arXiv:1108.3262v5 [stat.ME] 21 Feb 2014

Bayesian Inference in Nonparametric Dynamic State-Space Models

Anurag Ghosh, Soumalya Mukhopadhyay, Sandipan Roy and Sourabh Bhattacharya*

Abstract

We introduce state-space models where the functionals of the observational and the evolutionary equations are unknown, and treated as random functions evolving with time. Thus, our model is nonparametric and generalizes the traditional parametric state-space models. This random function approach also frees us from the restrictive assumption that the functional forms, although time-dependent, are of fixed forms. The traditional approach of assuming known, parametric functional forms is questionable, particularly in state-space models, since the validation of the assumptions require data on both the observed time series and the latent states; however, data on the latter are not available in state-space models.

We specify Gaussian processes as priors of the random functions and exploit the "look-up table approach" of Bhattacharya (2007) to efficiently handle the dynamic structure of the model. We consider both univariate and multivariate situations, using the Markov chain Monte Carlo (MCMC) approach for studying the posterior distributions of interest. We illustrate our methods with simulated data sets, in both univariate and multivariate situations. Moreover, using our Gaussian process approach we analyse a real data set, which has also been analysed by Shumway & Stoffer (1982) and Carlin, Polson & Stoffer (1992) using the linearity assumption. Interestingly, our analyses indicate that towards the end of the time series, the linearity assumption is perhaps questionable.

*Anurag Ghosh is a PhD student in Department of Statistical Science, Duke University, Soumalya Mukhopadhyay is a PhD student in Agricultural and Ecological Research Unit, Indian Statistical Institute, Sandipan Roy is a PhD student in Department of Statistics, University of Michigan, Ann Arbor, and Sourabh Bhattacharya is an Assistant Professor in Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108. Corresponding e-mail: sourabh@isical.ac.in.

Keywords: *Evolutionary equation; Gaussian process; Look-up table; Markov Chain Monte Carlo; Observational equation; State-space model.*

1. INTRODUCTION

The state-space models play important role in dealing with dynamic systems that arise in various disciplines such as finance, engineering, ecology, medicine, and statistics. The time-varying regression structure and the flexibility inherent in the sequential nature of state-space models make them very suitable for analysis and prediction of dynamic data. Indeed, as is well-known, most time series models of interest are expressible as state-space models; see Durbin & Koopman (2001) and Shumway & Stoffer (2011) for details. However, till date, the state-space models have considered only known forms of the equations, typically linear. But testing the parametric assumptions require data on both the observed time series and the unobserved states; unfortunately, data on the latter are not available in state-space models. Moreover, the regression structures of the state-space models may evolve with time, changing from linear to non-linear, and even the non-linear structure may also evolve with time, yielding further different non-linear structures. We are not aware of any nonparametric state-space approach in the statistical literature that can handle unknown functional forms, which may or may not be evolving with time. Another criticism of the existing state space models is the assumption that the (unobserved) states satisfy the Markov property. Although such Markov models have been useful in many situations where there are natural laws supporting such conditional independence, in general such assumption is not expected to hold. These arguments point towards the need for developing general, nonparametric, approaches to state-space models, and this indeed, is our aim in this article. We adopt the Bayesian paradigm for its inherent flexibility.

In a nutshell, in this work, adopting a nonparametric Bayesian framework, we treat the regression structures as unknown and model these as Gaussian processes, and develop the consequent theory in the Bayesian framework, considering both univariate and multivariate situations. Our Gaussian process approach of viewing the unknown functional forms allows very flexible modeling of the unknown structures, even though they might evolve with time. Also, as we discusss

in Section 4.7, as a consequence of our nonparametric approach, the unobserved state variables do not follow any Markov model. Thus our approach provides a realistic dependence structure between the state variables. We also develop efficient MCMC-based methods for simulating from the resulting posterior distributions. We demonstrate our methods in the case of both univariate and multivariate situations using simulated data. Application of our ideas to a real data set which has been analysed by Shumway & Stoffer (1982) and Carlin et al. (1992) assuming linearity, provided an interesting insight that, although the linearity assumption may not be unreasonable for most part of the time series, the assumption may be called in question towards the end of the time series. This vindicates that our approach is indeed capable of modeling unknown functions even if the forms are changing with time, without requiring any change point analysis and specification of functional forms before and after change points.

Before introducing our approach, we provide a brief overview of state-space models.

2. OVERVIEW OF STATE-SPACE MODELS

Generally, state-space models are of the following form: for $t = 1, 2, \ldots,$

$$y_t = f_t(x_t) + \epsilon_t \tag{1}$$

$$x_t = g_t(x_{t-1}) + \eta_t \tag{2}$$

In the above, f_t and g_t are assumed to be functions of known forms which may or may not explicitly depend upon t; η_t, ϵ_t are usually assumed to be zero mean iid normal variates. The choice $f_t(x_t) = F_t x_t$ and $g_t(x_{t-1}) = G_t x_{t-1}$, assuming known F_t, G_t, have found very wide use in the literature. Obviously, x_t, y_t may be univariate or multivariate. Matrix-variate dynamic linear models have been considered by Quintana & West (1987) and West & Harrison (1997) (see also Carvalho & West (2007)). Equation (1) is called the observational equation, while (2) is known as the evolutionary equation. Letting $\boldsymbol{D}_T = (y_1, \ldots, y_T)'$ denote the available data, the goal is to obtain inferences about y_{T+1} (single-step forecast), y_{T+k} (k-step forecast), x_{T+1} conditional on y_{T+1} (filtering), x_{T-k} (retrospection). In the Bayesian paradigm, the interests center upon analyzing the corresponding posteriors $[y_{T+1} \mid \boldsymbol{D}_T]$, $[y_{T+k} \mid \boldsymbol{D}_T]$, $[x_{T+1} \mid \boldsymbol{D}_T, y_{T+1}]$ (also, $[x_{T+1} \mid \boldsymbol{D}_T]$) and

$[x_{T-k} \mid \boldsymbol{D}_T]$.

In the non-Bayesian framework, solutions to dynamic systems are quite generally available via the well-known Kalman filter. However, the performance of Kalman filter is heavily dependent on the assumption of Gaussian errors and linearity of the functions in the observation and the evolution equations. In the case of non-linear dynamic models, various linearization techniques are used to obtain approximate solutions. For details on these issues, see Brockwell & Davis (1987), Meinhold & Singpurwala (1989), West & Harrison (1997) and the references therein. The Bayesian paradigm frees the investigator from restrictions of linear functions or Gaussian errors, and allows for very general dynamic model building through coherent combination of prior and the available time series data, and using Markov chain Monte Carlo (MCMC) for inference. Bayesian non-linear dynamic models with non-Gaussian errors, in conjunction with the Gibbs sampling approach for inference, have been considered in Carlin et al. (1992). For general details on non-linear and non-Gaussian approches to state space models, see Durbin & Koopman (2001).

However, even non-linear and non-Gaussian approaches assume that there is an underlying known natural phenomenon supporting some standard parametric model. Except in well-studied scientific contexts such assumptions are not unquestionable. In this work, we particularly concern ourselves with situations where parametric models are not established for the underlying scientific study. A case in point may be the context of climate change dynamics, where observed climate depends upon various factors in the forms of latent states, but in our knowledge, no clear parametric model is available for this extremely important and challenging problem. In medicine, growth of cancer cells at any time point may depend upon various unobserved factors (states), but no clear parametric model is available, in our knowledge. Similar challenges exist in sociology, in studies of dynamic social networks; in astrophysics, associated with the study of the evolution of the universe; in computer science, for target tracking and data-driven computer animation, and in various other fields. Thus, we expect our nonparametric approach to be quite relevant and useful for these investigations.

For the sake of clarity, in this main article, we consider only one-dimensional x_t and y_t, but we provide additional details of the univariate cases, and generalize our approach to accommodate

multivariate situations in the supplement Ghosh, Mukhopadhyay, Roy & Bhattacharya (2013), whose sections, figures and tables have the prefix "S-" when referred to in this paper. A brief description of the contents of the supplement can be found at the end of this article.

In Section 3 we introduce our novel nonparametric dynamic model where we use Gaussian processes to model the unknown functions f_t and g_t. Assuming the functions to be random allows us to accommodate even those functions the forms of which are changing with time. In order to describe the distribution of the unobserved states, we adopt the "look-up table" idea of Bhattacharya (2007). Since this idea is an integral part of the development of our methodology, we devote Section 4 to its detailed discussion. In Section 5 we provide the forms of the prior distributions of the hyperparameters of our model and build an MCMC based methodology for Bayesian inference. In Section 6 we include a brief discussion of two simulation studies, the details of which are reported in Section S-3 of the supplement. In Section 7 we consider application to a real, univariate data set. We present a summary of the current work, along with discussion of further work in Section 8.

3. NONPARAMETRIC DYNAMIC MODEL: UNIVARIATE CASE

In this section we model the unknown observational and the evolutionary functions using Gaussian processes assuming that the true functions are evolving with time t; our approach includes the time-invariant situation as a simple special case; see Section S-3.2.

In (1) and (2) we now assume f_t and g_t to be of unknown functional forms varying with time t. For convenience, we denote $f_t(x_t)$ and $g_t(x_{t-1})$ as $f(t, x_t)$ and $g(t, x_{t-1})$, respectively. That is, we treat time t as an input to both the observational and evolutionary functions, in addition to the other relevant inputs x_t and x_{t-1}. With this understanding, we re-write (1) and (2) as

$$y_t = f(t, x_t) + \epsilon_t, \;\; \epsilon_t \sim N(0, \sigma^2_\epsilon), \tag{3}$$

$$x_t = g(t, x_{t-1}) + \eta_t, \;\; \eta_t \sim N(0, \sigma^2_\eta). \tag{4}$$

We assume that $x_0 \sim N(\mu_{x_0}, \sigma^2_{x_0})$; $\mu_{x_0}, \sigma^2_{x_0}$ being known. Crucially, we allow $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$ to be of unknown functional forms, which we model as two independent Gaussian processes. To

present the details of the Gaussian processes, we find it convenient to use the notation $x^*_{t,t} = (t, x_t)'$, and $x^*_{t,t-1} = (t, x_{t-1})'$ so that (3) and (4) can be re-written as

$$y_t = f(x^*_{t,t}) + \epsilon_t, \ \ \epsilon_t \sim N(0, \sigma^2_\epsilon), \tag{5}$$

$$x_t = g(x^*_{t,t-1}) + \eta_t, \ \ \eta_t \sim N(0, \sigma^2_\eta); \tag{6}$$

in fact, more generally, we use the notation $x^*_{t,u} = (t, x_u)$. This general notation will be convenient for describing theoretical and computational details. Next, we provide details of the independent Gaussian processes used to model f and g.

3.1 Modeling the unknown observational and evolutionary time-varying functions using independent Gaussian processes

The functions f and g are modeled as independent Gaussian processes with mean functions $\mu_f(\cdot) = \boldsymbol{h}(\cdot)'\boldsymbol{\beta}_f$ and $\mu_g(\cdot) = \boldsymbol{h}(\cdot)'\boldsymbol{\beta}_g$ with $\boldsymbol{h}(x^*) = (1, x^*)'$ for any x^*, and covariance functions of the form $\sigma^2_f c_f(\cdot,\cdot)$ and $\sigma^2_g c_g(\cdot,\cdot)$, respectively. The process variances are σ^2_f and σ^2_g and c_f, c_g are the correlation functions. Typically, for any z_1, z_2, $c_f(z_1, z_2) = \exp\{-(z_1 - z_2)'\boldsymbol{R}_f(z_1 - z_2)\}$ and $c_g(z_1, z_2) = \exp\{-(z_1 - z_2)'\boldsymbol{R}_g(z_1 - z_2)\}$, where $\boldsymbol{R}_f$ and $\boldsymbol{R}_g$ are 2×2-dimensional diagonal matrices consisting of respective smoothness (or, roughness) parameters $\{r_{1,f}, r_{2,f}\}$ and $\{r_{1,g}, r_{2,g}\}$, which are responsible for the smoothness of the process realizations. These choices of the correlation functions imply that the functions, modeled by the process realizations, are infinitely smooth.

The sets of parameters $\boldsymbol{\theta}_f = (\boldsymbol{\beta}_f, \sigma^2_f, \boldsymbol{R}_f)$ and $\boldsymbol{\theta}_g = (\boldsymbol{\beta}_g, \sigma^2_g, \boldsymbol{R}_g)$ are assumed to be independent *a priori*. We consider the following form of prior distribution of the parameters: $[\boldsymbol{\beta}_f, \sigma^2_f, \boldsymbol{R}_f, \boldsymbol{\beta}_g, \sigma^2_g, R_g, \sigma^2_\epsilon, \sigma^2_\eta] = [\boldsymbol{\beta}_f, \sigma^2_f, \boldsymbol{R}_f][\boldsymbol{\beta}_g, \sigma^2_g, \boldsymbol{R}_g][\sigma^2_\epsilon, \sigma^2_\eta]$.

3.2 Hierarchical structure induced by our Gaussian process approach

In summary, our approach can be described in the following hierarchical form:

$$[y_t|f, \boldsymbol{\theta}_f, x_t] \sim N\left(f(x^*_{t,t}), \sigma^2_\epsilon\right);\ t = 1, \ldots, T, \tag{7}$$

$$[x_t|g, \boldsymbol{\theta}_g, x_{t-1}] \sim N\left(g(x^*_{t,t-1}), \sigma^2_\eta\right);\ t = 1, \ldots, T, \tag{8}$$

$$[x_0] \sim N\left(\mu_{x_0}, \sigma^2_{x_0}\right), \tag{9}$$

$$[f(\cdot)|\boldsymbol{\theta}_f] \sim GP\left(\boldsymbol{h}'(\cdot)'\boldsymbol{\beta}_f, \sigma^2_f c_f(\cdot,\cdot)\right), \tag{10}$$

$$[g(\cdot)|\boldsymbol{\theta}_g] \sim GP\left(\boldsymbol{h}'(\cdot)\boldsymbol{\beta}_g, \sigma^2_g c_g(\cdot,\cdot)\right), \tag{11}$$

$$[\boldsymbol{\beta}_f, \sigma^2_f, \boldsymbol{R}_f, \boldsymbol{\beta}_g, \sigma^2_g, R_g, \sigma^2_\epsilon, \sigma^2_\eta] = [\boldsymbol{\beta}_f, \sigma^2_f, \boldsymbol{R}_f][\boldsymbol{\beta}_g, \sigma^2_g, \boldsymbol{R}_g][\sigma^2_\epsilon, \sigma^2_\eta]. \tag{12}$$

In the above, GP stands for "Gaussian Process". Forms of the prior distributions in (12) are provided in (5) and specific details are provided in the relevant applications.

3.3 Conditional distribution of the observed data induced by the Gaussian process prior on the observational function and a brief discussion of the difficulty of obtaining the joint distribution of the state variables

It follows from the Gaussian process prior assumption on the unknown observational function f that the distribution of $\boldsymbol{D}_T$, conditional on $x^*_{1,1}, \ldots, x^*_{T,T}$ (equivalently, conditional on $x_1, \ldots, x_T$), and the other parameters is multivariate normal:

$$\boldsymbol{D}_T \sim N_T(\boldsymbol{H}_{D_T}\boldsymbol{\beta}_f, \sigma^2_f \boldsymbol{A}_{f,D_T} + \sigma^2_\epsilon \boldsymbol{I}_T), \tag{13}$$

where $\boldsymbol{H}'_{D_T} = \left[\boldsymbol{h}(x^*_{1,1}), \ldots, \boldsymbol{h}(x^*_{T,T})\right]$, $\boldsymbol{A}_{f,D_T}$ is a $T \times T$ matrix with (i,j)-th element $c_f(x^*_{i,i}, x^*_{j,j})$; $(i,j) = 1, \ldots, T$, and $\boldsymbol{I}_T$ is the T-th order identity matrix.

The joint distribution of the state variables $(x_1, \ldots, x_T)$, however, is much less straightforward. Observe that, although we have $[x_0] \sim N(\mu_{x_0}, \sigma^2_{x_0})$, $[x_1 \mid x_0] \sim N(\boldsymbol{h}(x_0)'\boldsymbol{\beta}_g, \sigma^2_g + \sigma^2_\eta)$, but $[x_2 \mid x_1, x_0]$=$[g(2, x_1) + \eta_2 \mid x_1, x_0]$=$[g(2, g(1, x_0) + \eta_1) + \eta_2 \mid g(1, x_0) + \eta_1, x_0]$, the rightmost expression suggesting that special techniques may be necessary to get hold of the conditional distribution. We adopt the procedure introduced by Bhattacharya (2007) to deal with this problem. The idea is to conceptually simulate the entire function g modeled by the Gaussian process, and

use the simulated process as a look-up table to obtain the conditional distributions of $\{x_i; i \geq 2\}$. The intuition behind the look-up table concept is briefly dicussed in the next subsection, while the detailed procedure of approximating the joint distribution of the state variables is provided in Section 4.

3.4 Intuition behind the look-up table idea for approximating the joint distribution of the state variables

For the purpose of illustration only let us assume that $\epsilon_t = 0$ for all t, yielding the model $x_t = g(x^*_{t,t-1})$. The concept of look-up table in this problem can be briefly explained as follows. Let us first assume that the entire process $g(\cdot)$ is available. This means that for every input z, the corresponding $g(z)$ is available, thus constituting a look-up table, with the first column representing z and the second column representing the corresponding $g(z)$. Conditional on $x^*_{t,t-1}$ (equivalently, conditional on x_{t-1}), $x_t = g(x^*_{t,t-1})$ can be obtained by simply picking the input $x^*_{t,t-1}$ from the first column of the look-up table and reporting the corresponding output value $g(x^*_{t,t-1})$, located in the second column of the look-up table. This hypothetical look-up table concept suggests that conditional on the simulated process g, it can be safely assumed that x_t depends only upon $x^*_{t,t-1}$ via $x_t = g(x^*_{t,t-1})$. Thus, if for all possible inputs, a simulation of the entire random function g, following the Gaussian process, is available, then for any input $x^*_{t,t-1}$, we only need to identify the corresponding $g(x^*_{t,t-1})$ in the look-up table. In practice, we can have a simulation of the Gaussian process g on a fine enough grid of inputs. Given this simulation on a fine grid, we can simulate from the conditional distribution of $g(x^*_{t,t-1})$, fixing $x^*_{t,t-1}$ as given. This simulation from the conditional distribution of $g(x^*_{t,t-1})$ will approximate x_t as accurately as we desire by making the grid as fine as required. By repeating this procedure for each t, we can approximate the joint distribution of the state variables as closely as we desire. In the next section we provide details regarding this approach.

4. DETAILED PROCEDURE OF APPROXIMATING THE JOINT DISTRIBUTION OF THE STATE VARIABLE USING THE LOOK-UP TABLE CONCEPT

4.1 Distribution of x_1

Note that given x_0 we can simulate $x_1 = g(x^*_{1,0}) \sim N(\boldsymbol{h}(x^*_{1,0})'\boldsymbol{\beta}_g, \sigma_g^2)$, which is the marginal distribution of the Gaussian process prior. Thus, x_1 is simulated without resorting to any approximation. It then remains to simulate the rest of the dynamic sequence, for which we need to simulate the rest of the process $\{g(x^*); x^* \neq x^*_{1,0}\}$.

4.2 Introduction of a set of auxiliary variables to act as proxy to the Gaussian process g

In practice, it is not possible to have a simulation of this entire set $\{g(x^*); x^* \neq x^*_{1,0}\}$.. We only have available a set of grid points $\boldsymbol{G}_n = \{z_1, \ldots, z_n\}$ obtained, perhaps, by Latin hypercube sampling (see, for example, Santner, Williams & Notz (2003)) and a corresponding simulation of g, given by $\boldsymbol{D}^*_n = \{g(z_1), \ldots, g(z_n)\}$, the latter having a joint multivariate normal distribution with mean

$$E\left[\boldsymbol{D}^*_n \mid \boldsymbol{\theta}_g\right] = \boldsymbol{H}_{D^*_n}\boldsymbol{\beta}_g \tag{14}$$

and covariance matrix

$$V\left[\boldsymbol{D}^*_n \mid \boldsymbol{\theta}_g\right] = \sigma_g^2 \boldsymbol{A}_{g,D^*_n}, \tag{15}$$

where $\boldsymbol{H}'_{D^*_n}$=$[\boldsymbol{h}(z_1), \ldots, \boldsymbol{h}(z_n)]$ and $\boldsymbol{A}_{g,D^*_n}$ is a correlation matrix with the (i,j)-th element $c_g(z_i, z_j)$. and $\boldsymbol{s}_{g,D^*_n}(\cdot) = (c_g(\cdot, z_1), \ldots, c_g(\cdot, z_n))'$.

Given $(x_0, g(x^*_{1,0}))$, we simulate $\boldsymbol{D}^*_n$ from $[\boldsymbol{D}^*_n \mid \boldsymbol{\theta}_g, g(x^*_{1,0}), x_0]$. Since the joint distribution of $\left[\boldsymbol{D}^*_n, g(x^*_{1,0}) \mid x_0\right]$ is multivariate normal with mean vector $\begin{pmatrix} \boldsymbol{H}_{D^*_n}\boldsymbol{\beta}_g \\ \boldsymbol{h}(x_0)'\boldsymbol{\beta}_g \end{pmatrix}$ and covariance matrix $\sigma_g^2 \boldsymbol{A}_{D^*_n, g(x^*_{1,0})}$ where

$$\boldsymbol{A}_{D^*_n, g(x^*_{1,0})} = \begin{pmatrix} \boldsymbol{A}_{g,D^*_n} & \boldsymbol{s}_{g,D^*_n}(x^*_{1,0}) \\ \boldsymbol{s}_{g,D^*_n}(x^*_{1,0})' & 1 \end{pmatrix}, \tag{16}$$

it follows that the conditional $[\boldsymbol{D}^*_n \mid g(x^*_{1,0}), x^*_{1,0}]$ has an n-variate normal distribution with mean

vector

$$E[\boldsymbol{D}_n^* \mid g(x_{1,0}^*), x_0, \boldsymbol{\theta}_g] = \boldsymbol{\mu}_{g,D_n^*} = \boldsymbol{H}_{D_n^*}\boldsymbol{\beta}_g + \boldsymbol{s}_{g,D_n^*}(x_{1,0}^*)(g(x_{1,0}^*) - \boldsymbol{h}(x_{1,0}^*)'\boldsymbol{\beta}_g) \tag{17}$$

and covariance matrix

$$V[\boldsymbol{D}_n^* \mid g(x_{1,0}^*), x_0, \boldsymbol{\theta}_g] = \sigma_g^2 \boldsymbol{\Sigma}_{g,D_n^*}, \tag{18}$$

where

$$\boldsymbol{\Sigma}_{g,D_n^*} = \boldsymbol{A}_{g,D_n^*} - \boldsymbol{s}_{g,D_n^*}(x_{1,0}^*)\boldsymbol{s}_{g,D_n^*}(x_{1,0}^*)'. \tag{19}$$

4.3 Distribution of each state variable conditional on the look-up table proxy $\boldsymbol{D}_n^*$

We now seek the conditional distribution $[x_t = g(x_{t,t-1}^*) \mid \boldsymbol{D}_n^*, x_{t-1}, x_{t-2}, \ldots, x_1]$. To notationally distinguish between the conditional distribution of $g(\cdot)$ from the elements of the set $\boldsymbol{D}_n^*$, we henceforth denote the elements of $\boldsymbol{D}_n^*$ as $g_{true}(\cdot)$. In other words, we henceforth write $\boldsymbol{D}_n^* = \{g_{true}(z_1), \ldots, g_{true}(z_n)\}$.

Recall that the look-up table idea supports conditional independence, that is, given a simulation of the entire random function g, x_t depends only upon x_{t-1} via $x_t = g(x_{t,t-1}^*)$, so that given g, x_t is conditionally independent of $\{x_k;\ k < t-1\}$. Indeed, given a fine enough grid $\boldsymbol{G}_n$, $\boldsymbol{D}_n^*$ approximates the random function g, which contains all information regarding the conditioned state variables $x_1, \ldots, x_{t-1}$. Hence,

$$[g(x_{t,t-1}^*) \mid \boldsymbol{D}_n^*, x_{t-1}, x_{t-2}, \ldots, x_1] \approx [g(x_{t,t-1}^*) \mid \boldsymbol{D}_n^*, x_{t-1}], \tag{20}$$

and this approximation can be made arbitrarily accurate by making the grid $\boldsymbol{G}_n$ as fine as desired. Hence, it is sufficient for our purpose to deal with the conditional distribution of $[g(x_{t,t-1}^*) \mid \boldsymbol{D}_n^*, x_{t-1}]$. This is easy to obtain: since given x_{t-1}, $(g(x_{t,t-1}^*), \boldsymbol{D}_n^*)$ is jointly multivariate normal, it is easily seen that $[g(x_{t,t-1}^*) \mid \boldsymbol{D}_n^*, x_{t-1}]$ is normal with mean

$$\mu_t = \boldsymbol{h}(x_{t,t-1}^*)'\boldsymbol{\beta}_g + \boldsymbol{s}_{g,D_n^*}(x_{t,t-1}^*)'\boldsymbol{A}_{g,D_n^*}^{-1}(\boldsymbol{D}_n^* - \boldsymbol{H}_{D_n^*}\boldsymbol{\beta}_g) \tag{21}$$

and variance

$$\sigma_t^2 = \sigma_g^2 \left\{1 - \boldsymbol{s}_{g,D_n^*}(x_{t,t-1}^*)'\boldsymbol{A}_{g,D_n^*}^{-1}\boldsymbol{s}_{g,D_n^*}(x_{t,t-1}^*)\right\}. \tag{22}$$

One subtlety involved in the assumption of conditional independence is that, conditional on $x^*_{1,0}$, $\boldsymbol{G}_n$ must not contain $x^*_{1,0}$; otherwise $\boldsymbol{D}^*_n$ would contain $g(x^*_{1,0})$ implying that $[g(x^*_{t,t-1}) \mid \boldsymbol{\theta}_g, \boldsymbol{D}^*_n, x_t]$ is dependent on $x_1 = g(x^*_{1,0})$, violating the conditional independence assumption.

