autoregressive coefficients vanish with $\phi_k = 0$, we obtain a Bayesian FOSR model (Montagna et al., 2012; Kowal and Bourgeois, 2019). Similarly, constraining $\alpha_{j,k,t} = \alpha_{j,k}$ but allowing $\phi_k \neq 0$ produces a Bayesian FOSR model with autoregressive errors. Each of these special cases was compared with the full DFOSR model on both simulated and real data in Sections 6 and 7, respectively.

Future work will extend model (2)–(5) for other important dependence structures, such as dynamic *functional* predictors $X_{j,t}(\boldsymbol{u})$ for $\boldsymbol{u} \in \mathcal{U}$, possibly with different domains $\mathcal{U} \neq \mathcal{T}$. The key simplification of the likelihood from Lemma 1 only requires the functional data likelihood (9) and the identifiability constraint $\boldsymbol{F}'\boldsymbol{F} = \boldsymbol{I}_K$. Alternative models for the factors $\{\beta_{k,t}\}$ in (3), including joint state space models for forecasting

Y_t and $\boldsymbol{x}_t$, may leverage the simplifications from Lemma 1 to achieve computational scalability in other modeling frameworks. Lastly, the ability to incorporate constraints for $\{f_k\}$ in Theorem 1, although not applicable for the yield curve models, offers a promising approach for modeling unknown yet *a priori* constrained basis functions in functional and spatial data analysis.

Supplementary Material

Supplementary Material to "Dynamic Regression Models for Time-Ordered Functional Data" (DOI: 10.1214/20-BA1213SUPPa; .pdf).

Supplementary Material to "Dynamic Regression Models for Time-Ordered Functional Data" (DOI: 10.1214/20-BA1213SUPPb; .zip). Supplementary materials include (i) a document with the MCMC algorithm details, the thin plate spline construction, and additional simulation results, (ii) R scripts for the yield curve analysis and the simulation studies, and (iii) an R package which implements the proposed model and all competing models from Sections 6 and 7.

Acknowledgments

We thank David Scott for providing feedback on an early version of the manuscript. We also thank the associated editor and two referees for their time and helpful comments, which have improved the readability of the manuscript.

References

Aguiar-Conraria, L., Martins, M. M., and Soares, M. J. (2012). "The yield curve and the macro-economy across time and frequencies." *Journal of Economic Dynamics and Control*, 36(12): 1950–1970. MR2982962. doi: https://doi.org/10.1016/j.jedc.2012.05.008. 475, 476

Altavilla, C., Giacomini, R., and Ragusa, G. (2017). "Anchoring the yield curve using survey expectations." *Journal of Applied Econometrics*, 32(6): 1055–1068. MR3714393. doi: https://doi.org/10.1002/jae.2588. 475, 476

Aue, A., Norinho, D. D., and Hörmann, S. (2015). "On the prediction of stationary functional time series." *Journal of the American Statistical Association*, 110(509): 378–392. MR3338510. doi: https://doi.org/10.1080/01621459.2014.909317. 459, 468, 481

Barber, R. F., Reimherr, M., and Schill, T. (2017). "The function-on-scalar LASSO with applications to longitudinal GWAS." *Electronic Journal of Statistics*, 11(1): 1351–1389. MR3635916. doi: https://doi.org/10.1214/17-EJS1260. 460

Belmonte, M. A., Koop, G., and Korobilis, D. (2014). "Hierarchical shrinkage in time-varying parameter models." *Journal of Forecasting*, 33(1): 80–94. MR3148281. doi: https://doi.org/10.1002/for.2276. 460, 468

Besse, P. C., Cardot, H., and Stephenson, D. B. (2000). "Autoregressive forecasting of some functional climatic variations." *Scandinavian Journal of Statistics*, 673–687. 459

Bhattacharya, A. and Dunson, D. B. (2011). "Sparse Bayesian infinite factor models." *Biometrika*, 291–306. MR2806429. doi: https://doi.org/10.1093/biomet/asr013. 469

