Dynamic Function-on-Scalars Regression

Daniel R. Kowal *

June 6, 2018

Abstract

We develop a modeling framework for dynamic function-on-scalars regression, in which a time series of functional data is regressed on a time series of scalar predictors. The regression coefficient function for each predictor is allowed to be dynamic, which is essential for applications where the association between predictors and a (functional) response is time-varying. For greater modeling flexibility, we design a non-parametric reduced-rank functional data model with an unknown functional basis expansion, which is both data-adaptive and, unlike most existing methods, modeled as unknown for appropriate uncertainty quantification. Within a Bayesian framework, we introduce shrinkage priors that simultaneously (i) regularize time-varying regression coefficient functions to be locally static, (ii) effectively remove unimportant predictor variables from the model, and (iii) reduce sensitivity to the selected rank of the model. A simulation analysis confirms the importance of these shrinkage priors, with substantial improvements over existing alternatives. We develop a novel projection-based Gibbs sampling algorithm, which offers unrivaled computational scalability for fully Bayesian functional regression. We apply the proposed methodology (i) to characterize the effects of demographic predictors on age-specific fertility rates in South and Southeast Asia, and (ii) to analyze the time-varying impact of macroeconomic variables on the U.S. yield curve.

KEYWORDS: time series; Bayesian methods; factor model; fertility; yield curve

* Assistant Professor, Department of Statistics, Rice University, Houston, TX 77251-1892 (E-mail: daniel.kowal@rice.edu)
1 Introduction

We are interested in modeling the association between a functional response and scalar predictors, commonly referred to as function-on-scalars regression (FOSR); see Ramsay and Silverman (2005) and Morris (2015). We address the additional complication that the functional response and the scalar predictors are both time-ordered. Applications of time-ordered functional data, or functional time series, are abundant, including: daily interest rate curves as a function of time to maturity (Hays et al., 2012; Kowal et al., 2017); yearly sea surface temperature as a function of time-of-year (Besse et al., 2000); yearly mortality rates as a function of age (Hyndman and Ullah, 2007); daily pollution curves as a function of time-of-day (Damon and Guillas, 2002; Aue et al., 2015); and a collection of spatio-temporal applications in which a time-dependent variable is measured as a function of spatial location (e.g., Cressie and Wikle, 2011). In these applications and others, there may be interest in modeling the relationship between the functional time series and dynamic predictors.

A core feature of functional data is the presence of within-curve dependence, often recognizable as smoothness, and represented formally via a covariance function (Ramsay and Silverman, 2005; Horváth and Kokoszka, 2012). In functional regression, a fundamental challenge is appropriately accounting for this within-curve dependence, while simultaneously modeling the effects—or lack thereof—of the predictor variables. Recent contributions to FOSR include Chen et al. (2016), Barber et al. (2017), and Fan and Reimherr (2017), which focus on the setting in which the number of predictors is large, and variable selection is a priority. In the dynamic setting, the time-ordering of the functional data and the predictors introduces further complications. Unmodeled (time) dependence produces statistically inefficient estimators and can lead to incorrect inference and spurious relationships. Further, in many applications, the association between predictors and the functional response may be time-varying. Dangl and Halling (2012) discuss the importance of time-varying parameter regression for macroeconomic data, but the concepts are broadly applicable: structural
shifts—which may occur with or without warning—obscure (dynamic) relationships and produce inferior estimates, predictions, and forecasts. It is therefore essential to account for both time-dependence and time-variation.

We propose a Bayesian dynamic function-on-scalars regression (DFOSR) model to jointly model within-curve (functional) dependence, between-curve (time) dependence, and dependence due to dynamic predictors. Within-curve dependence is modeled nonparametrically using a reduced-rank functional data model, which provides model flexibility for broad applicability. The unknown basis functions are endowed with a prior distribution that encourages smoothness, produces data-adaptive basis functions, and incorporates uncertainty quantification via the posterior distribution. We introduce an autoregressive structure for between-curve dependence and model the dynamic predictors by extending time-varying parameter regression to the functional data setting. Time-varying parameter regression has successfully improved estimation and forecasting for scalar time series (Dangl and Halling, 2012; Korobilis, 2013; Belmonte et al., 2014; Kowal et al., 2017b), but due to the large number of parameters and computational challenges has not yet been used for functional data.

We introduce Bayesian shrinkage priors for functional regression that simultaneously (i) regularize the time-varying regression coefficient functions to be locally static, (ii) effectively remove unimportant predictor variables from the model, and (iii) reduce sensitivity to the choice of the rank of the model. Our simulation study (Section 6) confirms the importance of these choices—the model for the unknown basis functions, the time-varying parameter regression, and the shrinkage priors—and demonstrates decisive improvements in statistical efficiency relative to state-of-the-art alternatives. For computationally scalable posterior inference, we develop an efficient Gibbs sampling algorithm. Using a novel projection scheme, we decouple the sampling blocks for functional dependence and non-functional dependences—i.e., time-dependence of the functional data and time-varying regression functions—such that the computational complexity of the non-functional sampling block is determined by the rank of the model $K$, rather than the number of observation
points $M$, typically with $K \ll M$. The general sampling scheme is broadly applicable for functional data models with (i) unknown (nonparametric) basis functions and (ii) complex (non-functional) dependences.

Our methodology is motivated by two applications. First, we analyze age-specific fertility rates (ASFRs) for developing nations in South and Southeast Asia. ASFRs measure fertility as a function of age within a population, which changes over time, and may depend on socioeconomic and demographic predictor variables. Fertility is an important factor in the health and welfare of women and their families and a fundamental component in population growth, with major implications for planning and allocation of resources. Our methodology provides a mechanism for understanding how various socioeconomic and demographic variables impact the shape of the ASFR, which allows for differential age-specific effects with appropriate uncertainty quantification via the posterior distribution.

Second, we study the impact of macroeconomic variables on the U.S. yield curve. For a given currency and level of risk of a debt, the yield curve describes the interest rate at time $t$ as a function of the length of the borrowing period, or time to maturity, $\tau$, and is closely linked to various aspects of the U.S. economy. Following Diebold et al. (2006), we consider real activity, monetary policy, and inflation as our predictor variables. In contrast with Diebold et al. (2006), we relax the parametric (Nelson-Siegel) assumption for the functional component and allow the macroeconomic associations with the yield curve to be time-varying. As a result, we gain insight into how these important macroeconomic variables are related to interest rates of different maturities, and how these relationships vary over time.

The remainder of the paper is organized as follows: we introduce the model in Section 2; the reduced-rank functional data model is in Section 3; the shrinkage priors are in Section 4; model properties are in Section 5; a simulation analysis is in Section 6; we apply the model to age-specific fertility rates in Section 7 and yield curves in Section 8; the MCMC algorithm is in Section 9; and we conclude in Section 10. Additional details on the MCMC algorithm, simulations, and the applications are provided in the Appendix.
2 Dynamic Function-on-Scalars Regression

Let $Y_1, \ldots, Y_T$ be a time-ordered sequence of random functions in $L^2(\mathcal{T})$, where $\mathcal{T} \subset \mathbb{R}^D$ is a compact index set and $D \in \mathbb{Z}^+$. Suppose we have a time-ordered sequence of predictors $x_t = (x_{1,t}, \ldots, x_{p,t})'$ for $t = 1, \ldots, T$, and we are interested in modeling the association between the scalar predictors $x_{j,t}$ and the functional response $Y_t$. Importantly, we consider the setting in which the relationship between $x_{j,t}$ and $Y_t$ may be time-varying, which is common for applications in which structural breaks may occur. The proposed dynamic function-on-scalars regression model (DFOSR) has three levels, which are jointly expressed via (1)-(3) below.

First, we decompose the functional time series $Y_t(\tau)$ into a linear combination of $K$ loading curves, $\{f_k(\tau)\}_{k=1}^K$, and factors, $\{\beta_k,t\}_{k=1}^K$, for each time $t = 1, \ldots, T$:

$$Y_t(\tau) \equiv \sum_{k=1}^{K} f_k(\tau) \beta_{k,t} + \epsilon_t(\tau), \quad \epsilon_t(\tau) \sim \text{indep } N(0, \sigma^2_\epsilon), \quad \tau \in \mathcal{T} \quad (1)$$

Model (1) is a dynamic functional factor model: the loadings $\{f_k\}$ are modeled as smooth unknown functions of $\tau$ to account for the within-curve correlation structure in $Y_t(\tau)$, and the factors $\{\beta_{k,t}\}$ are modeled dynamically to account for the between-curve time dependence in $Y_t(\tau)$. Equivalently, we may interpret $\{f_k\}$ as a time-invariant functional basis for $Y_t(\tau)$ with dynamic basis coefficients $\{\beta_{k,t}\}$, which we model using dynamic predictor variables (see (2) below). Despite the $K$-term expansion in (1), the contemporaneous covariance function of $\{Y_t\}$ is full rank (see Proposition 2); it is the autocovariance function of $\{Y_t\}$ that is reduced rank $K$ (see Proposition 3). Each $f_k$ is modeled nonparametrically using low-rank thin plate splines, which are well-defined for $\mathcal{T} \subset \mathbb{R}^D$ with $D \in \mathbb{Z}^+$ and are smooth, flexible, and efficient to compute (Ruppert et al., 2003; Wood, 2006). By modeling the $\{f_k\}$ as unknown, and imposing suitable identifiability constraints (see Section 3.2), our model incorporates the uncertainty of $\{f_k\}$ into the posterior distribution for all parameters of interest, which is necessary for valid inference.
Next, we introduce a dynamic regression component to incorporate the predictors $x_{j,t}$:

$$\beta_{k,t} = \mu_k + \sum_{j=1}^{p} x_{j,t} \alpha_{j,k,t} + \gamma_{k,t}, \quad \gamma_{k,t} = \phi_k \gamma_{k,t-1} + \eta_{k,t}, \quad \eta_{k,t} \overset{\text{indep}}{\sim} N(0, \sigma_{\eta_{k,t}}^2)$$

(2)

where $\mu_k$ is the intercept for factor $k$, $\alpha_{j,k,t}$ is the time-varying regression coefficient for predictor $j$ and factor $k$ at time $t$, and $\gamma_{k,t}$ is the regression error term, which we allow to be autocorrelated via an AR(1) process. Each regression coefficient $\alpha_{j,k,t}$ varies with $k$, and therefore its association with $Y_t(\tau)$ for a particular $\tau$ may be interpreted via the loading curve $f_k(\tau)$. The generality of (2) is useful for broad applicability, but important special cases exist. If $x_t = 0$ for all $t$, i.e., there are no predictors, model (1)-(2) produces 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). If $\alpha_{j,k,t} = \alpha_{j,k}$ for all $t$ and $\phi_k = 0$ for all $k$, model (1)-(2) is a (Bayesian) FOSR model. If $\alpha_{j,k,t} = \alpha_{j,k}$ for all $t$, model (1)-(2) is a Bayesian FOSR model with autoregressive errors (FOSR-AR).

Lastly, we specify the dynamics—and regularization—for the regression coefficients, $\alpha_{j,k,t}$:

$$\alpha_{j,k,t} = \alpha_{j,k,t-1} + \omega_{j,k,t}, \quad \omega_{j,k,t} \overset{\text{indep}}{\sim} N(0, \sigma_{\omega_{j,k,t}}^2)$$

(3)

For each $k$, (2)-(3) is a time-varying parameter regression for the dynamic predictors $x_{j,t}$, where the factors $\beta_{k,t}$ operate as the response variable. We select priors for $\sigma_{\omega_{j,k,t}}^2$ in Section 4 to encourage shrinkage of $\alpha_{j,k,t}$. Locally, we shrink $\omega_{j,k,t}$ toward zero, which implies that $\alpha_{j,k,t} \approx \alpha_{j,k,t-1}$ is locally constant at time $t$. Importantly, the factor-specific regression coefficients $\alpha_{j,k,t}$ are allowed to change at any time $t$, but the shrinkage prior encourages the more parsimonious FOSR-AR model. Globally, we shrink $\omega_{j,k,t}$ toward zero for all $t$, which, combined with shrinkage of the initial state $\alpha_{j,k,0}$, effectively removes factor $k$ for predictor $j$ from the model. Finally, we introduce ordered shrinkage across $k = 1, \ldots, K$ to cumulatively reduce $a priori$ the relative importance of the higher number factors $k$. By imposing ordered
shrinkage, we mitigate the impact of the choice of the number of factors $K$, as long as $K$ is chosen sufficiently large. The simulation analysis in Section 6 validates the importance of each of these shrinkage priors.