4.4 Accuracy of the Markov approximation of the distributions of the state variables conditional on $\boldsymbol{D}^*_n$

To first heuristically understand how the approximation (20) can be made arbitrarily accurate, note that thanks to the Gaussian process assumption, conditioning on $\boldsymbol{D}^*_n$ forces the random function $g(\cdot)$ to pass through the points in $(\boldsymbol{G}_n, \boldsymbol{D}^*_n)$ since the conditional $[g(x) \mid \boldsymbol{\theta}_g, \boldsymbol{D}^*_n, x]$ has zero variance if $x \in \boldsymbol{G}_n$ (see, for example, Bhattacharya (2007) and the references therein). In other words, if $x^*_{t,t-1} \in \boldsymbol{G}_n$, then $\sigma^2_t = 0$ so that

$$[g(x^*_{t,t-1}) \mid \boldsymbol{\theta}_g, \boldsymbol{D}^*_n, x_{t-1}] = \delta_{g_{true}(x^*_{t,t-1})}, \tag{23}$$

where δ_z denotes point mass at z. This property of the conditional associated with Gaussian process is in keeping with the insight gained from the discussion related to look-up table associated with prediction of the outputs of deterministic function having dynamic behaviour. However, $x^*_{t,t-1} \notin \boldsymbol{G}_n$ with probability 1 and the conditional $[g(x^*_{t,t-1}) \mid \boldsymbol{\theta}_g, \boldsymbol{D}^*_n, x_{t-1}]$ provides spatial interpolation within $(\boldsymbol{G}_n, \boldsymbol{D}^*_n)$ (see, for example, Cressie (1993), Stein (1999)). Finer the set $\boldsymbol{G}_n$, closer is $[g(x^*_{t,t-1}) \mid \boldsymbol{\theta}_g, \boldsymbol{D}^*_n, x_{t-1}]$ to $\delta_{g(x^*_{t,t-1})}$. The conditional independence assumption of $g(x^*_{t,t-1})$ of all $\{x^*_{t,k}; k < (t-1)\}$ given $(\boldsymbol{G}_n, \boldsymbol{D}^*_n)$ is in accordance with the motivation provided by the deterministic sequence and here $\boldsymbol{D}^*_n$ acts as a set of auxiliary variables, greatly simplifying computation, while not compromising on accuracy.

In Section S-1 of the supplement we formally prove a theorem stating that, given a particular design $\boldsymbol{G}_n$, the order of approximation of $g_{true}(\cdot)$ by the conditional distribution of $g(\cdot)$ given $\boldsymbol{D}^*_n$, within a finite region (but as large as required for all practical purposes), is $O\left(n^{-1}\right)$. Hence, given a judiciously chosen sufficiently fine grid $\boldsymbol{G}_n$ and corresponding $\boldsymbol{D}^*_n$, the conditioned state variables $\{x_k;\ k < t\}$ do not provide any extra information to $x_t = g(x^*_{t,t-1})$ regarding $g_{true}(x^*_{t,t-1})$ in an asymptotic sense with respect to $\boldsymbol{G}_n$. Hence, the Markov approximation (20) is valid for appropriate $\boldsymbol{G}_n$, and the accuracy of the approximation can be improved arbitrarily.

4.5 Summary of the look-up table procedure for obtaining the joint distribution of the state variables

To summarize the ideas, let $x_0 \sim \pi$, where π is some appropriate prior distribution of x_0. The entire dynamic sequence $\{x_0, x_1 = g(x^*_{1,0}), x_2 = g(x^*_{2,1}), \ldots\}$ can then be simulated using the following steps sequentially:

(1) Draw $x_0 \sim \pi$.

(2) Given x_0, draw $x_1 = g(x^*_{1,0}) \sim N(\boldsymbol{h}(x^*_{1,0})'\boldsymbol{\beta}_g, \sigma^2_g)$.

(3) Given x_0, and $x_1 = g(x^*_{1,0})$, draw $\boldsymbol{D}^*_n \sim [\boldsymbol{D}^*_n \mid \boldsymbol{\theta}_g, g(x^*_{1,0}), x_0]$.

(4) For $t = 2, 3, \ldots$, draw $x_t \sim [x_t = g(x^*_{t,t-1}) \mid \boldsymbol{\theta}_g, \boldsymbol{D}^*_n, x_{t-1}]$.

Step (1) is a simulation of x_0 from its prior, step (2) is simply drawn from the known marginal distribution of $g(x^*_{1,0})$ given x_0. In step (3) $\boldsymbol{D}^*_n$ is drawn conditional on $g(x^*_{1,0})$ (and $x^*_{1,0}$), conceptually implying that the rest of the process $\{g(x^*); x \neq x^*_{1,0}\}$ is drawn once $g(x^*_{1,0})$ is known. Step (4) then uses this simulated $\boldsymbol{D}^*_n$ to obtain the rest of the dynamic sequence, using the assumed conditional independence structure.

4.6 Explicit form of the look-up table induced joint distribution of $\{x_0, x_1, \ldots, x_{T+1}, \boldsymbol{D}^*_n\}$

Once $\boldsymbol{G}_n$ and $\boldsymbol{D}^*_n$ are available, we write down the joint distribution of $(\boldsymbol{D}^*_n, x_0, x_1, \ldots, x_T)$ conditional on the other parameters as

$$
\begin{aligned}
&[x_0, x_1, \ldots, x_{T+1}, \boldsymbol{D}^*_n \mid \boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma^2_\epsilon, \sigma^2_\eta] \\
&\qquad = [x_0][x_1 = g(x^*_{1,0}) + \eta_1 \mid x_0, \sigma^2_\eta][\boldsymbol{D}^*_n \mid \boldsymbol{\theta}_g] \\
&\qquad \prod_{t=1}^{T} [x_{t+1} = g(x^*_{t+1,t}) + \eta_{t+1} \mid \boldsymbol{D}^*_n, x_t, \boldsymbol{\theta}_g, \sigma^2_\eta] \qquad (24)
\end{aligned}
$$

Recall that $[x_0] \sim N(\mu_{x_0}, \sigma^2_{x_0})$, $[x_1 = g(x^*_{1,0}) + \eta_1 \mid x^*_{1,0}, \sigma^2_\eta] \sim N(\boldsymbol{h}(x^*_{1,0})'\boldsymbol{\beta}_g, \sigma^2_g + \sigma^2_\eta)$ and the distribution of $\boldsymbol{D}^*_n$ is multivariate normal with mean and variance given by (14) and (15). The conditional distribution $[x_{t+1} = g(x^*_{t+1,t}) + \eta_{t+1} \mid \boldsymbol{D}^*_n, x_t, \boldsymbol{\theta}_g, \sigma^2_\eta]$ is normal with mean

$$
\mu_{x_t} = \boldsymbol{h}(x^*_{t+1,t})'\boldsymbol{\beta}_g + \boldsymbol{s}_{g,\boldsymbol{D}^*_n}(x^*_{t+1,t})'\boldsymbol{A}^{-1}_{g,D^*_n}(\boldsymbol{D}^*_n - \boldsymbol{H}_{D^*_n}\boldsymbol{\beta}_g) \qquad (25)
$$

and variance

$$\sigma^2_{x_t} = \sigma^2_\eta + \sigma^2_g \left\{ 1 - \boldsymbol{s}_{g,\boldsymbol{D}^*_n}(x^*_{t+1,t})' \boldsymbol{A}^{-1}_{g,\boldsymbol{D}^*_n} \boldsymbol{s}_{g,\boldsymbol{D}^*_n}(x^*_{t+1,t}) \right\} \tag{26}$$

Observe that in this case even if $x^*_{t+1,t} \in \boldsymbol{G}_n$, due to the presence of the additive error term η_{t+1}, the conditional variance of x_{t+1} is non-zero, equalling $\sigma^2_{x_t} = \sigma^2_\eta$, the error variance.

4.7 Non-Markovian dependence structure of the marginalized joint distribution of the state variables $(x_0, x_1, \ldots, x_{T+1})$

As in Bhattacharya (2007), here also it is possible to marginalize out $\boldsymbol{D}^*_n$ from (24) to obtain the approximate joint distribution of $(x_0, x_1, \ldots, x_T)$. However, if $\boldsymbol{D}^*_n$ is integrated out, it is clear that the conditional indpendepence (Markov) property of x_t's given $\boldsymbol{D}^*_n$, as seen in (24), will be lost. Thus, the marginalized conditional distribution of x_{t+1} depends upon $\{x_k;\ k < t+1\}$, that is, the set of all the past state variables, unlike the non-marginalized case, where conditionally on $\boldsymbol{D}^*_n$, x_{t+1} depends only upon x_t. This makes it clear that even though for fixed (known) evolutionary function g the corresponding equation (2) satisfies the Markov property, such Markov property is lost when the function is modeled as Gaussian processes. Hence, in our approach based on Gaussian process, the state variables are non-Markovian.

4.8 To marginalize or not to marginalize with respect to $\boldsymbol{D}^*_n$?

As discussed in detail in Bhattacharya (2007), the complicated dependence structure associated with the marginalized joint distribution of $(x_0, x_1, \ldots, x_{T+1})$ is also the root of all numerical instabilities associated with MCMC implementation of our model. To understand this heuristically, note that evaluations of the conditionals $[x_{t+1}|x_t, \ldots, x_0, \boldsymbol{\theta}_g, \sigma^2_g]$ are required for MCMC implementation, but evaluations of the conditionals require inversions of covariance matrices involving the random states $x_0, x_1, \ldots, x_{t-1}$ in the correlation terms. By sample path continuity of the underlying Gaussian process, the sampled states $x_0, x_1, \ldots, x_{t-1}$ will be often close to each other with high probability, particularly if σ^2_g and σ^2_η are small, rendering the correlation matrix almost singular. Moreover, inversion of such correlation matrices at every iteration of MCMC is also very costly computationally. The problems are much aggravated for large t.

On the other hand, if $\boldsymbol{D}_n^*$ is retained, then such problem is avoided, since in that case we only need to compute the conditionals $[x_{t+1}|D_n^*, x_t, \boldsymbol{\theta}_g, \sigma_g^2]$, which involves inversion of the correlation matrix $\boldsymbol{A}_{g,D_n^*}$, which has (i,j)-th element of the form $c(z_i, z_j)$, where $z_1, \ldots, z_n$ are fixed constants selected by the user. Also, quite importantly, $\boldsymbol{A}_{g,D_n^*}^{-1}$ can be computed even before beginning MCMC simulations, and it remains fixed thereafter, saving a lot of computational time, in addition to providing protection against numerical instability. For further details, see Bhattacharya (2007).

Hence, we retain the set of auxiliary variables $\boldsymbol{D}_n^*$ in (24) for implementation of our MCMC methods, and finally discard them from the resultant MCMC samples to infer about the quantities of our interest.

In the next section we complete specification of our fully Bayesian model by choosing appropriate prior distributions of the parameters.

5. PRIOR SPECIFICATIONS AND BAYESIAN INFERENCE USING MCMC

We assume the following prior distributions:

$$[\log(r_{i,f})] \sim N\left(\mu_{r_{i,f}}, \sigma^2_{r_{i,f}}\right); \text{ for } i = 1, 2. \tag{27}$$

$$[\log(r_{i,g})] \sim N\left(\mu_{r_{i,g}}, \sigma^2_{r_{i,g}}\right); \text{ for } i = 1, 2. \tag{28}$$

$$[\sigma_\epsilon^2] \propto \left(\sigma_\epsilon^2\right)^{-\left(\frac{\alpha_\epsilon+2}{2}\right)} \exp\left\{-\frac{\gamma_\epsilon}{2\sigma_\epsilon^2}\right\}; \alpha_\epsilon, \gamma_\epsilon > 0 \tag{29}$$

$$[\sigma_\eta^2] \propto \left(\sigma_\eta^2\right)^{-\left(\frac{\alpha_\eta+2}{2}\right)} \exp\left\{-\frac{\gamma_\eta}{2\sigma_\eta^2}\right\}; \alpha_\eta, \gamma_\eta > 0 \tag{30}$$

$$[\sigma_f^2] \propto \left(\sigma_f^2\right)^{-\left(\frac{\alpha_f+2}{2}\right)} \exp\left\{-\frac{\gamma_f}{2\sigma_f^2}\right\}; \alpha_f, \gamma_f > 0 \tag{31}$$

$$[\sigma_g^2] \propto \left(\sigma_g^2\right)^{-\left(\frac{\alpha_g+2}{2}\right)} \exp\left\{-\frac{\gamma_g}{2\sigma_g^2}\right\}; \alpha_g, \gamma_g > 0 \tag{32}$$

$$[\boldsymbol{\beta}_f] \sim N_m\left(\boldsymbol{\beta}_{f,0}, \boldsymbol{\Sigma}_{\beta_{f,0}}\right) \tag{33}$$

$$[\boldsymbol{\beta}_g] \sim N_m\left(\boldsymbol{\beta}_{g,0}, \boldsymbol{\Sigma}_{\beta_{g,0}}\right) \tag{34}$$

All the prior parameters are assumed to be known. Now we discuss our approach to selecting the prior parameters for our application our Bayesian model in simulation studies and real data application in the univariate situations.

In order to choose the parameters of the log-normal priors of the smoothness parameters, we set the mean of the log-normal prior with parameters μ and σ^2, given by $\exp(\mu + \sigma^2/2)$, to 1. This yields $\mu = -\sigma^2/2$. Since the variance of this log-normal prior is given by $(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$, the relation $\mu = -\sigma^2/2$ implies that the variance is $\exp(\sigma^2) - 1 = \exp(-2\mu) - 1$. We set $\sigma^2 = 1$, so that $\mu = -0.5$. This implies that the mean is 1 and the variance is approximately 2, for the priors of each smoothness parameter $r_{i,f}$ and $r_{i,g}$; $i = 1, 2$.

For the choice of the parameters of the priors of σ_f^2, σ_g^2, σ_ϵ^2 and σ_η^2, we first note that the mean is of the form $\gamma/(\alpha - 2)$ and the variance is of the form $2\gamma^2/\{(\alpha - 2)^2(\alpha - 4)\}$. Thus, if we set $\gamma/(\alpha - 2) = a$, then the variance becomes $2a^2/(\alpha - 4)$. Here we set $a = 0.5, 0.5, 0.1, 0.1$, respectively, for σ_f^2, σ_g^2, σ_ϵ^2, σ_η^2. For each of these priors we set $\alpha = 4.01$, so that the variance is of the form $200a^2$.

We set the priors of $\boldsymbol{\beta}_f$ and $\boldsymbol{\beta}_g$ to be trivariate normal with zero mean and the identity matrix as the variance.

5.1 MCMC-based Bayesian inference

In this section we begin with the problem of forecasting y_{T+1}, given the data set $\boldsymbol{D}_T$. Interestingly, our approach to this problem provides an MCMC methodology which generates inference about all the posterior distributions required, either as by-products or by simple generalization of this MCMC approach using an augmentation scheme. Details follow.

The posterior predictive distribution of y_{T+1} given $\boldsymbol{D}_T$ is

$$\begin{aligned}[y_{T+1} \mid \boldsymbol{D}_T] \;=\; & \int [y_{T+1} \mid \boldsymbol{D}_T, x_0, x_1, \ldots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{R}_f, \sigma_\epsilon^2] \\ & \times [x_0, x_1, \ldots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \boldsymbol{R}_f, \boldsymbol{R}_g, \sigma_\epsilon^2 \sigma_\eta^2 \mid \boldsymbol{D}_T] d\boldsymbol{\theta}_f d\boldsymbol{\theta}_g d\sigma_\epsilon^2, \sigma_\eta^2 dx_0 dx_1, \ldots, dx_{t+1}. \end{aligned} \tag{35}$$

This posterior is not available analytically, and so simulation methods are necessary to make inferences. In particular, once a sample is available from the posterior $[x_0, x_1, \ldots, x_{t+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \boldsymbol{R}_f, \boldsymbol{R}_g, \sigma_\epsilon^2 \sigma_\eta^2 \mid \boldsymbol{D}_T]$, the corresponding samples drawn from $[y_{T+1} \mid \boldsymbol{D}_T, x_0, x_1, \ldots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{R}_f, \sigma_\epsilon^2]$ are from the posterior predictive (35), using which required posterior summaries can be obtained. Note that

the conditional distribution $[y_{T+1} = f(x^*_{T+1,T+1}) + \epsilon_{T+1} \mid \boldsymbol{D}_T, x_0, x_1, \ldots, x_{T+1}\boldsymbol{\beta}_f, \sigma_f^2, \boldsymbol{R}_f, \sigma_\epsilon^2]$ is normal with mean

$$\mu_{y_{T+1}} = \boldsymbol{h}(x^*_{T+1,T+1})'\boldsymbol{\beta}_f + \boldsymbol{s}_{f,\boldsymbol{D}_T}(x^*_{T+1,T+1})'\boldsymbol{A}^{-1}_{f,D_T}(\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{\beta}_f) \tag{36}$$

and variance

$$\sigma^2_{y_{T+1}} = \sigma^2_\epsilon + \sigma^2_f\left\{1 - \boldsymbol{s}_{f,\boldsymbol{D}_T}(x^*_{T+1,T+1})'\boldsymbol{A}^{-1}_{f,\boldsymbol{D}_T}\boldsymbol{s}_{f,\boldsymbol{D}_T}(x^*_{T+1,T+1})\right\} \tag{37}$$

Using $\boldsymbol{D}^*_n$, the conditional posterior $[x_0, x_1, \ldots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \boldsymbol{R}_f, \boldsymbol{R}_g, \sigma^2_\epsilon, \sigma^2_\eta \mid \boldsymbol{D}_T]$ can be written as

$$\begin{aligned}
&[x_0, x_1, \ldots, x_{T+1}, \boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma^2_\epsilon, \sigma^2_\eta \mid \boldsymbol{D}_T] \\
&= \int [x_0, x_1, \ldots, x_{T+1}, \boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma^2_\epsilon, \sigma^2_\eta, g(x^*_{1,0}), \boldsymbol{D}^*_n \mid \boldsymbol{D}_T] dg(x^*_{1,0}) d\boldsymbol{D}^*_n \qquad (38)\\
&\propto \int [x_0, x_1, \ldots, x_{T+1}, \boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma^2_\epsilon, \sigma^2_\eta, g(x^*_{1,0}), \boldsymbol{D}^*_n, \boldsymbol{D}_T] dg(x_0) d\boldsymbol{D}^*_n \qquad (39)\\
&= \int [\boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma^2_\epsilon, \sigma^2_\eta][x_0][g(x^*_{1,0}) \mid x^*_{1,0}][\boldsymbol{D}^*_n \mid g(x^*_{1,0}), x_0, \boldsymbol{\theta}_g] \\
&\quad \times [x_1 = g(x^*_{1,0}) + \eta_1 \mid g(x^*_{1,0}), x_0, \boldsymbol{\beta}_g, \sigma^2_g, \sigma^2_\eta] \\
&\quad \times [x_{T+1} = g(x^*_{T+1,T}) + \eta_T \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_T] \\
&\quad \times \prod_{t=1}^{T-1} [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_t][\boldsymbol{D}_T \mid x_1, \ldots, x_T, \boldsymbol{\theta}_f, \sigma^2_\epsilon] dg(x^*_{1,0}) d\boldsymbol{D}^*_n \qquad (40)
\end{aligned}$$

In the above, $[x_1 = g(x^*_{1,0}) + \eta_1 \mid g(x^*_{1,0}), \boldsymbol{\beta}_g, \sigma^2_g, \sigma^2_\eta] \sim N\left(g(x^*_{1,0}), \sigma^2_\eta\right)$. Although the analytic form of (40) is not available, MCMC simulation from $[x_0, x_1, \ldots, x_{T+1}, \boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma^2_\epsilon, \sigma^2_\eta, g(x^*_{1,0}), \boldsymbol{D}^*_n | \boldsymbol{D}_T]$, which is proportional to the integrand in (40), is possible. Ignoring $g(x^*_{1,0})$ and $\boldsymbol{D}^*_n$ in these MCMC simulations yields the desired samples from $[x_0, x_1, \ldots, x_{T+1}, \boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma^2_\epsilon, \sigma^2_\eta \mid \boldsymbol{D}_T]$. The details of the MCMC are provided in Section S-2.

5.2 MCMC-based single and mutiple step forecasts, filtering and retrospection

Observe that one can readily study the posteriors $[y_{T+1} \mid \boldsymbol{D}_T]$, $[x_{T+1} \mid \boldsymbol{D}_T]$ and $[x_{T-k} \mid \boldsymbol{D}_T]$ using the readily available MCMC samples of $\{y_{T+1}, x_{T+1}, x_{T-k}\}$ after ignoring the samples corresponding to the rest of the unknowns. To study the posterior $[x_{T+1} \mid \boldsymbol{D}_T, y_{T+1}]$, we only need

to augment y_{T+1} to $\boldsymbol{D}_T$ to create $\boldsymbol{D}_{T+1} = (\boldsymbol{D}'_T, y_{T+1})'$. Then our methodology can be followed exactly to generate samples from $[x_{T+1} \mid \boldsymbol{D}_{T+1}]$. Sample generation from $[y_{T+k} \mid \boldsymbol{D}_T]$ requires a slight generalization of this augmentation strategy. Here we use successive augmentation, adding each simulated y_{T+j} to the previous $\boldsymbol{D}_{T+j-1}$ to create $\boldsymbol{D}_{T+j} = (\boldsymbol{D}'_{T+j-1}, y_{T+j})$; $j = 1, 2, \ldots, k$. Then our MCMC methodology can be implemented successively to generate samples from y_{T+j+1} and all other variables. This implies that at each augmentation stage we need to draw a single MCMC sample from $[x_0, x_1, \ldots, x_{T+j+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \boldsymbol{R}_f, \boldsymbol{R}_g, \sigma_f^2, \sigma_g^2, \sigma_\epsilon^2, \sigma_\eta^2 \mid \boldsymbol{D}_{T+j}]$. Once this sample is generated, we can draw a single realization from y_{T+j+1} by drawing from $y_{T+j+1} \sim N\left(\mu_{y_{T+j+1}}, \sigma^2_{y_{T+j+1}}\right)$, where, analogous to (36) and (37),

$$\mu_{y_{T+j+1}} = \boldsymbol{h}(x^*_{T+j+1,T+j+1})'\boldsymbol{\beta}_f + \boldsymbol{s}_{f,\boldsymbol{D}_{T+j}}(x^*_{T+j+1,T+j+1})'\boldsymbol{A}^{-1}_{f,D_{T+j}}(\boldsymbol{D}_{T+j} - \boldsymbol{H}_{D_{T+j}}\boldsymbol{\beta}_f) \quad (41)$$

and variance

$$\sigma^2_{y_{T+j+1}} = \sigma^2_\epsilon + \sigma^2_f \left\{1 - \boldsymbol{s}_{f,\boldsymbol{D}_{T+j}}(x^*_{T+j+1,T+j+1})'\boldsymbol{A}^{-1}_{f,\boldsymbol{D}_{T+j}}\boldsymbol{s}_{f,\boldsymbol{D}_{T+j}}(x^*_{T+j+1,T+j+1})\right\}. \quad (42)$$

6. BRIEF DISCUSSION ON SIMULATION STUDIES

In Section S-3 of the supplement we present two detailed simulation experiments. In the first experiment, both the true observational and the evolutionary functions are of linear forms. In the second study, both these true functions are non-linear. In both the cases we fitted our Gaussian process based nonparametric model to the data generated from the true models. Whenever the true parameters are comparable to the parameters associated with our model, the posteriors of the parameters of our model successfully captured them. The true time series $\{x_0, \ldots, x_T\}$ fell well within their respective 95% Bayesian credible intervals in both the simulation studies. However, the lengths of the credible regions of the states seem to be somewhat larger than desired. Indeed, since the posterior distribution of the states depend upon the observational and evolutionary functions, both of which are treated as unknown, somewhat larger credible intervals only reflect the uncertainty associated with these unknown functions. In the context of our simulation studies where the true functions are known, the lengths of our credible intervals may not seem to be particularly encouraging, but as we already mentioned in the Section 1, we have in mind complex, realistic problems,

where true observational and evolutionary functions are extremely difficult to ascertain. In such problems, there exist large amounts of uncertainties regarding these functions, which would be coherently modeled by our Gaussian process approach, and relatively large Bayesian credible regions, that would arise as a result of acknowledging such uncertainties, would be coherent and make good practical sense.

We now apply our ideas based on Gaussian processes to a real data set.

7. APPLICATION TO A REAL DATA SET

Assuming a parametric, dynamic, linear model set up, Shumway & Stoffer (1982) and Carlin et al. (1992) analysed a data set consisting of estimated total physician expenditures by year $(y_t; t = 1, \ldots, 25)$ as measured by the Social Security Administration. The unobserved true annual physician expenditures are denoted by $x_t; t = 1, \ldots, 25$. Shumway & Stoffer (1982) used a maximum likelihood approach based on the EM algorithm, while Carlin et al. (1992) considered a Gibbs sampling based Bayesian approach.

We apply our Bayesian nonparametric Gaussian process based model and methodology on the same data set and check if the results of the former authors who analysed this data based on the linearity assumption, agree with those obtained by our far more general analysis.

7.1 Choice of prior parameters and MCMC implementation

We use the same prior structures as detailed in Section 5. We set $\alpha = 4.01$ so that the prior variance is of the form $200a^2$; we set $a = 0.5, 0.5, 100, 100$, respectively, for σ_f^2, σ_g^2, σ_ϵ^2 and σ_η^2. These imply that the prior expectations of the variances are 0.5, 0.5, 100, and 100, respectively, along with the aforementioned variances. The choices of high values of the prior expectations of σ_ϵ^2 and σ_η are motivated by Carlin et al. (1992), while the small variabilities of σ_f^2 and σ_g^2 reflects the belief that the true functional forms are perhaps not very different from linearity, the belief being motivated by the assumption of linearity by both Shumway & Stoffer (1982) and Carlin et al. (1992). Also motivated by the latter we chose $x_0 \sim N(2500, 100^2)$ to be the prior distribution of x_0. We set the priors of $\boldsymbol{\beta}_f$ and $\boldsymbol{\beta}_g$ to be trivariate normal distributions with zero mean and the

identity matrix as the covariance matrix. For the prior parameters of the smoothness parameters we set $\sigma^2 = 100$ and $\mu = -50$ in the log-normal prior distributions, so that the means are 1 and the variances are $\exp(100) - 1$. Here we set high variance for the smoothness parameters to account for much higher degree of uncertainty about smoothness in this real data situation compared to the simulation experiment.

To set up the grid $\boldsymbol{G}_n$ we noted that the MLEs of the time series obtained by Shumway & Stoffer (1982) using linear state space model are contained in $[2000, 30000]$. We divide this interval into 200 sub-intervals of equal length, and select a value randomly from each such sub-interval, obtaining values of the second component of the two-dimensional grid $\boldsymbol{G}_n$. For the first component, we generate a number uniformly from each of the 200 sub-intervals $[i, i+1]$; $i = 0, \ldots, 199$.