Bianchi, F., Mumtaz, H., and Surico, P. (2009). "The great moderation of the term structure of UK interest rates." *Journal of Monetary Economics*, 56(6): 856–871. 475, 476

Bolder, D., Johnson, G., and Metzler, A. (2004). *An empirical analysis of the Canadian term structure of zero-coupon interest rates*. Bank of Canada. 475

Byrne, J. P., Cao, S., and Korobilis, D. (2017). "Forecasting the term structure of government bond yields in unstable environments." *Journal of Empirical Finance*, 44: 209–225. 475, 476

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). "The horseshoe estimator for sparse signals." *Biometrika*, 465–480. MR2650751. doi: https://doi.org/10.1093/biomet/asq017. 469

Chen, K. and Müller, H.-G. (2012). "Modeling repeated functional observations." *Journal of the American Statistical Association*, 107(500): 1599–1609. MR3036419. doi: https://doi.org/10.1080/01621459.2012.734196. 463

Chen, Y., Goldsmith, J., and Ogden, R. T. (2016). "Variable selection in function-on-scalar regression." *Stat*, 5(1): 88–101. MR3478799. doi: https://doi.org/10.1002/sta4.106. 460

Coroneo, L., Giannone, D., and Modugno, M. (2016). "Unspanned macroeconomic factors in the yield curve." *Journal of Business & Economic Statistics*, 34(3): 472–485. MR3523788. doi: https://doi.org/10.1080/07350015.2015.1052456. 475, 476

Cruz-Marcelo, A., Ensor, K. B., and Rosner, G. L. (2011). "Estimating the term structure with a semiparametric Bayesian hierarchical model: an application to corporate bonds." *Journal of the American Statistical Association*, 106(494). MR2866969. doi: https://doi.org/10.1198/jasa.2011.ap09764. 476

Damon, J. and Guillas, S. (2002). "The inclusion of exogenous variables in functional autoregressive ozone forecasting." *Environmetrics*, 13: 759–774. 459, 468

Dangl, T. and Halling, M. (2012). "Predictive regressions with time-varying coefficients." *Journal of Financial Economics*, 106(1): 157–181. 460

Datta, J. and Ghosh, J. K. (2013). "Asymptotic properties of Bayes risk for the horseshoe prior." *Bayesian Analysis*, 8(1): 111–132. MR3036256. doi: https://doi.org/10.1214/13-BA805. 469

Diebold, F. X. and Li, C. (2006). "Forecasting the term structure of government bond yields." *Journal of Econometrics*, 130(2): 337–364. MR2211798. doi: https://doi.org/10.1016/j.jeconom.2005.03.005. 476

Diebold, F. X., Rudebusch, G. D., and Aruoba, B. S. (2006). "The macroeconomy and the yield curve: a dynamic latent factor approach." *Journal of Econometrics*, 131(1): 309–338. MR2276003. doi: https://doi.org/10.1016/j.jeconom.2005.01.011. 475

Durante, D. (2017). "A note on the multiplicative gamma process." *Statistics & Probability Letters*, 122: 198–204. MR3584158. doi: https://doi.org/10.1016/j.spl.2016.11.014. 469

Durbin, J. and Koopman, S. J. (2002). "A simple and efficient simulation smoother for state space time series analysis." *Biometrika*, 89(3): 603–616. MR1929166. doi: https://doi.org/10.1093/biomet/89.3.603. 470

Fama, E. F. and Bliss, R. R. (1987). "The information in long-maturity forward rates." *The American Economic Review*, 680–692. 478

Fan, Z. and Reimherr, M. (2017). "High-dimensional adaptive function-on-scalar regression." *Econometrics and Statistics*, 1: 167–183. MR3669995. doi: https://doi.org/10.1016/j.ecosta.2016.08.001. 460

Goldsmith, J., Greven, S., and Crainiceanu, C. (2013). "Corrected confidence bands for functional data using principal components." *Biometrics*, 69(1): 41–51. MR3058050. doi: https://doi.org/10.1111/j.1541-0420.2012.01808.x. 465