The DFOSR model (1)-(3) is completed by specifying prior distributions on the functions $\{f_k(\cdot)\}$, the intercepts $\{\mu_k\}$, the autoregressive coefficients $\{\phi_k\}$, and the variance components $\{\sigma^2_{\tilde{\epsilon}_t}\}$, $\{\sigma^2_{\eta_{k,t}}\}$, and $\{\sigma^2_{\omega_{j,k,t}}\}$. Note that the distributions for $\epsilon_t(\tau)$, $\eta_{k,t}$, and $\omega_{j,k,t}$ are conditionally Gaussian, so careful choice of priors for the accompanying variance components will produce a variety of marginal distributions (e.g., Wand et al., 2011) within the framework of (1)-(3).

The specification of model (1)-(3) using the $k$-specific factor terms, particularly in (2) and (3), is useful for interpretability, since each term indexed by $k$ may be interpreted via loading curve $f_k$, and for constructing computationally efficient MCMC sampling techniques. The DFOSR (1)-(3) also induces a model representation in the functional $\tau \in \mathcal{T}$ space. Let $\mathcal{G}\mathcal{P}(c,C)$ denote a Gaussian process with mean function $c$ and covariance function $C$.

**Proposition 1.** Model (1)-(3) implies the dynamic functional regression model

$$Y_t(\tau) = \tilde{\mu}(\tau) + \sum_{j=1}^{p} x_{j,t} \tilde{\alpha}_{j,t}(\tau) + \tilde{\gamma}_t(\tau) + \epsilon_t(\tau), \quad \epsilon_t(\tau) \overset{\text{indep}}{\sim} N(0, \sigma^2_{\tilde{\epsilon}_t}), \quad \tau \in \mathcal{T} \quad (4)$$

$$\tilde{\gamma}_t(\tau) = \int \tilde{\phi}(\tau, u) \tilde{\gamma}_{t-1}(u) \, du + \tilde{\eta}_t(\tau), \quad \tilde{\eta}_t(\cdot) \overset{\text{indep}}{\sim} \mathcal{G}\mathcal{P}(0, C_{\eta}) \quad (5)$$

$$\tilde{\alpha}_{j,t}(\tau) = \tilde{\alpha}_{j,t-1}(\tau) + \tilde{\omega}_t(\tau), \quad \tilde{\omega}_j(\cdot) \overset{\text{indep}}{\sim} \mathcal{G}\mathcal{P}(0, C_{\omega_{j,t}}) \quad (6)$$

under the expansions $\tilde{\mu}(\tau) = \sum_k f_k(\tau) \mu_k$, $\tilde{\alpha}_{j,t}(\tau) = \sum_k f_k(\tau) \alpha_{j,k,t}$, $\tilde{\gamma}_t(\tau) = \sum_k f_k(\tau) \gamma_{k,t}$, $\tilde{\phi}(\tau, u) = \sum_k f_k(\tau) f_k(u) \phi_k$, $\tilde{\eta}_t(\tau) = \sum_k f_k(\tau) \eta_{k,t}$, $\tilde{\omega}_{j,t}(\tau) = \sum_k f_k(\tau) \omega_{j,k,t}$, and the covariance functions $C_{\eta}(\tau, u) = \sum_k f_k(\tau) f_k(u) \sigma^2_{\eta_{k,t}}$, and $C_{\omega_{j,t}}(\cdot, u) = \sum_k f_k(\tau) f_k(u) \sigma^2_{\omega_{j,k,t}}$.

The predictors $x_{j,t}$ are directly associated with the functional time series $Y_t(\tau)$ via the dynamic regression coefficient functions $\tilde{\alpha}_{j,t}(\tau)$. Therefore, while we may interpret $\alpha_{j,k,t}$ in (2) using the $k$th loading curve $f_k$, we may instead examine the dynamic regression coefficient
function $\tilde{\alpha}_{j,t}(\tau)$ for predictor $j$ at time $t$ as a function of $\tau$. Since we obtain MCMC draws from the posterior distribution of $\{f_k\}$ and $\{\alpha_{j,k,t}\}$, we may conduct posterior inference on $\tilde{\alpha}_{j,t}(\tau)$ directly without modifying the MCMC sampling algorithm.

The error term $\tilde{\gamma}_t(\tau)$ in (4) captures the large-scale variability in $Y_t(\tau)$ at time $t$, and is autocorrelated, while the error term $\epsilon_t(\tau)$ models the small-scale variability, i.e., the observation error. In equation (5), the autocorrelated large-scale error terms $\tilde{\gamma}_t(\tau)$ follow a functional autoregressive model, which is the functional data analog of (vector) autoregression for time series data (e.g., Kowal et al., 2017c). Note that the representation in (5) requires a (functional) orthonormality constraint on the loading curves, $\int f_k(\tau)f_\ell(\tau)d\tau = I(k = \ell)$ for $I(\cdot)$ the indicator function. We modify this constraint for more efficient computations in Section 3 but our modification approximates this constraint.

Among existing FOSR methods, the most common approach is to represent the functional data $Y_t$ using a basis expansion. The functional regression model propagates to the basis coefficients—analogous to $\beta_{k,t}$ in (2)—so the dimensionality of the basis governs the dimensionality of the regression. Methods that use full basis expansions, such as splines (Ramsay and Silverman, 2005; Laurini, 2014) or wavelets (Morris and Carroll, 2006; Zhu et al., 2011), are neither parsimonious nor computationally scalable in the presence of other dependence or model complexity, such as autocorrelated functional data or time-varying regression functions. An alternative approach is to pre-compute a lower-dimensional basis, such as in functional principal components analysis (FPCA; Di et al., 2009; Goldsmith et al., 2013; Goldsmith and Kitago, 2016), which does provide model parsimony and computational scalability. However, methods that pre-compute a functional basis fail to account for the uncertainty in the unknown basis. This uncertainty is nontrivial: Goldsmith et al. (2013) demonstrate that FPC-based methods may substantially underestimate total variability, even for densely-observed functional data. In addition to the inferential impediments, such two-stage estimation techniques are often statistically inefficient (e.g., Di et al., 2009).

Several existing Bayesian reduced-rank functional data models do account for the uncer-
tainty in the dimension reduction. Suárez et al. (2017) propose a Bayesian FPCA, but do not incorporate predictors or dependence structures, and rely on a computationally expensive reversible-jump MCMC. Montagna et al. (2012) incorporate predictors, but the model is non-dynamic and does not include shrinkage priors to reduce the impact of unimportant variables. Kowal et al. (2017a) propose a functional dynamic linear model that may include dynamic predictors, but do not use shrinkage priors for the (possibly time-varying) regression coefficients, which results in substantially less accurate estimates with larger variability (see Section 6 and the Appendix). In addition, Kowal et al. (2017a) only consider functional data with observation points \( \tau \in \mathcal{T} \subset \mathbb{R}^D \) for \( D = 1 \), which limits applicability. Importantly, the MCMC algorithm in Kowal et al. (2017a) is also massively inefficient compared to our proposed MCMC algorithm (see Table 1).

3 Modeling the Loading Curves

The loading curves \( \{f_k\} \) capture the within-curve dependence (e.g., smoothness) of the functional data \( Y_t \) and provide the dimension reduction for both the autocorrelation and regression components in (2). Reduced-rank functional data models have been highly successful in a variety of applications and settings, including James et al. (2000), Di et al. (2009), Hays et al. (2012), Montagna et al. (2012), Jungbacker et al. (2013), and Kowal et al. (2017a). These methods and others demonstrate that a data-adaptive basis expansion can adequately model complex functional data with only a small number of components.

In our setting, there are several significant challenges to address. First, we must account for the inherent uncertainty in the \( \{f_k\} \): failure to account for the variability in \( \{f_k\} \) can lead to substantial underestimation of total variability—even for densely-observed curves—resulting in incorrect inference (Goldsmith et al., 2013). Next, the number of observation points, \( M \), may be extremely large for functional data (e.g., \( M > 10000 \)). Methods that scale poorly in \( M \), such as Gaussian processes (e.g., Montagna et al. 2012), will limit the
computational scalability of the MCMC algorithm. Further, the domain of the functional data is \( \mathcal{T} \subset \mathbb{R}^D \) with \( D \in \mathbb{Z}^+ \). While it is common for functional data models to focus on \( D = 1 \), we aim to develop methodology that applies for \( D > 1 \) as well. Lastly, we must enforce identifiability constraints on \( \{f_k\} \) in order to interpret each \( f_k \) as well as the \( k \)-specific parameters in (2)-(3).

To address these challenges—while accounting for the uncertainty in the unknown loading curves \( \{f_k\} \)—we model each \( f_k \) using low rank thin plate splines, which are are well-defined for \( \mathcal{T} \subset \mathbb{R}^D \) with \( D \in \mathbb{Z}^+ \). Low rank thin plate splines are smooth and flexible, and are known to be efficient in MCMC samplers \cite{Crainiceanu2005}. For identifiability, we design an orthonormality constraint for \( \{f_k\} \) that is both easy to enforce and, most importantly, may be used to simplify and substantially improve computational efficiency of the MCMC sampler for the parameters in (2)-(3).

### 3.1 Full Conditional Distributions: General Basis Functions

A common approach in nonparametric regression and functional data analysis 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. Popular choices include splines, Fourier basis functions, wavelets, and radial basis functions \cite{Ramsay2005, Morris2015}. We present results for arbitrary basis expansions, but provide details for our preferred low rank thin plate spline implementation in Section 3.3.

For each unknown loading curve \( f_k \), let \( f_k(\tau) = b'(\tau)\psi_k \), where \( b'(\tau) = (b_1(\tau), \ldots, b_{LM}(\tau)) \) is an \( L_M \)-dimensional vector of known basis functions and \( \psi_k \) is an \( L_M \)-dimensional vector of unknown basis coefficients. The choice of basis functions may be application-specific, and the number of basis functions \( L_M \) may depend on the selected basis and the number of observation points, \( M \); we provide default specifications for low rank thin plate splines in Section 3.3. Typically, basis expansions are combined with a suitable penalty function, such as \( P(f_k) = \int \left( \dddot{f}_k(\tau) \right)^2 d\tau \) for \( \dddot{f}_k \) the second derivative of \( f_k \) (assuming \( D = 1 \), which encourages
smoothness and guards against overfitting. For Bayesian implementations, such penalties correspond to prior distributions on the basis coefficients \( \psi_k \), or equivalently, the implied function \( f_k \). For example, the smoothness penalty above may be written \( P(f_k) = \psi_k' \Omega_b \psi_k \) for known \( L_M \times L_M \) penalty matrix \( \Omega_b \) with \((\ell, \ell')\) entry \( [\Omega_b]_{\ell, \ell'} = \int \hat{b}_\ell(\tau) \hat{b}_{\ell'}(\tau) d\tau \), which is commonly expressed as a Gaussian prior on \( \psi_k \) with prior precision \( \Omega_b \) (and usually includes a positive smoothing parameter). For generality, we assume the prior \( \psi_k \overset{\text{indep}}{\sim} N(0, \Sigma_{\psi_k}) \) for \( k = 1, \ldots, K \), which implies a Gaussian process prior on \( f_k \) with mean function zero and covariance function \( \text{Cov}(f_k(\tau), f_k(\tau')) = b'(\tau) \Sigma_{\psi_k} b(\tau') \).