We discard the first 10,000 MCMC iterations as burn-in and store the next 50,000 iterations for inference. We used the normal random walk proposal with variance 0.05 for updating σ_f, σ_g, σ_ϵ and σ_η and the normal random walk proposal with variance 10 for updating $\{x_0, x_1, \ldots, x_T\}$. These choices of the variances are based on pilot runs of our MCMC algorithm. As before, informal diagnostics indicated good convergence properties of our MCMC algorithm. It took around 17 hours to implement this application in an ordinary laptop machine.

7.2 Results of model-fitting

Figures 1, 2 and 3 show the posterior distributions of the unknowns. Our posterior time series of x_t has relatively narrow 95% highest posterior density credible intervals, vindicating that the linearity assumption is not unreasonable as claimed by Shumway & Stoffer (1982) and Carlin et al. (1992). Indeed, for such linear relationships, it is well-known that the Gaussian process priors that we adopted are expected to perform very well.

However, perhaps not all is well with the aforementioned linearity assumption of Shumway & Stoffer (1982) and Carlin et al. (1992). Note that although the MLE time series obtained by Shumway & Stoffer (1982) fall mostly within our Bayesian 95% highest posterior density credible intervals of x_t, five observations towards the end of the time series fall outside. These observations correspond to the years 1968, 1970, 1971, 1972, and 1973. The values of y_t in these years are

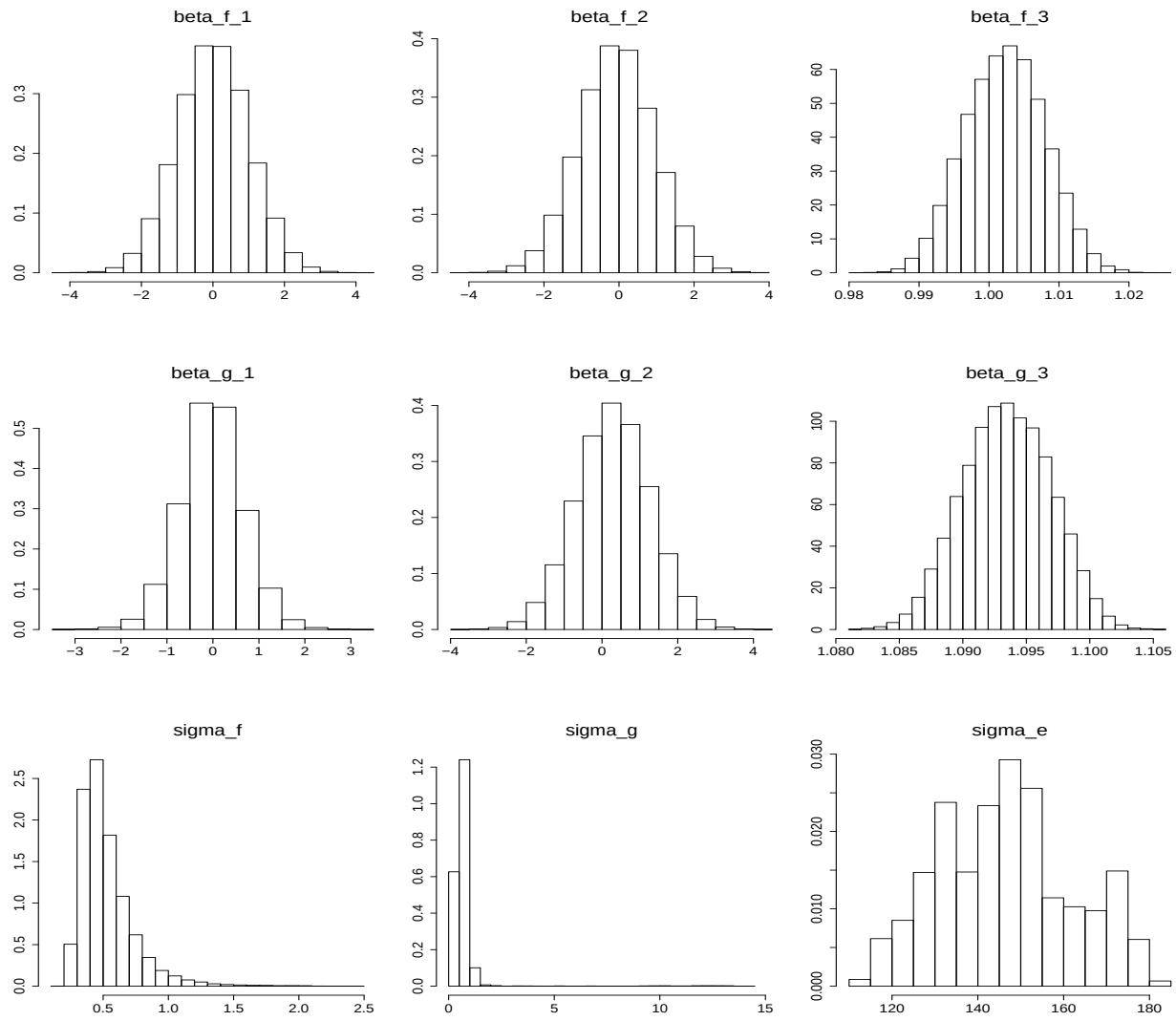

Figure 1: **Real data analysis**: Posterior densities of $\beta_{0,f}$, $\beta_{1,f}$, $\beta_{2,f}$, $\beta_{0,g}$, $\beta_{1,g}$, $\beta_{2,g}$, σ_f, σ_g, and σ_ϵ.

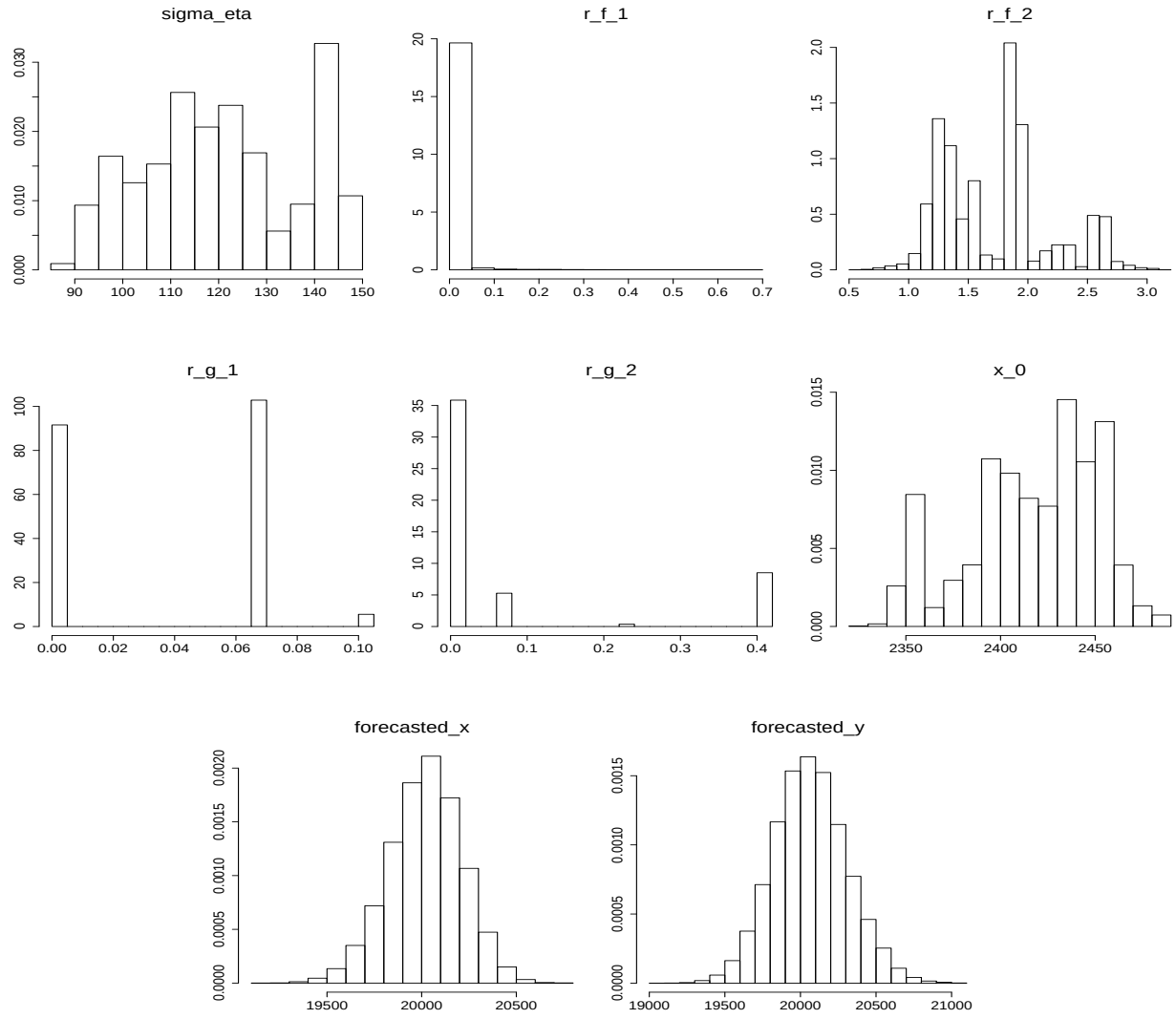

Figure 2: **Real data analysis**: Posterior densities of σ_η, $r_{1,f}$, $r_{2,f}$, $r_{1,g}$, $r_{2,g}$, x_0, x_{T+1} (one-step forecasted x), and y_{T+1} (one-step forecasted y).

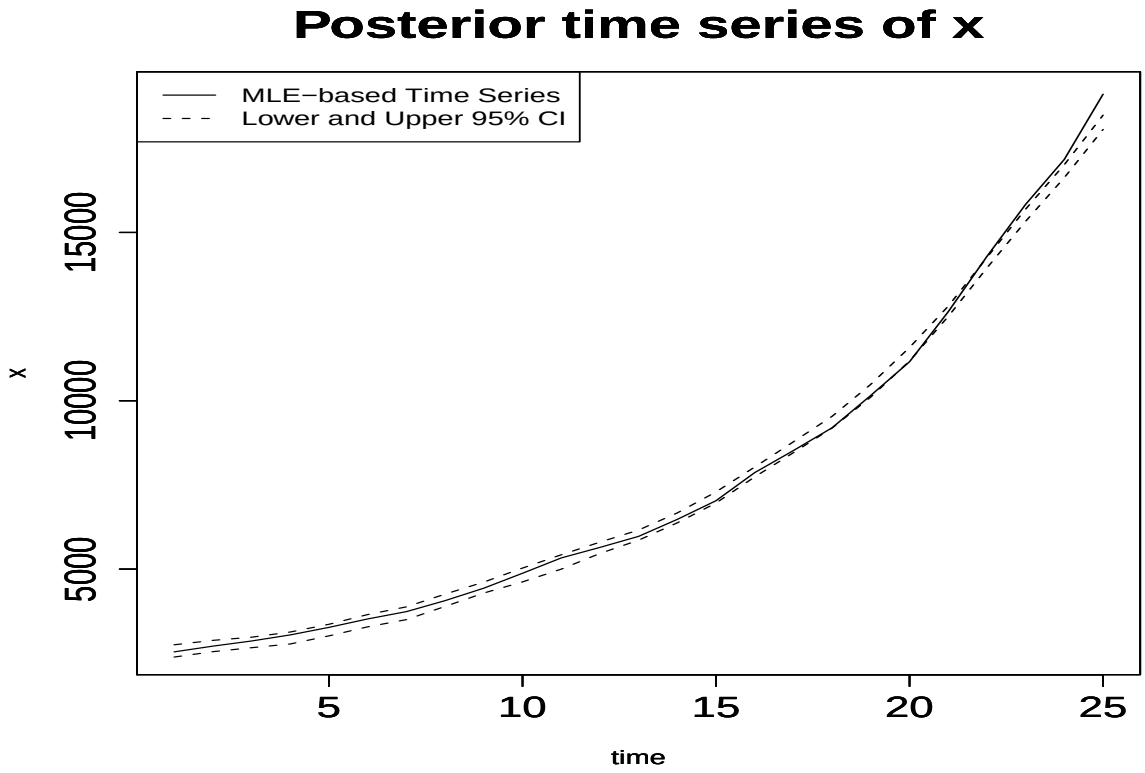

Figure 3: **Real data analysis**: 95% highest posterior density credible intervals of the time series $x_1, \ldots, x_T$. The solid line stands for the MLE time series obtained by Shumway and Stoffer (1982).

$11,099$, $14,306$, $15,835$, $16,916$, and $18,200$. We suspect that linearity breaks down towards the end of the time series, an issue which is perhaps overlooked by the linear model based approaches of Shumway & Stoffer (1982) and Carlin et al. (1992). On the other hand, without any change point analysis, our flexible nonparametric model, based on Gaussian processes, is able to accommodate changes in the regression structures.

8. CONCLUSIONS AND FUTURE WORK

In this article, using Gaussian process priors and the "look-up table" idea of Bhattacharya (2007) we proposed a novel methodology for Bayesian inference in nonparametric state space models, in both univariate and multivariate cases. The Gaussian process priors on the unknown functional forms of the observational and the evolutionary equations allow for very flexible modeling of time-varying random functions, where even the functional forms may change over time, without requiring any change point analysis. We have vindicated the effectiveness of our model and methodology with simulation experiments, and a real data analysis which provided interesting insight into nonlinearity of the underlying time series towards its end.

For our current purpose, we have assumed iid Gaussian noises ϵ_t and η_t; however, it is straightforward to generalize these to other parametric distributions (thick-tailed or otherwise). It may, however, be quite interesting to consider nonparametric error distributions, for example, mixtures of Dirichlet processes; such a work in the linear cases has been undertaken by Caron, Davy, Doucet, Duflos & Vanheeghe (2008). We shall also consider matrix-variate extensions to our current work, in addition to nonparametric error distributions.

ACKNOWLEDGMENT

We are extremely grateful to an anonymous reviewer whose suggestions resulted in improved presentation of our ideas.

DESCRIPTION OF THE SUPPLEMENT

In Section S-1 of the supplement we state and prove results on the accuracy of the look-up table idea and in Section S-2 we provide details of MCMC sampling in the univariate situation. In Section S-3 two simulation studies are considered in details. We provide extension of our Gaussian process based methodology for univariate cases to multivariate situations in Section S-4; followed by details of MCMC sampling in the multivariate set-up in Section S-5. In Section S-6 we conduct a detailed simulation study to illustrate our methodology in the multivariate set-up. Finally, in Section S-7 we provide detailed discussion of posteriors of the composite observational and evolutionary functions and their comparisons with the respective true composite functions; in particular, we provide the comparisons in the case of the three simulation studies (two univariate cases, and one multivariate case).

REFERENCES

Bhattacharya, S. (2007), "A Simulation Approach to Bayesian Emulation of Complex Dynamic Computer Models," *Bayesian Analysis*, 2, 783–816.

Brockwell, P. J., & Davis, R. A. (1987), *Time Series: Theory and Methods*, New York: Springer-Verlag.

Carlin, B. P., Polson, N. G., & Stoffer, D. S. (1992), "A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling," *Journal of the American Statistical Association*, 87, 493–500.

Caron, F., Davy, M., Doucet, A., Duflos, E., & Vanheeghe, P. (2008), "Bayesian Inference for Linear Dynamic Models with Dirichlet Process Mixtures," *IEEE Transactions on Signal Processing*, 56, 71–84.

Carvalho, C. M., & West, M. (2007), "Dynamic Matrix-Variate Graphical Models," *Bayesian Analysis*, 2, 69–98.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: Wiley.

Durbin, J., & Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford: Oxford University Press.

Ghosh, A., Mukhopadhyay, S., Roy, S., & Bhattacharya, S. (2013), "Supplement to "Bayesian Inference in Nonparametric Dynamic State-Space Models",". Submitted.

Meinhold, R. J., & Singpurwala, N. D. (1989), "Robustification of Kalman Filter Models," *Journal of the American Statistical Association*, 84, 479–486.

Quintana, J., & West, M. (1987), "Multivariate time series analysis: New techniques applied to international exchange rate data," *The Statistician*, 36, 275–281.

Santner, T. J., Williams, B. J., & Notz, W. I. (2003), *The design and analysis of computer experiments*, Springer Series in Statistics, New York, Inc.: Springer-Verlag.

Shumway, R. H., & Stoffer, D. S. (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis*, 3, 253–264.

Shumway, R. H., & Stoffer, D. S. (2011), *Time Series Analysis and Its Applications*, New York: Springer-Verlag.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York, Inc: Springer-Verlag.

West, M., & Harrison, P. (1997), *Bayesian Forecasting and Dynamic Models*, New York: Springer-Verlag.

arXiv:1108.3262v5 [stat.ME] 21 Feb 2014

Supplement to "Bayesian Inference in Nonparametric Dynamic State-Space Models"

Anurag Ghosh, Soumalya Mukhopadhyay, Sandipan Roy and Sourabh Bhattacharya*

Throughout, we refer to our main paper Ghosh, Mukhopadhyay, Roy & Bhattacharya (2013) as GMRB.

S-1. RESULT ON THE ACCURACY OF THE MARKOV APPROXIMATION OF THE CONDITIONAL DISTRIBUTION OF X_T GIVEN $\boldsymbol{D}_N^*$

Theorem S-1.1 *Let $g^{(t)}(\cdot)$ denote composition applied to $g(\cdot)$ with itself t $(t \geq 1)$ times, where $g^{(1)}(\cdot) = g(\cdot)$. Let $g_{true}^{(t)}(\cdot)$ be defined analogously. Let the inputs $\{z_1, \ldots, z_n\}$ satisfy $z_i - z_{i-1} = (b-a)/n$ for $i = 2, 3, \ldots, n$, where a, b are finite constants such that $a < \min\{z_1, \ldots, z_n\} < \min\{z_1, \ldots, z_n\} < b$. Also assume that $\boldsymbol{G}_n$ is of the form $\{z_1, \ldots, z_n, g_{true}(z_1), \ldots, g_{true}(z_n), g_{true}^{(2)}(z_1), \ldots, g_{true}^{(2)}(z_n), \ldots, g_{true}^{(T-1)}(z_1), \ldots, g_{true}^{(T-1)}(z_n)\}$, so that $\boldsymbol{D}_n^* = \{g(z_1), \ldots, g(z_n), g(g_{true}(z_1)), \ldots, g(g_{true}(z_n)), g(g_{true}^{(2)}(z_1)), \ldots, g(g_{true}^{(2)}(z_n)), \ldots, g(g_{true}^{(T-1)}(z_1)), \ldots, g(g_{true}^{(T-1)}(z_n))\}$. Then, for $t = 1, \ldots, T$, for $r > 0$,*

$$\int_a^b \left| E[g^{(t)}(z)|\boldsymbol{D}_n^*, \boldsymbol{\theta}_g] - g_{true}^{(t)}(z) \right|^r dz = O\left(n^{-1}\right). \tag{1}$$

Also, for any $z^* \in [a, b]$,

$$\left| E[g^{(t)}(z^*)|\boldsymbol{D}_n^*, \boldsymbol{\theta}_g] - g_{true}^{(t)}(z^*) \right| = O\left(n^{-1}\right), \tag{2}$$

and

$$Var[g^{(t)}(z^*)|\boldsymbol{D}_n^*, \boldsymbol{\theta}_g] = O\left(n^{-1}\right). \tag{3}$$

*Anurag Ghosh is a PhD student in Department of Statistical Science, Duke University, Soumalya Mukhopadhyay is a PhD student in Agricultural and Ecological Research Unit, Indian Statistical Institute, Sandipan Roy is a PhD student in Department of Statistics, University of Michigan, Ann Arbor, and Sourabh Bhattacharya is an Assistant Professor in Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108. Corresponding e-mail: sourabh@isical.ac.in.

Proof:

Note that for any $r > 0$, we have

$$\int_a^b \left| E[g^{(t)}(z)|\boldsymbol{D}_n^*, \boldsymbol{\theta}_g] - g_{true}^{(t)}(z) \right|^r dz = \frac{b-a}{n} \sum_{i=1}^{n} \left| E[g^{(t)}(z_i)|\boldsymbol{D}_n^*, \boldsymbol{\theta}_g] - g_{true}^{(t)}(z_i) \right|^r + O\left(n^{-1}\right),$$

by Riemann sum approximation. But by the choice of $\boldsymbol{G}_n$ and the interpolation property of Gaussian processes, it follows that, given $\boldsymbol{D}_n^*$, $g(z_i) = g_{true}(z_i)$ with probability one. Hence, given $\boldsymbol{D}_n^*$, $g(g(z_i)) = g(g_{true}(z_i)) = g_{true}(g_{true}(z_i))$ with probability 1 since $g_{true}(z_i)$ is in $\boldsymbol{G}_n$, and so on. In general, given $\boldsymbol{D}_n^*$, $g^{(t)}(z_i) = g_{true}^{(t)}(z_i)$ with probability 1. Hence, we have $\left| E[g^{(t)}(z_i)|\boldsymbol{D}_n^*, \boldsymbol{\theta}_g] - g_{true}^{(t)}(z_i) \right| = 0$ for $k = 1, \ldots, n$. The result (1) thus follows.

To see (2) note that if $z^* \in \boldsymbol{G}_n$, then $\left| E[g^{(t)}(z^*)|\boldsymbol{D}_n^*, \boldsymbol{\theta}_g] - g_{true}^{(t)}(z^*) \right| = 0$ by the interpolation property of Gaussian processes, so that (2) trivially holds. If $z^* \in [a, b]$ is not a design point in $\boldsymbol{G}_n$, then there exists $z_i \in \boldsymbol{G}_n$ such that $z^* \in [z_i, z_{i+1}]$. Hence, $z^* = z_i + h$, where $h < (b-a)/n$. Then, letting $\nu(\cdot) = E[g^{(t)}(\cdot)|\boldsymbol{D}_n^*, \boldsymbol{\theta}_g]$ and $\gamma(\cdot) = g_{true}^{(t)}(\cdot)$ we have

$$\nu(z_i + h) = \nu(z_i) + h\nu'(\xi_1) \quad \text{and} \tag{4}$$

$$\gamma(z_i + h) = \gamma(z_i) + h\gamma'(\xi_2), \tag{5}$$

where ξ_1 and ξ_2 lie between z_i and $z_i + h$. Since $\nu(z_i) = \gamma(z_i)$ and since $\nu'(\cdot)$ and $\gamma'(\cdot)$ are bounded on $[a, b]$ (as the exponential correlation function of our Gaussian process ensures that ν and γ are continuously differentiable), we have, using (4) and (5),

$$\left| E[g^{(t)}(z^*)|\boldsymbol{D}_n^*, \boldsymbol{\theta}_g] - g_{true}^{(t)}(z^*) \right| = |\nu(z_i + h) - \gamma(z_i + h)| = O(h).$$

Since $h < (b-a)/n$, result (2) follows.

To prove (3), let $\varphi(\cdot) = Var[g^{(t)}(\cdot)|\boldsymbol{D}_n^*, \boldsymbol{\theta}_g]$. Then

$$\varphi(z_i + h) = \varphi(z_i) + h\varphi'(\xi_3),$$

where ξ_3 lies between z_i and $z_i + h$. But $\varphi(z_i) = 0$ due to the interpolation property of Gaussian processes, and $\varphi'(\cdot)$ is bounded on $[a, b]$ (again, the exponential correlation function ensuring that φ is continuously differentiable). Moreover, since $h < (b-a)/n$, the result (3) follows. □

S-2. DETAILS OF MCMC SAMPLING IN THE UNIVARIATE SITUATION

Let $\boldsymbol{G}_{n+1} = \boldsymbol{G}_n \cup \{x^*_{1,0}\}$, $\boldsymbol{D}^*_{n+1} = \left(\boldsymbol{D}^{*\prime}_n, g(x^*_{1,0})\right)'$, $\boldsymbol{A}_{g,n+1} = \begin{pmatrix} \boldsymbol{A}_{g,D^*_n} & \boldsymbol{s}_{g,D^*_n}(x^*_{1,0}) \\ \boldsymbol{s}_{g,D^*_n}(x^*_{1,0})' & 1 \end{pmatrix}$ and $\boldsymbol{H}'_{g,D^*_{n+1}} = [\boldsymbol{H}'_{g,D^*_n}, \boldsymbol{h}(x^*_{1,0})]$. Clearly, $[\boldsymbol{D}^*_{n+1} \mid x^*_{1,0}, \boldsymbol{\theta}_g] = [g(x^*_{1,0}) \mid x_0, \boldsymbol{\theta}_g][\boldsymbol{D}^*_n \mid g(x^*_{1,0}), x_0, \boldsymbol{\theta}_g]$. With these definitions, the forms of the full conditional distributions of the unknowns are provided below.

$$[\boldsymbol{\beta}_f \mid \cdots] \propto [\boldsymbol{\beta}_f][\boldsymbol{D}_T \mid x_1, \ldots, x_T, \boldsymbol{\theta}_f, \sigma^2_\epsilon] \tag{6}$$

$$[\boldsymbol{\beta}_g \mid \cdots] \propto [\boldsymbol{\beta}_g][\boldsymbol{D}^*_{n+1} \mid x^*_{1,0}, \boldsymbol{\theta}_g] \prod_{t=1}^{T} [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_t] \tag{7}$$

$$[\sigma^2_f \mid \cdots] \propto [\sigma^2_f][\boldsymbol{D}_T \mid x_1, \ldots, x_T, \boldsymbol{\theta}_f, \sigma^2_\epsilon] \tag{8}$$

$$[\sigma^2_g \mid \cdots] \propto [\sigma^2_g][\boldsymbol{D}^*_{n+1} \mid x^*_{1,0}, \boldsymbol{\theta}_g] \prod_{t=1}^{T} [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_t] \tag{9}$$

$$[\sigma^2_\epsilon \mid \cdots] \propto [\sigma^2_\epsilon][\boldsymbol{D}_T \mid x_1, \ldots, x_T, \boldsymbol{\theta}_f, \sigma^2_\epsilon] \tag{10}$$

$$[\sigma^2_\eta \mid \cdots] \propto [\sigma^2_\eta][x_1 \mid g(x^*_{1,0}), x_0, \boldsymbol{\theta}_g, \sigma^2_\eta] \prod_{t=1}^{T} [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_t] \tag{11}$$

$$[r_{i,f} \mid \cdots] \propto [r_{i,f}][\boldsymbol{D}_T \mid x_1, \ldots, x_T, \boldsymbol{\theta}_f, \sigma^2_\epsilon]; \quad i = 1, 2 \tag{12}$$

$$[r_{i,g} \mid \cdots] \propto [r_{i,g}][\boldsymbol{D}^*_n \mid g(x^*_{1,0}), x_0, \boldsymbol{\theta}_g] \prod_{t=1}^{T} [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_t]; \quad i = 1, 2 \tag{13}$$

$$[g(x^*_{1,0}) \mid \cdots] \propto [g(x^*_{1,0}) \mid x_0, \boldsymbol{\beta}_g, \sigma^2_g][\boldsymbol{D}^*_n \mid g(x^*_{1,0}), x_0, \boldsymbol{\theta}_g][x_1 \mid g(x^*_{1,0}), x_0, \sigma^2_\eta] \tag{14}$$

$$[\boldsymbol{D}^*_n \mid \cdots] \propto \prod_{t=1}^{T} [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_t][\boldsymbol{D}^*_n \mid g(x^*_{1,0}), x_0, \boldsymbol{\theta}_g] \tag{15}$$

$$[x_0 \mid \cdots] \propto [x_0][\boldsymbol{D}^*_{n+1} \mid x_0, \boldsymbol{\theta}_g] \tag{16}$$

$$\begin{aligned}[x_1 \mid \cdots] \propto\ & [x_1 \mid g(x^*_{1,0}), x_0, \boldsymbol{\beta}_g, \sigma^2_g, \sigma^2_\eta][x_2 \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_1] \\ & \times [\boldsymbol{D}_T \mid x_1, \ldots, x_T, \boldsymbol{\theta}_f, \sigma^2_\epsilon]\end{aligned} \tag{17}$$

$$\begin{aligned}[x_{t+1} \mid \cdots] \propto\ & [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_t][x_{t+2} \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_{t+1}] \\ & \times [\boldsymbol{D}_T \mid x_1, \ldots, x_T, \boldsymbol{\theta}_f, \sigma^2_\epsilon]; \quad t = 1, \ldots, T-1\end{aligned} \tag{18}$$

$$[x_{T+1} \mid \cdots] \propto [x_{T+1} \mid \boldsymbol{\theta}_g, \sigma^2_\eta, \boldsymbol{D}^*_n, x_T] \tag{19}$$

Although some of the full conditionals are of standard forms permitting Gibbs sampling steps, others are non-standard and sampling requires Metropolis-Hastings (MH) steps in those situations. We describe below the Gibbs steps and construct proposal distributions when MH steps are needed.