Goldsmith, J. and Kitago, T. (2016). "Assessing systematic effects of stroke on motor control by using hierarchical function-on-scalar regression." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(2): 215–236. MR3456686. doi: https://doi.org/10.1111/rssc.12115. 465

Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C., and Reiss, P. T. (2016). *refund: Regression with Functional Data*. *R* package version 0.1-16. URL https://CRAN.R-project.org/package=refund 472

Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2011). "Longitudinal functional principal component analysis." In *Recent Advances in Functional Data Analysis and Related Topics*, 149–154. Springer. MR2815575. doi: https://doi.org/10.1007/978-3-7908-2736-1_23. 463

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). "The US Treasury yield curve: 1961 to the present." *Journal of Monetary Economics*, 54(8): 2291–2304. 476

Hays, S., Shen, H., and Huang, J. Z. (2012). "Functional dynamic factor models with application to yield curve forecasting." *The Annals of Applied Statistics*, 6(3): 870–894. MR3012513. doi: https://doi.org/10.1214/12-AOAS551. 476, 481

Hörmann, S., Kidziński, Ł., and Hallin, M. (2015). "Dynamic functional principal components." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2): 319–348. MR3310529. doi: https://doi.org/10.1111/rssb.12076. 465

Hyndman, R. J. and Ullah, M. S. (2007). "Robust forecasting of mortality and fertility rates: a functional data approach." *Computational Statistics & Data Analysis*,

51(10): 4942–4956. MR2364551. doi: https://doi.org/10.1016/j.csda.2006.07.028. 459

Jauch, M., Hoff, P. D., and Dunson, D. B. (2019). "Monte Carlo simulation on the Stiefel manifold via polar expansion." *arXiv preprint arXiv:1906.07684*. 467

Jungbacker, B., Koopman, S. J., and van der Wel, M. (2013). "Smooth dynamic factor analysis with application to the US term structure of interest rates." *Journal of Applied Econometrics*. MR3233733. doi: https://doi.org/10.1002/jae.2319. 476

Kim, S., Shephard, N., and Chib, S. (1998). "Stochastic volatility: likelihood inference and comparison with ARCH models." *The Review of Economic Studies*, 65(3): 361–393. 477

Koop, G. M. (2013). "Forecasting with medium and large Bayesian VARs." *Journal of Applied Econometrics*, 28(2): 177–203. MR3045863. doi: https://doi.org/10.1002/jae.1270. 475, 478

Korobilis, D. (2013). "Hierarchical shrinkage priors for dynamic regressions with many predictors." *International Journal of Forecasting*, 29(1): 43–59. MR3137689. doi: https://doi.org/10.1002/for.2268. 460, 468

Kowal, D. R. (2019). "Integer-valued functional data analysis for measles forecasting." *Biometrics*, 75(4): 1321–1333. MR4041833. doi: https://doi.org/10.1111/biom.13110. 459, 463

Kowal, D. R. (2020). "Supplementary Material of "Dynamic Regression Models for Time-Ordered Functional Data"." *Bayesian Analysis*. doi: https://doi.org/10.1214/20-BA1213SUPP. 461

Kowal, D. R. and Bourgeois, D. C. (2019). "Bayesian function-on-scalars regression for high-dimensional data." *Journal of Computational and Graphical Statistics*. Forthcoming. 460, 469, 481

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2017a). "A Bayesian multivariate functional dynamic linear model." *Journal of the American Statistical Association*, 112(518): 733–744. MR3671766. doi: https://doi.org/10.1080/01621459.2016.1165104. 465, 466, 467, 472, 477, 480, 481

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2017b). "Functional autoregression for sparsely sampled data." *Journal of Business & Economic Statistics*, 1–13. 461, 463, 467

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2019). "Dynamic shrinkage processes." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. MR3997101. doi: https://doi.org/10.1111/rssb.12325. 460, 468

Laurini, M. P. (2014). "Dynamic functional data analysis with non-parametric state space models." *Journal of Applied Statistics*, 41(1): 142–163. MR3291206. doi: https://doi.org/10.1080/02664763.2013.838663. 463