Given functional data observations \( Y_t = (Y_t(\tau_1), \ldots, Y_t(\tau_M))' \) at observation points \( \{\tau_j\}_{j=1}^M \), the likelihood in (1) becomes

\[
Y_t = \sum_{k=1}^K f_k \beta_{k,t} + \epsilon_t, \quad \epsilon_t \overset{\text{indep}}{\sim} N(0, \sigma^2_\epsilon I_M)
\] (7)

where \( f_k = (f_k(\tau_1), \ldots, f_k(\tau_M))' = B \psi_k \) are the loading curves evaluated at the observation points, with \( B = (b(\tau_1), \ldots, b(\tau_M))' \) the \( M \times L_M \) basis matrix. We construct a Bayesian backfitting sampling algorithm that iteratively draws from the full conditional distribution of each \( f_k \) conditional on \( \{f_\ell\}_{\ell \neq k} \). The full conditional distribution of the corresponding basis coefficients is \( [\psi_k | \cdots] \overset{\text{indep}}{\sim} N(Q_{\psi_k}^{-1} \ell_{\psi_k}, Q_{\psi_k}^{-1}) \), where \( Q_{\psi_k} = (B' B) \sum_{t=1}^T \beta_{k,t}^2 / \sigma^2_\epsilon + \Sigma_{\psi_k}^{-1} \) and \( \ell_{\psi_k} = B' \sum_{t=1}^T \beta_{k,t}^2 / \sigma^2_\epsilon \left( Y_t - \sum_{\ell \neq k} f_\ell \beta_{\ell,t} \right) \). Sampling \( \psi_k \) from this posterior distribution has computational complexity at most \( \mathcal{O}(L_M^3) \), and for some choices of bases and penalties (e.g., wavelets) is as fast as \( \mathcal{O}(L_M^2) \). By comparison, a full rank Gaussian process has computational complexity \( \mathcal{O}(M^3) \), and further requires computation of the inverse \( \Sigma_{\psi_k}^{-1} \), which in general will not have a simple structure. For thin plate splines, \( \Sigma_{\psi_k}^{-1} \) is known up to a scaling constant, which precludes the need to compute this inverse as part of the MCMC sampler.
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 allows us to interpret \( \{f_k\} \) and the \( k \)-specific model parameters in (2) and (3). Second, our particular choice of constraints provides massive computational improvements for sampling the parameters in (2) and (3). While Kowal et al. (2017a) impose the constraint \( \int f_k(\tau)f_\ell(\tau)d\tau = I\{k = \ell\} \), we instead use the constraint \( F'F = I_K \), where \( F = (f_1, \ldots, f_K) \) is the \( M \times K \) matrix of loading curves evaluated at the observation points \( \tau_1, \ldots, \tau_M \) and \( I_K \) is the \( K \times K \) identity matrix. This constraint, combined with a suitable ordering constraint on \( k = 1, \ldots, K \) (see Section 4), is sufficient for identifiability (up to sign changes, which in our experience are not problematic in the MCMC sampler).

The utility of our orthonormality constraint is illustrated with the following result:

**Lemma 1.** Under the identifiability constraint \( F'F = I_K \), the joint likelihood in (7) is

\[
p(Y_1, \ldots, Y_T|\{f_k, \beta_{k,t}, \sigma_{\epsilon_t}\}_k,t) = c_Y \prod_{t=1}^T \sigma_{\epsilon_t}^{-M} \exp \left\{ -\frac{1}{2\sigma_{\epsilon_t}^2} [Y_t'Y_t + \beta_t'\beta_t - 2\beta_t'(F'Y_t)] \right\}
\]

where \( c_Y = (2\pi)^{-MT/2} \) is a constant and \( \beta_t' = (\beta_{1,t}, \ldots, \beta_{K,t}) \).

The fundamental insight of Lemma 1 is that the factors, \( \beta_{k,t} \)—and therefore the model parameters that depend on the factors—depend on the functional data \( Y_t \) only via the projected data \( \tilde{Y}_t \equiv F'Y_t \). Importantly, the projected data \( \tilde{Y}_t \) are \( K \)-dimensional, whereas the functional data \( Y_t \) are \( M \)-dimensional, typically with \( K \ll M \). In particular, Lemma 1 implies the following working likelihood for the factors \( \beta_{k,t} \) and associated parameters:

**Lemma 2.** Under the identifiability constraint \( F'F = I_K \), the joint likelihood in (7) for \( \{\beta_{k,t}\} \) is equivalent to the working likelihood implied by

\[
\tilde{Y}_{k,t} = \beta_{k,t} + \tilde{\epsilon}_{k,t}, \quad \tilde{\epsilon}_{k,t} \overset{\text{indep}}{\sim} N(0, \sigma_{\epsilon_t}^2)
\]

up to a constant that does not depend on \( \beta_{k,t} \), where \( \tilde{Y}_{k,t} = f_k'Y_t \) and \( \tilde{\epsilon}_{k,t} = f_k'\epsilon_t \).
For sampling the factors $\beta_{k,t}$ (and associated parameters), we only need the likelihood (9), which only depends on $M$ via the projection $\tilde{\mathbf{Y}}_t \equiv \mathbf{F}' \mathbf{Y}_t$. The projection step is a one-time cost (per MCMC iteration) for all variables in (2) and (3). As a result, the model complexity for the dynamic components in (2) and (3) is not severely limited by the dimension of the functional data, $M$. These computational simplifications afford us the ability to incorporate the complex dynamics in (2)-(3) without sacrificing computational feasibility. Similar computational reductions are available in the more general case of (7) with $\epsilon_t \sim \mathcal{N}(0, \Sigma_{\epsilon_t})$ for general $M \times M$ covariance matrix $\Sigma_{\epsilon_t}$; see the Appendix for details.

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) (defined as DFOSR-NIG in Section 6). Notably, Kowal et al. (2017a) use a similar model for $\{f_k\}$, but do not use the identifiability constraint $\mathbf{F}' \mathbf{F} = \mathbf{I}_K$ to produce the simplifications in Lemmas 1 and 2. The improvements are substantial, particularly for the larger sample size.

| MCMC Algorithm | $T = 50, M = 20$ | $T = 200, M = 100$ |
|----------------|-------------------|-------------------|
| Proposed DFOSR | 48 seconds         | 3 minutes         |
| Kowal et al. (2017c) | 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 $K = 6$.

For each $f_k$, the orthonormality constraint may be decomposed into two sets of constraints: the linear constraints $f'_\ell f_k = 0$ for $\ell \neq k$ and the unit-norm constraint, $||f_k||^2 = 1$. Since the sampler in Section 3.1 conditions on $\{f_\ell\}_{\ell \neq k}$, the linearity constraint is fixed for each $f_k = \mathbf{B}\psi_k$. Therefore, given the full conditional distribution $[\psi_k | \cdots] \sim \mathcal{N}(Q_{\psi_k}^{-1} \ell_{\psi_k}, Q_{\psi_k}^{-1})$, we enforce the linear orthogonality constraint by conditioning on $C_k \psi_k = 0$, where $C_k = (f_1, \ldots, f_{k-1}, f_{k+1}, \ldots, f_K)' \mathbf{B} = (\psi_1, \ldots, \psi_{k-1}, \psi_{k+1}, \ldots, \psi_K)' \mathbf{B}' \mathbf{B}$. Conditioning on the constraint is particularly interpretable in a Bayesian setting, and produces desirable optimality properties for constrained penalized regression (see Theorem 1 of Kowal et al. 2017a). Since the full conditional distribution for $\psi_k$ is Gaussian, conditioning on $C_k \psi_k = 0$ produces a Gaussian distribution with easily computable mean and covariance. Sampling
from the constrained distribution is straightforward and efficient: given a draw from the unconstrained posterior, say \( \psi^0_k \sim N(Q^{-1}\psi_k, Q^{-1}) \), we retain the vector \( \psi_k^* = \psi^0_k - Q^{-1}\psi_k C_k' (C_k Q^{-1}\psi_k C_k')^{-1} C_k \psi^0_k \). Given the orthogonally-constrained sample \( f_k^* = B \psi_k^* \), we rescale to enforce the unit-norm constraint: \( f_k = f_k^*/||f_k^*|| \). This rescaling does not change the shape of the loading curve \( f_k \), and can be counterbalanced by an equivalent rescaling of the corresponding factor, i.e., \( \beta_{k,t} \leftarrow \beta_{k,t}||f_k^*|| \), to preserve exactly the likelihood (7). By applying this procedure iteratively for \( k = 1, \ldots, K \), the constraint \( F' F = I_K \) is satisfied for every MCMC iteration.

3.3 Low Rank Thin Plate Splines

Thin plate splines are designed for modeling an unknown smooth function of multiple inputs \( \tau \in T \subset \mathbb{R}^D \) with \( D \in \mathbb{Z}^+ \). Thin plate splines place a (known) basis function at every observation point, so \( L_M = M \); low rank thin plate splines (LR-TPS) select a smaller set of basis functions \( L_M < M \). LR-TPS can achieve similar estimation accuracy as thin plate splines in a fraction of the computing time, and demonstrate exceptional MCMC efficiency (Crainiceanu et al., 2005). Each LR-TPS \( f_k \) has only one hyperparameter \( \lambda_{fk} > 0 \), which is a prior precision corresponding to the smoothness parameter (Wahba, 1990).

Given observation points \( \tau_j \) for \( j = 1, \ldots, M \), we construct the basis and penalty matrices in three steps: (i) we build the LR-TPS basis and penalty matrices using the definitions in Wood (2006); (ii) we diagonalize the penalty matrix for an equivalent representation, following Ruppert et al. (2003) and Crainiceanu et al. (2005); and (iii) we orthonormalize the basis matrix (and adjust the penalty matrix accordingly). The diagonalization and orthonormalization steps (ii) and (iii) may accompany any choice of basis and penalty matrices, but substantially improve MCMC performance for LR-TPS. Note that while the diagonalization step (ii) is not strictly necessary given the orthonormalization step (iii), it ensures that the final penalty matrix—and therefore the prior precision matrix—is positive definite, which is not guaranteed for LR-TPS (Ruppert et al., 2003).
To build the basis and penalty matrices, we begin by selecting the number and location of knots. For a small number of observation points, $M \leq 25$, we use the full rank thin plate spline basis with knots at the unique observation points $\kappa_\ell = \tau_\ell$. When $M > 25$, we use $(L_M - D - 1) = \min\{M/4,150\}$ knots. In the case of $D = 1$, knots are selected using the quantiles of the observation points, i.e., $\kappa_\ell$ is the $(\ell/L_M)$th sample quantile of the unique $\tau_j$; for $D > 1$, we select knot locations using a space-filling algorithm, as in [Ruppert et al. (2003)]. Let $W_0$ be the $M \times (D + 1)$ matrix with $j$th row $[W_0]_j = (1, \tau_j')$, $Z_0$ be the $M \times (L_M - D - 1)$ matrix with $(j,\ell)$th entry $[Z_0]_{j,\ell} = b(||\tau_j - \kappa_\ell||)$, and $\Omega_{Z_0}$ be the $(L_M - D - 1) \times (L_M - D - 1)$ penalty matrix with $(\ell,\ell')$th entry $[\Omega_{Z_0}]_{\ell,\ell'} = b(||\kappa_\ell - \kappa_{\ell'}||)$, where $b(r) = r^{4-D} \log(r)$ for $D$ even and $b(r) = r^{4-D}$ for $D$ odd, $r > 0$, are the (nonlinear) cubic thin plate spline basis functions (Wood (2006)). The matrices $W_0$ and $Z_0$ constitute the LR-TPS basis matrix, while $\Omega_{Z_0}$ is the LR-TPS penalty matrix. To diagonalize the penalty matrix, let $B_0 = [W_0 : Z_0 \Omega_{Z_0}^{-1/2}]$ be the LR-TPS basis matrix and $\Omega_0 = \text{diag}(0'_{D+1}, 1'_{L_M-D-1})$ be the diagonalized LR-TPS penalty matrix, where $0_{D+1}$ is a $(D+1)$-dimensional vector of zeros and $1_{L_M-D-1}$ is a $(L_M - D - 1)$-dimensional vector of ones. Lastly, let $B_0 = QR$ be the QR decomposition of the initial basis matrix $B_0$, where $Q$ is $L_M \times L_M$ with $Q'Q = I_{L_M}$ and $R$ is $L_M \times L_M$ and upper triangular. Using the orthonormal basis matrix $B = Q$, we reparameterize the penalty matrix $\Omega = (R')^{-1}\Omega_0R^{-1}$ to obtain an equivalent representation. Notably, this basis matrix $B$ and penalty matrix $\Omega$ construction is a one-time cost.

For orthonormalized LR-TPS, the full conditional distribution simplifies to $[\psi_k | \cdots] \sim N\left(Q_{\psi_k}^{-1} \ell_{\psi_k}, Q_{\psi_k}^{-1}\right)$, where $Q_{\psi_k} = I_{L_M} \sum_{t=1}^T \beta_{k,t}^2/\sigma_{\epsilon_t}^2 + \lambda_{f_k} \Omega$ and $\ell_{\psi_k} = \sum_{t=1}^T [\beta_{k,t}/\sigma_{\epsilon_t}^2 (B'Y_t)] - \sum_{t=1}^T [\beta_{k,t}/\sigma_{\epsilon_t}^2 \sum_{t\neq k} \psi_t/\beta_{\ell,t}].$ Since $\lambda_{f_k} > 0$ corresponds to a prior precision parameter, we follow [Gelman (2006) and Kowal et al. (2017a)] and impose a uniform prior distribution on the corresponding standard deviation, $\lambda_{f_k}^{-1/2} \overset{iid}{\sim} \text{Uniform}(0,10^4)$. Sampling each $\psi_k$ does not require inverses to compute $Q_{\psi_k}$ and has computational complexity $O(L_M^3)$ using the Cholesky-based approach of [Rue (2001)]; see the Appendix for details. For smooth $f_k$, $L_M$ may be much smaller than $M$ without compromising the estimate of $f_k$. 