S-2.1 Updating $\boldsymbol{\beta}_f$ using Gibbs step

The full conditional of $\boldsymbol{\beta}_f$ is m-variate normal with mean

$$E\left[\boldsymbol{\beta}_f \mid \cdots\right] = \left\{\boldsymbol{H}'_{D_T}\left(\sigma_f^2 \boldsymbol{A}_{f,D_T} + \sigma_\epsilon^2 \boldsymbol{I}\right)^{-1} \boldsymbol{H}_{D_T} + \boldsymbol{\Sigma}_{\beta_f,0}^{-1}\right\}^{-1} \times \left\{\boldsymbol{H}'_{D_T}\left(\sigma_f^2 \boldsymbol{A}_{f,D_T} + \sigma_\epsilon^2 \boldsymbol{I}\right)^{-1} \boldsymbol{D}_T + \boldsymbol{\Sigma}_{\beta_f,0}^{-1}\boldsymbol{\beta}_{f,0}\right\}. \tag{20}$$

and variance

$$V\left[\boldsymbol{\beta}_f \mid \cdots\right] = \left\{\boldsymbol{H}'_{D_T}\left(\sigma_f^2 \boldsymbol{A}_{f,D_T} + \sigma_\epsilon^2 \boldsymbol{I}\right)^{-1} \boldsymbol{H}_{D_T} + \boldsymbol{\Sigma}_{\beta_f,0}^{-1}\right\}^{-1}. \tag{21}$$

S-2.2 Updating $\boldsymbol{\beta}_g$ using Gibbs step

The conditional distribution of $\boldsymbol{\beta}_g$ is m-variate normal with mean

$$\begin{aligned} &E\left[\boldsymbol{\beta}_g \mid \cdots\right] \\ &= \left\{\boldsymbol{\Sigma}_{\beta_g,0}^{-1} + \frac{\boldsymbol{H}'_{D^*_{n+1}} \boldsymbol{A}_{g,D^*_{n+1}}^{-1} \boldsymbol{H}_{D^*_{n+1}}}{\sigma_g^2} \right. \\ &\quad \left. + \sum_{t=1}^{T} \frac{\left(\boldsymbol{H}'_{D^*_n} \boldsymbol{A}_{g,D^*_n}^{-1} \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t}) - \boldsymbol{h}(x^*_{t+1,t})\right)\left(\boldsymbol{H}'_{D^*_n} \boldsymbol{A}_{g,D^*_n}^{-1} \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t}) - \boldsymbol{h}(x^*_{t+1,t})\right)'}{\sigma_{g,\eta,t}^2} \right\}^{-1} \\ &\quad \times \left\{\boldsymbol{\Sigma}_{\beta_g,0}^{-1}\boldsymbol{\beta}_{g,0} + \frac{\boldsymbol{H}'_{D^*_{n+1}} \boldsymbol{A}_{g,D^*_{n+1}}^{-1} \boldsymbol{D}^*_{n+1}}{\sigma_g^2} \right. \\ &\quad \left. + \sum_{t=1}^{T} \frac{\left(x_{t+1} - \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t})' \boldsymbol{A}_{g,D^*_n}^{-1} \boldsymbol{D}^*_n\right)\left(\boldsymbol{h}(x^*_{t+1,t}) - \boldsymbol{H}'_{D^*_n} \boldsymbol{A}_{g,D^*_n}^{-1} \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t})\right)}{\sigma_{g,\eta,t}^2} \right\}. \end{aligned} \tag{22}$$

and variance

$$
\begin{aligned}
&V\left[\boldsymbol{\beta}_g \mid \cdots\right] \\
&= \left\{ \boldsymbol{\Sigma}^{-1}_{\beta_g,0} + \frac{\boldsymbol{H}'_{D^*_{n+1}} \boldsymbol{A}^{-1}_{g,D^*_{n+1}} \boldsymbol{H}_{D^*_{n+1}}}{\sigma^2_g} \right. \\
&\quad \left. + \sum_{t=1}^{T} \frac{\left(\boldsymbol{H}'_{D^*_n} \boldsymbol{A}^{-1}_{g,D^*_n} \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t}) - \boldsymbol{h}(x^*_{t+1,t})\right)\left(\boldsymbol{H}'_{D^*_n} \boldsymbol{A}^{-1}_{g,D^*_n} \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t}) - \boldsymbol{h}(x^*_{t+1,t})\right)'}{\sigma^2_{g,\eta,t}} \right\}^{-1}
\end{aligned} \tag{23}
$$

In (22) and (23),

$$
\sigma^2_{g,\eta,t} = \sigma^2_g \left\{ 1 - \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t})' \boldsymbol{A}^{-1}_{g,D^*_n} \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t}) \right\} + \sigma^2_\eta. \tag{24}
$$

S-2.3 Updating σ^2_f and σ^2_g using MH steps

The full conditionals of σ^2_f and σ^2_g are not available in closed forms and MH steps are necessary here. For proposal distributions we first construct a new model by setting $\sigma^2_\epsilon = \sigma^2_f$ and $\sigma^2_\eta = \sigma^2_g$. For this model the full conditional distributions of σ^2_f and σ^2_g are inverse Gamma distributions, given by

$$
q_{\sigma^2_f}(\sigma^2_f) \propto \left(\sigma^2_f\right)^{-\left(\frac{T+\alpha_f+2}{2}\right)} \exp\left[-\frac{1}{2\sigma^2_f} \left\{ \gamma_f + \left(\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{\beta}_f\right)' \left(\boldsymbol{A}_{f,D_T} + \boldsymbol{I}_T\right)^{-1} \left(\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{\beta}_f\right) \right\} \right] \tag{25}
$$

and

$$
\begin{aligned}
q_{\sigma^2_g}(\sigma^2_g) \propto \left(\sigma^2_g\right)^{-\left(\frac{\alpha_g+4+n+T}{2}\right)} \exp\Bigg[-\frac{1}{2\sigma^2_g} \Big\{ &\gamma_g + (x_1 - g(x^*_{1,0}))^2 \\
&+ \sum_{t=1}^{T} \frac{\left\{ x_{t+1} - \boldsymbol{h}(x^*_{t+1,t})'\boldsymbol{\beta}_g - \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t})' \boldsymbol{A}^{-1}_{g,D^*_n} \left(\boldsymbol{D}^*_n - \boldsymbol{H}_{D^*_n}\boldsymbol{\beta}_g\right) \right\}^2}{\sigma^2_{g,t}} \\
&+ \left(\boldsymbol{D}^*_{n+1} - \boldsymbol{H}_{D^*_{n+1}}\boldsymbol{\beta}_g\right)' \boldsymbol{A}^{-1}_{g,D^*_{n+1}} \left(\boldsymbol{D}^*_{n+1} - \boldsymbol{H}_{D^*_{n+1}}\boldsymbol{\beta}_g\right) \Big\} \Bigg]
\end{aligned} \tag{26}
$$

In (26),

$$
\sigma^2_{g,t} = 2 - \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t})' \boldsymbol{A}^{-1}_{g,D^*_n} \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t}). \tag{27}
$$

In (26) the term $\left(\sigma_g^2\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma_g^2}(x_1 - g(x_{1,0}^*))^2\right\}$ is also taken into account since, given $\sigma_\eta^2 = \sigma_g^2$, $[x_1 \mid g(x_{1,0}^*), x_0] \sim N(g(x_{1,0}^*), \sigma_g^2)$. It is useful to remark here that unless $\sigma_f \approx \sigma_\epsilon$ and $\sigma_g \approx \sigma_\eta$ these proposal mechanisms may not be efficient. We shall discuss other proposal distributions in the context of applications.

S-2.4 Updating σ_ϵ^2 and σ_η^2 using MH steps

As before, the full conditionals of σ_ϵ^2 and σ_η^2 are not available in closed forms. For MH steps, we construct proposal distributions obtained by setting $\sigma_f^2 = \sigma_\epsilon^2$ and $\sigma_g^2 = \sigma_\eta^2$. The proposal distributions are given by

$$q_{\sigma_\epsilon^2}(\sigma_\epsilon^2) \propto \left(\sigma_\epsilon^2\right)^{-\left(\frac{T+\alpha_\epsilon+2}{2}\right)} \exp\left[-\frac{1}{2\sigma_\epsilon^2}\left\{\gamma_\epsilon + \left(\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{\beta}_f\right)' \left(\boldsymbol{A}_{f,D_T} + \boldsymbol{I}_T\right)^{-1} \left(\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{\beta}_f\right)\right\}\right] \tag{28}$$

$$\begin{aligned} q_{\sigma_\eta^2}(\sigma_\eta^2) \propto \left(\sigma_\eta^2\right)^{-\left(\frac{\alpha_\eta+4+n+T}{2}\right)} \exp\Bigg[-\frac{1}{2\sigma_\eta^2}\Bigg\{&\gamma_\eta + (x_1 - g(x_{1,0}^*))^2 \\ &+ \left(\boldsymbol{D}_{n+1}^* - \boldsymbol{H}_{D_{n+1}^*}\boldsymbol{\beta}_g\right)' \boldsymbol{A}_{g,D_{n+1}^*}^{-1} \left(\boldsymbol{D}_{n+1}^* - \boldsymbol{H}_{D_{n+1}^*}\boldsymbol{\beta}_g\right) \\ &+ \sum_{t=1}^{T} \frac{\left\{x_{t+1} - \boldsymbol{h}(x_{t+1,t}^*)'\boldsymbol{\beta}_g - \boldsymbol{s}_{g,D_n^*}(x_{t+1,t}^*)'\boldsymbol{A}_{g,D_n^*}^{-1}\left(\boldsymbol{D}_n^* - \boldsymbol{H}_{D_n^*}\boldsymbol{\beta}_g\right)\right\}^2}{\sigma_{g,t}^2}\Bigg\}\Bigg] \end{aligned} \tag{29}$$

In (29) the term $\left(\sigma_\eta^2\right)^{-(n+1)/2} \exp\left\{-\frac{1}{2\sigma_\eta^2}\left(\boldsymbol{D}_{n+1}^* - \boldsymbol{H}_{D_{n+1}^*}\boldsymbol{\beta}_g\right)' \boldsymbol{A}_{g,D_{n+1}^*}^{-1}\left(\boldsymbol{D}_{n+1}^* - \boldsymbol{H}_{D_{n+1}^*}\boldsymbol{\beta}_g\right)\right\}$ occurs since, given $\sigma_g^2 = \sigma_\eta^2$, $[\boldsymbol{D}_{n+1}^* \mid \boldsymbol{\beta}_g, \sigma_g^2, \boldsymbol{R}_g] \sim N_{n+1}\left(\boldsymbol{H}_{D_{n+1}^*}\boldsymbol{\beta}_g, \sigma_\eta^2 \boldsymbol{A}_{g,D_{n+1}^*}\right)$. Again, these proposals need not be efficient unless $\sigma_f \approx \sigma_\epsilon$ and $\sigma_g \approx \sigma_\eta$. In the context of specific applications we shall discuss other proposal distributions.

S-2.5 Updating $\boldsymbol{R}_f$ and $\boldsymbol{R}_g$ using MH steps

For $i = 1, 2$, the full conditionals of the smoothness parameters $r_{i,f}$ and $r_{i,g}$ are not available in closed forms, and we suggest MH steps with normal random walk proposals with adequately optimized variances.

S-2.6 Updating $g(x^*_{1,0})$ using Gibbs step

The full conditional of $g(x^*_{1,0})$ is univariate normal with mean and variance given, respectively, by

$$E[g(x^*_{1,0}) \mid \cdots] = \left\{ \frac{1}{\sigma^2_\eta} + \frac{1 + \boldsymbol{s}_{g,D^*_n}(x^*_{1,0})' \boldsymbol{\Sigma}^{-1}_{g,D^*_n} \boldsymbol{s}_{g,D^*_n}(x^*_{1,0})}{\sigma^2_g} \right\}^{-1} \times \left\{ \frac{x_1}{\sigma^2_\eta} + \frac{\boldsymbol{h}(x^*_{1,0})' \boldsymbol{\beta}_g + \boldsymbol{s}_{g,D^*_n}(x^*_{1,0})' \boldsymbol{\Sigma}^{-1}_{g,D^*_n} \boldsymbol{D}^*_z}{\sigma^2_g} \right\} \tag{30}$$

and

$$V[g(x^*_{1,0}) \mid \cdots] = \left\{ \frac{1}{\sigma^2_\eta} + \frac{1 + \boldsymbol{s}_{g,D^*_n}(x^*_{1,0})' \boldsymbol{\Sigma}^{-1}_{g,D^*_n} \boldsymbol{s}_{g,D^*_n}(x^*_{1,0})}{\sigma^2_g} \right\}^{-1} \tag{31}$$

In (30),

$$\boldsymbol{D}^*_z = \boldsymbol{D}^*_n - \boldsymbol{H}_{D^*_n} \boldsymbol{\beta}_g + \boldsymbol{s}_{g,D^*_n}(x^*_{1,0}) \boldsymbol{h}(x^*_{1,0})' \boldsymbol{\beta}_g \tag{32}$$

S-2.7 Updating $\boldsymbol{D}^*_n$ using Gibbs step

The full conditional distribution of $\boldsymbol{D}^*_n$ is n-variate normal with mean

$$E\left[\boldsymbol{D}^*_n \mid \cdots\right] = \left\{ \frac{\boldsymbol{\Sigma}^{-1}_{g,D^*_n}}{\sigma^2_g} + \boldsymbol{A}^{-1}_{g,D^*_n} \left(\sum_{t=1}^{T} \frac{\boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t}) \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t})'}{\sigma^2_{g,\eta,t}} \right) \boldsymbol{A}^{-1}_{g,D^*_n} \right\}^{-1} \times \left\{ \frac{\boldsymbol{\Sigma}^{-1}_{g,D^*_n} \boldsymbol{\mu}_{g,D^*_n}}{\sigma^2_g} + \boldsymbol{A}^{-1}_{g,D^*_n} \sum_{t=1}^{T} \frac{\boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t}) \{ x_{t+1} - \boldsymbol{\beta}'_g (\boldsymbol{h}(x^*_{t+1,t}) - \boldsymbol{H}'_{D^*_n} \boldsymbol{A}^{-1}_{g,D^*_n} \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t})) \}}{\sigma^2_{g,\eta,t}} \right\} \tag{33}$$

and variance

$$V\left[\boldsymbol{D}^*_n \mid \cdots\right] = \left\{ \frac{\boldsymbol{\Sigma}^{-1}_{g,D^*_n}}{\sigma^2_g} + \boldsymbol{A}^{-1}_{g,D^*_n} \left(\sum_{t=1}^{T} \frac{\boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t}) \boldsymbol{s}_{g,D^*_n}(x^*_{t+1,t})'}{\sigma^2_{g,\eta,t}} \right) \boldsymbol{A}^{-1}_{g,D^*_n} \right\}^{-1} \tag{34}$$

In (33) and (34), $\boldsymbol{\mu}_{g,D^*_n}$ and $\boldsymbol{\Sigma}_{g,D^*_n}$ are given by (9) and (10), respectively, of GMRB.

S-2.8 Updating x_0 using MH step

Let $\boldsymbol{\beta}_f = (\boldsymbol{\beta}'_{0,f}, \beta_{1,f})'$ and $\boldsymbol{\beta}_g = (\boldsymbol{\beta}'_{0,g}, \beta_{1,g})'$, where $\boldsymbol{\beta}_{0,f}$ and $\boldsymbol{\beta}_{0,g}$ are two-component vectors. Assuming the prior of x_0 to be normal with mean μ_{x_0} and variance $\sigma^2_{x_0}$, and setting $\sigma^2_g = 0$, the full conditional distribution of x_1 is univariate normal, with mean and variance given, respectively, by

$$\xi_0 = \left(\frac{1}{\sigma^2_{x_0}} + \frac{\beta^2_{1,g}}{\sigma^2_\eta}\right)^{-1} \left\{\frac{\mu_{x_0}}{\sigma^2_{x_0}} + \frac{(x_1 - \boldsymbol{k}'_1\boldsymbol{\beta}_{0,g})\beta_{1,g}}{\sigma^2_\eta}\right\}, \tag{35}$$

$$\phi^2_0 = \left(\frac{1}{\sigma^2_{x_0}} + \frac{\beta^2_{1,g}}{\sigma^2_\eta}\right)^{-1}. \tag{36}$$

In (35), for any t, $\boldsymbol{k}_t = (1, t)'$. We use $q_{x_0}(x_0) \equiv N(x_0 : \xi_0, \phi^2_0)$ as the proposal distribution for updating x_0 using MH step. Observe that, under $\sigma^2_g = 0$, $g(x^*_{1,0}) = \boldsymbol{h}(x^*_{1,0})'\boldsymbol{\beta}_g$ with probability one; hence $[x_1 \mid g(x^*_{1,0}), x_0, \boldsymbol{\beta}_g, \sigma^2_\eta] \sim N(\boldsymbol{h}(x^*_{1,0})'\boldsymbol{\beta}_g, \sigma^2_\eta)$, which has been taken into account while constructing the above proposal distribution. Note that this proposal will only be efficient when the $g(\cdot, \cdot)$ is close to linear. As a result, for non-linear applications, we shall often use other proposal mechanisms, such as the normal random walk.

S-2.9 Updating $\{x_1, \ldots, x_T\}$ using MH steps

We construct proposal distributions for simulating $\{x_1, \ldots, x_T\}$ based on linear observational and evolutionary equations, setting $\sigma^2_f = \sigma^2_g = 0$. Thus, for $t = 0, \ldots, T-1$, the proposal distributions of x_{t+1} are of the form $q_{x_{t+1}}(x_{t+1}) \equiv N(x_{t+1} : \xi_{t+1}, \phi_{t+1})$, that is, a normal distribution with mean ξ_{t+1} and variance ϕ^2_{t+1}, where the latter quantities are given by

$$\xi_{t+1} = \left(\frac{1 + \beta^2_{1,g}}{\sigma^2_\eta} + \frac{\beta^2_{1,f}}{\sigma^2_\epsilon}\right)^{-1} \left\{\frac{\boldsymbol{k}'_{t+1}\boldsymbol{\beta}_{0,g} + \beta_{1,g}x_t + (x_{t+2} - \boldsymbol{k}'_{t+2}\boldsymbol{\beta}_{0,g})\beta_{1,g}}{\sigma^2_\eta} + \frac{(y_{t+1} - \boldsymbol{k}'_{t+1}\boldsymbol{\beta}_{0,f})\beta_{1,f}}{\sigma^2_\epsilon}\right\} \tag{37}$$

$$\phi^2_{t+1} = \left(\frac{1 + \beta^2_{1,g}}{\sigma^2_\eta} + \frac{\beta^2_{1,f}}{\sigma^2_\epsilon}\right)^{-1} \tag{38}$$

These proposal mechanisms will be efficient only if both $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$ are close to linear. Hence, for non-linear situations, we shall consider other proposal distributions.

S-2.10 Updating x_{T+1} using Gibbs step

The full conditional distribution of x_{T+1} is normal with mean and variance given, respectively, by

$$E\left[x_{T+1} \mid \cdots\right] = \boldsymbol{h}(x^*_{T+1,T})'\boldsymbol{\beta}_g + \boldsymbol{s}_{g,\boldsymbol{D}^*_n}(x^*_{T+1,T})'\boldsymbol{A}^{-1}_{g,D^*_n}(\boldsymbol{D}^*_n - \boldsymbol{H}_{D^*_n}\boldsymbol{\beta}_g) \tag{39}$$

and variance

$$V\left[x_{T+1} \mid \cdots\right] = \sigma^2_\eta + \sigma^2_g\left\{1 - \boldsymbol{s}_{g,\boldsymbol{D}^*_n}(x^*_{T+1,T})'\boldsymbol{A}^{-1}_{g,\boldsymbol{D}^*_n}\boldsymbol{s}_{g,\boldsymbol{D}^*_n}(x^*_{T+1,T})\right\} \tag{40}$$

S-3. SIMULATION STUDIES IN THE UNIVARIATE SITUATIONS

In this section we consider two simulation studies: in the first study we consider univariate data generated from linear observational and evolutionary equations, the linear models being invariant with respect to time–we fit our Gaussian process based model on this linear data. In the second case we generate data from the parametric, univariate growth model of Carlin, Polson & Stoffer (1992) and fit our nonparametric model to the data, validating our model and methodology in the process when the data are governed by non-linear observational and evolutionary equations.

S-3.1 Data generated from a univariate linear model

We generate data from the following linear model:

$$x_t = \beta_{x,0} + \beta_{x,1}x_{t-1} + u_t \tag{41}$$

$$y_t = \beta_{y,0} + \beta_{y,1}x_t + v_t, \quad t = 1, \ldots, 100, \tag{42}$$

where $x_0 = 0$. As before, we assume $u_t \sim N(0, \sigma^2_\epsilon)$, $v_t \sim N(0, \sigma^2_\eta)$, independently for all t, and set $\sigma_\epsilon = 0.1 = \sigma_\eta$. We fix the true values of $(\beta_{x,0}, \beta_{x,1}, \beta_{y,0}, \beta_{y,1})$ to be $(1.0, 0.1, 8.0, 0.05)$, respectively. We then generated the data set consisting of 101 data points using these true values. As before, we set aside the last data point for the purpose of forecasting.

Note that under the above true model, the coefficient of time t is zero for both the observational and the evolutionary equations. In other words, the functions are not time-variant. Thus, while fitting our Gaussian process based model, we expect the corresponding coefficients in the mean

functions $\beta_{2,f}$, $\beta_{2,g}$ and the corresponding smoothness parameters $r_{1,f}$ and $r_{1,g}$ to have large posterior probabilities around zero. The results of our model implementation show that it is indeed the case.

S-3.1.1. Choice of grid and MCMC implementation To set up the grid $\boldsymbol{G}_n$ we first note the interval containing the entire true time series; in this example, the entire true time series fall within $[0.8, 1.5]$. We then consider a much larger interval, $[-30, 30]$, containing the aforementioned interval, and divide the larger interval into 100 sub-intervals of equal length. Then we select a value randomly from each sub-interval. This yields values of the second component of the two-dimensional grid $\boldsymbol{G}_n$. It is worth mentioning that we experimented with other reasonable grid choices, but the results suggested considerable robustness with respect to the grid choices. For the first component we generate a number uniformly from each of the 100 sub-intervals $[i, i+1]$; $i = 0, \ldots, 99$.

We discarded the first 10,000 MCMC iterations as burn-in and stored the next 50,000 iterations for inference. For updating $\{x_0, \ldots, x_T\}$ we used the linear model based proposals detailed in Sections S-2.8 and S-2.9. Since the true models are also linear, the performance of these proposal distributions, as measured by informal MCMC diagnostics, turned out to be adequate. However, updating $\{\sigma_f, \sigma_g, \sigma_\epsilon, \sigma_\eta\}$ using the linear model based proposals detailed in Sections S-2.3 and S-2.4 did not perform satisfactorily since the assumptions $\sigma_f \approx \sigma_\epsilon$ and $\sigma_g \approx \sigma_\eta$ do not hold here (clearly, since the true models are linear, $\sigma_f = \sigma_g = 0$, whereas σ_ϵ and σ_η are positive). So, to update the variance parameters we used the normal random walk proposal with variance 0.05, which worked adequately. We note, however, that it is possible to modify the proposals provided in Sections S-2.3 and S-2.4 so that the fact $\sigma_f = \sigma_g = 0$ is taken account of, but since our random walk proposals performed adequately here we refrained from further experiments on proposal distributions. It took around 15 hours in an ordinary laptop to implement this experiment.