Laurini, M. P. and Hotta, L. K. (2010). "Bayesian extensions to Diebold-Li term structure model." *International Review of Financial Analysis*, 19(5): 342–350. 476

Mönch, E. (2008). "Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach." *Journal of Econometrics*, 146(1): 26–43. MR2459641. doi: https://doi.org/10.1016/j.jeconom.2008.06.002. 475

Montagna, S., Tokdar, S. T., Neelon, B., and Dunson, D. B. (2012). "Bayesian latent factor regression for functional and longitudinal data." *Biometrics*, 68(4): 1064–1073. MR3040013. doi: https://doi.org/10.1111/j.1541-0420.2012.01788.x. 460, 465, 469, 481

Morris, J. S. (2015). "Functional regression." *Annual Review of Statistics and Its Application*, 2: 321–359. 460

Morris, J. S. and Carroll, R. J. (2006). "Wavelet-based functional mixed models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2): 179–199. MR2188981. doi: https://doi.org/10.1111/j.1467-9868.2006.00539.x. 463

Mumtaz, H. and Surico, P. (2009). "Time-varying yield curve dynamics and monetary policy." *Journal of Applied Econometrics*, 24(6): 895–913. MR2750183. doi: https://doi.org/10.1002/jae.1084. 475, 476

Nelson, C. R. and Siegel, A. F. (1987). "Parsimonious modeling of yield curves." *Journal of Business*, 60(4): 473. 475

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). "Stochastic volatility with leverage: Fast and efficient likelihood inference." *Journal of Econometrics*, 140(2): 425–449. MR2408914. doi: https://doi.org/10.1016/j.jeconom.2006.07.008. 477

Park, S. Y. and Staicu, A.-M. (2015). "Longitudinal functional data analysis." *Stat*, 4(1): 212–226. MR3405402. doi: https://doi.org/10.1002/sta4.89. 463

Piironen, J. and Vehtari, A. (2016). "On the hyperprior choice for the global shrinkage parameter in the horseshoe prior." *arXiv preprint arXiv:1610.05559*. 469

Polson, N. G. and Scott, J. G. (2010). "Shrink globally, act locally: sparse Bayesian regularization and prediction." *Bayesian Statistics*, 9: 501–538. MR3204017. doi: https://doi.org/10.1093/acprof:oso/9780199694587.003.0017. 468, 469

Polson, N. G. and Scott, J. G. (2012). "Local shrinkage rules, Lévy processes and regularized regression." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(2): 287–311. MR2899864. doi: https://doi.org/10.1111/j.1467-9868.2011.01015.x. 468

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer. MR2168993. 460

Reiss, P. T., Huang, L., and Mennes, M. (2010). "Fast function-on-scalar regression with penalized basis expansions." *The International Journal of Biostatistics*, 6(1). MR2683940. doi: https://doi.org/10.2202/1557-4679.1246. 472

Suarez, A. J., Ghosal, S., et al. (2017). "Bayesian estimation of principal components for functional data." *Bayesian Analysis*, 12(2): 311–333. MR3620735. doi: https://doi.org/10.1214/16-BA1003. 465

Svensson, L. E. (1994). "Estimating and interpreting forward interest rates: Sweden 1992–1994." Technical report, National Bureau of Economic Research. 476

van der Pas, S., Kleijn, B., and van der Vaart, A. (2014). "The horseshoe estimator: Posterior concentration around nearly black vectors." *Electronic Journal of Statistics*, 8(2): 2585–2618. MR3285877. doi: https://doi.org/10.1214/14-EJS962. 469

Van Dijk, D., Koopman, S. J., Van der Wel, M., and Wright, J. H. (2014). "Forecasting interest rates with shifting endpoints." *Journal of Applied Econometrics*, 29(5): 693–712. MR3258059. doi: https://doi.org/10.1002/jae.2358. 478

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer. MR1482232. 460, 466

Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. CRC Press. MR3726911. 464

Xiao, L., Li, Y., and Ruppert, D. (2013). "Fast bivariate P-splines: the sandwich smoother." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3): 577–599. MR3065480. doi: https://doi.org/10.1111/rssb.12007. 472