15
4 Shrinkage Priors for the Model

While the DFOSR (1)-(3) is highly flexible, it is also overparametrized: it is unlikely that the regression coefficients $\alpha_{j,k,t}$ change substantially for all times $t$, or that every predictor $x_{j,t}$ has a strong association with the functional response $Y_t$. Careful choices of priors for $\sigma_{\eta_{k,t}}^2$ and $\sigma_{\omega_{j,k,t}}^2$ offer shrinkage toward simpler models, which often improves estimation accuracy and reduces variability (see Section 6). We propose nested horseshoe priors (Carvalho et al., 2010) for shrinkage toward locally-static regression models with fewer predictors, and multiplicative gamma process priors (Bhattacharya and Dunson, 2011) for ordered shrinkage across factors $k = 1, \ldots, K$, which reduces the sensitivity to the choice of $K$. In (non-functional) time-varying parameter regression, shrinkage priors offer improvements in prediction and provide narrower posterior credible intervals (e.g., Kowal et al., 2017b). Zhu et al. (2011) illustrate the importance of shrinkage priors for functional data regression in the non-dynamic setting, which is further compounded by the substantial increase in the number of parameters in the time-varying parameter regression model (2)-(3).

For the dynamic regression coefficient innovations $\omega_{j,k,t} \overset{\text{ind}}{\sim} N(0, \sigma_{\omega_{j,k,t}}^2)$, we encourage shrinkage at multiple levels with the following hierarchy of half-Cauchy distributions:

$$
\sigma_{\omega_{j,k,t}} \overset{\text{ind}}{\sim} C^+(0, \lambda_{j,k}), \quad \lambda_{j,k} \overset{\text{ind}}{\sim} C^+(0, \lambda_j), \quad \lambda_j \overset{\text{ind}}{\sim} C^+(0, \lambda_0), \quad \lambda_0 \overset{\text{ind}}{\sim} C^+(0, 1/\sqrt{T-1}) \quad (10)
$$

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 $\lambda_j$ 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 $\lambda_j$ operators as a group shrinkage parameter that may effectively remove predictor $j$ from the model. Lastly, the global shrinkage parameter $\lambda_0$ 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: \( \omega_{j,k,t}^\text{indep} \sim N(0, \lambda_{j,k}^2) \). The simulation analysis of Section 6 clearly demonstrate the importance of these shrinkage priors, particularly for time-varying parameter regression.

The multiplicative gamma process (MGP) provides ordered shrinkage with respect to factor \( k \), which suggests that factors with larger \( k \) explain less variability in the data, and effectively reduces sensitivity to the choice of the total number of factors, \( K \). We assume MGP priors for the intercept terms \( \mu_k^\text{indep} \sim N(0, \sigma_{\mu_k}^2) \), which are given by the prior on the precisions, \( \sigma_{\mu_k}^{-2} = \prod_{\ell \leq k} \delta_{\mu_k^\ell} \), where \( \delta_{\mu_k^1} \sim \text{Gamma}(a_{\mu_1}, 1) \) and \( \delta_{\mu_k^\ell} \sim \text{Gamma}(a_{\mu_2}, 1) \) for \( \ell > 1 \). As discussed in Bhattacharya and Dunson (2011) and Durante (2017), selecting \( a_{\mu_1} > 0 \) and \( a_{\mu_2} \geq 2 \) produces stochastic ordering among the implied variances \( \sigma_{\mu_k}^2 \), which also satisfies the ordering requirement for model identifiability. Similarly, for the innovations \( \eta_k,t \sim N(0, \sigma_{\eta_k,t}^2) \) we follow Bhattacharya and Dunson (2011) and Montagna et al. (2012) and let \( \sigma_{\eta_k,t}^2 = \sigma_{\eta_k}^2 / \xi_{\eta_k,t} \) with \( \sigma_{\eta_k}^{-2} = \prod_{\ell \leq k} \delta_{\eta_k^\ell} \), \( \delta_{\eta_k^1} \sim \text{Gamma}(a_{\eta_1}, 1) \) and \( \delta_{\eta_k^\ell} \sim \text{Gamma}(a_{\eta_2}, 1) \) for \( \ell > 1 \), and \( \xi_{\eta_k,t} \sim \text{Gamma}(\nu_\eta/2, \nu_\eta/2) \). We allow the data to determine the rate of ordered shrinkage separately for \( \{\mu_k\} \) and \( \{\eta_k,t\} \) using the hyperpriors \( a_{\mu_1}, a_{\mu_2}, a_{\eta_1}, a_{\eta_2} \sim \text{Gamma}(2, 1) \). Finally, the hyperprior \( \nu_\eta \sim \text{Uniform}(2, 128) \) for the degrees of freedom parameter incorporates the possibility of heavy tails in the marginal distribution for \( \eta_k,t \).

5 Properties of the Model

The interactions among the functional response, the scalar predictors, and the model dynamics are illuminated by examining the implied contemporaneous and autocovariance functions of the functional data. Let \( \mathbf{f}(\tau) = (f_1(\tau), \ldots, f_K(\tau)) \) and \( \mathbf{\beta}_t = (\beta_{1,t}, \ldots, \beta_{K,t}) \).

Proposition 2. The contemporaneous covariance function of \( Y_t(\tau) \) in (1), conditional on \( \{f_k\} \), is \( \text{Cov}[Y_t(\tau), Y_t(\mathbf{u})] = f'(\tau) \text{Cov}(\mathbf{\beta}_t) f(\mathbf{u}) + \mathbb{I}\{\tau = \mathbf{u}\} \sigma_{\epsilon_t}^2 \), which is full rank.

The contemporaneous covariance function of \( Y_t(\cdot) \) is full rank, despite the \( K \) term summation in (1). The reduced-rank interpretation of the model derives from the lag-\( \ell \) autoco-
Proposition 3. The lag-\(\ell\) autocovariance function of \(Y_t(\tau)\) in (1), conditional on \(\{f_k\}\), is

\[
\text{Cov}[Y_t(\tau), Y_{t-\ell}(u)] = f'(\tau)\text{Cov}(\beta_t, \beta_{t-\ell}) f(u),
\]

which is rank \(K\).

Proposition 3 demonstrates that the lag-\(\ell\) autocovariance of the functional time series \(Y_t\) is driven by the lag-\(\ell\) autocovariance of the \(K\) factors, \(\{\beta_{k,t}\}\). Autocovariance among the factors may be induced by dynamic predictors \(\{x_t\}\) or modeled directly, as in (2). Note that both the contemporaneous covariance function in Proposition 2 and the autocovariance function in Proposition 3 may be time-varying.

As an illustrative example, consider the special case of the FOSR-AR model:

Proposition 4. For the likelihood (1) and the regression model \(\beta_{k,t} = \mu_k + \sum_{j=1}^p x_{j,t} \alpha_{j,k} + \gamma_{k,t}\) with stationary autoregressive errors \(\gamma_{k,t} = \phi_k \gamma_{k,t-1} + \eta_{k,t}\) and distributions \(\mu_k \overset{\text{indep}}{\sim} N(0, \sigma_{\mu_k}^2), \alpha_{j,k} \overset{\text{indep}}{\sim} N(0, \sigma_{\omega_{j,k}}^2), \text{and } \eta_{k,t} \overset{\text{indep}}{\sim} N(0, \sigma_{\eta_k}^2)\), the lag-\(\ell\) autocovariance function of \(Y_t(\tau)\) is

\[
\text{Cov}[Y_t(\tau), Y_{t-\ell}(u)] = \sum_{k=1}^K \left[ \sigma_{\mu_k}^2 f_k(\tau) f_k(u) \right] + \sum_{j=1}^p \left\{ x_{j,t} x_{j,t-\ell} \sum_{k=1}^K \left[ \sigma_{\omega_{j,k}}^2 f_k(\tau) f_k(u) \right] \right\} + \sum_{k=1}^K \left[ \phi_k^\ell \left( \frac{\sigma_{\eta_k}^2}{1-\phi_k^2} \right) f_k(\tau) f_k(u) \right].
\]

Proposition 4 decomposes the lag-\(\ell\) autocovariance of \(Y_t(\cdot)\) into three terms: a common mean term, a dynamic predictor term, and an autoregressive term. For fixed lag \(\ell\), the relative contributions of each term are determined by the variance components \(\sigma_{\mu_k}^2, \sigma_{\omega_{j,k}}^2, \text{ and } \sigma_{\eta_k}^2\), which are endowed with (shrinkage) priors in Section 4. Naturally, the common mean term is invariant to the lag \(\ell\), while the dynamic predictor term depends on \(x_{j,t} x_{j,t-\ell}\) and the autoregressive term depends on \(\phi_k^\ell\). Due to the assumed stationarity of \(\gamma_{k,t}\), the autocovariance is time-varying only via the dynamic predictors \(x_{j,t} x_{j,t-\ell}\). Note that Proposition 4 remains valid without the normality assumptions.

The reduced-rank model (1), while presented in a fully Bayesian framework, is nonetheless related to FPCA. By removing the measurement error variance and assuming a simpler model, we obtain an equivalence with a non-dynamic FPCA:
Theorem 1. Under (1) with $\sigma_{t} = \sigma_{\epsilon}$, $\beta_{k,t} \overset{\text{indep}}{\sim} N(0, \sigma_{\epsilon}^{2})$, and diffuse priors for $\{\sigma_{k}^{2}\}$ and $\{f_{k}\}$, the posterior mode of $(\{f_{k}\}, \{\beta_{k,t}\})$ is the FPC solution as $\sigma_{\epsilon} \to 0$.

The proof is given in the Appendix. The loading curves $\{f_{k}\}$ correspond to the FPCs and the factors $\{\beta_{k,t}\}$ correspond to the FPC scores.

While this comparison is useful for placing the proposed methodology in the context of existing FPCA procedures, we are ultimately not interested in these restrictive assumptions, or in computing posterior modes. In particular, our methodology offers a much more general class of models for the factors, including time dependence, predictors, and additional shrinkage, and incorporates smoothness priors for the loading curves. Importantly, we build these generalizations into a fully Bayesian framework, which provides simultaneous estimation of all model parameters and uncertainty quantification via the posterior distribution.

6 Simulations

6.1 Simulation Design

We conducted an extensive simulation study in order to (i) characterize the performance of the proposed methods relative to state-of-the-art alternatives for functional regression and (ii) assess the relative importance of our modeling choices, including the model for the loading curves in (1), the time-varying parameter regression in (2)-(3), and the shrinkage priors in Section 4. We consider simulation designs with dynamic and non-dynamic regression coefficients and different sample sizes: a small sample with $T = 50$ time points and $M = 20$ observation points, and a large sample with $T = 200$ and $M = 100$.

We incorporate two sources of sparsity in the regression: (i) some predictors are not associated with the functional response $Y_{t}(\tau)$ and (ii) some predictors are associated with $Y_{t}(\tau)$ exclusively via a small number of factors. We fix $p_{0} = 10$ regression coefficients to 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). For each nonzero predictor \( j = 1, \ldots, p_1 = 5 \), we uniformly sample \( p_j^* \) factors to be nonzero, where \( p_j^* \overset{iid}{\sim} \text{Poisson}(1) \) truncated to \([1, K^*] \). For dynamic regression coefficients, we simulate the nonzero factors \( k \) for predictor \( j \) 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 \( Z_{k,t} \overset{iid}{\sim} \mathcal{N}(0, 1/k^2) \) and \( I_{k,t} \overset{iid}{\sim} \text{Bernoulli}(0.01) \), which results in time-varying yet locally constant regression coefficients \( \alpha_{j,k,t}^* \). For non-dynamic regression coefficients, we simulate \( \alpha_{j,k}^* \overset{iid}{\sim} \mathcal{N}(0, 1/k^2) \). For all cases, the predictors are simulated from \( x_{j,t} \overset{iid}{\sim} \mathcal{N}(0, 1) \), and the intercepts are fixed at \( \mu_k^* = 1/k \). Finally, the autoregressive errors are \( \gamma_{k,t}^* = 0 \) and \( \eta_{k,t}^* \overset{iid}{\sim} \mathcal{N}(0, [1 - 0.8^2]/k^2) \), which are highly correlated yet stationary with marginal standard deviation \( 1/k \).