S-3.1.2. Results of model-fitting Figures S-1, S-2 and S-3 show the posterior distributions of the unknowns. Note that whenever applicable, the true value is well-captured by the posteriors in question. Also, as expected, the posteriors of $\beta_{2,f}$, $\beta_{2,g}$, $r_{1,f}$ and $r_{1,g}$, have high posterior probabili-

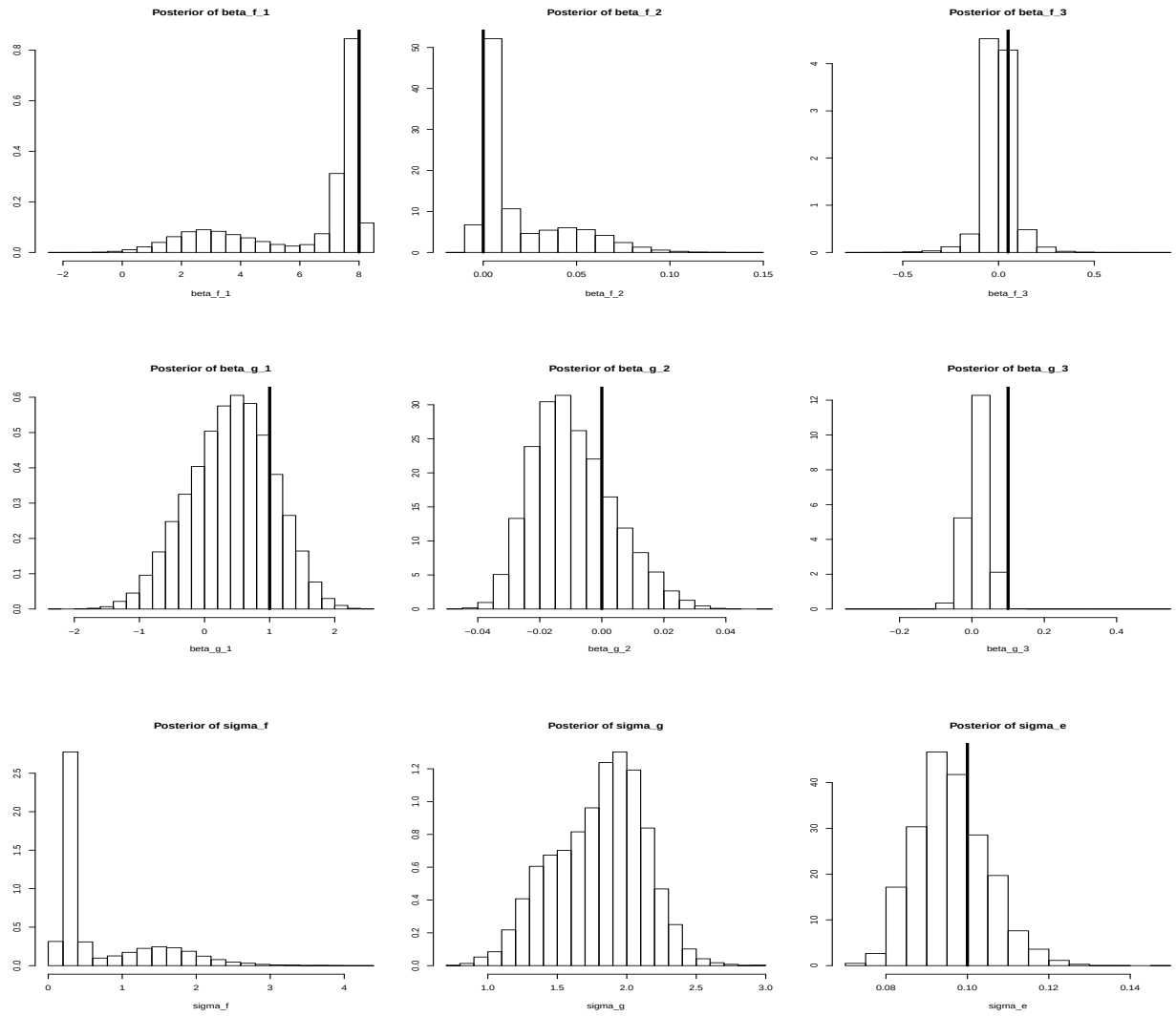

Figure S-1: **Simulation study with data generated from the univariate linear model**: Posterior densities of $\beta_{0,f}$, $\beta_{1,f}$, $\beta_{2,f}$, $\beta_{0,g}$, $\beta_{1,g}$, $\beta_{2,g}$, σ_f, σ_g, and σ_ϵ. The solid line stands for the true values of the respective parameters.

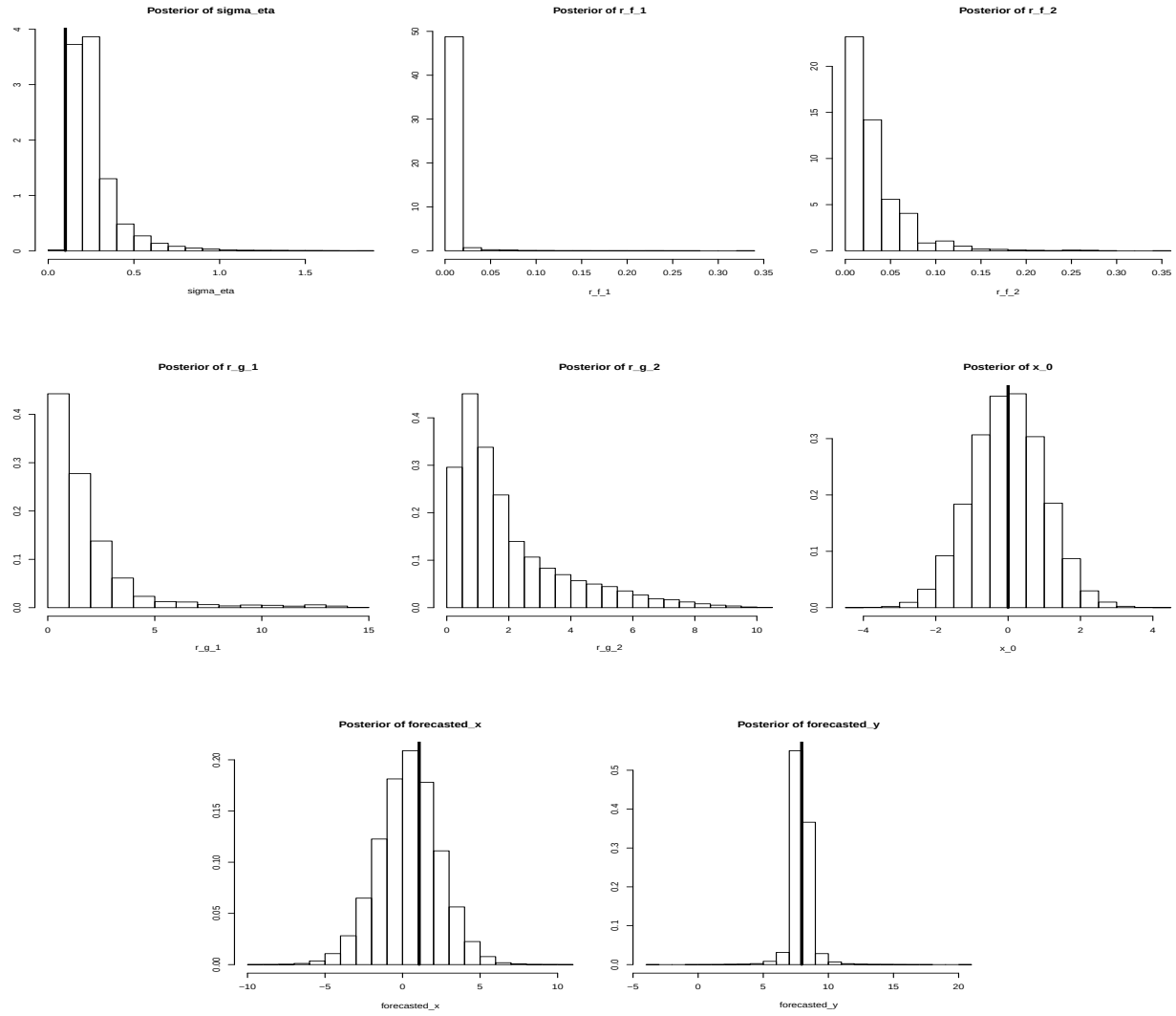

Figure S-2: **Simulation study with data generated from the univariate linear model**: Posterior densities of σ_η, $r_{1,f}$, $r_{2,f}$, $r_{1,g}$, $r_{2,g}$, x_0, x_{T+1} (one-step forecasted x), and y_{T+1} (one-step forecasted y). The solid line stands for the true values of the respective parameters.

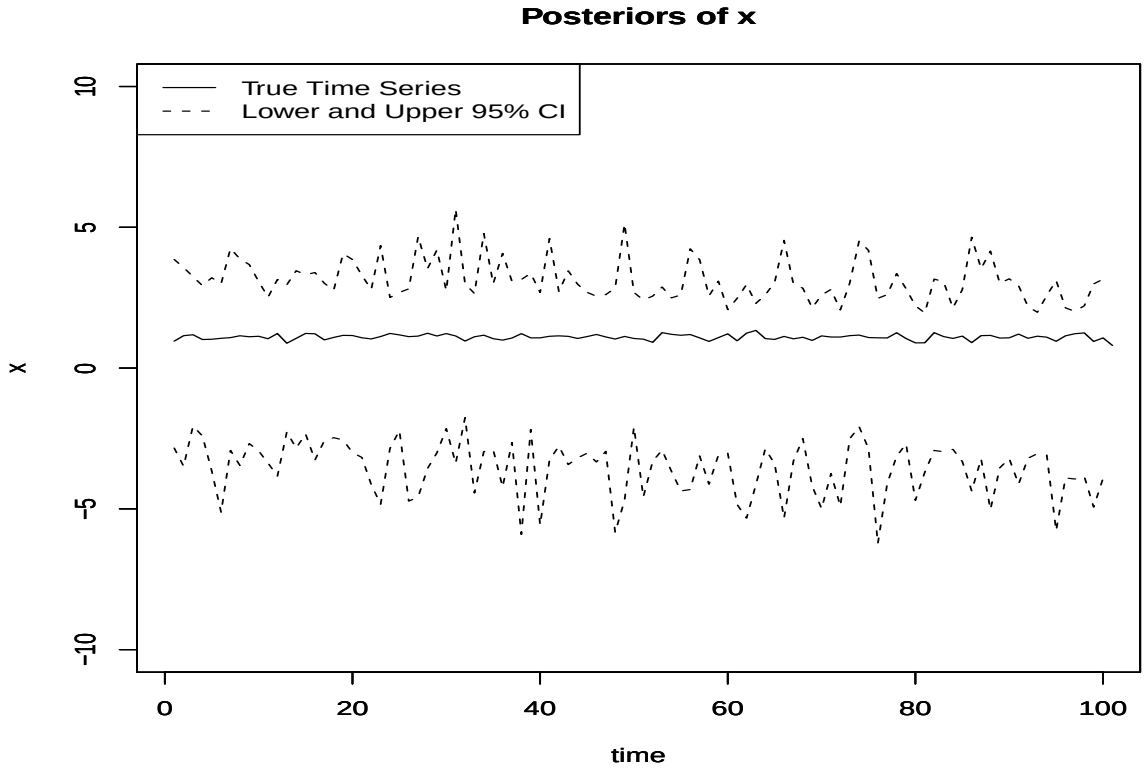

Figure S-3: **Simulation study with data generated from the univariate linear model**: 95% highest posterior density credible intervals of the time series $x_1, \ldots, x_T$. The solid line stands for the true time series.

ties around zero, supporting the time-invariant nature of the true observational and the evolutionary equations. The true values of σ_ϵ, σ_η and x_0 also fell well within the highest density regions of their respective posterior distributions. The true forecasted values of x_{101} and y_{101} are well-captured by their respective forecast posteriors. The true time series $\{x_1, \ldots, x_{100}\}$ again fell well within the lower and the upper 95% highest posterior density credible limits, but the credible regions seem to be wider than the usual parametric approaches. This is a consequence of acknowledging uncertainty about the functional forms of both f and g. Particularly because the posterior distribution of the state variables depends upon both the unknown functions f and g, it is perhaps not surprising that the credible intervals reflect the effect of uncertainties about the random functions. In light of the knowledge of the true time series in this simple simulated example, which has a linear structure, the relatively wide credible intervals may not seem to indicate very encouraging performance, but in complex, realistic situations, where any structure of the true time series is impossible to guess, the relatively large credible intervals make good practical sense.

S-3.2 Data generated from a univariate parametric growth model

Following Carlin et al. (1992) we generate data from the following model:

$$x_t = \alpha x_{t-1} + \beta x_{t-1}/(1 + x_{t-1}^2) + \gamma \cos(1.2(t-1)) + u_t \tag{43}$$

$$y_t = x_t^2/20 + v_t, \quad t = 1, \ldots, 100, \tag{44}$$

where $x_0 = 0$. Here we assume $u_t \sim N(0, \sigma_\epsilon^2)$ and $v_t \sim N(0, \sigma_\eta^2)$, independently for all t. We set $\sigma_\epsilon = 0.1 = \sigma_\eta$ and fix the true values of (α, β, γ) to be $(0.05, 0.1, 0.2)$, respectively. We then generated the data set consisting of 101 data points using these true values. As before, we set aside the last data point for the purpose of forecasting.

S-3.2.1. Choice of grid and MCMC implementation In this example the entire true time series falls within the space $[-0.33, 0.32]$. As before, we consider the much larger space, $[-30, 30]$, and divide it into 100 sub-intervals of equal length, and then select a value randomly from each such sub-interval. This yields values of the second component of the two-dimensional grid $\boldsymbol{G}_n$.

As in the linear model experiment, here also we experimented with other reasonable choices and even in this non-linear experiment, the results remained exhibited considerable robustness. For the first component we generate a number uniformly from each of the 100 sub-intervals $[i, i+1]$; $i = 0, \ldots, 99$.

We discarded the first 10000 MCMC iterations as burn-in and stored the next 50,000 iterations for inference. We used the normal random walk proposal with variance 0.05 for updating σ_f, σ_g, σ_ϵ and σ_η. For updating x_0 we used the normal random walk proposal with variance 1, but for updating $x_t; t = 1, \ldots, T$, we found in this example, using informal convergence diagnostics, that the random walk proposal with variance t provided much better mixing than those with constant variances. These proposals also performed much better than the linear model based proposals detailed in Sections S-2.8 and S-2.9. It took around 15 hours in an ordinary laptop to implement this experiment.

S-3.2.2. Results of model-fitting Figures S-4, S-5 and S-6 show the posterior distributions of the unknowns. Note that whenever applicable, the true value is well-captured by the posteriors in question. To avoid any confusion, we mention that the true values of (α, β, γ) associated with the true observational equation (43), for example, are not comparable with the parameters $(\beta_{0,f}, \beta_{1,f}, \beta_{2,f})$ associated with our Gaussian process model. Hence, the figures do not show any true value associated with the posteriors of the Gaussian process parameters. Note that this is in contrast with the previous example where the true model has a linear state space structure. There, the true functional forms are comparable with the linear mean functions of the underlying Gaussian processes, and that is the reason why the figures of the posterior distributions of the Gaussian process parameters also consisted of the vertical line denoting the true values of the parameters.

As before, the true values of x_{101} and y_{101} are captured very well by their respective posteriors. The true time series $\{x_1, \ldots, x_{100}\}$ again fell well within the lower and the upper 95% highest posterior density credible limits. However, note that the credible intervals in this case are somewhat wider than the corresponding linear model, which is to be expected. The true values of σ_ϵ, σ_η and x_0 also fell well within the highest density regions of their respective posterior distributions. That

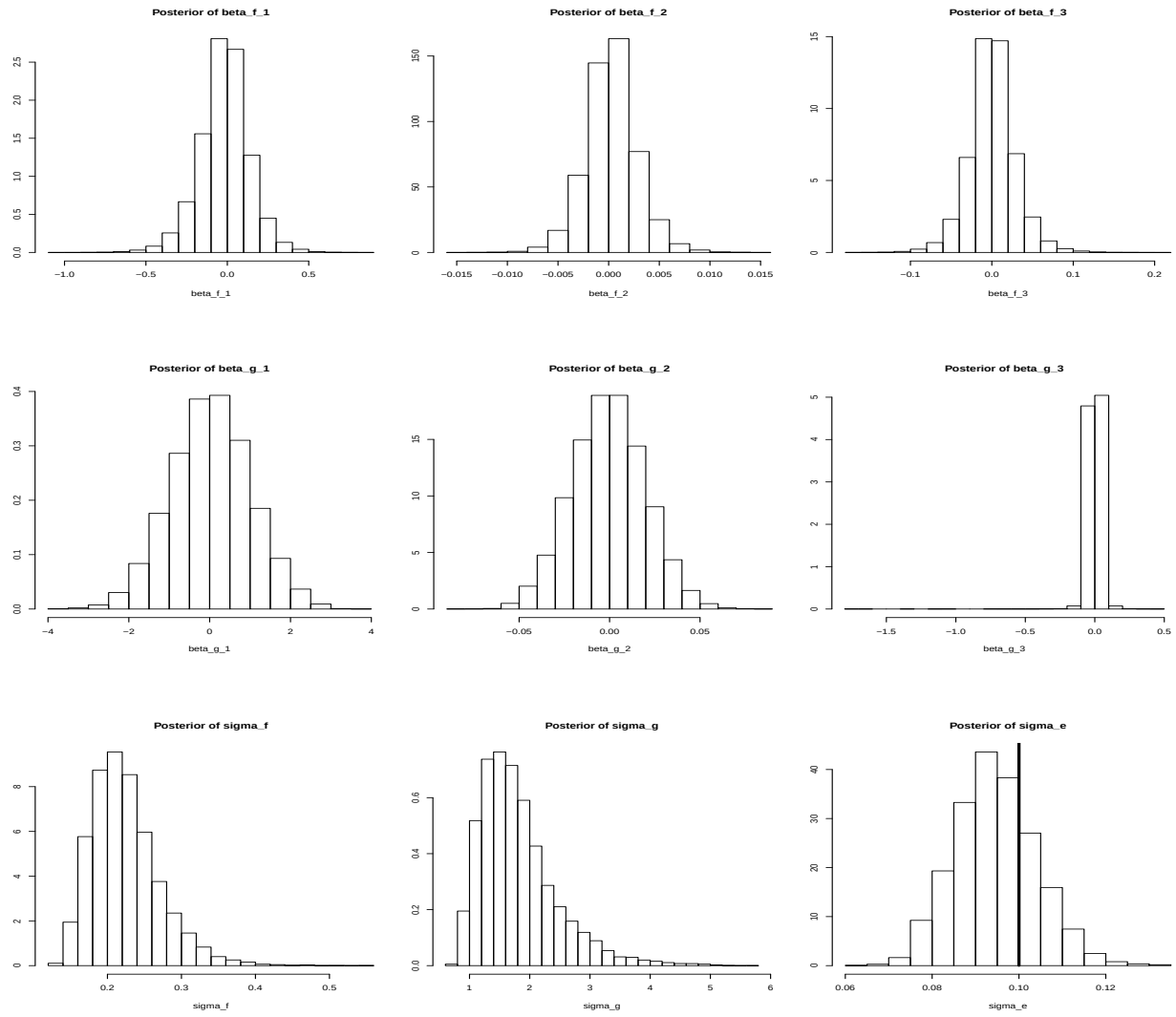

Figure S-4: **Simulation study with data generated from the univariate growth model of CPS**: Posterior densities of $\beta_{0,f}$, $\beta_{1,f}$, $\beta_{2,f}$, $\beta_{0,g}$, $\beta_{1,g}$, $\beta_{2,g}$, σ_f, σ_g, and σ_ϵ. The solid line stands for the true values of the respective parameters.

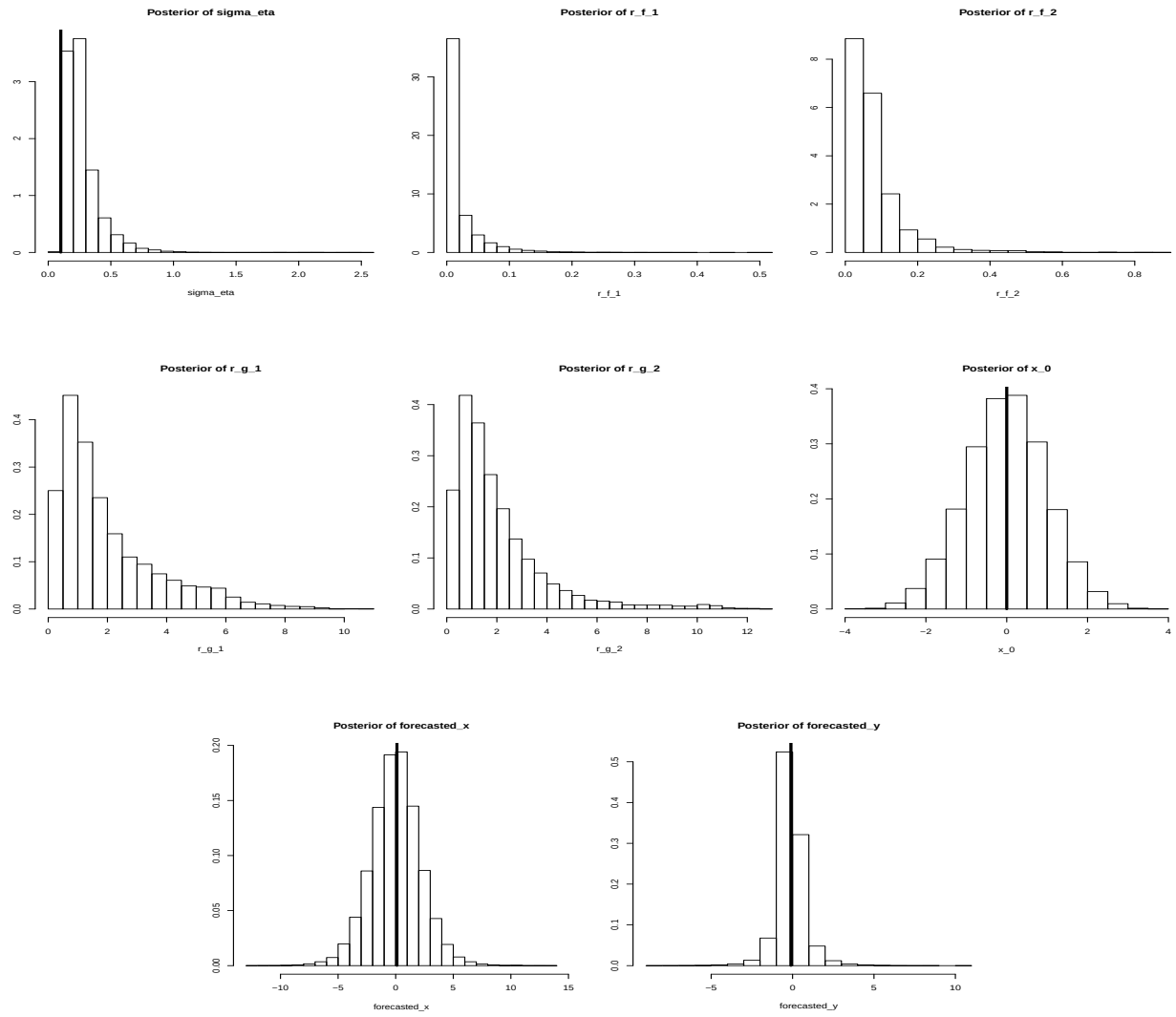

Figure S-5: **Simulation study with data generated from the univariate growth model of CPS**: Posterior densities of σ_η, $r_{1,f}$, $r_{2,f}$, $r_{1,g}$, $r_{2,g}$, x_0, x_{T+1} (one-step forecasted x), and y_{T+1} (one-step forecasted y). The solid line stands for the true values of the respective parameters.

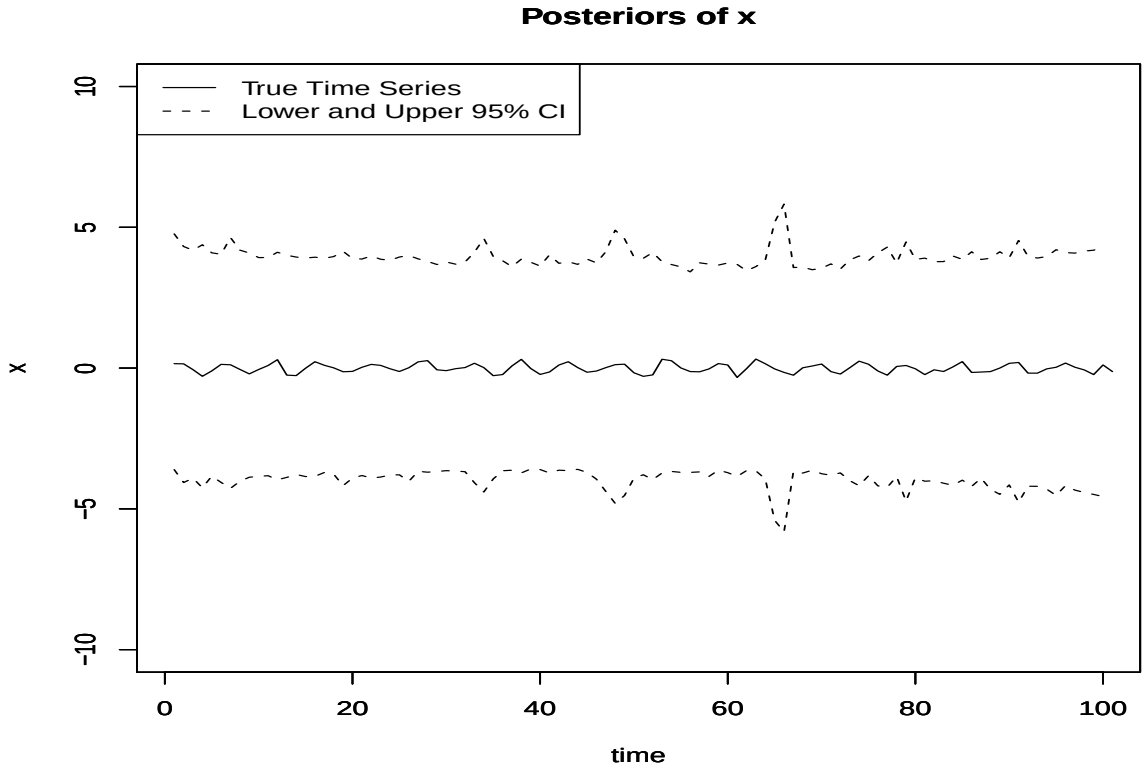

Figure S-6: **Simulation study with data generated from the univariate growth model of CPS**: 95% highest posterior density credible intervals of the time series $x_1, \ldots, x_T$. The solid line stands for the true time series.

our Gaussian process approach fits the data so well despite the fact that the functions $f(\cdot,\cdot)$ and $g(\cdot,\cdot)$ are highly non-linear, is quite encouraging.