For \( M \) equally-spaced points \( \tau \in [0, 1] \), the true loading curves are \( f_{k}^*(\tau) = 1/\sqrt{M} \) and for \( k = 2, \ldots, K^* = 4 \), \( f_k^* \) is an orthogonal polynomial of degree \( k \). Therefore, the model orthonormality constraint \( F'F = I_K \) is imposed for the true loading curves, \( f_k^* \). 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 \( f_k^*(\tau) \), the true curves are \( Y_t^*(\tau) = \sum_{k=1}^{K^*} f_k^*(\tau) \beta_{k,t}^* \) and the functional data are simulated from \( Y_t(\tau) = Y_t^*(\tau) + \sigma^* \epsilon_t^*(\tau) \), where \( \epsilon_t^*(\tau) \overset{iid}{\sim} \mathcal{N}(0, 1) \). After selecting a root-signal-to-noise ratio (RNSR), the observation error standard deviation is \( \sigma^* = \sqrt{\frac{\sum_{t=1}^{T} \sum_{j=1}^{M} (Y_t^*(\tau_j) - \bar{Y}^*)^2}{TM - 1}} / \text{RNSR} \) where \( \bar{Y}^* \) is the sample mean of \( \{Y_t^*(\tau_j)\}_{j,t} \). We select RNSR = 5, which produces moderately noisy functional data.

### 6.2 Methods For Comparison

We consider two variations of the proposed methodology: the DFOSR model (1)-(3) (DFOSR-HS) and the non-dynamic analog (FOSR-AR) with \( \alpha_{j,k,t} = \alpha_{j,k} \), in both cases using \( K = 6 > K^* = 4 \) to include more factors than necessary. We consider an alternative DFOSR model with normal-inverse-gamma innovations (DFOSR-NIG), i.e., we replace the horseshoe priors in (10) with \( \sigma_{\omega_{j,k}}^{-2} \overset{iid}{\sim} \text{Gamma}(0.001, 0.001) \). Originally proposed by Kowal et al. (2017a), this model does not provide aggressive shrinkage with respect to time \( t \), predictor \( j \), or factor \( k \), but otherwise retains the proposed DFOSR model characteristics. Next, to study the impor-
tance of estimating the loading curves \( f_k \), we implement a variation of DFOSR-NIG in which the loading curves \( f_k \) are estimated \emph{a priori} as functional principal components using Xiao et al. (2013), where \( K \) is selected to explain 99% of the variability in \( \{Y^*_j(\tau_j)\}_{j,t} \). For this method (Dyn-FPCA), we remove the ordered shrinkage by specifying \( \mu_k \overset{\text{iid}}{\sim} N(0, 100^2) \) and normal-inverse-gamma priors for \( \eta_{k,t} \) in (2) and \( \omega_{j,k,t} \) in (3). Among existing FOSR methods, we include Reiss et al. (2010), which is a FOSR estimated using least squares (FOSR-LS), and Barber et al. (2017), which is a FOSR with a group lasso penalty on each regression function (FOSR-Lasso), both implemented using the \texttt{refund} package in \texttt{R} (Goldsmith et al., 2016). These methods are non-Bayesian, and do not account for time-varying regression coefficients or autocorrelated errors.

6.3 Simulation Results

We compare methods using root mean squared errors of the dynamic regression coefficient functions, \( \text{RMSE} = \sqrt{\frac{1}{pTM} \sum_{j=1}^{p} \sum_{t=1}^{T} \sum_{\ell=1}^{M} (\hat{\alpha}_{j,t}(\tau_\ell) - \tilde{\alpha}^*_{j,t}(\tau_\ell))^2} \), where \( \hat{\alpha}_{j,t}(\tau_\ell) \) is the estimated regression coefficient for predictor \( j \) at time \( t \) and observation point \( \tau_\ell \) and \( \tilde{\alpha}^*_{j,t}(\tau_\ell) = \sum_{k=1}^{K^*} f_k(\tau_\ell) \alpha^*_{j,k,t} \) is the true regression coefficient. For the Bayesian methods, we use the posterior expectation of \( \hat{\alpha}_{j,t}(\tau_\ell) \) as our estimator. The RMSEs for the regression coefficients based on 50 simulations are in Figure 1.

In all cases, the proposed DFOSR-HS model performs better than existing methods, typically by a wide margin. Among time-varying parameter models, DFOSR-HS offers substantial improvements over DFOSR-NIG and Dyn-FPCA, which suggests that the shrinkage priors of Section 4 are an important component of the DFOSR model. DFOSR-NIG is uniformly better than Dyn-FPCA, which demonstrates that our model for the loading curves \( f_k \) in Section 3 improves upon an FPCA-based approach. For the dynamic simulations, the comparative performance of these methods depends on the sample size: when \( T = 200 \) and \( M = 100 \), the time-varying parameter regression models (DFOSR-HS, DFOSR-NIG, and Dyn-FPCA) are clearly preferable, but when \( T = 50 \) and \( M = 20 \), only the proposed
Figure 1: Root mean squared errors for the regression coefficient functions $\hat{\alpha}_{j,t}(\tau)$ under different simulation designs: the dynamic case (top row) and the non-dynamic case (bottom row) for large (left column) and small (right column) sample sizes. The proposed methods (DFOSR-HS and FOSR-AR) are marked with an asterisk and colored in light blue; simplifications of the proposed methods are in dark blue; and existing FOSR methods are in red.

DFOSR-HS performs well among dynamic models, and the proposed (non-dynamic) FOSR-AR performs best overall. For the non-dynamic simulations, the FOSR-AR performs best, followed by the DFOSR-HS, for both sample sizes. These results validate our DFOSR modeling choices for (i) the loading curves $f_k$ (Section 3), (ii) the shrinkage priors (Section 4), and (iii) the time-varying parameter model (3), particularly when $T$ is sufficiently large.

Additional simulation results comparing mean credible interval widths for time-varying parameter regression models (DFOSR-HS, DFOSR-NIG, and Dyn-FPCA) are in the Appendix. Notably, DFOSR-HS obtains substantially narrower credible intervals without sacrificing nominal coverage, which suggests greater power to detect functional associations.
7 Age-Specific Fertility Rates in South and Southeast Asia

We analyze age-specific fertility rates (ASFRs) for developing nations in South and Southeast Asia. Fertility is an important determinant of the health and welfare of women, their families, and their communities, and is a key factor in global and national population growth. Fertility rates may vary greatly between developed and less developed nations, and may depend on socioeconomic and demographic factors such as age, education, employment, marital status, and access to family planning. While it is common for studies to use total fertility rates, which aggregate over all age groups, important patterns and trends in the fertility rate may only be discoverable using *age-specific* fertility rates. The ASFR measures the annual number of births to women within a specific age group per 1000 women in that age group. Notably, equivalent total fertility rates may be attained using vastly different distributions of fertility among age groups (see Pantazis and Clark 2018, Fig. 2). Naturally, the distribution of fertility among age groups is a fundamental determinant of future fertility rates and population sizes. Therefore, it is appropriate to model the ASFR as a functional time series: the fertility rate is a *function* of age, and varies over *time* (year).

A particular challenge in modeling ASFRs for developing nations is the sparsity of survey data. The Demographic and Health Surveys (DHS) of the United States Agency for International Development (USAID) aggregates available survey data, which may be accessed via STATcompiler (Casterline and Lazarus 2010). We consider DHS survey data from 1994-2016 for 12 nations in South and Southeast Asia: Afghanistan, Bangladesh, Cambodia, India, Indonesia, Maldives, Myanmar, Nepal, Pakistan, Philippines, Timor-Leste, and Vietnam. During this time period, four nations only have one available survey, and there are at most two surveys available each year; for years with two surveys, we use the average ASFRs. For each survey, the reported ASFR is the ASFR over the three years preceding the survey.

Fertility rates are readily available only for a small number of age groups: the DHS survey data provides ASFRs for the seven age groups 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49.
For modeling purposes, we use the midpoints of each age group, so the observation points are $\tau_j \in T_{\text{obs}} \equiv \{17, 22, 27, 32, 37, 42, 47\}$. However, we are interested in the age-specific fertility rates over the entire domain, $T = [15, 49]$. Within the DFOSR framework of (1), we propose a model-based imputation approach to obtain estimates and inference for $M = 31$ ages within the range of observed values: $\tau = 17, \ldots, 47$. In the Gibbs sampler, we draw $[Y_t(\tau^*) | \{f_k\}, \{\beta_{k,t}\}, \{\sigma_{\epsilon_t}\}] \sim \text{N}(\sum_k f_k(\tau^*)\beta_{k,t}, \sigma_{\epsilon_t}^2)$ for each unobserved $\tau^* \not\in T_{\text{obs}}$. As a result, we obtain (i) model-based interpolated fertility rate curves with posterior credible bands and (ii) inference for regression functions over a denser grid of points.

In addition to the dynamic and functional aspects of ASFR data, we are interested in modeling the association between age-specific fertility and important socioeconomic and demographic predictor variables. In particular, we include the following predictor variables for each year $t$, provided by DHS and accessed via STATcompiler: (i) the percentage of currently married or in union women currently using any method of contraception, (ii) the median age of first marriage or union in years among women (age 25-49), (iii) the percentage of women with secondary or higher education, and (iv) the percentage of currently married or in union women employed in the 12 months preceding the survey. For each predictor variable of interest, the proposed DFOSR framework allows us to estimate and perform inference on an accompanying regression function, which may be time-varying. Therefore, our methodology provides a mechanism for understanding how each predictor impacts the shape of the ASFR, which allows for differential effects on different age groups.

Using the MCMC algorithm of Section 9 we sample from the posterior distribution of the FOSR-AR model: (1)-(3) with $\alpha_{j,k,t} = \alpha_{j,k}$, the shrinkage priors of Section 4 and the LR-TPS loading curve model of Section 3. We set $\sigma_{\epsilon_t} = \sigma_\epsilon$ with a Jeffreys’ prior $[\sigma_\epsilon^2] \propto 1/\sigma_\epsilon^2$ and impose stationarity via the AR coefficient priors $[(\phi_k + 1)/2] \sim \text{Beta}(a_\phi, b_\phi)$ with $a_\phi = 5$ and $b_\phi = 2$. The time-varying parameter DFOSR produced similar results (the simulations of Section 6 suggest that, even when the true model is a DFOSR, the non-dynamic parameter model FOSR-AR may be preferable for small sample sizes $T \leq 50$). We report results for
$K = 3$; larger values of $K$ produce nearly identical results. The MCMC is highly efficient: the computation time for 25000 iterations of the Gibbs sampling algorithm (with $T = 20$, $M = 31$, and $p = 6$), implemented in R (on a MacBook Pro, 2.7 GHz Intel Core i5), is less than 3 minutes. We discard the first 10000 simulations as a burn-in and retain every 3rd sample. Traceplots indicate good mixing and suggest convergence (see the Appendix).

In Figure 2, we plot the ASFRs with the model-imputed ASFR curves $\hat{Y}_t(\tau) = \sum_{k=1}^{K} f_k(\tau)\beta_{k,t}$ and the loading curves $f_k(\tau)$ for $\tau = 17, \ldots, 47$ with 95% simultaneous credible bands (Ruppert et al., 2003). The fitted ASFR curves $\hat{Y}_t$ demonstrate an overall decrease in the fertility rate from 2000 to 2016, but this effect is not uniform: the largest decrease occurs for ages 27-37, while the fertility for ages less than 20 actually increased. Importantly, the 95% simultaneous credible bands for $\hat{Y}_t$ do not overlap, which confirms that these ASFR curves have indeed changed over time. The loading curves are smooth and describe the dominant modes of variability in the ASFRs. Much of the variability in the $\{f_k\}$ occurs between the ages of 20-40, which further supports the use of age-specific, rather than total, fertility rates.

**Figure 2:** (Left) Age-specific fertility rates for South and Southeast Asia in 2000 and 2016. For each year $t$, the solid lines are the posterior means of $\hat{Y}_t(\tau) = \sum_{k=1}^{K} f_k(\tau)\beta_{k,t}$ and the gray bands are 95% simultaneous credible bands for $\hat{Y}_t(\tau)$, where $\tau = 17, \ldots, 47$ years of age. (Right) Estimated loading curves $f_k$. For each curve $f_k(\tau)$, the solid line is the posterior mean, the light gray bands are 95% pointwise highest posterior density intervals, and the dark gray bands are 95% simultaneous credible bands.
In Figure 3, we plot the (static) regression functions $\tilde{\alpha}_j(\tau) = \sum_{k=1}^K f_k(\tau)\alpha_{j,k}$ for each predictor $j = 1, \ldots, p$, which may be interpreted via model \[4\]. The 95% simultaneous credible bands exclude zero for both (i) the percentage of currently married or in union women currently using any method of contraception and (ii) the median age of first marriage or union in years among women (age 25-49), which indicates that these variables are highly important for ASFRs. The U-shaped $\tilde{\alpha}_1(\tau)$ in Figure 3 suggests that a greater percentage of married women with access to contraceptives corresponds to a decline in the expected fertility rate, specifically among women aged 22-45. The S-shaped $\tilde{\alpha}_2(\tau)$ in Figure 3 suggests that a larger median age of first marriage corresponds to a decrease in the expected fertility rate among women aged 17-23 and an increase in the expected fertility rate among women aged 30-40. Importantly, these results are age-specific: the association between each predictor and the fertility rate varies by age, while the smoothness of loading curves $f_k$ implies that similar ages should have similar associations.

Figure 3: Estimated regression function for the percentage of married women using contraceptives (left), the median age of first marriage among women (left center), the percentage of women with secondary or higher education (right center), and the percentage of married women employed in the 12 months preceding the survey (right). For each (static) regression function $\tilde{\alpha}_j(\tau) = \sum_{k=1}^K f_k(\tau)\alpha_{j,k}$, the solid line is the posterior mean, the light gray bands are 95% pointwise highest posterior density intervals, and the dark gray bands are 95% simultaneous credible bands.

8 Macroeconomy and the Yield Curve

The yield curves describes the time-varying term structure of interest rates: at each time $t$, the yield curve $Y_t(\tau)$ characterizes how interest rates vary over the length of the bor-
rowing period, or maturity, $\tau$. 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). Due to these fundamental economic connections, we are interested in modeling the associations between the yield curve and key macroeconomic variables, namely, real activity, monetary policy, and inflation. Importantly, the DFOSR modeling framework allows us to associate these variables with particular maturities $\tau$ along the yield curve, and to study how the associations may change over time.

Dynamic yield curve models commonly adopt the Nelson-Siegel parameterization (Nelson and Siegel 1987), usually within a state space framework (e.g., Diebold and Li 2006; Diebold et al., 2006; Koopman et al., 2010). These parametric approaches are less flexible and introduce unnecessary bias in estimation and forecasting, and often require solving computationally intensive nonlinear optimization problems. Existing nonparametric methods include Hays et al. (2012) and Jungbacker et al. (2013), but these approaches do not provide the uncertainty quantification, time-varying parameter regression, and shrinkage capabilities of the proposed DFOSR model.

We obtain zero-coupon yield curve data from Gürkaynak et al. (2007), which are pre-smoothed using Svensson (1994) for $M = 30$ maturities $\tau_j \in T_{\text{obs}} \equiv \{1, \ldots, 30\}$ years. The macroeconomic predictors are manufacturing capacity utilization (CU) for real activity, the federal funds rate (FFR) for monetary policy, and annualized price inflation (PCE) for inflation, which are available from the FRED database. We compute monthly averages of the yield curve data for common frequency with the macroeconomic variables, and consider the time period from January 1986 to February 2018 ($T = 386$). We sample from the posterior distribution of the DFOSR model (1)-(3) using the MCMC algorithm of Section 9. To incorporate volatility clustering, we include a stochastic volatility model for $\sigma_{t}^{2}$, which is an important component in many financial and economic applications (see the Appendix for details and a supporting figure). We report results for $K = 6$, but larger values of $K$
produce nearly identical results. We ran the MCMC for 25000 iterations, discarded the first 10000 simulations as a burn-in, and retained every 3rd sample.

In Figure 4, we plot the posterior expectation of the dynamic regression functions $\tilde{\alpha}_{j,t}(\tau) = \sum_{k=1}^{K} f_k(\tau) \alpha_{j,k,t}$ for CU, FFR, and PCE. During the late 1980s and 1990s, CU appears to impact the curvature of the yield curve, with a prominent hump for maturities around 10 years, but this effect dissipates during the 2000s. FFR has the largest estimated effect, almost entirely for small maturities, which impacts the slope of the yield curve. Notably, the FFR effect is mostly time-invariant during this period (1986-2018). PCE has a moderate impact on the slope of the yield curve—in the opposite direction of FFR—but only until the 1990s. The results in Figure 4 demonstrate that these macroeconomic associations with the yield curve are maturity-specific (functional) and may change over time (dynamic), which confirms the utility of the DFOSR model (1)-(3).

![Figure 4: Posterior expectation of the time-varying regression coefficient functions $\tilde{\alpha}_{j,t}(\tau)$ for capacity utilization (CU, left), federal funds rate (FFR, center), and personal consumption expenditures (PCE, right). The FFR has the largest estimated effect, particularly for smaller maturities. The impact of CU and PCE has declined substantially since the late 1980s.]

9 MCMC Sampling Algorithm

We develop an efficient Gibbs sampling algorithm for model (1)-(3) based on four essential components: (i) the loading curve sampler for $\{f_k\}$ with the identifiability constraint $F'F = \ldots$
\( I_K \) enforced; (ii) the projection-based simplification of the likelihood (1) from Lemma 1; (iii) a state space simulation smoother for the dynamic regression parameters in (2) and (3); and (iv) parameter expansion techniques for the variance components in (1), (2), and (3). For sparsely observed functional data, in which the functional data \( Y_t \) are not observed at the same observation points \( \tau_1, \ldots, \tau_M \) for all times \( t \), we include a sampling-based imputation step as in Section 7. Since components (i) and (ii) are discussed in Section 3 and component (iv) uses standard techniques for Bayesian shrinkage, we focus on (iii) here. The details of the full Gibbs sampling algorithm are provided in the Appendix.

Using Lemma 1, we project the functional data \( Y_t \) on the loading curves \( f_k \) to obtain the working likelihood (9) in Lemma 2. The projection eliminates dependence in the likelihood on the number of observation points, \( M \), which may be large, and the location of the observation points, \( \tau \). Combining the dynamic terms from (2)-(3) into state variables with likelihood (9), we have

\[
\tilde{Y}_{k,t} = \mu_k + \begin{pmatrix} x_t' & 1 \end{pmatrix} \begin{pmatrix} \alpha_{k,t} \\ \gamma_{k,t} \end{pmatrix} + \tilde{\epsilon}_{k,t}
\]

\[
\begin{pmatrix} \alpha_{k,t} \\ \gamma_{k,t} \end{pmatrix} = \begin{pmatrix} I_p & 0 \\ 0 & \phi_k \end{pmatrix} \begin{pmatrix} \alpha_{k,t-1} \\ \gamma_{k,t-1} \end{pmatrix} + \begin{pmatrix} \omega_{k,t} \\ \eta_{k,t} \end{pmatrix}
\]

where \( \alpha_{k,t} = (\alpha_{1,k,t}, \ldots, \alpha_{p,k,t})' \). The resulting model is a dynamic linear model [West and Harrison, 1997] in the state variables \( (\alpha_{k,t}', \gamma_{k,t}') \), and therefore the parameters \( \{\alpha_{k,t}, \gamma_{k,t}\}_{t=1}^T \) may be sampled jointly across all \( t = 1, \ldots, T \) using efficient state space simulation methods, such as Durbin and Koopman (2002). These samplers are also valid for FOSR-AR with \( \alpha_{j,k,t} = \alpha_{j,k} \). Note that the model (11)-(12) 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 \( \{\alpha_{k,t}\} \) and all autoregressive regression error terms \( \{\gamma_{k,t}\} \) jointly has
computational complexity $O(KTp^3)$. For small to moderate number of predictors $p < 30$, the algorithm is efficient; for sufficiently small $K$, the sampler is nearly computationally equivalent to the analogous non-functional time-varying parameter regression model.

In addition to the loading curve sampler for $\{f_k\}$ in Section 3 and the state space simulation sampler for $\{\alpha_{k,t}, \gamma_{k,t}\}_{t=1}^T$ via (11)-(12), the Gibbs sampler proceeds by iteratively sampling the intercepts $\{\mu_k\}$, the autoregressive coefficients $\{\phi_k\}$, and the variance components $\sigma_{\epsilon_t}^2$, $\sigma_{\eta_{k,t}}^2$, and $\sigma_{\omega_{j,k,t}}^2$—as well as any relevant hyperparameters—from their full conditional distributions (see the Appendix). Posterior inference is therefore available for these quantities as well as the time-varying parameter regression functions $\tilde{\alpha}_j(\tau) = \sum_{k=1}^K f_k(\tau)\alpha_{j,k}$ from Proposition 1 and the fitted curves $\hat{Y}_t(\tau) = \sum_{k=1}^K f_k(\tau)\beta_{k,t}$ with $\beta_{k,t}$ defined in (2).

10 Discussion and Future Work

The proposed dynamic function-on-scalars regression model provides a fully Bayesian framework for simultaneously modeling functional dependence, time dependence, and dynamic predictors. We incorporate a nonparametric model for functional dependence, an autoregressive model for time-dependence, and a time-varying parameter regression model for dynamic predictors. The model is highly flexible, yet incorporates appropriate shrinkage and smoothness priors to guard against overfitting. A simulation study validates our model for the loading curves $f_k$ (Section 3) and our choice of shrinkage priors (Section 4) by demonstrating substantial improvements in estimation accuracy relative to existing methods as well as simpler submodels. Applications in age-specific fertility rates and yield curves illustrate the utility of our approach: in particular, we provide estimation, uncertainty quantification, and imputation for regression coefficient functions, which may be time-varying.

Future work will extend model (1) for other important dependence structures, such as dynamic functional predictors $X_{j,t}(u)$ for $u \in \mathcal{U}$, possibly with different domains $\mathcal{U} \neq \mathcal{T}$. Notably, our highly efficient projection-based Gibbs sampler only requires the likelihood (1).
and the identifiability constraint $F'F = I_K$ to obtain the working likelihood \([9]\). Therefore, it is straightforward to combine our nonparametric model for the loading curves $f_k$—and the accompanying uncertainty quantification—with alternative models for $\beta_{k,t}$ in \([2]-[3]\), while maintaining computational scalability. Lastly, given the success of time-varying parameter regression in forecasting scalar time series \cite{Dangl2012, Korobilis2013, Belmonte2014}, the proposed DFOSR model and its extensions offer a promising approach for forecasting functional time series data.

A MCMC Algorithm

The dynamic function-on-scalars regression model (DFOSR), with all prior distributions, is

\begin{align*}
Y_t(\tau) &= \sum_{k=1}^{K} f_k(\tau) \beta_{k,t} + \epsilon_t(\tau), \quad \epsilon_t(\tau) \overset{\text{iid}}{\sim} N(0, \sigma_\epsilon^2), \quad \sigma_\epsilon^2 \propto 1/\sigma_\epsilon^2 \\
f_k(\tau) &= b'(\tau) \psi_k, \quad \psi_k \overset{\text{iid}}{\sim} N\left(\mathbf{0}, \lambda^{-1}_{f_k} \Omega^{-1}\right), \quad \lambda^{-1/2}_{f_k} \sim \text{Uniform}(0, 10^4) \\
\beta_{k,t} &= \mu_k + \sum_{j=1}^{p} x_{j,t} \alpha_{j,k,t} + \gamma_{k,t}, \quad \gamma_{k,t} = \phi_k \gamma_{k,t-1} + \eta_{k,t}, \quad \eta_{k,t} \overset{\text{iid}}{\sim} N(0, \sigma_{\eta,k,t}^2) \\
\mu_k &\overset{\text{iid}}{\sim} N(0, \sigma_{\mu,k}^2), \quad [(\phi_k + 1)/2] \overset{\text{iid}}{\sim} \text{Beta}(5, 2) \\
\sigma_{\mu,k}^{-2} &= \prod_{\ell \leq k} \delta_{\mu,\ell}, \quad \delta_{\mu,1} \sim \text{Gamma}(a_{\mu,1}, 1), \quad \delta_{\mu,\ell} \sim \text{Gamma}(a_{\mu,2}, 1), \quad \ell > 1 \\
\sigma_{\eta,k,t}^{-2} &= \sigma_{\eta,k}^{-2}/\xi_{\eta,k,t}, \quad \xi_{\eta,k,t} \overset{\text{iid}}{\sim} \text{Gamma}(\nu_\eta/2, \nu_\eta/2), \quad \nu_\eta \sim \text{Unif}(2, 128) \\
\sigma_{\eta,k}^{-2} &= \prod_{\ell \leq k} \delta_{\eta,\ell}, \quad \delta_{\eta,1} \sim \text{Gamma}(a_{\eta,1}, 1), \quad \delta_{\eta,\ell} \sim \text{Gamma}(a_{\eta,2}, 1), \quad \ell > 1 \\
a_{\mu,1}, a_{\mu,2}, a_{\eta,1}, a_{\eta,2} &\overset{\text{iid}}{\sim} \text{Gamma}(2, 1) \\
\alpha_{j,k,t} &= \alpha_{j,k,t-1} + \omega_{j,k,t}, \quad \omega_{j,k,t} \overset{\text{iid}}{\sim} N(0, \sigma_{\omega,j,k,t}^2) \\
\sigma_{\omega,j,k,t}^{-1} &\overset{\text{iid}}{\sim} C^+(0, \lambda_j), \quad \lambda_j \sim C^+(0, \lambda_0), \quad \lambda_0 \sim C^+(0, 1/\sqrt{T-1}) \\
\eta_{k,0} &\overset{\text{iid}}{\sim} t_3(0, 1), \quad \omega_{j,k,0} \overset{\text{iid}}{\sim} t_3(0, 1)
\end{align*}
for \( \tau \in T, j = 1, \ldots, p, k = 1, \ldots, K, \) and \( t = 1, \ldots, T. \) The details for each level are described in the main paper. Note that \( \Omega \) in (14) may not be invertible, but for (low rank) thin plate splines the posterior distribution of \( \psi_k \) will be proper.