S-4. EXTENSION OF OUR GAUSSIAN PROCESS BASED APPROACH TO THE MULTIVARIATE SITUATION

Now we extend our nonparametric dynamic model to the case where both $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ are multivariate. In particular, we assume that they are p-component and q-component vectors, respectively. Then our multivariate model is of the form

$$\boldsymbol{y}_t = \boldsymbol{f}(\boldsymbol{x}^*_{t,t}) + \boldsymbol{\epsilon}_t, \;\; \boldsymbol{\epsilon}_t \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon), \tag{45}$$

$$\boldsymbol{x}_t = \boldsymbol{g}(\boldsymbol{x}^*_{t,t-1}) + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim N_q(\boldsymbol{x}_0, \boldsymbol{\Sigma}_\eta), \tag{46}$$

where $\boldsymbol{x}_0 \sim N_q(\boldsymbol{\mu}_{x_0}, \boldsymbol{\Sigma}_{x_0})$; $\boldsymbol{\mu}_{x_0}, \boldsymbol{\Sigma}_{x_0}$ assumed known. In the above, $\boldsymbol{f}(\cdot) = (f_1(\cdot), \ldots, f_p(\cdot))'$ is a function with p components and $\boldsymbol{g}(\cdot) = (g_1(\cdot), \ldots, g_q(\cdot))'$ is a function consisting of q components. We assume that $\boldsymbol{f}(\cdot)$ is a p-variate Gaussian process with mean $E[\boldsymbol{f}(\cdot)] = \boldsymbol{B}'_f \boldsymbol{h}(\cdot)$ and with covariance function $cov(\boldsymbol{f}(\boldsymbol{z}_1), \boldsymbol{f}(\boldsymbol{z}_2)) = c_f(\boldsymbol{z}_1, \boldsymbol{z}_2)\boldsymbol{\Sigma}_f$, for any q-dimensional inputs $\boldsymbol{z}_1, \boldsymbol{z}_2$. Here $\boldsymbol{h}(\cdot) = (h_1(\cdot), \ldots, h_m(\cdot))'$ and $\boldsymbol{B}_f = (\boldsymbol{\beta}_{1,f}, \ldots, \boldsymbol{\beta}_{p,f})$, where, for $j = 1, \ldots, p$, $\boldsymbol{\beta}_{j,f}$ are m-dimensional column vectors; clearly, $m = q+2$. Also, $c_f(\boldsymbol{z}_1, \boldsymbol{z}_2) = \exp\{-(\boldsymbol{z}_1 - \boldsymbol{z}_2)'\boldsymbol{R}_f(\boldsymbol{z}_1 - \boldsymbol{z}_2)\}$, where $\boldsymbol{R}_f$ is a diagonal matrix consisting of $(q+1)$ smoothness parameters, denoted by $\{r_{1,f}, \ldots, r_{(q+1),f}\}$. Similarly, we assume that $\boldsymbol{g}(\cdot)$ is a Gaussian process with mean $E[\boldsymbol{g}(\cdot)] = \boldsymbol{B}'_g \boldsymbol{h}(\cdot)$ and covariance function $c_g(\boldsymbol{z}_1, \boldsymbol{z}_2)\boldsymbol{\Sigma}_g = \exp\{-(\boldsymbol{z}_1 - \boldsymbol{z}_2)'\boldsymbol{R}_g(\boldsymbol{z}_1 - \boldsymbol{z}_2)\}$, the notation used being analogous to those used for description of the Gaussian process $\boldsymbol{f}(\cdot)$.

S-4.1 Multivariate data and its distribution

The multivariate data is given by the following $T \times p$ matrix: $\boldsymbol{D}_T = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)'$. Writing $\boldsymbol{D}_T$ vectorically as a Tp-vector for convenience, as $\boldsymbol{D}_{Tp} = (\boldsymbol{y}'_1, \boldsymbol{y}'_2, \ldots, \boldsymbol{y}'_T)'$, it follows that $[\boldsymbol{D}_{Tp} \mid$

$\boldsymbol{B}_f, \boldsymbol{\Sigma}_f, \boldsymbol{R}_f, \boldsymbol{\Sigma}_\epsilon]$ is a Tp-variate normal with mean vector

$$E[\boldsymbol{D}_{Tp} \mid \boldsymbol{B}_f, \boldsymbol{\Sigma}_f, \boldsymbol{R}_f, \boldsymbol{\Sigma}_\epsilon] = \begin{pmatrix} \boldsymbol{B}_f'\boldsymbol{h}(\boldsymbol{x}_{1,1}^*) \\ \boldsymbol{B}_f'\boldsymbol{h}(\boldsymbol{x}_{2,2}^*) \\ \vdots \\ \boldsymbol{B}_f'\boldsymbol{h}(\boldsymbol{x}_{T,T}^*) \end{pmatrix} = \boldsymbol{\mu}_{D_{Tp}} \text{ (say)} \tag{47}$$

and covariance matrix

$$V[\boldsymbol{D}_{Tp} \mid \boldsymbol{B}_f, \boldsymbol{\Sigma}_f, \boldsymbol{R}_f, \boldsymbol{\Sigma}_\epsilon] = \boldsymbol{A}_{f,D_T} \otimes \boldsymbol{\Sigma}_f + \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma}_{D_{Tp}} \text{ (say)}. \tag{48}$$

Since (48) does not admit a Kronecker product form, the distribution of $\boldsymbol{D}_{Tp}$ can not be written as matrix normal, which requires right and left covariance matrices forming Kronecker product in the corresponding multivariate distribution of the vector. It follows that $[\boldsymbol{y}_{T+1} = \boldsymbol{f}(\boldsymbol{x}_{T+1,T+1}^*) + \boldsymbol{\epsilon}_{T+1} \mid \boldsymbol{D}_T, \boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{T+1}\boldsymbol{B}_f, \boldsymbol{\Sigma}_f, \boldsymbol{R}_f, \boldsymbol{\Sigma}_\epsilon]$ is p-variate normal with mean

$$\boldsymbol{\mu}_{y_{T+1}} = \boldsymbol{B}_f'\boldsymbol{h}(\boldsymbol{x}_{T+1,T+1}^*) + (\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{B}_f)'\boldsymbol{A}_{f,D_T}^{-1}\boldsymbol{s}_{f,\boldsymbol{D}_T}(\boldsymbol{x}_{T+1,T+1}^*) \tag{49}$$

and variance

$$\boldsymbol{\Sigma}_{y_{T+1}} = \left\{1 - \boldsymbol{s}_{f,\boldsymbol{D}_T}(\boldsymbol{x}_{T+1,T+1}^*)'\boldsymbol{A}_{f,\boldsymbol{D}_T}^{-1}\boldsymbol{s}_{f,\boldsymbol{D}_T}(\boldsymbol{x}_{T+1,T+1}^*)\right\}\boldsymbol{\Sigma}_f + \boldsymbol{\Sigma}_\epsilon \tag{50}$$

S-4.2 Distributions of $\boldsymbol{g}(\boldsymbol{x}_{1,0}^*)$ and $\boldsymbol{D}_n^*$

Conditional on $\boldsymbol{x}_0$, $\boldsymbol{g}(\boldsymbol{x}_{1,0}^*)$ is q-variate normal with mean $\boldsymbol{B}_g'\boldsymbol{h}(\boldsymbol{x}_{1,0}^*)$ and covariance matrix $\boldsymbol{\Sigma}_g$.

In contrast with the distribution of $\boldsymbol{D}_T$, $\boldsymbol{D}_{z,nq} = (\boldsymbol{g}'(\boldsymbol{z}_1), \boldsymbol{g}'(\boldsymbol{z}_2), \ldots, \boldsymbol{g}'(\boldsymbol{z}_n))'$ has an nq-variate normal distribution with mean

$$E[\boldsymbol{D}_{z,nq} \mid \boldsymbol{B}_g, \boldsymbol{\Sigma}_g, \boldsymbol{R}_g] = \begin{pmatrix} \boldsymbol{B}_g'\boldsymbol{h}(\boldsymbol{z}_1) \\ \boldsymbol{B}_g'\boldsymbol{h}(\boldsymbol{z}_2) \\ \vdots \\ \boldsymbol{B}_g'\boldsymbol{h}(\boldsymbol{z}_n) \end{pmatrix} = \boldsymbol{\mu}_{D_{z,nq}} \text{ (say)} \tag{51}$$

and covariance matrix

$$V[\boldsymbol{D}_{z,nq} \mid \boldsymbol{B}_g, \boldsymbol{\Sigma}_g, \boldsymbol{R}_g] = \boldsymbol{A}_{g,D_n^*} \otimes \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}_{D_{z,nq}} \text{ (say)}. \tag{52}$$

Hence, the distribution of the $n \times q$-dimensional matrix $\boldsymbol{D}_n^* = (\boldsymbol{g}(\boldsymbol{z}_1), \boldsymbol{g}(\boldsymbol{z}_2), \ldots, \boldsymbol{g}(\boldsymbol{z}_n))'$ is matrix normal:

$$[\boldsymbol{D}_n^* \mid \boldsymbol{B}_g, \boldsymbol{\Sigma}_g, \boldsymbol{R}_g] \sim \mathcal{N}_{n,q}\left(\boldsymbol{H}_z \boldsymbol{B}_g, \boldsymbol{A}_{g,D_n^*}, \boldsymbol{\Sigma}_g\right) \tag{53}$$

Conditionally on $(\boldsymbol{x}_0, \boldsymbol{g}(\boldsymbol{x}_{1,0}^*))$, it follows that $\boldsymbol{D}_n^*$ is $n \times q$-dimensional matrix-normal:

$$[\boldsymbol{D}_n^* \mid \boldsymbol{g}(\boldsymbol{x}_{1,0}^*), \boldsymbol{x}_0, \boldsymbol{B}_g, \boldsymbol{\Sigma}_g, \boldsymbol{R}_g, \boldsymbol{\Sigma}_\eta] \sim \mathcal{N}_{n,q}\left(\boldsymbol{\mu}_{g,D_n^*}, \boldsymbol{\Sigma}_{g,D_n^*}, \boldsymbol{\Sigma}_g\right) \tag{54}$$

In (54) $\boldsymbol{\mu}_{g,D_n^*}$ is the mean matrix, given by

$$\boldsymbol{\mu}_{g,D_n^*} = \boldsymbol{H}_{D_n^*} \boldsymbol{B}_g + \boldsymbol{s}_{g,D_n^*}(\boldsymbol{x}_{1,0}^*)(\boldsymbol{g}(\boldsymbol{x}_{1,0}^*)' - \boldsymbol{h}(\boldsymbol{x}_{1,0}^*)' \boldsymbol{B}_g), \tag{55}$$

and

$$\boldsymbol{\Sigma}_{g,D_n^*} = \boldsymbol{A}_{g,D_n^*} - \boldsymbol{s}_{g,D_n^*}(\boldsymbol{x}_{1,0}^*) \boldsymbol{s}_{g,D_n^*}(\boldsymbol{x}_{1,0}^*).' \tag{56}$$

Here we slightly abuse notation to denote both univariate and multivariate versions of the mean matrix and the right covariance matrix by $\boldsymbol{\mu}_{g,D_n^*}$ and $\boldsymbol{\Sigma}_{g,D_n^*}$, respectively (see (9) and (11) of GMRB).

S-4.3 Joint distribution of $\{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{T+1}, \boldsymbol{D}_n^*\}$

Note that

$$[\boldsymbol{x}_1 \mid \boldsymbol{g}(\boldsymbol{x}_0), \boldsymbol{x}_0, \boldsymbol{B}_g, \boldsymbol{\Sigma}_g] \sim N_q\left(\boldsymbol{g}(\boldsymbol{x}_{1,0}^*), \boldsymbol{\Sigma}_\eta\right), \tag{57}$$

and for $t = 1, \ldots, T$, the conditional distribution $[\boldsymbol{x}_{t+1} = \boldsymbol{g}(\boldsymbol{x}_{t+1,t}^*) + \boldsymbol{\eta}_{t+1} \mid \boldsymbol{D}_n^*, \boldsymbol{x}_t, \boldsymbol{B}_g, \boldsymbol{\Sigma}_g, \boldsymbol{R}_g, \boldsymbol{\Sigma}_\eta]$ is q-variate normal with mean

$$\boldsymbol{\mu}_{x_t} = \boldsymbol{B}_g' \boldsymbol{h}(\boldsymbol{x}_{t+1,t}^*) + (\boldsymbol{D}_n^* - \boldsymbol{H}_{D_n^*} \boldsymbol{B}_g)' \boldsymbol{A}_{g,D_n^*}^{-1} \boldsymbol{s}_{g,\boldsymbol{D}_n^*}(\boldsymbol{x}_{t+1,t}^*) \tag{58}$$

and variance

$$\boldsymbol{\Sigma}_{x_t} = \left\{1 - \boldsymbol{s}_{g,\boldsymbol{D}_n^*}(\boldsymbol{x}_{t+1,t}^*)' \boldsymbol{A}_{g,\boldsymbol{D}_n^*}^{-1} \boldsymbol{s}_{g,\boldsymbol{D}_n^*}(\boldsymbol{x}_{t+1,t}^*)\right\} \boldsymbol{\Sigma}_g + \boldsymbol{\Sigma}_\eta. \tag{59}$$

Since $[\boldsymbol{x}_0] \sim N_p\left(\boldsymbol{\mu}_{x_0}, \boldsymbol{\Sigma}_{x_0}\right)$ and the distribution of $\boldsymbol{D}_n^*$ is given by (53) the joint distribution is obtained by taking products of the individual distributions.

S-4.4 Prior distributions

We assume the following prior distributions:

$$\text{For } \; i = 1, \ldots, (q+1),$$

$$[\log(r_{i,f})] \overset{iid}{\sim} N\left(\mu_{R_f}, \sigma^2_{R_f}\right) \tag{60}$$

$$\text{For } \; i = 1, \ldots, (p+1),$$

$$[\log(r_{i,g})] \overset{iid}{\sim} N\left(\mu_{R_f}, \sigma^2_{R_g}\right) \tag{61}$$

$$[\boldsymbol{\Sigma}_\epsilon] \propto |\boldsymbol{\Sigma}_\epsilon|^{-\frac{\nu_\epsilon+p+1}{2}} \exp\left[-\frac{1}{2} tr\left(\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\Sigma}_{\epsilon,0}\right)\right]; \;\; \nu_\epsilon > p-1 \tag{62}$$

$$[\boldsymbol{\Sigma}_\eta] \propto |\boldsymbol{\Sigma}_\eta|^{-\frac{\nu_\eta+q+1}{2}} \exp\left[-\frac{1}{2} tr\left(\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_{\eta,0}\right)\right]; \;\; \nu_\eta > q-1 \tag{63}$$

$$[\boldsymbol{\Sigma}_f] \propto |\boldsymbol{\Sigma}_f|^{-\frac{\nu_f+p+1}{2}} \exp\left[-\frac{1}{2} tr\left(\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\Sigma}_{f,0}\right)\right]; \;\; \nu_f > p-1 \tag{64}$$

$$[\boldsymbol{\Sigma}_g] \propto |\boldsymbol{\Sigma}_g|^{-\frac{\nu_g+q+1}{2}} \exp\left[-\frac{1}{2} tr\left(\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Sigma}_{g,0}\right)\right]; \;\; \nu_g > q-1 \tag{65}$$

$$[\boldsymbol{B}_f \mid \boldsymbol{\Sigma}_f] \sim \mathcal{N}_{m,p}\left(\boldsymbol{B}_{f,0}, \boldsymbol{\Sigma}_{B_f,0}, \psi\boldsymbol{\Sigma}_f\right) \tag{66}$$

$$[\boldsymbol{B}_g \mid \boldsymbol{\Sigma}_g] \sim \mathcal{N}_{m,q}\left(\boldsymbol{B}_{g,0}, \boldsymbol{\Sigma}_{B_g,0}, \psi\boldsymbol{\Sigma}_g\right) \tag{67}$$

All the prior parameters are assumed to be known; and their choices will be discussed in the specific applications to be considered.

The full conditionals are of analogous forms as those in the univariate case, provided by equations from (6) to (19), but now the univariate distributions must be replaced with multivariate distributions and multivariate distributions with matrix-variate distributions, and interestingly in some cases the full conditionals are not available in closed form although these were available in the one-dimensional version of the problem. For example, although the full conditional of $\boldsymbol{\beta}_f$ and $\boldsymbol{\beta}_g$ were available in the one-dimensional problem, the corresponding distributions of $\boldsymbol{B}_f$ and $\boldsymbol{B}_g$ are no longer available in closed forms. The problem of obtaining the closed form of the full conditional of $\boldsymbol{B}_f$ can be attributed to the fact that (48) can not be represented as a single Kronecker product. The non-availability of the closed form in the case of $\boldsymbol{B}_g$ is due to fact that the covariance matrices of $\boldsymbol{x}_t$ are additive. In fact, it turns out that only the full conditional of $\boldsymbol{x}_{T+1}$ is of standard form. Details of our MCMC methods for the multivariate case are provided in the next section.

S-5. DETAILS OF MCMC SAMPLING IN THE MULIVARIATE SITUATION

To obtain good proposal distributions, we use our old strategy of ignoring the error terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$. Below we provide details of the proposal distributions used in each case. For our purpose, we abuse notation slightly as in Section 4 of GMRB, that is, we denote by $\boldsymbol{G}_{n+1}$, $\boldsymbol{D}^*_{n+1}$, $\boldsymbol{A}_{g,n+1}$ and $\boldsymbol{H}_{g,D^*_{n+1}}$ the multivariate analogues of the quantities corresponding to the univariate situation described in Section 4 of GMRB. In other words, let $\boldsymbol{G}_{n+1} = \boldsymbol{G} \cup \{\boldsymbol{x}^*_{1,0}\}$, $\boldsymbol{D}^*_{n+1} = \left(\boldsymbol{D}^{*\prime}_n, \boldsymbol{g}(\boldsymbol{x}^*_{1,0})\right)'$, $\boldsymbol{A}_{g,n+1} = \begin{pmatrix} \boldsymbol{A}_{g,D^*_n} & \boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{1,0}) \\ \boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{1,0})' & 1 \end{pmatrix}$ and $\boldsymbol{H}'_{g,D^*_{n+1}} = [\boldsymbol{H}'_{g,D^*_n}, \boldsymbol{h}(\boldsymbol{x}^*_{1,0})]$.

S-5.1 Proposal distribution for updating $\boldsymbol{B}_f$

Assuming $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma}_f$ in our multivariate dynamic state-space model, the full conditional of $\boldsymbol{B}_f$ is $m \times p$-variate matrix-normal:

$$\boldsymbol{B}_f \sim \mathcal{N}_{m,p}\left(\boldsymbol{\mu}_{B_f}, \boldsymbol{\Sigma}_{B_f}, \boldsymbol{\Sigma}_f\right), \tag{68}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{B_f} = &\left(\boldsymbol{H}'_{D_T}\left(\boldsymbol{A}_{f,D_T} + \boldsymbol{I}_T\right)^{-1}\boldsymbol{H}_{D_T} + \psi^{-1}\boldsymbol{\Sigma}^{-1}_{B_f,0}\right)^{-1} \\ &\times \left(\boldsymbol{H}'_{D_T}\left(\boldsymbol{A}_{f,D_T} + \boldsymbol{I}_T\right)^{-1}\boldsymbol{D}_T + \psi^{-1}\boldsymbol{\Sigma}^{-1}_{B_f,0}\boldsymbol{B}_{f,0}\right), \end{aligned} \tag{69}$$

and

$$\boldsymbol{\Sigma}_{B_f} = \left(\boldsymbol{H}'_{D_T}\left(\boldsymbol{A}_{f,D_T} + \boldsymbol{I}_T\right)^{-1}\boldsymbol{H}_{D_T} + \psi^{-1}\boldsymbol{\Sigma}^{-1}_{B_f,0}\right)^{-1}. \tag{70}$$

S-5.2 Proposal distribution for $\boldsymbol{B}_g$

Assuming $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Sigma}_g$, the conditional distribution of $\boldsymbol{B}_g$ is $m \times q$-variate matrix:

$$\boldsymbol{B}_g \sim \mathcal{N}_{m,q}\left(\boldsymbol{\mu}_{B_g}, \boldsymbol{\Sigma}_{B_g}, \boldsymbol{\Sigma}_g\right), \tag{71}$$

where

$$\boldsymbol{\mu}_{B_g} = \Big\{\psi^{-1}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{B}_g,0} + \boldsymbol{H}'_{D^*_{n+1}}\boldsymbol{A}^{-1}_{g,D^*_{n+1}}\boldsymbol{H}_{D^*_{n+1}} + \sum_{t=1}^{T}\frac{\left(\boldsymbol{H}'_{D^*_n}\boldsymbol{A}^{-1}_{g,D^*_n}\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t}) - \boldsymbol{h}(\boldsymbol{x}^*_{t+1,t})\right)\left(\boldsymbol{H}'_{D^*_n}\boldsymbol{A}^{-1}_{g,D^*_n}\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t}) - \boldsymbol{h}(\boldsymbol{x}^*_{t+1,t})\right)'}{\sigma^2_{g,t}}\Big\}^{-1}$$
$$\times\Big\{\psi^{-1}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{B}_g,0}\boldsymbol{B}_{g,0} + \boldsymbol{H}'_{D^*_{n+1}}\boldsymbol{A}^{-1}_{g,D^*_{n+1}}\boldsymbol{D}^*_{n+1} + \sum_{t=1}^{T}\frac{\left(\boldsymbol{h}(\boldsymbol{x}^*_{t+1,t}) - \boldsymbol{H}'_{D^*_n}\boldsymbol{A}^{-1}_{g,D^*_n}\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t})\right)\left(\boldsymbol{x}_{t+1} - \boldsymbol{D}'_z\boldsymbol{A}^{-1}_{g,D^*_n}\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t})\right)'}{\sigma^2_{g,t}}\Big\}. \tag{72}$$

and

$$\boldsymbol{\Sigma}_{B_g} = \Big\{\psi^{-1}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{B}_g,0} + \boldsymbol{H}'_{D^*_{n+1}}\boldsymbol{A}^{-1}_{g,D^*_{n+1}}\boldsymbol{H}_{D^*_{n+1}} + \sum_{t=1}^{T}\frac{\left(\boldsymbol{H}'_{D^*_n}\boldsymbol{A}^{-1}_{g,D^*_n}\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t}) - \boldsymbol{h}(\boldsymbol{x}^*_{t+1,t})\right)\left(\boldsymbol{H}'_{D^*_n}\boldsymbol{A}^{-1}_{g,D^*_n}\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t}) - \boldsymbol{h}(\boldsymbol{x}^*_{t+1,t})\right)'}{\sigma^2_{g,t}}\Big\}^{-1} \tag{73}$$

In (72) and (73), $\sigma^2_{g,t}$ is given by (27), with obvious notational change from the univariate x_t to the multivariate $\boldsymbol{x}_t$. The corresponding changes from x^* to $\boldsymbol{x}^*$ are also self-explanatory.

S-5.3 Proposal distributions for updating $\boldsymbol{\Sigma}_f$ and $\boldsymbol{\Sigma}_g$

Setting $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma}_f$ and $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Sigma}_g$, we obtain the following inverse Wishart proposal distributions of $\boldsymbol{\Sigma}_f$ and $\boldsymbol{\Sigma}_g$:

$$q_{\boldsymbol{\Sigma}_f}(\boldsymbol{\Sigma}_f) \propto |\boldsymbol{\Sigma}_f|^{-\left(\frac{\nu_f+p+1+T+m}{2}\right)}\exp\Big[-\frac{1}{2}tr\Big\{\boldsymbol{\Sigma}_f^{-1}\Big(\boldsymbol{\Sigma}_{f,0} + \psi^{-1}(\boldsymbol{B}_f - \boldsymbol{B}_{f,0})'\boldsymbol{\Sigma}^{-1}_{B_f,0}(\boldsymbol{B}_f - \boldsymbol{B}_{f,0}) + (\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{B}_f)'(\boldsymbol{A}_{f,D_T} + \boldsymbol{I}_T)^{-1}(\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{B}_f)\Big)\Big\}\Big] \tag{74}$$

and

$$q_{\boldsymbol{\Sigma}_g}(\boldsymbol{\Sigma}_g) \propto |\boldsymbol{\Sigma}_g|^{-\left(\frac{\nu_g+q+2+m+n+T}{2}\right)} \exp\left[-\frac{1}{2} tr \boldsymbol{\Sigma}_g^{-1}\left\{\boldsymbol{\Sigma}_{g,0} + \psi^{-1}(\boldsymbol{B}_g - \boldsymbol{B}_{g,0})'\boldsymbol{\Sigma}_{B_g,0}^{-1}(\boldsymbol{B}_g - \boldsymbol{B}_{g,0})\right.\right.$$
$$+(\boldsymbol{x}_1 - \boldsymbol{g}(\boldsymbol{x}_{1,0}^*))(\boldsymbol{x}_1 - \boldsymbol{g}(\boldsymbol{x}_{1,0}^*))'$$
$$+\sum_{t=1}^{T}\frac{1}{\sigma_{g,t}^2}\left\{\boldsymbol{x}_{t+1} - \boldsymbol{B}_g'\boldsymbol{h}(\boldsymbol{x}_{t+1,t}^*) - \left(\boldsymbol{D}_n^* - \boldsymbol{H}_{D_n^*}\boldsymbol{B}_g\right)'\boldsymbol{A}_{g,D_n^*}^{-1}\boldsymbol{s}_{g,D_n^*}(\boldsymbol{x}_{t+1,t}^*)\right\}$$
$$\times\left\{\boldsymbol{x}_{t+1} - \boldsymbol{B}_g'\boldsymbol{h}(\boldsymbol{x}_{t+1,t}^*) - \left(\boldsymbol{D}_n^* - \boldsymbol{H}_{D_n^*}\boldsymbol{B}_g\right)'\boldsymbol{A}_{g,D_n^*}^{-1}\boldsymbol{s}_{g,D_n^*}(\boldsymbol{x}_{t+1,t}^*)\right\}'$$
$$\left.\left.+\left(\boldsymbol{D}_{n+1}^* - \boldsymbol{H}_{D_{n+1}^*}\boldsymbol{B}_g\right)'\boldsymbol{A}_{g,D_{n+1}^*}^{-1}\left(\boldsymbol{D}_{n+1}^* - \boldsymbol{H}_{D_{n+1}^*}\boldsymbol{B}_g\right)\right\}\right] \tag{75}$$

S-5.4 Proposal distributions for updating $\boldsymbol{\Sigma}_\epsilon$, $\boldsymbol{\Sigma}_\eta$, $\boldsymbol{R}_f$ and $\boldsymbol{R}_g$

As in the univariate case, here we set $\boldsymbol{\Sigma}_f = \boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}_\eta$. Then the proposal distributions of $\boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{\Sigma}_\eta$ are given by the following:

$$q_{\boldsymbol{\Sigma}_\epsilon}(\boldsymbol{\Sigma}_\epsilon) \propto |\boldsymbol{\Sigma}_\epsilon|^{-\left(\frac{\nu_\epsilon+p+1+T}{2}\right)}$$
$$\times \exp\left[-\frac{1}{2} tr \boldsymbol{\Sigma}_\epsilon^{-1}\left\{\boldsymbol{\Sigma}_{\epsilon,0} + (\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{B}_f)'(\boldsymbol{A}_{f,D_T} + \boldsymbol{I}_T)^{-1}(\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{B}_f)\right\}\right] \tag{76}$$

$$q_{\boldsymbol{\Sigma}_\eta}(\boldsymbol{\Sigma}_\eta) \propto |\boldsymbol{\Sigma}_\eta|^{-\left(\frac{\nu_\eta+q+3+n+T}{2}\right)} \exp\left[-\frac{1}{2} tr \boldsymbol{\Sigma}_\eta^{-1}\left\{\boldsymbol{\Sigma}_{\eta,0} + (\boldsymbol{x}_1 - \boldsymbol{g}(\boldsymbol{x}_{1,0}^*))(\boldsymbol{x}_1 - \boldsymbol{h}(\boldsymbol{x}_{1,0}^*))'\right.\right.$$
$$+\left(\boldsymbol{D}_{n+1}^* - \boldsymbol{H}_{D_{n+1}^*}\boldsymbol{B}_g\right)'\boldsymbol{A}_{g,D_{n+1}^*}^{-1}\left(\boldsymbol{D}_{n+1}^* - \boldsymbol{H}_{D_{n+1}^*}\boldsymbol{B}_g\right)$$
$$+\sum_{t=1}^{T}\frac{1}{\sigma_{g,t}^2}\left\{\boldsymbol{x}_{t+1} - \boldsymbol{B}_g'\boldsymbol{h}(\boldsymbol{x}_{t+1,t}^*) - \left(\boldsymbol{D}_n^* - \boldsymbol{H}_{D_n^*}\boldsymbol{B}_g\right)'\boldsymbol{A}_{g,D_n^*}^{-1}\boldsymbol{s}_{g,D_n^*}(\boldsymbol{x}_{t+1,t}^*)\right\}$$
$$\left.\left.\times\left\{\boldsymbol{x}_{t+1} - \boldsymbol{B}_g'\boldsymbol{h}(\boldsymbol{x}_{t+1,t}^*) - \left(\boldsymbol{D}_n^* - \boldsymbol{H}_{D_n^*}\boldsymbol{B}_g\right)'\boldsymbol{A}_{g,D_n^*}^{-1}\boldsymbol{s}_{g,D_n^*}(\boldsymbol{x}_{t+1,t}^*)\right\}'\right\}\right] \tag{77}$$

As in the case of the corresponding univariate proposals, in (77) the factor

$$|\boldsymbol{\Sigma}_\eta|^{-\left(\frac{n}{2}\right)} \exp\left[-\frac{1}{2} tr \boldsymbol{\Sigma}_\eta^{-1}\left(\boldsymbol{D}_{n+1}^* - \boldsymbol{H}_{D_{n+1}^*}\boldsymbol{B}_g\right)'\boldsymbol{A}_{g,D_{n+1}^*}^{-1}\left(\boldsymbol{D}_{n+1}^* - \boldsymbol{H}_{D_{n+1}^*}\boldsymbol{B}_g\right)\right]$$

occurs because under $\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}_\eta$, $[\boldsymbol{D}_{n+1}^* \mid \boldsymbol{B}_g, \boldsymbol{\Sigma}_g, \boldsymbol{R}_g] \sim \mathcal{N}_{n,q}\left(\boldsymbol{H}_{D_{n+1}^*}\boldsymbol{B}_g, \boldsymbol{A}_{g,D_{n+1}^*}, \boldsymbol{\Sigma}_\eta\right)$.