We construct a Gibbs sampling algorithm that primarily features draws from known full conditional distributions with a small number of slice sampling steps (Neal, 2003). For the half-Cauchy and t-distributions in (22) and (23), respectively, we use the following scale mixture of Gaussian parameter expansions. The hierarchy of half-Cauchy distributions may be written on the precision scale with Gamma expansions:

\[
\sigma_{\omega, j, k, t}^{-2} | \xi_{\omega, j, k, t} \sim \text{Gamma}(1/2, \xi_{\omega, j, k, t}), \\
[\xi_{\omega, j, k, t} | \xi_{\sigma, j, k, t} \sim \text{Gamma}(1/2, \lambda_j), \\
[\lambda_j | \lambda_0 \sim \text{Gamma}(1/2, \lambda_0^{-2}), \\
[\lambda_0 | \lambda_0 \sim \text{Gamma}(1/2, \lambda_0^{-2}), \\
[\xi_{\lambda_0} \sim \text{Gamma}(1/2, \lambda_0), \\
[\xi_{\lambda_0} \sim \text{Gamma}(1/2, T - 1). \]
\]

The t-distributions are expanded as

\[
[\eta_{k, 0} | \xi_{\eta, k, 0} \sim N(0, 1/\xi_{\eta, k, 0}) \text{ and } \xi_{\eta, k, 0} \sim \text{Gamma}(3/2, 3/2) \text{ and similarly, } [\omega_{j, k, 0} | \xi_{\omega, j, k, 0} \sim N(0, 1/\xi_{\omega, j, k, 0}) \text{ and } \xi_{\omega, j, k, 0} \sim \text{Gamma}(3/2, 3/2). \]
\]

In all cases, the full conditional distributions are Gamma (on the precision scale).

Gibbs Sampling Algorithm

1. **Imputation:** for all unobserved \( Y_t(\tau^*_t) \), sample each \( Y_t(\tau^*_t) | \{f_k\}, \{\beta_{k,t}\}, \{\sigma_{\epsilon_t}\} \) indep

\[
N\left(\sum_k f_k(\tau^*_t)\beta_{k,t}, \sigma_{\epsilon_t}^2\right).
\]

2. **Loading curves and smoothing parameters:** for \( k = 1, \ldots, K, \)

   (a) Sample \( [\lambda_{f_k} | \cdots] \sim \text{Gamma}((L_M - D + 1 + 1)/2, \psi_k' \Omega \psi_k/2) \) truncated to \((10^{-8}, \infty)\).

   (b) Sample \( [\psi_k | \cdots] \sim N(Q^{-1}_\psi \ell_{\psi_k}, Q^{-1}_\psi) \text{ conditional on } C_k \psi_k = 0, \text{ where } C_k = (f_1, \ldots, f_{k-1}, f_{k+1}, \ldots, f_K)' \text{ and } B = (\psi_1, \ldots, \psi_{k-1}, \psi_{k+1}, \ldots, \psi_K)' \), using a modified version of the efficient Cholesky decomposition approach of Wand and Ormerod (2008):

      i. Compute the (lower triangular) Cholesky decomposition \( Q_{\psi_k} = Q_L Q'_L \).
ii. Use forward substitution to obtain \(\ell\) as the solution to \(Q_L \ell = \ell_\psi\), then use backward substitution to obtain \(\psi_0^k\) as the solution to \(Q'_L \psi^0_k = \ell + z\), where \(z \sim N(0, I_{LM})\);

iii. Use forward substitution to obtain \(\bar{C}\) as the solution to \(Q_L \bar{C} = C_k\), then use backward substitution to obtain \(\tilde{\bar{C}}\) as the solution to \(Q'_L \tilde{\bar{C}} = \bar{C}\);

iv. Set \(\psi^*_k = \psi_0^k - \tilde{\bar{C}}(C_k \bar{C})^{-1} C_k \psi_0^k\);

v. Retain the vectors \(\psi_k = \psi^*_k / \sqrt{\psi^*_k B' B \psi^*_k} = \psi^*_k / ||\psi^*_k||\) and \(f_k = B \psi_k\) and update \(\beta_{k,t} \leftarrow \beta_{k,t} ||\psi^*_k||\).

3. **Project:** update \(\tilde{Y}_{k,t} = f_k' Y_t = \psi_k' (B' Y_t)\) for all \(k, t\).

4. **Dynamic state variables:** sample \(\{\alpha_{j,k,t}\}, \{\gamma_{k,t}\}|\{\tilde{Y}_{k,t}\}, \cdots\) jointly, including the initial states \(\{\eta_{k,0}\}\) and \(\{\omega_{j,k,0}\}\), using Durbin and Koopman (2002).

   *Note:* we condition on \(\{\mu_k\}\) for computational efficiency (i.e., a smaller state vector), but \(\mu_k\) could be included in this joint sampler.

5. **Unconditional mean and AR coefficients:** for \(k = 1, \ldots, K\),

   (a) Using the centered AR parametrization with \(\gamma^c_{k,t} = \gamma_{k,t} + \mu_k\) (computed with the previous simulated value of \(\mu_k\)), so \(\gamma^c_{k,t} = \mu_k + \phi_k (\gamma^c_{k,t-1} - \mu_k) + \eta_{k,t}\), sample \(\mu_k, \cdots \) \(\sim\) \(N(Q^{-1}_{\mu_k}, \ell^{-1}_{\mu_k})\) where \(Q_{\mu_k} = \sigma_{\mu_k}^{-2} + (1 - \phi_k)^2 \sum_{t=2}^T \sigma_{\eta_{k,t}}^{-2}\) and \(\ell_{\mu_k} = (1 - \phi_k) \sum_{t=2}^T (\gamma^c_{k,t} - \phi_k \gamma^c_{k,t-1}) \sigma_{\eta_{k,t}}^{-2}\).

   (b) Sample \(\phi_k\) using the slice sampler (Neal 2003).

      *Note:* if instead we assume the prior \(\phi_k \iddot Uniform(-1, 1)\), then the full conditional distribution for \(\phi_k\) is available in closed form (a truncated-normal distribution).

6. **Variance parameters:**

   (a) **Observation error variance:** \([\sigma_e^{-2} | \cdots] \sim Gamma \left(\frac{MT}{2}, \frac{1}{2} \sum_{t=1}^T ||Y_t - F \beta_t||^2\right)\)

   (b) **Multiplicative Gamma Process Parameters:** given \(\mu_k\) and \(\eta_{k,t} = \gamma_{k,t} - \phi_k \gamma_{k,t-1}\) for \(\gamma_{k,t} = \gamma^c_{k,t} - \mu_k\) (after sampling \(\mu_k\) above),
i. Sample \([\delta_{\mu_1}| \cdots] \sim \text{Gamma}(a_{\mu_1} + \frac{K}{2}, 1 + \frac{1}{2} \sum_{k=1}^{K} \tau_{\mu_k}^{(1)} \mu_k^2)\) and \([\delta_{\mu_e}| \cdots] \sim \text{Gamma}(a_{\mu_2} + \frac{K-\ell+1}{2}, 1 + \frac{1}{2} \sum_{k=1}^{K} \tau_{\mu_k}^{(\ell)} \mu_k^2)\) for \(\ell > 1\) where \(\tau_{\mu_k}^{(k)} = \prod_{h=1,h \neq k}^{\ell} \delta_{\mu_h}\).

ii. Set \(\sigma_{\mu_k} = \prod_{\ell \leq k} \delta_{\mu_\ell}^{-1/2}\).

iii. Sample \([\delta_{\eta_1}| \cdots] \sim \text{Gamma}(a_{\eta_1} + \frac{K(T-1)}{2}, 1 + \frac{1}{2} \sum_{k=1}^{K} \tau_{\eta_k}^{(1)} \sum_{t=2}^{T} \eta_{k,t}^2 \xi_{\eta_k,t})\) and \([\delta_{\eta_e}| \cdots] \sim \text{Gamma}(a_{\eta_2} + \frac{(K-\ell+1)(T-1)}{2}, 1 + \frac{1}{2} \sum_{k=1}^{K} \tau_{\eta_k}^{(\ell)} \sum_{t=2}^{T} \eta_{k,t}^2 \xi_{\eta_k,t})\) for \(\ell > 1\) where \(\tau_{\eta_k}^{(k)} = \prod_{h=1,h \neq k}^{\ell} \delta_{\eta_h}\).

iv. Set \(\sigma_{\eta_k} = \prod_{\ell \leq k} \delta_{\eta_\ell}^{-1/2}\).

v. Sample \([\xi_{\eta_k,t}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(\nu_{\eta} + \frac{1}{2}, \nu_{\eta} + \frac{\eta_{\eta,t}^2}{2\sigma_{\eta_k}^2})\)

vi. Set \(\sigma_{\eta_k,t} = \sigma_{\eta_k}/\sqrt{\xi_{\eta_k,t}}\).

(c) Hierarchical Half-Cauchy Parameters: for \(\omega_{j,k,t} = \alpha_{j,k,t} - \alpha_{j,k,t-1}\),

i. Sample \([\sigma_{\omega_{j,k,t}}^{-2}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(1, \xi_{\omega_{j,k,t}} + \omega_{j,k,t}^2/2)\) and \([\xi_{\omega_{j,k,t}}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(1, \lambda_{j,k,t}^{-2} + \sigma_{\omega_{j,k,t}}^{-2})\).

ii. Sample \([\lambda_{j,k,t}^{-2}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(\frac{T}{2}, \lambda_{j,k,t} + \sum_t \xi_{\omega_{j,k,t}})\) and \([\xi_{\lambda_{j,k}}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(1, \lambda_{j,k}^{-2} + \lambda_{j,k,t}^{-2})\).

iii. Sample \([\lambda_{j,t}^{-2}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(\frac{K+1}{2}, \lambda_{j} + \sum_{k=1}^{K} \xi_{\lambda_{j,k}})\) and \([\xi_{\lambda_j}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(1, \lambda_{j}^{-2} + \lambda_{j,t}^{-2})\).

iv. Sample \([\lambda_{0,t}^{-2}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(\frac{p+1}{2}, \lambda_{0} + \sum_{j=1}^{p} \xi_{\lambda_j})\) and \([\xi_{\lambda_0}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(1, (T-1) + \lambda_{0,t}^{-2})\).

(d) Parameter-expanded initial values:

i. Sample \([\xi_{\eta_{k,0}}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(\frac{3}{2}, \frac{3}{2} + \frac{\eta_{\eta,0}^2}{2})\).

ii. Sample \([\xi_{\omega_{j,k,0}}| \cdots] \overset{\text{indep}}{\sim} \text{Gamma}(\frac{3}{2}, \frac{3}{2} + \frac{\omega_{j,k,0}^2}{2})\).

7. Hyperparameters: sample \(a_{\mu_1}, a_{\mu_2}, a_{\eta_1}, a_{\eta_2},\) and \(\nu_{\eta}\) independently using the slice sampler [Neal, 2003].