The full conditionals of the smoothness parameters in $\boldsymbol{R}_f$ and $\boldsymbol{R}_g$ are not available in closed forms, and as before we suggest Metropolis-Hastings steps with normal random walk proposals with adequately optimized variances.

S-5.5 Proposal distribution for updating $\boldsymbol{g}(\boldsymbol{x}^*_{1,0})$

Assuming $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Sigma}_g$, the full conditional of $\boldsymbol{g}(\boldsymbol{x}^*_{1,0})$ is q-variate normal with mean and variance given, respectively, by

$$\begin{aligned}\boldsymbol{\mu}^*_g = E[\boldsymbol{g}(\boldsymbol{x}^*_{1,0}) \mid \cdots] = \{2 + \boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{1,0})'\boldsymbol{\Sigma}^{-1}_{g,D^*_n}\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{1,0})\}^{-1}\boldsymbol{\Sigma}_g \\ \times \left\{\boldsymbol{x}_1 + \boldsymbol{B}'_g\boldsymbol{h}(\boldsymbol{x}^*_{1,0}) + \boldsymbol{D}^{*\prime}\boldsymbol{\Sigma}^{-1}_{g,D^*_n}\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{1,0})\right\}\end{aligned} \tag{78}$$

and

$$\boldsymbol{\Sigma}^*_g = V[\boldsymbol{g}(\boldsymbol{x}^*_{1,0}) \mid \cdots] = \{2 + \boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{1,0})'\boldsymbol{\Sigma}^{-1}_{g,D^*_n}\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{1,0})\}^{-1}\boldsymbol{\Sigma}_g \tag{79}$$

In the above, $\boldsymbol{\Sigma}_{g,D^*_n}$ is given by (59) of GMRB, and $\boldsymbol{D}^*_z$ is given by

$$\boldsymbol{D}^*_z = \boldsymbol{D}^*_n - \boldsymbol{H}_{D^*_n}\boldsymbol{B}_g + \boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{1,0})\boldsymbol{h}(\boldsymbol{x}^*_{1,0})'\boldsymbol{B}_g \tag{80}$$

We consider $q_{g(x^*_{1,0})}(\boldsymbol{g}(\boldsymbol{x}^*_{1,0})) \equiv N_q\left(\boldsymbol{g}(\boldsymbol{x}^*_{1,0}) : \boldsymbol{\mu}^*_g, \boldsymbol{\Sigma}^*_g\right)$ as the proposal distribution for updating $\boldsymbol{g}(\boldsymbol{x}^*_{1,0})$.

S-5.6 Proposal distribution for updating $\boldsymbol{D}^*_n$

The full conditional distribution of $\boldsymbol{D}^*_n$ in our dynamic model after setting $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Sigma}_g$, is matrix-normal:

$$[\boldsymbol{D}^*_n] \sim \mathcal{N}_{n,q}\left(\boldsymbol{\mu}^*_{D^*_n}, \boldsymbol{\Sigma}^*_{D^*_n}, \boldsymbol{\Sigma}_g\right), \tag{81}$$

where

$$\boldsymbol{\mu}^*_{D^*_n} = \left\{ \boldsymbol{\Sigma}^{-1}_{g,D^*_n} + \boldsymbol{A}^{-1}_{g,D^*_n} \left(\sum_{t=1}^{T} \frac{\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t}) \boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t})'}{\sigma^2_{g,t}} \right) \boldsymbol{A}^{-1}_{g,D^*_n} \right\}^{-1}$$
$$\times \left\{ \boldsymbol{\Sigma}^{-1}_{g,D^*_n} \boldsymbol{\mu}_{g,D^*_n} + \boldsymbol{A}^{-1}_{g,D^*_n} \sum_{t=1}^{T} \frac{\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t}) \{ \boldsymbol{x}'_{t+1} - (\boldsymbol{h}(\boldsymbol{x}^*_{t+1,t})' - \boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t})' \boldsymbol{A}^{-1}_{g,D^*_n} \boldsymbol{H}_{D^*_n}) \boldsymbol{B}_g \}}{\sigma^2_{g,t}} \right\} \tag{82}$$

and

$$\boldsymbol{\Sigma}^*_{D^*_n} = \left\{ \boldsymbol{\Sigma}^{-1}_{g,D^*_n} + \boldsymbol{A}^{-1}_{g,D^*_n} \left(\sum_{t=1}^{T} \frac{\boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t}) \boldsymbol{s}_{g,D^*_n}(\boldsymbol{x}^*_{t+1,t})'}{\sigma^2_{g,t}} \right) \boldsymbol{A}^{-1}_{g,D^*_n} \right\}^{-1} \tag{83}$$

In the above, $\boldsymbol{\mu}_{g,D^*_n}$ and $\boldsymbol{\Sigma}_{g,D^*_n}$ are given by (58) and (59) of GMRB, respectively. The distribution (81) will be used as proposal distribution to update $\boldsymbol{D}^*_n$ using Metropolis-Hastings step.

S-5.7 MH proposal for updating $\boldsymbol{x}_0$

For $j = 1, \ldots, p$, let $\boldsymbol{\beta}_{j,f} = (\boldsymbol{\beta}'_{j,f,1}, \boldsymbol{\beta}'_{j,f,2})'$. Likewise, for $j = 1, \ldots, q$, let $\boldsymbol{\beta}_{j,g} = (\boldsymbol{\beta}'_{j,g,1}, \boldsymbol{\beta}'_{j,g,2})'$. Here $\boldsymbol{\beta}_{j,f,1}$ and $\boldsymbol{\beta}_{j,g,1}$ are two-dimensional and $\boldsymbol{\beta}_{j,f,2}$ and $\boldsymbol{\beta}_{j,g,2}$ are q-dimensional vectors. Also let $\boldsymbol{a}_f = (\boldsymbol{\beta}_{1,f,1}, \ldots, \boldsymbol{\beta}_{p,f,1})'$ and $\boldsymbol{b}_f = (\boldsymbol{\beta}_{1,f,2}, \ldots, \boldsymbol{\beta}_{p,f,2})'$. Similarly, $\boldsymbol{a}_g = (\boldsymbol{\beta}_{1,g,1}, \ldots, \boldsymbol{\beta}_{q,g,1})'$ and $\boldsymbol{b}_g = (\boldsymbol{\beta}_{1,g,2}, \ldots, \boldsymbol{\beta}_{q,g,2})'$. Thus, $\boldsymbol{a}_f$ and $\boldsymbol{a}_g$ are $p \times 2$ and $q \times 2$-dimensional matrices respectively, while $\boldsymbol{b}_f$ and $\boldsymbol{b}_g$ are, respectively, $p \times q$ and $q \times q$ dimensional matrices. Then, $\boldsymbol{B}'_f \boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{a}_f \boldsymbol{k}_1 + \boldsymbol{b}_f \boldsymbol{x}$ and $\boldsymbol{B}'_g \boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{a}_g \boldsymbol{k}_1 + \boldsymbol{b}_g \boldsymbol{x}$, for any q-dimensional $\boldsymbol{x}$. With these representations, and setting $\boldsymbol{\Sigma}_g = 0$, the full conditional distribution of $\boldsymbol{x}_0$ is q-variate normal with mean and variance given, respectively, by

$$E[\boldsymbol{x}_0 \mid \cdots] = \left\{ \boldsymbol{\Sigma}^{-1}_{x_0} + \boldsymbol{b}'_g {\boldsymbol{\Sigma}_\eta}^{-1} \boldsymbol{b}_g \right\}^{-1} \left\{ \boldsymbol{\Sigma}^{-1}_{x_0} \boldsymbol{\mu}_{x_0} + \boldsymbol{b}'_g {\boldsymbol{\Sigma}_\eta}^{-1} (\boldsymbol{x}_1 - \boldsymbol{a}_g \boldsymbol{k}_1) \right\} \tag{84}$$

$$V[\boldsymbol{x}_0 \mid \cdots] = \left\{ \boldsymbol{\Sigma}^{-1}_{x_0} + \boldsymbol{b}'_g {\boldsymbol{\Sigma}_\eta}^{-1} \boldsymbol{b}_g \right\}^{-1}. \tag{85}$$

S-5.8 MH proposals for $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$

Ignoring the errors of the functions $\boldsymbol{f}$ and $\boldsymbol{g}$, that is, setting $\boldsymbol{\Sigma}_f = \boldsymbol{\Sigma}_g = 0$, it turns out that the proposal distribution of $\boldsymbol{x}_{t+1}$, for $t = 0, \ldots, T-1$, can be taken as a q-variate normal distribution

with mean and variance given, respectively, by

$$E[\boldsymbol{x}_{t+1} \mid \cdots] = \left(\boldsymbol{\Sigma}_\eta^{-1} + \boldsymbol{b}_g'\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{b}_g' + \boldsymbol{b}_f'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{b}_f\right)^{-1} \times \left\{\boldsymbol{\Sigma}_\eta^{-1}\left(\boldsymbol{a}_g\boldsymbol{k}_{t+1} + \boldsymbol{b}_g\boldsymbol{x}_t\right) + \boldsymbol{b}_g'\boldsymbol{\Sigma}_\eta^{-1}\left(\boldsymbol{x}_{t+2} - \boldsymbol{a}_g\boldsymbol{k}_{t+2}\right) + \boldsymbol{b}_f'\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{y}_{t+1} - \boldsymbol{a}_f\boldsymbol{k}_{t+1})\right\} \tag{86}$$

$$V[\boldsymbol{x}_{t+1} \mid \cdots] = \left(\boldsymbol{\Sigma}_\eta^{-1} + \boldsymbol{b}_g'\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{b}_g' + \boldsymbol{b}_f'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{b}_f\right)^{-1}. \tag{87}$$

S-5.9 Gibbs step for updating $\boldsymbol{x}_{T+1}$

The full conditional distribution of $\boldsymbol{x}_{T+1}$ is q-variate normal with mean

$$E[\boldsymbol{x}_{T+1} \mid \cdots] = \boldsymbol{B}_g'\boldsymbol{h}(\boldsymbol{x}^*_{T+1,T}) + (\boldsymbol{D}^*_n - \boldsymbol{H}_{D^*_n}\boldsymbol{B}_g)'\boldsymbol{A}^{-1}_{g,D^*_n}\boldsymbol{s}_{g,\boldsymbol{D}^*_n}(\boldsymbol{x}^*_{T+1,T}) \tag{88}$$

and variance

$$V[\boldsymbol{x}_{T+1} \mid \cdots] = \boldsymbol{\Sigma}_\eta + \boldsymbol{\Sigma}_g\left\{1 - \boldsymbol{s}_{g,\boldsymbol{D}^*_n}(\boldsymbol{x}^*_{T+1,T})'\boldsymbol{A}^{-1}_{g,\boldsymbol{D}^*_n}\boldsymbol{s}_{g,\boldsymbol{D}^*_n}(\boldsymbol{x}^*_{T+1,T})\right\}. \tag{89}$$

The caveat for using these proposal distributions is that if the underlying assumptions $\Sigma_f = \Sigma_\epsilon$, $\Sigma_g = \Sigma_\eta$ and $\Sigma_g = \mathbf{0}$ do not hold, then the proposal mechanisms will turn out to be much less effective, and more so compared to the univariate cases, due to the curse of dimensionality. So, we shall also consider other appproaches to these updating procedures, to be discussed in the context of the specific applications.

S-6. SIMULATION STUDY: MULTIVARIATE CASE

We consider a simulation study where we generate the data from a 4-variate model, constructed using the univariate nonstationary growth model of Carlin et al. (1992). To reduce computational burden in this multi-dimensional set up we generated 50 data points using the following model:

for $t = 1, \ldots, 50$,

$$x_{t,1} = \alpha x_{t-1,1} + \beta x_{t-1,1}/(1 + x_{t-1,1}^2) + \gamma \cos(1.2(t-1)) + \eta_{t,1} \tag{90}$$

$$x_{t,2} = \alpha x_{t-1,2} + \beta x_{t-1,2}/(1 + x_{t-1,2}^2) + \eta_{t,2} \tag{91}$$

$$x_{t,3} = \alpha + \beta x_{t-1,3} + \eta_{t,3} \tag{92}$$

$$x_{t,4} = \gamma \cos(1.2(t-1)) + \eta_{t,4} \tag{93}$$

$$y_{t,i} = x_{t,i}^2/20 + \epsilon_{t,i}; \;\; i = 1, \ldots, 4, \tag{94}$$

where $\epsilon_{t,i}$ and $\eta_{t,i}$ are iid zero-mean normal random variables with variances 0.1. We set $\alpha = 0.05$, $\beta = 0.1$ and $\gamma = 0.2$, and $\boldsymbol{x}_0 = \mathbf{0}$.

S-6.1 Choice of prior parameters

For the priors of $\boldsymbol{B}_f$ and $\boldsymbol{B}_g$, we set $\psi = 1$, $\boldsymbol{B}_{f,0}$ and $\boldsymbol{B}_{g,0}$ to be null matrices, and $\boldsymbol{\Sigma}_{B_f,0}$ and $\boldsymbol{\Sigma}_{B_g,0}$ to be identity matrices. For the priors of $\boldsymbol{\Sigma}_\epsilon$, $\boldsymbol{\Sigma}_\eta$, $\boldsymbol{\Sigma}_f$, and $\boldsymbol{\Sigma}_g$ we set $\nu_\epsilon = p$, $\nu_\eta = q$, $\nu_f = p$, and $\nu_g = q$. We chose $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma}_\eta = 0.1\boldsymbol{I}$, and $\boldsymbol{\Sigma}_f = \boldsymbol{\Sigma}_g = 0.5\boldsymbol{I}$, where $\boldsymbol{I}$ denotes the identity matrix. For the log-normal priors of the smoothness parameters we set $\mu_{R_f} = \mu_{R_g} = -0.5$ and $\sigma_{R_f}^2 = \sigma_{R_g}^2 = 1$. The choices imply that the prior mean and the prior variance of each of the smoothness parameters are, respectively, 1 and 2 (approximately). In our simulation experiment these choices seemed adequate.

S-6.2 MCMC implementation

For updating the smoothness parameters we used normal random walk proposals with variances 0.005. To set up the grid $\boldsymbol{G}_z$ for the model-fitting purpose, we considered $[-30, 30]^4$ to be a grid space for the 4-dimensional variable $\boldsymbol{x}$. Indeed, this grid-space is much larger than, and contains the region where the entire true 4-dimensional time series lie. We divide $[-30, 30]$ into 100 equal sub-intervals and chose a point randomly from each of 100 sub-intervals, in each dimension, yielding $n = 100$ 4-dimensional points corresponding to $\boldsymbol{x}$. As before, we chose the first component of the grid (corresponding to the time component) by uniformly drawing from each subinterval $[i, i+1]$; $i = 0, \ldots, 99$.

The block updating proposal for updating $\boldsymbol{B}_f$, described in Section S-2.2 worked quite well, but those for block updating $\boldsymbol{B}_g$, $\boldsymbol{g}(\boldsymbol{x}^*_{1,0})$, $\boldsymbol{D}^*_n$, and those for the covariance matrices, as decribed in Section S-2 yielded poor acceptance rates. In order to update the covariance matrices $\boldsymbol{\Sigma}_\epsilon$, $\boldsymbol{\Sigma}_\eta$, $\boldsymbol{\Sigma}_g$ and $\boldsymbol{\Sigma}_f$, we considered the following strategy: we re-wrote the matrices in the form $\boldsymbol{CC}'$, where $\boldsymbol{C}$ is a lower triangular matrix, and used normal random walk with variance 0.005, to update the non-zero elements in a single block. This improved the acceptance rates. For $\boldsymbol{B}_g$, $\boldsymbol{g}(\boldsymbol{x}^*_{1,0})$ and $\boldsymbol{D}^*_n$, the strategy of block updating using random walk failed. As a remedy we considered the transformation-based MCMC (TMCMC), recently developed by Dutta & Bhattacharya (2013); in particular, we used the additive transformation, which requires much less number of move types and hence computationally less expensive compared to other, non-additive move types. Briefly, for each of the blocks, we generated $\xi \sim N(0, 0.05)I(\xi > 0)$. Then, for each parameter in the block, we either added or subtracted ξ with equal probability. In other words, we used the same ξ to update all the parameters in a block, unlike the block random walk proposal. This considerably improved the acceptance rates. For theoretical and implementation details, see Dutta & Bhattacharya (2013).

We discarded the first 10,000 iterations of the MCMC run as burn-in and stored the subsequent 50,000 iterations for inference. Convergence is assessed by informal diagnostics, as before. It took an ordinary laptop about 24 hours to implement this multivariate experiment.

S-6.3 Results of model-fitting

Figures S-7 and S-8 show that our Gaussian process-based nonparametric model performs well in spite of the multidimensional situation—the true values are well-captured by the posterior distributions, and the true forecast values are also well-supported by the corresponding forecast distributions of $\boldsymbol{x}_{51}$ and $\boldsymbol{y}_{51}$. Moreover, each component of the true 4-variate time series fall entirely within the 95% highest posterior densities of the corresponding component of the 4-variate posterior time series.

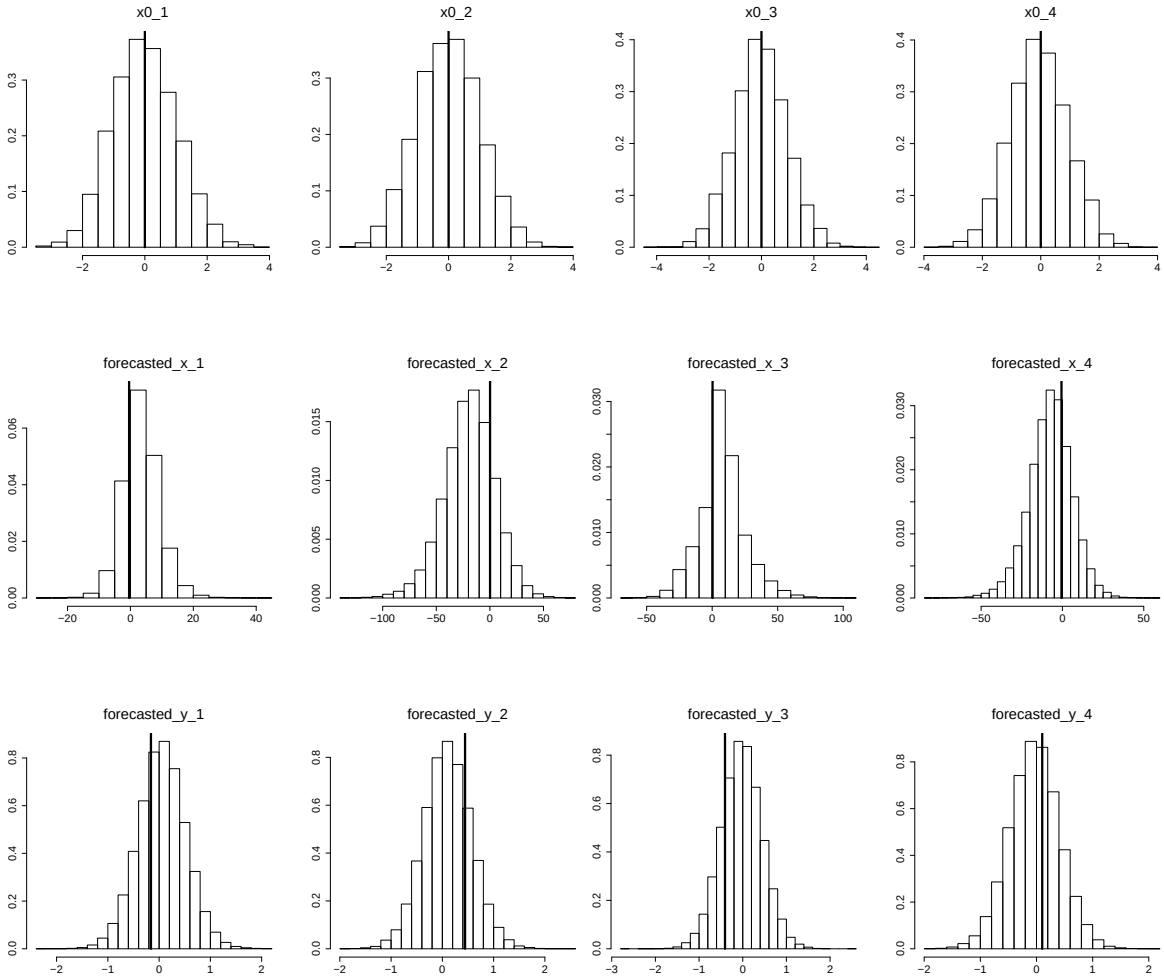

Figure S-7: **Simulation study with data generated from the 4-variate modification of the model of CPS**: Posterior distributions of $\boldsymbol{x}_0$, $\boldsymbol{x}_{T+1}$ (one-step forecasted $\boldsymbol{x}$), and $\boldsymbol{y}_{T+1}$ (one-step forecasted $\boldsymbol{y}$). The solid line stands for the true values of the respective parameters.

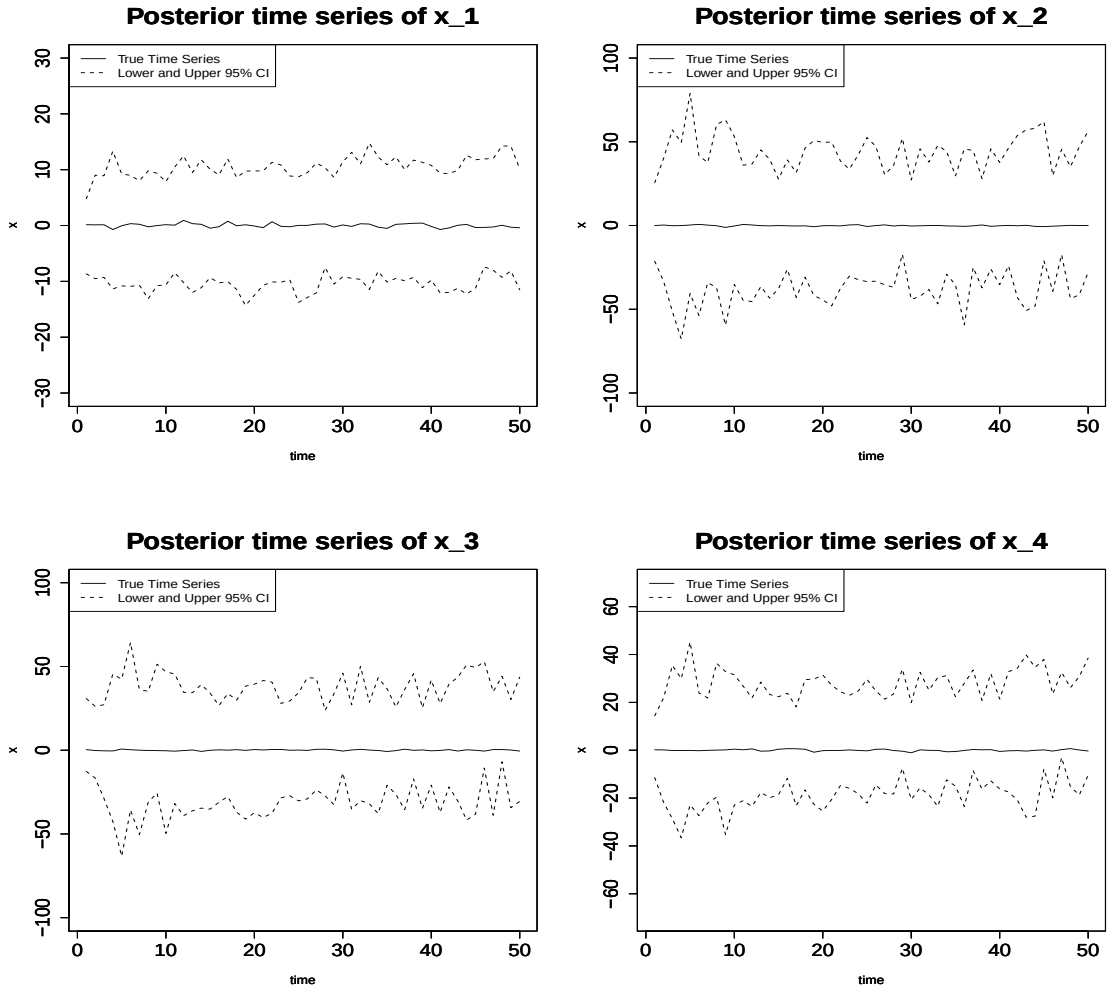

Figure S-8: **Simulation study with data generated from the 4-variate modification of the model of CPS**: 95% highest posterior density credible intervals of the time series $x_{1,j}, \ldots, x_{T,j}$; $j = 1, 2, 3, 4$. The solid line stands for the true time series.