In the yield curve application of Section 8, the Jeffreys prior in (13) is replaced by a stochastic volatility model for the variance \(\sigma_{\ell t}^2\). Specifically, the model is an AR(1) for
the log-variance $h_t = \log \sigma^2_{t,i}$: 

$$h_{t+1} = \mu_h + \phi_h (h_t - \mu_h) + \nu_{ht},$$

where $\mu_h \sim N(-10, 100)$ is the unconditional mean of log-volatility, 

$$[(\phi_h + 1)/2] \sim \text{Beta}(20, 1.5)$$

is the autoregressive parameter, and

$$\nu_{ht} \overset{iid}{\sim} N(0, \sigma^2_{\nu})$$

is the log-volatility innovation with standard deviation $\sigma_{\nu} \sim \text{Uniform}(0, 100)$. Sampling $\{h_t\}$ is a straightforward modification of the algorithm in Kastner and Frühwirth-Schnatter (2014), and conditional on $\{h_t\}$, the parameters $\mu_h, \phi_h,$ and $\sigma_{\nu}$ may be sampled iteratively using standard procedures for Bayesian autoregressive models.

**B Additional Proofs**

Given functional data observations $Y_t = (Y_t(\tau_1), \ldots, Y_t(\tau_M))'$ at observation points $\{\tau_j\}_{j=1}^M$, consider the generalization of the likelihood (7) from the main paper:

$$Y_t = \sum_{k=1}^K f_k \beta_{k,t} + \epsilon_t, \quad \epsilon_t \overset{indep}{\sim} N(0, \Sigma_{\epsilon_t})$$

(24)

where $\Sigma_{\epsilon_t}$ is a general $M \times M$ covariance matrix. In the more general case, we may modify Lemma 1 from the main paper:

**Lemma 3.** Under the identifiability constraint $F'F = I_K$, the joint likelihood in (24) is

$$p(Y_1, \ldots, Y_T | \{f_k, \beta_{k,t}, \Sigma_{\epsilon_t}\}_{k,t}) = c_Y \prod_{t=1}^T |\Sigma_{\epsilon_t}|^{-1/2} \exp \left\{ -\frac{1}{2} \left[ Y_t' \Sigma_\epsilon^{-1} Y_t + \beta_t' \left( F' \Sigma_\epsilon^{-1} F \right) \beta_t - 2 \beta_t' \left( F' \Sigma_\epsilon^{-1} Y_t \right) \right] \right\}$$

(25)

where $c_Y = (2\pi)^{-MT/2}$ is a constant and $\beta_t' = (\beta_{1,t}, \ldots, \beta_{K,t})$.

Analogous to the results in the main paper, Lemma 3 implies the following working likelihood for the factors $\beta_{k,t}$ and associated parameters:

**Lemma 4.** Under the identifiability constraint $F'F = I_K$, the joint likelihood in (24) for
\{\beta_{k,t}\} \text{ is equivalent to the working likelihood implied by}

\[ \tilde{Y}_t = \beta_t + \tilde{\epsilon}_t, \quad \tilde{\epsilon}_t \overset{\text{indep}}{\sim} N(0, Q^{-1}_\beta) \] (26)

up to a constant that does not depend on \( \beta_t \), where \( \tilde{Y}_t = Q^{-1}_\beta \ell_{\beta t} \) for \( Q_{\beta t} = F' \Sigma^{-1} \epsilon \) and \( \ell_{\beta t} = F' \Sigma_{\epsilon t}^{-1} Y_t \).

The most useful case of Lemma 4 is when \( \Sigma_{\epsilon t} \) is diagonal, so that the error covariance function is \( C_{\epsilon_t}(\tau, u) = \text{Cov}(\epsilon_t(\tau), \epsilon_t(u)) = \mathbb{I}(\tau = u) V_{\epsilon_t}(\tau) \) and \( V_{\epsilon_t}(\cdot) \) is the variance function. In this case, computing the inverse \( \Sigma_{\epsilon t}^{-1} \) is efficient, and the projection step to obtain \( \tilde{Y}_t \) only requires the inverse of a \( K \times K \) matrix, \( Q_{\beta t} \). Furthermore, if \( V_{\epsilon_t}(\cdot) = V_{\epsilon}(\cdot) \) is non-dynamic, then computing \( Q^{-1}_{\beta t} = Q^{-1}_\beta \) is a one-time cost per MCMC iteration.

**Theorem 2.** Under (1) with \( \sigma_{\epsilon_t} = \sigma_\epsilon, \beta_{k,t} \overset{\text{indep}}{\sim} N(0, \sigma_k^2) \), and diffuse priors for \( \{\sigma_k^2\} \) and \( \{f_k\} \), the posterior mode of \((\{f_k\}, \{\beta_{k,t}\})\) is the FPC solution as \( \sigma_\epsilon \to 0 \).

**Proof.** Under the conditions of Theorem 2, negative twice the joint full conditional log-posterior for the factors and loading curve coefficients \( \psi_k \) is, up to a constant,

\[
-2 \log p(\{\beta_t\}, \{\psi_k\}|\{Y_t\}, \cdots) = \sigma_\epsilon^{-2} \sum_{t=1}^T \left\{ \|Y_t - B\Psi \beta_t\|^2 + \sigma_\epsilon^2 \sum_{k=1}^K \beta_{k,t}^2/\sigma_k^2 \right\} (27)
\]

where \( \Psi = (\psi_1, \ldots, \psi_K) \). Notably, this expression matches equation (12) in James et al. (2000). James et al. (2000) show that as \( \sigma_\epsilon \to 0 \), the maximum likelihood estimators of \( \Psi \), \( \{\beta_t\} \), \( \{\sigma_k^2\} \), and \( \sigma_\epsilon^2 \) reproduce the functional principal components solution, or equivalently, the classical principal components solution for the coefficients of \( Y_t \) projected on the basis \( B \), i.e., \( (B'B)^{-1} B'Y_t \). The posterior mode therefore produces an equivalent solution under diffuse priors on \( \{f_k\} \) and \( \{\sigma_k^2\} \). \( \square \)
C Additional Simulation Results

Using the same simulation designs as in Section 6, we compare mean credible interval widths (MCIWs) for the time-varying parameter regression models (DFOSR-HS, DFOSR-NIG, and Dyn-FPCA). The MCIWs are defined as

\[
\text{MCIW} = \frac{1}{pTM} \sum_{j=1}^{p} \sum_{t=1}^{T} \sum_{\ell=1}^{M} \left[ \tilde{\alpha}_{j,t}^{(95)}(\tau_{\ell}) - \tilde{\alpha}_{j,t}^{(5)}(\tau_{\ell}) \right]
\]  

(28)

where \(\tilde{\alpha}_{j,t}^{(95)}(\tau_{\ell})\) and \(\tilde{\alpha}_{j,t}^{(5)}(\tau_{\ell})\) are the 95% and 5% quantiles, respectively, of the posterior distribution for \(\tilde{\alpha}_{j,t}(\tau_{\ell})\) for predictor \(j\) at time \(t\) and observation point \(\tau_{\ell}\). For completeness, we report nominal frequentist coverage in Table 2. In each case, the coverage is more conservative than the 90% level, but the results are similar for all methods.

The MCIW results are in Figure 5. The proposed DFOSR-HS obtains substantially narrower credible intervals than competing methods, which suggests greater power to detect functional associations.

| Simulation Design | DFOSR-HS* | DFOSR-NIG | Dyn-FPCA |
|-------------------|-----------|-----------|----------|
| Dynamic: \(T = 200, M = 100\) | 0.97      | 0.98      | 0.99     |
| Dynamic: \(T = 50, M = 20\)   | 0.98      | 0.99      | 0.99     |
| Non-dynamic: \(T = 200, M = 100\) | 0.96      | 0.99      | 0.98     |
| Non-dynamic: \(T = 200, M = 100\) | 0.96      | 0.98      | 0.98     |

Table 2: Nominal frequentist coverage for the dynamic regression coefficient functions \(\alpha_{j,t}(\tau)\) under different simulation designs for each time-varying parameter regression model.

D MCMC Diagnostics

We include MCMC diagnostics for the fertility application (Section 7). Traceplots for \(\hat{Y}_t(\tau) = \sum_{k=1}^{K} f_k(\tau)\beta_{k,t}\) and \(\tilde{\alpha}_j(\tau) = \sum_{k=1}^{K} f_k(\tau)\alpha_{j,k}\) are in Figures 6 and 7, respectively. These traceplots indicate good mixing and suggest convergence.
Figure 5: Mean credible interval widths for the regression coefficient functions $\hat{\alpha}_{j,t}(\tau)$ under different simulation designs: the dynamic case (top row) and the non-dynamic case (bottom row) for large (left column) and small (right column) sample sizes. The proposed method (DFOSR-HS) is marked with an asterisk.

### E Additional Application Details

Figure 8 plots the observation error standard deviation, $\sigma_{\epsilon_t}$, for the yield curve application. To incorporate volatility clustering, we include a stochastic volatility model for $\sigma^2_{\epsilon_t}$, following Kastner and Frühwirth-Schnatter (2014). There is strong evidence that the observation error standard deviation is time-varying. Importantly, the proposed DFOSR model framework can incorporate the stochastic volatility model with minimal modifications.

### References

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.
Figure 6: Traceplots for the model-imputed ASFR curves $\hat{Y}_t(\tau) = \sum_{k=1}^{K} f_k(\tau)\beta_{k,t}$ at various ages $\tau$ for various years $t$ in the fertility application. The traceplots indicate good mixing.

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.

Belmonte, M. A., Koop, G., and Korobilis, D. (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, 33(1):80–94.

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

Bhattacharya, A. and Dunson, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika*, pages 291–306.

Bolder, D., Johnson, G., and Metzler, A. (2004). *An empirical analysis of the Canadian term structure of zero-coupon interest rates*. Bank of Canada.
Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, pages 465–480.

Casterline, J. B. and Lazarus, R. (2010). Determinants and consequences of high fertility: a synopsis of the evidence. *Addressing the Neglected MDG: World Bank Review of Population and High Fertility*, World Bank publications.

Chen, Y., Goldsmith, J., and Ogden, R. T. (2016). Variable selection in function-on-scalar regression. *Stat*, 5(1):88–101.

Crainiceanu, C., Ruppert, D., and Wand, M. P. (2005). Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software*, 14(14):1–24.

Cressie, N. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons.
Figure 8: Observation error standard deviation, $\sigma_{\epsilon_t}$, for the yield curve data. The solid line is the posterior mean, the light gray bands are 95% pointwise highest posterior density intervals, and the dark gray bands are 95% simultaneous credible bands. There is strong evidence that the observation error standard deviation is time-varying.

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

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

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2009). Multilevel functional principal component analysis. *The Annals of Applied Statistics*, 3(1):458.

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

Durante, D. (2017). A note on the multiplicative gamma process. *Statistics & Probability Letters*, 122:198–204.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.

Fan, Z. and Reimherr, M. (2017). High-dimensional adaptive function-on-scalar regression. *Econometrics and statistics*, 1:167–183.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515–534.

Goldsmith, J., Greven, S., and Crainiceanu, C. (2013). Corrected confidence bands for functional data using principal components. *Biometrics*, 69(1):41–51.

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.

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.

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.

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.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.
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.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, 87(3):587–602.

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*.

Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408–423.

Koopman, S. J., Mallee, M. I., and Van der Wel, M. (2010). Analyzing the term structure of interest rates using the dynamic Nelson–Siegel model with time-varying parameters. *Journal of Business & Economic Statistics*, 28(3):329–343.

Korobilis, D. (2013). Hierarchical shrinkage priors for dynamic regressions with many predictors. *International Journal of Forecasting*, 29(1):43–59.

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.

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2017b). Dynamic shrinkage processes. *arXiv preprint arXiv:1707.00763*.

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

Laurini, M. P. (2014). Dynamic functional data analysis with non-parametric state space models. *Journal of Applied Statistics*, 41(1):142–163.
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.

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

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.

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, pages 705–741.

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

Pantazis, A. and Clark, S. J. (2018). A parsimonious characterization of change in global age-specific and total fertility rates. *PloS one*, 13(1):e0190574.

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

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

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).

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):325–338.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Number 12. Cambridge University Press.

Suarez, A. J., Ghosal, S., et al. (2017). Bayesian estimation of principal components for functional data. *Bayesian Analysis*, 12(2):311–333.
Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. Technical report, National Bureau of Economic Research.

Wahba, G. (1990). *Spline models for observational data*, volume 59. SIAM.

Wand, M. and Ormerod, J. (2008). On semiparametric regression with O’Sullivan penalized splines. *Australian & New Zealand Journal of Statistics*, 50(2):179–198.

Wand, M. P., Ormerod, J. T., Padoan, S. A., and Früahrwirth, R. (2011). Mean field variational Bayes for elaborate distributions. *Bayesian Analysis*, 6(4):847–900.

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

Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.

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.

Zhu, H., Brown, P. J., and Morris, J. S. (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association*, 106(495):1167–1179.