S-7. POSTERIORS OF THE FUNCTION COMPOSITIONS AND COMPARISONS WITH THE TRUE FUNCTION COMPOSITIONS

It may be of interest to check how the posterior distributions of the function compositions at different time points compare with the true function compositions generating the data. To clarify, for any x, let $g_0^*(x) = x$, and let $g_t^*(x) = g(t, g_{t-1}^*(x))$ for $t \geq 1$ be the t-step composition of $g(\cdot, \cdot)$. It may be of interest to compare the posterior of $g_t^*(\cdot)$ with the true t-step functional composition of the data-generating evolutionary equation. Similarly, for $t \geq 1$ it may be of interest to compare the true t-step composition of the observational equation with the posterior of $f_t^*(\cdot) = f(t, g_t^*(\cdot))$.

Although technically the comparisons are possible, there is a subtle issue that needs to be understood. It is important to note that the observed data has been generated by a single latent time series, guided by the evolutionary equation, beginning at a single value x_0. Since we consider x_0 unknown, we attempt to learn about this unknown through the posterior distribution of x_0. Given the observed data $\boldsymbol{D}_T$, the support of the posterior distribution contains the likely values of x_0 that might have given rise to the observed data via the compositions of the observational and the evolutionary equations. But beyond the support of the posterior of x_0 there is no imformation about the composite functions. Hence, it is reasonable to compare the true function compositions with $g_t^*(x)$ and $f_t^*(x)$ when x belongs to the support of the posterior distribution of x_0, but not if x is not in the support of the posterior of x_0. Since the support of the posterior of x_0 is not likely to be large if the data and the model contain sufficient information, the composite functions are likely to be close to linear on that relatively small region. Below we provide details of computing the posteriors of $g_t^*(\cdot)$ and $f_t^*(\cdot)$.

We note that it follows on similar lines as in (16) and (17) of GMRB, that for $t \geq 1$, $[g_t^*(x) \mid \boldsymbol{D}_n^*, g_{t-1}^*(x), \boldsymbol{\theta}_g]$ is normal with mean

$$\mu_{g,t}(x) = \boldsymbol{h}(t, g_{t-1}^*(x))'\boldsymbol{\beta}_g + \boldsymbol{s}_{g,\boldsymbol{D}_n^*}(t, g_{t-1}^*(x))'\boldsymbol{A}_{g,D_n^*}^{-1}(\boldsymbol{D}_n^* - \boldsymbol{H}_{D_n^*}\boldsymbol{\beta}_g)$$

and variance

$$\sigma_{g,t}^2(x) = \sigma_g^2 \left\{1 - \boldsymbol{s}_{g,\boldsymbol{D}_n^*}(t, g_{t-1}^*(x))'\boldsymbol{A}_{g,\boldsymbol{D}_n^*}^{-1}\boldsymbol{s}_{g,\boldsymbol{D}_n^*}(t, g_{t-1}^*(x))\right\}.$$

Noting that the posterior of $g_t^*(x)$ can be expressed (using conditional independence) as

$$[g_t^*(x) \mid \boldsymbol{D}_T] = \int [g_t^*(x) \mid \boldsymbol{D}_n^*, g_{t-1}^*(x), \boldsymbol{\theta}_g][\boldsymbol{D}_n^*, g_{t-1}^*(x), \boldsymbol{\theta}_g \mid \boldsymbol{D}_T] d\boldsymbol{D}_n^* dg_{t-1}^*(x) d\boldsymbol{\theta}_g,$$

given the available posterior samples of $\boldsymbol{D}_n^*, g_{t-1}^*(x), \boldsymbol{\theta}_g$, draws from $[g_t^*(x) \mid \boldsymbol{D}_n^*, g_{t-1}^*(x), \boldsymbol{\theta}_g]$ yields posterior samples from $[g_t^*(x) \mid \boldsymbol{D}_T]$. Thus, for any x, we can compute the posterior distribution of $g_t^*(x); t = 1, 2, \ldots$.

Similarly, it follows as in (27) and (28) of GMRB, that for $t \geq 1$, the full conditional of $f_t^*(x) = f(t, g_t^*(x))$, given $g_t^*(x)$ is normal, with mean

$$\mu_{f,t}(x) = \boldsymbol{h}(t, g_t^*(x))'\boldsymbol{\beta}_f + \boldsymbol{s}_{f,\boldsymbol{D}_T}(t, g_t^*(x))'\boldsymbol{A}_{f,D_T}^{-1}(\boldsymbol{D}_T - \boldsymbol{H}_{D_T}\boldsymbol{\beta}_f)$$

and variance

$$\sigma_{f,t}^2(x) = \sigma_f^2 \left\{1 - \boldsymbol{s}_{f,\boldsymbol{D}_T}(t, g_t^*(x))'\boldsymbol{A}_{f,\boldsymbol{D}_T}^{-1}\boldsymbol{s}_{f,\boldsymbol{D}_T}(t, g_t^*(x))\right\}.$$

Following the method discussed for generating posterior samples of $g_t^*(x)$, we can easily generate posterior samples of $f_t^*(x)$.

The procedure in the multivariate situation is analogous; we only modify the notations $f_t^*(x)$ and $g_t^*(x)$ to $f_{jt}^*(x)$ and $g_{kt}^*(x)$ to denote the corresponding observational and evolutionary t-step composite functions for the j-th and the k-th components of the functions espectively, where $j = 1, 2, \ldots, p$ and $k = 1, \ldots, q$.

Figures S-9 and S-10 display the 95% highest posterior density intervals of $f_t^*(x)$ and $g_t^*(x)$ for $t = 1, 2, 3$ and for $x \in [-4, 4]$ in the simulation studies concerning the univariate linear model and the non-linear growth model of CPS respectively. The domain $[-4, 4]$ is chosen because the posterior of x_0 is supported on this interval in both the linear and the non-linear cases (see Figures 2 and 5 of GMRB). As already mentioned, the true functions are almost linear within this short interval. That actually the true functions can be far from linear on a much wider domain is already clear from the observational and the evolutionary equations in the non-linear example; the true function $f_1^*(\cdot)$ in the case of the non-linear example as shown in Figure S-11 is clearly non-linear over a much wider domain even though it is close to linear (indeed, almost constant) on $[-4, 4]$. Figures S-12 and S-13 show the true functions f_{jt}^* and g_{kt}^* for $j, k = 1, 2, 3, 4$ and $t = 1, 2, 3$

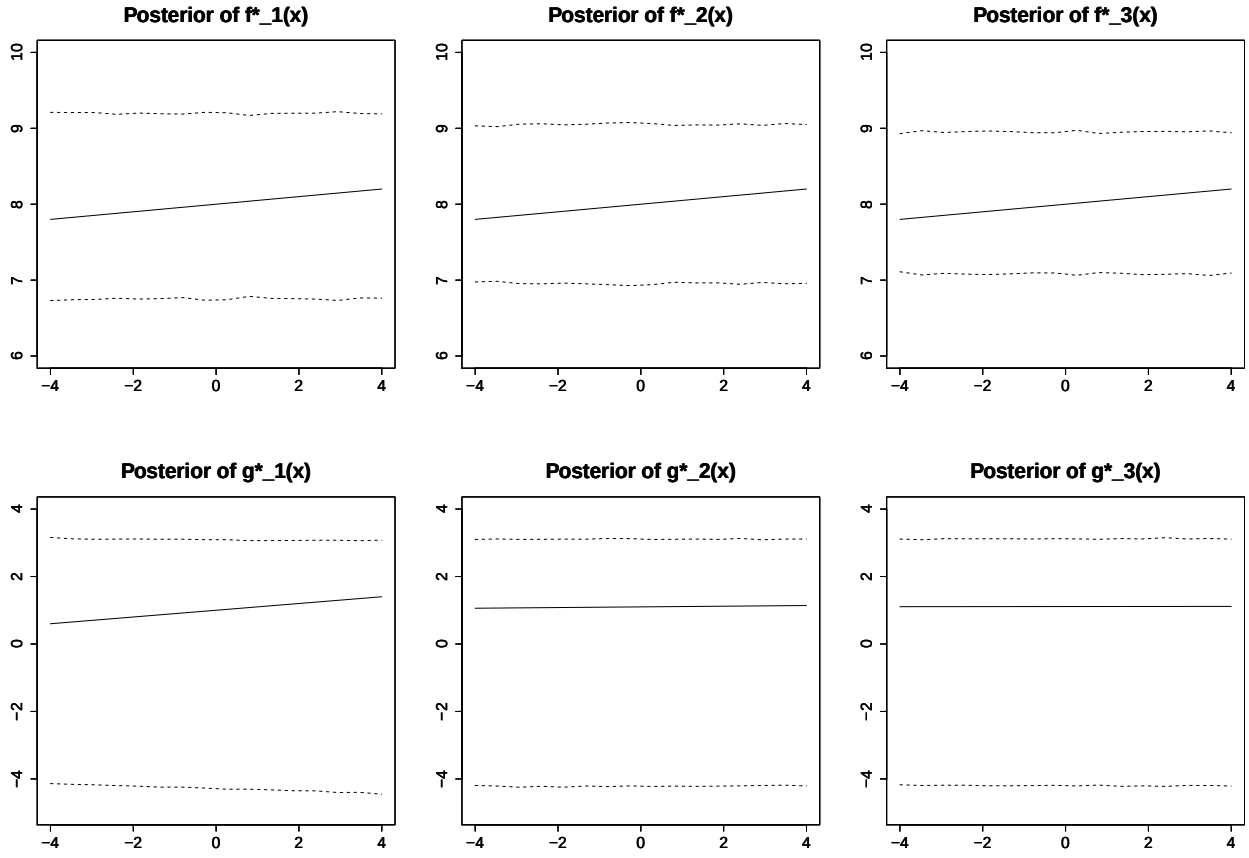

Figure S-9: **Simulation study with data generated from the univariate linear model**: The broken lines denote the 95% credible regions of the posterior distributions of $f_t^*(x)$ and $g_t^*(x)$ for $t = 1, 2, 3$; $x \in [-4, 4]$, which is the support of the posterior of x_0. The solid lines denote the true composite observational and evolutionary functions.

along with the corresponding 95% highest posterior density credible intervals in the multivariate situation. Here also we select the domain of x as $[-4, 4]$ because Figure S-7 shows that *a posteriori* each co-ordinate of x_0 has support $[-4, 4]$.

In all the cases the true, composite, observational and the evolutionary functions fall comfortably within their respective 95% highest posterior density credible intervals when $x \in [-4, 4]$ is considered. We also experimented with $x \in [-30, 30]$, but in this case, as already anticipated, in many cases (not reported here) the true values are excluded from the credible intervals when $x \notin [-4, 4]$, particularly towards the boundaries of the range.

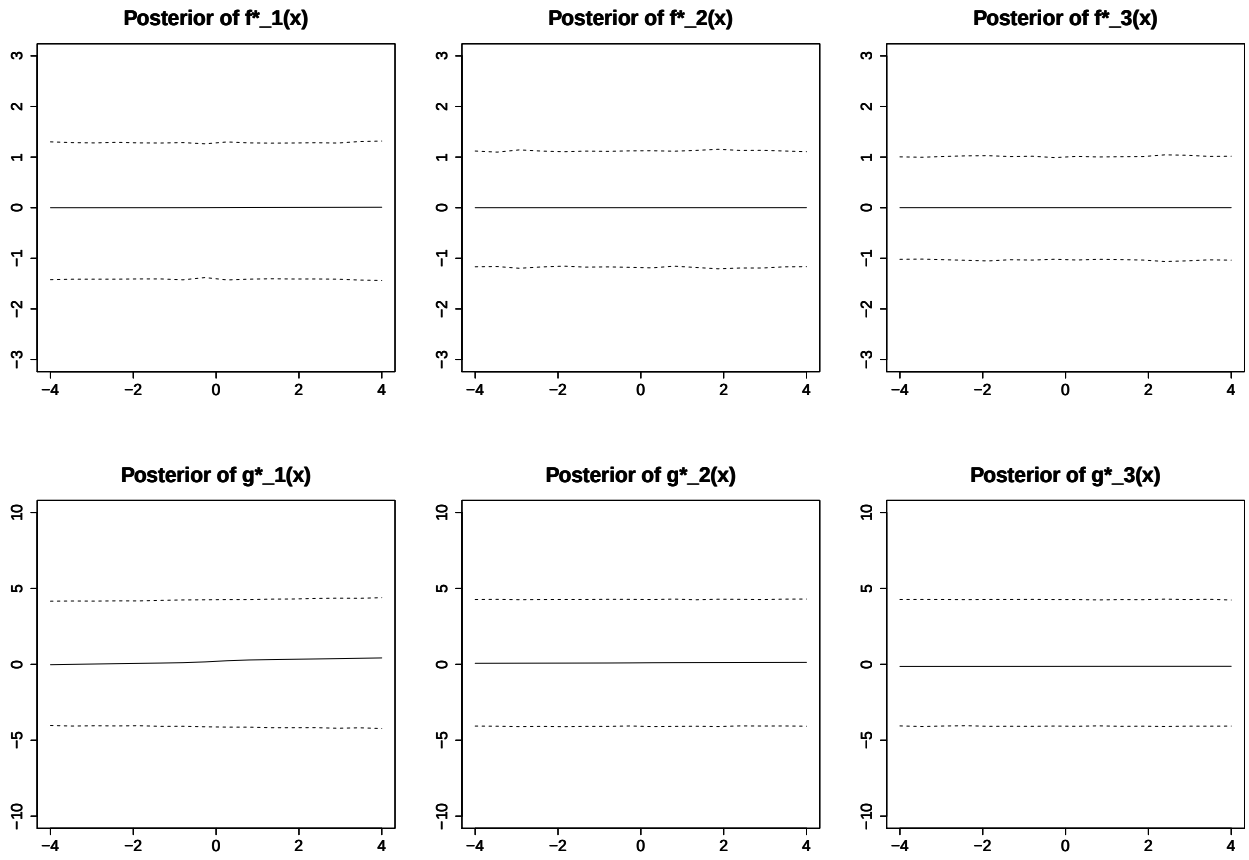

Figure S-10: **Simulation study with data generated from the univariate growth model of CPS**: The broken lines denote the 95% highest posterior density credible regions of the posterior distributions of $f_t^*(x)$ and $g_t^*(x)$ for $t = 1, 2, 3$; $x \in [-4, 4]$, which is the support of the posterior of x_0. The solid lines denote the true composite observational and evolutionary functions.

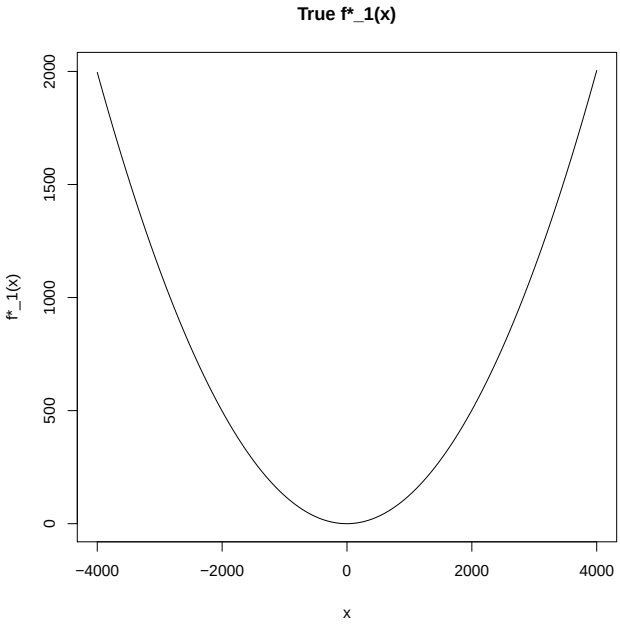

Figure S-11: **Simulation study with data generated from the univariate growth model of CPS**: Displayed is the true function $f_1^*(\cdot)$ on the much wider domain $[-4000, 4000]$.

REFERENCES

Carlin, B. P., Polson, N. G., & Stoffer, D. S. (1992), "A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling," *Journal of the American Statistical Association*, 87, 493–500.

Dutta, S., & Bhattacharya, S. (2013), "Markov Chain Monte Carlo Based on Deterministic Transformations," *Statistical Methodology (to appear)*, . Available at arxiv:1106.5850v3 with supplementary section in arxiv.org/pdf/1306.6684.

Ghosh, A., Mukhopadhyay, S., Roy, S., & Bhattacharya, S. (2013), "Bayesian Inference in Nonparametric Dynamic State-Space Models,". Submitted.

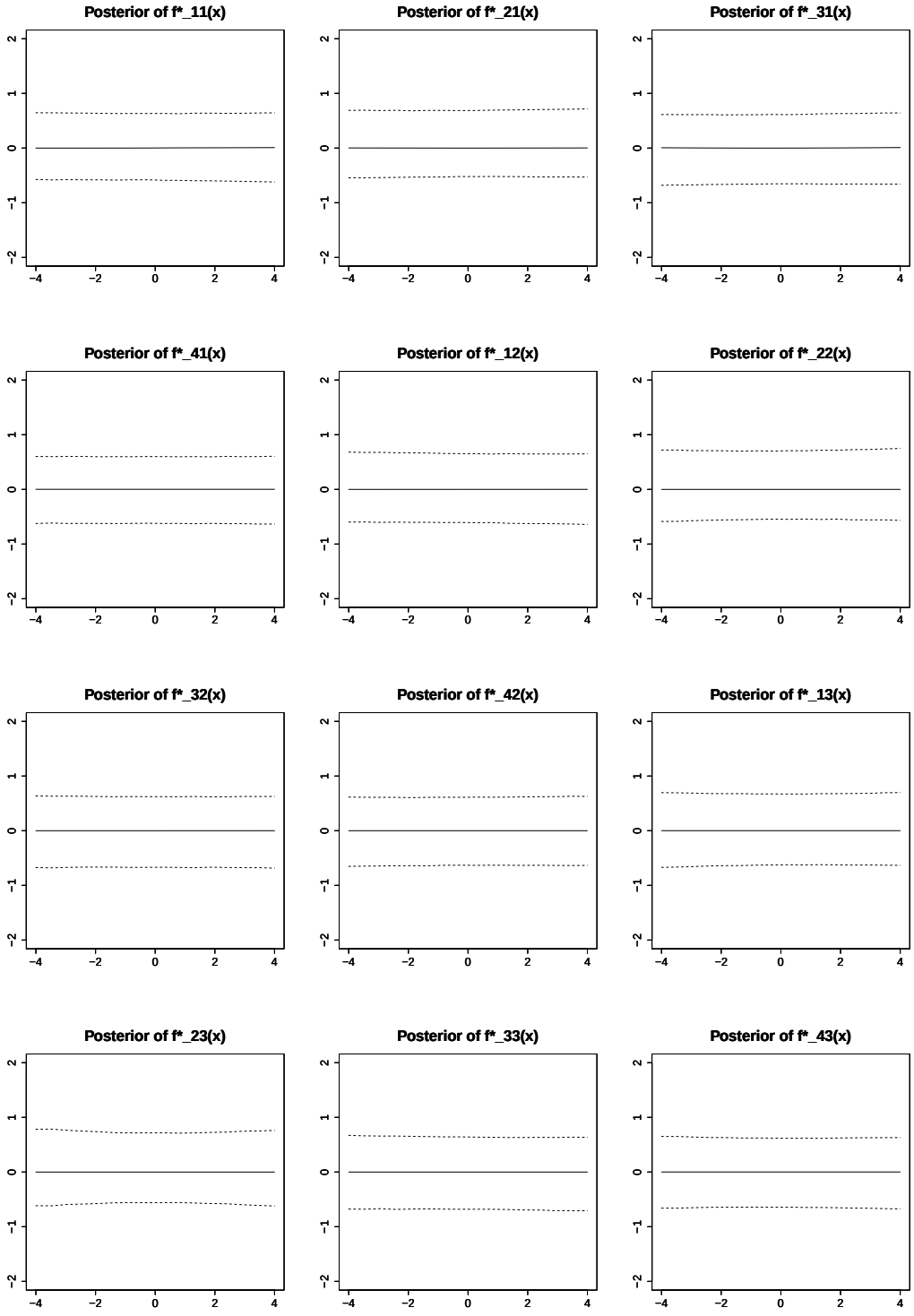

Figure S-12: **Simulation study with data generated from the 4-variate modification of the model of CPS**: The broken lines denote the 95% highest posterior density credible regions of the posterior distributions of $f^*_{jt}(x)$ for $j = 1, 2, 3, 4$ and $t = 1, 2, 3$; $x \in [-4, 4]$, which is the support of the posterior of x_0. The solid lines denote the true composite observational and evolutionary functions.

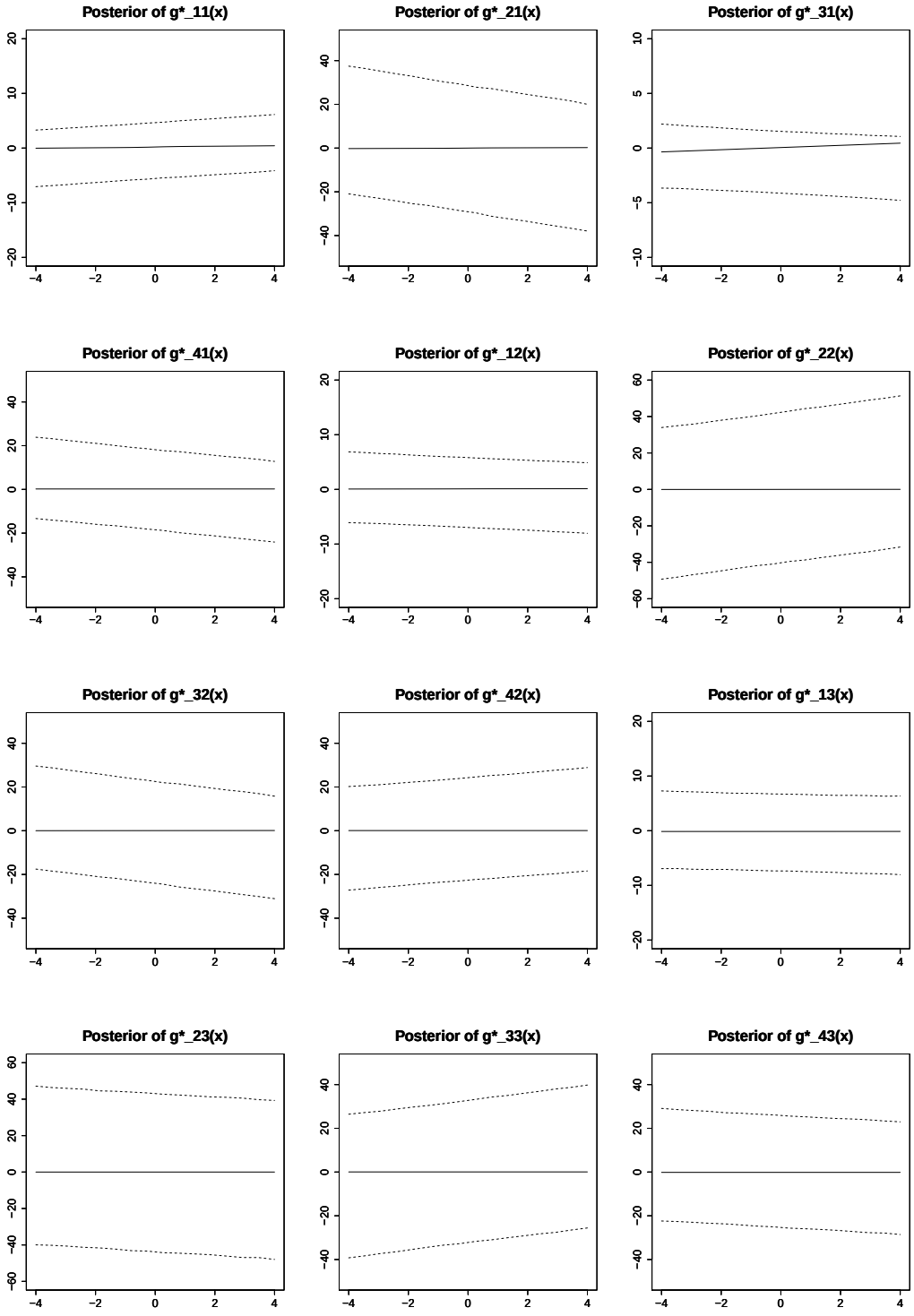

Figure S-13: **Simulation study with data generated from the 4-variate modification of the model of CPS**: The broken lines denote the 95% highest posterior density credible regions of the posterior distributions of $g^*_{jt}(x)$ for $j = 1, 2, 3, 4$ and $t = 1, 2, 3$; $x \in [-4, 4]$, which is the support of the posterior of x_0. The solid lines denote the true composite observational and evolutionary functions.